Incidence des leucémies de l'enfant en fonction de la proximité et des caractéristiques générales de diverses sources d'expositions environnementales

Claire Faure Sermage-Faure

▶ **To cite this version:**

Claire Faure Sermage-Faure. Incidence des leucémies de l'enfant en fonction de la proximité et des caractéristiques générales de diverses sources d'expositions environnementales. Santé publique et épidémiologie. Université Paris Sud - Paris XI, 2012. Français. NNT : 2012PA11T028 . tel-00779494

HAL Id: tel-00779494

https://theses.hal.science/tel-00779494

Submitted on 22 Jan 2013

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.

UNIVERSITÉ PARIS XI
FACULTÉ DE MÉDECINE PARIS-SUD

Année 2012 N°

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS SUD

Spécialité : Épidémiologie

présentée et soutenue publiquement

par

Claire SERMAGE-FAURE

le 21 juin 2012

Incidence des leucémies de l'enfant en fonction de la proximité et des caractéristiques générales de diverses sources d'expositions environnementales

Directeur de thèse : Monsieur Denis HÉMON

JURY

Monsieur le Professeur Jean-Christophe THALABARD	Président
Monsieur Marc COLONNA	Rapporteur
Monsieur le Professeur Denis ZMIROU-NAVIER	Rapporteur
Monsieur le Professeur Gilles BRÜCKER	Examinateur
Madame Agnès LEFRANC	Examinateur
Monsieur Denis HÉMON	Directeur de thèse

A Remerciements

Tout d'abord un grand merci à mon directeur de thèse Denis Hémon, pour tout ce que vous m'a appris en épidémiologie et transmis de votre passion pour la recherche ! Ces quelques années ont été très riches professionnellement, scientifiquement et humainement. Merci aussi, comme à Jacqueline Clavel, de votre enthousiasme et de votre disponibilité tout au long de ces années de travail passionnant et très varié sur le programme GEOCAP! Enfin, merci à tous les deux pour votre confiance et de m'avoir permis de continuer ce travail malgré l'éloignement géographique de ces dernières années !

Je voudrais également remercier les différents membres du jury et notamment le Professeur Jean-Christophe Thalabard qui m'a fait l'honneur d'en être le président; le docteur Marc Colonna et le Professeur Denis Zmirou d'avoir été les rapporteurs de ma thèse, et qui m'ont permis grâce à leurs remarques enrichissantes de bien approfondir mon travail; ainsi que Mme Agnès Lefranc et le Professeur Gilles Brücker qui ont acceptés d'être examinateurs.

Un immense merci à toute l'équipe du RNHE avec qui nous avons beaucoup travaillé pour le recueil des adresses et autres informations sur les cas : merci aux enquêteurs qui ont mis beaucoup d'énergie à retrouver un maximum d'adresses, merci aussi aux secrétaires du registre Souad, Brigitte, Soizic et Eva pour votre patience et votre disponibilité lors de mes TRES fréquentes sollicitations et bien sûr un grand merci à Aurélie pour ton aide précieuse dans au long de ces années !
Merci à Annie pour cette collaboration très enrichissante dans notre travail commun sur les modèles de lissage spatial, merci pour tout ce que tu m'as appris! Merci à Stéphanie pour ton aide et ces échanges toujours très intéressants sur tous nos sujets d'études.

Je voudrais également remercier tous nos collaborateurs dans le programme GEOCAP et spécialement Mathieu Carrere, Laetitia Codet, Olivier Lamy et tous les personnes de Géocible qui nous ont beaucoup aidé, écouté dans nos demandes et qui nous ont permis de mener à bien ce projet. Merci à Magda Tomasini et Laurent

Auzet de l'INSEE pour leur collaboration dans les différentes étapes de la construction de l'échantillon témoin. Un grand merci aussi à l'équipe du comité technique de Rte et spécialement Jérôme Bessou et François Deschamps avec qui nous avons travaillé depuis 6 ans : merci pour votre disponibilité et pour tout ce que vous nous avez appris ! Enfin, merci à Dominique Laurier et Michel Chartier qui nous ont été d'une aide précieuse pour tout le travail concernant les centrales nucléaires.

Merci à tous les membres de l'équipe 6 du CESP et avant de l'U754 avec qui cela a toujours été un plaisir de travailler. Merci spécialement à mes collègues de bureaux Florence M, Sara, Jérémie, Alicia et Claire : je garderai un très bon souvenir de tous nos échanges ! Merci aussi à Laurent, Sofien, Guillaume et Frédéric pour leur aide précieuse (!) et leurs dépannages en informatique et à Evelyne et Linda pour leur disponibilité dans toutes les démarches administratives. Et à tous : MERCI pour tous ces bons moments passés ensemble !

MERCI Marie, Anne-C, Malix, Soph, Caro, Anne et Naïs pour votre amitié fidèle et votre soutien de statisticienne (ou non ;-)) pendant cette thèse et merci Hélène pour ta présence et ton aide spécialement ces derniers temps. Eh oui, ça y est, je peux le dire : j'ai fini ma thèse !!

Bien sûr, merci à toute ma famille : mes frères et sœurs, beau-frère et belles-sœurs pour votre soutien et aux grands-parents de Joséphine d'avoir été là pour elle ces derniers temps, votre aide a été plus que précieuse !

Enfin, merci à ma petite Joséphine pour sa patience et sa joie de vivre et à Xavier mon mari, pour son soutien quotidien et sa présence depuis le début!

B Sommaire

A Remerciements 2
B Sommaire 4
C Productions scientifiques issues du travail de thèse 7
C.1 Articles publiés: 7
C.2 Article en préparation 7
C.3 Chapitres d'ouvrages 8
C.4 Communication orale : 8
C.5 Communications affichées : 8
C.6 Recueil de données effectué par la doctorante dans le cadre de ce travail de thèse 9
D Résumé 14
E Abstract 15
F Liste des tableaux 16
G Liste des figures 19
H Liste des abréviations 20
II. Problématique scientifique générale et objectifs 21
A Introduction 22
A.1 Epidémiologie descriptive des leucémies de l'enfant 22
A.2 Facteurs de risque établis 26
A.3 Facteurs de risque suspectés 27
A.4 Facteurs de risques environnementaux 29
B Objectifs de la thèse et approche générale 34
III. Matériel et méthodes 37
A Les cas 38
B Données démographiques 40
B.1 Estimations de population 40
B.2 Données contextuelles 41
C Témoins 43
C.1 Constitution de l'échantillon 43
C.2 Représentativité de l'échantillon témoin 44
D Géolocalisation 49
D.1 Géolocalisation des domiciles 49
D.1.a Principe du géocodage 49
D.1.b Géocodage des adresses GEOCAP 49
D.1.c Précision du géocodage 50
D.1.d Validation du géocodage 54
D.2 Positionnement des domiciles par rapport aux sources 55
D.2.a Positionnement des adresses par rapport aux sources 55
D.2.b Géocodage « au domicile » 58
E Méthodes statistiques générales 63
E.1 Incidence 64

E.2 Cas-témoins 65
IV. Variations d'incidence au niveau départemental 66
A Introduction 67
B Matériel 69
B.1 Registre des cas 69
B.2 Population 69
C Méthodes 70
C.1 Cartographie "brute" 70
C.1.a Tests de signification des risques relatifs 70
C.1.b Test d'hétérogénéité globale des risques relatifs 71
C.2 Cartographie "lissée" 71
C.2.a Modèles d'estimation des risques relatifs 72
C.2.b Lois hyper a priori 75
C.2.c Ajustement des modèles aux observations 77
C.2.d Corrélations entre SIR lissés 79
D Résultats 80
D.1 Cartographie "brute" 80
D.1.a Hétérogénéité des risques relatifs 80
D.1.b Cartes des SIR 82
D.1.c Tests de signification des SIR 82
D.1.d Corrélations des SIR entre types histologiques, périodes, sexe, classes d'âge 85
D.2 Cartographie "lissée" 87
D.2.a Adéquation des modèles de lissage 87
D.2.b Lissage des SIR 88
D.2.c Cartes des SIR lissés 90
D.2.d Corrélations entre SIR lissés 94
E Discussion 95
V. Proximité des centrales nucléaires de production d'électricité et leucémies aigües de l'enfant 100
A Introduction 101
B Matériel et méthodes 103
B.1 Sujets 103
B.1.a Etude cas-témoins 2002-2007 103
B.1.b Etudes d'incidence: 1990-2001, 2002-2007 et 1990-2007 103
B.2 Caractéristiques des CNPE 104
B.3 Distance à la CNPE la plus proche 105
B.4 Dose à la moelle estimée due aux rejets gazeux des CNPE 106
B.5 Analyse statistique 110
B.6 Puissance statistique 112
C Résultats 113
C.1 Distance à la CNPE la plus proche 113
C.1.a Etude cas-témoins 2002-2007 113
C.1.b Etude d'incidence 2002-2007 116

C.1.c Etude d'incidence 1990-2001 118
C.1.d Etude d'incidence 1990-2007 119
C.1.e Analyses de sensibilité, prise en compte de facteurs de confusion potentiels et analyses par sous-groupe 120
C.2 Estimations de doses liées aux rejets gazeux radioactifs des CNPE ... 127
D Discussion 129
VI. Proximité aux lignes à haute tension et leucémies aigües de l'enfant . 134
A Introduction 135
B Matériel et méthodes 138
B.1 Sujets : 138
B.2 Exposition aux LHT 138
B.2.a Données utilisées 138
B.2.b Distance à la ligne la plus proche 142
B.2.c Analyse statistique 143
B.2.d Puissance de l'étude 144
C Résultats 145
C.1.a Distance « automatique » 145
C.1.b Distance « au domicile » 148
C.1.c Analyses de sensibilité 151
D Discussion 160
VII. Conclusions et perspectives 165
A Conclusion 166
B Perspectives 168
VIII. Bibliographie 170
IX. Articles 186

C Productions scientifiques issues du travail de thèse

C.1 Articles publiés:

Faure C, Mollié A**,** Bellec S**,** Guyot-Goubin A**,** Clavel J**,** Hémon D. Geographic variations in childhood acute leukaemia incidence in France 1990-2004. European Journal of Cancer Prevention, 2009 Aug;18(4):267-79.

Amigou A, **Sermage-Faure C**, Orsi L, Leverger G, Baruchel A, Bertrand Y, Nelken B, Robert A, Michel G, Margueritte G, Perel Y, Mechinaud F, Bordigoni P, Hémon D, Clavel J. Road traffic and childhood leukemia: The ESCALE study (SFCE). Environ Health Perspect. 2011 Apr;119(4):566-72.

Sermage-Faure C, Laurier D, Goujon-Bellec S, Chartier M, Guyot-Goubin A, Rudant J, Hémon D & Clavel J. Childhood leukemia around french nuclear power plants - the geocap study, 2002-2007. *Int J Cancer* (2012)

C.2 Article en préparation

Childhood leukemia around high voltage power lines – the Geocap study, 2002-2007. **Sermage Claire**, Demoury Claire, Guyot-Goubin A, Goujon-Bellec S, Rudant J, Hémon D, Clavel J.

C.3 Chapitres d'ouvrages

Sommelet D, Clavel J, Lacour B. Epidémiologie des Cancers de l'Enfant. Springer Verlag ed, Paris, France, janvier 2009 : 371 pages ; chapitre Facteurs de risque liés à l'environnement, Benzène, hydrocarbures polycycliques, pollution atmosphérique p 345

C.4 Communication orale :

Conférence IGN *« L'information géographique au cœur des chantiers du Grenelle de l'environnement » ;* Table ronde : Instaurer un environnement respectueux de la santé ; Cartes épidémiologiques par l'U754 de l'Inserm et Géocible, **Claire Faure** (Inserm), Olivier Lamy (Géocible), Jacqueline Clavel (Inserm), Denis Hémon (Inserm). Le 9 avril 2008 à Paris Porte de Versailles.

C.5 Communications affichées :

Association entre l'incidence des leucémies et tumeurs solides de l'enfant et certaines sources d'expositions environnementales. Estimation de l'exposition au trafic routier et exposition aux champs magnétiques dus à la proximité des lignes à haute tension. Méthodes et premiers résultats ; **Claire Sermage**, Denis Hémon, Jacqueline Clavel ; les rencontres scientifiques de l'ANSES ; exposition aux contaminants de l'environnement ; le 6 décembre 2010 à la Cité universitaire de Paris.

Childhood leukemia around high voltage power lines – the Geocap study, 2002-2007 **Sermage Claire**, Bessou Jérôme, Cougnaud Damien, Deschamps François, Hémon Denis, Clavel Jacqueline*;* Childhood Cancer 2012.

C.6 Recueil de données effectué par la doctorante dans le cadre de ce travail de thèse

RNHE : Ce travail a nécessité l'utilisation des données concernant les 8 149 cas de leucémies aigues enregistrés dans le RNHE sur la période 1990-2007, notamment les adresses au diagnostic recueillies et améliorées avant le début de la thèse.

Echantillon témoin : L'échantillon témoin, constitué pour mette en œuvre les études cas-témoins faisant partie de ce travail, comprend 30 000 sujets : 5 000 par année pour la période 2002-2007. Sa constitution, à laquelle nous avons largement contribué, est détaillée dans le chapitre Matériel et Méthodes et a nécessité plusieurs étapes clés : des discussions préalables avec l'INSEE ont d'abord permis de cibler les bases de données utiles et nécessaires pour la constitution de l'échantillon. Par la suite, une procédure a été mise en place avec les services juridiques de l'INSEE et de la DGI afin de garantir la confidentialité des données venant de la DGI pendant les différents échanges de données prévus. Enfin, des vérifications sur la représentativité de l'échantillon ont été réalisées.

Géolocalisation des sujets : La géolocalisation des adresses des sujets a été faite en étroite collaboration entre notre équipe et l'entreprise "Géocible", nous nous sommes également impliqués de façon importante dans la conduite de cette collaboration à toutes ses étapes. Quand cela a été possible, ce géocodage a été fait de manière automatique, une autre partie de ces adresses a été géocodée manuellement et enfin, certaines ont été géocodées « au domicile » à partir d'examens photographiques. Géocible a signé une convention avec l'INSEE puisque les adresses des sujets témoins INSEE sont livrées directement par l'INSEE à Géocible. Pour mettre en œuvre les conditions de confidentialités édictées par

l'INSEE à propos de la constitution de cet échantillon témoin, nous avons accompagné l'élaboration d'une convention de partenariat entre Géocible, l'INSEE et l'Inserm qui cadre les modalités d'échanges de données conformes à ces conditions de confidentialité.

Par ailleurs, afin de travailler avec des bases de données très précises que sont celles de l'IGN, nous avons accompagné l'élaboration et la signature d'une convention tripartite entre Géocible, l'IGN et notre équipe en janvier 2009 afin que Géocible puisse avoir accès à la BD ADRESSE (version de septembre 2008) de l'IGN pour géocoder les adresses de nos sujets, ceci avant que l'IGN donne accès aux bases de données du Référentiel Grande Echelle (RGE) dont la BD ADRESSE à tous les instituts de recherche, ce qui nous a ensuite permis d'utiliser les mises à jour de ces bases de données (mars 2011).

CNPE : En collaboration avec l'IRSN et par ailleurs en lien avec le groupe de travail mis en place par l'ASN sur les leucémies de l'enfant autour des sites nucléaires, nous avons pu définir la liste des sites nucléaires à prendre en compte dans notre étude, c'est-à-dire l'ensemble des centrales nucléaires de production d'électricité (CNPE). L'IRSN nous a également fourni les coordonnées géographiques exactes des CNPE. Nous avons également du travailler sur la cohérence de notre étude avec les précédentes : suite à la mise à jour des données du RNHE et la rectification de certaines dates de mise en service de CNPE à considérer.

Lignes à haute-tension : Un groupe de travail a été mis en place depuis 5 ans avec plusieurs experts de Rte afin d'évaluer au mieux l'exposition des sujets aux champs magnétiques dus à la proximité des lignes à haute tension. Cette

collaboration a été décrite dans une convention détaillant et précisant les tâches et responsabilités de chacune des deux entités dans nos travaux communs :

- dans un premier temps et avant toute analyse d'une éventuelle association exposition - LA, le groupe de travail "exposition" de Rte a défini les méthodes de caractérisation des expositions liées à une adresse en général, certes de façon concertée avec les épidémiologistes de l'Inserm, mais de façon totalement indépendante de toute analyse d'association exposition-LA,
- dans un second temps, et après que ces méthodes aient été considérées de façon concertée comme fixées, le groupe de travail "épidémiologie » de l'Inserm a réalisé les études d'association entre paramètres d'exposition et LA.

Ces travaux comprennent aussi bien la définition des lignes concernées, leurs caractérisations et leurs géolocalisations, que le positionnement de nos sujets par rapport à ces lignes et la définition d'indicateurs d'exposition aux champs magnétiques pertinents et suffisamment simples pour permettre le traitement du nombre important de sujets inclus dans l'étude.

Trafic routier : Un autre volet du programme GEOCAP dont nous avons initialisé l'étude à l'occasion de la préparation de cette thèse est l'exposition au benzène liée au trafic routier et donc à la proximité des voies de circulation automobile. Nous avons notamment interagi avec l'ADEME (Agence De l'Environnement et de la Maîtrise de l'Energie), AIRPARIF (Association de surveillance de la qualité de l'air en Ile de France) et Géocible pour mettre en place des indicateurs d'exposition pertinents et bénéficier des résultats de leur calcul. Nous avons pu obtenir la localisation précise des routes par type de route (NAVTEQ), l'intensité du trafic sur ces routes (données recueillies par le Centre d'Etudes

Techniques de l'Equipement-CETE pour l'ADEME) ainsi que la concentration en NO_2 dans l'air (ADEME) qui est un bon indicateur du trafic routier.

Une fois collectées, ces données nous ont servi à mettre en place différents indicateurs d'exposition au trafic automobile liée à la proximité des voies de circulation :

- distance à la voie la plus proche (par type de voie),
- densité de voies (nombre de km de voies dans un rayon de 500 m autour de l'adresse du domicile par exemple) et de trafic à proximité de l'habitation
- estimation de la concentration atmosphérique en NO_2 à proximité de l'habitation.

L'association entre certaines de ces variables et l'incidence de LA a été étudiée par Alicia Amigou dans l'enquête cas-témoins ESCALE et a fait l'objet d'un article publié en 2011 dans le revue "Environmental Health Perspectives" et renforçant les évidences publiées à ce jour en faveur de l'existence d'une association positive entre indicateurs d'exposition au benzène liée à la proximité de trafic automobile et accroissement de l'incidence des leucémies de l'enfant[1].

Par ailleurs, nous avons travaillé avec AIRPARIF sur l'élaboration d'un indice d'exposition précis de la pollution atmosphérique due au trafic routier pour tenir compte de différents facteurs (l'intensité du trafic, le parc automobile, la qualité des carburants, les variables météorologiques) dans l'évaluation de l'exposition des sujets. Nous avons ainsi étudié dans ce cadre les expositions au benzène, PM_{10} et NO_2 des sujets d'Ile de France de la période 2002-2004, et leurs corrélations avec les indicateurs simples d'exposition au trafic routier en cours d'élaboration. Nous cherchons à évaluer les expositions des autres sujets (France entière) pour lesquels

[1] Amigou A, **Sermage-Faure C**, Orsi L, Leverger G, Baruchel A, Bertrand Y, Nelken B, Robert A, Michel G, Margueritte G, Perel Y, Mechinaud F, Bordigoni P, Hémon D, Clavel J. Road traffic and childhood leukemia: The ESCALE study (SFCE). Environ Health Perspect. 2011 Apr;119(4):566-72.

les données sont moins nombreuses et moins précises en collaboration avec l'InVS et en tirant profit des travaux d'estimation de la pollution atmosphérique actuellement développés au niveau européen dans le cadre du projet ESCAPE.

D Résumé

Objectifs : Le rôle de l'environnement dans l'étiologie des leucémies aigües de l'enfant (LA) reste encore largement à documenter. Plusieurs facteurs de risque environnementaux "candidats" font aujourd'hui l'objet de recherches intenses, dans un contexte où la caractérisation des expositions environnementales fait des progrès constants, notamment pour celles d'entre elles qui présentent une forte hétérogénéité spatiale à exploiter sur le plan épidémiologique. Dans ce contexte, le présent travail a pour objectif d'étudier la relation entre l'incidence des leucémies de l'enfant et la proximité des centrales nucléaires de production d'électricité (CNPE) et des lignes de transport d'électricité à haute tension (LHT). Avant cette analyse fine, un premier travail a consisté à étudier les variations départementales de l'incidence de LA.

Matériel et méthodes : Les cas inclus dans ces études sont toutes les LA du Registre National des Hémopathies malignes de l'Enfant, qui enregistre l'ensemble des cas diagnostiqués en France métropolitaine, sur la période étudiée : 1990-2004 pour l'étude de l'incidence départementale et 2002-2007 pour les études de l'association entre incidence de LA et facteurs d'exposition environnementale. Pour ces dernières, l'approche principale utilisée est une approche cas-témoins. Les 30 000 sujets témoins constitue un échantillon représentatif de la population pédiatrique française sur la période d''intérêt, construit en lien avec l'INSEE. Dans cette approche la géolocalisation des adresses des cas et des témoins ainsi que la géolocalisation des sources d'exposition et leur caractérisation permet de définir des critères de proximité en relation avec la probabilité et/ou l'intensité d'exposition aux facteurs d'intérêt et de les mettre en relation avec le statut cas *vs* témoins des sujets. Toutes les informations concernant les CNPE et les LHT ont été étudiées en lien étroit avec l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN) et le Réseau de Transport d'Électricité (Rte).

Résultats :

• L'étude des LA par département n'a pas mis en évidence de tendance ni de structure spatiale dans l'incidence à ce niveau géographique : que ce soit globalement, par classe d'âge, par sexe ou par sous-type de leucémie. Ainsi, l'utilisation pour les études ultérieures de taux d'incidence de référence au niveau départemental plutôt que national n'est pas nécessaire.

• Sur la période 2002-2007 contrairement aux périodes précédentes, un quasi-doublement de l'incidence des LA à moins de 5 km des CNPE a été mis en évidence, avec l'approche cas-témoin fondée sur les distance des adresses des sujets aux CNPE (OR= 1,9 [1,0-3,3]) comme avec l'étude d'incidence fondée sur la commune de résidence (O=14 ; E=7,4 ; SIR= 1,9 [1,0-3,2]). Cette augmentation de risque n'était pas spécifique d'une CNPE ou d'un type de CNPE. Non liée à la cartographie des émissions aériennes de radioactivité par les CNPE et incompatible avec les relations dose-risque établies dans la littérature pour ces doses de radioactivité, elle suggère l'influence d'autres facteurs de risque qui restent à déterminer tant au niveau français que dans le cadre d'une collaboration européenne. Ces facteurs pourraient par ailleurs être prévalents non seulement à proximité des CNPE mais également dans d'autres zones et sont donc d'autant plus intéressants à identifier, d'un point de vue scientifique comme d'un point de vue de santé publique.

• L'étude de l'incidence des LA à proximité des LHT a montré une légère association entre la proximité aux lignes de 225 kV et plus et l'incidence des LA mais pas avec la proximité aux lignes de 150 kV et moins. Ces résultats sont concordants avec ceux de la littérature mais devront être confirmés et/ou affinés par l'étude d'estimations modélisées des expositions complétant l'approche par exploitation des critères de distance aux lignes et de tension de lignes (63 à 400 kV) développée ici.

E Abstract

Objectives: The role of the environment in the etiology of childhood acute leukemia (AL) is only partly explored. Many studies are currently ongoing to investigate the role of a few "candidate" environmental risk factors. Meanwhile, the characterization of very localized environmental exposures are made easier, notably thanks to geographic information systems. In this context, the aim of the present work is to study the association between the incidence of AL and the proximity no nuclear power plants (NPP) and to high voltage overhead power lines (HV OLs). At first, the geographical variations of AL have been studied at the *Département* level.

Material and methods: The cases included in the studies are all the cases of AL of the French National Registry of Childhood Haemopatopoietic Malignancies diagnosed in mainland France during the studied periods: 1990-2004 for the study of incidence on *Départements* and 2002-2007 for the studies of association between incidence of AL and environmental exposure factors. Concerning those latter studies, a case-control approach has been used. The control sample, representative of the French pediatric population, contains 30,000 subjects and has been drawn by the INSEE. The precise localization of addresses of subjects and of exposure sources in relation with the type of sources are essential to build indicators of proximity of exposure sources reflecting the probability and/or intensity of exposure and to study the association of theses with the case-control status of the subjects. All data concerning NPP and HV OLs have been studied in close collaboration with the French Institute of Radioprotection and Nuclear Safety (IRSN) and the French Network of Electricity (Rte).

Results:

• The study of AL by *Département* has highlighted neither any trend nor any spatial structure in the incidence at this geographical level globally as well as by age, gender and subtype of leukemia. This lead to the conclusion that the use of departmental reference incidence rates rather than national ones was not justified.

• On 2002-2007, on the contrary of on previous periods, the incidence of AL at less than 5 km from a NPP was nearly twice higher than expected, using both the case-control approach (OR= 1.9 [1,0-3.2]) and the incidence approach (O=14; E=7.4; SIR= 1.9 [1.0-3.2]). This result was not specific to any age group, NPP or a type of NPP. As the incidence of AL was not associated with the geographic zoning of gaseous discharges of NPPs and a doubling of the incidence of LA cannot be expected from the available reference dose-response relationships and estimated radioactive discharges from the NPPs, it can be supposed that there are other risk factors involved which have to be identified in France as in Europe by using a collaborative analysis. Such factors may not only be more frequent in the close vicinity of NPPs but also prevalent in other places and are therefore very interesting to identify, from a scientific as well as public health point of view.

• The study of the proximity to HV OLs highlighted a slight association between the incidence of AL and the close proximity (< 50 m) of very high voltage lines (≥ 225 kV) but not with the proximity to lines of lower voltage (≤ 150 kV). These results are globally concordant with those found in other countries but must be confirmed and narrowed by the ongoing study which uses, as indicator of exposure to magnetic fields (MF) due to proximity of HV OLs, precise estimation of residential exposure to MF. This approach takes in account other determinants of exposure to MF than distance to line by voltage level.

F Liste des tableaux

Tableau 1 : Survie des leucémies aigües diagnostiquées selon l'âge, le sexe et la prévalence d'une trisomie 21 (RNHE, 1990-2000) 24
Tableau 2 : Taux d'incidence par million, standardisé sur l'âge, des leucémies de l'enfant dans le monde..... 25
Tableau 3 : distribution des sujets de l'échantillon témoin 2002-2007 selon différentes caractéristiques en comparaison avec celle de sa population source..... 44
Tableau 4 : Indicateurs de dispersion et d'hétérogénéité des sujets témoins (2002-2004) au niveau communal, en stratifiant sur le département de résidence..... 47
Tableau 5 : Précision du géocodage obtenu en fonction de la précision d'adresse fournie en entrée 52
Tableau 6 : Type de géocodage obtenu pour les cas et les témoins 53
Tableau 7 : Distribution de la distance entre géocodage « automatique » et géocodage « au domicile », en fonction de la précision du géocodage « automatique », chez les premiers sujets géocodés « au domicile »..... 55
Tableau 8 : Buffers choisis pour déterminer les sujets à géocoder « au domicile », par type de géocodage « automatique » 59
Tableau 9 : Géocodage « au domicile » pour les adresses cherchées..... 60
Tableau 10 : Distribution de la distance entre le géocodage "automatique" et le géocodage « au domicile » parmi les sujets à moins de 200m d'une ligne en géocodage « automatique » 61
Tableau 11 : Précision du géocodage « automatique » en fonction du code d'affectation 62
Tableau 12 : Choix des distributions hyper a priori dans le modèle CAR pour l'inverse de la variance 77
Tableau 13 : Hétérogénéité des SIR bruts par sous-population 81
Tableau 14 : SIR des leucémies aigües (1990-2004, garçons et filles, 0-14 ans) par département 86
Tableau 15 : Etendue et ratio des max/min des SIR lissés par sous-population 89
Tableau 16 : Caractéristiques des 19 centrales nucléaires de production d'électricité françaises 105

Tableau 17 : Puissance de l'étude cas-témoins et des études d'incidence pour détecter différentes valeurs (θ) du risque relatif en utilisant des tests unilatéraux au risque d'erreur 5%. 112
Tableau 18 : Distance à la CNPE la plus proche et LA : étude cas-témoin Geocap et études d'incidences 114
Tableau 19 : Description des cas enregistrés à moins de 20 km d'une CNPE sur la période 2002-2007 115
Tableau 20 : Association entre leucémie aigüe et 2 indicateurs d'exposition: distance de l'adresse à la CNPE la plus proche et exposition aux radiations estimée par un zonage géographique des rejets gazeux autour des CNPE dans l'étude cas-témoins Geocap (2002-2007) 116
Tableau 21 : Distance à la CNPE la plus proche et leucémie aigüe de l'enfant - analyses de l'étude cas-témoin Geocap (adresse individuelle) et de l'étude d'incidence (adresse de la mairie), 2002-2007, stratifiées sur la puissance électrique du CNPE, l'année d'étude 121
Tableau 22 : LA à moins de 5 km d'une CNPE, 2002-2007 –estimations après exclusion, une à une, de chaque CNPE 123
Tableau 23 : Période 2002-2007: Classification croisée des témoins, selon la distance « mairie - CNPE la plus proche » et l'estimation de dose à la moelle due aux rejets gazeux radioactifs des CNPE à la localisation de la mairie 127
Tableau 24 : Association entre l'estimateur de dose basé sur un zonage géographique (Dose Based Geographic Zoning - DBGZ [a]) autour des CNPE et les leucémies aigues de l'enfant – Etude cas-témoins Geocap et études d'incidence. 128
Tableau 25 : Réseau de lignes aériennes à haute et très haute tension de Rte en France 140
Tableau 26 : Puissance de l'étude cas-témoins pour détecter différentes estimations (θ) du risque relatif 144
Tableau 27 : Association entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile » 146
Tableau 28 : Association entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne 147

Tableau 29 : Distribution des cas et des témoins à moins de 200m d'une LHT en distance « automatique », et qui ont été géocodés « au domicile » selon différentes variables individuelles et contextuelles 148
Tableau 30 : Croisement des distances « automatique » et « au domicile » chez les cas à moins de 200m d'une ligne ≥225kV en distance « automatique » et ayant un géocodage « au domicile » 149
Tableau 31 : Association, par classe d'âge, entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne 153
Tableau 32 : Association, chez les sujets précis à 20 m, entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne 155
Tableau 33 : Association entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne et de la taille de l'unité urbaine de la commune de résidence 156
Tableau 34 : Association entre incidence de LAL et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne 159

G Liste des figures

Figure 1 : Taux d'incidence des leucémies aigües de l'enfant selon l'âge en France (RNHE, 2000-2004) 23
Figure 2 : Distribution des paramètres de sur-dispersion, de Potthoff-Witthinghill "réduits", pour tous les départements français 46
Figure 3 : Schéma de différentes situations d'exposition de sujets par rapport à une source linéaire 57
Figure 4 : Exemple d'examen photographique montrant la position du géocodage « automatique » et du géocodage « au domicile » d'une adresse 58
Figure 5 : Carte des SIR bruts pour les LA, LAL et LAM 83
Figure 6 : Carte des SIR des LA, lissés par les modèles BYM, CAR, MIX et KHR.. 91
Figure 7 : Carte des SIR lissés par le modèle KHR, pour les LAL et les LAM, par sous-période de 5 ans : 1990-1994 ; 1995-1999 et 2000-2004 93
Figure 8 : Exemple de zonage pour la CNPE de Chinon à partir des estimations de la dose due aux rejets radioactifs gazeux de cette CNPE *(Source : IRSN/DPRH/SER/UETP)* 107
Figure 9 : Localisation des CNPE qui sont proches les unes des autres ou proches d'autres sites nucléaires 109
Figure 10 : Corrélation entre la distance de l'adresse individuelle à la CNPE la plus proche et la distance de la mairie de la commune à la CNPE la plus proche, chez les 995 témoins domiciliés à moins de 20 km d'une CNPE 118
Figure 11 : SIR et OR < 5km d'une CNPE, par CNPE, sur la période 2002-2007 . 124
Figure 12 : Réseau français appartenant à Rte, 225 et 400kV 139

H Liste des abréviations

Bq : Becquerel

CM-EBF : champ magnétique à extrêmement basse fréquence

CNPE : Centrale Nucléaire de Production d'Electricité

DBGZ : « Dose Based Geographic Zoning », c'est-à-dire le zonage géographique établi à partir des estimations de doses à la moelle dues aux rejets des CNPE

GEOCAP : GEOlocalisation des CAncers Pédiatriques, programme de recherche sur les cancers de l'enfant dirigé par J. Clavel (Equipe 6 du CESP – U1018 Inserm)

Géocible : Société d'études et de conseil spécialisée dans le géomarketing

IC95% : intervalle de confiance à 95%

IGN: Institut Géographique National

INSEE : Institut National de la Statistique et des Etudes Economiques

IRSN : Institut de Radioprotection et de Sûreté Nucléaire

LA : leucémie aigue de l'enfant

LAL : leucémie aiguë lymphoblastique de l'enfant

LAM : leucémie aiguë myéloblastique de l'enfant

MWe : méga watt électrique

µT : micro Tesla

O (resp. E): Nombre de cas observés (resp. attendus)

OR : Odds Ratio

RNHE : Registre National des Hémopathies malignes de l'Enfant

Rte : Réseau de Transport d'Electricité

SIG : système d'information géographique

SIR : Ratio d'incidence standardisé

Sv : sievert

II. Problématique scientifique générale et objectifs

A Introduction

A.1 Epidémiologie descriptive des leucémies de l'enfant

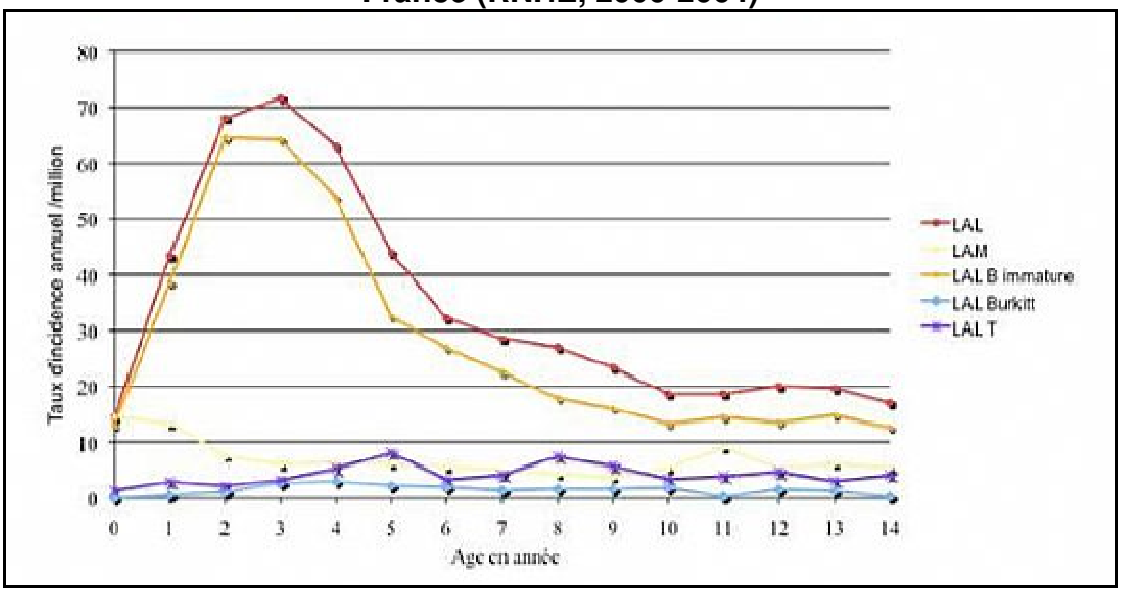
Incidence et survie en France

En France, environ 1 700 enfants de moins de 15 ans sont atteints chaque année d'un cancer, les leucémies sont les plus fréquents d'entre eux avec 40% des cas (470 cas en moyenne par année). Le taux d'incidence brut annuel des leucémies chez l'enfant est de 43,8 par million dans notre pays et le taux standardisé de 45,9.

Les leucémies de l'enfant se développent très majoritairement aux dépens des précurseurs des lignées lymphoïdes. Il s'agit presque uniquement de leucémies aiguës (LA). Les leucémies lymphoïdes chroniques ne se voient pas chez l'enfant ; les leucémies myéloïdes chroniques et les leucémies myélomonocytaires chroniques sont extrêmement rares (0,6 et 0,7 cas par million par an).

Parmi les LA, environ 80% sont des leucémies aigues lymphoblastiques (LAL) et parmi ces dernières, 80% sont des LAL B dites communes. La distribution des LAL B (Figure 1) décrit un pic maximum vers l'âge de 2-3 ans, plus marqué chez les garçons que chez les filles, et encore nettement visible un peu après 6 ans. Moins de 20% des LA sont des leucémies aigues myéloblastiques (LAM). Les garçons sont plus atteints que les filles par la leucémie avec un sex-ratio de 1,3.

Figure 1: Taux d'incidence des leucémies aigües de l'enfant selon l'âge en France (RNHE, 2000-2004)



Source : site web du RNHE : http://www.u754.idf.inserm.fr/page.asp?page=4611

Pour la période 1990-2000, la survie à 5 ans était de 82% pour les LAL et de 58% pour les LAM [Goubin *et al.*,2006]. 18% des enfants atteints de LAL et 42% de ceux atteints de LAM décèdent donc avant 5 ans après le diagnostic. Cependant, ces taux de survie se sont sensiblement améliorés puisque la survie à 5 ans était quasi-nulle il y a 50 ans. Par ailleurs, les pourcentages de survie des leucémies en France sont proches de ceux obtenus dans les pays occidentaux. On peut observer des différences de survie selon l'âge au diagnostic de leucémies et notamment les moins de 1 an ont le pronostic le moins favorable (Tableau 1). La trisomie 21 est un facteur pronostic défavorable dans les cas de LAL.

Tableau 1 : Survie des leucémies aigües diagnostiquées selon l'âge, le sexe et la prévalence d'une trisomie 21 (RNHE, 1990-2000)

	1 an		5 ans	
	%	p	%	p
LAL (N=3960)				
Sexe				
Garçons	94 [93-95]		81 [79-82]	0.07
Filles	95 [94-96]		83 [81-85]	
Age au diagnostic				
< 1 an	77 [69-84]	$<10^{-4}$	48 [39-57]	$<10^{-4}$
1-4 ans	96 [95-97]		87 [85-88]	
5-9 ans	95 [94-96]		83 [81-85]	
10-14 ans	91 [89-94]		72 [69-76]	
Trisomie 21				
Non	94 [94-95]		82 [81-83]	$<10^{-2}$
Oui	92 [86-99]		66 [54-78]	
LAM (N=809)				
Sexe				
Garçons	78 [74-82]		56 [51-61]	
Filles	77 [73-81]		59 [55-64]	
Age au diagnostic				
< 1 an	61 [53-70]	$<10^{-4}$	45 [37-54]	$<10^{-3}$
1-4 ans	80 [75-84]		62 [56-68]	
5-9 ans	84 [78-89]		61 [54-68]	
10-14 ans	81 [75-86]		57 [51-64]	
Trisomie 21				
Non	79 [76-81]	0.06	58 [54-61]	
Oui	67 [55-80]		56 [42-69]	

Source : site web du RNHE : http://www.u754.idf.inserm.fr/page.asp?page=4611

Avant 1990, plusieurs pays ont enregistré une augmentation de l'incidence des LA pour laquelle il a été impossible de faire la différence entre une réelle augmentation et l'amélioration de l'enregistrement. En France, l'incidence des leucémies est stable sur une première période couverte par le RNHE (1990-1999) quelles que soient les tranches d'âge. En revanche, une légère augmentation d'incidence a été détectée sur la période 1990-2004 avec une augmentation annuelle de 0,7 % (p = 0,01) du groupe des « leucémies, syndromes myélodysplasiques et myéloprolifératifs ». En fait, cette augmentation concerne spécifiquement l'incidence des syndromes myélodisplasiques (+4,1 % par an) et elle est vraisemblablement due à l'amélioration des méthodes de recueil des données et d'enregistrement de ces hémopathies.

Variations géographiques

L'incidence des leucémies aiguës connaît moins de variations géographiques que les cancers de l'adulte et elles sont assez bien établies. Dans les pays en voie de développement, l'incidence est plus faible que dans les pays industrialisés (Tableau 2) et le pic des LAL observé avant 5 ans dans ces derniers y est moins marqué voire quasiment absent [Parkin,1998]. Même si l'incidence varie relativement peu en Europe, elle est légèrement plus élevée dans les pays nordiques et les pays de l'est que dans les pays du sud : le taux d'incidence varie de 39,1 cas par million d'enfants dans l'est à 48,0 dans le nord. En moyenne en Europe, il est de 44,0 cas par million et par an [Stiller *et al.*,2006]. Ces variations peuvent évoquer des facteurs environnementaux, socio-économiques ou ethniques. Il reste toutefois possible que certaines d'entre elles reflètent au moins en partie des disparités de diagnostic ou d'enregistrement.

Tableau 2 : Taux d'incidence par million, standardisé sur l'âge, des leucémies de l'enfant dans le monde

	MALE		FEMALE		TOTAL		
	Cases	ASR(W)	Cases	ASR(W)	Cases	M/F	ASR(W)
Africa, North							
Algeria, 4 registries	108	**25.6**	63	**16.0**	171	1.7	**20.9**
*Egypt, Alexandria (1980-1989)	162	**24.5**	97	**15.3**	259	1.7	**20.0**
Tunisia, 3 registries	32	**26.6**	33	**29.0**	65	1.0	**27.8**
Africa, West							
The Gambia (1988-1998)	8	**3.4**	5	**2.1**	13	1.6	**2.7**
Guinea, Conakry (1993-1999)	7	**4.4**	4	**2.5**	11	1.8	**3.5**
Mali, Bamako (1988-1997)	1	**0.6**	5	**2.8**	6	0.2	**1.8**
Niger, Niamey (1993-1999)	4	**5.1**	6	**6.9**	10	0.7	**6.0**
Nigeria, Ibadan (1993-1999)	3	**4.8**	2	**2.1**	5	1.5	**3.0**
Africa, Central							
Congo, Brazzaville (1996-1999)	6	**11.3**	-	-	6	-	**5.8**
Africa, East							
France, La Reunion (1988-1994)	17	**30.4**	17	**29.7**	34	1.0	**30.1**
Kenya, Eldoret (1998-2000)	11	**26.9**	6	**14.5**	17	1.8	**20.7**
Malawi, Blantyre (1991-2001)	3	**1.6**	6	**3.2**	9	0.5	**2.5**
Uganda, Kyadondo County (1993-1997)	9	**8.7**	9	**7.2**	18	1.0	**7.9**
Zimbabwe, Harare: African (1990-1997)	41	**24.4**	34	**19.0**	75	1.2	**21.6**
Africa, South							
*Namibia (1983-1992)	17	**6.3**	16	**6.0**	33	1.1	**6.2**
South Africa: Black (1989-1992)	*259*	**11.8**	*191*	**8.9**	*450*	*1.4*	**10.4**
South Africa: Indian (1989-1992)	*41*	**67.5**	*21*	**34.1**	*62*	*2.0*	**51.0**
South Africa: Mixed race (1989-1992)	*40*	**18.8**	*25*	**11.5**	*65*	*1.6*	**15.2**
South Africa: White (1989-1992)	*111*	**50.4**	*84*	**42.5**	*195*	*1.3*	**46.5**
Swaziland (1996-1999)	6	**6.7**	5	**6.5**	11	1.2	**6.6**
Europe/USA							
USA, SEER: White (1993-1997)	492	**49.3**	402	**42.4**	894	1.2	**45.9**
USA, SEER: Black (1993-1997)	65	**36.0**	46	**26.2**	111	1.4	**31.2**
France, 8 registries (1993-1997)	122	**45.7**	86	**33.4**	208	1.4	**39.7**
The Netherlands (1993-1997)	342	**49.2**	231	**34.8**	573	1.5	**42.2**
UK, England (1993-1997)	1057	**45.5**	835	**37.2**	1892	1.3	**41.4**

**International Incidence of Childhood Cancer Volume II*
In italics: histopathology-based registries

Source : Publication scientifique du CIRC N.153 : Cancer in Africa – childhood cancer

En France, seule une légère hétérogénéité spatiale au niveau communal a été détectée [Bellec *et al.*,2006]. Une étude plus récente n'a trouvé aucune hétérogénéité au niveau des bassins de vie (il y a 1 895 bassins de vie en France) et aucun cluster stable n'a été détecté [Demoury *et al.*,2011]. La recherche d'agrégations de cas, la détection de clusters entre autres, est utilisée pour tenter de mettre en évidence, *a priori* ou *a posteriori*, certains facteurs de risque potentiels qui auraient des variations spatiales.

A.2 Facteurs de risque établis

L'étiologie des leucémies de l'enfant est mal connue. Des **facteurs génétiques** ont été identifiés, mais ils sont à même d'expliquer très peu de cas. Ainsi, les leucémies surviennent exceptionnellement dans le cadre de syndromes génétiques identifiés, comme le syndrome de Li-Fraumeni, certains syndromes d'instabilité chromosomique (anémie de Fanconi, ataxie télangiectasie, syndrome de Bloom) ou une neurofibromatose de von Recklinghausen [Sommelet,2009]. En dehors de ces situations, il faut ajouter la trisomie 21, associée à un risque très élevé de leucémie. En revanche, il est maintenant prouvé que la forte concordance du risque de leucémie de l'enfant entre jumeaux monozygotes résulte de la transmission transplacentaire de clones préleucémiques.

Deux études d'association génome entier publiées récemment convergent sur 2 gènes ARID5B et IKZF1 impliqués dans la différentiation des lymphocytes, dont la mutation hétérozygote serait associée à un doublement du risque de leucémie [Papaemmanuil *et al.*,2009] ; [Treviño *et al.*,2009]. Ces mutations ont des fréquences alléliques de 25 à 70% en fonction des SNPs étudiés. Les résultats de notre équipe confirment ces associations nettes.

L'exposition aux **rayonnements ionisants** à forte dose est un facteur de risque reconnu de leucémie infantile [IARC,2000] ; [IARC,2001].

Les autres facteurs de risque bien établis sont essentiellement **iatrogènes**, puisqu'il s'agit des radiations ionisantes à forte dose, des chimiothérapies anticancéreuses, et à moindre degré des immunosuppresseurs.

A.3 Facteurs de risque suspectés

L'hypothèse que les leucémies de l'enfant pourraient être **d'origine infectieuse** a été envisagée de longue date, mais elle a longtemps été délaissée devant l'absence d'une contagion évidente de la maladie [Alexander,1993]. Au cours des années 1960, des virus du type rétrovirus et herpès virus ont été identifiés dans plusieurs leucémies animales, notamment chez les chats, les poulets, et les bovins. Plus récemment, on a découvert la responsabilité du virus d'Epstein Barr dans les lymphomes de Burkitt et la maladie de Hodgkin et du virus HTLV-1 dans une forme spécifique de LAL T de l'adulte. Cependant, aucun virus spécifique n'a jamais été mis en cause dans les leucémies de l'enfant. A la fin des années 1980, deux hypothèses ont été proposées, mettant en cause de manière différente les infections dans l'étiologie des leucémies de l'enfant. Elles concernaient particulièrement les LAL. Selon l'hypothèse proposée par Kinlen, les leucémies pourraient être une séquelle rare d'une infection fréquente et spécifique, probablement d'origine virale, et le modèle infectieux ressemblerait à celui de la leucémie féline. Cette hypothèse, évoquée notamment devant certains agrégats spatiotemporels de leucémie [Kinlen,1988], a été appuyée par l'observation récurrente d'une augmentation d'incidence dans différents sites soumis à d'importants mouvements de population

[Kinlen *et al.*,1995]. Dans le prolongement de cette hypothèse, notre équipe a mis en évidence une association entre l'incidence des leucémies de l'enfant et les mouvements de population du lieu de résidence à la naissance [Rudant *et al.*,2006] ou à l'interview [Bellec *et al.*,2008]. Dans la dernière étude, les 15% de la population qui était dans les communes isolées ayant connu les mouvements de population les plus importants avaient un SIR de 1,4.

Selon l'hypothèse proposée par Greaves, les leucémies de l'enfant, en particulier les LAL B communes, seraient la conséquence, rare, d'une réponse anormale à des infections banales. Les infections banales des premiers mois de la vie semblent jouer un rôle majeur dans la maturation du système immunitaire. En l'absence de stimulation suffisante, les cellules lymphoïdes préleucémiques qui apparaissent in utero chez environ 1% des fœtus auraient un risque accru de transformation leucémique au cours d'épisodes infectieux de la petite enfance [Greaves,2006]. Plusieurs études montrent une association négative entre les leucémies de l'enfant et un nombre élevé d'infections banales avant 1 an, ou la mise en collectivité précoce [Rudant *et al.*,2010]; [Urayama *et al.*,2010] ; [Urayama *et al.*,2011] . Dans la dernière méta-analyse, l'odds ratio (OR) de l'incidence de LAL associée à la mise en collectivité est de 0,8 (IC95% = [0,7-0,9]). Dans la dernière étude française, les 10% de la population des enfants de plus de 1 an mis en collectivité à plein temps avant l'âge de 1 an ont un OR de 0,8 (IC95% = [0,6-1,1]) et les 30% de cette population ayant eu des infections banales répétées avant l'âge de 1 an ont un OR de 0,7 (IC95% = [0,6-0,9]). Il s'agirait donc d'une hypothèse infectieuse indirecte et non spécifique [Greaves,1988].

Certaines caractéristiques périnatales ont été trouvées associées au risque de leucémie comme un poids de naissance élevé. L'allaitement maternel prolongé semble diminuer le risque de leucémie avec, pour la dernière méta-analyse [Martin *et al.*,2005], des OR pour les LAL de 0,8 (IC95% = [0,7-0,9]) et pour les LAM de 0,7 (IC95% = [0,6-0,9]).

A.4 Facteurs de risques environnementaux

Le rôle de l'environnement dans l'étiologie des leucémies de l'enfant est encore très partiellement documenté. Les principales données disponibles mettent en cause les expositions aux pesticides, aux hydrocarbures et aux rayonnements ionisants et non ionisants.

Les **expositions professionnelles parentales** ont été beaucoup étudiées, mais les expositions y sont généralement mal définies. Les expositions maternelles aux solvants organiques pendant la grossesse pourraient être des facteurs de risque des leucémies de l'enfant.

L'exposition domestique aux **pesticides** a été trouvée liée au risque de leucémie de l'enfant à plusieurs reprises [Lowengart *et al.*,1987] ; [Buckley *et al.*,1989] ; [Leiss &Savitz,1995] ; [Infante-Rivard *et al.*,1999] ; [Ma *et al.*,2002] ; [Menegaux *et al.*,2006] ; [Rudant *et al.*,2007]. La plupart des études suggèrent un doublement de risque chez l'enfant dont la mère a utilisé des pesticides pendant la grossesse. Dans la dernière étude française, les 40% de la population qui ont été exposés aux pesticides pendant la grossesse avaient un OR de 2,1 (IC95% = [1,7-2,5]). L'association est moins nette pour l'utilisation de ces produits pendant l'enfance.

Il est établi que le **benzène** est leucémogène chez l'adulte [IARC,1982]. Cette hypothèse a été étudiée chez l'enfant. Les études sur le trafic automobile sont

majoritairement en faveur d'un risque de leucémie de l'enfant accru à proximité immédiate des routes à fort trafic ([Savitz &Feingold,1989]; [Nordlinder &Järvholm,1997]; [Feychting *et al.*,1998]; [Pearson *et al.*,2000]; [Crosignani *et al.*,2004]; [Visser *et al.*,2004] ; [Amigou *et al.*,2011]). Dans la dernière étude française sur ce sujet [Amigou *et al.*,2011], pour les 10% de la population ayant la densité de voies à fort trafic la plus élevée à moins de 500 m de la résidence, l'OR était de 2,2 (IC95% = [1,1-4,2]). Le registre de cancers lombard rapporte un risque de leucémie significativement lié au niveau estimé de benzène produit par le trafic automobile au lieu de résidence au diagnostic, avec un OR de 3,9 (IC95% = [1,4-11,3]) associé aux expositions au benzène estimées à plus de 10 µg/m^3 [Crosignani *et al.*,2004]. Du point de vue des expositions au benzène, on peut en rapprocher les résultats de deux études cas-témoins successives réalisées dans notre équipe, convergentes et évoquant l'augmentation possible du risque de leucémie à proximité immédiate de garages ou de stations-service (OR entre 2 et 4, concernant 3 à 6% de la population pédiatrique) [Steffen *et al.*,2004] ; [Brosselin *et al.*,2009] ainsi que les résultats d'une autre étude rapportant une augmentation de risque à moins de 100 mètres des stations-service [Harrison *et al.*,1999].

Peu d'études ([Lyons *et al.*,1995]; [Sans *et al.*,1995] ; [Knox &Gilman,1997]; [Wilkinson *et al.*,1999]) ont exploré la possibilité d'une influence de la pollution liée à la présence d'**installations industrielles non nucléaires** sur le risque de leucémie de l'enfant. Aucune n'a trouvé d'augmentation de risque à proximité de sites de pétrochimie. On ne dispose pas encore de données sur le risque de cancer de l'enfant à proximité des incinérateurs et de sources industrielles de dioxines et de PCB.

Le rôle des **radiations ionisantes à forte dose** est établi pour la plupart des cancers de l'enfant, notamment pour les leucémies. Concernant le rôle des **radiations ionisantes à faible dose**, les travaux les plus nombreux ont concerné le risque de leucémie à proximité des sites nucléaires. Dans la recherche des causes possibles de ces cancers, la plupart des études, qu'elles aient été locales et postérieures à la découverte de clusters ou multi-sites et planifiées *a priori*, ont abouti à des résultats ne permettant ni d'affirmer l'existence d'une incidence plus élevée au voisinage des sites nucléaires, ni de pouvoir l'exclure de façon certaine [Laurier &Bard,1999]. Ces travaux ont contribué à renforcer l'hypothèse d'une origine infectieuse de certaines leucémies de l'enfant, notamment à la suite des observations de Kinlen sur les excès d'incidence des leucémies de l'enfant observés dans les zones où un afflux massif de population avait été observé, par exemple sur les sites, comme les sites nucléaires, ayant impliqué l'arrivée d'une population très nombreuse pour leur construction. La littérature concernant ces expositions sera détaillée dans le chapitre correspondant de cette thèse.

En France, 58% de l'exposition aux radiations ionisantes est d'origine naturelle (3 mSv en moyenne par an : 34% due à des radiations alpha liées à l'exposition au radon, 11% due aux rayons gamma d'origine tellurique, 7% due aux rayons gamma d'origine cosmique et 6% due à l'ingestion d'eau et de nourriture). Le radon constitue la deuxième source d'exposition, après la source médicale. Depuis la fin des années 80, une vingtaine d'études écologiques ont mis en relation les variations géographiques d'incidence des leucémies avec celles de la concentration de radon dans l'habitat. La plupart concluent à une corrélation positive [Laurier *et al.*,2001]. Une revue récente portant spécifiquement sur les leucémies infantiles dresse un constat similaire (11 associations positives, dont 8 significatives, sur 12 études

écologiques [Raaschou-Nielsen,2008]). En France, une étude écologique a montré une association modérée mais significative entre l'exposition au radon domestique et l'incidence des LAM chez les enfants : l'incidence était de 24% supérieure dans les zones correspondant au quintile supérieur d'exposition au radon (en moyenne, il y avait une différence de 100 Bq/m^3 entre le quintile d'exposition supérieur et le quintile inférieur pris comme référence) [Evrard *et al.*,2005]. Les études cas-témoins réalisées en population générale ont en revanche rapporté des résultats contrastés et pour la plupart non significatifs [Raaschou-Nielsen,2008]. Cependant, l'association attendue d'après les études écologiques est faible et sa mise en évidence dans le contexte d'une étude cas-témoins exigerait des études d'une taille sensiblement supérieure à celles qui ont été publiées. Une étude récente a toutefois montré une association entre concentration de radon domestique estimée par modélisation et risque de LAL [Raaschou-Nielsen *et al.*,2008]. Récemment, il a été estimé qu'en Grande Bretagne, entre 8 et 30% des leucémies infantiles pourraient être attribuables à la radioactivité d'origine naturelle [Wakeford *et al.*,2009]. Cette estimation a été faite en combinant les données connues de niveaux de radioactivité naturelle en Grande-Bretagne, une conversion de ces doses en dose à la moelle reçue par des enfants et enfin grâce aux modèles de prédiction d'excès de risque de leucémies radio-induites construits à partir de l'étude des survivants des bombardements d'Hiroshima et de Nagasaki, Cette estimation intègre notamment les effets des rayonnements gamma d'origines tellurique et cosmique, qui n'ont pas systématiquement été considérés dans les études citées ci-dessus [Laurier *et al.*,2001].

Les expositions aux **champs magnétiques à extrêmement basse fréquence (CM-EBF) qui sont des radiations non ionisantes** ont également fait l'objet d'un

nombre important de travaux. Deux méta-analyses [Ahlbom *et al.*,2000], [Greenland *et al.*,2000]) ont amené le CIRC à classer l'exposition aux CM de 0,4 µT ou plus comme "Cancérogène possible chez l'Homme" (2B) pour les leucémies de l'enfant [IARC ,2002]. Cette exposition doublerait le risque de leucémie de l'enfant et concernerait environ 1% de la population pédiatrique. La proximité des lignes à haute tension n'explique qu'une partie des expositions élevées aux CM. Toutefois, ce sont des expositions qu'il est possible d'identifier alors que les autres installations à l'origine de ces niveaux de champs sont plus difficiles à localiser et à dénombrer. La littérature concernant ces expositions sera détaillée dans le chapitre correspondant de cette thèse.

Une étude récente sur les **antennes-relais** (radio-fréquences) fréquemment mises en cause par le public en santé publique, amène plutôt à écarter la vraisemblance d'un lien avec les leucémies de l'enfant[Elliott *et al.*,2010].

B Objectifs de la thèse et approche générale

Certains facteurs environnementaux sont toujours suspectés comme à risque pour les leucémies de l'enfant, c'est le cas de la proximité des CNPE, des lignes à haute-tension et des voies de circulation à fort trafic routier. C'est pour approfondir ces questions que le programme GEOCAP a été mis en place : il a pour objectif de savoir s'il existe en France, un excès de cancers chez les enfants de moins de 15 ans domiciliés à proximité (i) des lignes à haute tension (LHT), lié aux champs magnétiques induits par ces lignes, (ii) des voies de circulation, lié à la pollution atmosphérique liée au trafic routier, (iii) des stations service, lié au benzène rejeté dans l'air, (iv) des sites nucléaires et notamment les centrales nucléaires de production d'électricité (CNPE).

Comme ingénieur d'études sur ce programme depuis 2006, j'ai débuté la réflexion sur l'estimation de l'exposition dans ses deux dimensions complémentaires: la distance aux sources et leur caractérisation générale d'une part et l'utilisation d'indicateurs d'exposition fondés sur des modélisations de termes source et processus de diffusion d'autre part.

Pour cette étude, une approche cas-témoins a été choisie. Cette approche a l'intérêt de permettre une estimation de l'exposition individualisée et plus précise que dans une étude écologique, plus objective que dans une étude réalisée par questionnaire et ceci pour un nombre très important de cas et en exploitant les importants contrastes d'exposition qu'on ne peut généralement trouver qu'à l'échelle d'un pays tout entier, analysé sur une maille géographique fine. Dans l'étude d'expositions dont la portée est très courte comme la proximité aux LHT ou au trafic routier, c'est l'unique approche possible puisqu'une étude d'incidence n'est pas faisable à cette

échelle mais seulement à l'échelle des communes. Cependant, certains inconvénients demeurent : comme le fait de ne pas disposer de l'histoire résidentielle des sujets, ce qui aurait pu être possible dans une enquête individuelle. D'autre part, il a été nécessaire de construire un échantillon témoin, ce qui n'est pas utile dans une étude d'incidence qui permet d'étudier l'incidence dans des unités géographiques homogènes qui ont une population de référence bien connue. Cette approche pourrait être considérée comme une étude d'incidence à échelle très fine dans laquelle le groupe témoin donne l'exposition de la population de référence.

Par ailleurs, quand cela a été possible une étude d'incidence a été également faite.

En effet, il existe des phénomènes qu'il est raisonnable d'étudier au niveau communal, comme la pollution atmosphérique de fond ou la proximité aux CNPE. On peut alors associer l'incidence au niveau communal avec ces facteurs d'exposition.

L'analyse de ces données commence naturellement par un premier volet : l'étude de l'incidence des cancers de l'enfant autour des sources d'exposition, en utilisant la première métrique d'exposition : la distance aux sources considérée conjointement avec les caractéristiques générales de ces sources (par exemple la tension des LHT). Par ailleurs ces recherches ne peuvent évidemment pas considérer l'ensemble des cancers pédiatriques comme une seule entité pathologique à étiologie nécessairement commune et nous avons commencé les travaux de ce programme par l'étude des leucémies aigues, étude qui fait l'objet de ma thèse.

Préalablement à cette étude, le premier volet de ma thèse a pour objet d'étudier les variations d'incidence des leucémies de l'enfant entre départements, sans *a priori* sur un facteur de risque. L'étude de ces variations et la recherche d'agrégations de cas peut refléter une propension générale de la maladie à survenir de façon agrégée

dans l'espace, être secondaire à la présence locale d'expositions cancérogènes, ou encore témoigner des fluctuations d'incidence attendues sous l'hypothèse d'une distribution poissonienne des cas.

Le présent travail a pour objectif d'étudier la relation entre l'incidence des leucémies de l'enfant et la proximité des centres nucléaires de production d'électricité et des lignes à haute tension principalement à l'aide d'une approche cas-témoins et à l'aide d'une étude d'incidence quand cela a été possible. Avant cette analyse fine, un premier travail a consisté à étudier les variations départementales de cette incidence.

III. Matériel et méthodes

L'étude de l'incidence des leucémies de l'enfant décrite ci-dessus nécessite de déterminer les cas à inclure, d'avoir certaines informations sur ceux-ci, de déterminer les nombres de cas attendus dans les zones d'intérêt pour les études d'incidence et, pour les études cas-témoins, de sélectionner des témoins appropriés et d'obtenir les informations permettant de les comparer aux cas. Par ailleurs, en ce qui concerne les études cas-témoins utilisant la distance aux sources d'exposition, il est essentiel de bien géolocaliser les adresses des sujets de l'étude et de les positionner par rapport aux sources.

A Les cas

Le Registre National des Hémopathies malignes de l'Enfant (RNHE) a été créé à l'INSERM en 1995 par le Dr J. Clavel au sein de l'Inserm-U170 (actuellement CESP UMRS 1018 – Equipe 6) et a obtenu sa qualification lors de la session 1998 du Comité National des Registres, qualification qui a été renouvelée depuis, tous les 4 ans. Ce registre a été mis en place grâce à la collaboration des pédiatres onco-hématologistes réunis au sein de la Société française de lutte contre les Cancers et les leucémies de l'Enfant et de l'adolescent (SFCE) et est soutenu par l'InVS et l'INCa.

Il contient tous les cas d'hémopathies malignes ou à la limite de la malignité diagnostiqués depuis 1990 chez les enfants de moins de 15 ans habitant sur le territoire français métropolitain au moment du diagnostic de la maladie: cas de leucémies (lymphoïdes aiguës, myéloïdes aiguë myélomonocytaires chroniques, myéloïdes chroniques), de lymphomes (maladies de Hodgkin ou lymphomes non hodgkiniens) et autres hémopathies (syndromes myélodysplasiques, histiocytoses langerhansiennes).

Le nombre moyen de sources pas cas de leucémie diagnostiquée est de 2,8 sur la période 2000-2004. L'exhaustivité du RNHE a été testée par une méthode de capture-recapture et par une comparaison cas par cas avec les fichiers des quatre registres pédiatriques français couvrant les régions Auvergne-Limousin, Bretagne, Lorraine et Rhône-Alpes sur la période 1990-1997. Elle a été estimée à 99,2% pour les leucémies [Clavel *et al.*,2004] [Lacour *et al.*,2010].

Les cas sont principalement identifiés par des enquêteurs qualifiés qui les recherchent de manière active à partir de différentes sources d'information : services d'hématologie et d'oncologie pédiatriques; laboratoires de cytologie et d'anatomo-pathologie; listes nominatives des Départements d'Information Médicale des centres Hospitaliers Universitaires et des Centres de Lutte contre le Cancer; service INSERM des causes médicales de décès ou Cépi-DC (informations non nominatives); registres départementaux et régionaux des cancers du réseau FRANCIM. En France, tous les cas (à quelques exceptions près) sont traités dans des services pédiatriques.

Pour chaque cas, le RNHE dispose d'informations précises sur l'enfant (nom, prénom, sexe, date et commune de naissance, adresse au moment du diagnostic), sur le diagnostic (morphologie et immunologie pour les leucémies, histologie, cytologie et stade pour les lymphomes, caryotype tumoral pour les leucémies et les lymphomes) ainsi que sur la date de diagnostic, le statut vital et la date de dernières nouvelles. Le RNHE essaie d'enregistrer les évènements en suivant activement les enfants. La base de données du RNHE inclut des procédures permettant de détecter des doublons et de vérifier les règles d'exclusion (âge supérieur à 15 ans, résidence

hors de France métropolitaine au moment du diagnostic, premier diagnostic avant 1990).

Dans l'étude des variations géographiques au niveau départemental de l'incidence, les cas inclus sont toutes les LA enregistrées pour la période 1990-2004 telles qu'elles étaient recueillies au 26 septembre 2007.

Pour les études d'incidence autour des CNPE, les cas inclus sont les LA des périodes spécifiques de ces analyses (1990-2004 ; 1990-2001 ; 2002-2007 ; 1990-2007).

Pour GEOCAP, les cas inclus sont toutes les LA de la période 2002-2007 qui avaient moins de 15 ans à la fin de l'année du diagnostic. Le nombre de cas varie légèrement entre l'étude de la proximité aux CNPE et celle de la proximité aux LHT étant donné que quelques cas ont été exclus (suite à une vérification des données du diagnostic) et d'autres inclus (suite à l'inclusion de données plus tardives) entre les deux études.

B Données démographiques

B.1 Estimations de population

L'estimation des ratios d'incidence standardisés (SIR) nécessite de connaître les populations des unités géographiques étudiées auxquelles rapporter le nombre de cas incidents. Nous disposions des données de recensement de l'INSEE de 1990, 1999, 2006 et 2007, qui sont renseignées par commune et par classe d'âge de un an, d'une part et, d'autre part des interpolations intercensitaires de l'INSEE pour les autres années de la période 1990 – 2007, au niveau des communes par classe d'âge de 5 ans et au niveau des départements par classe d'âge de 1 an. Ces interpolations s'appuient sur les taux de migration, de naissance et de décès de la population.

A partir de ces données, nous avons fait des interpolations pour estimer les populations par commune et par classe d'âge de 1 an ou 5 ans pour les années intercensitaires. Pour les enfants qui n'étaient pas nés lors du premier recensement, c'est le nombre de naissances vivantes de l'année considérée qui est utilisé pour estimer les populations de cette génération.

Les personnes-années à risque sont calculées comme la moyenne des effectifs de la population à un âge donné au 1er janvier de l'année concernée et au 1er janvier de l'année suivante, ce sont donc des personnes-années à un âge en années révolues.

Le nombre de cas de LA attendus dans une unité géographique s'obtient ensuite en multipliant les taux d'incidence par classe d'âge de la même année au niveau de l'ensemble de la métropole, choisis comme référence, par le nombre de personnes-années estimées pour les mêmes classes d'âge dans cette unité géographique.

B.2 Données contextuelles

La France métropolitaine, dont la population était de 11 258 747 habitants de moins de 15 ans en 2006 est divisée en 96 départements, dont la population variait en 2006 de 12 338 à 518 280 habitants de moins de 15 ans avec une moyenne de 117 279.
Il y avait 36 569 communes en France en 2006 dont la population de 0 à 14 ans variait de 0 à 313 870 habitants (pour la commune de Paris).

La commune est généralement l'unité administrative française la plus petite pour laquelle des données socio-économiques sont disponibles conjointement aux données d'estimations de population : parmi celles-ci, nous avons utilisé dans ce

travail le taux de chômage, le revenu médian, la proportion de bacheliers chez les plus de 15 ans, la proportion d'ouvriers dans la population active et la tranche d'unité urbaine. Certaines communes de plus de 30 000 habitants sont découpées en plusieurs cantons et les données citées ci-dessus sont disponibles à ce niveau découpage inférieur. Les analyses d'incidence présentées ici seront faites au niveau communal (ou départemental).

La notion d'unité urbaine repose sur la continuité de l'habitat : est considérée comme telle une commune ou ensemble de communes présentant une continuité dans le bâti (les constructions ne sont pas espacées de plus de 200 m) et contenant au moins 2 000 habitants. La condition est que plus de la moitié de la population de chaque commune réside dans la zone agglomérée. Cette classification comporte 9 catégories (la première classe contenant les communes rurales et la dernière étant l'unité urbaine de Paris) et a été faite lors du recensement de 1999.

C Témoins

Nous avons construit, en collaboration avec l'INSEE, un échantillon de 30 000 témoins sur la période 2002-2007, représentatif de la population pédiatrique française, à partir des bases de données de la Direction Générale des Impôts (DGI). Pour chacune de ces années 2002 à 2007, un échantillon de 5 000 enfants a été extrait de la base totale des enfants habitant en France métropolitaine.

C.1 Constitution de l'échantillon

Afin de tirer un échantillon d'adresses représentatif des domiciles des enfants de moins de quinze ans en France et d'avoir toutes les variables nécessaires à l'échantillonnage et à l'étude ultérieure, l'INSEE a croisé les fichiers de déclarations d'impôts sur le revenu de l'année « n » et de taxe d'habitation de l'année « n + 1 » qui permettent de faire le tirage de l'échantillon correspondant à l'année n.

Le tirage a été stratifié par département en fonction de la taille de la population des enfants de moins de quinze ans dans chacun, essentiellement pour faciliter la manipulation de ces fichiers qui sont extrêmement volumineux. Le tirage a été fait par la procédure SURVEYSELECT de SAS par tirage au sort aléatoire des unités statistiques que sont les membres des foyers fiscaux âgés de moins de 15 ans, après stratification sur le département.

Les données antérieures à l'année 2002 n'ont pas pu être utilisées car elles n'étaient plus accessibles sur un plan technique ou conservées de façon non exhaustive par l'INSEE. Notre échantillon de sujets témoins recouvre donc la période 2002-2007 et servira de groupe de référence aux seuls cas contemporains 2002-2007.

C.2 Représentativité de l'échantillon témoin

La représentativité de cet échantillon témoin d'enfants de moins de 15 ans par rapport à la population source du tirage a été vérifiée et est excellente selon différents critères (Tableau 3) : individuels (âge atteint au 31/12, nombre d'enfants dans le ménage, revenu du foyer fiscal) et contextuels au niveau de la commune d'habitation (tranche d'unité urbaine, taux de chômage, revenu médian, proportion de bacheliers chez les plus de 15 ans et proportion d'ouvriers dans la population active). Tous les écarts entre fréquences observées dans l'échantillon témoin et attendues d'après la population source sont très faibles sur le plan quantitatif. Ils sont tous non - significatifs sur le plan statistique sauf la variable "revenu du foyer fiscal" (p=0,01) mais dans ce cas encore les fréquences observées dans l'échantillon témoin (10%± 0,4% au maximum) sont très proches des déciles de la population source.

Tableau 3 : distribution des sujets de l'échantillon témoin 2002-2007 selon différentes caractéristiques en comparaison avec celle de sa population source

		Base DGI 2002-2007		Echantillon 2002-2007		p (Khi-2)
		Nombre	%	Nombre	%	
Age révolu	***0-4ans***	21 076318	*32,9%*	9 717	*32,4%*	
	5-9ans	21 338 878	*33,3%*	10 098	*33,7%*	
	10-14ans	21 634 088	*33,8%*	10 185	*34,0%*	
	0-14 ans	64 049 284	***100,00%***	30 000	***100,00%***	*0,15*
Nombre d'enfants du ménage	***1***	20 164 790	*31,5%*	92 31	*30,8%*	
	2	27 073 337	*42,3%*	12 788	*42,6%*	
	3	12 460 634	*19,5%*	5 903	*19,7%*	
	4	3 203 772	*5,0%*	1 529	*5,1%*	
	5 et +	1 146 756	*1,8%*	549	*1,8%*	
		64 049 289	***100,00%***	30 000	***100,00%***	*0,12*
déciles de revenus de la pop générale	***D1***			2 914	*9,7%*	
	D2			2 937	*9,8%*	
	D3			2 922	*9,7%*	
	D4			2 934	*9,8%*	
	D5			3 138	*10,5%*	
	D6			3 088	*10,3%*	
	D7			3 109	*10,4%*	
	D8			3 022	*10,1%*	
	D9			2 970	*9,9%*	
	D10			2 966	*9,9%*	
				30 000	*100,00%*	*0,01*

Taille d'unité urbaine	**Commune rurale**	16 690 948	26,1%	7 738	25,8%	
	Unité urbaine <5 000 habs	4 044 851	6,3%	1 877	6,36%	
	Unité urbaine 5 000-9 999 habs	3 496 151	5,5%	1 655	5,5%	
	Unité urbaine 10 000-19 999 habs	3 219 437	5,0%	1 504	5,0%	
	Unité urbaine 20 000-49 999 habs	3 827 433	6,0%	1 798	6,0%	
	Unité urbaine 50 000-99 999 habs	4233456	6,6%	1 951	6,5%	
	Unité urbaine 100 000-199 999 habs	3 191 842	5,0%	1 539	5,1%	
	Unité urbaine 200 000-1 999 999 habs	13 902 582	21,7%	6 571	21,9%	
	Unité urbaine de Paris	11 442 589	17,9%	5 367	17,9%	
		64 049 289	***100,00%***	30 000	***100,00%***	*0,90*
Revenu médian de la commune	***Q1***			7339	24,7%	
	Q2			7452	25,0%	
	Q3			7441	25,0%	
	Q4			7534	25,3%	*0,46*
Part de bacheliers dans la population de plus de 15 ans de la commune	***Q1***			7420	24,7%	
	Q2			7502	25,0%	
	Q3			7551	25,2%	
	Q4			7527	25,1%	*0,73*
Part d'ouvriers dans la population active de la commune	***Q1***			7611	25,4%	
	Q2			7490	25,0%	
	Q3			7452	24,8%	
	Q4			7447	24,8%	*0,51*
Taux de chômage de la commune	***Q1***			7499	25,0%	
	Q2			7536	25,1%	
	Q3			7518	25,1%	
	Q4			7447	24,8%	*0,90*

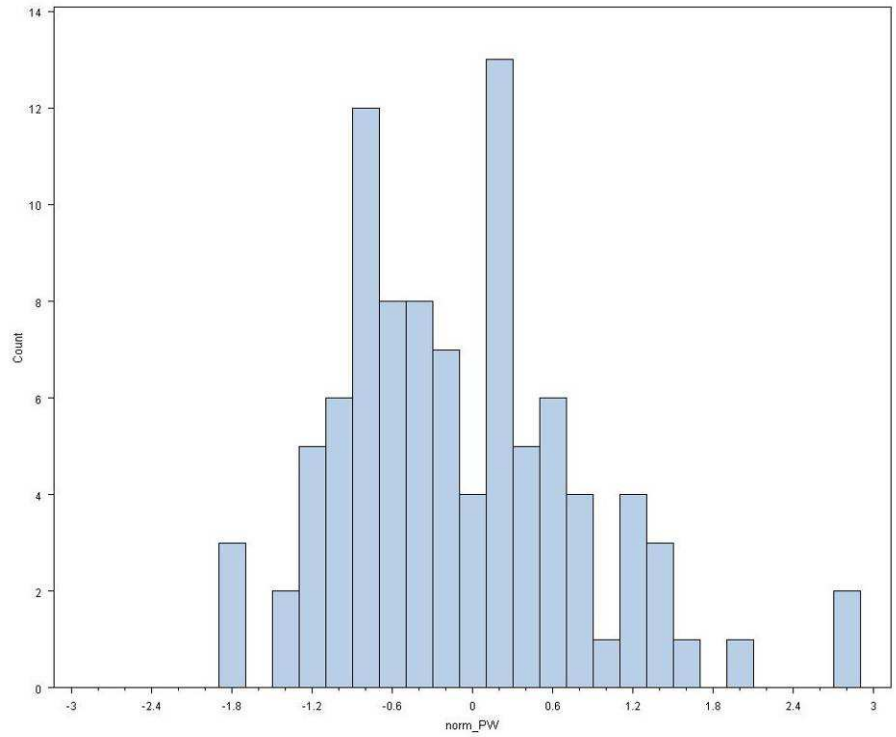
La conformité de l'échantillon avec les données de recensement a également été vérifiée : par âge et par département.

Nous avons aussi étudié la **dispersion** des 15 000 témoins des années 2002 à 2004 afin de vérifier qu'il n'existe pas de sur ou sous dispersion des sujets de l'échantillon par commune, en stratifiant cette analyse sur les départements.

Dans tous les départements où cela était possible - tous les départements métropolitains sauf Paris qui ne contient qu'une commune -, la sur-dispersion a été

testée au niveau communal par le test classique de Potthoff-Wittinghill (voir partie III chapitre E) et une éventuelle hétérogénéité a été recherchée par un test de Khi-2 de Pearson. Seuls 3 des 95 paramètres de sur-dispersion *β^dep* estimés par Potthoff-Wittinghill (Figure 2) sont significativement différents de 1 à 5%, ce qui représente 3% des départements. Parmi les 95 départements métropolitains, 12 sont significativement hétérogènes à 5% au niveau communal avec le test du Khi-2 de Pearson, ce qui représente 12,7% des départements.

Figure 2 : Distribution des paramètres de sur-dispersion, de Potthoff-Witthinghill "réduits", pour tous les départements français



Afin d'étudier la dispersion pour la France entière, plusieurs indicateurs ont été construits en prenant en compte la stratification de l'échantillon faite au niveau départemental :

T1, qui estime la sur-dispersion globale par la somme des *β^dep* dont la variance est la somme des variances de chacun des ß ce qui conduit à la statistique

T1 = ∑ß/√∑ var(ß), dont la distribution peut être approchée par la distribution d'une loi normale centrée réduite.

T2, qui estime la sur-dispersion par la somme des *β^dep*, pondérée par l'inverse de la variance de cet estimateur (prenant ainsi en compte les variations de précisions d'estimations des *β^dep* dues à la taille des départements), ce qui conduit à la statistique

T2 = ∑pondérée des ß / √var(∑ pondérée des ß).

T3, qui tient également compte de la taille des départements mais en considérant la somme des carrés des β^dep/√var(β^dep), cet estimateur suit donc une loi de Khi-2 à 95 degrés de liberté.

Enfin, T4, qui permet de tester l'hétérogénéité globale par la somme des statistiques de tests du Khi-2 de Pearson calculées pour chaque département, avec comme nombre de degrés de liberté « *le nombre total de communes - le nombre de départements* ».

Tableau 4 : Indicateurs de dispersion et d'hétérogénéité des sujets témoins (2002-2004) au niveau communal, en stratifiant sur le département de résidence

	T1 **∑β^dep/√(∑var(β^dep))**	T2 **∑(β^dep/var)/∑(1/var) *√(1/var)**	T3 **∑(β^dep/√var (β^dep))^2**	T4 **∑T(Khi-2^dep)**
valeur	**-1,06**	**-0,61**	**80,15**	35945,90
p [a]	***0,14***	***0,27***	***0,86***	***1,00***

Le coefficient d'hétérogénéité β est défini par $1+\beta=\frac{\mathrm{var}(y_i)}{E(y_i)}$ pour *i*=1,… *n*

[a] *p*-value du test

En utilisant ces différents estimateurs (Tableau 4), aucune sous ou sur dispersion significative ni hétérogénéité dans la distribution des sujets témoins par commune n'a

été trouvée. Le tirage aléatoire effectué par département ne semble donc pas avoir été associé à une répartition non-aléatoire des témoins par commune pour le sous-échantillon des années 2002 à 2004. Même s'il n'a pas été possible avec les données fournies de le vérifier sur les années 2005-2007, cette analyse nous conforte dans l'idée que l'échantillon global de 30 000 témoins sur les années 2002 à 2007 ne révèle pas de caractéristiques géographiques particulières au niveau communal.

Par ailleurs, l'échantillon témoin ayant été construit avec une stratification sur le département avec un effectif par département proportionnel – aux arrondis près - à la taille de la population d'enfants de moins de 15 ans dans le département, ceci nous a amenés à stratifier toutes les analyses cas-témoins sur le département de résidence des sujets.

D Géolocalisation

Dans l'étude cas-témoins, la première étape essentielle pour estimer l'exposition des sujets est le géocodage de leur adresse.

D.1 Géolocalisation des domiciles

D.1.a Principe du géocodage

Le géocodage consiste à déterminer, à partir d'une adresse, la localisation précise en coordonnées géographiques du lieu correspondant, à l'aide d'un système d'information géographique (SIG). "Un SIG est un ensemble de matériels informatiques, de logiciels, de données géographiques, et de personnel capable de saisir, stocker, mettre à jour, manipuler, analyser et présenter toutes formes d'informations géographiquement référencées (F. de Blomac, 1994)".

Le système de coordonnées utilisé pour cette étude est le système de coordonnées projetées Lambert II étendu qui est un système français de projection permettant de définir des coordonnées (X,Y) sur un même plan pour tout point en France. Grâce à ce système, pour deux points P1(x_1,y_1) et P2(x_2,y_2), la distance entre les points P1 et P2 peut être approchée de façon satisfaisante par la valeur :

$P1P2 = \sqrt{((x_2-x_1)^2+(y_2-y_1)^2)}$.

D.1.b Géocodage des adresses GEOCAP

C'est l'adresse au diagnostic pour les cas et l'adresse de l'année considérée pour les témoins qui est géocodée puisque ce sont les seules adresses disponibles. En effet, nous ne disposons de l'histoire résidentielle ni pour les cas, ni pour les témoins.

Le géocodage a été effectué par l'entreprise Géocible au moyen du logiciel de système d'information géographique MAPINFO (Mapinfo, Pitney Bowes Software Inc., Troy, NY, USA) et des cartes vectorisées NAVTEQ (Navteq, Paris, France;

http://www.navteq.com/) et BD ADRESSE (données IGN). Les algorithmes mis en place par Géocible aboutissent, quand c'est possible, à une localisation fournie sous forme de coordonnées Lambert II étendu (X,Y). L'algorithme de géolocalisation permet de corriger de manière automatique les erreurs d'orthographe ou autres erreurs minimes qui peuvent exister dans les adresses fournies en entrée. Cependant, il arrive que malgré cela, les adresses ne soient pas reconnues dans les bases de données. Un travail manuel est alors effectué pour tenter de trouver la localisation de l'adresse fournie : vérification manuelle de l'adresse et/ou utilisation des plans des communes.

Tout ce travail, y compris la dernière partie « manuelle » sera appelée géocodage « automatique ».

D.1.c Précision du géocodage

La meilleure précision possible de géocodage est essentielle pour bien estimer l'exposition des sujets puisque la décroissance de l'exposition à partir des sources peut être très rapide. Cependant, les adresses d'entrée sont de qualité variable : adresses sans numéro s'il n'existe pas ou s'il n'est pas essentiel en utilisation courante et donc non précisé dans les données. Un lieu-dit peut également être renseigné comme seule adresse ce qui ne peut pas toujours permettre une géolocalisation précise. Enfin, certains renseignements peuvent être manquants.

Différents niveaux de précision de localisation des adresses sont obtenus suite à ce géocodage; ils dépendent de la précision de l'adresse d'entrée mais également de l'information disponible dans les bases de données utilisées pour le géocodage.

La meilleure précision est obtenue quand l'adresse est complète (la commune, la voie et le numéro dans la voie sont fournis) et qu'elle est identifiable exactement

dans une des bases de données utilisées : la localisation de l'adresse est faite aux coordonnées exactes indiquées dans la base de données pour la boîte aux lettres de cette adresse précise. Quand l'adresse n'est pas complète ou imprécise ou que les bases de données sont moins renseignées pour cette adresse, la localisation est nécessairement moins précise :

- Quand le numéro n'est pas dans les bases, les numéros approchants (les numéros dans les voies sont souvent indiqués aux croisements avec d'autres voies) peuvent être utilisés pour approcher au mieux la position du numéro de l'adresse.
- Si la voie est fournie mais sans numéro ou si aucun numéro n'est donné pour cette voie dans les bases de données, le géocodage est fait au centre de la voie. La précision dépend alors de la longueur de la voie, qui est parfois renseignée dans les bases de données. Elle a été classée en 4 types : les voies courtes (≤200 m), les voies de longueur moyenne (200-1000 m), les voies longues (>1000 m) et les voies de longueur inconnue.
- Les bases de données incluent des informations sur les lieux-dits et si l'adresse fournie en entrée est un lieu-dit sans indication de voie, le géocodage est fait au centre du lieu-dit.
- Enfin, si seule la commune est fournie en adresse ou si la voie fournie n'est trouvée ni dans les bases de données ni manuellement, la localisation est faite à la mairie de la commune de résidence.

Le détail des affectations de géocodage est indiqué dans le tableau 11, le code d'affectation étant formé de trois parties :

- La première indique la base utilisée pour le géocodage « IGN/NAVTEQ/MANUEL »
- La deuxième partie indique le détail du géocodage obtenu NUMERO/NUMERO_A (numéro approchant)/VOIE/LIEU_DIT/COMMUNE
- La troisième partie indique la longueur de voie de l'adresse (COURTE/MOYENNE/LONGUE/INCONNUE).

Par exemple, un sujet géocodé avec la base NAVTEQ, au numéro approchant et sur une voie longue a le code d'affectation NAVTEQ_NUMERO_A_LONGUE.

Le tableau 5 détaille globalement les précisions obtenues par le géocodage en fonction de l'adresse fournie en entrée.

Tableau 5 : Précision du géocodage obtenu en fonction de la précision d'adresse fournie en entrée

		DONNEES DISPONIBLES DANS LES BASES				
	GEOCODAGE OBTENU	**complètes NUMERO +VOIE +COMMUNE**	**incomplètes: NUMEROS APPROCHANTS +VOIE +COMMUNE**	**incomplètes: VOIE +COMMUNE**	**incomplètes: LIEU-DIT +COMMUNE**	**incomplètes: COMMUNE**
ADRESSE D'ENTREE	**complète : NUMERO +VOIE +COMMUNE**	NUMERO	NUMERO_APPROCHANT	VOIE	/	COMMUNE
	incomplète : VOIE +COMMUNE	Centre de la VOIE	VOIE	VOIE	/	COMMUNE
	incomplète : VOIE +LIEU-DIT +COMMUNE	VOIE	VOIE	VOIE	Centre du LIEU-DIT	COMMUNE
	imprécise : LIEU-DIT +COMMUNE	LIEU-DIT	LIEU-DIT	LIEU-DIT	LIEU-DIT	COMMUNE
	incomplète : COMMUNE	Mairie de la COMMUNE	COMMUNE	COMMUNE	COMMUNE	COMMUNE

Le tableau 6 indique la part des cas et témoins géocodés selon ces différents codes d'affectations globaux.

Tableau 6 : Type de géocodage obtenu pour les cas et les témoins

Géocodage obtenu	Témoins		Cas	
	n	%	n	%
NUMERO	23 171	77,2%	1 940	69,9%
NUMERO APPROCHANT	802	2,7%	130	4,7%
VOIE	4 405	14,7%	470	16,9%
LIEU-DIT	1 330	4,4%	152	5,5%
COMMUNE	292	1%	82	3%
TOTAL	30 000		2 774	

Le nombre de cas est celui de l'étude de proximité aux LHT, légèrement différent de celui de l'étude de proximité aux CNPE.

Comme le montre le tableau 6, la précision du géocodage est légèrement meilleure chez les témoins, issus des bases de données de la Direction Générale des Impôts, que chez les cas issus du RNHE. Cet écart de qualité reflète un écart de qualité des adresses recueillies chez les cas et chez les témoins. Il est cependant limité et en analysant la sensibilité de nos résultats selon la catégorie de précision du géocodage, nous avons vérifié à toutes les étapes de ce travail que cet écart n'expliquait aucun de nos résultats.

Suite à ce premier géocodage, une étape de vérification a été effectuée en lien avec Géocible afin de corriger d'éventuelles erreurs :

- la première étape consiste à vérifier que la géolocalisation finale est bien dans la commune fournie en entrée,
- la deuxième étape permet de vérifier que la géolocalisation est à la voie indiquée en entrée,
- et la troisième étape permet de vérifier, le cas échéant, que la géolocalisation correspond au numéro de voie indiqué en entrée.

D.1.d Validation du géocodage

Une validation de la géolocalisation des sujets a été faite en cherchant manuellement la position exacte du bâtiment de l'adresse des 1 287 sujets initialement considérés comme potentiellement exposés à >0,1 µT du fait de leur proximité des lignes à haute tension du réseau Rte ; c'est-à-dire à moins de 200 m d'une ligne 400kV, à moins de 120 m d'une ligne 225 kV, à moins de 100 m d'une ligne 150 kV ou à moins de 70 m d'une ligne 90 ou 63 kV.

Parmi les 1 287 sujets sélectionnés, 822 ont pu être géolocalisés « au domicile », c'est-à-dire sur l'habitation. En effet, parmi les 464 restants, 333 avaient une adresse initiale incomplète (sans numéro de rue par exemple) qui ne permettait pas de localisation « au domicile » et pour 131 sujets, les données des outils de géocodage étaient insuffisantes.

Une première estimation de la précision du géocodage avait été faite a *priori* par Géocible en fonction de catégories d'affectation initialement fournies par eux : 15 m pour les NUMERO, 50 m pour les NUMERO APPROCHANT sur une VOIE COURTE, 100 m pour les NUMERO APPROCHANT sur une VOIE LONGUE ou au centre d'une VOIE COURTE ou au centre d'un LIEU-DIT, 500 m pour les géocodés au centre d'une VOIE LONGUE et 1 000 m pour les géocodés à la mairie de la COMMUNE. La distribution des distances entre le géocodage « automatique » et le géocodage « au domicile » pour les 822 sujets concernés fait l'objet du tableau 7. Cette distance étudiée ici peut être considérée comme une bonne estimation de la précision du géocodage « automatique ».

Tableau 7 : Distribution de la distance entre géocodage « automatique » et géocodage « au domicile », en fonction de la précision du géocodage « automatique », chez les premiers sujets géocodés « au domicile »

Précision *a priori* du géocodage « automatique »	n	min	p5%	p25%	médiane	p75%	p95%	p99%	max
A (15 m)	667	1,0	3,2	8,2	17,0	34,0	96,6	251,5	521,5
B (50 m)	10	17,0	17,0	36,8	60,3	89,3	219,3	219,3	219,3
C (100 m)	98	7,1	22,4	46,7	82,7	201,9	1 291,4	4 668,3	4 668,3
D (500 m)	47	19,2	20,8	44,4	120,4	439,0	1 646,9	9 881,6	9 881,6
TOTAL	822								

px% : $x^{ème}$ percentile de la distribution

Cette étape confirme que l'estimation de la précision du géocodage « automatique » énoncée *a priori* par Géocible était correcte en médiane. En effet, pour les quatre catégories de précision définies a priori par Géocible (respectivement 15 m, 50 m, 100 m et 500 m) les médianes des distances par catégorie sont très proches de ces valeurs (respectivement 17 m, 60 m, 83 m, 120 m). Cependant, il existe des situations dans lesquelles les distances entre les deux géocodages sont bien plus importantes qu'attendu. Par exemple, pour certains sujets qui avaient été géocodés avec une précision *a priori* à 15 m, la position exacte du bâtiment est à plus de 200 m (p 99% de la distribution = 251,5).

D.2 Positionnement des domiciles par rapport aux sources

D.2.a Positionnement des adresses par rapport aux sources

Une fois que la localisation des adresses des sujets dans l'espace effectuée, il faut les positionner par rapport aux sources d'intérêt afin de déterminer l'exposition des sujets. Les sources d'exposition étudiées ici ont une localisation qui est connue de manière très exacte.

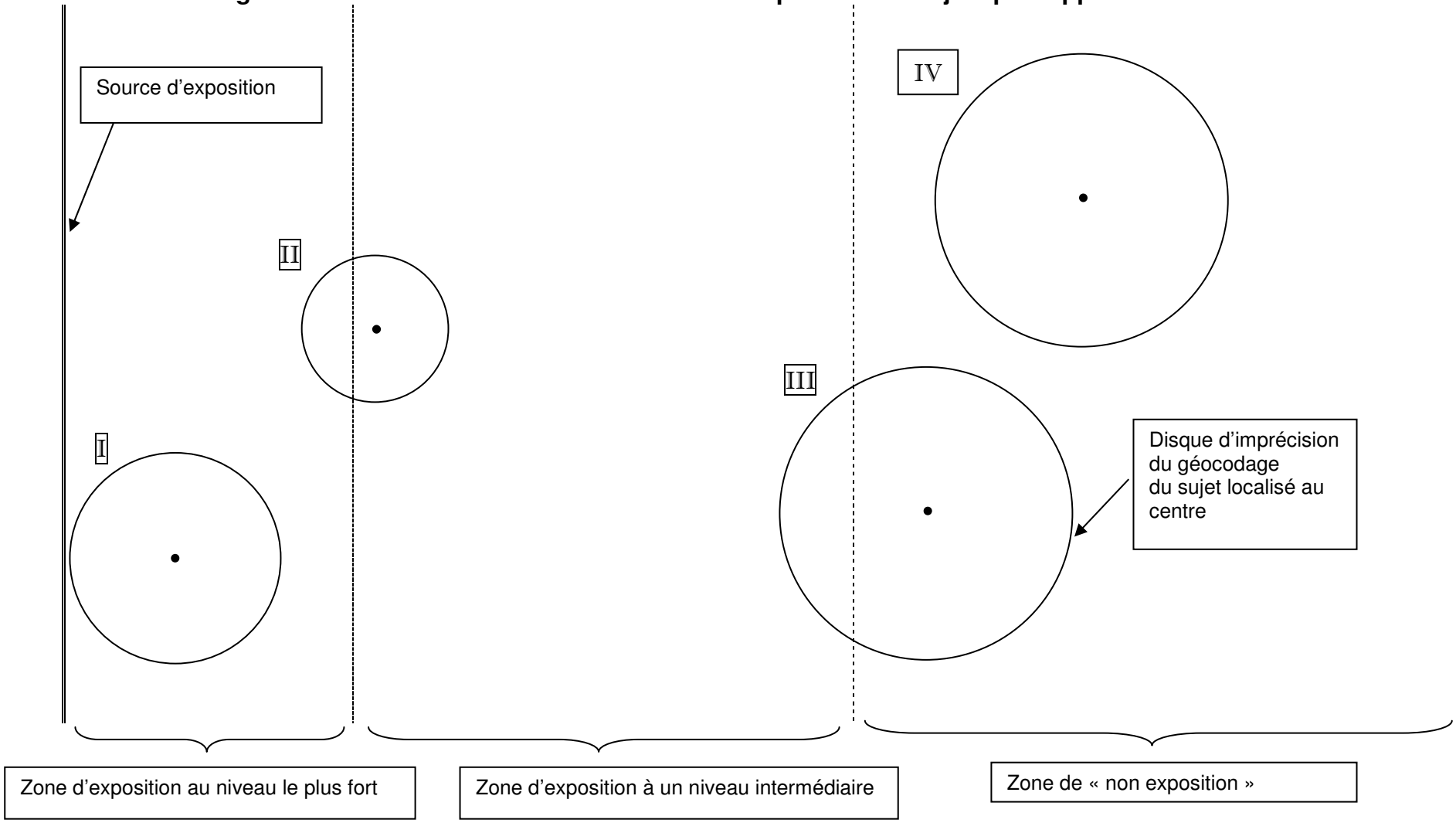
Pour classer correctement un sujet dans une catégorie d'exposition par rapport à une source, il est important de prendre en compte la précision du géocodage

« automatique » de l'adresse du sujet. Pour cela, un disque correspondant à la précision du géocodage de chaque sujet lui est associé dont le centre est le point géocodé et s'il recoupe les buffers d'exposition autour des sources (par exemple, des classes de distance par rapport aux sources - Figure 3), le sujet est considéré comme potentiellement exposé (situations I, II et III). Si le disque d'imprécision du sujet recoupe un seul buffer d'exposition (situation I), la classe d'exposition est connue avec une très bonne précision. Si le disque recoupe plusieurs buffers d'exposition (situations II et III), on peut souhaiter une meilleure localisation de l'adresse du sujet pour connaître avec plus d'exactitude l'exposition du sujet. S'il n'y a pas de recoupement, il peut être considéré comme non exposé (situation IV) et une géolocalisation plus précise pour ce sujet n'est pas nécessaire.

En ce qui concerne la proximité des CNPE, à partir des coordonnées Lambert des centres de gravité des cheminées d'évacuation des rejets gazeux de CNPE dans l'atmosphère, nous avons pu déterminer la distance de l'adresse de chaque sujet à la CNPE la plus proche : les sujets potentiellement exposés étant ceux dont le buffer de précision recoupe des buffers de 20 km autour des CNPE. Les classes d'exposition étant larges (cercles concentriques de 5 km), peu de sujets avaient des disques d'imprécision qui recoupaient deux catégories de distances nécessitant une vérification de géocodage. Nous avons donc vérifié manuellement la géolocalisation de ces sujets.

En ce qui concerne la proximité aux LHT, la décroissance des champs magnétiques induits par celles-ci est proportionnelle au carré de l'inverse de la distance, donc est très rapide (entre 0 et 200 m au maximum), il était donc important, pour tous les sujets proches des lignes d'avoir une localisation très précise des domiciles.

Figure 3 : Schéma de différentes situations d'exposition de sujets par rapport à une source linéaire



D.2.b Géocodage « au domicile »

Le géocodage « au domicile », c'est-à-dire sur le bâtiment, que nous souhaitions obtenir pour les sujets proches des LHT a été réalisé, par Géocible, à partir de bases de données (Google Map, Street View) permettant de rapprocher informations photographiques et adresses (Figure 4). Il est différent d'un géocodage « automatique » puisque d'une part, il peut permettre de positionner de manière précise ou corrigée un bâtiment le long de la voir en levant d'éventuelles incertitudes ou erreur liées à la connaissance de la seule adresse et d'autre part, il fait aussi passer d'un géocodage « à la boîte au lettres », c'est-à-dire au bord de la voie, à un géocodage sur l'habitation, c'est-à-dire parfois en retrait par rapport à la voie, exactement sur le bâtiment.

Figure 4 : Exemple d'examen photographique montrant la position du géocodage « automatique » et du géocodage « au domicile » d'une adresse

Point jaune : position du géocodage « automatique »
Point vert : position du géocodage « au domicile »

Les sujets pour lesquels nous souhaitions obtenir ce géocodage sont tous ceux dont le bâtiment du domicile était possiblement situé à moins de 200 m des lignes. En termes de géocodage « automatique », ils pouvaient effectivement avoir été positionnés à moins de 200 m d'une ligne mais ils pouvaient aussi avoir été positionnés plus loin avec une imprécision du géocodage « automatique » élevée.

Une première sélection de sujets a été faite en prenant les sujets pour lesquels le disque d'imprécision associé au géocodage « automatique » recoupe les buffers d'exposition de 200 m autour des LHT. Ces disques d'imprécision ont été choisis larges de manière à faire une sélection la plus sensible possible des sujets habitant réellement à moins de 200 m d'une ligne, quelle qu'ait pu être la localisation suite au géocodage « automatique » et son imprécision. Pour cela, nous avons utilisé la validation du géocodage « automatique » précédemment faite et la distribution de la distance entre les deux types de géocodage « automatique » et « au domicile » pour déterminer une précision maximaliste du géocodage « automatique » en fonction du code d'affectation détaillé. A partir de ces données, nous avons choisi des disques d'imprécision du géocodage correspondant aux 99èmes percentiles de ces distributions (Tableau 8).

Tableau 8 : Buffers choisis pour déterminer les sujets à géocoder « au domicile », par type de géocodage « automatique »

Type de géocodage	p99% de la distribution de la distance géocodage « automatique » - géocodage « au domicile » (m) *
Numéro sur voie courte ou moyenne	180
Numéro sur voie longue	350
Numéro sur voie de longueur inconnue	450
Lieu-dit	700
Numéro approchant sur une voie courte moyenne ou de longueur inconnue ou au centre d'une **voie** courte	1 000
Numéro approchant sur une voie longue ou au centre d'une **voie** moyenne, longue ou de longueur inconnue	5 000

* chez les 822 sujets géocodés « au domicile » dans la première étude de validation du géocodage « automatique »

Afin de déterminer d'une manière un peu plus spécifique les sujets potentiellement proches des lignes, une deuxième sélection a été faite en parallèle à partir de la commune d'habitation. Nous avons retiré de la première liste des sujets pour lesquels on souhaite un géocodage « au domicile », tous les sujets dont la commune ne recoupait pas les buffers d'exposition des lignes. Ceci est intéressant dans la mesure où certains sujets peuvent être proches de la frontière de leur commune d'habitation mais avoir un géocodage « automatique » imprécis et donc un disque d'imprécision associé très large, dépassant les frontières de la commune. Si ces sujets appartiennent à une commune ne recoupant pas le réseau à 200 m, il n'est pas utile de chercher pour eux un géocodage « au domicile ».

Parmi les 3 607 adresses pour lesquelles cela a été cherché, 2 846 (78,9%) ont pu faire l'objet d'un géocodage « au domicile » (Tableau 9), le géocodage est donc fait sur la maison, ou sur le bâtiment si le logement n'est pas individuel.

Tableau 9 : Géocodage « au domicile » pour les adresses cherchées

Type de géocodage	n	%
Maison	**2 229**	**61,8%**
Bâtiment	**617**	**17,1%**
Non trouvée	761	21,1%
Total des adresses cherchées	3 607	100,0%

Globalement, la part de sujets imprécis chez les sujets pour lesquels on a un géocodage « au domicile » est plus importante que chez tous les sujets de l'étude puisqu'en dehors de la bande stricte de 200 m autour des lignes, le géocodage « au domicile » était cherché uniquement si le disque d'imprécision recoupait la bande de

200 m autour des lignes. Afin d'avoir une estimation de la distribution de la distance entre le géocodage « automatique » et le géocodage « au domicile » chez tous les sujets, en évitant de surreprésenter les sujets dont les géolocalisations "automatiques" étaient les moins précises, nous n'avons apprécié cette distribution que chez les seuls sujets dont le géocodage « automatique » était strictement inférieur à 200 m d'une ligne. Le tableau 10 décrit cette distribution par type de géocodage : il montre que le géocodage « automatique » avait une bonne précision en médiane et bien corrélée au type de géocodage, ceci même si des valeurs extrêmes sont observées.

Tableau 10 : Distribution de la distance entre le géocodage "automatique" et le géocodage « au domicile » parmi les sujets à moins de 200m d'une ligne en géocodage « automatique »

Distance entre le géocodage "automatique" et le géocodage « au domicile » (m)							
Type de géocodage	**n**	**min**	**p25%**	**médiane**	**p75%**	**p95%**	**max**
Numéro	806	0	5,4	**12,0**	25,3	**76,1**	521,5
Numéro approchant	27	10	22,9	**56,7**	196,6	**1383**	1987,6
Voie	33	22	41,9	**88,3**	136,5	**997,7**	1868,6
Lieu-dit	6	40.3	45,9	**148,1**	517,9	**695,2**	695,2
TOTAL	872	0	5,92	**13,2**	29,6	119,7	1987,6

De meilleures estimations de la précision du géocodage « automatique » ont pu être faites à partir de ce travail, le tableau 11 les décrit par code d'affectation détaillé.

Tableau 11: Précision du géocodage « automatique » en fonction du code d'affectation

Géocodage « automatique »	Précision estimée (m)	
Code d'affectation	p50%	p99%
IGN_NUMERO_COURTE	20	180
IGN_NUMERO_MOYENNE	20	180
NAVTEQ_NUMERO_COURTE	20	180
NAVTEQ_NUMERO_MOYENNE	20	180
IGN_NUMERO_LONGUE	20	350
NAVTEQ_NUMERO_LONGUE	20	350
IGN_NUMERO_INCONNUE	20	450
NAVTEQ_NUMERO_INCONNUE	20	450
MANUEL_NUMERO_COURTE	50	180
MANUEL_NUMERO_MOYENNE	50	180
MANUEL_NUMERO_LONGUE	50	450
IGN_VOIE_COURTE	50	1000
NAVTEQ_VOIE_COURTE	50	1000
MANUEL_VOIE_COURTE	50	1000
MANUEL_VOIE_MOYENNE	50	5000
MANUEL_NUMERO_INCONNUE	100	2500
IGN_NUMERO_A_COURTE	100	1000
IGN_NUMERO_A_INCONNUE	100	1000
NAVTEQ_NUMERO_A_COURTE	100	1000
NAVTEQ_NUMERO_A_INCONNUE	100	1000
NAVTEQ_NUMERO_A_MOYENNE	100	5000
NAVTEQ_VOIE_INCONNUE	100	5000
NAVTEQ_VOIE_MOYENNE	100	5000
IGN_VOIE_MOYENNE	100	5000
IGN_NUMERO_A_MOYENNE	100	5000
IGN_VOIE_INCONNUE	100	5000
MANUEL_VOIE_INCONNUE	100	10 000
IGN_LIEU_DIT_INCONNUE	250	1600
MANUEL_LIEU_DIT_INCONNUE	250	1600
NAVTEQ_LIEU_DIT_INCONNUE	250	1600
IGN_NUMERO_A_LONGUE	250	5000
IGN_VOIE_LONGUE	250	5000
NAVTEQ_NUMERO_A_LONGUE	250	5000
NAVTEQ_VOIE_LONGUE	250	5000
MANUEL_VOIE_LONGUE	250	5000
NAVTEQ_VOIE_INCONNUE	commune	commune

E Méthodes statistiques générales

Les analyses des données recueillies incluent la tabulation des résultats, des représentations graphiques afin de détecter des éléments aberrants et l'ajustement de modèles aux données : régression logistique pour les études cas-témoins et régression de Poisson pour les études d'incidence.

Les analyses statistiques ont été réalisées en utilisant le logiciel SAS software package (version 9 ; SAS Institue, Cary, NC) ainsi que le logiciel R 2.8.1 (2008).

Pour tester l'hétérogénéité des SIR dans les analyses ou l'hétérogénéité de la répartition des témoins par commune, **le test classique du Khi-2 de Pearson** $Q^2 = \sum_i \frac{(y_i - E_i)^2}{E_i}$ a été utilisé. Le test du **rapport de vraisemblance** (de Poisson) $L = 2\sum_i y_i \, ln(SIR_i)$ a également été utilisé pour l'étude des SIR. Ces statistiques suivent asymptotiquement une loi du Khi-2 à (n–1) degrés de liberté sous l'hypothèse nulle d'homogénéité.

La présence d'une sur-dispersion Poissonnienne a été testée par le test **de Potthoff et Whittinghill** en postulant que le nombre de sujets ou de cas observés dans une unité géographique est distribué indépendamment pour chacune selon une loi binomiale négative, telle que le rapport entre sa variance et sa moyenne est de 1+β. Ainsi, la sur-dispersion est quantifiée par le coefficient d'hétérogénéité β : β=0 correspond à une absence de sur-dispersion (loi de Poisson) alors que β>0 indique que la variance est supérieure à la moyenne (sur-dispersion). Sous l'hypothèse d'homogénéité, la statistique de test proposée par Potthoff et Whittinghill [Potthoff &Wittinghill,1966] ; [Potthoff &Wittinghill,1966] suit asymptotiquement une loi du Khi-2 et il a été prouvé que ce test était localement plus puissant que le test de Khi-2 de

Pearson, sous cette hypothèse alternative de sur-dispersion Poissonnienne. Muirhead et Butland [Muirhead &Butland,1996] ont proposé une estimation du coefficient d'hétérogénéité β et de son écart-type sous l'hypothèse nulle.

$$\hat{\beta} = \frac{y_+ \left[\sum_i \left(\frac{y_i (y_i - 1)}{E_i} \right) - y_+ + 1 \right]}{(n-1)(y_+ - 1)} \text{ avec } y_+ = \sum_i y_i \text{ ;}$$

E.1 Incidence

Dans les études d'incidence, le risque relatif de LA a été estimé par le **ratio d'incidence standardisé (SIR),** défini comme le ratio du nombre de cas observé (O) sur le nombre de cas attendus (E). L'intervalle de confiance (IC) à 90% dans l'étude des variations d'incidence au niveau départemental est l'intervalle exact de Poisson [Estève *et al.*,1993] et celui à 95% pour l'étude d'incidence autour des CNPE a été calculé en utilisant l'approximation de Byar [Breslow &Day,1987].

Dans l'étude de la proximité aux CNPE, l'hétérogénéité des SIR par année et par CNPE a été testée en prenant une référence externe puis interne (à une zone de 20 km autour de chaque CNPE). Pour prendre en compte le fait que les nombres de cas sont faibles, ce qui aurait pu compromettre la validité des tests, les degrés de significativité ont été estimés par simulation. 50,000 distributions des cas dans les communes ont été générées sous l'hypothèse nulle d'une distribution de Poisson avec les nombres attendus correspondants et les statistiques observées ont été comparées aux distributions respectives des 50 000 statistiques issues des simulations.

E.2 Cas-témoins

Dans les études cas-témoins, le risque relatif de LA a été estimé par **l'odds ratio (OR)** calculé par régression logistique non conditionnelle ajusté sur l'âge en 3 classes de 5 ans et stratifié sur le département de résidence.

L'âge considéré pour ces études est l'âge au 31 décembre puisque la seule information disponible dans les fichiers de l'INSEE sur l'âge des témoins est leur année de naissance : c'est donc l'âge en années révolues au 31 décembre à la fin de l'année du diagnostic pour les cas et à la fin de l'année d'échantillonnage pour les témoins qui a été utilisé.

Des tests de tendance associés aux variables catégorielles de distance ont été effectués. Pour cela, nous avons créé une variable prenant comme valeur, pour chaque classe des variables catégorielles, la médiane de l'inverse de la distance : c'est un choix souvent fait dans ce type d'études, pour transformer de façon simple une distance en une proximité variant en sens inverse et ramenant notamment à des valeurs proches de 1/d = 0 tous les points "éloignés", ceci sans considérations théoriques ou empiriques sur la forme de la relation distance-odds ratio. L'écart à la linéarité a été testé grâce au test du maximum de vraisemblance et la variable créée a été introduite dans le modèle comme variable continue. La pente de la droite était ainsi testée.

IV. Variations d'incidence au niveau départemental

A Introduction

En France, le taux d'incidence annuel de LA est de 13,8 par million. Récemment, une légère hétérogénéité a été détectée chez les LA à l'échelle des communes (généralement la plus petite unité administrative en France qui contient 36 000 communes pour 60 millions d'habitants) sur la période 1990-1994, et plus particulièrement chez les enfants de moins de 5 ans [Bellec *et al.*,2006]. Cette observation rejoignait certaines de l'étude EUROCLUS [Alexander *et al.*,1998].

Ce chapitre concerne la recherche de variations spatiales de l'incidence de LA, par département, en France métropolitaine et sur la période 1990-2004. Les sous-types de LA (LAL et LAM) ont entre autres été étudiés séparément puisque les différentes cytologies peuvent présumer des différences de facteurs de risque. Si l'incidence de LA varie au niveau départemental en France, il peut être important de le prendre en compte dans la recherche de facteurs de risque. Un autre intérêt de cette étude est qu'elle peut, s'il existe une variation de l'incidence de LA au niveau départemental, inciter à la recherche de facteurs de risque à la même échelle comme le comportement de santé (nutrition, consommation d'alcool et de tabac, infections), le degré d'urbanisation du département ou certaines caractéristiques environnementales comme le radon d'origine naturelle. La suspicion de tels facteurs de risque environnementaux nous a amenés à étudier les variations départementales de l'incidence de LA en fonction de différentes sous-périodes, du sexe de l'enfant ainsi que de leur âge.

Les modèles utilisés ici pour estimer les risques relatifs d'incidence des LA dans chaque département sont bien adaptés pour l'étude de maladies non-contagieuses et rares, c'est-à-dire donnant lieu à un petit nombre de cas observés. Ils prennent en

compte l'existence possible de dépendances spatiales locales des risques entre eux (ici ceux des différents départements) qui seraient dues à l'influence de facteurs de risque non mesurés ou inconnus de la maladie et partagés par des unités géographiques voisines. Ces modèles font appel à une formulation hiérarchique bayésienne en postulant *a priori* la combinaison d'un processus spatial sous-jacent des vrais taux d'incidence inconnus et d'une variabilité poissonienne des cas incidents observés [Wakefield *et al.*,2000].

Parmi les modèles spatiaux qui proposent de prendre en compte cette dépendance spatiale [Best *et al.*,2005], deux modèles spatiaux paramétriques basés sur des champs aléatoires de Markov ont été choisis : les modèles BYM [Besag *et al.*,1991] et CAR [Stern &Cressie,2000]. Ce sont des modèles couramment utilisés dans la cartographie de maladies. Cependant, ce type de modèles paramétriques appliquent un lissage global sur les risques ce qui peut lisser plus que souhaité la structure de risque si celle-ci n'est pas continue. Afin de permettre la mise en évidence d'éventuelles discontinuités locales (avec des îlots de départements à risques élevés) et de mettre moins d'hypothèses sur les distributions *a priori*, deux modèles spatiaux semi-paramétriques ont aussi été utilisés : un modèle mixte spatial appelé ici modèle MIX [Green &Richardson,2002], et un modèle spatial de partition propose par Knorr-Held et Raßer [Knorr-Held *et al.*,2000] appelé ici modèle KHR.

La principale caractéristique partagée par ces deux types de modèles est qu'ils visent à amortir la composante aléatoire des variations spatiales des risques relatifs observés localement pour mieux révéler une éventuelle structure sous-jacente qu'elle soit continue (modèles paramétriques) ou discontinue (modèles non-paramétriques).

B Matériel

B.1 Registre des cas

Les données utilisées pour cette étude sont les cas de LA du RNHE, diagnostiquées entre le 1er janvier 1990 et le 31 décembre 2004 en France métropolitaine, telles qu'elles étaient enregistrées dans le registre au 26 septembre 2007 (8 cas ont été retirés car leur département de résidence n'était pas connu, et parmi ceux-là, 4 n'avait comme source d'information que la fiche d'admission à l'hôpital).

Cette base de données comportait donc 6 796 cas de LA sur les 94 départements français métropolitains. Parmi elles, 81% (5 526) étaient des LAL et 17% (1 160) des LAM, les 2% (110) restants étant des LA d'autres types.

Le pic bien connu d'incidence de LA (et de LAL) a été bien observé sur ces données à 2-3 ans et l'incidence est également légèrement plus élevée, avec un pic plus étendu et plus tardif pour les garçons que pour les filles.

B.2 Population

Les personnes-années 1990 à 2003 ont été calculées à partir des données de population fournies par l'INSEE. Pour calculer les personnes-années en 2004, les populations par département, sexe et classe d'âge de 1 an, aux 1er janvier 2004 et 2005 sont nécessaires. Les données de 2005 n'étaient connues, au moment de cette analyse, que par département, sexe et classe d'âge de 5 ans. Une estimation de la répartition souhaitée a été effectuée à partir de ces données et des répartitions de la population par classes d'âge de 1 an dans celles de 5 ans en 2004. Cette estimation a été vérifiée et jugée juste pour l'utilisation qui en serait faire par la suite.

Dans cette étude, l'effet important de l'âge et de l'année sur l'incidence a justifié l'utilisation des taux d'incidence de la France entière par classe d'âge de 1 an et par année comme taux de référence pour estimer les nombres de cas attendus chaque année, pour chaque classe d'âge.

Différentes sous-populations ont été utilisées pour étudier la cartographie des LA en France métropolitaine: par sexe, tranche d'âge de 5 ans : 0-4ans/5-9ans/10-14ans, et sous-périodes de 5 années 1990-1994/1995-1999/2000-2004.

C Méthodes

C.1 Cartographie "brute"

Les méthodes d'inférence sur les risques relatifs sont fondées sur la modélisation classique d'apparition d'événements rares par une loi de Poisson : $y_i \sim$ Poisson $(E_i \lambda_i)$ indépendamment sur chaque département *i,* i=1,…n=94 où n est le nombre de départements y_i le nombre de cas incidents observés dans le département i, E_i le nombre de cas attendus dans ce département après ajustement sur le sexe et l'âge par tranche de 1 an et λ_i le risque relatif d'incidence propre à chaque département i. Le risque relatif est estimé par le SIR "local" du département i : $SIR_i = y_i / E_i$ indépendamment pour chaque département.

C.1.a Tests de signification des risques relatifs

Nous avons testé indépendamment pour chaque département la signification de l'écart de chaque *SIR* à 1 par le test exact de Poisson, et calculé l'intervalle de

confiance à 90% du risque relatif [Estève *et al.*,1993]. Pour tenir compte de la multiplicité des tests, nous avons utilisé la correction de Bonferroni.

C.1.b Test d'hétérogénéité globale des risques relatifs

Trois tests ont été utilisés pour tester l'hétérogénéité globale des SIR, sur tout le territoire globalement, sans rechercher de localisation dans des départements particuliers, pour lesquels l'hypothèse nulle est l'égalité des *n* SIR à 1 (homogénéité) : le test du rapport de vraisemblance et celui du Khi-2 de Pearson et pour tester la présence d'une sur-dispersion extra-Poissonnienne le test de Potthoff et Whittinghill.

C.2 Cartographie "lissée"

Dans ce paragraphe, les méthodes sont décrites en fonction du problème posé, c'est-à-dire, la France métropolitaine découpée en 94 départements.

Les méthodes de lissage utilisées ici font appel à une formulation hiérarchique bayésienne, habituellement prise pour modéliser des données rares dans l'espace, le premier niveau de hiérarchie représentant la variabilité locale (ici intra département) modélisée par une loi de Poisson. La variabilité des risques relatifs entre les départements est représentée au second niveau par un modèle introduisant une dépendance spatiale sur l'ensemble des *n* risques relatifs $\boldsymbol{\lambda}=(\lambda_i,\ i=1,\ldots n)$.

Dans un cadre d'inférence bayésienne, ces méthodes de lissage permettent de minimiser l'erreur quadratique moyenne en faisant différentes hypothèses *a priori*, sur la structure sous-jacente de la variabilité géographique des risques relatifs à travers différents modèles paramétriques (BYM et CAR) ou non paramétriques (KHR

et MIX) et sur l'amplitude de cette variabilité. Nous avons considéré qu'elle se situait raisonnablement en moyenne entre [0,8 – 1,25] et au maximum entre [0,25 – 4,0].

C.2.a Modèles d'estimation des risques relatifs

Modèle BYM

Dans le modèle BYM [Besag *et al.*,1991] le logarithme du risque relatif "exact" est représenté comme la somme de deux composantes : l'une modélisant la variation hétérogène (représentant la sur-dispersion extra-Poissonnienne, les erreurs de mesure, les covariables non spatialement structurées) et l'autre spatialement structurée, modélisant les variations spatiales locales entre départements voisins:

$ln(\boldsymbol{\lambda}) = \boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$

où $\boldsymbol{v}$ est la composante hétérogène, un bruit gaussien d'espérance nulle et de variance τ_v^2 et $\boldsymbol{u}$ est la composante spatialement structurée, un processus autorégessif gaussien de variance τ_u^2.

Pour un département donné, le processus $\boldsymbol{u}$ a pour espérance conditionnelle la moyenne de ses départements voisins et une variance conditionnelle inversement proportionnelle au nombre de ses voisins.

Dans une optique bayésienne les deux paramètres de variance sont supposés suivre des distributions hyper *a priori*, qui classiquement sont des lois Gamma sur les inverses des variances τ_u^2 et de τ_v^2.

Les estimateurs des risques obtenus avec ce modèle sont le résultat d'un lissage des risques observés soit vers une moyenne globale en l'absence de dépendance

spatiale, soit vers une moyenne locale calculée à partir des risques des départements voisins, lorsqu'une dépendance spatiale est présente.

Modèle CAR

Le modèle CAR [Stern &Cressie,2000] est une adaptation des modèles conditionnels autorégressifs classiques développés initialement pour des domaines réguliers, à un domaine irrégulier en tenant compte des risques relatifs des départements voisins et de leurs précisions relatives à travers leurs nombres de cas attendus.

Il est défini comme un processus autorégressif conditionnel gaussien sur $\boldsymbol{x}=ln(\boldsymbol{\lambda})$ avec trois paramètres : μ la moyenne globale, ϕ un paramètre d'autocorrélation variant de $-0{,}3365$ à 0.1717 pour le voisinage des départements français et τ^2 un paramètre de variance.

On choisira comme distributions hyper *a priori* pour μ et ϕ des lois uniformes sur leurs ensembles de définition respectifs : $]-\infty ; +\infty[$ et $]-0{,}3365 ; +0{,}1717[$, et pour l'inverse de la variance τ^2 une loi Gamma.

Modèle MIX

Pour sortir du cadre paramétrique strict, le modèle MIX [Green &Richardson,2002] utilise des modèles de mélange sur la structure des risques relatifs $\lambda_i = \lambda_{z_i}$ où (λ_j, j=1,… k) caractérisent k différentes composantes et (z_i, i=1,… n) sont des variables d'allocation qui associent à chaque département une composante (de 1 à k).

On suppose ainsi l'existence de k niveaux de risque ou « composantes », k n'étant pas fixé. Les différents départements sont alloués de manière stochastique à ces catégories par la variable d'allocation z_i.

La dépendance spatiale est intégrée au niveau de ces variables d'allocations (elle intervient de manière plus locale) par l'intermédiaire du modèle de Potts via un paramètre d'interaction ψ qui sera estimé. La probabilité d'allocation d'un département *i* à une catégorie de risque *j* prend en compte ψ et le nombre de départements voisins appartenant à cette même catégorie.

Ce mécanisme d'allocation ne tient pas seulement compte d'un *a priori* sur les similarités spatiales des risques des départements voisins, mais permet également à des départements non contigus d'appartenir à la même composante.

On suppose *a priori* que les composantes (λ_j, j=1,… k) suivent des lois Gamma indépendantes de moyenne 1, que ψ suit une loi discrète uniforme sur {0 ; 0,1 ; 0,2 ; … ; ψ_{max}}

où ψ_{max} =1,0 correspond à un haut niveau d'interaction sur le voisinage des départements français, et que k suit une loi uniforme sur {1, 2, …, k_{max}} où k_{max} = 10 semble être en général suffisamment flexible. Dans notre cas, puisque k_{max} devrait refléter la variabilité *a priori* du risque, 10 composantes au maximum devraient être largement suffisantes pour décrire les différents niveaux de risque des leucémies aigües de l'enfant sur les départements français.

Modèle KHR

Le modèle KHR [Knorr-Held *et al.*,2000] est un modèle de partition qui suppose que les logarithmes des risques relatifs $\boldsymbol{x}$ sont indépendants et identiquement distribués suivant une loi normale de moyenne μ et de variance σ^2 avec $x_i = ln(\lambda_{z_i})$

où (λ_j, j=1,… k) caractérisent k différents niveaux de risque, $ln(\lambda_j)$ est normale de moyenne μ et de variance σ^2 et (z_i, i=1,… n) sont des variables d'allocation qui associent à chaque département un niveau de risque (de 1 à k).

Pour prendre en compte de manière non paramétrique une information *a priori* spatiale, le modèle suppose que les départements contigus de même risque relatif peuvent être agrégés en clusters, le nombre *k* de ces clusters n'étant pas fixé.

Parmi les départements, *k* sont tirés au sort, chacun d'eux définissant le "centre d'un cluster" *j*, *j*=1,…*k*. Les départements restants sont alloués au cluster *j* si le centre de ce cluster est plus proche du département que tous les autres centres, la distance considérée étant le nombre minimum de frontières à franchir entre le département et le centre d'un cluster.

Les lois hyper *a priori* choisies sont pour μ une loi uniforme sur $]-\infty ; +\infty[$, pour *k* une loi géométrique tronquée de paramètre 0,02 (correspondant à un nombre de clusters *a priori* égal à 50, c'est-à-dire approximativement la moitié du nombre de départements), et pour l'inverse de la variance σ^2 une loi Gamma.

C.2.b Lois hyper a priori

L'intensité du lissage obtenu dépend du choix des paramètres des lois hyper *a priori* *[*Mollié,2000*]*. Les paramètres des différents modèles intervenant à des niveaux différents dans les modèles hiérarchiques, les effets de ces hyper paramètres ne sont pas directement comparables.

Nous avons choisi ces hyper paramètres (Tableau 12) de manière à ce qu'*a priori* les degrés de lissage soient semblables pour les quatre modèles. Pour cela nous avons considéré que pour l'étude des leucémies aigües, l'intervalle de variation à 95% du risque relatif pouvait raisonnablement varier en moyenne de 0,80 à 1,25 et à 5% de 0,25 à 4.

Pour les modèles KHR et MIX, le calcul des intervalles de fluctuation *a priori* ne dépend ni de la structure de voisinage, ni des nombres attendus, pour le modèle

BYM il ne dépend que de la structure de voisinage considérée, ici celle des départements français : pour ces trois modèles le choix *a priori* peut donc être constant quelle que soit la pathologie étudiée.

En revanche pour le modèle CAR, le calcul fait intervenir la moyenne (harmonique) des nombres attendus *mh*(***E***) : le choix *a priori* dépendra donc de la rareté du sous-type de leucémie pour obtenir *a priori* un intervalle de fluctuation à 95% des risques similaire.

Pour le modèle MIX, un choix simple qui satisfait les conditions de marges de variations du risque relatif est a=b=100.

Pour chacun des trois modèles KHR, BYM et CAR, l'intensité du lissage des variations aléatoires dépend de la moyenne et de la variance de la loi Gamma hyper *a priori* sur les inverses des variances : plus la moyenne est élevée ou la variance faible, plus le lissage est important. Nous avons choisi des lois Gamma pour lesquelles l'écart-type était égal à la moyenne, donc la dispersion relative (coefficient de variation *a*) est égale à 1.

Pour le choix du paramètre de dispersion *b*, nous avons utilisé deux estimations des intervalles de fluctuation à 95% des risques relatifs *a priori* : l'un fondé sur la moyenne du paramètre de la loi Γ qui donne un intervalle moyen *a priori* et l'autre dérivé du 97,5^{ème} percentile de la même loi pour obtenir un intervalle *a priori* à 95%. Pour les modèles KHR et BYM nous avons choisi la valeur *b*=0,01.

Pour le modèle CAR, les valeurs du paramètre de dispersion *b* de la loi Gamma varient, elles sont résumées dans le Tableau 12 .Pour les localisations fréquentes, toutes LA et LAL (*mh*(***E***) compris entre 30 et 40) *b*=0,5, pour les moins fréquentes, LA et LAL par période ou par sexe ou pour la classe d'âge de 0 à 4 ans et LA pour la classe d'âge de 5 à 9 ans (*mh*(***E***) compris entre 10 et 22) *b*=0,2 ; pour les plus rares,

LAM, LA pour la classe d'âge 10-14 ans et LAL pour les classes d'âge 5-9 ans et 10-14 ans (*mh*(***E***) compris entre 6 et 10) *b*=0,1 ; et enfin pour les très rares LAM par sexe et par période (*mh*(***E***) compris entre 2 et 4) *b*=0,04.

Tableau 12 : Choix des distributions hyper a priori dans le modèle CAR pour l'inverse de la variance

		$\Gamma(1,b)$ [c]		A priori à 95% de l'intervalle de fluctuation $I_{95\%}$	
	$hm_b(E)$	*b*	moyenne	Étendue moyenne [d]	95% [d]
LA[a]	39,5	0,5	2	[0,80 ; 1,25]	[0,25 ; 4,00]
LAL[a]	32,1			[0,78 ; 1,28]	[0,21 ; 4,65]
LA 1990-1994	13,1	0,2	5	[0,78 ; 1,27]	[0,22 ; 4,58]
LA 1995-1999	12,9			[0,78 ; 1,28]	[0,22 ; 4,64]
LA 2000-2004	13,5			[0,79 ; 1,27]	[0,22 ; 4,48]
LA garçons	21,9			[0,83 ; 1,21]	[0,31 ; 3,25]
LA filles	17,6			[0,81 ; 1,23]	[0,27 ; 3,72]
LA 0-4 ans	19,3			[0,82 ; 1,22]	[0,29 ; 3,50]
LA 5-9 ans	11,8			[0,77 ; 1,29]	[0,20 ; 4,97]
LAL 1990-1994	10,7			[0,76 ; 1,31]	[0,19 ; 5,39]
LAL 1995-1999	10,5			[0,76 ; 1,31]	[0,18 ; 5,47]
LAL 2000-2004	10,9			[0,77 ; 1,30]	[0,19 ; 5,30]
LAL garçons	18,1			[0,81 ; 1,23]	[0,27 ; 3,65]
LAL filles	14,0			[0,79 ; 1,26]	[0,23 ; 4,36]
LAL 0-4 ans	16,0			[0,80 ; 1,24]	[0,25 ; 3,96]
LA 10-14 ans	8,4	0,1	10	[0,81 ; 1,24]	[0,26 ; 3,83]
LAL 5-9 ans	9,9			[0,82 ; 1,22]	[0,29 ; 3,45]
LAL 10-14 ans	6,2			[0,78 ; 1,28]	[0,21 ; 4,78]
LAM[a]	6,7			[0,79 ; 1,27]	[0,22 ; 4,50]
LAM 1990-1994	2,2	0,04	25	[0,77 ; 1,30]	[0,19 ; 5,26]
LAM 1995-1999	2,2			[0,77 ; 1,30]	[0,19 ; 5,26]
LAM 2000-2004	2,3			[0,77 ; 1,29]	[0,20 ; 5,08]
LAM garçons	3,4			[0,81 ; 1,24]	[0,26 ; 3,80]
LAM filles	3,3			[0,81 ; 1,24]	[0,26 ; 3,88]

[a] LA: leucémie aigüe; LAL: leucémie aigüe lymphoblastique; LAM: leucémie aigüe myéloblastique
[b] *hm(E)* moyenne du nombre de cas attendus
[c] $\Gamma(1,b)$ distribution Gamma de paramètres de forme 1 et de dispersion *b* (écart-type= moyenne)
[d] Etendue moyenne et étendue à 95% sont les intervalles de fluctuations à 95% a priori du risque relatif dérivés de la moyenne et du quantile 97,5% de la loi Gamma respectivement.

C.2.c Ajustement des modèles aux observations

La technique d'ajustement de ces modèles aux observations fait appel à des algorithmes MCMC permettant de simuler la distribution *a posteriori* des paramètres et des risques relatifs. Le risque relatif d'un département est alors estimé par la

moyenne *a posteriori* des risques simulés pour ce département. Nous avons également utilisé les intervalles de crédibilité à 90% du risque relatif et la probabilité pour que le risque relatif soit supérieur à 1, estimés à partir de cette distribution *a posteriori*. Le critère *Deviance Information Criterion* DIC [Spiegelhalter *et al.*,2002] a permis de comparer la complexité et la qualité des ajustements des différents modèles hiérarchiques bayésiens. Ce critère est la somme de la moyenne *a posteriori* de la déviance D_, qui reflète l'ajustement du modèle et d'une estimation du nombre effectif de paramètres p_D, qui peut être considéré comme un terme de pénalité reflétant la complexité du modèle.

Le logiciel Winbugs a été utilisé pour les estimations des modèles BYM et CAR [Lunn *et al.*,2000]. Un algorithme MCMC a permis d'estimations les différents paramètres sur 3 000 itérations : une sur 10 des 30 000 itérations faites, après un temps de chauffe de 1000 à 3000 itérations.

L'estimation des modèles MIX et KHR fait appel à des algorithmes MCMC à sauts réversibles [Green,1995]. Les codes Fortran et C permettant l'utilisation du modèle MIX ont été gracieusement mis à disposition par leurs auteurs. Le programme *Bayesian Detection of Clusters and Discontinuities in Disease Maps* BDCD [Knorr-Held *et al.*,2000] a permis l'estimation des paramètres du modèle KHR. 500 000 itérations ont été générées pour estimer les paramètres du modèle MIX, après un temps de chauffe de 20 000 itérations. 5 000 000 itérations ont été générées pour le modèle KHR après un temps de chauffe de 100 000 itérations, (en prenant 2 500 itérations pour l'estimation des paramètres, une toutes les 2 000 itérations).

C.2.d Corrélations entre SIR lissés

Les *n* observations obtenues pour chaque modèle n'étant pas indépendantes, le test de Student modifié [Clifford *et al.*,1989] a été utilisé pour tester la corrélation spatiale observée entre les estimations obtenues par sexe, par groupe d'âge ou par groupe de périodes. Ce test permet de corriger le nombre de degrés de liberté du test classique en considérant l'autocorrélation spatiale des estimations lissées. Nous avons pris en compte les autocorrélations des ordres 1 à 5 estimées à partir des matrices de voisinage d'ordre 1 à 5 [Cliff &Ord,1981].

D Résultats

Vingt-quatre cartes ont été examinées (par chacune des 5 approches : SIR bruts, SIR respectivement lissés par les modèles BYM, CAR, MIX et KHR) présentant les variations d'incidence de toutes les LA, des LAL et des LAM, globalement, ou par période (1990-94, 1995-99, 2000-04), par classe d'âge (3 classes : 0-4 ans, 5-9 ans, 10-14 ans) et par sexe.

La Corse n'a pas été prise en compte dans cette étude car tous les modèles de lissage utilisés ne pouvaient pas l'inclure du fait de l'absence de tout département voisin. Pour vérification, les SIR des tous les départements ont tout de même été calculés en prenant en compte la Corse (considérée comme un seul département) dans le calcul du taux d'incidence France entière. Le SIR de ce département est alors de 1,11, non significativement différent de 1 et les SIR bruts des autres départements ne sont pas différents de ceux calculés sans tenir compte de la Corse.

D.1 Cartographie "brute"

D.1.a Hétérogénéité des risques relatifs

Les trois tests d'hétérogénéité (rapport de vraisemblance, Khi-2 et Potthoff et Witthingill) donnent des résultats concordants très proches et non significatifs. Sur les différentes sous-populations étudiées, une sur-dispersion Poissonnienne de l'ordre de 20% à 30% est observée lorsque l'hétérogénéité globale est à la limite de la signification (p entre 0,06 et 0,10), mais aucune stabilité dans ces résultats n'a été observée par sous-population (Tableau 13).

Tableau 13 : Hétérogénéité des SIR bruts par sous-population

	$\hat{\beta}$ [a]	*p* [b]	nb cas[c]	*mh(E)*[d]	*étendue(E)* [d]		*étendue(SIR*[e]*)*	
Leucémie aigüe								
LA	**0,19**	*0,10*	6796	39,5	[7,3;	340,6]	[0,5;	1,5]
LA 1990-1994	0,05	0,33	2235	13,1	[2,4;	115,6]	[0,4;	2,0]
LA 1995-1999	**0,22**	*0,07*	2218	12,9	[2,4;	111,0]	[0,0;	2,0]
LA 2000-2004	0,16	0,13	2343	13,5	[2,5;	113,8]	[0,5;	1,8]
LA garçons	0,04	0,35	3766	21,9	[4,1;	188,9]	[0,2;	1,5]
LA filles	0,04	0,37	3030	17,6	[3,2;	151,6]	[0,3;	2,1]
LA 0-4 ans	0,11	0,21	3408	19,3	[3,5;	169,5]	[0,5;	1,9]
LA 5-9 ans	0,13	0,17	1999	11,8	[2,2;	100,5]	[0,3;	1,9]
LA 10-14 ans	**0,24**	*0,06*	1389	8,4	[1,60;	70,5]	[0,0;	2,1]
Leucémie aigüe lymphoblastique								
LAL	**0,23**	*0,06*	5526	32,1	[5,9;	276,9]	[0,5;	1,5]
LAL 1990-1994	–0,03	0,57	1828	10,7	[2,0;	94,6]	[0,3;	1,9]
LAL 1995-1999	0,14	0,15	1805	10,5	[1,9;	90,2]	[0,0;	2,3]
LAL 2000-2004	**0,20**	*0,09*	1893	10,9	[2,0;	91,9]	[0,3;	1,9]
LAL garcons	0,01	0,45	3113	18,1	[3,4;	156,2]	[0,0;	1,5]
LAL filles	0,13	0,17	2413	14,0	[2,6;	120,7]	[0,3;	2,2]
LAL 0-4 ans	–0,01	0,49	2808	16,0	[2,9;	139,7]	[0,3;	2,0]
LAL 5-9 ans	**0,20**	*0,08*	1690	9,9	[1,8;	85,0]	[0,1;	2,1]
LAL 10-14 ans	0,08	0,25	1028	6,2	[1,2;	52,2]	[0,0;	2,1]
Leucémie aigüe myéloblastique								
LAM	0,06	0,30	1160	6,7	[1,2;	58,2]	[0,0;	2,6]
LAM 1990-1994	**0,28**	*0,03*	377	2,2	[0,4;	19,4]	[0,0;	5,4]
LAM 1995-1999	–0,002	0,46	384	2,2	[0,4;	19,3]	[0,0;	3,3]
LAM 2000-2004	–0,02	0,52	399	2,3	[0,4;	19,4]	[0,0;	3,3]
LAM garçons	0,11	0,22	586	3,4	[0,6;	29,4]	[0,0;	2,8]
LAM filles	–0,04	0,57	574	3,3	[0,6;	28,8]	[0,0;	3,0]

n= 94, le nombre de départements; y_i nombre observé de cas incidents et E_i le nombre de cas attendus incidents dans le département *i*

[a] le coefficient d'hétérogénéité β est défini par $1+\beta=\frac{\mathrm{var}(y_i)}{E(y_i)}$ avec *i*=1,… *n*

[b] *p* = degré de significativité statistique du test de Potthoff et Whittinghill de l'hypothèse β = 0

[c] nombre total de cas y_+

[d] *mh(E)* moyenne harmonique; *étendue (E)* étendue du nombre de cas attendus

[e] $SIR_i = y_i / E_i$

Pour les SIR des LA, on observe une sur-dispersion de 18,6% sur l'ensemble des cas (tests d'hétérogénéité p=0,10), une sur-dispersion de 22,2% pour les SIR de la seule période 1995-99 (p=0,07), enfin une sur-dispersion de 24,3% pour la classe d'âge 10-14 ans ($IC_{90\%}$=[0,2%; 48,4%] p= 0,06).

Les LAL ont une sur-dispersion de 23,5% (p=0,06) globalement, une sur-dispersion de 19,9% (p=0,09) pour la période 2000-2004 et une sur-dispersion de 20,2% (p=0,08) pour la classe d'âge 5-9 ans. Pour les LAM seules les SIR de la période 1990-94 ont une sur-dispersion notable de 27,9% ($IC_{90\%}$= [3,8%; 52,1%]), (p=0,04).

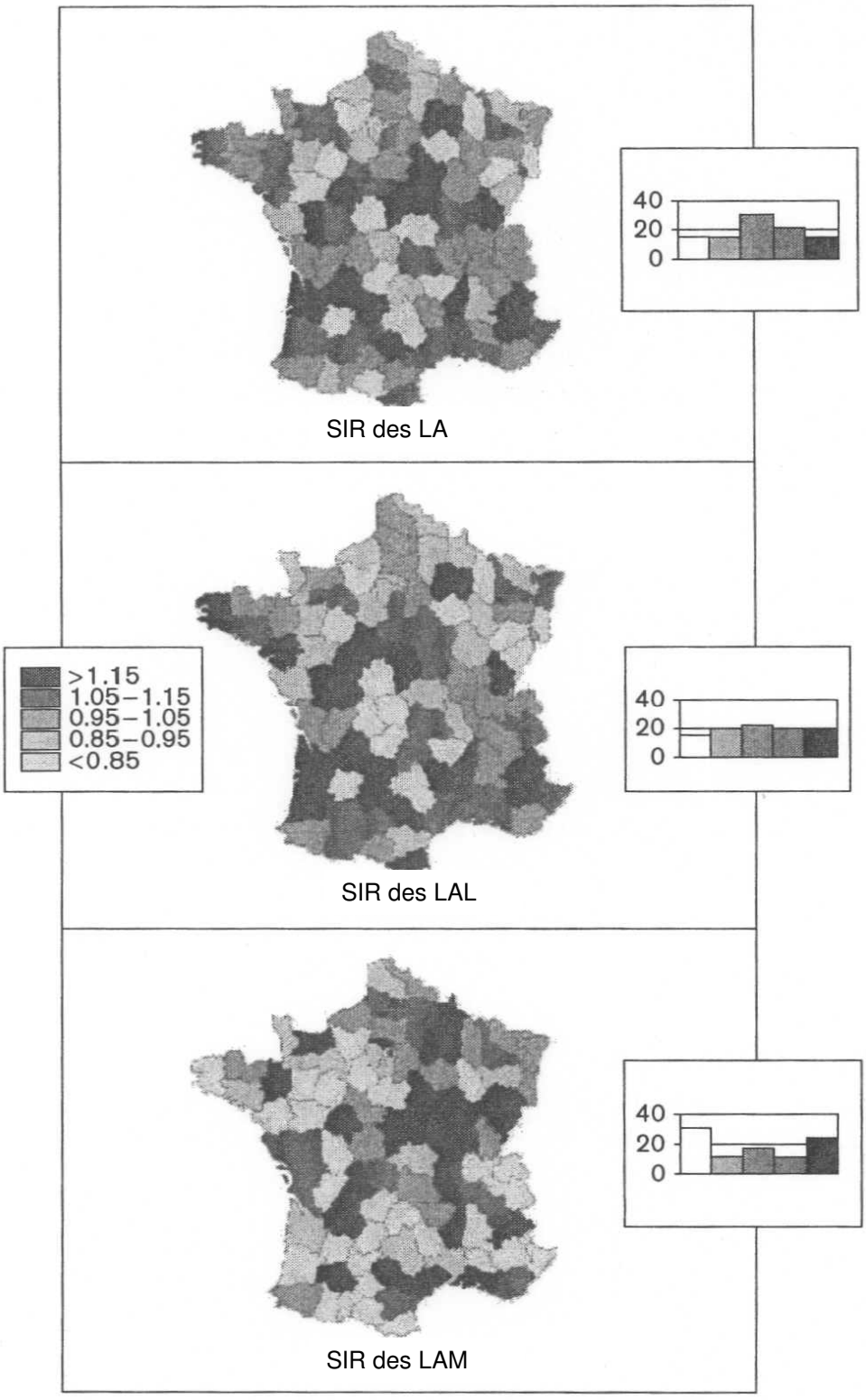
D.1.b Cartes des SIR

Les cartes des SIR ne révèlent pas l'existence d'une structure globale, ni globalement, ni par sous-type de LA (Figure 5), ni par période ou par sexe.

D.1.c Tests de signification des SIR

Nous n'avons pas trouvé d'excès localement significatif qui présente une stabilité par type de LA, période, âge et sexe. Aucun excès n'est localement significatif en tenant compte de la multiplicité des tests.

Figure 5 : Carte des SIR bruts pour les LA, LAL et LAM



Pour les LA (Tableau 14) parmi les six SIR significativement supérieurs à 1 (Bouches-du-Rhône n°13, Dordogne n°24, Gironde n°33, Indre-et-Loire n°37, Jura n°39, Tarn n°81) : deux seulement se retrouvent sur deux périodes (département n°24 périodes 1995-99 et 2000-04 et département n°39 périodes 1990-94 et 2000-04), quatre (départements n°33, 37, 39, 81) se retrouvent pour les LAL et un (département n°81) pour les classes d'âge 5-9 et 10-14 ans. Les SIR du département n°39 sont significativement supérieurs à 1 chez les filles et les garçons.

D'autre part les sept SIR significativement inférieurs à 1 (Aveyron n°12, Corrèze n°19, Eure n°27, Lot-et-Garonne n°47, Haut-Rhin n°68, Sarthe n°72, Seine-Saint-Denis n°93) le sont aussi pour les seules LAL, seul un département (n°72) l'est pour deux classes d'âge (0-4 et 5-9 ans) et aucun n'est stable par sous-période.

Pour les LAL, aucun des sept SIR significativement supérieurs à 1 ne l'est sur plus d'une sous-période ou chez les filles et les garçons simultanément. Seul celui du département de la Gironde (n°33) se retrouve pour les classes d'âge 0-4 et 10-14 ans. D'autre part, parmi les huit SIR significativement inférieurs à 1, seul celui du département n°72 l'est également sur deux périodes (1995-99 et 2000-04).

Pour les LAM, aucun des cinq SIR significativement supérieurs à 1 et aucun des trois SIR significativement inférieurs à 1 ne le sont chez les filles et les garçons simultanément, et aucun ne l'est sur plus d'une sous-période.

Aucune structuration spatiale ne s'avérait finalement présenter une quelconque stabilité par sexe, par âge, ou par période.

D.1.d Corrélations des SIR entre types histologiques, périodes, sexe, classes d'âge

Les SIR des LA sont très corrélés avec ceux des LAL (r =0,89 ; $p<10^{-5}$) et ceux des LAM (r = 0,340 ; $p<10^{-4}$). Les SIR des LAL et des LAM ne sont pas corrélés entre eux (r = –0,04 ; ns). On retrouve des résultats similaires plus affirmés chez les filles (LA-LAL r = 0,89 ; LA-LAM r = 0,42) et chez les garçons (LA-LAL r = 0,91 ; LA-LAM r = 0,522).

Aucune corrélation significative n'est observée entre les SIR des trois périodes, entre ceux des filles et des garçons, ou entre ceux des trois classes d'âge ni pour les LA, ni par sous-type de LA.

Tableau 14 : SIR des leucémies aigües (1990-2004, garçons et filles, 0-14 ans) par département

département *i*	y_i[a]	E_i[b]	SIR_i[c]		$IC_{90\%}(\theta_i)$[d]	
01 – Ain	63	65,9	0,96		[0,77	1,18]
02 – Aisne	63	68,0	0,93		[0,74	1,14]
03 – Allier	28	33,3	0,84		[0,60	1,15]
04 - Alpes-de Haute-Provence	19	15,2	1,25		[0,82	1,83]
05-Hautes-	20	13,6	1,47		[0,97	2,13]
06 - Alpes-Maritimes	110	101,5	1,08		[0,92	1,27]
07 – Ardèche	37	31,1	1,19		[0,89	1,56]
08 – Ardennes	38	36,5	1,04		[0,78	1,36]
09 – Ariège	11	13,4	0,82		[0,46	1,36]
10 – Aube	28	34,2	0,82		[0,58	1,12]
11 – Aude	30	31,5	0,95		[0,68	1,29]
12 – Aveyron	16	25,3	**0,63**	*ns	[0,40	0,96]
13 – Bouches-du-Rhône	236	208,4	**1,13**	*ns	[1,01	1,26]
14 – Calvados	84	77,3	1,09		[0,90	1,30]
15 – Cantal	13	14,0	0,93		[0,55	1,47]
16 – Charente	34	35,4	0,96		[0,71	1,28]
17 – Charente-Maritime	58	56,5	1,03		[0,82	1,28]
18 – Cher	42	33,5	1,25		[0,95	1,62]
19 – Corrèze	11	21,0	**0,52**	*ns	[0,29	0,87]
21 - Côte-d'Or	59	57,2	1,03		[0,82	1,28]
22 - Côtes-[illegible]	60	58,0	1,03		[0,82	1,28]
23 – Creuse	10	10,7	0,94		[0,51	1,59]
24 – Dordogne	49	36,4	**1,35**	*ns	[1,05	1,71]
25 – Doubs	55	60,0	0,92		[0,72	1,15]
26 – Drôme	49	52,3	0,94		[0,73	1,19]
27 – Eure	51	70,2	**0,73**	**ns	[0,57	0,92]
28 - Eure-et-	46	51,7	0,89		[0,66	1,14]
29 – Finistère	104	94,2	1,10		[0,93	1,30]
30 – Gard	78	69,8	1,12		[0,92	1,35]
31 - Haute-Garonne	111	112,1	0,99		[0,84	1,16]

département *i*	y_i	E_i	SIR_i		$IC_{90\%}(\theta_i)$	
32 – Gers	18	16,1	1,1		[0,72	1,66]
33 – Gironde	167	138,6	**1,2**	*ns	[1,06	1,37]
34 – Hérault	102	95,2	1,0		[0,90	1,26]
35 - Ille-et-Vilaine	109	102,4	1,0		[0,90	1,25]
36 - Indre	19	22,6	0,8		[0,55	1,24]
37 - Indre-et-	77	61,1	**1,2**	*ns	[1,03	1,52]
38 - Isère	134	134,3	1,0		[0,86	1,15]
39 - Jura	45	29,3	**1,5**	**ns	[1,18	1,97]
40 - Landes	35	33,0	1,0		[0,78	1,40]
41 - Loir-et-Cher	39	34,3	1,1		[0,86	1,49]
42 - Loire	83	83,7	0,9		[0,82	1,19]
43 - Haute-Loire	18	22,7	0,7		[0,51	1,18]
44 - Loire-Atlantique	143	136,1	1,05		[0,91	1,21]
45 - Loiret	73	73,9	0,9		[0,81	1,20]
46 - Lot	21	15,1	1,3		[0,93	2,00]
47 - Lot-et-Garonne	18	31,2	**0,58**	**ns	[0,37	0,86]
48 - Lozère	7	7,3	0,9		[0,45	1,81]
49 - Maine-et-	75	90,1	0,8		[0,68	1,01]
50 - Manche	57	57,1	1,0		[0,79	1,24]
51 - Marne	79	67,3	1,1		[0,97	1,42]
52 - Haute-Marne	22	22,9	0,9		[0,65	1,37]
53 - Mayenne	31	34,8	0,8		[0,64	1,20]
54 - Meurthe-et-Moselle	95	82,9	1,15		[0,96	1,36]
55 - Meuse	18	23,2	0,7		[0,50	1,15]
56 - Morbihan	76	72,8	1,0		[0,86	1,26]
57 - Moselle	112	121,8	0,9		[0,78	1,08]
58 - Nièvre	27	22,4	1,2		[0,85	1,67]
59 - Nord	310	340,6	0,9		[0,87	1,00]
60 - Oise	104	103,2	1,0		[0,85	1,19]
61 - Orne	36	33,8	1,0		[0,79	1,41]
62 - Pas-de-	183	189,0	0,9		[0,85	1,09]

département i	y_i	E_i	SIR_i		$CI_{90\%}(\theta_i)$	
63 - Puy-de-Dôme	65	61,3	1,06		[0,85	1,30]
64- Pyrénées-	63	61,0	1,03		[0,83	1,27]
65 - Hautes-	19	21,2	0,89		[0,56	1,31]
66 - Pyrénées-	43	39,3	1,09		[0,83	1,41]
67 - Bas-Rhin	124	119,	1,04		[0,89	1,20]
68 - Haut-Rhin	69	85,4	**0,81**	*/ns	[0,66	0,99]
69 - Rhône	197	190,	1,03		[0,91	1,16]
70 - Haute-Saône	22	26,9	0,82		[0,55	1,17]
71 - Saône-et-Loire	64	58,7	1,09		[0,88	1,34]
72 - Sarthe	43	62,0	**0,69**	**ns	[0,53	0,89]
73 - Savoie	43	43,2	0,99		[0,76	1,28]
74 - Haute-Savoie	76	79,8	0,95		[0,78	1,15]
75 - Paris	189	190,	0,99		[0,88	1,12]
76 - Seine-Maritime	137	153,	0,89		[0,77	1,03]
77 - Seine-et-	158	164,	0,96		[0,84	1,10]
78 - Yvelines	168	182,	0,92		[0,81	1,05]
79 - Deux-Sèvres	44	37,9	1,16		[0,89	1,49]
80 - Somme	70	66,4	1,05		[0,86	1,29]
81 - Tarn	47	34,8	**1,35**	*ns	[1,04	1,72]
82 - Tarn-et-	24	22,2	1,08		[0,75	1,52]
83 - Var	99	96,5	1,03		[0,86	1,21]
84 - Vaucluse	59	59,2	1,00		[0,79	1,24]
85 - Vendée	57	60,6	0,94		[0,75	1,17]
86 - Vienne	44	41,2	1,07		[0,82	1,37]
87 - Haute-Vienne	34	32,4	1,05		[0,77	1,40]
88 - Vosges	45	45,1	1,00		[0,77	1,28]
89 - Yonne	47	37,9	1,24		[0,96	1,58]
90 - Territoire-de-	18	16,5	1,09		[0,71	1,62]
91 - Essonne	161	150,	1,07		[0,94	1,22]
92 - Hauts-de-	151	170,	0,89		[0,77	1,01]
93 - Seine-Saint-Denis	161	192,	**0,84**	*ns	[0,73	0,95]
94 - Val-de-Marne	158	151,	1,04		[0,91	1,19]
95 - Val-d'Oise	151	154,	0,98		[0,85	1,12]

[a] y_i nombre de cas incidents observés dans le département *i* [b] E_i nombre de cas incidents attendus dans le département *i* [c] $SIR_i = y_i / E_i$ SIR dans le département *i*

*ns / **ns *:* degré de significativité statistique *p* du test exact de Poisson <5% / <1% mais non significatif après correction de Bonferroni

[d] $IC_{90\%}(\theta_i)$ 90% intervalle de confiance du risque relatif θ_i dans le département *i* (basé sur la distribution du test exact de Poisson)

D.2 Cartographie "lissée"

D.2.a Adéquation des modèles de lissage

Aucun des quatre modèles de lissage ne conduit à une adéquation entre SIR estimés et observés majoritairement meilleure que les trois autres sur les 24 cartes examinées. On peut cependant observer que les modèles non paramétriques KHR et MIX présentent plus fréquemment le critère DIC le plus faible, respectivement 9 et 8 fois sur 24.

Le DIC reste toujours inférieur à 188 qui est la valeur correspondant au modèle saturé basé sur la vraisemblance de Poisson. Il est généralement inférieur à *L*, sa valeur correspondant au modèle "poolé" avec tous les risques relatifs égaux à 1; l'amélioration du DIC est de 0,5 à 8,2. Cependant, le DIC est légèrement meilleur (0,2 à 3,3) que *L*, généralement quand le résultat du test d'hétérogénéité globale (test du rapport de vraisemblance) n'était pas significatif et/ou quand les données étaient très rares: LAL et les LAM par sous-population. En ce qui concerne la performance des modèles en fonction du sous-type de leucémie, le modèle CAR semble obtenir le plus petit DIC spécialement pour les LA, le modèle MIX pour les LAM et le modèle KHR pour les LAL. Pour le modèle KHR, le DIC ne diffère jamais de plus de 5 du DIC minimum ce qui n'est jamais le cas des trois autres modèles. Pour le modèle MIX, le critère DIC diffère 3 fois de plus de 5 du DIC minimum (jusqu'à 7,5). En revanche, les modèles paramétriques présentent des DIC moins souvent minimaux (BYM 1 fois, CAR 6 fois) et plus souvent supérieurs de plus de 5 au DIC minimal (BYM 4 fois, CAR 4 fois).

Les corrélations entre les SIR et les estimations des modèles varient de 0,35 à 0,99 : dans tous les cas, elles sont les plus fortes avec le modèle CAR (de 0,92 à 0,99) puis avec le modèle MIX (coefficient de corrélation de 0,66 à 0,81) et le modèle BYM (coefficient de corrélation de 0,60 à 0,80), elles diminuent avec le modèle CLUS (coefficient de corrélation de 0,35 à 0,74).

On obtient de bonnes corrélations entre les estimations par les différents modèles : les estimations du modèle BYM sont les plus corrélées avec celles des deux modèles non paramétriques, le modèle MIX (0,94 à 0,98), et le modèle KHR (de 0,85 à 0,97), les corrélations les plus faibles sont observées entre le modèle CAR et les deux modèles non paramétriques MIX (de 0,52 à 0,83) et KHR (0,23 à 0,69). Les estimations données par les deux modèles non paramétriques KHR et MIX ont des corrélations fortes (de 0,76 à 0,93).

D.2.b Lissage des SIR

Il est difficile de déceler des différences stables entre les cartes fournies par les quatre modèles, d'une part parce qu'elles sont toutes très peu contrastées, et d'autre part parce qu'elles sont le plus souvent fortement corrélées entre elles. On observe que lorsqu'une structure semble apparaître dans l'une des 24 cartes examinées on la retrouve le plus souvent avec les autres modèles pour la même sous-population, sauf avec la méthode CAR. Pour cette dernière méthode, les SIR sont moins corrélés à ceux obtenus avec les autres modèles, ne font jamais apparaître de structure et sont les plus corrélée aux SIR bruts.

Quel que soit le modèle, l'étendue des estimations est très restreinte par rapport à celle des SIR bruts (Tableau 15), au maximum d'environ 0,81 à 1,22, le rapport entre

le maximum et le minimum étant de 1,10 à 1,85 pour le modèle CAR, de 1,17 à 1,40 pour BYM, de 1,07 à 1,22 pour KHR et de 1,06 à 1,39 pour MIX.

Tableau 15 : Etendue et ratio des max/min des SIR lissés par sous-population

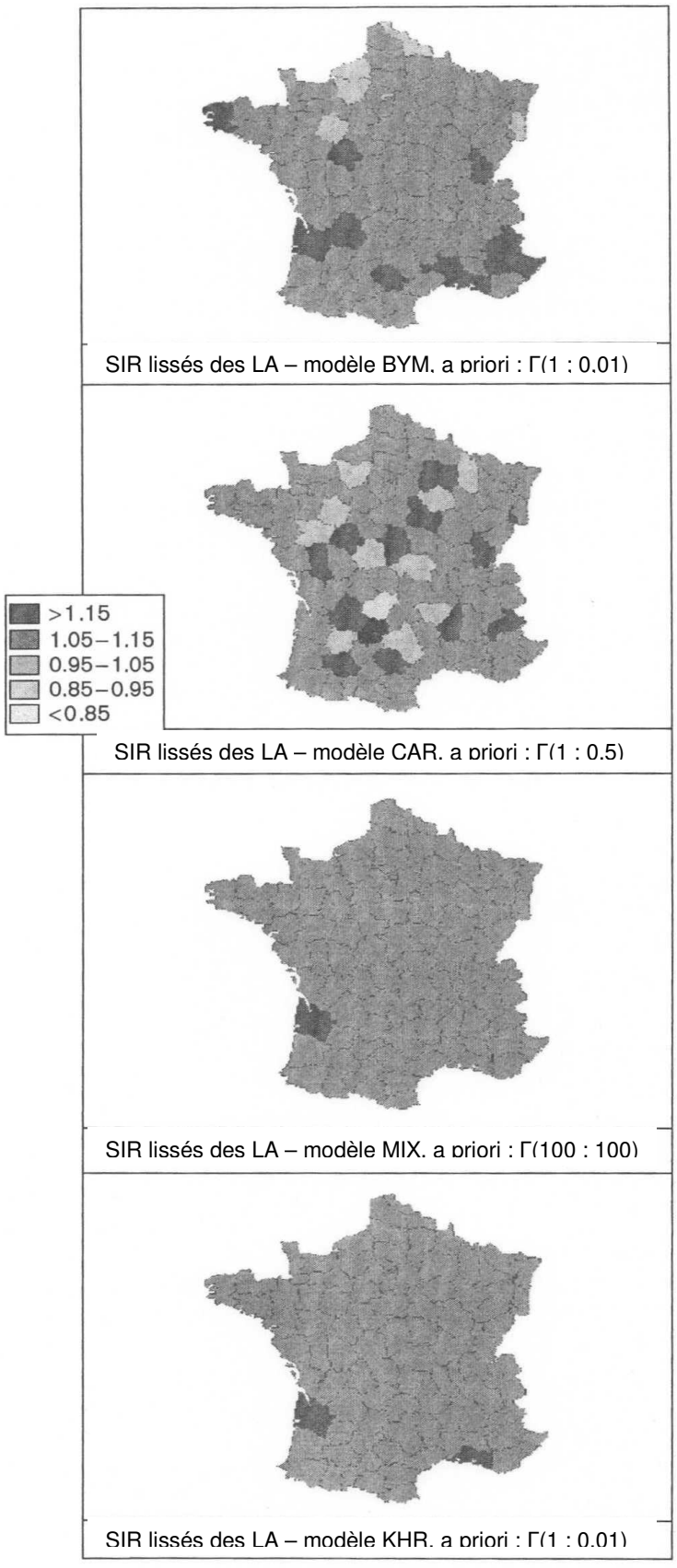
	BYM		CAR		KHR		MIX	
	Etendue	Ratio	Etendue	Ratio	Etendue	Ratio	Etendue	Ratio
Toutes LA	0,91-1,1	1,21	0,84-1,17	1,39	0,95-1,06	1,12	0,95-1,05	1,11
1990-1994	0,94-1,1	1,17	0,90-1,16	1,29	0,97-1,04	1,07	0,97-1,04	1,07
1995-1999	0,89-1,18	1,33	0,86-1,21	1,41	0,93-1,12	1,20	0,93-1,10	1,19
2000-2004	0,92-1,14	1,24	0,87-1,22	1,41	0,96-1,08	1,13	0,96-1,08	1,16
M	0,92-1,12	1,21	0,91-1,07	1,18	0,94-1,06	1,12	0,96-1,05	1,09
F	0,90-1,10	1,22	0,82-1,31	1,61	0,94-1,05	1,12	0,95-1,05	1,11
0-4ans	0,89-1,14	1,28	0,90-1,19	1,32	0,95-1,09	1,15	0,94-1,07	1,14
5-9ans	0,87-1,11	1,25	0,89-1,19	1,33	0,95-1,05	1,10	0,95-1,05	1,10
10-14ans	0,86-1,18	1,38	0,87-1,18	1,35	0,91-1,09	1,20	0,91-1,08	1,18
LAL	0,89-1,14	1,28	0,95-1,09	1,15	0,93-1,09	1,16	0,86-1,19	1,39
1990-1994	0,92-1,08	1,17	0,96-1,06	1,10	0,96-1,03	1,07	0,90-1,13	1,25
1995-1999	0,89-1,22	1,37	0,93-1,15	1,24	0,94-1,11	1,18	0,88-1,22	1,39
2000-2004	0,86-1,2	1,40	0,92-1,13	1,23	0,91-1,11	1,21	0,87-1,19	1,38
M	0,89-1,14	1,28	0,89-1,07	1,19	0,95-1,08	1,14	0,95-1,06	1,12
F	0,87-1,15	1,32	0,82-1,32	1,60	0,91-1,09	1,20	0,92-1,08	1,17
0-4ans	0,88-1,13	1,28	0,92-1,16	1,27	0,94-1,10	1,17	0,94-1,07	1,14
5-9ans	0,89-1,14	1,29	0,88-1,18	1,34	0,94-1,07	1,14	0,94-1,06	1,13
10-14ans	0,82-1,15	1,40	0,90-1,14	1,26	0,90-1,10	1,22	0,90-1,07	1,18
LAM	0,9-1,13	1,26	0,91-1,12	1,30	0,94-1,07	1,14	0,95-1,06	1,11
1990-1994	0,91-1,13	1,245	0,90-1,67	1,85	0,96-1,07	1,12	0,97-1,03	1,06
1995-1999	0,92-1,13	1,22	0,94-1,15	1,23	0,95-1,15	1,21	0,97-1,05	1,08
2000-2004	0,90-1,18	1,31	0,95-1,14	1,20	0,93-1,1	1,18	0,97-1,06	1,09
M	0,91-1,13	1,24	0,94-1,14	1,21	0,95-1,11	1,17	0,97-1,05	1,09
F	0,93-1,15	1,24	0,94-1,15	1,22	0,96-1,06	1,10	0,98-1,05	1,08
Min		1,17		1,19		1,07		1,06
Max		1,40		1,85		1,24		1,21

Les deux modèles paramétriques réduisent moins que les autres les variabilités empiriques des SIR bruts (Tableau 15) : principalement le modèle CAR (14 fois sur 24 c'est ce modèle qui les réduit le moins) avec une amplitude des estimations au maximum de [0,82; 1,31] et dans les autres cas c'est le modèle BYM, les étendues des estimations des risques étant du même ordre de grandeur. En revanche les deux modèles non paramétriques réduisent plus les variabilités des données brutes et spécialement le modèle MIX (amplitude maximum = [0,88; 1,22]).

D.2.c Cartes des SIR lissés

Aucune structure spatiale n'est observée sur la cartographie des SIRs lissés de LA (Figure 6), que ce soit globalement ou séparément par sexe, par période ou par tranche d'âge. Par ailleurs, tous les intervalles de crédibilité à 90% des SIR contenaient 1 sauf pour le département 93 avec le modèle BYM [$ICr_{90\%}$ = (0,83-0,99)] mais ce résultat n'est pas retrouvé par sous-population.

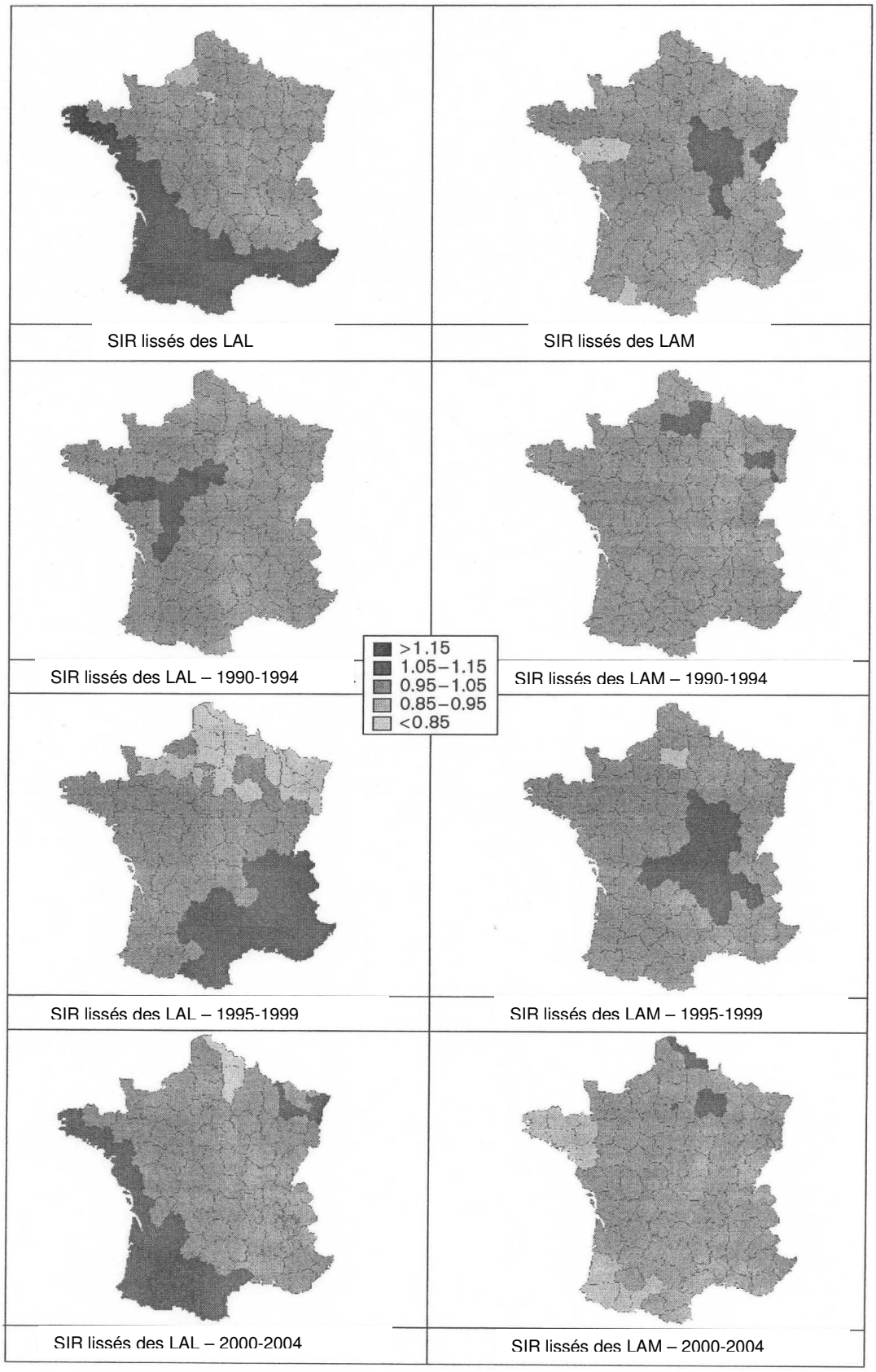
Figure 6 : Carte des SIR des LA, lissés par les modèles BYM, CAR, MIX et KHR



Concernant la cartographie des LAL étudiées globalement, les SIR lissés du Sud-Ouest de la France, sont très légèrement plus élevés (avec le modèle BYM de l'ordre de 1,05-1,15, MIX de l'ordre de 1,02 et KHR de l'ordre de 1,07) que ceux du Nord-Est, (avec le modèle BYM de l'ordre de 0,9-0,95, MIX de l'ordre de 0,98 et KHR de l'ordre de 0,95) (Figure 7). Cette structure est retrouvée dans la carte des SIR des filles et de la tranche d'âge 0-4 ans mais pas systématiquement par période, par sexe ni par âge. Les quelques SIR lissés trouvés significativement différents de 1 (principalement ceux des départements 13, 33, 59 et 93) pour les LAL avec les modèles BYM et KHR, ne le restaient pas par sous-population de LAL.

Concernant la cartographie des LAM étudiées globalement, les SIR lissés du Centre-Est de la France sont très légèrement plus élevées, (avec le modèle BYM de l'ordre de 1,05-1,08, CAR 1,08 et KHR 1,06), que ceux du reste de la France (Figure 7). Cependant, cette structure n'est stable ni par période, ni par sexe: elle n'apparaît que sur les cartes des SIR de la période 1995-1999 et des garçons. Par ailleurs, tous les SIR lissés avaient un intervalle de crédibilité à 90% contenant 1.

Figure 7 : Carte des SIR lissés par le modèle KHR, pour les LAL et les LAM, par sous-période de 5 ans : 1990-1994 ; 1995-1999 et 2000-2004



D.2.d Corrélations entre SIR lissés

Nous avons examiné les corrélations entre trois périodes, deux sexes et trois classes d'âges des cartes des SIR lissés par une même méthode : ceci pour les toutes LA, les LAL seules ainsi que pour les LAM seules.

On observe que les modèles de lissages présentent d'autant plus de corrélations entre les cartes qu'ils réduisent fortement les variabilités des données brutes : les corrélations les plus faibles sont observées pour les estimations CAR, ce sont BYM et KHR qui présentent les corrélations les plus fortes. Cependant, aucune de ces corrélations n'étaient significatives avec le t-test modifié. Par exemple, l'estimation du coefficient de corrélation entre les SIR des LAL filles et garçons, lissés avec le modèle BYM était de 0,41 ($p<10^{-4}$), coefficient largement significatif avec le test de Student classique (92 degrés de liberté) mais non-significatif avec le test de Student modifié prenant en compte les autocorrélations spatiales à l'ordre 3 (20 degrés de liberté). Les nuages de dispersion entre les cartes KHR (les plus corrélées) ne sont pas évocateurs de l'existence d'un lien solide entre ces cartes.

E Discussion

Aucune structure spatiale stable n'a été observée sur les données brutes et lissées étudiées ici. Même si une légère hétérogénéité a été détectée globalement sur les LA, elle n'était pas associée à une structure spatiale stable par sous-type de leucémie, période de temps, sexe ou groupe d'âge.
Aucune méthode de lissage n'a mis en évidence de structure géographique particulière de l'incidence.

Les modèles bayésiens développés pour cartographier l'incidence des maladies ont été utilisés pour estimer la structure spatiale globale des risques relatifs, avec une bonne puissance d'étude due au fait que l'étude est faite sur le pays entier et avec une variance réduite grâce à l'utilisation du lissage des estimateurs. Différentes hypothèses *a priori* ont été faites, d'abord sur la structure spatiale de la variabilité du risque relatif à travers premièrement les modèles paramétriques BYM et CAR et les modèles semi-paramétriques MIX et KHR, puis sur l'amplitude de cette variabilité.
Les deux modèles paramétriques BYM et CAR sont tous les deux conditionnels autorégressifs mais ils diffèrent non seulement dans la manière de définir et d'interpréter la dépendance spatiale mais aussi sur leur structure de voisinage, c'est-à-dire sur la manière dont sont pondérés les risques des départements voisins dans l'estimation du risque.
Cependant, l'intensité du lissage de ces modèles paramétriques, dépendante de la variance des paramètres pour le modèle BYM et des paramètres de variance et d'autocorrélation dans le modèle CAR, est affectée par les paramètres de tous les

départements et n'est donc pas très adaptée localement. Afin d'éviter un lissage trop fort d'une éventuelle structure sous-jacente discontinue, les modèles semi-paramétriques MIX et KHR ont été utilisés. Dans ces modèles, la distribution spatiale continue des risques relatifs qui est supposée dans les modèles paramétriques est remplacée par une allocation discrète [Richardson *et al.*,2004] ou des modèles de partition [Knorr-Held *et al.*,2000].

Le mécanisme d'allocation discrète du modèle MIX permet *a priori* une continuité dans la structure spatiale en autorisant l'allocation des départements voisins à la même catégorie de risque. Il permet également les discontinuités locales de la structure de risque en allouant à des départements non voisins la même catégorie de risque.

En revanche dans le modèle de partition KHR, ce sont des clusters de départements contigus qui sont définis *a priori*. Tous les départements d'un cluster sont voisins, ce qui n'est pas le cas pour les départements appartenant à la même composante dans le modèle MIX. En fonction du modèle utilisé, des nombres différents d'itérations ont été utilisés pour estimer les SIR lissés. Cependant, les résultats ne changeaient pas quand le même nombre d'itérations était utilisé pour tous les modèles.

En ce qui concerne l'étendue de variabilité, dans le contexte épidémiologique des risques relatifs de leucémie de l'enfant au niveau départemental, il semble raisonnable de considérer qu'*a priori*, à 95%, l'intervalle de fluctuation du risque relatif pour un département ait une étendue moyenne allant de 0,80 à 1,25 et une étendue à 95% de 0,25 à 4,00. Par ailleurs, en utilisant une étendue de variabilité de 0,1 à 9,0 en moyenne et jusqu'à 0 à + ∞, les résultats obtenus sont similaires. Les résultats ne dépendent donc pas des hyper a priori.

Les modèles utilisés dans cette étude couvrent un large panel des modèles disponibles pour la détection de différentes sortes de structures spatiales [Best *et al.*,2005]. Cependant, quelques études de simulations ([Knorr-Held &Rasser,1999] ; [Green &Richardson,2002]; [Richardson *et al.*,2004] ; [Best *et al.*,2005]) ont montré qu'il pouvait y avoir des risques existants élevés sous-estimés par les modèles bayésiens. Il peut y avoir un manque de puissance pour détecter des risques potentiels modérés, comme la plupart des risques environnementaux et particulièrement quand les nombres de cas attendus sont faibles. Ici, c'est particulièrement le cas pour les LAM étudiées globalement [étendue E= (1,2-58,2)] et *a fortiori* quand elles sont étudiées par sous-populations (Tableau 15).

A notre connaissance, très peu d'études se sont focalisées sur l'hétérogénéité géographique de l'incidence des leucémies de l'enfant à grande échelle, sur une période de 15 ans et en utilisant une telle approche. Le RNHE est une base de données quasi-exhaustive et très fiable qui couvre la France entière depuis 1990 et a permis de réaliser une telle étude.

Aucune autre étude n'a utilisé l'ensemble de ces méthodes pour détecter une hétérogénéité à partir de données sur les cancers de l'enfant, mais quelques études à l'étranger, en utilisant moins de méthodes ont trouvé des résultats semblables. Dans certaines, aucune hétérogénéité spatiale n'a été détectée (GCCR registry website - annual report 1999 - http://www.kinderkrebsregister.de/english/.; [Muirhead,1995]; [Williams *et al.*,2001]) et dans d'autres, quelques variations spatiales à la limite de la significativité ([Petridou *et al.*,1997] ; [Alexander *et al.*,1998] [Wakefield *et al.*,2000][McNally *et al.*,2009]) ont été trouvées.

Des analyses sur le registre des cancers de l'enfant d'Allemagne n'ont pas trouvé de différence significative entre les SIR des 328 districts de l'Allemagne de l'ouest, ni de tendance spatiale dans cette région sur les périodes 1982-1994 et 1990-1999 (site du registre *German Childhood Cancer Register*, rapport annuel 1999). Des variations spatiales à la limite de la significativité ont été trouvées entre les « counties » d'Angleterre et du Pays de Galles sur la période 1969-1988 : elles semblaient expliquées complètement (pour toutes LA) et en partie (pour les LAL) par les variations socio-économiques entre ces régions (Spatial Epidemiology, J.F. Bithell, T.J. Vincent chap18.3.).

Une étude précédente menée en France à une échelle plus fine (la commune) avait détecté une légère hétérogénéité (sur-dispersion significative de 1,2%) dans l'incidence des LA sur la période 1990-1994, résultat comparable à ce qui a été trouvé ici sur une période différente [Bellec *et al.*,2006]. Les résultats rapportés ici sont également concordants avec les résultats de l'étude EUROCLUS dans laquelle une composante extra-poissonienne de la variance de 1,7% a été détectée sur une échelle intermédiaire entre le département et la commune [Alexander *et al.*,1998]. L'étude française précédente a également cherché à détecter des clusters spécifiques ce qui n'a pas mené à des résultats significatifs[Bellec *et al.*,2006]. Une étude plus récente n'a trouvé aucune hétérogénéité au niveau des bassins de vie (il y a 1 895 bassins de vie en France) et aucun cluster stable n'a été détecté [Demoury *et al.*,2011]

Le fait que nous n'ayons pas détecté d'hétérogénéité de l'incidence des LA à l'échelle départementale n'implique pas qu'il n'existe aucune hétérogénéité à

d'autres échelles. Dans la population d'étude et dans d'autres, des facteurs de risque suspectés ont été étudiés à une échelle plus fine et ont été trouvés associés à l'incidence de LA : cela a été décrit globalement plus haut et le sera plus en détail dans les chapitres suivants pour les deux expositions étudiées dans le cadre de cette thèse. Par ailleurs, des facteurs de risque individuels des leucémies de l'enfant ont été identifiés comme les radiations ionisantes à forte dose, des facteurs iatrogènes comme la chimiothérapie et des facteurs génétiques comme la trisomie 21. Cependant, ces facteurs de risque ne sont pas susceptibles de créer une hétérogénéité spatiale à l'échelle du département.

En conclusion, cette étude menée en utilisant plusieurs modèles de lissage bayésien complémentaires n'a pas mis en évidence de variations géographiques nettes dans l'incidence des leucémies de l'enfant étudiée sur la France métropolitaine à l'échelle départementale et sur la période 1990-2004. Ce résultat n'exclut ni qu'il pourrait exister une hétérogénéité spatiale de cette incidence à des échelles plus fines ni que l'incidence de LA pourrait être corrélée spatialement avec certaines caractéristiques populationnelles qui auraient une variabilité interdépartementale. Cependant, cela implique qu'il n'y a pas de raison d'estimer des taux d'incidence de référence à un niveau départemental plutôt que national, ni d'ajuster des études ultérieures qui examineraient les variations spatiales de l'incidence de LA sur le département.

V. Proximité des centrales nucléaires de production d'électricité et leucémies aigües de l'enfant

A Introduction

De nombreuses études épidémiologiques ont analysé l'incidence des leucémies de l'enfant autour des sites nucléaires en France ([Viel *et al.*,1993]; [Bouges *et al.*,1999], [White-Knoning *et al.*, 2004], [Evrard *et al.*, 2006]) et ailleurs ([Bithell *et al.*,2008]; [Laurier *et al.*,2008]) et plus spécialement autour des centrales nucléaires ([Kaatsch *et al.*,1998]; [Mangano *et al.*,2003]; [Laurier *et al.*,2008]; [Heinävaara *et al.*,2010]; [Spycher *et al.*,2011]). Quelques excès ont été trouvés: autour des usines de retraitement britanniques de Sellafield ([Goldsmith,1992]; [Bithell *et al.*,1994]; [Sharp *et al.*,1996]) et Dounreay ([Heasman *et al.*,1986]; [Black *et al.*,1994]) et autour de la centrale nucléaire allemande de Krümmel ([Hoffmann *et al.*,1997]; [Kaatsch *et al.*,1998]; [Hoffmann *et al.*,2007]). Aucune hypothèse n'a été identifiée qui permette réellement d'expliquer ces associations même si, particulièrement pour les usines britanniques, l'hypothèse des mouvements de population est étayée par plusieurs observations convergentes ([Kinlen,1995]; [Kinlen *et al.*,1995]; [Dickinson &Parker,1999]; [Kinlen,2011]). Par ailleurs, aucune excès de risque "général" au voisinage des CNPE (centrales nucléaires de production d'électricité) n'a été mis en évidence avant la publication allemande de 2008 sur les cancers de l'enfant à proximité des CNPE allemandes cas-témoins "Epidemiologische Studie zu Kinderkrebs in der Umgebung von Kernkraftwerken; KiKK" ([Kaatsch *et al.*,2008]; [Spix *et al.*,2008]), excès qui était observé chez les seuls 0-4 ans et à moins de 5 km d'une CNPE.

En France, deux précédentes études réalisées au sein de notre équipe ont analysé l'incidence des leucémies de l'enfant autour des sites nucléaires: la première étude d'incidence [White-Koning *et al.*,2004], portait sur la période 1990-1998 et utilisait

comme indicateur d'exposition la distance des communes aux sites nucléaires, elle n'a pas mis en évidence d'excès d'incidence des leucémies de l'enfant à moins de 20km d'un site nucléaire ou d'une centrale nucléaire. La deuxième étude d'incidence [Evrard *et al.*,2006] portait sur la période 1990-2001 et utilisait comme indicateur d'exposition un zonage géographique d'estimations de doses dues aux rejets gazeux des sites nucléaires. Elle n'a pas non plus mis en évidence d'association entre cet indicateur d'exposition et l'incidence des leucémies de l'enfant.

Nous avons étudié l'incidence des leucémies de l'enfant de moins de 15 ans autour des 19 CNPE françaises, sur une nouvelle période d'observation 2002-2007 mais également sur l'ensemble de la période 1990-2007, pour laquelle nous disposons des données d'incidence de cette pathologie. Nous avons utilisé une approche cas-témoins pour la période 2002-2007 et des études d'incidence au niveau communal pour les périodes 1990-2001, 2002-2007 et 1990-2007. La distance de l'adresse précise du domicile de chaque sujet à la CNPE la plus proche a été utilisée dans l'analyse cas-témoins, dans les études d'incidence c'est la distance du centre de la commune à la centrale la plus proche qui a été utilisée. Nous avons également étudié comme indicateur d'exposition l'estimation de dose à la moelle osseuse due aux rejets gazeux des CNPE réalisée par l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN) pour la précédente étude, 1990-2001, de l'équipe.

B Matériel et méthodes

B.1 Sujets

B.1.a Etude cas-témoins 2002-2007

L'analyse cas témoins porte sur la période 2002-2007 et inclut les 2 753 enfants atteints de leucémies aigues, diagnostiquées en France métropolitaine et qui avaient moins de 15 ans au 31 décembre de l'année au diagnostic. Les sujets témoins étaient les 30 000 enfants constituant l'échantillon témoin contemporain des cas, réalisé par l'INSEE, précédemment décrit et qui forment un échantillon représentatif de la population pédiatrique française (voir partie III chapitre C) pour chacune des 6 années 2002 à 2007 (5 000 enfants par année).

B.1.b Etudes d'incidence: 1990-2001, 2002-2007 et 1990-2007

Pour les études d'incidence, tous les cas de LA, âgés de moins de 15 ans, enregistrés dans le RNHE pendant la période correspondante ont été inclus. L'âge utilisé dans les études d'incidence est l'âge à la date de diagnostic en années révolues.

A partir des données de recensement de l'INSEE et des interpolations que nous avons réalisées pour les années pour lesquelles les estimations par classe d'âge de un an n'étaient pas données au niveau communal, les nombres de cas attendus de LA ont été calculés pour chacune des 18 années de la période 1990-2007 par commune et par classe d'âge de 5 ans en utilisant le taux d'incidence national de LA pour l'année correspondante comme taux de référence.

B.2 Caractéristiques des CNPE

Les 19 CNPE françaises (Tableau 16) peuvent être classées suivant leur puissance électrique, qui dépend de leur nombre de réacteurs: 1800 Mega Watts (MWe) pour *Fessenheim* et *Saint-Laurent,* 2600 MWe pour *Belleville*, *Flamanville*, *Golfech*, *Nogent*, *Penly* et *Saint-Alban*, 2800 MWe pour *Chooz* et *Civaux*, 3600 MWe pour *Bugey*, *Chinon*, *Cruas*, *Dampierre*, *Le Blayais* et *Tricastin*, 5200 MWe pour *Cattenom* et *Paluel* et 5400 MWe for *Paluel*.

Elles ont toutes été mises en service avant 1990 sauf *Chooz* (1996) et *Civaux* (1997).

Trois CNPE (*Flamanville*, *Paluel* and *Penly*) sont situées sur la côte de la Manche et très proches de l'usine de retraitement de combustible irradié de La Hague, la CNPE de *Le Blayais* est située à l'estuaire de la *Gironde*, en bordure de l'Océan Atlantique et les autres CNPE sont situées le long d'une rivière.

Enfin, trois CNPE étant situées à moins de 20 km d'une frontière (*Cattenom*, *Chooz* and *Fessenheim*) les études cas-témoins et d'incidence que nous avons réalisées n'ont porté que sur la partie française de leur voisinage.

Ces 19 CNPE françaises ont été incluses dans l'étude. Pour chacune d'elles, les informations concernant l'année de mise en service des réacteurs, la puissance électrique et les estimations de doses de rejets gazeux nous ont été fournies par l'IRSN. Les cas et personnes x années autour des CNPE de Chooz et Civaux n'ont été pris en compte dans les études d'incidence 1990-2001 et 1990-2007 qu'à partir du début de leur année de mise en service qui est postérieure à 1990 (1996 et 1997 respectivement).

Tableau 16: Caractéristiques des 19 centrales nucléaires de production d'électricité françaises

	Date de mise en service	Puissance électrique (Mega Watts)	Localisation par rapport à une côte	Localisation frontalière
Belleville	1987	2600		
Bugey	1971	3600		
Cattenom	1986	5200		*Luxembourg*
Chinon	1963	3600		
Chooz	1996	2800		*Belgique*
Civaux	1997	2800		
Cruas	1983	3600		
Dampierre	1980	3600		
Fessenheim	1977	1800		*Allemagne*
Flamanville	1985	2600	*Manche*	
Golfech	1990	2600		
Gravelines	1980	5400	*Mer du nord*	
Le Blayais	1981	3600		
Nogent	1987	2600		
Paluel	1984	5200	*Manche*	
Penly	1990	2600	*Manche*	
Saint-Alban	1985	2600		
Saint-Laurent	1969	1800		
Tricastin	1980	3600		

B.3 Distance à la CNPE la plus proche

Le premier indicateur d'exposition utilisé a été la distance du domicile des sujets à la CNPE la plus proche.

Dans l'étude **cas-témoin**, la distance précise du domicile du sujet à la CNPE la plus proche a pu être utilisée.

Les sujets ont été classés en cinq catégories [0-5 km[, [5-10km[, [10-15km[, [15-20km[et ≥20km.

Afin de prendre en compte la précision du géocodage des adresses, nous avons représenté un sujet par un point associé à un disque ayant pour rayon la précision du géocodage. Pour 28 sujets, le disque centré au point du géocodage et de rayon égal à la précision de celui-ci recoupait deux classes de distances (voir partie III chapitre

D 2). Cependant, une vérification manuelle sur des cartes très précises nous a finalement permis de déterminer à quelle classe de distance devait être affectés ces 28 sujets. Les 413 sujets pour lesquels seule la commune d'habitation était connue ont été classés en fonction de la localisation de la mairie de la commune.

Concernant l'approche **incidence**, nous avons calculé l'incidence au niveau de la commune, qui est la plus petite unité administrative pour laquelle les données de population sont disponibles (sauf dans les quelques situations exceptionnelles où une commune est subdivisée en cantons). Toutes les communes ont été classées dans une des cinq catégories de distance [0-5 km[, [5-10km[, [10-15km[, [15-20km[et ≥20km en fonction de la distance de leur mairie à la CNPE la plus proche. Les communes dont les mairies étaient à moins de 20 km de plusieurs CNPE ont été considérées comme appartenant à la zone d'influence de la seule CNPE la plus proche, le but étant de maintenir une partition stricte de la zone d'étude pour assurer l'indépendance des analyses statistiques effectuées.

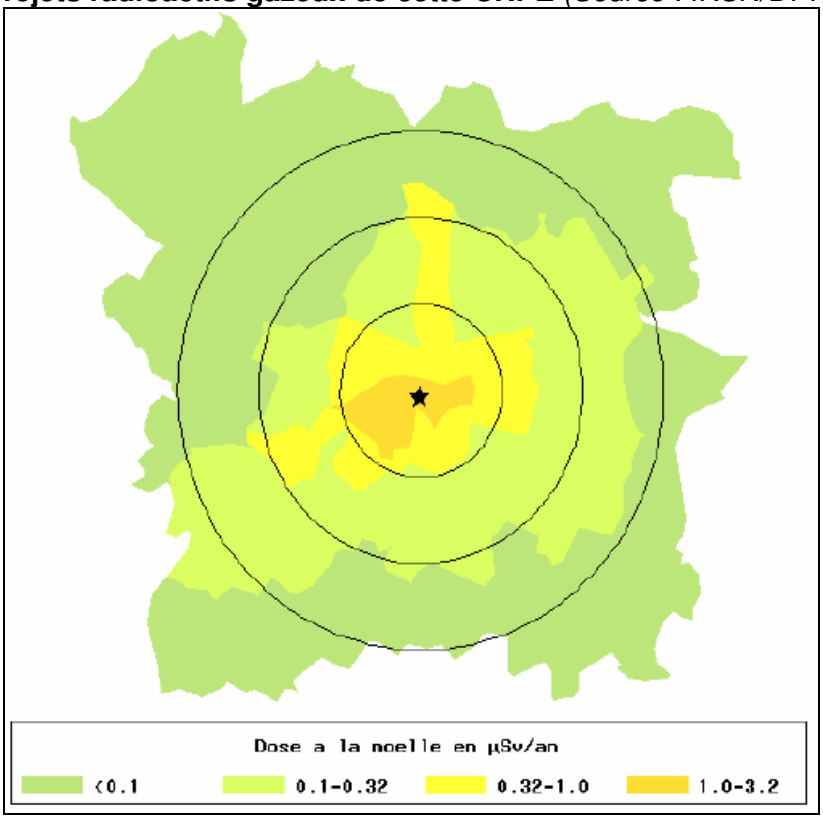
B.4 Dose à la moelle estimée due aux rejets gazeux des CNPE

Le second indicateur d'exposition est l'estimation de la dose à la moelle due aux rejets radioactifs gazeux des CNPE : **DBGZ ("Dose Based Geographic Zoning" : zonage géographique fondé sur l'estimation de la dose liée aux rejets gazeux de la CNPE)**.

Ces doses ont été estimées par l'IRSN au niveau communal [Evrard *et al.*,2006]. La dose estimée par l'IRSN attribuée à un sujet est la dose équivalente à la moelle osseuse reçue chez l'enfant, estimée pour sa commune et exprimée en sievert (Sv) unité de mesure de la radioactivité permettant d'évaluer le risque biologique pour un

organe donné. Cette dose équivalente est calculée à partir de la dose absorbée (exprimée en gray) et pondérée par un facteur tenant compte du type de rayonnement, elle permet d'estimer de façon quantitative l'exposition de la population considérée aux rayonnements ionisants à partir de la modélisation des transferts des rejets radioactifs réels gazeux, des installations nucléaires en fonctionnement normal dans l'environnement. Les rejets utilisés sont les rejets réels annuels moyens type par type de réacteur pour les CNPE (une CNPE pouvant avoir plusieurs réacteurs de différents types) et les rejets annuels par site pour les autres sites nucléaires. Les données postérieures à 1995 (1995-2000) sont utilisées pour les CNPE, et autour de 1995 pour les autres sites suivant la disponibilité des données. Nous disposions d'une estimation de dose (Figure 8) pour chaque commune incluse dans un carré de 20 km de côté centré sur 24 sites nucléaires (les 19 CNPE, La Hague, Cadarache, Pierrelatte, Romans-sur-Isère et Saclay).

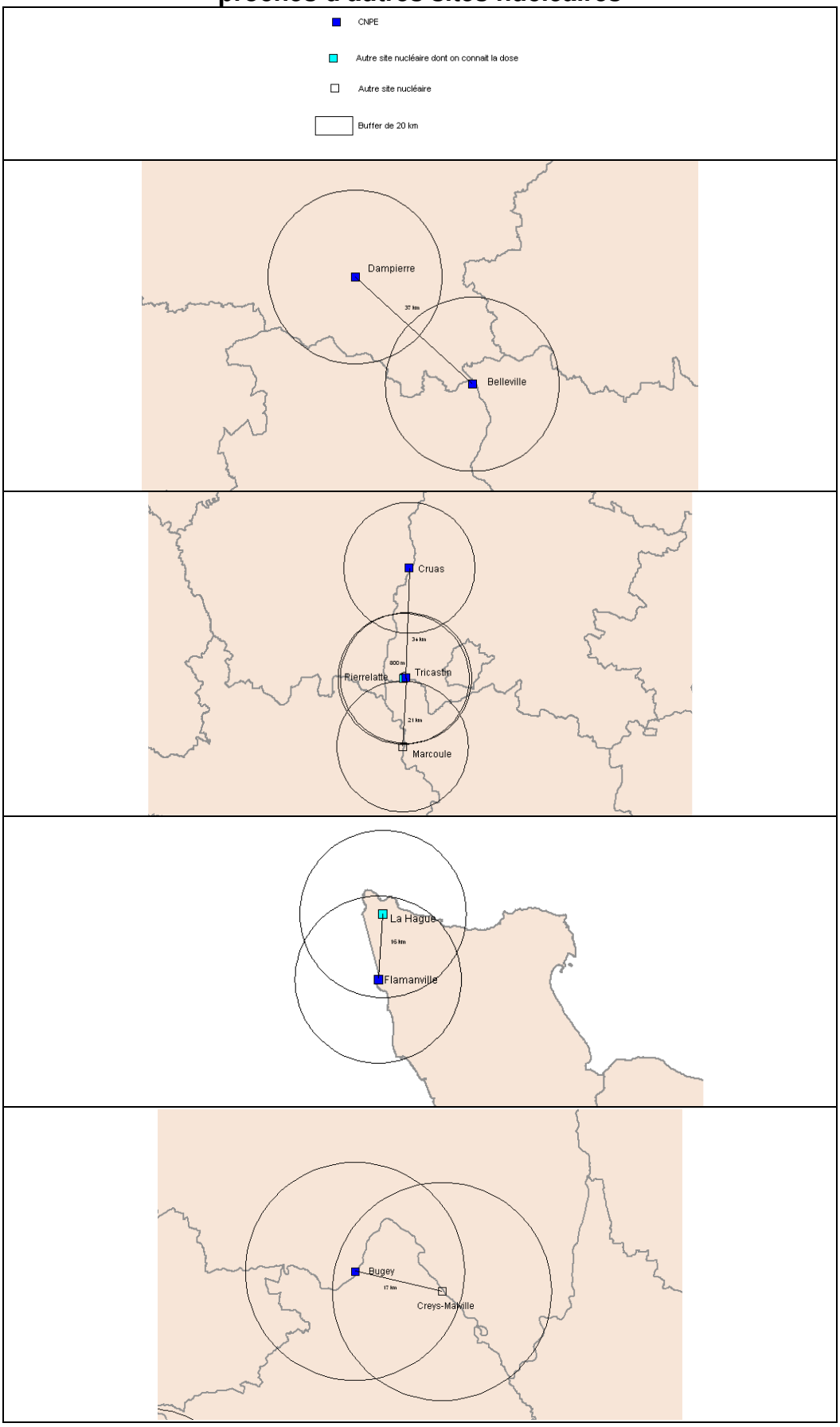
Figure 8: Exemple de zonage pour la CNPE de Chinon à partir des estimations de la dose due aux rejets radioactifs gazeux de cette CNPE *(Source : IRSN/DPRH/SER/UETP)*



Quand une commune était proche de deux CNPE, la dose estimée attribuée aux sujets de cette commune était la somme des doses correspondantes à chacune des CNPE (par exemple Belleville et Dampierre, Tricastin et Cruas, Figure 9). Certaines communes proches d'une CNPE sont aussi proches d'autres sites nucléaires pour lesquels l'IRSN avait également estimé les doses de rejets gazeux: c'est le cas de la CNPE de Flamanville proche de l'usine de retraitement de La Hague et des CNPE de Cruas et Tricastin, proches du site nucléaire de Pierrelatte. Dans ce cas, la dose totale était attribuée à la commune et à ses sujets.

Enfin, certaines CNPE sont proches d'autres sites nucléaires pour lesquels nous ne disposons pas d'estimations de doses (comme Bugey qui est proche du site de Creys-Malville et Tricastin qui est proche de Marcoule, Figure 9). Dans ce type de situation, la seule dose prise en compte était donc celle de la CNPE.

Figure 9 : Localisation des CNPE qui sont proches les unes des autres ou proches d'autres sites nucléaires



Afin de réaliser des calculs sur des voisinages de CNPE disjoints et présentant donc des nombre d'observations statistiquement indépendants, lorsqu'une commune était proche de plusieurs CNPE, les sujets y habitant ont été rattachés à la CNPE qui contribuait le plus à la dose totale la concernant.

Pour que cette approche soit cohérente avec l'approche en distance, nous avons considéré que les communes dont la mairie était située à plus de 20 km d'une CNPE avaient une dose estimée égale à l'exposition de fond « 0 µSv/an », qui correspond à la catégorie de référence.

L'estimation de dose attribuée à une commune étant la somme des estimations provenant de différents sites pour lesquels ces estimations étaient disponibles (autre CNPE ou autre type de site nucléaire), certaines communes très proches d'un autre site nucléaire qu'une CNPE et relativement éloignées de cette CNPE –par exemple entre 15 et 20 km - peuvent tout de même avoir une dose estimée relativement élevée, notamment les communes proches de La Hague avec des estimations de doses $>0{,}7.10^{-6}$µSv.

B.5 Analyse statistique

Les analyses ont été effectuées globalement sur tous les cas, par classe d'âge de 5 ans (0-4, 5-9, 10-14 ans) et par sous-type de LA: LAL et LAM.

La CNPE de *Le Blayais* n'a pas été considérée pour les tests à moins de 5 km car il n'y avait pas de cas attendus dans cette zone.

Nous avons effectué :

- des analyses séparées par année et en excluant une année, par CNPE et en les excluant tour à tour et par puissance électrique de CNPE.

- des analyses séparées sur les CNPE qui sont situées au bord de la Manche et les autres ainsi que sur les CNPE voisines d'une frontière et les autres.

L'environnement des CNPE a également été pris en compte en utilisant différentes variables fournies par l'INSEE. Les analyses ont été stratifiées sur la taille d'unité urbaine en trois classes : les zones "rurales" (communes rurales et unités urbaines de moins de 5 000 habitants) ; les zones "semi-urbaines" (unités urbaines de 5 000 à 99 999 habitants) ; les zones "urbaines" (unités urbaines de plus de 100 000 habitants). Les variables contextuelles caractérisant le statut socio-économique des communes, notamment le taux de chômage, le revenu médian, la part d'ouvriers dans la population active et la part de bacheliers chez les plus de 15 ans ont également été prises en compte par stratification des analyses. Par ailleurs, des analyses de sensibilité visant à utiliser des zones de référence proches des CNPE ont été faites en excluant les sujets à plus de 50 km d'un CNPE. Les données ont également été analysées après exclusion des sujets qui habitaient à moins de 200 m d'une ligne à haute tension (LHT) puisqu'une association entre la proximité à ces sources d'exposition et le risque de LA a été suggérée dans la littérature ([Ahlbom *et al.*,2000]; [Greenland *et al.*,2000]).

Pour tester l'existence d'une tendance entre l'incidence des LA et la distance à une CNPE, nous avons utilisé comme covariable la fonction inverse de la distance, pour des classes de 2 km. La significativité des paramètres de la régression a été évaluée en simulant 10 000 réplications aléatoires du jeu de données sous l'hypothèse H_0 d'uniformité spatiale des taux d'incidence des LA. Pour l'étude cas-témoins, 10 000 permutations du statut cas-témoins des sujets ont été simulées, par département et indépendamment de la distance aux CNPE. Dans les études d'incidence, les 10 000

jeux de données ont été construits sous l'hypothèse d'une distribution de Poisson du nombre de cas par commune ayant comme espérance le nombre attendu calculé à partir des taux d'incidence de référence de l'année et des estimations des populations pédiatriques par classe d'âge de la commune considérée.

B.6 Puissance statistique

Pour un test unilatéral au risque d'erreur 5%, la puissance statistique de l'étude cas-témoins 2002-2007 pour détecter un OR égal à 2 était de 40% pour la classe d'âge 0-4 ans et de 70% pour les 0-14 ans (Tableau 17). Pour un test de Poisson exact unilatéral au risque d'erreur 5%, la puissance de détection d'un doublement d'incidence chez les 0-14 ans était de 70% pour la période 2002-2007 et de plus de 80% pour les études d'incidence sur 1990-2001 et 1990-2007.

La puissance des analyses réalisées chez tous les sujets de moins de 15 ans est donc bonne quelle que soit la période d'étude, alors que celle des analyses réalisées uniquement chez les sujets de moins de 5 ans est limitée pour la période 2002-2007.

Tableau 17 : Puissance de l'étude cas-témoins et des études d'incidence pour détecter différentes valeurs (θ) du risque relatif en utilisant des tests unilatéraux au risque d'erreur 5%.

	Etude cas-témoins	Etudes d'incidence		
	2002-2007	2002-2007	1990-2001	1990-2007
Age < 15 ans				
θ=1,2	0,13	0,13	0,17	0,22
θ=1,5	0,32	0,34	0,50	0,66
θ=2,0	0,70	0,73	0,92	1,00
Age < 5 ans				
θ=1,2	0,10	0,10	0,13	0,15
θ=1,5	0,20	0,22	0,31	0,42
θ=2,0	0,42	0,47	0,68	0,85

C Résultats

C.1 Distance à la CNPE la plus proche

C.1.a Etude cas-témoins 2002-2007

Le tableau 18 donne la distribution des cas et des témoins autour des CNPE, par catégorie de distance de l'adresse individuelle à la CNPE la plus proche. Chez les 0-14 ans, l'OR correspondant aux enfants habitant à moins de 5 km d'une CNPE comparés à ceux habitant à 20 km ou plus d'une CNPE est proche de 2 et diffère significativement de 1 : OR=1,85 [1,04-3,32]. Dans les zones entre 5 et 20 km autour des CNPE, aucun OR ne diffère significativement de 1, ni non plus globalement dans la zone 0-20 km. Chez les enfants de moins de 5 ans, aucun OR significativement différent de 1 n'a été observé même si l'OR correspondant à la zone 0-5km est supérieur à 1 (OR=1,64 [0,65-4,10], p=0,15) et peut être considéré comme proche de l'OR observé chez les 0-14 ans. Les observations réalisées chez les 5-9 ans et les 10-14 ans sont similaires.

Tableau 18 : Distance à la CNPE la plus proche et LA : étude cas-témoin Geocap et études d'incidences

	Etude cas-témoins Geocap, 2002-2007[1] (distance: adresse-CNPE la plus proche)					Etude d'incidence, 2002-2007[2] (distance: mairie –CNPE la plus proche)					Etude d'incidence, 1990-2001[2] (distance: mairie –CNPE la plus proche)					Etude d'incidence, 1990-2007[2] (distance: mairie –CNPE la plus proche)				
	Cas	**Témoins**	**OR**	**IC 95%**	**p**	**O**	**E**	**SIR**	**IC 95%**	**p**	**O**	**E**	**SIR**	**IC 95%**	**p**	**O**	**E**	**SIR**	**IC 95%**	**p**
Age < 15 ans																				
0-4,99 km	14	80	1,9	[1,0-3,3]	*	14	7,4	1,9	[1,0-3,2]	*	10	13,6	0,7	[0,4-1,4]		24	21,0	1,1	[0,7-1,7]	
5-9,99 km	17	213	0,9	[0,5-1,5]		19	20,6	0,9	[0,6-1,4]		40	39,2	1,0	[0,7-1,4]		59	59,8	1,0	[0,8-1,3]	
10-14,99 km	27	320	0,9	[0,6-1,4]		30	25,4	1,2	[0,8-1,7]		50	48,5	1,0	[0,8-1,4]		80	73,9	1,1	[0,9-1,4]	
15-19,99 km	41	447	1,0	[0,7-1,4]		36	42,4	0,9	[0,6-1,2]		73	81,5	0,9	[0,7-1,1]		109	124,0	0,9	[0,7-1,1]	
≥ 20 km	2654	28940	1,0																	
<20 km	99	1060	1,0	[0,8-1,3]		99	95,7	1,0	[0,8-1,3]		173	182,9	0,9	[0,8-1,1]		272	278,6	1,0	[0,9-1,1]	
≥ 20 km	2654	28940	1,0																	
Age < 5 ans[3]																				
0-4,99 km	6	27	1,6	[0,7-4,1]		8	3,6	2,2	[1,0-4,4]	*	6	6,6	0,9	[0,3-2,0]		14	10,2	1,4	[0,8-2,3]	
5-9,99 km	7	55	1,0	[0,5-2,3]		10	10,2	1,0	[0,5-1,8]		21	19,3	1,1	[0,7-1,7]		31	29,5	1,1	[0,7-1,5]	
10-14,99 km	11	103	0,8	[0,4-1,4]		11	12,6	0,9	[0,4-1,6]		18	23,5	0,8	[0,5-1,2]		29	36,1	0,8	[0,5-1,2]	
15-19,99 km	17	136	1,0	[0,6-1,7]		18	20,8	0,9	[0,5-1,4]		39	39,7	1,0	[0,7-1,3]		57	60,6	0,9	[0,7-1,2]	
≥ 20 km	1248	9396	1,0																	
Total <20 km	41	321	1,0	[0,7-1,4]		47	47,3	1,0	[0,7-1,3]		84	89,1	0,9	[0,8-1,2]		131	136,4	1,0	[0,8-1,1]	
≥ 20 km	1248	9396	1,0																	

[1] Odds ratios (OR) et intervalles de confiance à 95% (IC 95%) estimés par régression logistique ajustée sur l'âge en fin d'année (classes d'âge de 5 ans pour les 0-14 ans, classes d'âge de 1 an pour les 0-4 ans) et sur le département de résidence ; [2] Ratio d'incidence standardisé (SIR) calculé comme le radio du nombre de cas observés (O) sur attendus (E) et intervalle de confiance à 95% estimé avec l'approximation de Byar (95%IC); [3] Age en fin d'année pour l'étude Geocap; âge en années révolues pour les études d'incidence

La distribution des cas par sexe, type de LA et classe d'âge dans les différentes catégories de distance aux CNPE est très proche de celle de la catégorie de référence "≥ 20 km" (Tableau 19).

Tableau 19 : Description des cas enregistrés à moins de 20 km d'une CNPE sur la période 2002-2007

	Distance : adresse – CNPE la plus proche											
	0-4,99 km		5-9,99 km		10-14,99 km		15-19,99 km		< 20 km		≥ 20 km	
	n	%	n	%	n	%	n	%	n	%	n	%
Sexe	**14**		**17**		**27**		**41**		**99**		**2654**	
Fille	7	50,0%	8	47,1%	13	48,1%	20	48,8%	48	48,5%	1202	45,3%
Garçon	7	50,0%	9	52,9%	14	51,9%	21	51,2%	51	51,5%	1452	54,7%
Age												
< 5 ans	6	42,9%	7	41,2%	11	40,7%	17	41,5%	41	41,4%	1248	47,0%
5-9 ans	5	35,7%	6	33,3%	11	40,7%	14	34,1%	36	36,4%	829	31,2%
10-14 ans	3	21,4%	4	23,5%	5	18,6%	10	24,4%	22	22,2%	577	21,7%
Syndrome de Down	0	0%	0	0%	0	0%	1	2,4%	1	1,0%	41	1,5%
Type de leucémie aigüe												
LAL	14	100,0%	13	76,5%	23	85,2%	30	73,2%	80	80,8%	2179	82,1%
LAL à cellules précurseur B *	11	78,6%	10	58,8%	19	70,4%	26	63,4%	66	66,7%	1810	68,2%
LAL à cellules T	2	14,3%	3	17,6%	3	11,1%	4	9,8%	12	12,1%	249	9,4%
autre LAL	1	7,1%	0	0,0%	1	3,7%	0	0,0%	2	2,0%	120	4,5%
LAM	0	0,0%	4	23,5%	3	11,1%	10	24,0%	17	17,2%	407	15,3%
Autre LA	0	0,0%	0	0,0%	1	3,7%	1	2,0%	2	2,0%	68	2,6%

LA: leucémie aigüe; LAL: leucémie aigüe lymphoblastique; LAM: leucémie aigüe myéloblastique
* ce sous-groupe inclut les LAL B communes et pré-B

Des analyses séparées par sous-type de LA (LAL and LAM) n'ont pas mis en évidence de résultats spécifiques pour l'un ou l'autre sous-type (Tableau 20) même si chez les LAL l'association avec la proximité aux CNPE semble légèrement renforcée à moins de 5 km d'une CNPE (OR=2,4 [1,3-4,2]). Le nombre de LAM était très faible : 17 cas à moins de 20 km et aucun à moins de 5 km.

Tableau 20 : Association entre leucémie aigüe et 2 indicateurs d'exposition: distance de l'adresse à la CNPE la plus proche et exposition aux radiations estimée par un zonage géographique des rejets gazeux autour des CNPE dans l'étude cas-témoins Geocap (2002-2007)

	Leucémie aigüe lymphoblastique				Leucémie aigüe myéloblastique			
	Cas	Témoins	OR[1]	IC 95%	Cas	Témoins	OR[1]	IC 95%
Age < 15 ans								
0-4,99 km	14	80	**2,4**	[1,3-4,2]	0	80	**0,0**	-
5-9,99 km	13	213	0,9	[0,5-1,5]	4	213	1,4	[0,5-3,8]
10-14,99 km	23	320	1,0	[0,6-1,5]	3	320	0,6	[0,2-1,9]
15-19,99 km	30	447	0,9	[0,6-1,4]	10	447	1,4	[0,7-2,8]
Total <20 km	80	1060	**1,0**	[0,8-1,3]	17	1060	**1,0**	[0,6-1,8]
≥ 20 km	2173	28940	**1,0**		405	28940		
'>0,72 µSV'	8	97	1,3	[0,6-2,8]	0	97	**0,0**	-
'0,21-0,72 µSV'	16	213	1,0	[0,6-1,7]	3	213	0,9	[0,3-2,8]
'0,094-0,21 µSV'	24	317	1,1	[0,7-1,6]	5	317	0,9	[0,4-2,3]
'0-0,094 µSV'	33	417	1,1	[0,7-1,5]	6	417	0,9	[0,3-2,2]
Total <20 km	81	1044	**1,1**	[0,8-1,4]	14	1044	**0,8**	[0,5-1,5]
≥ 20 km	2172	28956	**1,0**		408	28956		
Age < 5 ans								
0-4,99 km	6	27	**2,1**	[0,8-5,4]	0	27	**0,0**	-
5-9,99 km	6	55	1,1	[0,5-2,7]	1	55	0,7	[0,1-5,5]
10-14,99 km	9	103	0,8	[0,4-1,6]	1	103	0,4	[0,1-2,8]
15-19,99 km	10	136	0,8	[0,4-1,5]	6	136	1,7	[0,7-4,1]
Total <20 km	31	321	**0,9**	[0,6-1,4]	8	321	**0,9**	[0,4-2,1]
≥ 20 km	1024	9396	**1,0**		194	9396		
'>0,72 µSV'	4	28	**1,5**	[0,5-4,6]	0	28	**0,0**	-
'0,21-0,72 µSV'	6	59	1,2	[0,5-2,8]	0	59	**0,0**	-
'0,094-0,21 µSV'	12	88	1,3	[0,7-2,6]	3	88	1,1	[0,3-3,3]
'0-0,094 µSV'	11	137	0,8	[0,4-1,4]	4	137	1,2	[0,4-3,3]
Total <20 km	**33**	312	1,0	[0,7-1,4]	7	312	**0,8**	[0,4-1,9]
≥ 20 km	1022	9405	**1,0**		195	9405		

[1] Odds ratios (OR) et intervalles de confiance à 95% (IC 95%) estimés par régression logistique ajustée sur l'âge en fin d'année (classes d'âge de 5 ans pour les 0-14 ans, classes d'âge de 1 an pour les 0-4 ans) et sur le département de résidence ; [2] Age en fin d'année

C.1.b Etude d'incidence 2002-2007

Sur la période 2002-2007 (Tableau 18), 2 831 cas de LA de moins de 15 ans ont été inclus dans l'étude d'incidence. Parmi eux, 14 sont classés dans la zone la plus proche des CNPE, puisqu'ils habitaient dans une commune dont la mairie est située à moins de 5 km d'une CNPE. Dans ces communes, 7,4 cas étaient attendus et le SIR correspondant est donc proche de 2. Il diffère significativement de 1 (SIR=1,9 [1,0-3,2], p unilatéral= 0,02). De même chez les 0-4 ans, 8 cas ont été

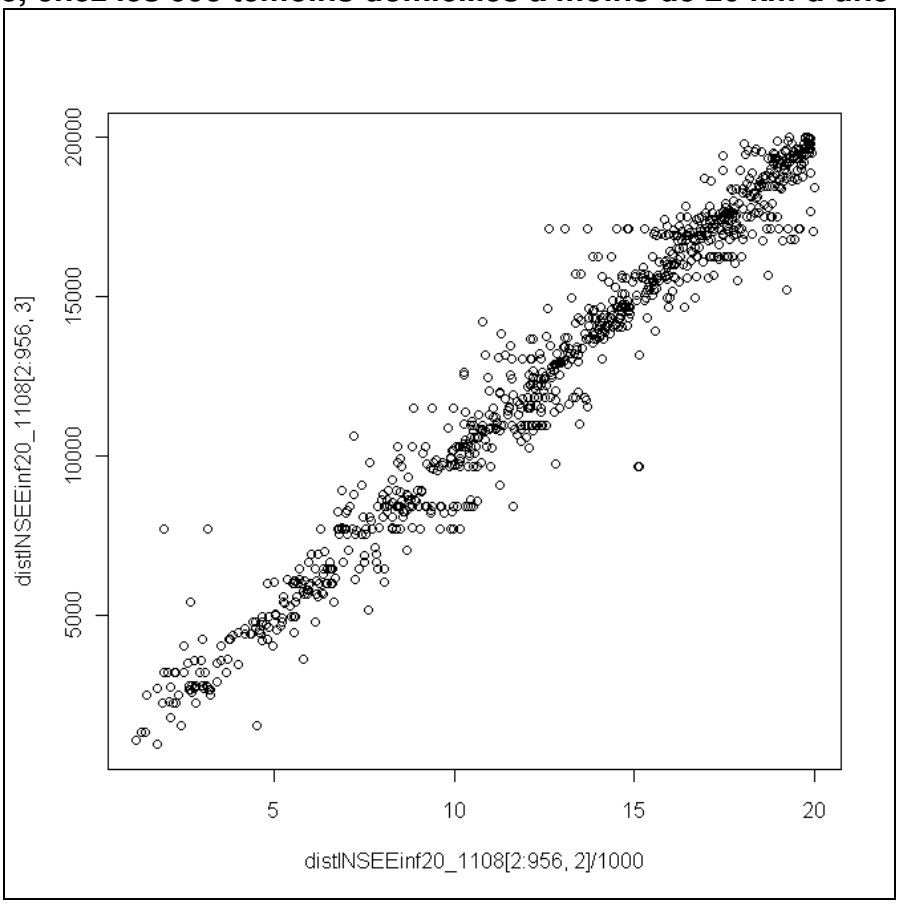
observés alors que 3,6 étaient attendus (SIR =2,2 [1,0-4,4], p unilatéral= 0,03). Les enfants qui avaient 14 ans en années révolues au diagnostic mais 15 ans à la fin de l'année du diagnostic ne sont pas inclus dans l'étude cas-témoins basée sur l'âge à la fin de l'année du diagnostic, c'est la raison pour laquelle il y a un peu plus de cas inclus dans l'étude d'incidence que dans l'étude cas-témoins.

Aucun excès particulier n'a été détecté ni dans les autres zones qui entourent les CNPE à plus de 5 km ni globalement dans la zone 0-20 km. A moins de 20 km, les nombres de cas observés sont très proches des nombres de cas attendus (O = 99 ; E = 95,7). Même si une tendance croissante significative a été détectée pour le SIR avec l'inverse de la distance aux CNPE (p unilatéral=0,001), les SIR observés à plus de 5 km d'une CNPE sont tous proches de 1 et n'en diffèrent pas significativement. Les différences statistiquement significatives observées entre les nombres de cas observés et attendus sont donc mieux décrits par une augmentation d'incidence à moins de 5 km d'une CNPE que par une tendance croissante du taux d'incidence avec la proximité croissante d'une CNPE.

Les résultats obtenus en utilisant l'approche "incidence" basée sur la distance de la mairie de la commune à la CNPE la plus proche et ceux obtenus en utilisant l'approche cas-témoins basée sur la distance de l'adresse individuelle à la CNPE la plus proche sont quantitativement et qualitativement très proches puisque les OR et SIR sont très proches tout comme leurs intervalles de confiance respectifs et le degré de significativité statistique de leurs écarts à 1. Ceci est lié au fait que ces deux types de distances sont très corrélés (Figure 10). Chez les 995 témoins dont la distance de l'adresse individuelle à la CNPE la plus proche et la distance de la mairie

de la commune de résidence à la CNPE la plus proche sont inférieures à 20 km, ces deux distances sont en effet fortement corrélées (r = + 0,97).

Figure 10 : Corrélation entre la distance de l'adresse individuelle à la CNPE la plus proche et la distance de la mairie de la commune à la CNPE la plus proche, chez les 995 témoins domiciliés à moins de 20 km d'une CNPE



C.1.c Etude d'incidence 1990-2001

Il n'y a que quelques différences d'effectifs rares et limitées entre les résultats publiés précédemment par notre équipe sur la période 1990-1998 [White-Koning *et al.*,2004] et ceux calculés *a posteriori* sur la même période. Elles sont expliquées par les quelques améliorations qui ont été apportées au RNHE comme l'exclusion ou l'ajout de certains cas (peu nombreux puisque le registre était déjà quasiment exhaustif lors des premiers calculs), la correction d'un diagnostic, d'une date de diagnostic ou d'une adresse. Par ailleurs, des fusions ou réorganisations de communes pouvant

avoir eu lieu, la localisation communale de quelques cas a pu également légèrement changer.

L'étude mise à jour montre, sur la période 1990-2001, 5 356 cas de LA diagnostiqués en France chez les enfants de moins de 15 ans. Parmi ceux-ci, 173 habitaient à moins de 20 km d'une CNPE alors que 182,9 étaient attendus et 10 à moins de 5 km alors que 13,6 étaient attendus. Le SIR pour cette dernière zone était de 0,7 [0,4-1,4], il ne différait pas significativement de 1 (cf Tableau 18 ci-dessus). L'analyse par groupe d'âge ne montre pas de résultats spécifiques pour une catégorie.

Par rapport à l'étude précédente de notre équipe sur l'incidence de LA et la distance aux CNPE sur la période 1990-1998, les résultats sont inchangés et nous retrouvons une absence d'augmentation d'incidence des leucémies de l'enfant autour des centrales nucléaires quelque soit la tranche d'âge observée entre 0 et 14 ans.

Dans une étude précédente, la même période 1990-2001 a été étudiée en utilisant un indicateur différent: l'estimation des doses dues aux rejets gazeux des CNPE [Evrard *et al.*,2006]. Aucune association entre cet indicateur et l'incidence des LA n'avait été trouvée et l'étude mise à jour donne les mêmes résultats (Tableau 18).

C.1.d Etude d'incidence 1990-2007

Sur la période 1990-2007, 8 187 cas de LA ont été diagnostiqués en France chez les enfants de moins de 15 ans (Tableau 18). Au total, 272 cas ont été diagnostiqués chez des enfants habitant à moins de 20 km de l'une des 19 CNPE entre 1990 et 2007 et 278,6 cas étaient attendus (SIR=1,0 [0,9–1,1]). Parmi les cas incidents, 24 habitaient à moins de 5 km d'une CNPE alors que 21,0 étaient attendus. Le SIR correspondant était de 1,1 [0,7-1,7]. Bien qu'il soit légèrement supérieur au SIR de la

période 1990-2001, il ne diffère pas significativement de 1 (p unilatéral = 0,3) (Tableau 18). Chez les 0-4 ans, le même résultat est observé: un SIR légèrement supérieur à 1 (1,4 [0,8-2,3]) mais non significatif (p unilatéral=0,15).

Aucune des quatre zones définies autour des CNPE ne présentait d'excès significatif de cas (Tableau 18) et aucune variation tendancielle des SIR avec la distance aux CNPE n'a été observée ni pour la totalité des enfants, ni pour aucune des trois tranches d'âge étudiées.

C.1.e Analyses de sensibilité, prise en compte de facteurs de confusion potentiels et analyses par sous-groupe

Par année

L'analyse 2002-2007 par année d'étude, en incidence comme en cas-témoins montre des SIR et OR élevés principalement pour les années 2003, 2006 et 2007 (Tableau 21). Le nombre de sujets par année étant faible, nous avons aussi considéré les observations obtenues sur 5 années, en excluant séquentiellement chacune des 6 années 2002 à 2007. Les SIR et OR (0-5 km) obtenus restent du même ordre de grandeur quand les années 2006 ou 2007 sont exclues même si leur écart à 1 n'est plus statistiquement significatif.

Pour la zone 0-5 km et après correction de Bonferroni pour tenir compte du fait que 6 tests ont été réalisés, le SIR de l'année 2006 est significativement supérieur à 1 (p unilatéral $_{Bonf}$ = 0,05). S'il n'y a pas d'hétérogénéité statistiquement significative entre les 6 SIR (p = 0,14 et par simulation p = 0,20), ils ne peuvent cependant pas être considérés comme tous égaux à 1 (p = 0,001 et par simulation p = 0,008).

Entre 1990 et 2007 seules les années 2006 et 2007 ont un SIR significativement supérieur à 1 à moins de 5 km d'une CNPE.

Cependant, après correction de Bonferroni (18 années d'observation), ces excès n'étaient plus statistiquement significatifs. Pour la période 1990-2007, l'hétérogénéité des 18 SIR observés était proche de la significativité statistique (p = 0,12 et par simulation p = 0,09) et l'ensemble de leurs écarts à 1 était proche de la significativité statistique (p = 0,06 and par simulation p = 0,09).

Par CNPE

Les résultats détaillés par CNPE sur la période 2002-2007 ne révèlent pas qu'une CNPE ou un groupe de CNPE aient une incidence particulièrement élevée à moins de 5 km (Tableau 21 – Figure 11).

Tableau 21: Distance à la CNPE la plus proche et leucémie aigüe de l'enfant - analyses de l'étude cas-témoin Geocap (adresse individuelle) et de l'étude d'incidence (adresse de la mairie), 2002-2007, stratifiées sur la puissance électrique du CNPE, l'année d'étude

	Etude cas-témoins Geocap										**Etude d'incidence, 2002-2007[2]**									
	< 5 km *vs,* ≥ 20 km					< 20 km *vs,* ≥ 20 km					< 5 km					< 20 km				
	Cas	Témoins	OR	IC 95%	p	Cas	Témoins	OR	IC 95%	p	O	E	SIR	IC 95%	p	O	E	**SIR**	IC 95%	p
Puissance électrique																				
5200-5400 MWe	3	22	**1,4**	[0,4-4,9]		28	303	**0,9**	[0,6-1,4]		3	1,8	**1,7**	[0,3-4,8]		27	26,2	**1,0**	[0,7-1,5]	
3600 MWe	4	22	**2,1**	[0,7-6,3]	*	28	325	**1,0**	[0,6-1,4]		3	1,9	**1,6**	[0,3-4,5]		27	31,6	**0,9**	[0,6-1,2]	
2600-2800 MWe	6	29	**2,1**	[0,8-5,1]	#	33	345	**1,1**	[0,7-1,5]		7	3,0	**2,3**	[0,9-4,8]	*	34	30,8	**1,1**	[0,8-1,5]	
1800 MWe	1	7	**1,4**	[0,2-11,6]		10	87	**1,3**	[0,6-2,6]		1	0,6	**1,7**	[0,0-9,3]		11	7,1	**1,5**	[0,8-2,8]	
Année																				
2002	0	12	**0,0**			15	180	**0,8**	[0,5-1,5]		1	1,2	**0,8**	[0,0-4,5]		15	16,2	**0,9**	[0,5-1,5]	
2003	3	15	**2,1**	[0,6-7,5]		18	182	**1,2**	[0,7-2,0]		3	1,2	**2,4**	[0,5-7,1]		17	16,2	**1,1**	[0,6-1,7]	
2004	1	13	**0,8**	[0,1-6,4]		19	168	**1,2**	[0,7-2,0]		1	1,3	**0,8**	[0,0-4,4]		18	16,3	**1,1**	[0,7-1,7]	
2005	1	13	**0,8**	[0,1-6,6]		14	161	**1,1**	[0,6-1,9]		0	1,2	**0,0**	[0,0-3,1]		17	15,5	**1,1**	[0,6-1,8]	
2006	5	13	**4,9**	[1,6-14,8]	*	18	175	**1,1**	[0,6-1,9]		5	1,2	**4,1**	[1,3-9,6]	*	17	15,7	**1,1**	[0,6-1,7]	
2007	4	14	**3,9**	[1,2-12,9]	*	15	195	**0,8**	[0,4-1,4]		4	1,2	**3,3**	[0,9-8,4]	*	15	15,7	**1,0**	[0,5-1,6]	

Localisation côtière																				
Oui	3	18	**1,7**	[0,5-6,0]		26	258	**1,1**	[0,7-1,3]	3	1,5	**2,0**	[0,4-5,9]		29	24,7	**1,2**	[0,8-1,7]		
Non	11	62	**1,9**	[1,0-3,6]	*	73	802	**1,0**	[0,8-1,3]	11	5,9	**1,9**	[0,9-3,4]	*	70	71,1	**1,0**	[0,8-1,2]		
Localisation frontalière																				
Oui	4	19	**1,9**	[0,6-5,9]		17	187	**0,9**	[0,5-1,5]	4	1,3	**3,0**	[0,8-7,6]	*	19	16,4	**1,2**	[0,7-1,8]		
Non	10	61	**1,8**	[0,9-3,6]		82	873	**1,1**	[0,8-1,3]	10	6,0	**1,7**	[0,8-3,1]	#	80	79,3	**1,0**	[0,8-1,3]		
Taille d'unité urbaine de la commune																				
Rurale	9	49	**2,2**	[1,0-4,7]	*	49	445	**1,2**	[0,8-1,6]	9	3,9	**2,3**	[1,1-4,4]	*	51	41,8	**1,2**	[0,9-1,6]	#	
Semi-urbaine	3	16	**2,2**	[0,6-7,9]		26	388	**0,8**	[0,5-1,3]	3	2,5	**1,2**	[0,2-3,5]		26	34,6	**0,8**	[0,5-1,1]		
Urbaine	2	15	**1,5**	[0,3-6,6]		24	227	**1,3**	[0,8-2,1]	2	0,9	**2,1**	[0,2-7,7]		22	19,4	**1,1**	[0,7-1,7]		
Revenu médian de la commune																				
< médiane	11	51	**2,3**	[1,2-4,4]	*	63	703	**1,0**	[0,8-1,3]	11	4,6	**2,4**	[1,2-4,3]	*	63	64,7	**1,0**	[0,8-1,3]		
≥ médiane	3	27	**1,1**	[0,3-3,6]		35	352	**1,1**	[0,7-1,6]	3	2,8	**1,1**	[0,2-3,2]		36	31,0	**1,2**	[0,8-1,6]		
Part d'ouvriers dans la commune																				
< médiane	2	15	**1,6**	[0,4-7,3]		18	175	**1,2**	[0,7-2,1]	2	1,4	**1,5**	[0,2-5,2]		18	15,8	**1,1**	[0,7-1,8]		
≥ médiane	12	65	**1,9**	[1,0-3,5]	*	81	885	**1,0**	[0,7-1,2]	12	6,0	**2,0**	[1,0-3,5]	*	81	79,9	**1,0**	[0,8-1,3]		
Part de bacheliers dans la commune																				
< médiane	13	70	**1,9**	[1,0-3,6]	*	79	819	**1,0**	[0,8-1,3]	13	6,0	**2,2**	[1,2-3,7]	*	78	71,5	**1,0**	[0,9-1,4]		
≥ médiane	1	10	**1,4**	[0,2-11,6]		20	241	**1,0**	[0,6-1,6]	1	1,4	**0,7**	[0,0-4,1]		21	21,3	**0,9**	[0,5-1,3]		

[1] Odds ratios (OR) et intervalles de confiance à 95% (IC95%) estimés par régression logistique ajustée sur l'âge en fin d'année (classes d'âge de 5 ans pour les 0-14 ans, classes d'âge de 1 an pour les 0-4 ans) et sur le département de résidence ; [2] Ratio d'incidence standardisé (SIR) calculé comme le radio du nombre de cas observés (O) sur le nombre de cas attendus (E) et intervalle de confiance à 95% estimé avec l'approximation de Byar (95%IC); * $p_{unilatéral}$ <0,05; # $p_{unilatéral}$ <0,10

Aucune CNPE n'a un OR significativement supérieur à 1 et même si les CNPE de Chooz, Civaux et Cruas ont un SIR supérieur à 1 de façon significative ou proche de la significativité (respectivement, p unilatéral = 0,01; 0,09; 0,06), ils ne le restent pas quand la multiplicité des tests (18 CNPE considérées isolément parmi les 19, sauf *Le Blayais*) est prise en compte par la correction de Bonferroni. Par ailleurs, quand chaque CNPE est exclue tour à tour, les SIR et OR sur les 17 autres CNPE restent très proches des SIR et OR globaux et significativement (ou quasi-significativement) différents de 1 (Tableau 22).

Tableau 22 : LA à moins de 5 km d'une CNPE, 2002-2007 –estimations après exclusion, une à une, de chaque CNPE

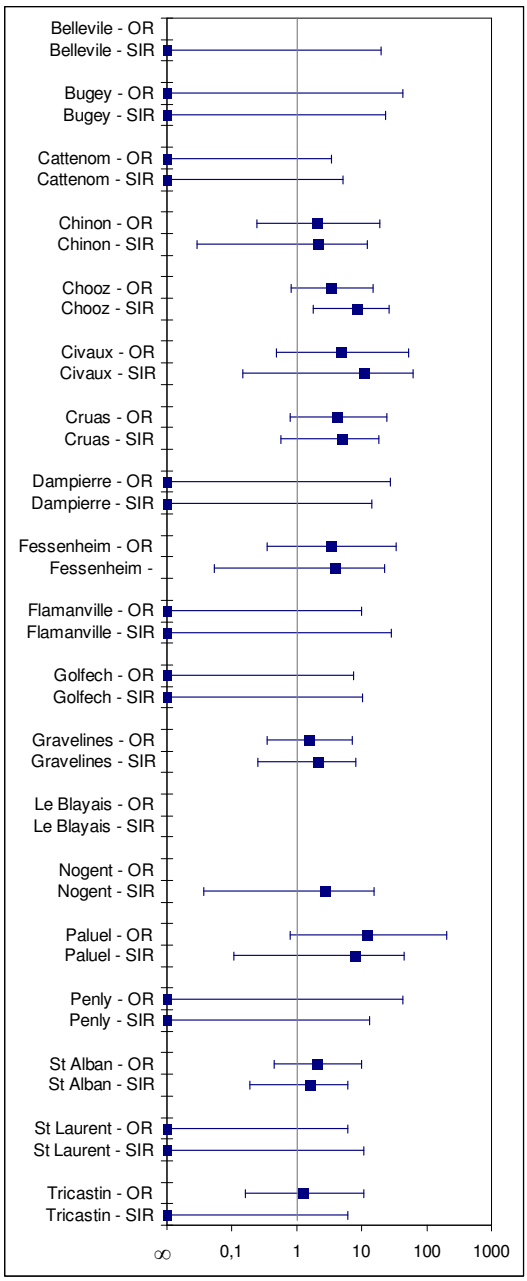
	Etude cas-témoins [a]		Etude d'incidence [b]	
	OR	**IC 95%**	**SIR**	**IC 95%**
sans Bellevile	1,9	[1,0-3,3]	2,0	[1,1-3,3]
sans Bugey	1,9	[1,1-3,4]	1,9	[1,1-3,3]
sans Cattenom	2,1	[1,2-3,7]	2,1	[1,2-3,6]
sans Chinon	1,8	[1,0-3,4]	1,9	[1,0-3,2]
sans Chooz	1,7	[0,9-3,2]	1,6	[0,8-2,8]
sans Civaux	1,8	[1,0-3,2]	1,8	[1,0-3,1]
sans Cruas	1,7	[0,9-3,1]	1,7	[0,9-3,0]
sans Dampierre	1,9	[1,1-3,4]	2,0	[1,1-3,3]
sans Fessenheim	1,8	[1,0-3,3]	1,8	[1.0-3,1]
sans Flamanville	1,9	[1,1-3,4]	1,9	[1,1-3,3]
sans Golfech	2,0	[1,1-3,5]	2,0	[1,1-3,4]
sans Gravelines	1,9	[1,0-3,6]	1,9	[1,0-3,3]
sans Le Blayais	1,9	[1,0-3,3]	1,9	[1,0-3,2]
sans Nogent	1,9	[1,0-3,3]	1,9	[1,0-3,2]
sans Paluel	1,7	[1,0-3,2]	1,8	[1,0-3,1]
sans Penly	1,9	[1,1-3,4]	2,0	[1,1-3,3]
sans Saint-Alban	1,8	[1,0-3,4]	2,0	[1,0-3,4]
sans Saint-Laurent	2,0	[1,1-3,5]	2,0	[1,1-3,3]
sans Tricastin	1,9	[1,1-3,5]	2,1	[1,1-3,5]

[a] Odds ratios (OR) et intervalles de confiance à 95% (IC95%) estimés par régression logistique ajustée sur l'âge en fin d'année (classes d'âge de 5 ans pour les 0-14 ans, classes d'âge de 1 an pour les 0-4 ans) et sur le Département de résidence; [b] Ratios d'incidence standardisés (SIR) calculés comme le ratio du nombre de cas observés (O) sur le nombre de cas attendus (E) avec l'approximation de Byar pour l'intervalle de confiance à 95% (IC95%);

Le SIR et l'OR de Chooz à moins de 20 km sont tous deux significativement supérieurs à 1 mais ce résultat n'est pas très différent du résultat sur 0-5 km puisque

la configuration de cette CNPE est très particulière : les zones 0-5 km et 0-20 km autour de cette CNPE sont presque identiques car cette CNPE est très proche de la frontière franco-belge et nichée dans l'angle qu'elle forme à cet endroit.

Figure 11: SIR et OR < 5km d'une CNPE, par CNPE, sur la période 2002-2007



L'étude d'incidence à moins de 5 km par CNPE sur la période globale 1990-2007 n'a pas mis en évidence de CNPE avec un risque particulièrement supérieur à 1. En effet même si certains SIR sont significativement supérieurs à 1, ils ne le restent pas après correction de Bonferroni. Il y a une hétérogénéité non significative entre les 18 CNPE (p = 0,13 et par simulation p = 0,11) et le test d'hypothèse selon laquelle ils sont tous égaux à 1 est quasi-significatif (p = 0,07 et par simulation p = 0,11).

Caractéristiques des CNPE

Les analyses ont été stratifiées sur la puissance électrique des CNPE et aucun résultat spécifique n'a été trouvé sur un groupe particulier, ni aucune tendance avec la puissance électrique (Tableau 21).

Sur la période 2002-2007, les observations réalisées sur les CNPE situées au bord de la Manche sont identiques aux observations faites sur l'ensemble des CNPE: OR =1,7 [0,5-6,0] et il en est de même pour les observations faites sur les CNPE situées au bord d'une rivière : OR =1,9 [1,0-3,6]. Ces deux groups de CNPE, qui ont des structures de rejets liquides *vs* gazeux particulières, ne présentent donc pas de particularité par rapport à l'ensemble des CNPE.

Enfin, les analyses séparées des CNPE à moins de 20km d'une frontière et des autres n'ont pas montré de différence entre ces groupes: OR (CNPE frontalières)=1,9 [0,6-5,9] et OR (CNPE non frontalières)=1,8 [0,9-3,6].

Prise en compte des caractéristiques contextuelles communales

Les analyses stratifiées sur le type d'unité urbaine des communes n'ont pas mis en évidence de résultats particuliers. En effet, même si les SIR et OR calculés sur les sujets habitants dans les communes les plus urbaines n'étaient pas significativement

différents de 1, il y avait deux fois moins d'exposés dans les communes urbaines par rapport aux communes les plus rurales (Tableau 21).

Les analyses réalisées en excluant les sujets situés à plus de 50 km d'une CNPE n'ont pas non plus montré de résultats particuliers : sur la période 2002-2007, l'OR associé à la proximité à moins de 5 km d'une CNPE était similaire à celui obtenu dans les études principales : OR=1,7 [0,95-3,1]. Par ailleurs, le SIR entre 20 et 50 km des CNPE est proche de 1 : SIR=1,05 [0,96-1,15].

Les variables socio-économiques étudiées étaient par contre associées à la distance aux CNPE : les communes situées à moins de 20 km d'une CNPE étaient, en moyenne, moins avantagées que les autres communes françaises, mais ce résultat n'était pas spécifique de la zone 0-5 km. Parmi ces caractéristiques contextuelles des communes, seul le revenu médian est associé significativement au statut cas-témoin et cette association est faible : OR=1,1 [1,0-1,2]. L'augmentation de risque détectée à moins de 5 km d'une CNPE semble plus marquée dans les communes ayant un revenu médian ou une part de bacheliers chez les plus de 15 ans inférieur à la médiane des communes françaises. Les analyses ajustées sur ces 4 variables ne modifiaient cependant pas les résultats (Tableau 21).

Proximité à une ligne à haute tension

Dans l'étude cas-témoins 2002-2007, la distance à une CNPE et celle à une ligne à haute tension sont corrélées : il y a plus de sujets témoins proches d'une LHT à moins de 20 km d'un CNPE (22% à moins de 400 m de la LHT la plus proche) qu'au-delà (11% à moins de 400 m de la LHT la plus proche). Quand les analyses sont stratifiées ou ajustées sur la distance à la LHT la plus proche (< 400 m de la LHT vs

≥ 400 m), les résultats restent cependant inchangés, tout comme quand la coupure est choisie à 200 m ou 600 m d'une LHT.

C.2 Estimations de doses liées aux rejets gazeux radioactifs des CNPE

La distribution des sujets selon les estimations de dose à la moelle liées aux rejets radioactifs gazeux (DBGZ) est sensiblement différente de leur distribution selon la distance aux CNPE (Tableau 23) : 40% des témoins de la catégorie des plus exposés (> 0,72 µSv) étaient dans la zone 0-5 km d'une CNPE, 44% dans la zone 5-10 km, 11% dans la zone 10-15 km et 4% dans la zone 15-20 km.

Tableau 23 : Période 2002-2007: Classification croisée des témoins, selon la distance « mairie - CNPE la plus proche » et l'estimation de dose à la moelle due aux rejets gazeux radioactifs des CNPE à la localisation de la mairie

distance *mairie*	Dose à la moelle estimée pour la *commune*					
	Référence	>0 µSv et ≤0,093 µSv	0,094-0,20 µSv	0,21-0,72 µSv	> 0,72 µSv	Total
0-4,99 km	0	5	5	38	39	87
5-9,99 km	0	33	50	93	43	219
10-14,99 km	0	119	104	62	11	296
15-19,99 km	1	260	158	20	4	443
≥ 20 km	28 955	0	0	0	0	28 955
Total	28 956	417	317	213	97	30 000

[a] L'estimateur "Dose Based Geographic Zoning" est basé sur l'estimation de la dose de radioactivité à la moelle associée aux rejets gazeux des CNPE à la mairie de la Commune et exprimée en µSv. La "Référence" de l'estimateur DBGZ inclut tous les sujets qui ont une estimation de dose nulle et ceux qui résident dans des communes dont la mairie est située à plus de 20km d'une CNPE.

Aucune association entre l'incidence des LA et DBGZ n'a été observée dans l'étude cas-témoins ni dans l'étude d'incidence sur la période 2002-2007 (Tableau 24). Les SIR et OR étaient proches de 1 pour toutes les catégories de DBGZ.

Tableau 24 : Association entre l'estimateur de dose basé sur un zonage géographique (Dose Based Geographic Zoning - DBGZ [a]) autour des CNPE et les leucémies aigues de l'enfant – Etude cas-témoins Geocap et études d'incidence

	Etude cas-témoins, 2002-2007[b] (distance: mairie-CNPE la plus proche)				Etude d'incidence, 2002-2007[c] (distance: mairie-CNPE la plus proche)				Etude d'incidence, 1990-2001[c] (distance: mairie-CNPE la plus proche)				Etude d'incidence 1990-2007[c] (distance: mairie-CNPE la plus proche)			
	Cas	**Témoins**	**OR**	**IC95%**	**O**	**E**	**SIR**	**IC95%**	**O**	**E**	**SIR**	**IC95%**	**O**	**E**	**SIR**	**IC95%**
Age < 15 ans[d]																
> 0,72 µSv [a]	8	97	1,0	[0,5-2,1]	8	8,3	1,0	[0,4-1,9]	13	16,4	0,8	[0,4-1,4]	21	24,7	0,9	[0,5-1,3]
0,21-0,71 µSv	19	213	1,0	[0,6-1,6]	20	18,5	1,1	[0,7-1,7]	42	36,7	1,1	[0,8-1,6]	62	55,2	1,1	[0,9-1,4]
0,094-0,20 µSv	29	317	1,0	[0,7-1,5]	31	30,0	1,0	[0,7-1,5]	54	56,8	1,0	[0,7-1,2]	85	86,7	1,0	[0,8-1,2]
>0 µSv et ≤ 0,093 µSv	40	417	1,0	[0,7-1,4]	40	39,0	1,0	[0,7-1,4]	64	73,1	0,9	[0,7-1,1]	104	112,0	0,9	[0,8-1,1]
Référence	2657	28956	1,0													
<20 km et >0 µSv	96	1044	1,0	[0,8-1,3]	99	95,7	1,0	[0,8-1,3]	173	182,9	1,0	[0,8-1,1]	272	278,6	1,0	[0,9-1,1]
Référence	2657	28956	1,0													
Age < 5 ans[d]																
> 0,72 µSv [a]	4	28	1,1	[0,4-3,2]	5	4,0	1,2	[0,4-2,9]	7	8,0	0,9	[0,4-1,8]	12	12,0	1,0	[0,5-1,8]
0,21-0,71 µSv	6	59	0,9	[0,4-2,0]	8	9,2	0,9	[0,4-1,7]	23	17,9	1,3	[0,8-1,9]	31	27,1	1,1	[0,8-1,6]
0,094-0,20 µSv	15	88	1,3	[0,7-2,3]	16	14,9	1,1	[0,6-1,7]	29	27,9	1,0	[0,7-1,5]	45	42,8	1,1	[0,8-1,4]
>0 µSv et ≤ 0,093 µSv	16	137	0,9	[0,5-1,5]	18	19,1	0,9	[0,6-1,5]	25	35,3	0,7	[0,5-1,1]	43	54,5	0,8	[0,6-1,1]
Référence	1248	9396	1,0													
< 20 km et>0 µSv	41	312	1,0	[0,7-1,4]	47	47,3	1,0	[0,7-1,3]	84	89,1	1,0	[0,8-1,1]	131	136,4	1,0	[0,8-1,1]
Référence	1248	9405	1,0													

[a] L'estimateur "Dose Based Geographic Zoning" est basé sur l'estimation de la dose d'irradiation à la moelle associée aux rejets gazeux des CNPE à la mairie de la Commune et exprimée en µSv. La "Référence" de l'estimateur DBGZ inclut tous les sujets qui ont une estimation de dose nulle ou résidante dans des communes dont la mairie est située à plus de 20 km d'une CNPE.

[b] Odds ratios (OR) et intervalles de confiance à 95% (IC95%) estimés par régression logistique ajustée sur l'âge en fin d'année (classes d'âge de 5 ans pour les 0-14 ans, classes d'âge de 1 an pour les 0-4 ans) et sur le Département de résidence; [c] Ratios d'incidence standardisés (SIR) calculés comme le ratio du nombre de cas observés (O) sur le nombre de cas attendus (E) avec l'approximation de Byar pour l'intervalle de confiance à 95% (IC95%);

[d] Age en fin d'année dans l'étude Geocap; âge en années révolues dans les études d'incidence

Les résultats étaient similaires avec des catégories de DBGZ faites à partir des quartiles de la distribution de l'exposition des témoins ou avec les catégories utilisées dans l'étude précédente [Evrard *et al.*,2006] et l'exclusion des CNPE proches d'autres sites nucléaires ne modifiait pas ces observations. Enfin, aucune association entre LA et DBGZ n'a été trouvée pour la période globale 1990-2007 (Tableau 24).

D Discussion

L'étude cas-témoins GEOCAP met en évidence une association entre l'incidence des LA de l'enfant et le fait d'habiter à moins de 5 km d'une CNPE sur la période 2002-2007. Cette association a aussi été observée avec l'étude d'incidence sur la même période. Cependant, l'étude d'incidence sur la période précédente 1990-2001 n'a mis en évidence aucune association. L'utilisation d'estimations de doses de radiation délivrée à la moelle et provenant des déchets gazeux des CNPE fournissent des résultats très différents puisque les SIR et OR étaient très proches de 1 et n'en différaient pas significativement dans toutes les catégories d'exposition. L'association observée sur la période 2002-2007 avec le fait de résider à moins de 5 km d'une CNPE n'était pas limitée à catégorie d'âge, à une CNPE ou un groupe de CNPE, à une ou plusieurs années.

Une des forces de l'étude est que les cas de LA sont tous ceux identifiés dans le RNHE, dont le recrutement a couvert la France entière depuis 1990 avec une très bonne exhaustivité [Clavel *et al.*,2004]. D'autre part, les témoins ont été échantillonnés à partir d'une base quasi-exhaustive puisqu'elle contient tous les enfants appartenant à un foyer fiscal. Cet échantillon témoin était bien représentatif de la base de tirage initiale. Les deux approches, cas-témoins et incidence, auraient

pu comporter différentes sortes de biais. Cependant, le fait que ces deux approches aboutissent à des résultats très similaires est un argument fort en faveur de la validité de l'estimation des OR et SIR, puisque même si, dans ces deux approches, les cas ont été inclus à partir du même registre le RNHE, les populations auxquelles ils sont rapportés sont estimées de deux façons très différentes (estimation des populations communales fondées sur les données des recensements pour l'approche "incidence", estimation fondée sur le groupe témoin constitué à partir des déclarations de revenus et taxes d'habitation pour l'approche "cas-témoins". Ces résultats nous confortent donc la représentativité de l'échantillon témoin de l'étude GEOCAP et la validité des estimations de populations utilisées dans l'étude d'incidence.

Les ajustements faits sur les caractéristiques socio-démographiques de la commune de résidence et sur l'âge de l'enfant ont permis d'écarter un certain nombre de facteurs de confusion contextuels potentiels.

Une des principales limites de cette étude est que les données n'ont pas permis l'ajustement sur des facteurs de risque individuels des LA comme le rang de naissance, l'allaitement, le mode de garde ou l'exposition aux pesticides. Cependant, il n'y a aucune raison évidente pour que ces expositions individuelles diffèrent selon le fait d'habiter à moins ou plus de 5 km d'une CNPE, conditionnellement au degré d'urbanisation ou à d'autres caractéristiques socio-démographiques de la commune prises en compte. Les mouvements de population n'ont pas été étudiés ici.

Comme beaucoup d'études de l'incidence des leucémies aigues de l'enfant au voisinage des CNPE, nous n'avons pas pu prendre en compte dans l'étude GEOCAP l'histoire résidentielle des sujets, ce qui est bien sûr une autre limite importante dans l'évaluation de l'exposition des sujets aux radiations et à tout autre facteur lié à la

proximité des CNPE. Le fait de restreindre les analyses aux sujets de moins de 5 ans, qui ont une probabilité plus faible d'avoir changé de domicile depuis leur naissance que les autres enfants plus âgés, a conduit à des OR et SIR du même ordre de grandeur.

Trois CNPE étaient situées à moins de 20 km d'une frontière. Or seules les données françaises sont disponibles sur les disques de 20 km de rayon autour de ces CNPE. Cependant, l'exclusion de ces trois CNPE ne modifie pas les résultats globaux. Par ailleurs, aucune CNPE d'un pays frontalier n'est située à moins de 20 km d'une frontière française.

La proximité des CNPE situées en bord de Manche de l'usine de retraitement de la Hague a été prise en compte en analysant ces CNPE séparément, ce qui n'a pas montré de résultat particulier.

Enfin, il est important de noter que les résultats des études complémentaires faites par CNPE ou par année*,* même si elles permettent vraisemblablement d'écarter l'existence d'un excès localisé à une unique CNPE et persistant, comme on a pu en observer en Angleterre à Sellafield, en Ecosse à Dounreay et en Allemagne à Krümmel, sont à interpréter avec précaution compte tenu de la faiblesse des effectifs en jeu et de la multiplicité des tests mis en œuvre.

L'estimation des doses de radiation délivrées à la moelle provenant des déchets gazeux des CNPE a été réalisée à partir des niveaux moyens de rejets annuels de déchets, de la composition des déchets et de certains paramètres météorologiques locaux. Les données concernant les rejets réels de carbone 14 n'étant pas disponibles, ils ont été estimés comme égaux aux limites réglementaires de rejets. La contribution du carbone 14 à la dose de radiation totale peut donc avoir été

surestimée. Cependant, ceci a été fait de la même façon pour toutes les CNPE. Par conséquent, malgré certaines erreurs potentielles d'estimations de doses, les communes ont probablement été correctement ordonnées en termes d'exposition, liées aux CNPE. Le modèle n'a pas pris en compte les rejets liquides dont la contribution à la dose totale réelle de radiations pour les individus peut être du même ordre de grandeur que les rejets gazeux. Cependant, celle-ci peut être considérée comme homogène sur la zone étudiée. D'autre part, alors même que les rejets liquides des CNPE situées proches d'une côte et celles des CNPE proches d'une rivière sont différents, les analyses stratifiées sur cette variable n'ont pas fait apparaître de différence dans l'association avec l'incidence de LA.

Globalement, les doses estimées dues aux rejets gazeux des CNPE étaient extrêmement faibles comparées à celles dues à des sources de radiation naturelles et on ne s'attend pas, d'après ce qui est connu à ce jour, à ce qu'elles soient à l'origine d'un excès de risque de leucémie [UNSCEAR,2000].

Dans les deux études d'incidence françaises multi-sites précédentes [White-Koning *et al.*,2004]; [Evrard *et al.*,2006] aucune association entre la proximité aux CNPE et l'incidence de LA n'a été observée. Ce résultat était concordant avec la plupart des études multi-sites précédentes [Bithell *et al.*,2008]; [Laurier *et al.*,2008]; [COMARE, 2011]. En Allemagne, l'étude cas-témoin KiKK qui concernait les enfants de moins de 5 ans a mis en évidence une association entre l'incidence de LA et la proximité aux CNPE [Kaatsch *et al.*,2008]. Ce résultat allemand semblait spécifique de cette tranche d'âge 0-4 ans, ce qui n'est pas le cas dans notre étude. Dans l'étude KiKK, une tendance croissante avec l'inverse de la distance aux CNPE, considérée en variable continue avait été observée, mais pas quand elle était considérée en

variable catégorielle. Dans l'étude GEOCAP, une tendance croissante légère et non-significative des OR et SIR avec l'inverse de la distance a été observée.

A notre connaissance, aucune autre étude n'a utilisé de modèle quantitatif utilisant une estimation des rejets gazeux des CNPE pour classer les lieux de résidence en fonction de l'exposition aux radiations.

Les CNPE sont les seuls sites nucléaires étudiés ici, les autres sites faisant partie du cycle nucléaire (enrichissement de l'uranium, stockage et retraitement des déchets) ne sont pas pris en compte, une définition précise de ces sites faisant encore l'objet de discussions chez les spécialistes de la radioprotection et de la sûreté nucléaire.

En conclusion, l'augmentation d'incidence observée dans cette étude à moins de 5 km des CNPE ne conforte que partiellement les derniers résultats allemands puisqu'elle ne concerne que les années récentes et n'est pas limitée aux enfants les plus jeunes (0-4 ans). L'absence d'association entre l'incidence des LA et l'estimation locale de l'exposition aux radiations liées aux rejets gazeux des CNPE semble indiquer que l'association avec la distance aux CNPE n'est pas expliquée par l'exposition aux radiations dues aux CNPE en fonctionnement normal. Au total, les résultats suggèrent que l'excès de risque observé sur 2002-2007 pourrait être du à un ou plusieurs facteurs non identifiés liés à la proximité des CNPE, y compris des expositions aux radiations non ou mal cernées par les estimations utilisées ici. La présence de tels facteurs potentiels pourraient être transitoire ou au contraire persistante. Ces résultats appellent donc une poursuite de l'analyse de l'incidence dans le temps, une recherche approfondie des facteurs de risque potentiellement liés à la proximité de CNPE ainsi qu'une analyse collaborative de toutes les données disponibles provenant des études multi-sites faites dans le monde.

VI. Proximité aux lignes à haute tension et leucémies aigües de l'enfant

A Introduction

Les champs magnétiques à extrêmement basse fréquence (CM-EBF) sont générés à proximité de lignes, caténaires, transformateurs et appareils électriques dans lesquels passe du courant électrique alternatif 50 Hz (ou 60 Hz en Amérique du Nord). Les lignes de transport d'électricité à haute tension (LHT) n'expliquent qu'une partie des expositions élevées aux CM-EBF. Toutefois, ce sont des sources d'exposition qu'il est facile d'identifier alors que les autres installations à l'origine de mêmes niveaux de champ sont plus difficiles à localiser et à dénombrer. Dans une maison proche d'une ligne, c'est cette ligne qui est la source principale du CM. Parmi les enfants résidant à moins de 30 mètres des lignes, la proportion d'enfants exposés à des champs d'au moins 0,4 µT est beaucoup plus élevée que dans le reste de la population.

L'étude du lien entre expositions aux CM-EBF et incidence des leucémies de l'enfant a fait l'objet d'un nombre important de travaux depuis l'étude de Weirthemer en 1979[Wertheimer &Leeper,1979].

Différents indicateurs d'exposition ont été utilisés :. exposition par des mesures sur 24h dans différentes pièces de la maison, mesures de plus courtes durées, ou **mesures** de l'exposition personnelle du sujet par un port de dosimètre pendant un temps défini [Tomenius,1986] ; [Savitz *et al.*,1988] ; [London *et al.*,1991] [Feychting &Ahlbom,1993] ; [Coghill *et al.*,1996] ; [Linet *et al.*,1997] ; [Michaelis *et al.*,1997] ; [Michaelis *et al.*,1998] ; [Dockerty *et al.*,1999]; [Green *et al.*,1999] ; [Green *et al.*,1999] ; [McBride *et al.*,1999] ; [UKCCS,1999] ; [Schüz *et al.*,2001] ; [Kabuto *et al.*,2006]. D'autres études ont porté sur l'exposition aux CM-EBF spécifiquement induite par la proximité aux LHT et utilisé des **méthodes de calculs avancés**

prenant en compte les différentes caractéristiques de la ligne pour calculer l'exposition des sujets proches de ces lignes ([Myers *et al.*,1990] [Feychting &Ahlbom,1993] [Olsen *et al.*,1993] [Tynes &Haldorsen,1997] [Bianchi *et al.*,2000] [UKCCS,2000] [Feizi &Arabi,2007] [Malagoli *et al.*,2010]). Certaines études ont utilisé des indicateurs plus simples sans estimer précisément les champs induits comme la **distance aux lignes**, parfois en prenant en compte la tension de la ligne, [Fulton *et al.*,1980] [Myers *et al.*,1990] [Petridou *et al.*,1993] [Tynes &Haldorsen,1997] [Kleinerman *et al.*,2000] [UKCCS,2000] [Draper *et al.*,2005] [Abdul Rahman *et al.*,2008], **l'incidence** au voisinage des lignes [Verkasalo *et al.*,1993] [Lin &Lee,1994] [Li *et al.*,1998], [Mizoue *et al.*,2004] ou encore la **visibilité d'une ligne** autour de l'habitation [Tomenius,1986] [Fajardo-Gutiérrez *et al.*,1993]. Enfin, d'autres études ont utilisé les « **wire codes** », décrits par Wertheimer [Wertheimer &Leeper,1979], comme indicateur d'exposition [Wertheimer &Leeper,1979], [Savitz *et al.*,1988] [London *et al.*,1991] [Savitz &Kaune,1993] [Petridou *et al.*,1997] [Linet *et al.*,1997] [Green *et al.*,1999] [Green *et al.*,1999] [McBride *et al.*,1999], qui définissent 5 classes d'exposition en prenant en compte le type de ligne et de la distance de l'habitation à la ligne. Certaines études utilisent simultanément plusieurs indicateurs.

Globalement, les associations trouvées dans ces études avec l'incidence des leucémies de l'enfant sont moins fortes quand l'exposition a été estimée par mesure directe que par calcul des champs induits par les lignes à haute tension.

L'histoire résidentielle des sujets était connue dans la plupart de ces études, elle a été prise en compte dans le calcul de l'exposition mais parfois seule l'adresse à la naissance ou l'adresse au diagnostic était disponible.

Les principales méta-analyses d'études dans lesquelles l'exposition aux CM est calculée ou mesurée [Ahlbom *et al.*,2000] ; [Greenland *et al.*,2000] et mise en

relation avec l'incidence de LA ont amené le Centre International de Recherche sur le Cancer (CIRC) à classer l'exposition aux champs magnétiques à extrêmement basse fréquence comme "Cancérogène possible chez l'Homme" (2B) [IARC ,2002]. Une exposition de plus de 0,4 µT doublerait le risque de leucémie de l'enfant et concernerait environ 1% de la population pédiatrique.

Depuis cette publication, plusieurs études ont apporté des estimations du même ordre même si dans certaines études les augmentations de risque mises en évidence ne sont pas significatives [Bianchi *et al.*,2000] ; [Mizoue *et al.*,2004] ; [Abdul Rahman *et al.*,2008] ; [Malagoli *et al.*,2010] ou inexistante [Kleinerman *et al.*,2000]. Draper et al ont rapporté une augmentation de risque de 1,7 chez les enfants résidant à moins de 200 m de LHT, ce qui correspond à des champs plus faibles en moyenne que ceux qui étaient pointés par les analyses poolées [Draper *et al.*,2005]. Les études postérieures aux deux méta-analyses de 2000 ont fait l'objet d'une méta-analyse en 2010 [Kheifets *et al.*,2010]. Elle rapporte comme les méta-analyses précédentes une augmentation de risque de LA chez les sujets les plus exposés aux CM mais l'association trouvée n'était pas statistiquement significative. L'étude anglaise analysant la distance aux lignes à haute tension [Draper *et al.*,2005] a depuis été complétée par une étude portant sur un sous-ensemble de sujets de la première étude pour lesquels les CM-EBF ont pu être mesurés [Kroll *et al.*,2010]. Cette étude n'a pas trouvé d'augmentation d'incidence de LA chez les sujets ayant une exposition de plus de 0,2 µT.

B Matériel et méthodes

L'étude que nous avons réalisé porte sur l'incidence des LA au voisinage des LHT en France métropolitaine sur la période 2002-2007.

B.1 Sujets :

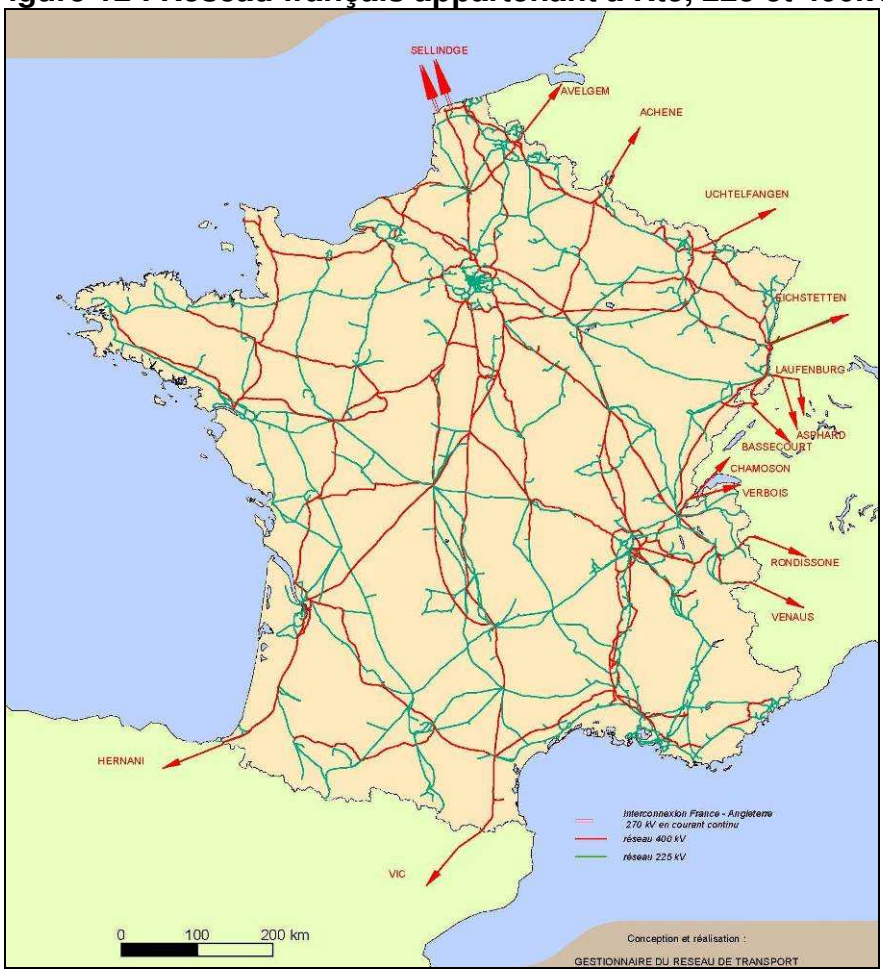
Les sujets inclus dans cette étude sont tout d'abord les 30 000 témoins de l'échantillon INSEE (voir partie III.C). Concernant les cas, ce sont les 2 774 cas de LA de la période 2002-2007 enregistrés dans le RNHE en novembre 2011.

B.2 Exposition aux LHT

B.2.a Données utilisées

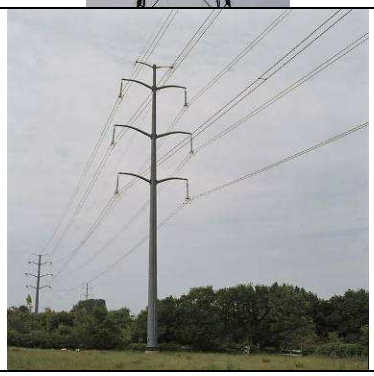
Le Réseau de transport d'électricité (Rte), avec qui nous collaborons étroitement depuis plusieurs années, possède un service de système d'information géographique qui indique l'emplacement de tous les supports (ou pylônes) de LHT, et tous les tronçons de lignes (Figure 12). Ces données sont continuellement mises à jour et leur précision continuellement améliorée. Les lignes considérées sont celles existantes dans la base de données de septembre 2011. Plus de 90% des supports ont une géolocalisation précise à 2,5 m près. Les autres ont une incertitude moyenne de géolocalisation qui est de l'ordre de 25 m suivant la tension de la ligne donc du type de pylône.

Figure 12 : Réseau français appartenant à Rte, 225 et 400kV



En France, les tensions des lignes à haute et très haute tension (Tableau 25) sont 400kV, 225kV, 150kV (tension qui tend à disparaître), 90kV et 63kV.

Tableau 25 : Réseau de lignes aériennes à haute et très haute tension de Rte en France

Réseau	Tension	Km de lignes (en 2005)	
Très haute tension	400 kV	13 350	
	225 kV	21 200	
Haute tension	150 kV	1 050	
	90 kV	12 650	
	63 kV	29 150	
TOTAL		77 400	

Il est couramment reconnu que l'exposition à un CM de moins de 0,1 µT peut être considérée comme une exposition "de fond", le CM induit par la ligne n'étant alors pas discernable de celui généré par toute autre source d'exposition environnante.

Nous ne considérons donc ici que les lignes aériennes car les lignes souterraines génèrent des expositions aux CM-EBF sensiblement plus faibles et dont la portée à la surface du sol (0,1 µT à environ 15 m) est à la fois faible et de l'ordre de grandeur de la précision du géocodage des habitations et de leur taille.

Nous avons choisi de ne pas prendre en compte la date de mise en service des lignes ou de construction des pylônes, ce qui aurait exclu les tronçons qui ont été

construits après l'année considérée pour le sujet. En effet, ces dates ne sont pas renseignées pour toutes les lignes et ne reflètent pas toujours une date de construction de ligne mais bien plus souvent une modification de celle-ci (renouvellement de pylônes, ajout d'un poste intermédiaire). Par ailleurs, depuis 2 000, le kilométrage total des lignes en France fait l'objet d'un accord entre Rte et l'Etat ce qui implique que très peu de nouvelles lignes sont construites. Pour ces différentes raisons, nous avons choisi de garder tous les croisements de lignes x sujets dans l'étude. A posteriori, nous avons constaté que les sujets qui étaient "proches" de pylônes dont les dates d'implantations étaient postérieures à leurs années d'étude l'étaient à des distances relativement importantes. L'impact de ce choix dans notre étude qui se concentre sur la proximité directe aux LHT ne doit donc pas être important.

Nous avons exclu la Corse de l'étude puisque les LHT de cette île ne sont pas gérées par Rte : il aurait été très long et difficile d'obtenir et de géocoder le tracé de ces lignes, alors qu'elles sont très peu nombreuses et ne concernent qu'un très petit nombre d'enfants.

Géocible a calculé pour tous les sujets la distance de l'adresse à toutes les LHT à proximité du domicile : la zone prise en compte est de 600 m augmenté de l'imprécision du géocodage « automatique » à 99%, ceci afin de connaître toutes les lignes potentiellement à moins de 600m du sujet (partie III.D). En ce qui concerne les sujets pour lesquels la seule localisation connue est la commune, une autre approche a été utilisée : si la totalité de la commune est située à plus de 600 m de toute LHT du réseau Rte, les sujets sont considérés comme n'ayant pas de LHT à

moins de 600 m. En revanche, on ne connaît pas la distance aux lignes des sujets géocodés à la commune et pour lesquels celle-ci est située à moins de 600 m de la ligne Rte la plus proche.

B.2.b Distance à la ligne la plus proche

Le premier indicateur d'exposition est la distance à la LHT la plus proche. Celle-ci est calculée premièrement à partir du géocodage « automatique » des adresses, distance qui sera appelée « distance automatique ». Elle est aussi calculée à partir du géocodage « au domicile », effectué pour les sujets possiblement domicilié à moins de 200 m des lignes en géocodage « automatique » (voir partie III. D), distance qui sera appelée « distance au domicile ».

La classe de référence choisie pour cette variable est constituée des sujets dont la commune est entièrement située à plus de 600 m de la LHT la plus proche: (Référence : "Communes ≥ 600 m des lignes") puisqu'on peut considérer qu'à plus de 600 m d'une ligne l'exposition aux CM liée à cette ligne est certainement inférieure à 0,1 µT et donc que ces sujets sont non exposés à la proximité de LHT. Nos analyses ont également été réalisées en prenant comme référence les sujets « très certainement à plus de 600 m de toute ligne », c'est-à-dire les sujets dont la «distance automatique » diminuée de l'imprécision du géocodage du sujet à 99% était supérieure à 600m (Référence : "Domiciles ≥ 600 m des lignes").

Quatre classes de distance ont été choisies *a priori* pour cerner des expositions décroissantes aux CM: 0-49 m ; 50-99 m ; 100-199 m ; 200-600 m.

Un indicateur continu usuel de proximité aux LHT a également été considéré avec l'inverse de la distance.

B.2.c Analyse statistique

La distance aux lignes a d'abord été étudiée pour l'ensemble des tensions. Les CM induits par une ligne dépendant globalement de sa tension (l'intensité du courant étant d'autant plus forte que les lignes sont de tension élevée) nous avons également étudié séparément la distance aux lignes 225 kV et 400 kV d'une part et la distance aux lignes 63 ; 90 et 150 kV d'autre part.

Afin de tester l'existence d'une tendance avec la distance, l'inverse de la distance (km^{-1}) a été incluse dans le modèle : jusqu'à 600 m pour la distance « automatique » et jusqu'à 200 m pour la distance « au domicile ».

Les analyses ont également été faites en ne gardant que les sujets précis à 20 m puisque les autres ont une géolocalisation moins précise. Par ailleurs, nous avons fait des analyses stratifiées sur l'âge des sujets : ceux âgés de moins de 5 ans ont été étudiés indépendamment des sujets ayant entre 5 et 14 ans, les premiers ayant une plus faible probabilité d'avoir changé de domicile depuis leur naissance.

Les LAL ont également été analysées séparément.

Enfin, nous avons analysé les résultats par taille d'unité urbaine en trois classes : zones rurales, zones semi-urbaines et zones urbaines.

B.2.d Puissance de l'étude

La puissance de l'étude (Tableau 26) dépend fortement de la catégorie de distance concernée : l'OR minimum détectable pour une puissance de 80% en test unilatéral quand on étudie l'exposition « habiter à moins de 50 m d'une LHT de plus de 225 kV en géocodage automatique » est de 2,47.

Tableau 26 : Puissance de l'étude cas-témoins pour détecter différentes estimations (θ) du risque relatif

	Exposition		témoins exposés		OR minimum détectable unilatéral
Géocodage "automatique" Référence [a]: 7 061 témoins 609 cas	distance ≥225 kV	0 – 50m	60	0,8%	2,47
		0 – 100m	157	2,2%	1,85
		0 – 200m	424	6,0%	1,50
	distance ≤150 kV	0 – 50m	164	2,3%	1,83
		0 – 100m	367	5,2%	1,54
		0 – 200m	828	11,7%	1,36
Géocodage "au domicile" Référence [a]: 7 061 témoins 609 cas	distance ≥225 kV	0 – 50m	38	0,5%	2,96
		0 – 100m	112	1,6%	2,02
		0 – 200m	284	4,0%	1,81
	distance ≤150 kV	0 – 50m	127	1,8%	1,96
		0 – 100m	248	3,5%	1,66
		0 – 200m	570	8,1%	1,43

[a] Référence : Communes entièrement situées à plus de 600 m de le ligne Rte la plus proche

C Résultats

Il y avait 7 061 témoins et 609 cas appartenant à des communes ne recoupant pas le réseau à 600 m : ces sujets constituent donc la référence (Référence : Communes ≥ 600 m des lignes"). Nous avons trouvé, comme attend puisque la taille d'une commune est bien supérieure à celle de la portée de l'exposition à une LHT, que le fait d'habiter dans une commune qui recoupe le réseau Rte à moins de 600 m n'était pas associé au statu cas-témoins: OR= 1,07 [0,96-1,19].

C.1.a Distance « automatique »

Parmi les 30 000 témoins et 2 774 cas, 203 témoins et 67 cas ont été exclus parce qu'étant géocodés à la mairie de la commune et que ces communes recoupaient le réseau à moins de 600 m, on ne pouvait pas déterminer s'ils habitaient près d'une ligne ou non. Moins de 0,4% de la totalité sujets habitaient à moins de 200 m d'une ligne et seulement 0,1% à moins de 50 m d'une ligne.

Quand on analyse la distance « automatique » à la LHT la plus proche, quelle que soit la tension de la ligne chez tous les sujets, aucune association nette n'est trouvée entre la proximité aux LHT et l'incidence de LA (Tableau 27), ceci en analysant la distance en classes ou continu dans le modèle.

Tableau 27: Association entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile »

	Distance « automatique »													Distance "au domicile"					
	Tous sujets						Sujets géocodés "au domicile"							Tous sujets					
	Cas	**Témoins**	**OR**[a]	IC 95%		***P***[b]	**Cas**	**Témoins**	**OR**[a]	IC 95%		***p***[b]		**Cas**	**Témoins**	**OR**[a]	IC 95%		***p***[b]
0-49m	23	213	**1,2**	0,8	2,0	*16%*	17	153	**1,2**	0,7	2,1	*21%*	**0-49m**	14	158	**1,1**	0,6	1,9	*40%*
50-99m	25	284	**1,0**	0,7	1,6	*45%*	20	202	**1,2**	0,7	1,9	*29%*	**50-99m**	22	184	**1,3**	0,8	2,1	*13%*
100-199m	44	669	**0,8**	0,6	1,1	*95%*	29	449	**0,7**	0,5	1,1	*93%*	**100-199m**	36	455	**0,9**	0,6	1,3	*68%*
0-199m	92	1166	**0,9**	0,7	1,2	*78%*	66	804	**0,9**	0,7	1,2	*67%*	**0-199m**	72	797	**1,0**	0,6	1,4	*39%*
200-599m	343	3633	**1,1**	0,9	1,2	*21%*	147	1123	**1,5**	1,2	1,8	*0%*	**200-599m**	137	1133	**1,4**	1,1	1,7	*0%*
≥ 600m et non référence "Commune"	1663	17937	**1,1**	1,0	1,2	*18%*	40	161	**3,0**	2,1	4,4	*0%*	**≥ 600m et non référence "Commune"**	59	179	**3,7**	2,7	5,1	*0%*
Géocodage "Commune" et non référence	67	203	**3,5**	2,6	4,7	*0%*	15	21	**6,9**	3,5	13,8	*0%*	**Pas de géocodage "au domicile" et non référence**	1897	20830	**1,0**	0,9	1,1	*27%*
Référence "Commune"	**609**	**7061**	**1,0**				**609**	**7061**	**1,0**				**Référence "Commune"**	**609**	**7061**	**1,0**			
TOTAL	**2774**	**30000**					**877**	**9170**					**TOTAL**	**2774**	**30000**				

[a] Odds ratios (OR) et intervalles de confiance à 95% (IC 95%) estimés par régression logistique ajustée sur l'âge en fin d'année (classes d'âge de 5 ans) et sur le département de résidence ; [b] p : $p_{unilatéral}$

Une association à la limite de la significativité statistique est trouvée à moins de 50 m en « distance automatique » des seules lignes 225 kV et 400 kV : OR=1,7 [0,9-3,6] (Tableau 28). En revanche, aucune association n'est trouvée pour la proximité en distance « automatique » des lignes de moins de 150 kV (Tableau 28). Quand on ne considère que les sujets qui ont été géocodés « au domicile », l'association trouvée à moins de 50 m en distance « automatique » est renforcée : OR=2,2 [0,96-5,0] p=0,03.

Tableau 28: Association entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne

	Distance "automatique"												Distance "au domicile"					
	Tous sujets						Sujets géocodés "au domicile"						Tous sujets					
	Cas	Témoins	OR [a]	IC 95%		p [b]	Cas	Témoins	OR [a]	IC 95%		p [b]		Cas	Témoins	OR [a]	IC 95%		p [b]
≥225 kV																			
0-49 m	9	60	**1,8**	0,9	3,6	*6%*	7	38	**2,2**	1,0	5,0	*3%*	**0-49m**	4	38	**1,3**	0,5	3,7	*32%*
50-99 m	8	97	**1,0**	0,5	2,1	*50%*	7	68	**1,2**	0,5	2,6	*35%*	**50-99m**	8	74	**1,3**	0,6	2,7	*27%*
100-199 m	16	267	**0,7**	0,4	1,2	*92%*	12	181	**0,8**	0,4	1,5	*77%*	**100-199m**	18	172	**1,2**	0,8	2,1	*20%*
0-199 m	33	424	**0,9**	0,6	1,3	*69%*	26	287	**1,1**	0,7	1,6	*37%*	**0-199m**	30	284	**1,3**	0,8	1,9	*13%*
200-599 m	144	1416	**1,2**	1,0	1,4	*8%*	61	516	**1,4**	1,0	1,8	*2%*	**200-599m**	54	520	**1,2**	0,9	1,6	*12%*
≥ 600m et non référence "Commune"	1921	20896	**1,0**	0,9	1,2	*22%*	166	1285	**1,5**	1,2	1,8	*0%*	**≥ 600m et non référence "Commune"**	184	1305	**1,6**	1,3	1,9	*0%*
Géocodage "Commune" et non référence	67	203	**3,5**	2,6	4,7	*0%*	15	21	**6,7**	3,4	13,4	*0%*	**Pas de géocodage "au domicile" et non référence**	1897	20830	**1,0**	0,9	1,1	*28%*
Référence "Commune"	**609**	**7061**	**1,0**				**609**	**7061**	**1,0**				**Référence "Commune"**	**609**	**7061**	**1,0**			
TOTAL	**2774**	**30000**					**877**	**9170**					**TOTAL**	**2774**	**30000**				
≤150 kV																			
0-49 m	14	164	**0,9**	0,5	1,6	*62%*	10	121	**0,9**	0,5	1,8	*62%*	**0-49m**	10	127	**0,9**	0,5	1,8	*59%*
50-99 m	21	203	**1,1**	0,7	1,8	*30%*	16	143	**1,3**	0,7	2,2	*20%*	**50-99m**	15	121	**1,3**	0,7	2,3	*19%*
100-199 m	33	461	**0,8**	0,5	1,2	*89%*	20	309	**0,7**	0,5	1,2	*90%*	**100-199m**	23	322	**0,8**	0,5	1,2	*85%*
0-199 m	68	828	**0,9**	0,7	1,2	*66%*	46	573	**0,9**	0,6	1,2	*76%*	**0-199m**	48	570	**0,9**	0,7	1,3	*67%*
200-599 m	241	2740	**1,0**	0,8	1,2	*57%*	107	863	**1,4**	1,1	1,8	*0%*	**200-599m**	103	864	**1,3**	1,1	1,7	*1%*
≥ 600m et non référence "Commune"	1789	19168	**1,1**	1,0	1,2	*14%*	100	652	**1,8**	1,4	2,3	*0%*	**≥ 600m et non référence "Commune"**	117	675	**2,0**	1,6	2,5	*0%*
Géocodage "Commune" et non référence	67	203	**3,5**	2,59	4,7	*0%*	15	21	**6,7**	3,4	13,4	*0%*	**Pas de géocodage "au domicile" et non référence**	1897	20830	**1,0**	0,9	1,1	*30%*
Référence "Commune"	**609**	**7061**	**1,0**				**609**	**7061**	**1,0**				**Référence "commune"**	**609**	**7061**	**1,0**			
TOTAL	**2774**	**30000**					**877**	**9170**					**TOTAL**	**2774**	**30000**				

[a] Odds ratios (OR) et intervalles de confiance à 95% (IC 95%) estimés par régression logistique ajustée sur l'âge en fin d'année (classes d'âge de 5 ans) et sur le département de résidence ; [b] p : $p_{unilatéral}$

C.1.b Distance « au domicile »

Parmi les sujets à moins de 200 m d'une ligne en distance « automatique », 72% des cas et 69% des témoins ont pu être géocodés « au domicile » et ont donc une distance « au domicile » (Tableau 29). Cet indicateur d'exposition n'est donc pas disponible pour environ 30% de tous les sujets. Cependant cette sélection ne semble pas particulière puisqu'elle ne dépend pas de l'âge des sujets, du type d'unité urbaine ni de la distance « automatique » aux lignes. Comme attendu, les sujets géocodés « au domicile » étaient plus précis en géocodage « automatique » puisque la précision du géocodage dépendait de la précision de l'adresse fournie en entrée.

Tableau 29 : Distribution des cas et des témoins à moins de 200m d'une LHT en distance « automatique », et qui ont été géocodés « au domicile » selon différentes variables individuelles et contextuelles

		Cas					Témoins				
		Total		Géocodés "au domicile"		% "au domicile"	Total		Géocodés "au domicile"		% "au domicile"
Age	**0-4 ans**	45	49%	35	52%	78%	359	31%	241	30%	67%
	5-9 ans	30	33%	21	31%	70%	389	33%	266	33%	68%
	10-14 ans	17	18%	10	16%	59%	418	36%	297	37%	71%
Unité urbaine de la commune	**communes rurales < 5 000 habs**	19	21%	13	20%	68%	286	25%	146	18%	51%
	5 000 - 100 000 habs	32	35%	20	30%	63%	306	26%	203	25%	66%
	> 100 000 habs	41	45%	33	50%	80%	574	49%	455	57%	79%
Distance "automatique"	**0-50 m**	23	25%	17	26%	74%	213	18%	153	19%	72%
	50-100 m	25	27%	20	30%	80%	284	24%	202	25%	71%
	100-200 m	44	48%	29	44%	66%	669	57%	449	56%	67%
Précision du géocodage "automatique"	**20 m**	68	74%	57	86%	84%	863	74%	736	92%	85%
	50 m	4	4%	2	3%	50%	75	6%	13	2%	17%
	100 m	10	11%	5	8%	50%	145	12%	37	5%	26%
	250 m	10	11%	2	3%	20%	83	7%	18	2%	22%
TOTAL	**0-200 m**	**92**	**100%**	**66**	**100%**	**72%**	**1166**	**100%**	**804**	**100%**	**69%**

Le passage d'une distance « automatique » à celle « au domicile » opère bien un reclassement des sujets. Globalement, seulement 20% des sujets à moins de 200 m des lignes changent de classe de distance (en considérant les classes utilisées pour la distance « automatique » plus haut). Cependant, quand on considère la distance à la ligne ≥225 kV et notamment les cas proches des lignes, 3 parmi 7 (plus de 40%) sont reclassés (Tableau 30).

Tableau 30 : Croisement des distances « automatique » et « au domicile » chez les cas à moins de 200m d'une ligne ≥225kV en distance « automatique » et ayant un géocodage « au domicile »

		Distance « au domicile »			Total	Non géocodés « au domicile »
CAS		**< 50 m**	**50-199 m**	**≥ 200 m**		
Distance « automatique »	**< 50 m**	4	3	0	7	2
	50-199 m	0	19	0	19	5
	Total	4	22	0	26	7
TEMOINS		**< 50 m**	**50-199 m**	**≥ 200 m**		
Distance « automatique »	**< 50 m**	30	7	1	38	22
	50-199 m	6	224	19	248	115
	Total	36	231	20	286	137

Quand on analyse la distance « au domicile », aucune association significative n'est trouvée, notamment à moins de 50 m pour les lignes 225 kV et 400 kV, même si l'OR est légèrement supérieur à 1 : OR=1,3 [0,5-3,7] (Tableau 28). Par ailleurs, nous avons vérifié manuellement les 2 cas et les 22 témoins à moins de 50m en distance « automatique » qui n'ont pas pu être géocodés « au domicile »: pour 1 cas et 5 témoins, même si la localisation précise du bâtiment n'a pas pu être définie, nous

avons pu vérifier qu'ils étaient bien, en distance « domicile », à moins de 50 m de ce type de lignes. L'inclusion de ces 6 sujets dans l'étude en distance « au domicile » renforce légèrement l'association à moins de 50 m d'une ligne de plus de 225 kV : OR=1,4[0,5-3,5].

Chez les moins de 5 ans (Tableau 31), l'association précédemment trouvée chez les 0-14 ans à moins de 50 m est significative : OR=2,6 [1,0-7,0] et on observe une tendance à la limite de la significativité (p unilatéral =0,1). Même si l'association n'est pas significative, l'OR à moins de 50m en distance « au domicile » est supérieur à 1 : OR=2,5 [0,6-10] et une tendance à la limite de la significativité est observée :
p unilatéral = 0,1.
Pour les LHT de moins de 150 kV, à nouveau aucune association n'est observée.
Chez les sujets de 5 à 14 ans, aucune association n'est trouvée entre la proximité des lignes et l'incidence de LA, que ce soit en distance « automatique » ou en distance « au domicile » et que ce soit la proximité aux lignes de plus de 225 kV et celles de moins de 150 kV. L'OR correspondant à moins de 50 m d'une ligne ≥ 225 kV en distance « automatique » est OR=1,0 [0,3-3,3].
Le test d'interaction entre le fait d'habiter à moins de 50 m d'une ligne de plus de 225 kV et celui d'être âgé de moins de 5 ans est significatif à 11% pour la distance « automatique » et à 12% pour la distance « au domicile ». L'association trouvée semble donc bien limitée aux 0-4 ans.

C.1.c Analyses de sensibilité

Quand on restreint les analyses aux sujets qui ont un géocodage précis (Tableau 32), l'association trouvée à moins de 50m pour les lignes ≥225 kV est renforcée et notamment l'OR correspondant à la distance « automatique » est significativement supérieur à 1 : OR=2,1 [0,9-4,7] p unilatéral=0,04.

L'étude de la distance aux lignes par taille d'unité urbaine semble montrer des associations plus fortes (toujours pour la distance aux lignes ≥225 kV à moins de 50 m) dans les zones « rurales » et « semi-urbaines » même si c'est dans le milieu urbain (taille d'unité urbaine de plus de 100 000 habitants) qu'on a le plus de témoins exposés donc plus de puissance *a priori* pour détecter une augmentation de risque (Tableau 33). L'interaction entre le fait d'habiter à moins de 50 m d'une ligne ≥225 kV et celui d'habiter dans une zone « rurale » ou « semi-urbaine » est significative à 4% pour la distance « automatique » et à 17% pour la distance « au domicile ».

L'association entre LA et le fait d'habiter à moins de 50 m des lignes ≥225 kV est plus forte pour les LAL (Tableau 34), en distance « automatique » : OR=1,9 [0,9-4] p unilatéral = 0,05, mais on n'observe pas de tendance significative avec la proximité en continu (p unilatéral=0,17). En distance « au domicile », on n'observe pas de résultats significatifs même si l'OR est supérieur à 1 : OR=1,6 [0,6-1,6] et la tendance avec la proximité en continu n'est pas significative (p unilatéral =0,2).

Nous avons cherché à prendre en compte la proximité aux CNPE mais aucun sujet n'est simultanément à moins de 50 m d'une ligne 225 ou 400 kV et à moins de 5 km d'une CNPE, zone dans laquelle nous avons détecté un doublement de risque dans

la même population (partie V). Aucune association entre proximité des lignes et proximité des CNPE qui expliquerait ces OR supérieurs à 1 n'est donc possible ici.

Des analyses ont également été faites en faisant varier la limite de la classe des sujets les plus exposés (< 50 m) de - 20 m ; - 10 m ; + 10 m ; + 20 m et les résultats étaient très semblables de ceux trouvés avec une bande de 50 m. Par ailleurs, en prenant comme référence les sujets « très certainement à plus de 600 m de toute ligne », les résultats étaient très proches de ceux présentés ici.

Tableau 31 : Association, par classe d'âge, entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne

≥225 kV	Distance "automatique"													Distance "au domicile"					
	Tous sujets						Sujets géocodés "au domicile"							Tous sujets					
0-4 ans	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	6	14	**2,6**	1,0	7,0	*3%*	6	7	**4,9**	1,5	15,2	*0%*	**0-49m**	3	7	**2,6**	0,6	10,3	*9%*
50-99m	6	27	**1,6**	0,7	4,1	*14%*	5	18	**1,9**	0,6	5,6	*12%*	**50-99m**	6	21	**2,1**	0,8	5,4	*6%*
100-199m	5	87	**0,4**	0,2	1,0	*97%*	5	60	**0,7**	0,3	1,8	*78%*	**100-199m**	9	58	**1,2**	0,6	2,4	*33%*
0-199m	17	128	**0,9**	0,6	1,6	*59%*	16	85	**1,4**	0,8	2,5	*12%*	**0-199m**	18	86	**1,7**	1,0	2,9	*4%*
200-599m	73	444	**1,2**	0,9	1,6	*13%*	32	153	**1,7**	1,1	2,6	*1%*	**200-599m**	28	160	**1,3**	0,8	2,0	*13%*
≥ 600m et non réf."Commune"	871	6734	**0,9**	0,8	1,1	*87%*	68	416	**1,2**	0,9	1,7	*9%*	**≥ 600m et non réf. "Commune"**	76	422	**1,3**	1,0	1,7	*4%*
Géocodage ""Commune" et non référence	35	85	**2,6**	1,7	4,1	*0%*	6	14	**3,6**	1,3	10,1	*1%*	**Pas de géocodage "au domicile" et non référence**	874	6723	**0,9**	0,8	1,1	*87%*
Référence "Commune"	**311**	**2326**	**1,0**				**311**	**2326**	**1,0**				**Référence "Commune"**	**311**	**2326**	**1,0**			
Total	**1307**	**9717**					**433**	**2994**					**TOTAL**	**1307**	**9717**				
5-14 ans	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	3	46	**1,0**	0,3	3,3	*49%*	1	31	**0,4**	0,1	3,3	*79%*	**0-49m**	1	31	**0,5**	0,1	3,7	*75%*
50-99m	2	70	**0,5**	0,1	2,0	*85%*	2	50	**0,6**	0,1	2,4	*77%*	**50-99m**	2	53	**0,6**	0,1	2,6	*75%*
100-199m	11	180	**1,0**	0,5	1,8	*56%*	7	121	**0,9**	0,4	2,0	*58%*	**100-199m**	9	114	**1,2**	0,6	2,5	*29%*
0-199m	16	296	**0,8**	0,5	1,4	*73%*	10	202	**0,7**	0,4	1,4	*81%*	**0-199m**	12	198	**0,9**	0,5	1,7	*62%*
200-599m	71	972	**1,2**	0,9	1,5	*15%*	29	363	**1,2**	0,8	1,8	*18%*	**200-599m**	26	360	**1,1**	0,8	1,8	*26%*
≥ 600m et non réf."Commune"	1050	14162	**1,2**	1,0	1,4	*1%*	98	869	**1,7**	1,3	2,2	*0%*	**≥ 600m et non réf. "Commune"**	108	883	**2,0**	1,5	2,5	*0%*
Géocodage "Commune" et non référence	32	118	**4,3**	2,9	6,6	*0%*	9	7	**17,2**	6,2	47,8	*0%*	**Pas de géocodage "au domicile" et non référence**	1023	14107	**1,2**	1,0	1,3	*3%*
Référence "Commune"	**298**	**4735**	**1,0**				**298**	**4735**	**1,0**				**Référence "Commune"**	**298**	**4735**	**1,0**			
Total	**1467**	**20283**					**444**	**6176**					**TOTAL**	**1467**	**20283**				
TOTAL 0-14ans	**2774**	**30000**					**877**	**9170**						**2774**	**30000**				

≤150kV	Distance "automatique"											Distance "au domicile"					
	Tous sujets					Sujets géocodés "au domicile"						Tous sujets					
0-4ans	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	8	52	**1,1**	0,5	2,3	*44%*	5	37	**0,8**	0,3	2,0	*71%*	**0-49m**	5	33	**1,1**	0,4	2,9	*41%*
50-99m	11	61	**1,3**	0,7	2,5	*24%*	8	38	**1,6**	0,7	3,6	*13%*	**50-99m**	8	35	**1,7**	0,8	3,8	*10%*
100-199m	13	145	**0,6**	0,4	1,2	*93%*	9	95	**0,8**	0,4	1,5	*78%*	**100-199m**	9	99	**0,6**	0,3	1,3	*89%*
0-199m	32	258	**0,9**	0,6	1,3	*73%*	22	170	**0,9**	0,6	1,5	*61%*	**0-199m**	22	167	**0,9**	0,6	1,5	*58%*
200-599m	115	902	**0,9**	0,7	1,1	*82%*	46	270	**1,3**	0,9	1,9	*7%*	**200-599m**	46	273	**1,3**	0,9	1,8	*9%*
≥ 600m et non réf."Commune"	814	6146	**0,9**	0,8	1,1	*79%*	48	214	**1,8**	1,2	2,6	*0%*	**≥ 600m et non réf. "Commune"**	54	228	**1,7**	1,2	2,3	*0%*
Géocodage "Commune" et non référence	35	85	**2,6**	1,7	4,1	*0%*	6	14	**3,6**	1,3	10,2	*1%*	**Pas de géocodage "au domicile" et non référence**	874	6723	**0,9**	0,8	1,1	*88%*
Référence "Commune"	**311**	**2326**	**1,0**				**311**	**2326**	**1,0**				**Référence "Commune"**	**311**	**2326**	**1,0**			
Total	**1307**	**9717**					**433**	**2994**					**TOTAL**	**1307**	**9717**				
5-14ans	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	6	112	**0,9**	0,4	2,1	*59%*	5	84	**0,9**	0,4	2,2	*60%*	**0-49m**	5	94	**0,9**	0,4	2,2	*60%*
50-99m	10	142	**1,1**	0,6	2,2	*37%*	8	105	**1,1**	0,5	2,2	*45%*	**50-99m**	7	86	**1,3**	0,6	2,8	*29%*
100-199m	20	316	**1,0**	0,6	1,7	*44%*	11	214	**0,8**	0,4	1,5	*78%*	**100-199m**	14	223	**0,9**	0,5	1,6	*62%*
0-199m	36	570	**1,0**	0,7	1,5	*43%*	24	403	**0,9**	0,6	1,4	*71%*	**0-199m**	26	403	**1,0**	0,6	1,8	*46%*
200-599m	126	1838	**1,1**	0,9	1,4	*25%*	61	593	**1,5**	1,1	2,1	*0%*	**200-599m**	57	591	**1,5**	1,1	2,0	*0%*
≥ 600m et non réf. "Commune"	975	13022	**1,2**	1,0	1,4	*1%*	52	438	**1,9**	1,4	2,6	*0%*	**≥ 600m et non réf. "Commune"**	63	447	**2,2**	1,7	3,0	*0%*
Géocodage "Commune" et non référence	32	118	**4,3**	2,8	6,6	*0%*	9	7	**17,0**	6,1	47,2	*0%*	**Pas de géocodage "au domicile" et non référence**	1023	14107	**1,2**	1,0	1,3	*3%*
Référence "Commune"	**298**	**4735**	**1,0**				**298**	**4735**	**1,0**				**Référence "Commune"**	**298**	**4735**	**1,0**			
Total	**1467**	**20283**					**444**	**6176**					**TOTAL**	**1467**	**20283**				
TOTAL 0-14ans	**2774**	**30000**					**877**	**9170**						**2774**	**30000**				

Tableau 32 : Association, chez les sujets précis à 20 m, entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne

sujets "précis" à 20m *	Distance "automatique"										Distance "au domicile"					
	Tous sujets					Sujets géocodés "au domicile"						Tous sujets				
	Cas	Témoins	OR	IC 95%		p_{uni}	Cas	Témoins	OR	IC 95%	p_{uni}		Cas	Témoins	OR	IC 95%	p_{uni}
≥225 kV																	
0-49m	7	39	**2,1**	0,9	4,7	*4%*	6	34	**2,0**	0,8	4,9	*6%*	**0-49m**	4	34	**1,4**	0,5	4,0	*28%*
50-99m	6	68	**1,0**	0,4	2,4	*46%*	6	62	**1,2**	0,5	2,7	*37%*	**50-99m**	8	70	**1,3**	0,6	2,8	*22%*
100-199m	13	200	**0,8**	0,4	1,4	*82%*	10	171	**0,7**	0,4	1,4	*85%*	**100-199m**	13	160	**0,9**	0,5	1,7	*53%*
0-199m	26	307	**1,0**	0,7	1,5	*51%*	22	267	**1,0**	0,6	1,6	*53%*	**0-199m**	25	264	**1,1**	0,7	1,7	*29%*
200-599m	104	1122	**1,1**	0,8	1,3	*32%*	52	456	**1,3**	1,0	1,8	*4%*	**200-599m**	48	460	**1,2**	0,9	1,7	*12%*
≥ 600m et non réf. "Commune"	1352	16279	**0,9**	0,8	1,1	*85%*	106	1045	**1,2**	0,9	1,5	*10%*	**≥ 600m et non réf. "Commune"**	107	1044	**1,2**	0,9	1,5	*8%*
Géocodage "Commune" et non référence	0	0					0	0					**Pas de géocodage "au domicile" et non référence**	1302	15940	**0,9**	0,8	1,0	*91%*
Référence "Commune"	**609**	**7061**	**1,0**				**609**	**7061**	**1,0**				**Référence "Commune"**	**609**	**7061**	**1,0**			
Total	**2091**	**24769**					**789**	**8829**					**Total**	**2091**	**24769**				
≤150 kV																			
0-49m	9	120	**0,8**	0,4	1,7	*68%*	8	107	**0,8**	0,4	1,8	*67%*	**0-49m**	8	118	**0,8**	0,4	1,7	*68%*
50-99m	16	155	**1,2**	0,7	2,0	*27%*	15	134	**1,2**	0,7	2,1	*24%*	**50-99m**	13	113	**1,2**	0,7	2,2	*29%*
100-199m	22	346	**0,7**	0,5	1,1	*91%*	16	283	**0,6**	0,4	1,1	*96%*	**100-199m**	20	296	**0,8**	0,5	1,3	*88%*
0-199m	47	621	**0,9**	0,6	1,2	*81%*	39	524	**0,8**	0,6	1,2	*85%*	**0-199m**	41	527	**0,9**	0,6	1,2	*79%*
200-599m	168	2103	**0,9**	0,7	1,1	*87%*	84	786	**1,2**	1,0	1,6	*6%*	**200-599m**	80	779	**1,2**	0,9	1,5	*11%*
≥ 600m et non réf. "Commune"	1267	14984	**1,0**	0,9	1,1	*75%*	57	458	**1,5**	1,1	2,0	*0%*	**≥ 600m et non réf. "Commune"**	59	462	**1,5**	1,1	2,0	*0%*
Géocodage "Commune" et non référence	0	0					0	0					**Pas de géocodage "au domicile" et non référence**	1302	15940	**0,9**	0,8	1,0	*91%*
Référence "Commune"	**609**	**7061**	**1,0**				**609**	**7061**	**1,0**				**Référence "Commune"**	**609**	**7061**	**1,0**			
Total	**2091**	**24769**					**789**	**8829**					**Total**	**2091**	**24769**				

(*) les sujets "précis" à 20 m sont les sujets dont le géocodage automatique a été trouvé grâce à la base IGN ou NAVTEQ au numéro; la médiane de la distribution de la distance du point géocodé "automatique" au point géocodé "au domicile" pour ces codes d'affectation est de 20m environ

Tableau 33 : Association entre incidence de LA et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne et de la taille de l'unité urbaine de la commune de résidence

≥225 kV	Distance "automatique"													Distance "au domicile"					
	Tous sujets						Sujets géocodés "au domicile"							Tous sujets					
TUU																			
< 5 000 habs	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	4	19	**2,5**	0,8	7,7	*5%*	3	5	**7,7**	1,7	33,6	*0%*	**0-49m**	1	7	**2,1**	0,3	17,8	*24%*
50-99m	1	20	**0,6**	0,1	4,6	*69%*	1	9	**0,9**	0,1	7,6	*53%*	**50-99m**	2	7	**2,6**	0,5	12,7	*12%*
100-199m	2	71	**0,3**	0,1	1,4	*93%*	1	38	**0,3**	0,0	2,2	*89%*	**100-199m**	3	40	**0,8**	0,2	2,7	*64%*
0-199m	7	110	**0,8**	0,4	1,7	*73%*	5	52	**1,0**	0,4	2,6	*48%*	**0-199m**	6	54	**1,2**	0,5	2,8	*35%*
200-599m	29	393	**0,8**	0,6	1,2	*81%*	9	116	**0,9**	0,5	1,9	*57%*	**200-599m**	7	116	**0,7**	0,3	1,5	*81%*
≥ 600m et non réf. "Commune"	524	5630	**1,0**	0,9	1,2	*29%*	37	297	**1,5**	1,0	2,1	*2%*	**≥ 600m et non réf. "Commune"**	45	306	**1,7**	1,2	2,3	*0%*
Géocodage "Commune" et non référence	36	67	**5,7**	3,7	8,7	*0%*	7	11	**5,7**	2,1	15,1	*0%*	**Pas de géocodage "au domicile" et non référence**	538	5724	**1,1**	0,9	1,2	*26%*
Référence "commune"	**308**	**3415**	**1,0**				**308**	**3415**	**1,0**				**Référence "Commune"**	**308**	**3415**	**1,0**			
Total	**904**	**9615**					**366**	**3891**					**Total**	**904**	**9615**				
5 000- 99 999 habs	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	4	6	**4,9**	1,2	19,1	*1%*	3	5	**3,3**	0,7	15,5	*6%*	**0-49m**	2	6	**2,9**	0,5	15,4	*11%*
50-99m	1	19	**0,6**	0,1	4,8	*68%*	1	11	**1,0**	0,1	9,3	*50%*	**50-99m**	2	14	**1,7**	0,4	7,8	*25%*
100-199m	5	33	**1,5**	0,5	3,9	*23%*	3	20	**1,4**	0,4	5,1	*30%*	**100-199m**	4	17	**2,1**	0,7	6,5	*10%*
0-199m	10	58	**1,7**	0,8	3,5	*9%*	7	36	**1,7**	0,7	4,4	*13%*	**0-199m**	8	37	**1,8**	0,7	4,3	*10%*
200-599m	27	219	**1,2**	0,7	2,0	*23%*	11	86	**1,1**	0,5	2,3	*38%*	**200-599m**	9	85	**1,0**	0,5	2,1	*51%*
≥ 600m et non réf. "Commune"	511	5811	**0,8**	0,6	1,1	*88%*	57	429	**1,2**	0,8	1,9	*16%*	**≥ 600m et non réf. "Commune"**	62	432	**1,3**	0,9	1,9	*10%*
Géocodage "Commune" et non Référence	18	63	**2,5**	1,4	4,5	*0%*	4	3	**19,2**	3,3	111,2	*0%*	**Pas de géocodage "au domicile" et non référence**	487	5597	**0,8**	0,6	1,1	*90%*
Référence "Commune"	**71**	**756**	**1,0**				**71**	**756**	**1,0**				**Référence "Commune"**	**71**	**756**	**1,0**			
Total	**637**	**6907**					**150**	**1310**					**Total**	**637**	**6907**				

≥100 000 habs	Cas	Témoins	OR	IC 95%		p_{uni}	Cas	Témoins	OR	IC 95%		p_{uni}		Cas	Témoins	OR	IC 95%		p_{uni}
0-49m	1	35	**0,4**	0,1	2,9	*82%*	1	28	**0,5**	0,1	3,7	*75%*	**0-49m**	1	25	**0,6**	0,1	4,2	*71%*
50-99m	6	58	**1,4**	0,6	3,3	*24%*	5	48	**1,2**	0,5	3,2	*35%*	**50-99m**	4	53	**1,0**	0,4	2,9	*48%*
100-199m	9	163	**0,7**	0,4	1,4	*84%*	8	123	**0,8**	0,4	1,7	*71%*	**100-199m**	11	115	**1,2**	0,6	2,4	*26%*
0-199m	16	256	**0,8**	0,5	1,4	*77%*	14	199	**0,9**	0,5	1,6	*69%*	**0-199m**	16	193	**1,0**	0,6	1,8	*48%*
200-599m	88	804	**1,3**	1,0	1,8	*2%*	41	314	**1,5**	1,0	2,3	*3%*	**200-599m**	38	319	**1,5**	1,0	2,2	*2%*
≥ 600m et non réf. "Commune"	886	9455	**1,1**	0,9	1,4	*11%*	72	559	**1,5**	1,0	2,2	*2%*	**≥ 600m et non réf. "Commune"**	77	567	**1,7**	1,3	2,4	*0%*
Géocodage "Commune" et non Référence	13	73	**2,1**	1,1	3,9	*1%*	4	7	**4,7**	1,2	18,3	*1%*	**Pas de géocodage "au domicile" et non référence**	872	9509	**1,1**	0,9	1,4	*15%*
Référence "Commune"	**230**	**2890**	**1,0**				**230**	**2890**	**1,0**				**Référence "Commune"**	**230**	**2890**	**1,0**			
Total	**1233**	**13478**					**361**	**3969**					**Total**	**1233**	**13478**				
TOTAL 0-14ans	**2774**	**30000**					**877**	**9170**						**2774**	**30000**				

≤150kV	**Distance "automatique"**												**Distance "au domicile"**						
	Tous sujets						**Sujets géocodés "au domicile"**							**Tous sujets**					
< 5 000 habs	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	1	34	**0,4**	0,1	3,0	*81%*	1	21	**0,6**	0,1	4,6	*69%*	**0-49m**	1	21	**0,7**	0,1	5,1	*65%*
50-99m	4	45	**1,1**	0,4	3,0	*45%*	2	23	**1,0**	0,2	4,3	*50%*	**50-99m**	2	19	**1,0**	0,2	4,6	*47%*
100-199m	7	109	**0,7**	0,3	1,6	*80%*	5	55	**0,9**	0,4	2,4	*55%*	**100-199m**	5	57	**1,0**	0,4	2,4	*54%*
0-199m	12	188	**0,8**	0,4	1,4	*81%*	8	99	**0,9**	0,4	1,9	*62%*	**0-199m**	8	97	**0,9**	0,4	1,9	*60%*
200-599m	68	724	**1,0**	0,8	1,4	*38%*	21	209	**1,2**	0,7	1,9	*23%*	**200-599m**	21	207	**1,2**	0,7	1,9	*27%*
≥ 600m et non réf. "Commune"	480	5221	**1,0**	0,9	1,2	*34%*	22	157	**1,6**	1,0	2,6	*2%*	**≥ 600m et non réf. "Commune"**	29	172	**1,9**	1,2	2,9	*0%*
Géocodage "Commune" et non Référence	36	67	**5,7**	3,7	8,8	*0%*	7	11	**5,7**	2,1	15,2	*0%*	**Pas de géocodage "au domicile" et non référence**	538	5724	**1,1**	0,9	1,2	*25%*
Référence "Commune"	**308**	**3415**	**1,0**				**308**	**3415**	**1,0**				**Référence "Commune"**	**308**	**3415**	**1,0**			
Total	**904**	**9615**					**366**	**3891**					**Total**	**904**	**9615**				

5 000- 99 999 habs	Cas	Témoins	OR	IC 95%		p_{uni}	Cas	Témoins	OR	IC 95%		p_{uni}		Cas	Témoins	OR	IC 95%		p_{uni}
0-49m	7	53	**1,2**	0,5	2,7	*37%*	4	36	**0,7**	0,2	2,1	*75%*	**0-49m**	3	41	**0,7**	0,2	2,2	*74%*
50-99m	5	73	**0,7**	0,3	1,7	*80%*	5	56	**0,7**	0,3	1,9	*76%*	**50-99m**	6	45	**1,2**	0,5	3,1	*32%*
100-199m	15	144	**1,0**	0,5	1,8	*53%*	8	89	**0,7**	0,3	1,7	*76%*	**100-199m**	10	96	**1,0**	0,5	2,1	*48%*
0-199m	27	270	**0,9**	0,6	1,5	*60%*	17	181	**0,7**	0,4	1,3	*86%*	**0-199m**	19	182	**0,8**	0,4	1,5	*75%*
200-599m	71	876	**0,7**	0,5	1,1	*95%*	36	264	**1,2**	0,7	1,9	*25%*	**200-599m**	34	264	**1,1**	0,7	1,7	*40%*
≥ 600m et non réf. "Commune"	450	4942	**0,9**	0,7	1,2	*81%*	22	106	**2,3**	1,3	4,1	*0%*	**≥ 600m et non réf. "Commune"**	26	108	**2,4**	1,4	4,0	*0%*
Géocodage "Commune" et non Référence	18	63	**2,5**	1,4	4,5	*0%*	4	3	**19,9**	3,4	116,7	*0%*	**Pas de géocodage "au domicile" et non référence**	487	5597	**0,8**	0,6	1,1	*90%*
Référence "Commune"	**71**	**756**	**1,0**				**71**	**756**	**1,0**				**Référence "Commune"**	**71**	**756**	**1,0**			
Total	**637**	**6907**					**150**	**1310**					**Total**	**637**	**6907**				
≥100 000 habs	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}	**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}		**Cas**	**Témoins**	**OR**	IC 95%		p_{uni}
0-49m	6	77	**1,0**	0,4	2,4	*49%*	5	64	**0,9**	0,3	2,3	*61%*	**0-49m**	6	65	**1,3**	0,5	3,0	*31%*
50-99m	12	85	**1,7**	0,9	3,3	*5%*	9	64	**1,5**	0,7	3,4	*14%*	**50-99m**	6	57	**1,3**	0,6	3,2	*25%*
100-199m	11	208	**0,7**	0,4	1,3	*89%*	7	165	**0,5**	0,2	1,1	*97%*	**100-199m**	9	169	**0,7**	0,3	1,4	*85%*
0-199m	29	370	**1,0**	0,6	1,5	*51%*	21	293	**0,8**	0,4	1,3	*84%*	**0-199m**	21	291	**0,8**	0,4	1,3	*83%*
200-599m	102	1140	**1,1**	0,8	1,4	*32%*	50	390	**1,3**	0,9	2,0	*9%*	**200-599m**	48	393	**1,5**	1,0	2,1	*1%*
≥ 600m et non réf. "Commune"	859	9005	**1,2**	0,9	1,4	*8%*	56	389	**1,7**	1,2	2,5	*0%*	**≥ 600m et non réf. "Commune"**	62	395	**2,0**	1,4	2,8	*0%*
Géocodage "Commune" et non Référence	13	73	**2,1**	1,1	3,9	*1%*	4	7	**4,6**	1,2	17,6	*1%*	**Pas de géocodage "au domicile" et non référence**	872	9509	**1,1**	0,9	1,3	*18%*
Référence "Commune"	**230**	**2890**	**1,0**				**230**	**2890**	**1,0**				**Référence "Commune"**	**230**	**2890**	**1,0**			
Total	**1233**	**13478**					**361**	**3969**					**Total**	**1233**	**13478**				
TOTAL 0-14ans	**2774**	**30000**					**877**	**9170**						**2774**	**30000**				

Tableau 34 : Association entre incidence de LAL et 2 indicateurs de distance aux LHT : distance « automatique » et distance « au domicile », en fonction de la tension de la ligne

LAL	Distance "automatique"												Distance "au domicile"					
	Tous sujets						Sujets géocodés "au domicile"						Tous sujets					
	Cas	Témoins	OR	IC 95%		p_{uni}	Cas	Témoins	OR	IC 95%		p_{uni}		Cas	Témoins	OR	IC 95%		p_{uni}
≥225 kV																			
0-49m	8	60	**1,9**	0,9	4,0	*5%*	7	38	**2,8**	1,2	6,3	*1%*	**0-49m**	4	38	**1,6**	0,6	4,6	*19%*
50-99m	6	97	**0,9**	0,4	2,1	*57%*	6	68	**1,3**	0,5	3,0	*29%*	**50-99m**	7	74	**1,4**	0,6	3,1	*20%*
100-199m	15	267	**0,8**	0,5	1,4	*80%*	11	181	**0,9**	0,5	1,7	*63%*	**100-199m**	14	172	**1,1**	0,7	2,0	*28%*
0-199m	29	424	**1,0**	0,7	1,5	*53%*	24	287	**1,2**	0,8	1,9	*17%*	**0-199m**	25	284	**1,3**	0,8	2,0	*12%*
200-599m	116	1416	**1,1**	0,9	1,4	*12%*	48	516	**1,3**	0,9	1,8	*5%*	**200-599m**	44	520	**1,2**	0,9	1,7	*14%*
≥ 600m et non réf. "Commune"	1583	20896	**1,1**	0,9	1,2	*20%*	139	1285	**1,5**	1,2	1,9	*0%*	**≥ 600m et non réf. "Commune"**	152	1305	**1,6**	1,3	2,0	*0%*
Géocodage "Commune" et non Référence	54	203	**3,4**	2,5	4,7	*0%*	10	21	**5,5**	2,5	12,0	*0%*	**Pas de géocodage "au domicile" et non référence**	1561	20830	**1,0**	0,9	1,2	*26%*
Référence "Commune"	**488**	**7061**	**1,0**				**488**	**7061**	**1,0**				**Référence "Commune"**	**488**	**7061**	**1,0**			
TOTAL	**2270**	**30000**					**709**	**9170**					**TOTAL**	**2270**	**30000**				
≤150kV																			
0-49m	13	164	**1,1**	0,6	1,9	*39%*	9	121	**1,0**	0,5	2,0	*53%*	**0-49m**	9	127	**1,0**	0,5	2,0	*48%*
50-99m	18	203	**1,2**	0,8	2,0	*20%*	15	143	**1,5**	0,8	2,5	*9%*	**50-99m**	14	121	**1,5**	0,8	2,7	*9%*
100-199m	28	461	**0,8**	0,6	1,3	*79%*	18	309	**0,8**	0,5	1,3	*82%*	**100-199m**	21	322	**0,9**	0,6	1,4	*71%*
0-199m	59	828	**1,0**	0,7	1,3	*52%*	42	573	**1,0**	0,7	1,4	*52%*	**0-199m**	44	570	**1,0**	0,6	1,5	*41%*
200-599m	201	2740	**1,0**	0,8	1,2	*50%*	92	863	**1,5**	1,2	1,9	*0%*	**200-599m**	88	864	**1,4**	1,1	1,8	*0%*
≥ 600m et non réf. "Commune"	1468	19168	**1,1**	1,0	1,2	*14%*	77	652	**1,8**	1,3	2,3	*0%*	**≥ 600m et non réf. "Commune"**	89	675	**1,9**	1,4	2,4	*0%*
Géocodage "Commune" et non Référence	54	203	**3,4**	2,5	4,7	*0%*	10	21	**5,5**	2,5	12,0	*0%*	**Pas de géocodage "au domicile" et non référence**	1561	20830	**1,0**	0,9	1,2	*27%*
Référence "Commune"	**488**	**7061**	**1,0**				**488**	**7061**	**1,0**				**Référence "Commune"**	**488**	**7061**	**1,0**			
TOTAL	**2270**	**30000**					**709**	**9170**					**TOTAL**	**2270**	**30000**				

D Discussion

Nous avons trouvé une association à la limite de la significativité statistique entre le fait d'habiter à moins de 50 m d'une LHT de plus de 225 kV et le statut cas *vs* témoin. Cette association est nette quand on étudie la distance « automatique », chez tous les sujets, mais l'est moins en distance « au domicile », chez les 70% des sujets pour lesquels le géocodage « au domicile » a été possible. Aucune association n'a été trouvée avec la proximité aux lignes de plus faible tension (63kV, 95 kV, 150 kV) alors même qu'elle concernait un nombre plus important de sujets.

Nous avons aussi étudié séparément les enfants de moins de 5 ans, notamment parce qu'ils sont moins susceptibles d'avoir changé de domicile depuis leur naissance et ceux de 5 à 14 ans pour lesquels un tel changement est plus probable : dans l'enquête cas-témoin française en population générale (2003-2004) Escale, seuls 30% des témoins de 5-14 ans n'ont pas déménagé entre la naissance et le diagnostic alors que 60% des 0-4 ans sont dans le même domicile depuis la naissance. L'association trouvée chez les 0-14 ans semble limitée aux enfants de 0 à 4 ans puisqu'elle est plus forte pour cette classe d'âge, y compris en distance « au domicile » et n'est pas retrouvée chez les sujets de 5 à 14 ans. Le test d'interaction entre le fait d'habiter à moins de 50 m d'une LHT de plus de 225 kV et le fait d'avoir moins de 5 ans est significatif à 11% en distance « automatique » et à 12% en distance « au domicile ». Par ailleurs, l'association semble plus forte dans les tranches d'unité urbaine « rurales » et « semi-urbaines » que dans les tranches d'unité urbaine de plus de 100 000 habitants : l'interaction entre le fait d'habiter à moins de 50 m en distance « automatique » d'une ligne de plus de 225 kV et

d'habiter dans une commune de taille d'unité urbaine de moins 100 000 habitants est significative à 4% et à 17% en distance « au domicile ».

Nous n'avions pas d'information permettant d'ajuster nos résultats sur l'ensemble des facteurs de risque de LA de l'enfant établis ou fortement suspectés, potentiellement associés de ce fait à des effets de confusion. Cependant, nous avons pu montrer que nos observations restaient stables en excluant les sujets domiciliés à proximité des centrales nucléaires de production d'électricité.

Ce sont nos observations faites sur les LHT de plus de 225 kV qui doivent être comparées à celles de l'étude Draper [Draper *et al.*,2005] puisque les LHT incluses dans cette étude étaient principalement de tension 275 et 400 kV (et en très faible proportion 132 kV). L'association trouvée dans notre étude montre une association chez les sujets à moins de 50 m des lignes alors que dans l'étude anglaise, l'association était significative jusqu'à 200 m et plus faiblement jusqu'à 600 m. L'adresse étudiée dans notre étude est l'adresse au diagnostic alors que celle étudiée par Draper est l'adresse à la naissance : cette différence peut contribuer à l'observation de résultats pour partie différents dans ces deux études.

Parmi les hypothèses possibles pouvant induire un lien entre distance aux lignes et incidence de leucémie de l'enfant, la principale étudiée jusqu'à ce jour est celle qui met en cause les CM-EBF induits par ces lignes. Cependant, il n'existe pas d'argument expérimental en faveur de la cancérogénicité directe de ces champs qui, n'induisent pas de lésion du génome, même à forte dose, contrairement aux radiations et rayonnements à très haute fréquence qui sont ionisants. Par ailleurs,

une relation quantitative exposition-risque a très rarement été retrouvée dans les études.

D'autres hypothèses doivent donc être aussi envisagées. Un des facteurs parfois suspecté comme inducteur de cette association est celle des ions corona : cette hypothèse part du fait que la présence de LHT peut créer une ionisation de l'air, des ions corona étant ainsi présents dans l'environnement des lignes (en moyenne jusqu'à 400 m) qui se fixent sur des particules (potentiellement polluantes) et seront plus facilement inhalées du fait de leur ionisation [Fews *et al.*,1999]. Une autre hypothèse soulevée est celle d'un environnement socio-économique associé à la présence des LHT ou qui aurait conduit à la construction de ces lignes : éloignement des pôles urbains, milieu potentiellement défavorisé ([UKCCS,1999] ; [Hatch *et al.*,2000]). Il reste à évaluer dans quelle mesure ces différents facteurs, dont la portée en distance est bien plus importante que celles recouvrant uniquement la zone de décroissance des champs magnétiques, pourraient expliquer une association entre leucémies de l'enfant et proximité aux lignes dans des zones autour des lignes.

La localisation des lignes souterraines étant connue par Rte (de manière moins précise que les lignes aériennes) la prise en compte de ces données sera faite dans des analyses ultérieures, au moins comme paramètre d'ajustement. La localisation des lignes SNCF à courant alternatif n'a pas été prise en compte mais les champs induits par ces lignes ont des caractéristiques sensiblement différentes de ceux induits par les lignes Rte : décroissance plus lente et très dépendante du trafic puisque le courant passe dans la ligne uniquement au moment du passage d'un train.

La simple « distance aux LHT » permet certes de définir des sujets qui sont dans des zones d'exposition aux CM-EBF liés à ces lignes, mais est loin de fournir une estimation quantitative de l'exposition aux CM-EBF qui en résulte puisque de nombreux autres paramètres influencent l'intensité des CM. Plusieurs études ont cherché à valider le fait que la proximité directe des LHT était associée à une forte exposition aux CM. Globalement, elles trouvent que les mesures sur 24h des enfants (prenant donc en compte l'exposition à l'école et par d'autres sources) sont corrélées au fait d'habiter proche d'une LHT puisque l'exposition des sujets proches des lignes est globalement plus élevée que celles des sujets ne vivant pas à proximité des lignes[Kleinerman et al.,1997] ; [Vistnes et al.,1997] ; [Schüz et al.,2000] ; [Forssén et al.,2002]. Cependant, de nombreuses erreurs de classement sont faites en considérant que la seule distance [Maslanyj et al.,2008] ; [Kroll 2010]. Le fait d'étudier séparément les lignes 225 – 400 kV des lignes de plus faible tension permet au moins de ne pas classer de la même manière des sujets proches de ces types de lignes induisant des CM d'intensités très différentes à une même distance.

Dans la continuité de cette étude, un indice quantitatif d'exposition résidentielle est en cours d'élaboration en étroite collaboration avec Rte. Il sera utilisé pour estimer l'exposition précise aux CM-EBF liée à la proximité des LHT des sujets de GEOCAP. Cet indicateur prend en compte le nombre et le type de configuration des lignes à proximité des habitations étudiées, l'intensité du courant qui parcourt la ligne, la hauteur des conducteurs, le dénivelé du terrain ainsi que les distances de l'habitation à la ligne et au pylône le plus proche qui sont les déterminants importants du CM. Dans cette démarche, non seulement l'intensité du CM-EBF lié à la proximité et aux

caractéristiques des LHT en un point donné est estimée mais aussi son incertitude. Quand les sujets habitent très près des lignes (30 ou 50 m selon les situations) et dans certains cas particuliers de fort dénivelé ou de hauteur de conducteur spécialement élevée, l'estimation de leur exposition est traitée manuellement afin de bien l'estimer en fonction de la situation particulière rencontrée. Ce travail a donné lieu à un rapport méthodologique rédigé par Rte qui explicite très précisément la méthodologie choisie pour estimer l'exposition aux champs magnétiques au mieux dans le cadre de cette étude. Ce rapport fera prochainement l'objet d'un article soumis pour publication dans une revue spécialisée de façon antérieure à l'analyse cas-témoins qui sera réalisée ultérieurement et donc de façon indépendante de celle-ci.

VII. Conclusions et perspectives

A Conclusion

• La première partie de ce travail concernait l'étude de **l'incidence des LA par département** sur la période 1990-2004. Nous avons utilisé différentes méthodes de lissage adaptées à l'étude de maladies rares et permettant de tenir compte de potentielles autocorrélations spatiales. Même si une légère hétérogénéité du taux d'incidence des LA a été détectée, elle n'était pas associée à une tendance ou structure spatiale stable par sous-type de leucémie, période de temps, sexe ou groupe d'âge sur la période 1990-2004. Ce résultat n'exclut ni qu'il pourrait exister une hétérogénéité spatiale de cette incidence à des échelles plus fines ni que l'incidence des LA pourrait être corrélée spatialement avec certaines caractéristiques populationnelles qui auraient une variabilité interdépartementale. Cependant, cela implique qu'il n'y a pas de raison d'estimer des taux d'incidence de référence à un niveau départemental plutôt que national, ni d'ajuster sur le département des études ultérieures qui examineraient les variations spatiales de l'incidence des LA, sauf si elles utilisaient des échantillons eux même stratifiés sur le département.

• L'étude de l'influence de facteurs environnementaux dans l'incidence des LA grâce au programme GEOCAP a fait l'objet des deux parties. **Concernant les CNPE**, sur la période 2002-2007 contrairement aux périodes précédentes (1990-2001), un quasi-doublement de l'incidence des LA à moins de 5 km des CNPE a été mise en évidence, avec l'approche cas-témoin comme avec l'étude d'incidence. Cette augmentation, statistiquement significative et relativement conséquente semblait être plus importante sur les années 2006 et 2007 bien qu'elle persistait après l'exclusion une par une de chacune des 6 années 2002 à 2007. Elle n'était pas spécifique d'une CNPE ou d'un type de CNPE. Par ailleurs l'incidence des LA observée au voisinage

des CNPE ne présentait aucun lien avec la cartographie des émissions aériennes de radioactivité par les CNPE. Enfin, ce quasi doublement d'incidence de LAà moins de 5 km des CNPE est largement supérieur à l'excès de risque attendu au voisinage des CNPE sur la base des modélisations de rejets gazeux et liquides et des relations dose-risque actuellement considérées comme de référence. Au total, ces observations suggèrent l'influence d'autres facteurs de risques, dont la prévalence doit être pus élevée au voisinage immédiat des CNPE, et qui peuvent bien évidemment aussi être présents dans tout autre environnement. Ces facteurs seraient donc intéressants à bien cerner aussi bien pour mieux comprendre les variations d'incidence des LA au voisinage des CNPE que pour mieux comprendre les facteurs de variations spatiales et de risque de LA en général. Ils restent à déterminer tant au niveau français que dans le cadre d'une collaboration européenne qui permettra de regrouper toutes les études déjà faites pour avoir une meilleure puissance d'analyse.

• **L'étude de la proximité des LHT** a montré une légère association entre l'incidence des LA et la proximité aux lignes de plus de 225 kV mais pas avec la proximité aux LHT de 150 kV et moins, alors même que les effectifs de cas et témoins impliqués par la proximité à ces lignes à plus faible tension étaient plus importants que ceux impliqués par la proximité aux lignes de plus forte tension. L'association observée avec la proximité aux LHT de 225 kV et plus, était limitée aux 0-4 ans, qui sont aussi les sujets qui sont le moins susceptibles d'avoir déménagé depuis la naissance, ainsi que aux enfants des zones les moins urbaines : communes de tranche d'unité urbaine de moins de 100 000 habitants. Ces résultats sont concordants avec ceux de la littérature mais devront être confirmés et/ou affinés par l'étude d'estimations

modélisées des expositions complétant l'approche par exploitation des critères de distance aux lignes et de tension de lignes (63 à 450 kV) développée ici.

B Perspectives

Une première perspective suite à ces études est de **confirmer ces résultats sur des périodes ultérieures**. Le RNHE sera exhaustif pour les années 2008 à 2010 en fin d'année 2012 et nous mettons en place avec l'INSEE le tirage d'un nouvel échantillon témoin pour ces mêmes années. L'étude de l'incidence des LA à proximité des CNPE et des LHT pourra donc être mise à jour sur ces années.

L'extension de l'enregistrement des cas du RNHE est envisagée à moyen terme aux adolescents de 15 à 18 ans et aux enfants et adolescents des DOM-TOM.

Dans le programme HOPE-Epidémiologie développé par J. Clavel, il est prévu que les **histoires résidentielles** des cas soient enregistrées de manière systématique. Ainsi, l'exposition à différentes sources d'exposition à un moment précis (par exemple « au moment de la conception ») pourra être étudiée, tout comme les expositions cumulées entre la conception et le diagnostic. Enfin, le recueil des observations d'une nouvelle enquête cas-témoins par questionnaire (ESTELLE) vient de s'achever qui comporte le relevé de l'histoire résidentielle des sujets, depuis leur naissance, qui permettra une estimation de l'histoire de l'exposition d'intérêt de la naissance au diagnostic et aussi de mettre en corrélation les différents types de trajectoires résidentielles avec d'éventuelles particularités des sujets et/ou de leur parents.

D'autres volets de l'exposition environnementale en lien avec l'incidence des cancers de l'enfant sont et vont être étudiés grâce au programme GEOCAP et notamment **l'exposition au benzène du au trafic routier**. En relation avec AIRPARIF et l'InVS,

nous avons déjà bien avancé le travail sur la recherche d'indicateurs pertinents : prise en compte de la pollution de fond qui est globalement très différente en milieu rural et en milieu urbain, distance à la voie la plus proche - par type de voie-, densité de voies et de trafic dans un disque de rayon donné autour de l'habitation et estimation au NO_2 et au benzène. Ces premiers travaux ont permis de mettre en évidence une association positive nette entre les divers indicateurs d'exposition au trafic routier et/ou au NO_2 et l'incidence des leucémies de l'enfant dans l'étude cas - témoins ESCALE [Amigou *et al.*,2011]. Ces indicateurs d'exposition à la pollution atmosphérique liée au trafic routier vont maintenant être étudiés sur les sujets GEOCAP.

Par ailleurs, la thèse en cours de Claire Demoury comporte une partie sur l'étude de l'association entre l'incidence de LA et **l'exposition au radon** d'origine domestique chez les sujets GEOCAP.

Enfin, le Registre National des Tumeurs Solides de l'Enfant (RNTSE), dirigé par B. Lacour, Nancy, enregistre les tumeurs de l'enfant depuis 2000 (950 cas par an dont 350 tumeurs du système nerveux central). Le RNHE et le RNTSE ont fonctionné de façon étroitement coordonnée depuis leur mise en place et ont maintenant fusionné. L'association entre l'incidence des tumeurs solides de l'enfant et les différents facteurs d'exposition environnementale décrits dans cette thèse sera étudiée dans ce cadre.

VIII. Bibliographie

[Abdul_Rahman2008] Abdul Rahman HI, Shah SA, Alias H & Ibrahim HM. A case-control study on the association between environmental factors and the occurrence of acute leukemia among children in Klang valley, Malaysia.. *Asian Pac J Cancer Prev* (2008) **9**: pp. 649-652.

[Ahlbom2000] Ahlbom A, Day N, Feychting M, Roman E, Skinner J, Dockerty J, Linet M, McBride M, Michaelis J, Olsen JH, Tynes T & Verkasalo PK. A pooled analysis of magnetic fields and childhood leukaemia. *Br J Cancer* (2000) **83**: pp. 692-698.

[Alexander1993] Alexander FE. Viruses, clusters and clustering of childhood leukaemia: a new perspective?. *Eur J Cancer* (1993) **29A**: pp. 1424-1443.

[Alexander1998] Alexander FE, Boyle P, Carli PM, Coebergh JW, Draper GJ, Ekbom A, Levi F, McKinney PA, McWhirter W, Michaelis J, Peris-Bonet R, Petridou E, Pompe-Kirn V, Plìsko I, Pukkala E, Rahu M, Storm H, Terracini B, Vatten L & Wray N. Spatial clustering of childhood leukaemia: summary results from the Euroclus project. *Br J Cancer* (1998) **77**: pp. 818-824.

[Amigou2011] Amigou A, Sermage-Faure C, Orsi L, Leverger G, Baruchel A, Bertrand Y, Nelken B, Robert A, Michel G, Margueritte G, Perel Y, Mechinaud F, Bordigoni P, Hémon D & Clavel J. Road traffic and childhood leukemia: the escale study (sfce). *Environ Health Perspect* (2011) **119**: pp. 566-572.

[Bellec2006] Bellec S, Hémon D, Rudant J, Goubin A & Clavel J. Spatial and space-time clustering of childhood acute leukaemia in France from 1990 to 2000: a nationwide study. *Br J Cancer* (2006) **94**: pp. 763-770.

[Bellec2008#150] Bellec S, Baccaïni B, Goubin A, Rudant J, Ripert M, Hémon D & Clavel J. Childhood leukaemia and population movements in France, 1990-2003. *Br J Cancer* (2008) **98**: pp. 225-231.

[Besag1991] Besag J, York J & Mollié A. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* (1991) **43**: pp. 1-59.

[Best2005] Best N, Richardson S & Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* (2005) **14**: pp. 35-59.

[Bianchi2000] Bianchi N, Crosignani P, Rovelli A, Tittarelli A, Carnelli CA, Rossitto F, Vanelli U, Porro E & Berrino F. Overhead electricity power lines and

childhood leukemia: a registry-based, case-control study. *Tumori* (2000) **86**: pp. 195-198.

[Bithell1994] Bithell JF, Dutton SJ, Draper GJ & Neary NM. Distribution of childhood leukaemias and non-Hodgkin's lymphomas near nuclear installations in England and Wales. *BMJ* (1994) **309**: pp. 501-505.

[Bithell2008] Bithell JF, Keegan TJ, Kroll ME, Murphy MFG & Vincent TJ. Childhood leukaemia near British nuclear installations: methodological issues and recent results. *Radiat Prot Dosimetry* (2008) : .

[Bithell2008#334] Bithell JF, Keegan TJ, Kroll ME, Murphy MFG & Vincent TJ. Childhood leukaemia near British nuclear installations: methodological issues and recent results. *Radiat Prot Dosimetry* (2008) **132**: pp. 191-197.

[Black1994] Black RJ, Sharp L, Harkness EF & McKinney PA. Leukaemia and non-Hodgkin's lymphoma: incidence in children and young adults resident in the Dounreay area of Caithness, Scotland in 1968-91. *J Epidemiol Community Health* (1994) **48**: pp. 232-236.

[Bouges1999] Bouges S, Daurès JP & Hébrard M. [incidence of acute leukemias, lymphomas and thyroid cancers in children under 15 years, living around the Marcoule nuclear site from 1985 to 1995]. *Rev Epidemiol Sante Publique* (1999) **47**: pp. 205-217.

[Breslow1987] Breslow NE & Day NE. Statistical methods in cancer research. volume ii--the design and analysis of cohort studies. *IARC Sci Publ* (1987) : pp. 1-406.

[Brosselin2009] Brosselin P, Rudant J, Orsi L, Leverger G, Baruchel A, Bertrand Y, Nelken B, Robert A, Michel G, Margueritte G, Perel Y, Mechinaud F, Bordigoni P, Hemon D & Clavel J. Childhood acute leukemia and residence next to gas stations and automotive repair garages: the ESCALE study (SFCD*). *Occup Environ Med* (2009).

[Buckley1989] Buckley JD, Robison LL, Swotinsky R, Garabrant DH, LeBeau M, Manchester P, Nesbit ME, Odom L, Peters JM, Woods WG & et al. Occupational exposures of parents of children with acute non lymphocytic leukemia: a report from the children cancer study group.. *Cancer Res* (1989) **49**: pp. 4030-4037.

[COMARE2011] Committee on Medical Aspects of Radiation in the Environment (COMARE). Fourteenth report. further consideration of the incidence of childhood

leukaemia around nuclear power plants in Great Britain. Health protection agency. Chilton 2011. (2011) .

[Clavel2004] Clavel J, Goubin A, Auclerc MF, Auvrignon A, Waterkeyn C, Patte C, Baruchel A, Leverger G, Nelken B, Philippe N, Sommelet D, Vilmer E, Bellec S, Perrillat-Menegaux F & Hémon D. Incidence of childhood leukaemia and non-Hodgkin's lymphoma in France: national registry of childhood leukaemia and lymphoma, 1990-1999. *Eur J Cancer Prev* (2004) **13**: pp. 97-103.

[Cliff1981] Cliff A & Ord J. Spatial processes. models and applications. Pion L (Ed.). , 1981.

[Clifford1989] Clifford P, Richardson S & Hémon D. Assessing the significance of the correlation between two spatial processes. *Biometrics* (1989) **45**: pp. 123-134.

[Coghill1996] Coghill RW, Steward J & Philips A. Extra low frequency electric and magnetic fields in the bedplace of children diagnosed with leukaemia: a case-control study. *Eur J Cancer Prev* (1996) **5**: pp. 153-158.

[Crosignani2004] Crosignani P, Tittarelli A, Borgini A, Codazzi T, Rovelli A, Porro E, Contiero P, Bianchi N, Tagliabue G, Fissi R, Rossitto F & Berrino F. Childhood leukemia and road traffic: a population-based case-control study. *Int J Cancer* (2004) **108**: pp. 596-599.

[Crosignani2004#36] Crosignani P, Tittarelli A, Borgini A, Codazzi T, Rovelli A, Porro E, Contiero P, Bianchi N, Tagliabue G, Fissi R, Rossitto F & Berrino F. Childhood leukemia and road traffic: a population-based case-control study. *Int J Cancer* (2004) **108**: pp. 596-599.

[Demoury2011] Demoury C, Goujon-Bellec S, Guyot-Goubin A, Hémon D & Clavel J. Spatial variations of childhood acute leukaemia in France, 1990-2006: global spatial heterogeneity and cluster detection at 'living-zone' level. *Eur J Cancer Prev* (2011).

[Dickinson1999] Dickinson HO & Parker L. Quantifying the effect of population mixing on childhood leukaemia risk: the Seascale cluster. *Br J Cancer* (1999) **81**: pp. 144-151.

[Dockerty1999] Dockerty JD, Elwood JM, Skegg DC & Herbison GP. Electromagnetic field exposures and childhood leukaemia in New Zealand. *Lancet* (1999) **354**: pp. 1967-1968.

[Draper2005] Draper G, Vincent T, Kroll ME & Swanson J. Childhood cancer in relation to distance from high voltage power lines in England and Wales: a case-control study.. *BMJ* (2005) **330**: p. 1290.

[Elliott2010] Elliott P, Toledano MB, Bennett J, Beale L, de Hoogh K, Best N & Briggs DJ. Mobile phone base stations and early childhood cancers: case-control study. *BMJ* (2010) **340**: p. c3077.

[Estève1993] Estève J, Benhamou E & Raymond L. Méthodes statistiques en épidémiologie descriptive. Les Editions INSERM P (Ed.). , 1993.

[Evrard2005] Evrard AS, Hémon D, Billon S, Laurier D, Jougla E, Tirmarche M & Clavel J. Ecological association between indoor radon concentration and childhood leukaemia incidence in France, 1990-1998. *Eur J Cancer Prev* (2005) **14**: pp. 147-157.

[Evrard2006#91] Evrard A, Hémon D, Morin A, Laurier D, Tirmarche M, Backe J, Chartier M & Clavel J. Childhood leukaemia incidence around French nuclear installations using geographic zoning based on gaseous discharge dose estimates.. *Br J Cancer* (2006) **94**: pp. 1342-1347.

[Fajardo-Gutiérrez1993] Fajardo-Gutiérrez A, Garduño-Espinosa J, Yamamoto-Kimura L, Hernández-Hernández DM, Gómez-Delgado A, Mejía-Aranguré M, Cartagena-Sandoval A & Martínez-García MC. [residence close to high-tension electric power lines and its association with leukemia in children]. *Bol Med Hosp Infant Mex* (1993) **50**: pp. 32-38.

[Feizi2007] Feizi AAHP & Arabi MAA. Acute childhood leukemias and exposure to magnetic fields generated by high voltage overhead power lines - a risk factor in Iran.. *Asian Pac J Cancer Prev* (2007) **8**: pp. 69-72.

[Fews1999] Fews AP, Henshaw DL, Wilding RJ & Keitch PA. Corona ions from power lines and increased exposure to pollutant aerosols. *Int J Radiat Biol* (1999) **75**: pp. 1523-1531.

[Feychting1993] Feychting M & Ahlbom A. Magnetic fields and cancer in children residing near Swedish high-voltage power lines. *Am J Epidemiol* (1993) **138**: pp. 467-481.

[Feychting1998] Feychting M, Svensson D & Ahlbom A. Exposure to motor vehicle exhaust and childhood cancer. *Scand J Work Environ Health* (1998) **24**: pp. 8-11.

[Forssén2002] Forssén UM, Ahlbom A & Feychting M. Relative contribution of residential and occupational magnetic field exposure over twenty-four hours among people living close to and far from a power line. *Bioelectromagnetics* (2002) **23**: pp. 239-244.

[Fulton1980] Fulton JP, Cobb S, Preble L, Leone L & Forman E. Electrical wiring configurations and childhood leukemia in Rhode island.. *Am J Epidemiol* (1980) **111**: pp. 292-296.

[Goldsmith1992] Goldsmith JR. Nuclear installations and childhood cancer in the uk: mortality and incidence for 0-9-year-old children, 1971-1980. *Sci Total Environ* (1992) **127**: p. 13-35, 43-55.

[Goubin2006] Goubin A, Auclerc MF, Auvrignon A, Patte C, Bergeron C, Hémon D & Clavel J. Survival in France after childhood acute leukaemia and non-Hodgkin's lymphoma (1990-2000). *Eur J Cancer* (2006) **42**: pp. 534-541.

[Greaves1988] Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia.. *Leukemia* (1988) **2**: pp. 120-125.

[Greaves2006] Greaves M. Infection, immune responses and the etiology of childhood leukaemia. *Nat Rev Cancer* (2006) **6**: pp. 193-203.

[Green1995] Green P. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* (1995) **82**: pp. 711-732.

[Green1999] Green LM, Miller AB, Villeneuve PJ, Agnew DA, Greenberg ML, Li J & Donnelly KE. A case-control study of childhood leukemia in southern Ontario, Canada, and exposure to magnetic fields in residences. *Int J Cancer* (1999) **82**: pp. 161-170.

[Green1999#68] Green LM, Miller AB, Agnew DA, Greenberg ML, Li J, Villeneuve PJ & Tibshirani R. Childhood leukemia and personal monitoring of residential exposures to electric and magnetic fields in Ontario, Canada. *Cancer Causes Control* (1999) **10**: pp. 233-243.

[Green2002] Green P & Richardson S. Hidden Markov models and disease mapping. *Journal of the American Statistical Association* (2002) **97**: pp. 1055-1070.

[Greenland2000] Greenland S, Sheppard AR, Kaune WT, Poole C & Kelsh MA. A pooled analysis of magnetic fields, wire codes, and childhood leukemia. childhood leukemia-EMF study group. *Epidemiology* (2000) **11**: pp. 624-634.

[Harrison1999] Harrison RM, Leung PL, Somervaille L, Smith R & Gilman E. Analysis of incidence of childhood cancer in the west midlands of the United Kingdom

in relation to proximity to main roads and petrol stations. *Occup Environ Med* (1999) **56**: pp. 774-780.

[Hatch2000] Hatch EE, Kleinerman RA, Linet MS, Tarone RE, Kaune WT, Auvinen A, Baris D, Robison LL & Wacholder S. Do confounding or selection factors of residential wiring codes and magnetic fields distort findings of electromagnetic fields studies?. *Epidemiology* (2000) **11**: pp. 189-198.

[Heasman1986] Heasman M, Kemp I, Urquhart J & Black R. Childhood leukaemia in northern Scotland. *Lancet* (1986) **1**: p. 266.

[Heinävaara2010] Heinävaara S, Toikkanen S, Pasanen K, Verkasalo PK, Kurttio P & Auvinen A. Cancer incidence in the vicinity of Finnish nuclear power plants: an emphasis on childhood leukemia. *Cancer Causes Control* (2010) **21**: pp. 587-595.

[Hoffmann1997] Hoffmann W, Dieckmann H, Dieckmann H & Schmitz-Feuerhake I. A cluster of childhood leukemia near a nuclear reactor in northern Germany. *Arch Environ Health* (1997) **52**: pp. 275-280.

[Hoffmann2007] Hoffmann W, Terschueren C & Richardson DB. Childhood leukemia in the vicinity of the Geesthacht nuclear establishments near Hamburg, Germany. *Environ Health Perspect* (2007) **115**: pp. 947-952.

[IARC1982] IARC. Monographs on the evaluation of carcinogenic risk to humans: some industrial chemicals and dyestuffs. Lyon: International Agency for Research on Cancer, IARC, vol 29, 1982.

[IARC2000] IARC. Monographs on the evaluation of carcinogenic risk to humans ionizing radiation, part 1 x- and gamma (y)-radiation, and neutrons. . Lyon: International Agency for Research on Cancer, IARC vol 75, 2000.

[IARC2001] IARC. Monographs on the evaluation of carcinogenic risk to humans: ionizing radiation, part ii: some internally deposited radionuclides. Lyon: International Agency for Research on Cancer, IARC vol 78, 2001.

[IARC2002] IARC. Monograph on the evaluation of carcinogenic risks to humans vol.80 non-ionizing radiation, part 1: static and extremely low frequency (elf) electric and magnetic fields/ IARC working group on the evaluation of carcinogenic risks to humans. IARC press Lyon F (Ed.). , 2002.

[Infante-Rivard1999] Infante-Rivard C, Labuda D, Krajinovic M & Sinnett D. Risk of childhood leukemia associated with exposure to pesticides and with gene polymorphisms.. *Epidemiology* (1999) **10**: pp. 481-487.

[Kaatsch1998] Kaatsch P, Kaletsch U, Meinert R & Michaelis J. An extended study on childhood malignancies in the vicinity of German nuclear power plants. *Cancer Causes Control* (1998) **9**: pp. 529-533.

[Kaatsch2008] Kaatsch P, Spix C, Schulze-Rath R, Schmiedel S & Blettner M. Leukaemia in young children living in the vicinity of German nuclear power plants. *Int J Cancer* (2008) **122**: pp. 721-726.

[Kabuto2006] Kabuto M, Nitta H, Yamamoto S, Yamaguchi N, Akiba S, Honda Y, Hagihara J, Isaka K, Saito T, Ojima T, Nakamura Y, Mizoue T, Ito S, Eboshida A, Yamazaki S, Sokejima S, Kurokawa Y & Kubo O. Childhood leukemia and magnetic fields in Japan: a case-control study of childhood leukemia and residential power-frequency magnetic fields in Japan.. *Int J Cancer* (2006) **119**: pp. 643-650.

[Kheifets2010] Kheifets L, Ahlbom A, Crespi CM, Draper G, Hagihara J, Lowenthal RM, Mezei G, Oksuzyan S, Schüz J, Swanson J, Tittarelli A, Vinceti M & Wunsch Filho V. Pooled analysis of recent studies on magnetic fields and childhood leukaemia.. *Br J Cancer* (2010) **103**: pp. 1128-1135.

[Kinlen1988] Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. *Lancet* (1988) **2**: pp. 1323-1327.

[Kinlen1995] Kinlen LJ, Dickson M & Stiller CA. Childhood leukaemia and non-Hodgkin's lymphoma near large rural construction sites, with a comparison with Sellafield nuclear site. *BMJ* (1995) **310**: pp. 763-768.

[Kinlen1995#310] Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* (1995) **71**: pp. 1-5.

[Kinlen2011] Kinlen L. Childhood leukaemia, nuclear sites, and population mixing. *Br J Cancer* (2011) **104**: pp. 12-18.

[Kleinerman1997] Kleinerman RA, Linet MS, Hatch EE, Wacholder S, Tarone RE, Severson RK, Kaune WT, Friedman DR, Haines CM, Muirhead CR, Boice JDJ & Robison LL. Magnetic field exposure assessment in a case-control study of childhood leukemia.. *Epidemiology* (1997) **8**: pp. 575-583.

[Kleinerman2000] Kleinerman RA, Kaune WT, Hatch EE, Wacholder S, Linet MS, Robison LL, Niwa S & Tarone RE. Are children living near high-voltage power lines at increased risk of acute lymphoblastic leukemia?. *Am J Epidemiol* (2000) **151**: pp. 512-515.

[Knorr-Held1999] Knorr-Held L & Rasser G. Bayesian detection of clusters and discontinuities in disease maps: simulations. (1999) : .

[Knorr-Held2000] Knorr-Held L, Rasser G. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* (2000) **56**: pp. 13-21.

[Knox1997] Knox EG & Gilman EA. Hazard proximities of childhood cancers in Great Britain from 1953-80.. *J Epidemiol Community Health* (1997) **51**: pp. 151-159.

[Kroll2010] Kroll ME, Swanson J, Vincent TJ & Draper GJ. Childhood cancer and magnetic fields from high-voltage power lines in England and Wales: a case-control study. *Br J Cancer* (2010) **103**: pp. 1122-1127.

[Lacour2010] Lacour B, Guyot-Goubin A, Guissou S, Bellec S, Désandes E & Clavel J. Incidence of childhood cancer in France: national children cancer registries, 2000-2004. *Eur J Cancer Prev* (2010) **19**: pp. 173-181.

[Laurier1999] Laurier D & Bard D. Epidemiologic studies of leukemia among persons under 25 years of age living near nuclear sites. *Epidemiol Rev* (1999) **21**: pp. 188-206.

[Laurier2001#127] Laurier D, Valenty M & Tirmarche M. Radon exposure and the risk of leukemia: a review of epidemiological studies.. *Health Phys* (2001) **81**: pp. 272-288.

[Laurier2008] Laurier D, Hémon D & Clavel J. Childhood leukaemia incidence below the age of 5 years near French nuclear power plants. *J Radiol Prot* (2008) **28**: pp. 401-403.

[Laurier2008#142] Laurier D, Jacob S, Bernier MO, Leuraud K, Metz C, Samson E & Laloi P. Epidemiological studies of leukaemia in children and young adults around nuclear facilities: a critical review. *Radiat Prot Dosimetry* (2008) **132**: pp. 182-190.

[Leiss1995] Leiss JK & Savitz DA. Home pesticide use and childhood cancer: a case-control study. *Am J Public Health* (1995) **85**: pp. 249-252.

[Li1998] Li CY, Lee WC & Lin RS. Risk of leukemia in children living near high-voltage transmission lines. *J Occup Environ Med* (1998) **40**: pp. 144-147.

[Lin1994] Lin RS & Lee WC. Risk of childhood leukemia in areas passed by high power lines. *Rev Environ Health* (1994) **10**: pp. 97-103.

[Linet1997] Linet MS, Hatch EE, Kleinerman RA, Robison LL, Kaune WT, Friedman DR, Severson RK, Haines CM, Hartsock CT, Niwa S, Wacholder S & Tarone RE. Residential exposure to magnetic fields and acute lymphoblastic leukemia in children. *N Engl J Med* (1997) **337**: pp. 1-7.

[London1991] London SJ, Thomas DC, Bowman JD, Sobel E, Cheng TC & Peters JM. Exposure to residential electric and magnetic fields and risk of childhood leukemia.. *Am J Epidemiol* (1991) **134**: pp. 923-937.

[Lowengart1987] Lowengart RA, Peters JM, Cicioni C, Buckley J, Bernstein L, Preston-Martin S & Rappaport E. Childhood leukemia and parents' occupational and home exposures. *J Natl Cancer Inst* (1987) **79**: pp. 39-46.

[Lunn2000] Lunn D, Thomas A, Best N & Spiegelhalter D. Winbugs -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* (2000) **10**: pp. 325-337.

[Lyons1995] Lyons RA, Monaghan SP, Heaven M, Littlepage BN, Vincent TJ & Draper GJ. Incidence of leukaemia and lymphoma in young people in the vicinity of the petrochemical plant at Baglan bay, south Wales, 1974 to 1991. *Occup Environ Med* (1995) **52**: pp. 225-228.

[Ma2002] Ma X, Buffler PA, Gunier RB, Dahl G, Smith MT, Reinier K & Reynolds P. Critical windows of exposure to household pesticides and risk of childhood leukemia.. *Environ Health Perspect* (2002) **110**: pp. 955-960.

[Malagoli2010] Malagoli C, Fabbi S, Teggi S, Calzari M, Poli M, Ballotti E, Notari B, Bruni M, Palazzi G, Paolucci P & Vinceti M. Risk of hematological malignancies associated with magnetic fields exposure from power lines: a case-control study in two municipalities of northern Italy. *Environ Health* (2010) **9**: p. 16.

[Mangano2003] Mangano JJ, Sherman J, Chang C, Dave A, Feinberg E & Frimer M. Elevated childhood cancer incidence proximate to U.S. nuclear power plants. *Arch Environ Health* (2003) **58**: pp. 74-82.

[Martin2005] Martin RM, Gunnell D, Owen CG & Smith GD. Breast-feeding and childhood cancer: a systematic review with meta-analysis. *Int J Cancer* (2005) **117**: pp. 1020-1031.

[Maslanyj2008] Maslanyj M, Simpson J, Roman E & Schüz J. Power frequency magnetic fields and risk of childhood leukaemia: misclassification of exposure from the use of the 'distance from power line' exposure surrogate. *Bioelectromagnetics* (2008) : .

[McBride1999] McBride ML, Gallagher RP, Thériault G, Armstrong BG, Tamaro S, Spinelli JJ, Deadman JE, Fincham S, Robson D & Choi W. Power-frequency electric and magnetic fields and risk of childhood leukemia in Canada.. *Am J Epidemiol* (1999) **149**: pp. 831-842.

[McNally2009] McNally RJQ, Alexander FE, Vincent TJ & Murphy MFG. Spatial clustering of childhood cancer in Great Britain during the period 1969-1993. *Int J Cancer* (2009) **124**: pp. 932-936.

[Menegaux2006] Menegaux F, Baruchel A, Bertrand Y, Lescoeur B, Leverger G, Nelken B, Sommelet D, Hémon D & Clavel J. Household exposure to pesticides and risk of childhood acute leukaemia. *Occup Environ Med* (2006) **63**: pp. 131-134.

[Michaelis1997] Michaelis J, Schüz J, Meinert R, Menger M, Grigat JP, Kaatsch P, Kaletsch U, Miesner A, Stamm A, Brinkmann K & Kärner H. Childhood leukemia and electromagnetic fields: results of a population-based case-control study in Germany. *Cancer Causes Control* (1997) **8**: pp. 167-174.

[Michaelis1998] Michaelis J, Schüz J, Meinert R, Zemann E, Grigat JP, Kaatsch P, Kaletsch U, Miesner A, Brinkmann K, Kalkner W & Kärner H. Combined risk estimates for two German population-based case-control studies on residential magnetic fields and childhood acute leukemia. *Epidemiology* (1998) **9**: pp. 92-94.

[Mizoue2004] Mizoue T, Onoe Y, Moritake H, Okamura J, Sokejima S & Nitta H. Residential proximity to high-voltage power lines and risk of childhood hematological malignancies.. *J Epidemiol* (2004) **14**: pp. 118-123.

[Mollié2000] Mollié A. Bayesian mapping of Hodgkin's disease in France. In *Spatial epidemiology, methods and applications*. Elliott, P., Wakefield J., Best, N., Briggs, D. (Ed.). 2000. p. 267.

[Muirhead1995] Muirhead CR. Childhood leukemia in metropolitan regions in the united states: a possible relation to population density?. *Cancer Causes Control* (1995) **6**: pp. 383-388.

[Muirhead1996] Muirhead C & Butland B. Testing for over-dispersion using an adapted form of the Potthoff-Wittinghill method. In *Methods for investigating localized clustering of disease*. Alexander F BP (Ed.). 1996. pp. 40-52.

[Myers1990] Myers A, Clayden AD, Cartwright RA & Cartwright SC. Childhood cancer and overhead powerlines: a case-control study.. *Br J Cancer* (1990) **62**: pp. 1008-1014.

[Nordlinder1997] Nordlinder R & Järvholm B. Environmental exposure to gasoline and leukemia in children and young adults--an ecology study. *Int Arch Occup Environ Health* (1997) **70**: pp. 57-60.

[Olsen1993] Olsen JH, Nielsen A & Schulgen G. Residence near high voltage facilities and risk of cancer in children. *BMJ* (1993) **307**: pp. 891-895.

[Papaemmanuil2009] Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, Kinsey SE, Lightfoot T, Roman E, Irving JAE, Allan JM, Tomlinson IP, Taylor M, Greaves M & Houlston RS. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia.. *Nat Genet* (2009) **41**: pp. 1006-1010.

[Parkin1998] Parkin, D.M, Kramarova, E., Draper, G.J., Masuyer, E., Michaelis, J., Neglia, J, Qureshi, S. & Stiller, CA.. International incidence of childhood cancer, vol ii. IARC Scientific Publications No 144, Lyon, IARC (Ed.). , 1998.

[Pearson2000] Pearson RL, Wachtel H & Ebi KL. Distance-weighted traffic density in proximity to a home is a risk factor for leukemia and other childhood cancers.. *J Air Waste Manag Assoc* (2000) **50**: pp. 175-180.

[Petridou1993] Petridou E, Kassimos D, Kalmanti M, Kosmidis H, Haidas S, Flytzani V, Tong D & Trichopoulos D. Age of exposure to infections and risk of childhood leukaemia. *BMJ* (1993) **307**: p. 774.

[Petridou1997] Petridou E, Alexander FE, Trichopoulos D, Revinthi K, Dessypris N, Wray N, Haidas S, Koliouskas D, Kosmidis H, Piperopoulou F & Tzortzatou F. Aggregation of childhood leukemia in geographic areas of Greece. *Cancer Causes Control* (1997) **8**: pp. 239-245.

[Petridou1997#140] Petridou E, Trichopoulos D, Kravaritis A, Pourtsidis A, Dessypris N, Skalkidis Y, Kogevinas M, Kalmanti M, Koliouskas D, Kosmidis H, Panagiotou JP, Piperopoulou F, Tzortzatou F & Kalapothaki V. Electrical power lines and childhood leukemia: a study from Greece. *Int J Cancer* (1997) **73**: pp. 345-348.

[Potthoff1966] Potthoff R & Wittinghill M. Testing for homogeneity. ii. the Poisson distribution. *Biometrika* (1966) **53**: pp. 183-190.

[Potthoff1996] Potthoff R & Wittinghill M. Testing for homogeneity. i. the binomial and multinomial distributions. *Biometrika* (1966) **53**: pp. 167-182.

[Raaschou-Nielsen2008] Raaschou-Nielsen O, Andersen CE, Andersen HP, Gravesen P, Lind M, Schüz J & Ulbak K. Domestic radon and childhood cancer in Denmark.. *Epidemiology* (2008) **19**: pp. 536-543.

[Raaschou-Nielsen2008#203] Raaschou-Nielsen O. Indoor radon and childhood leukaemia. *Radiat Prot Dosimetry* (2008) **132**: pp. 175-181.

[Richardson2004] Richardson S, Thomson A, Best N & Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* (2004) **112**: pp. 1016-1025.

[Rudant2006] Rudant J, Baccaïni B, Ripert M, Goubin A, Bellec S, Hémon D & Clavel J. Population-mixing at the place of residence at the time of birth and incidence of childhood leukaemia in France. *Eur J Cancer* (2006) **42**: pp. 927-933.

[Rudant2007] Rudant J, Menegaux F, Leverger G, Baruchel A, Nelken B, Bertrand Y, Patte C, Pacquement H, Vérité C, Robert A, Michel G, Margueritte G, Gandemer V, Hémon D & Clavel J. Household exposure to pesticides and risk of childhood hematopoietic malignancies: the ESCALE study (SFCE). *Environ Health Perspect* (2007) **115**: pp. 1787-1793.

[Rudant2010] Rudant J, Orsi L, Menegaux F, Petit A, Baruchel A, Bertrand Y, Lambilliotte A, Robert A, Michel G, Margueritte G, Tandonnet J, Mechinaud F, Bordigoni P, Hémon D & Clavel J. Childhood acute leukemia, early common infections, and allergy: the ESCALE study. *Am J Epidemiol* (2010).

[Sans1995] Sans S, Elliott P, Kleinschmidt I, Shaddick G, Pattenden S, Walls P, Grundy C & Dolk H. Cancer incidence and mortality near the Baglan bay petrochemical works, south Wales.. *Occup Environ Med* (1995) **52**: pp. 217-224.

[Savitz1988] Savitz DA, Wachtel H, Barnes FA, John EM & Tvrdik JG. Case-control study of childhood cancer and exposure to 60-hz magnetic fields. *Am J Epidemiol* (1988) **128**: pp. 21-38.

[Savitz1989] Savitz DA & Feingold L. Association of childhood cancer with residential traffic density. *Scand J Work Environ Health* (1989) **15**: pp. 360-363.

[Savitz1993] Savitz DA & Kaune WT. Childhood cancer in relation to a modified residential wire code. *Environ Health Perspect* (1993) **101**: pp. 76-80.

[Schüz2000] Schüz J, Grigat JP, Störmer B, Rippin G, Brinkmann K & Michaelis J. Extremely low frequency magnetic fields in residences in Germany. distribution of measurements, comparison of two methods for assessing exposure, and predictors for the occurrence of magnetic fields above background level. *Radiat Environ Biophys* (2000) **39**: pp. 233-240.

[Schüz2001#251] Schüz J, Grigat JP, Brinkmann K & Michaelis J. Residential magnetic fields as a risk factor for childhood acute leukaemia: results from a German population-based case-control study. *Int J Cancer* (2001) **91**: pp. 728-735.

[Sharp1996] Sharp L, Black RJ, Harkness EF & McKinney PA. Incidence of childhood leukaemia and non-Hodgkin's lymphoma in the vicinity of nuclear sites in Scotland, 1968-93. *Occup Environ Med* (1996) **53**: pp. 823-831.

[Sommelet2009] Sommelet, D., Clavel, J., Lacour, B. Epidémiologie des cancers de l'enfant. Springer (Ed.). Broché, 2009.

[Spiegelhalter2002] Spiegelhalter D, Best N, Carlin B & van der Linde A. Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B* (2002) **64**: pp. 583-639.

[Spix2008] Spix C, Schmiedel S, Kaatsch P, Schulze-Rath R & Blettner M. Case-control study on childhood cancer in the vicinity of nuclear power plants in Germany 1980-2003. *Eur J Cancer* (2008) **44**: pp. 275-284.

[Spycher2011] Spycher BD, Feller M, Zwahlen M, Röösli M, von der Weid NX, Hengartner H, Egger M, Kuehni CE, . Childhood cancer and nuclear power plants in Switzerland: a census-based cohort study. *Int J Epidemiol* (2011) : .

[Steffen2004] Steffen C, Auclerc MF, Auvrignon A, Baruchel A, Kebaili K, Lambilliotte A, Leverger G, Sommelet D, Vilmer E, Hémon D & Clavel J. Acute childhood leukaemia and environmental exposure to potential sources of benzene and other hydrocarbons; a case-control study. *Occup Environ Med* (2004) **61**: pp. 773-778.

[Stern2000] Stern HS & Cressie N. Posterior predictive model checks for disease mapping models. *Stat Med* (2000) **19**: pp. 2377-2397.

[Stiller2006] Stiller CA, Marcos-Gragera R, Ardanaz E, Pannelli F, Almar Marqués E, Cañada Martinez A & Steliarova-Foucher E. Geographical patterns of childhood cancer incidence in Europe, 1988-1997. report from the automated childhood cancer information system project. *Eur J Cancer* (2006) **42**: pp. 1952-1960.

[Tomenius1986] Tomenius L. 50-hz electromagnetic environment and the incidence of childhood tumors in Stockholm county. *Bioelectromagnetics* (1986) **7**: pp. 191-207.

[Treviño2009] Treviño LR, Shimasaki N, Yang W, Panetta JC, Cheng C, Pei D, Chan D, Sparreboom A, Giacomini KM, Pui C, Evans WE & Relling MV. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol* (2009) **27**: pp. 5972-5978.

[Tynes1997] Tynes T & Haldorsen T. Electromagnetic fields and cancer in children residing near Norwegian high-voltage power lines. *Am J Epidemiol* (1997) **145**: pp. 219-226.

[UKCCS1999] UKCCS. Exposure to power-frequency magnetic fields and the risk of childhood cancer. UK childhood cancer study investigators. *Lancet* (1999) **354**: pp. 1925-1931.

[UKCCS2000] UKCCS. Childhood cancer and residential proximity to power lines. UK childhood cancer study investigators. *Br J Cancer* (2000) **83**: pp. 1573-1580.

[UNSCEAR2000] United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). Exposure from natural radiation sources. in: UNSCEAR: sources and effects of ionizing radiation. (2000) Volume I: Sources, Annex B. New-York: United Nations, 133: .

[Urayama2010] Urayama KY, Buffler PA, Gallagher ER, Ayoob JM & Ma X. A meta-analysis of the association between day-care attendance and childhood acute lymphoblastic leukaemia. *Int J Epidemiol* (2010) **39**: pp. 718-732.

[Urayama2011] Urayama KY, Ma X, Selvin S, Metayer C, Chokkalingam AP, Wiemels JL, Does M, Chang J, Wong A, Trachtenberg E & Buffler PA. Early life exposure to infections and risk of childhood acute lymphoblastic leukemia. *Int J Cancer* (2011) **128**: pp. 1632-1643.

[Verkasalo1993] Verkasalo PK, Pukkala E, Hongisto MY, Valjus JE, Järvinen PJ, Heikkilä KV & Koskenvuo M. Risk of cancer in Finnish children living close to power lines. *BMJ* (1993) **307**: pp. 895-899.

[Viel1993] Viel JF, Richardson S, Danel P, Boutard P, Malet M, Barrelier P, Reman O & Carré A. Childhood leukemia incidence in the vicinity of La Hague nuclear-waste reprocessing facility (France). *Cancer Causes Control* (1993) **4**: pp. 341-343.

[Visser2004] Visser O, van Wijnen JH & van Leeuwen FE. Residential traffic density and cancer incidence in Amsterdam, 1989-1997. *Cancer Causes Control* (2004) **15**: pp. 331-339.

[Vistnes1997] Vistnes AI, Ramberg GB, Bjørnevik LR, Tynes T & Haldorsen T. Exposure of children to residential magnetic fields in Norway: is proximity to power lines an adequate predictor of exposure?. *Bioelectromagnetics* (1997) **18**: pp. 47-57.

[Wakefield2000] Wakefield J, Best N & Waller L. Bayesian approaches to disease mapping. In *Spatial epidemiology. methods and applications.*. Elliott P, Wakefield J, Best N, Briggs D (Eds.). 2000. pp. 104-127.

[Wakeford2009] Wakeford R, Kendall GM & Little MP. The proportion of childhood leukaemia incidence in Great Britain that may be caused by natural background ionizing radiation.. *Leukemia* (2009) **23**: pp. 770-776.

[Wertheimer1979] Wertheimer N & Leeper E. Electrical wiring configurations and childhood cancer. *Am J Epidemiol* (1979) **109**: pp. 273-284.

[White-Koning2004] White-Koning ML, Hémon D, Laurier D, Tirmarche M, Jougla E, Goubin A & Clavel J. Incidence of childhood leukaemia in the vicinity of nuclear sites in France, 1990-1998. *Br J Cancer* (2004) **91**: pp. 916-922.

[Wilkinson1999] Wilkinson P, Thakrar B, Walls P, Landon M, Falconer S, Grundy C & Elliott P. Lymphohaematopoietic malignancy around all industrial complexes that include major oil refineries in Great Britain. *Occup Environ Med* (1999) **56**: pp. 577-580.

[Williams2001] Williams JR, Alexander FE, Cartwright RA & McNally RJQ. Methods for eliciting aetiological clues from geographically clustered cases of disease, with application to leukaemia-lymphoma data. *J.R.Statist.Soc.A.* (2001) **164**: pp. 49-60.

IX. Articles

Geographical variations in the incidence of childhood acute leukaemia in France over the period 1990–2004

Claire Faure[a,b,c,*], Annie Mollié[d,*], Stéphanie Bellec[a,b,c], Aurélie Guyot-Goubin[a,b,c], Jacqueline Clavel[a,b,c] and Denis Hémon[a,b]

Spatial variations in childhood acute leukaemia (AL) incidence rates were investigated by département, in mainland France, over the period 1990–2004. This is the first spatial study of this incidence to cover a 15-year period. French National Registry of Childhood Haematological Malignancies data and population counts by type of leukaemia (AL, acute lymphoblastic leukaemia, acute myeloblastic leukaemia), time period (1990–2004, 1990–1994, 1995–1999 and 2000–2004), sex, and age group (0–14, 0–4, 5–9 and 10–14 years of age) were considered. The overall homogeneity of the relative risks of leukaemia was tested, as well as comparison to 1 of each relative risk by the exact Poisson test. To give a more stable estimate of the underlying relative risk pattern than that provided by the local standardized incidence ratios (SIRs), Bayesian hierarchical models using four different spatial priors have been produced: the parametric BYM and CAR models, and two semiparametric models. Very slight overall heterogeneity was observed on the whole AL data set (SIR overdispersion 18.6%, $P=0.10$). Irrespective of the model, the ranges of the smoothed SIRs exhibited considerable shrinkage relative to the ranges of the local SIRs. The associated maps were slightly heterogeneous; the smoothed SIRs of overall acute lymphoblastic leukaemia of the south-west départements were slightly higher than those of the north-east. The results, however, did not remain stable when investigated by leukaemia type, time period, sex or age group. No spatial heterogeneity of childhood AL incidence on the département scale was observed but that does not exclude spatial heterogeneity on other scales. *European Journal of Cancer Prevention* **18:267–279 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

European Journal of Cancer Prevention 2009, 18:267–279

Keywords: Bayesian hierarchical models, childhood acute leukaemia, France, incidence, mapping, spatial models

[a]Institut National de la Santé et de la Recherche Médicale (INSERM), [b]Université Paris-Sud, [c]French National Registry of Childhood Haematological malignancies, Villejuif and [d]Université Paris-Ouest Nanterre La Défense, Nanterre, France

Correspondence to Jacqueline Clavel, MD, PhD, U754-INSERM, 16 Avenue Paul Vaillant-Couturier, Villejuif Cedex 94807, France
Tel: +33 145 59 50 38; fax: +33 145 59 51 51;
e-mail: jacqueline.clavel@inserm.fr

*Claire Faure and Annie Mollié are equally contributed to this study

Received 15 September 2008 Accepted 23 February 2009

Introduction

The annual incidence rate of childhood leukaemia (CL) varies significantly between European regions, from 39.1 cases per million in the east to 48.0 in the north, and is on average 44.0 cases per million (Stiller *et al.*, 2006). In France, the annual incidence rate of CL is 43.1 cases per million. Recently, a slight heterogeneity on the commune scale (the smallest administrative unit in France: 36 600 communes for 60 million inhabitants), over the period 1990–1994, was reported for acute leukaemia (AL), particularly for children aged 0–4 years (Bellec *et al.*, 2006). This observation was similar to that of the EUROCLUS study (Alexander *et al.*, 1998).

This study investigates spatial variations of AL incidence rates, by département, in mainland France over the period 1990–2004 (metropolitan France is administratively divided into 96 départements, the population of which ranges from 73 818 to 2 572 224 inhabitants and has 622 487 inhabitants on average in 1999). As the different cytological leukaemia types may suggest different risk factors, acute lymphoblastic leukaemia (ALL) and acute myeloblastic leukaemia (AML) have been investigated separately. If the incidence of AL in France varies on the département scale, it may be important to take this into account in the search for AL risk factors. Another interest of that study is that it can involve the search for corresponding suspected CL risk factors varying on the same scale, such as health behaviour (nutrition, consumption of tobacco and alcohol, infections), urban for residence or environmental characteristics such as terrestrial radon. Such suspected environmental risk factors led us to study the stability of AL spatial heterogeneity over time, sex and age.

The models used to estimate the relative risk of childhood AL incidence in each département were most relevant to the study of small observed counts of non-contagious diseases and aimed to take into account potential local spatial dependence between risks, because of the influence of unmeasured or unknown risk factors. The models involved the combination of an underlying spatial process of unknown 'true' incidence rates and a Poisson variability of observed incidence counts

0959-8278 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins

DOI: 10.1097/CEJ.0b013e32832bf43a

formulated within the Bayesian hierarchical framework (Wakefield *et al.*, 2000).

Among the variety of spatial models that have been proposed for the latent (second) level of the hierarchy (Best *et al.*, 2005), two parametric spatial priors based on Gaussian Markov random fields, the BYM (Besag *et al.*, 1991) and CAR models (Stern and Cressie, 2000), were selected, the models most commonly used in disease mapping. Such parametric models, however, result in an overall shrinkage of the risks that may oversmooth the relative risk surface when it is not continuous. In order to allow for potential local discontinuities in the risk surface (e.g. hotspots) and make fewer distributional assumptions, two semiparametric spatial models were also fitted: a spatial mixture model for the risks referred to as the MIX model (Green and Richardson, 2002), and a spatial partition model proposed by Knorr-Held and Rasser (Knoor-Held and Rasser, 2000) henceforth referred to as the KHR model. The main feature of these models is that they provide some shrinkage and spatial smoothing of the local relative risk estimates, which give a more stable estimation of the underlying relative risk pattern than that provided by the local estimates.

Materials and methods

The study concerned the 94 départements of mainland France. Corsica, a French Mediterranean island divided into two départements, was excluded from this study because the spatial models used did not include the particular case of isolated départements without neighbour.

Cases

The French National Registry of Childhood Haematological malignancies (NRCH) includes among others all the cases of leukaemia, since the beginning of 1990, that occurred in children below 15 years of age and living in metropolitan France at the time of diagnosis. The exhaustiveness of the registry is about 99% for leukaemia (Clavel *et al.*, 2004). The data used for this study were the 6796 cases of AL registered in the NRCH diagnosed between 1 January 1990 and 31 December 2004: 81% (5526) consisted of ALL, 11% (1160) of AML and 8% (110) of other AL. The incidence of AL in France computed from those data was 40.7 cases per million person–years. The sex ratio was 1.18, which is close to the sex ratios reported in other industrialized countries (Stiller *et al.*, 2006).

During the period 1990–2004, the observed distributions of the incidence of AL by leukaemia type, sex and age were similar to those described earlier over the period 1990–1999 (Clavel *et al.*, 2004) and were comparable with those of neighbouring countries (Alexander *et al.*, 1998).

Population

The 1990 and 1999 French population censuses conducted by the French Institute of Statistics and Economic Studies were used to compute person–years by département. The French Institute of Statistics and Economic Studies has also reported interpolations for the years between censuses and extrapolations for the years after 1999. Person–years were computed by sex and 1-year age group as the mean population count on 1st January of the relevant year and the next year. Population counts by time period (1990–2004, 1990–1994, 1995–1999 and 2000–2004), sex, and age group [0–14, 0–4, 5–9 and 10–14 years old (y.o.)] were also considered.

Standardized incidence ratios: local relative risk estimates

The basic inference of the relative risks for rare events was based on a Poisson model for the observed number of incident cases y_i in département i: $y_i \sim \text{Poisson}(E_i\, \theta_i)$ for $i = 1,\ldots n = 94$, independently for the n départements, where E_i denotes the expected number of incident cases (adjusted for age and sex when appropriate) and θ_i the 'incidence relative risk' in département i, assuming a multiplicative effect between age × sex and département on the incidence rate. Under the Poisson model, θ_i is estimated by the 'local' observed standardized incidence ratio (SIR), $SIR_i = y_i/E_i$.

Testing for the overall homogeneity of the relative risks

The overall homogeneity of the relative risks (i.e. $\theta_i = 1$ for $i = 1,\ldots n$) was tested in three ways.

First, the Poisson log-likelihood ratio statistic L and the Pearson's chi-square statistic were used, both asymptotically distributed as a χ^2 with $n-1$ degrees of freedom under the null hypothesis of homogeneity.

Investigation for a Poisson overdispersion was then conducted based on the assumption of a negative binomial distribution of the observed number of cases independently in each département, with a variance to the mean ratio equal to $1+\beta$; $\beta = 0$ indicates the absence of overdispersion, while $\beta > 0$ indicates a variance greater than the mean. Under the null hypothesis of homogeneity, the score statistic of Potthoff and Wittinghill (Potthoff and Wittinghill, 1996a, 1996b) has been proved to be asymptotically χ^2 distributed and more powerful than the Pearson's χ^2 statistic with the above form of overdispersion. Moreover, an estimate of the heterogeneity coefficient β is available (Muirhead and Butland, 1996).

Standardized incidence ratio significance tests

Exact Poisson significance tests comparing each relative risk to 1 (Estève *et al.*, 1993) and the corresponding 90% relative risk confidence interval ($CI_{90\%}$) were performed independently for each département. To allow for multiple

testing, the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) was used to control the 'false discovery rate'.

Smoothed standardized incidence ratios: Bayesian relative risk estimates

Bayesian disease mapping models

The three-level hierarchical formulation usually used is a typical model for disease mapping based on aggregated data for rare events, considering the Poisson model for the observed number of cases at the lowest level of the hierarchy.

At the second stage, the overall spatial relative risk variability is described by a spatial model defined for the set of n relative risks $\boldsymbol{\theta} = (\theta_i, i = 1,\ldots n)$ modelled as random variables, using both parametric BYM and CAR models and semiparametric MIX and KHR models. The spatial neighbourhood structure needs to be prescribed: two French départements were considered first-order neighbours if they were adjacent.

The top (third) level specifies the distributions of hyperparameters, that is, the parameters considered as random variables at the second stage model.

The BYM model: the BYM model (Besag *et al.*, 1991) specifies the log relative risk as the sum of two independent components expressed as:

$$\log(\theta_i) = x_i = v_i + u_i \text{ for } i = 1,\ldots n$$

where $\boldsymbol{v} = (v_i\ (i = 1,\ldots n)$ is a Gaussian noise with null expectation and unknown variance parameter σ_v^2, describing unstructured heterogeneity and $\boldsymbol{u} = (u_i\ (i = 1,\ldots n)$ is an intrinsic auto-regressive Gaussian process with unknown variance parameter σ_u^2, representing the local spatially structured variation between neighbouring départements. In a fully Bayesian perspective, both inverse variances are given specified Γ-hyperprior distributions.

The relative risk estimate obtained by this model is shrunk towards either a local mean depending on the neighbouring départements or the overall mean, depending on whether there is spatial dependency or not.

The CAR model: the CAR model was adapted (Stern and Cressie, 2000) from the Gaussian conditional autoregressive models previously developed for regular lattices with constant conditional variance, to irregular sets of areas specifying the conditional variance to be inversely proportional to E_i. A CAR model for $\boldsymbol{x} = \log(\boldsymbol{\theta})$ defined by the following three unknown parameters was used: the overall mean μ assumed to be constant across départements, the autocorrelation parameter φ, which reflects the overall strength of spatial dependence between neighbouring départements, and the variance parameter σ^2.

The autocorrelation parameter is constrained to lie in a range determined by the neighbourhood structure (Cressie, 1993) that is (−0.3365, 0.1717) for the French départements. Vague uniform hyperprior distributions on the whole real line and on]−0.3365, 0.1717[were chosen for μ and φ, respectively, and a Γ-distribution for the inverse variance.

The MIX model: the MIX model (Green and Richardson, 2002) is a mixture model assuming that the relative risk $\theta_i = \theta'_{z_i}$, where $(\theta'_j, j = 1,\ldots k)$ characterize k different risk classes, and that $(z_i, i = 1,\ldots n)$ are allocation variables taking values of 1, 2,…k assigning each département to a risk class. The number of risk classes k is unknown. Given k, the allocation variables follow a spatial process, the Potts model, which involves an unknown positive interaction parameter ψ that controls the strength of the spatial dependence of the allocations. Thus, the probability of allocating a département to a risk class increases with both the value of ψ and the number of neighbouring départements already allocated to that class.

Prior assumptions are independent γ-distributions for the risk classes $(\theta'_j, j = 1,\ldots k)$, a discrete uniform distribution on $\{0, 0.1, 0.2,\ldots,\psi_{max}\}$ for the interaction parameter ψ, where $\psi_{max} = 1.0$ is a very high level of interaction for the neighbourhood structure of the French départements, and a discrete uniform distribution on $\{1, 2, \ldots, k_{max}\}$ for the number of risk classes k where $k_{max} = 10$ seems adequately large for the prior maximum number of risk classes for relative risks of CL in the French départements.

The KHR model: the KHR model (Knorr-Held and Rasser, 2000) is a spatial partition model which assumes that the n départements can be combined in a set of k ($k \leq n$) nonoverlapping clusters (subsets of neighbouring départements), each with a constant relative risk, k, being unknown.

Thus the log-relative risks x are assumed to be independent and identically distributed Gaussian (μ, σ^2) random variables with $x_i = \log(\theta'_{z_i})$, where $(\theta'_j, j = 1,\ldots k)$ represent k different cluster risk levels, and $(z_i, i = 1,\ldots n)$ are the allocation variables assigning each département to a cluster (from 1 to k).

A vague uniform prior distribution on the whole real line for μ, a truncated geometric prior distribution with parameter 0.02 (corresponding to a prior mean of 50 clusters, i.e. approximately half the number of départements) for the number of clusters k, and a Γ-prior distribution for the inverse variance were chosen.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Hyperprior distribution choices

The overall amount of smoothing depends on the hyperprior distribution choices. The effects of those choices are not directly comparable because the parameters interact at different stages of the hierarchical Bayesian models.

The forms of those distributions, which have previously been described (*cf.* Bayesian disease mapping models) for each spatial prior, will not be addressed further. Instead, this section will focus on the choice of hyperprior Γ-distribution parameters (shape a and inverse scale b) enabling, *a priori*, similar degrees of smoothing for the four prior spatial models.

It was considered that the 95% fluctuation interval of a given *département*-specific relative risk of CL could reasonably be considered to range approximately from 0.8 to 1.25, on average, and have a wider range than 0.25–4.0 in approximately 5% of cases.

For the MIX model, a simple choice satisfying the above conditions was $a = b = 100$.

Table 1 Hyperprior distribution choices and the inverse variance of the CAR model

		Γ(1,b)[b]		Prior 95% fluctuation interval $I_{95\%}$	
	hm(E)[a]	b	Mean	Mean range[c]	95% range[c]
AL	39.5	0.5	2	0.80–1.25	0.25–4.00
ALL	32.1			0.78–1.28	0.21–4.65
AL 1990–1994	13.1	0.2	5	0.78–1.27	0.22–4.58
AL 1995–1999	12.9			0.78–1.28	0.22–4.64
AL 2000–2004	13.5			0.79–1.27	0.22–4.48
AL boys	21.9			0.83–1.21	0.31–3.25
AL girls	17.6			0.81–1.23	0.27–3.72
AL 0–4 y.o.	19.3			0.82–1.22	0.29–3.50
AL 5–9 y.o.	11.8			0.77–1.29	0.20–4.97
ALL 1990–1994	10.7			0.76–1.31	0.19–5.39
ALL 1995–1999	10.5			0.76–1.31	0.18–5.47
ALL 2000–2004	10.9			0.77–1.30	0.19–5.30
ALL boys	18.1			0.81–1.23	0.27–3.65
ALL girls	14.0			0.79–1.26	0.23–4.36
ALL 0–4 y.o.	16.0			0.80–1.24	0.25–3.96
AL 10–14 y.o.	8.4	0.1	10	0.81–1.24	0.26–3.83
ALL 5–9 y.o.	9.9			0.82–1.22	0.29–3.45
ALL 10–14 y.o.	6.2			0.76–1.28	0.21–4.78
AML	6.7			0.79–1.27	0.22–4.50
AML 1990–1994	2.2	0.04	25	0.77–1.30	0.19–5.26
AML 1995–1999	2.2			0.77–1.30	0.19–5.26
AML 2000–2004	2.3			0.77–1.29	0.20–5.08
AML boys	3.4			0.81–1.24	0.26–3.80
AML girls	3.3			0.81–1.24	0.26–3.88

AL, acute leukaemia; ALL, acute lymphoblastic leukaemia; AML, acute myeloblastic leukaemia; SIR, standardized incidence ratio; y.o., years old.
[a]$hm(E)$ harmonic mean of the expected numbers of cases.
[b]Γ(1,b) Γ-distribution with shape parameter 1 and inverse scale parameter b (standard deviation = mean).
[c]The mean range and 95% range are the *a priori* 95% relative risk fluctuation intervals based on the mean and the 97.5% quantile of the γ-prior, respectively.

For each of the BYM, CAR and KHR models, a high degree of smoothing is related to either a high mean or a low variance. Thus, $a = 1$ was selected to obtain standard deviation equal to the mean. For the choice of b, two approximate *a priori* 95% relative risk fluctuation intervals were used: one based on the mean of the Γ-prior, in order to obtain the prior 'mean range'; the other derived from the 97.5% quantile of the Γ-prior giving the prior '95% range'. For the BYM and KHR models, this led us to $b = 0.01$, whereas for the CAR model, the choice of prior (Table 1) depended on the number of expected cases in the considered population.

Bayesian relative risk estimates and model-comparison criteria

The estimation techniques for the hierarchical Bayesian models use Markov chain Monte Carlo (MCMC) algorithms that simulate from the joint posterior distribution of parameters and relative risks. Thus the département-specific relative risk was estimated from the posterior mean of the simulated risks for the département and the 90% credible interval of the relative risk ($CrI_{90\%}$). Besides, the posterior probability that the relative risk exceeds 1 was considered.

Deviance information criteria (DIC) have been reported to enable comparison of the complexity and fit of the four models (Spiegelhalter *et al.*, 2002). The DIC is the sum of the posterior mean deviance $\overline{D}$, which reflects model fit, and an estimate of the 'effective' number of parameters p_D, which may be viewed as a penalty term reflecting the complexity of the model.

The BYM and CAR models were implemented using WinBUGS (Lunn *et al.*, 2000). A single MCMC chain of 30 000 iterations was run to obtain an estimate of the parameters on 3000 iterations: one out of 10 was taken after a burn-in of 1000–3000 iterations.

Reversible jump MCMC implementation (Green, 1995) was required to fit both the MIX and KHR models. The Fortran and C codes for the MIX model were kindly supplied by their authors. The Bayesian Detection of Clusters and Discontinuities in Disease Maps program is available at *http://www.stat.uni-muenchen.de/sfb386/software/bdcd/* for the KHR model. The results for the MIX model were generated by runs of 500 000 iterations of a single chain after a burn-in of 20 000 iterations (with satisfactory acceptance rates for the split and merge moves, about 10–14%). A single chain of 5 000 000 iterations was run for the KHR model after a burn-in of 100 000 iterations (using 2500 iterations to estimate parameters, taking one of 2000 iterations).

Associations between smoothed standardized incidence ratios

As the sets of smoothed relative risk estimates were spatially autocorrelated, significance tests were conducted on the correlation coefficients between two sets (between time periods, sexes or age groups) using the modified *t*-test (Clifford *et al.*, 1989). This test enabled correction of the d.f. of the classic *t*-distribution by considering the spatial autocorrelations of each set of smoothed estimates. The modified d.f. were calculated considering the estimated spatial autocorrelations of order 1–5 (Cliff and Ord, 1981) from the first-order to fifth-order neighbourhood structures of the départements.

Results

Twenty-four maps depicting the spatial variations in the incidence relative risks for childhood AL, ALL and AML overall, and by time period (1990–1994, 1995–1999 and 2000–2004), sex, and age group (0–4, 5–9 and 10–14 y.o., except for AML) were estimated using each of the following five methods: the SIRs, and smoothed SIRs derived using the BYM, CAR, MIX and KHR models, respectively. Although all the results were analyzed, only an illustrative selection of maps and tables will be addressed herein.

As Corsica was excluded from the study and with a view to verification, we checked that the SIRs for the mainland départements remained unchanged when Corsica was taken into account. Moreover, the SIR for Corsica was 1.1, not significantly different from 1.

Standardized incidence ratios: local relative risk estimates

The three overall homogeneity tests yielded similar nonsignificant results. An estimated Poisson overdispersion of approximately 20–30% was observed in accordance with an overall heterogeneity on the borderline of significance (*P* values from 0.04 to 0.10) but no consistency on subpopulations (by leukaemia type, time period, sex or age group) was observed (Table 2).

For overall AL (1990–2004, boys and girls and 0–14 y.o.), an overdispersion of 18.6% ($P = 0.10$) was observed, 22.2% ($P = 0.07$) for the period 1995–1999 and 24.3%

Table 2 SIR heterogeneity coefficient

	$\hat{\beta}$[a]	*P* value[b]	nb cases[c]	*hm*(*E*)[d]	Range(*E*)[d]	Range (SIR)[e]
Acute leukaemia						
AL	**0.186**	0.097	6796	39.5	7.27–340.55	0.52–1.54
AL 1990–1994	0.054	0.329	2235	13.1	2.41–115.55	0.36–2.04
AL 1995–1999	**0.222**	0.069	2218	12.9	2.36–110.98	0.00–2.03
AL 2000–2004	0.159	0.130	2343	13.5	2.49–113.80	0.45–1.81
AL boys	0.039	0.350	3766	21.9	4.05–188.91	0.17–1.52
AL girls	0.040	0.375	3030	17.6	3.23–151.64	0.34–2.08
AL 0–4 y.o.	0.110	0.209	3408	19.3	3.51–169.52	0.49–1.92
AL 5–9 y.o.	0.127	0.175	1999	11.8	2.16–100.53	0.32–1.95
AL 10–14 y.o.	**0.243**	0.056	1389	8.4	1.60–70.50	0.00–2.07
Acute lymphoblastic leukaemia						
ALL	0.235	0.060	5526	32.1	5.92–276.87	0.47–1.54
ALL 1990–1994	−0.034	0.571	1828	10.7	1.97–94.58	0.29–1.92
ALL 1995–1999	0.138	0.155	1805	10.5	1.92–90.25	0.00–2.25
ALL 2000–2004	**0.199**	0.091	1893	10.9	2.02–91.89	0.33–1.86
ALL boys	0.009	0.451	3113	18.1	3.35–156.15	0.00–1.48
ALL girls	0.131	0.166	2413	14.0	2.57–120.72	0.27–2.24
ALL 0–4 y.o.	−0.009	0.486	2808	16.0	2.91–139.71	0.32–1.98
ALL 5–9 y.o.	**0.202**	0.084	1690	9.9	1.83–84.99	0.13–2.05
ALL 10–14 y.o.	0.061	0.253	1028	6.2	1.18–52.17	0.00–2.10
Acute myeloblastic leukaemia						
AML	0.062	0.297	1160	6.7	1.24–58.17	0.00–2.59
AML 1990–1994	**0.279**	0.035	377	2.2	0.41–19.41	0.00–5.41
AML 1995–1999	−0.002	0.463	384	2.2	0.41–19.28	0.00–3.28
AML 2000–2004	−0.017	0.518	399	2.3	0.42–19.43	0.00–3.29
AML boys	0.107	0.222	586	3.4	0.63–29.40	0.00–2.83
AML girls	−0.042	0.570	574	3.3	0.61–28.77	0.00–2.97

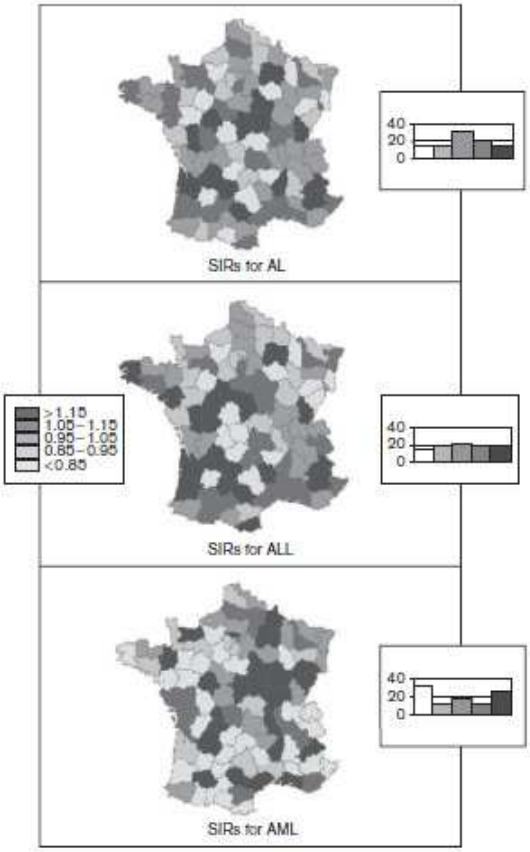
$n = 94$, the number of départements; y_i observed number of incident cases and E_i expected number of incident cases in département i.
AL, acute leukaemia; ALL, acute lymphoblastic leukaemia; AML, acute myeloblastic leukaemia; SIR, standardized incidence ratio; y.o., years old.

The heterogeneity coefficient β is defined as $1 + \beta = \frac{\mathrm{var}(y_i)}{E(y_i)}$ for $i = 1, \ldots n$.

[a] $\hat{\beta} = \frac{y_+ \left[\sum_i \left(\frac{y_i (y_i - 1)}{E_i} \right) - y_+ + 1 \right]}{(n-1)(y_+ - 1)}$ with $y_+ = \sum_i y_i$

[b] *P* value of Potthoff and Whittinghill test of $\beta = 0$.
[c] Total number of cases y_+.
[d] *hm*(*E*), harmonic mean; *range* (*E*) range of the expected numbers of cases.
[e] Range (SIR) with $(SIR)_i = y_i / E_i$.

Fig. 1



Standardized incidence ratio (SIR) maps for acute leukaemia (AL), acute lymphoblastic leukaemia (ALL) and acute myeloblastic leukaemia (AML).

($P = 0.06$) for the 10–14 y.o. age group. ALL showed an estimated overdispersion of 23.5% ($P = 0.06$) overall, 19.9% ($P = 0.09$) for the period 2000–2004 and 20.2% ($P = 0.08$) for the 5–9 y.o. age group. For AML, only the first period showed a significant estimated overdispersion of 27.9% ($P = 0.04$).

Although the SIR ranges were rather wide when based on the smallest counts (Table 2), the SIR maps did not show any spatial pattern in the relative risks for AL, ALL or AML overall (Fig. 1), or by subpopulation.

For overall AL (Table 3), six SIRs were significantly greater than 1 (départements No. 13, 24, 33, 37, 39 and 81) and seven SIRs were significantly less than 1 (No. 12, 19, 27, 47, 68, 72 and 93). However, no consistency in the significant local excesses of relative risks by subpopulation was observed. Moreover, when allowing for multiple testing, no local excess remained significant. No significant SIR correlation among time periods, sexes or age groups was observed for AL, ALL or AML overall.

Smoothed standardized incidence ratios

Bayesian hierarchical model complexity and fit

None of the four models appeared 'superior' to the others over the 24 sets of smoothed SIRs (Table 4). The smallest DIC values were mainly observed with the semiparametric KHR and MIX models (nine and eight times out of 24, respectively).

The DIC was always less than 188, its approximate value for the 'saturated' model based on Poisson likelihood and was mainly smaller than L, its approximate value for the 'pooled' model with all relative risks was equal to 1; the improvement ranged from 0.5 to 8.2. However, the DIC was found to be slightly (from 0.2 to 3.3) greater than L, many times when the result of the overall heterogeneity test (log-likelihood ratio test) was not significant and/or for very sparse data, that is, especially overall AML and AML by subpopulation. With regard to the performance of the models by leukaemia type (Table 4), the CAR model appeared to yield the smallest DIC particularly for AL, the MIX model especially for AML and the KHR model especially for ALL.

When the separate contributions of fit $\overline{D}$ and complexity p_D were considered, all the models were found to have resulted in considerable amount of shrinkage: between approximately 3 and 26 'effective' parameters were needed to fit 94 relative risks. The smallest values of p_D obtained for the rarer data (Table 4) indicated that the prior structure contributed much information. Both the semiparametric MIX and KHR models had the smallest p_D values, whereas the BYM model had the highest values. In contrast, both the parametric BYM and CAR models resulted in the best fits with values of $\overline{D}$ ranging from approximately 74 to 103 (Table 4).

Strong correlations between the smoothed SIRs obtained using the various models were observed; the most strongly correlated SIRs were those obtained with the BYM and the MIX model (from 0.94 to 0.98). The weakest correlations were between the CAR and the KHR models (from 0.23 to 0.69).

Overall amount of smoothing

When a spatial pattern emerged in a given map (e.g. ALL, 0–4 y.o., 1990–1994) derived from one model (e.g. KHR), it frequently occurred in the maps of the same subpopulation derived from other models. In that respect, the CAR model was an exception. The resulting maps did not often show any spatial pattern and were less

Table 3 SIRs of overall acute leukaemia (1990–2004, boys and girls, 0–14 y.o.) by French département

Département *i*	y_i[a]	E_i[b]	SIR_i[c]		$CI_{95\%}(\theta)$[d]
01 – Ain	63	65.9	0.96		0.77–1.18
02 – Aisne	63	68.0	0.93		0.74–1.14
03 – Allier	28	33.3	0.84		0.60–1.15
04 – Alpes-de Haute-Provence	19	15.2	1.25		0.82–1.83
05 – Hautes-Alpes	20	13.6	1.47		0.97–2.13
06 – Alpes-Maritimes	110	101.5	1.08		0.92–1.27
07 – Ardèche	37	31.1	1.19		0.89–1.56
08 – Ardennes	38	36.5	1.04		0.78–1.36
09 – Ariège	11	13.4	0.82		0.46–1.36
10 – Aube	28	34.2	0.82		0.58–1.12
11 – Aude	30	31.5	0.95		0.68–1.29
12 – Aveyron	16	25.3	0.63	*NS	0.40–0.96
13 – Bouches-du-Rhône	236	208.4	1.13	*NS	1.01–1.26
14 – Calvados	84	77.3	1.09		0.90–1.30
15 – Cantal	13	14.0	0.93		0.55–1.47
16 – Charente	34	35.4	0.96		0.71–1.28
17 – Charente-Maritime	58	56.5	1.03		0.82–1.28
18 – Cher	42	33.5	1.25		0.95–1.62
19 – Corrèze	11	21.0	0.52	*NS	0.29–0.87
21 – Côte-d'Or	59	57.2	1.03		0.82–1.28
22 – Côtes-d'Armor	60	58.0	1.03		0.82–1.28
23 – Creuse	10	10.7	0.94		0.51–1.59
24 – Dordogne	49	36.4	1.35	*NS	1.05–1.71
25 – Doubs	55	60.0	0.92		0.72–1.15
26 – Drôme	49	52.3	0.94		0.73–1.19
27 – Eure	51	70.2	0.73	**NS	0.57–0.92
28 – Eure-et-Loir	46	51.7	0.89		0.66–1.14
29 – Finistère	104	94.2	1.10		0.93–1.30
30 – Gard	78	69.8	1.12		0.92–1.35
31 – Haute-Garonne	111	112.1	0.99		0.84–1.16
32 – Gers	18	16.1	1.12		0.72–1.66
33 – Gironde	167	138.6	1.20	*NS	1.06–1.37
34 – Hérault	102	95.2	1.07		0.90–1.26
35 – Ille-et-Vilaine	109	102.4	1.06		0.90–1.25
36 – Indre	19	22.6	0.84		0.55–1.24
37 – Indre-et-Loire	77	61.1	1.26	*NS	1.03–1.52
38 – Isère	134	134.3	1.00		0.86–1.15
39 – Jura	45	29.3	1.54	**NS	1.18–1.97
40 – Landes	35	33.0	1.06		0.78–1.40
41 – Loir-et-Cher	39	34.3	1.14		0.86–1.49
42 – Loire	83	83.7	0.99		0.82–1.19
43 – Haute-Loire	18	22.7	0.79		0.51–1.18
44 – Loire-atlantique	143	136.1	1.05		0.91–1.21
45 – Loiret	73	73.9	0.99		0.81–1.20
46 – Lot	21	15.1	1.39		0.93–2.00
47 – Lot-et-Garonne	18	31.2	0.58	**NS	0.37–0.86
48 – Lozère	7	7.3	0.96		0.45–1.81
49 – Maine-et-Loire	75	90.1	0.83		0.68–1.01
50 – Manche	57	57.1	1.00		0.79–1.24
51 – Marne	79	67.3	1.17		0.97–1.42
52 – Haute-Marne	22	22.9	0.96		0.65–1.37
53 – Mayenne	31	34.8	0.89		0.64–1.20
54 – Meurthe-et-Moselle	95	82.9	1.15		0.96–1.36
55 – Meuse	18	23.2	0.78		0.50–1.15
56 – Morbihan	76	72.8	1.04		0.86–1.26
57 – Moselle	112	121.8	0.92		0.78–1.08
58 – Nièvre	27	22.4	1.21		0.85–1.67
59 – Nord	310	340.6	0.91		0.87–1.00
60 – Oise	104	103.2	1.01		0.85–1.19
61 – Orne	36	33.8	1.06		0.79–1.41
62 – Pas-de-Calais	183	189.0	0.97		0.85–1.09
63 – Puy-de-Dôme	65	61.3	1.06		0.85–1.30
64 – Pyrénées-Atlantiques	63	61.0	1.03		0.83–1.27
65 – Hautes-Pyrénées	19	21.2	0.89		0.56–1.31
66 – Pyrénées-Orientales	43	39.3	1.09		0.83–1.41
67 – Bas-Rhin	124	119.7	1.04		0.89–1.20
68 – Haut-Rhin	69	85.4	0.81	*NS	0.66–0.99
69 – Rhône	197	190.8	1.03		0.91–1.16
70 – Haute-Saône	22	26.9	0.82		0.55–1.17
71 – Saône-et-Loire	64	58.7	1.09		0.86–1.34
72 – Sarthe	43	62.0	0.69	**NS	0.53–0.89
73 – Savoie	43	43.2	0.99		0.76–1.28
74 – Haute-Savoie	76	79.8	0.95		0.78–1.15
75 – Paris	189	190.5	0.99		0.86–1.12
76 – Seine-Maritime	137	153.9	0.89		0.77–1.03

Table 3 (continued)

Département i	y_i[a]	E_i[b]	SIR$_i$[c]		CI$_{90\%}$(θ_i)[d]
77 – Seine-et-Marne	158	164.4	0.96		0.84–1.10
78 – Yvelines	168	182.0	0.92		0.81–1.05
79 – Deux-Sèvres	44	37.9	1.16		0.89–1.49
80 – Somme	70	66.4	1.05		0.86–1.29
81 – Tarn	47	34.8	1.35	*NS	1.04–1.72
82 – Tarn-et-Garonne	24	22.2	1.08		0.75–1.52
83 – Var	99	96.5	1.03		0.86–1.21
84 – Vaucluse	59	59.2	1.00		0.79–1.24
85 – Vendée	57	60.6	0.94		0.75–1.17
86 – Vienne	44	41.2	1.07		0.82–1.37
87 – Haute-Vienne	34	32.4	1.05		0.77–1.40
88 – Vosges	45	45.1	1.00		0.77–1.28
89 – Yonne	47	37.9	1.24		0.96–1.58
90 – Territoire-de-Belfort	18	16.5	1.09		0.71–1.62
91 – Essonne	161	150.0	1.07		0.94–1.22
92 – Hauts-de-Seine	151	170.4	0.89		0.77–1.01
93 – Seine-Saint-Denis	161	192.2	0.84	*NS	0.73–0.95
94 – Val-de-Marne	158	151.5	1.04		0.91–1.19
95 – Val-d'Oise	151	154.8	0.98		0.85–1.12

CI, confidence interval; SIR, standardized incidence ratio.
[a]y_i observed number of incident cases in département i.
[b]E_i expected number of incident cases in département i.
[c]SIR$_i$=y_i/E_i standardized incidence ratio in département i.
[d]CI$_{90\%}$(θ_i) 90% confidence interval of relative risk θ_i in *département i* (based on exact Poisson distribution).
*NS/**NS: P value of exact Poisson test <5/<1% but nonsignificant according to Benjamini-Hochberg test.

correlated with those obtained with the other models, but were strongly correlated with the local SIRs.

Nevertheless, for a given model, it was difficult to detect any stable relative risk patterns shown by the 24 maps, indeed those maps showed very little contrast and were frequently poorly correlated with each other (e.g. the maps of ALL, 0–4 y.o. using the BYM model, for 1990–1994 and for 1995–1999).

Irrespective of the model, the ranges of the smoothed SIRs exhibited considerable shrinkage relative to the ranges of the local SIRs. A minimum range of 0.97–1.03 was obtained with the MIX model, and a maximum range of 0.90–1.67 was obtained with the CAR model, both for AML over the 1990–1994 period, the subpopulation which showed the greatest SIR range and inaccuracy (Table 2). The parametric models induced less shrinkage in line with the greater p_D values.

Smoothed standardized incidence ratio maps

No spatial structure of the AL maps was observed, irrespective of the model used, considering AL overall (Fig. 2) or by time period, by sex or by age group. For overall AL, all the 90% credibility intervals included 1, except for département No. 93 with the BYM model [CrI$_{90\%}$ = (0.83–0.99)] but this particular result did not remain stable by subpopulation.

For overall ALL, the smoothed SIRs of the south-west départements were slightly higher (with BYM, approximately 1.05–1.15; with MIX, about 1.02; and with KHR, about 1.07) than those of the north-east départements (with BYM, about 0.9–0.95; with MIX, about 0.98; and with KHR about 0.95) (Fig. 3). This structure was also observed on the smoothed SIR maps for girls and for the 0–4 y.o. age group, but did not recur in the other subpopulations. The few smoothed SIRs found significantly different from 1 (mainly for départements No. 13, 33, 59 and 93), on overall ALL with BYM and KHR models, did not remain significant when investigated by subpopulation.

On the maps of overall AML for models BYM, CAR, and KHR, the smoothed SIRs of the centre-east départements were slightly higher (with BYM, about 1.05–1.08; with CAR, about 1.08; and with KHR, 1.06) than the others (Fig. 3). However, the structure only recurred on a few subpopulation maps: for the period 1995–1999 and for boys. All the smoothed SIRs had a 90% credibility interval including 1.

Smoothed standardized incidence ratio correlations between time periods, sexes and age groups

The correlations of the SIRs (smoothed using a same model) for the three time periods, two sexes and three age groups (for overall AL, ALL and AML) were investigated.

The greater the smoothing, the more correlated the smoothed SIRs. The lowest correlations were found with the CAR model. In contrast, some correlations, especially for the SIRs smoothed by BYM or KHR, were strong but none was significant with the modified t-test. For instance, the estimate of the boy–girl correlation coefficient of the SIRs for ALL, smoothed by BYM, was 0.41 and highly significant with the classic t-test (92 d.f.) but no longer significant with the modified t-test (20.2

Table 4 Harmonic mean of expected number of cases *hm(E)*, likelihood ratio test statistic *L*, DIC, effective number of parameters p_D and posterior mean deviance $\overline{D}$ by leukaemia type, time period, sex and age group

			Minimum DIC[c] or difference from the minimum					
	hm(E)[a]	Likelihood ratio *L*[b]	BYM	CAR	MIX	KHR	Minimum p_D[d]	Minimum $\overline{D}$[d]
Acute leukaemia								
AL	39.5	112.1	+5.7	**109.6**	+4.8	+2.6	11.6	85.1
AL 1990–1994	13.1	99.2	+6.1	**102.0**	+1.7	+2.3	7.3	87.4
AL 1995–1999	12.9	117.9	+3.1	+6.5	+4.4	**111.0**	9.2	95.3
AL 2000–2004	13.5	106.1	+4.7	**105.6**	+2.8	+3.8	10.2	87.3
AL boys	21.9	99.3	+3.5	+1.0	+0.6	**102.2**	11.8	85.8
AL girls	17.6	98.3	+5.1	**97.4**	+3.8	+1.6	8.6	79.7
AL 0–4 y.o.	19.3	102.2	+2.5	+3.2	+2.9	**100.8**	10.0	83.5
AL 5–9 y.o.	11.8	104.2	+3.4	**105.4**	+1.5	+3.3	9.8	89.0
AL 10–14 y.o.	8.4	119.5	+1.1	+1.0	**118.3**	+2.2	10.1	101.8
Acute lymphoblastic leukaemia								
ALL	32.1	116.8	+2.9	+6.3	+5.6	**108.8**	10.6	87.1
ALL 1990–1994	10.7	92.1	+3.5	+1.2	+0.7	**95.4**	6.9	83.6
ALL 1995–1999	10.5	109.4	+2.1	+6.8	+4.7	**104.5**	8.5	90.1
ALL 2000–2004	10.9	111.7	**108.7**	+2.0	+1.1	+2.2	11.8	89.6
ALL boys	18.1	99.2	+1.8	+2.9	+0.7	**100.6**	11.4	83.3
ALL girls	14.0	107.7	+5.0	+7.7	+7.5	**99.5**	8.4	86.3
ALL 0–4 y.o.	16.0	92.7	+4.1	+8.4	+5.4	**88.5**	8.6	74.5
ALL 5–9 y.o.	9.9	114.7	+1.3	+0.03	**114.9**	+1.7	10.8	98.7
ALL 10–14 y.o.	6.2	104.8	+0.9	+3.1	**103.7**	+1.9	7.8	89.8
Acute myeloblastic leukaemia								
AML	6.7	101.7	+1.8	+0.6	**103.8**	+0.7	6.4	92.9
AML 1990–1994	3.3	115.8	+2.8	**115.3**	+1.5	+3.2	2.9	103.6
AML 1995–1999	3.4	98.5	+2.8	+1.6	**99.6**	+0.8	2.9	94.7
AML 2000–2004	2.2	93.3	+1.5	+2.5	**94.0**	+1.2	3.0	87.0
AML boys	2.3	108.4	+3.0	+0.3	**110.2**	+1.5	3.8	103.3
AML girls	2.2	95.3	+2.8	+0.9	**97.4**	+1.7	3.5	91.5

AL, acute leukaemia; ALL, acute lymphoblastic leukaemia; AML, acute myeloblastic leukaemia; DIC, deviance information criteria; SIR, standardized incidence ratio; y.o., years old.

[a] *hm(E)* harmonic mean of the expected numbers of cases.

[b] $L = 2\sum_i y_i \log(SIR)$ likelihood ratio statistic with approximate $\chi^2_{(n-1)}$; n=94; y_i observed number of incident cases in département i; E_i expected number of incident cases in département i; $SIR_i = y_i / E_i$.

[c] $DIC = p_D + \bar{D}$; minimum **DIC** is greater than the sum of the two minima.

[d] Minimum p_D and minimum $\bar{D}$ are the minima between the four models.

modified d.f.). Besides, scatterplots of the SIRs smoothed by the KHR model (the most correlated) did not show evidence of any correlation between the smoothed SIRs.

Hyperprior choice sensitivity

The smoothed SIRs of overall AL, ALL and AML were investigated by considering other priors implying a much greater variability. A prior 95% fluctuation interval of the relative risk enabled variation between 0.1 and 9 on average, up to 0 to + ∞. The range 0.1–9 is 20-fold wider than that of the chosen *a priori* (0.8–1.25).

The variability of the smoothed SIRs was found to be slightly greater in this sensitivity analysis than that observed with the chosen *a priori*, with the BYM and CAR models: for instance, for overall AL, the extreme values were 0.84 and 1.20 with the BYM model [$\Gamma(0.1,0.1)$ on τ_u^2 and τ_v^2 in line with Green and Richardson (Green and Richardson, 2002)] and 0.79 and 1.25 with the CAR model [$\Gamma(0.1,4)$ on τ^2]. In contrast, with both semi-parametric models, the range of smoothed SIRs for AL underwent greater shrinkage. For the KHR model, the range was from 0.99 to 1.02 [$\Gamma(1,1.2)$ on τ^2]. With the MIX model and a $\Gamma(0.1,0.1)$ distribution *a priori*, the smoothing was stronger than with the *a priori* $\Gamma(100,100)$: for overall AL, between 0.998 and 1.001. In any event, the maps showed no structure at all and the results for local excesses remained unchanged.

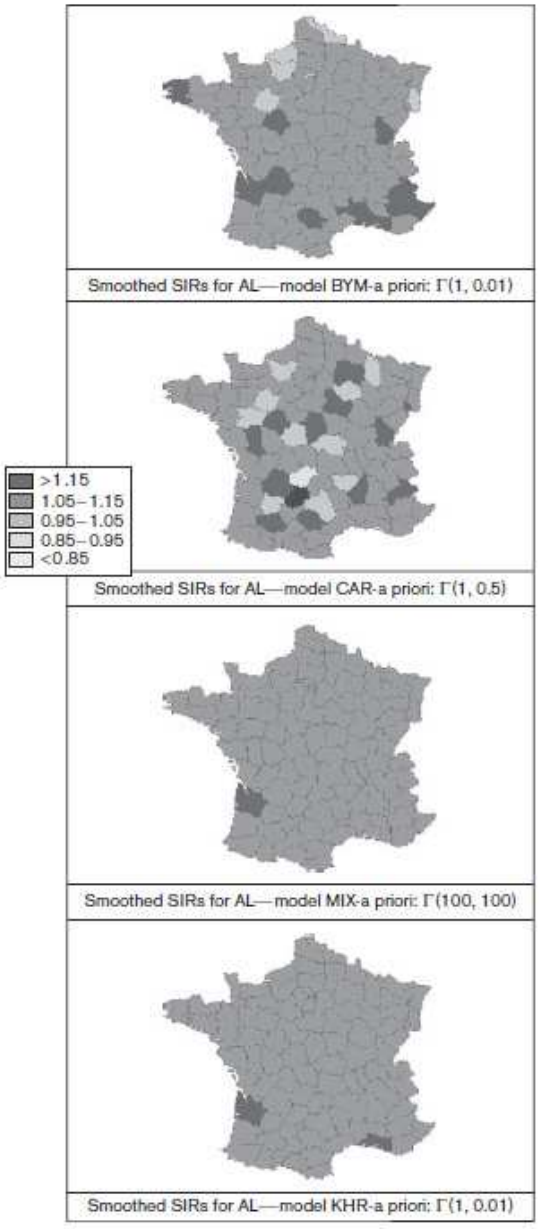
Discussion

No spatial structure was observed with the local or smoothed estimates. Although very slight overall heterogeneity in the overall AL was observed, the results did not remain stable when investigated by leukaemia type, time period, sex or age group.

None of the smoothing methods used suggested the presence of a geographical pattern.

Fully Bayesian models previously developed in the context of disease mapping were used since they estimate the overall spatial pattern of relative risks, borrowing strength across the whole country and reducing the variance through the use of shrinkage estimators. The effect of shrinkage was thus dependent on the prior assumptions. Various prior assumptions were made, first, on the latent spatial

Fig. 2



Standardized incidence ratios (SIRs) of overall acute leukaemia (AL) smoothed by BYM, CAR, MIX and KHR.

structure of relative risk variability via both the parametric BYM and CAR models and the semiparametric, MIX and KHR models, and, second, on the range of the variability.

The two parametric spatial BYM and CAR models are both conditional autoregressions on the neighbours, but they differ not only in the manner in which spatial dependence is defined and interpreted, but also in the manner in which neighbouring relative risks are taken into account in the estimation.

However, the amount of smoothing with these parametric models, controlled by the variance parameters in the BYM model, or by the autocorrelation and variance parameters in the CAR model, is affected globally by all of the départements and is not locally adapted. In order to overcome the problem of potential oversmoothing, the semiparametric MIX and KHR models were used. In those models, the continuous spatial distribution of the relative risks assumed in the parametric models is replaced by discrete allocation (Richardson *et al.*, 2004) or partition models (Knorr-Held and Rasser, 2000).

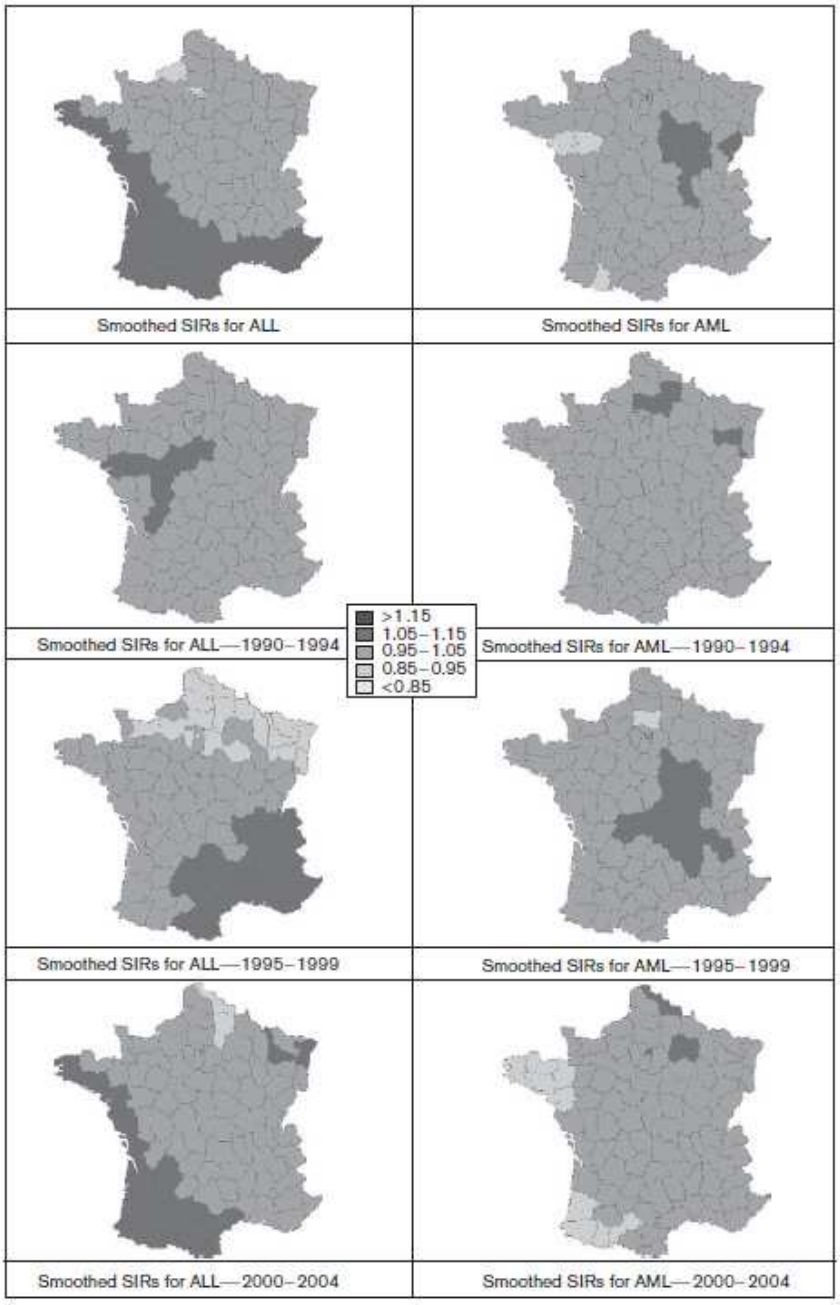
The discrete allocation mechanism of the MIX model defines risk classes, *a priori*, which first allows for continuous spatial pattern in promoting the allocation of neighbouring départements to the same risk class, and then also adapts to local discontinuities in risk pattern by allowing non-neighbouring départements to belong to the same risk class.

In contrast, in the KHR partition model, clusters of contiguous départements are defined *a priori*. All départements in a cluster are neighbours, whereas that is not the case for départements allocated to the same risk class in the MIX model. Depending on the model used, different numbers of iterations were carried out to estimate the smoothed SIRs. However, the results were unchanged when the smoothed SIRs were computed with the same number of iterations for each of the models.

Considering the range of variability in the epidemiological context of département-specific relative risks of CL, it seemed reasonable to consider that, *a priori*, a 95% fluctuation interval of the risk for a département had a mean range from 0.8 to 1.25 and a 95% range from 0.25 to 4.0. Using a wider range of variability from 0.1 to 9 on average, up to 0 to $+\infty$, similar results were obtained. The results were thus not dependent on the hyperprior selected.

Thus, the models used in this study covered a wide range of the models available for the detection of various kinds of spatial patterns (Best *et al.*, 2005). However, simulation studies (Knorr-Held and Rasser, 1999; Green and Richardson, 2002; Richardson *et al.*, 2004; Best *et al.*, 2005) have shown that true high risks are likely to be underestimated by these Bayesian models. There may be a lack of power to detect moderate potential risks, such as most environmental risks, especially when expected

Fig. 3



Standardized incidence ratios (SIRs) of acute lymphoblastic leukaemia (ALL) and acute myeloblastic leukaemia (AML), by time period, smoothed by KHR-Γ(1,0.01).

numbers of cases are small. Herein, that might be the case particularly for overall AML [range (E) = (1.24–58.17)] and *a fortiori* for AML by subpopulations (Table 2).

To the authors' knowledge, very few studies have focused on the geographical heterogeneity of CL incidence on a large scale and a 15-year period using the approach described. The NRCH is an exhaustive and therefore very reliable data set that covers the whole country since 1990. Such exhaustiveness is rare and makes the study innovative.

No other study has used all the above methods to detect heterogeneity in childhood cancer data, although some studies conducted in other countries and using fewer methods have found similar results. No spatial heterogeneity (GCCR registry website – annual report 1999 – *http://www.kinderkrebsregister.de/english/*; Muirhead, 1995; Williams *et al.*, 2001) or spatial variations on the borderline of significance was observed (Alexander *et al.*, 1997; Petridou *et al.*, 1997; Wakefield *et al.*, 2000; McNally *et al.*, 2009).

A previous study conducted on a finer scale in France (commune) found a slight heterogeneity (1.2% significant overdispersion) in the incidence of AL over the period 1990–1994, which is comparable with that found in this study, even though for a different period (Bellec *et al.*, 2006). The results reported herein are also consistent with the results of the EUROCLUS study which detected an extra-Poisson component of the variance equal to 1.7%, on an intermediate scale between the commune and département scales (Alexander *et al.*, 1998). The French study (Bellec *et al.*, 2006) also conducted to detect specific clusters did not lead to significant results.

However, the nonspatial heterogeneity of childhood AL incidence on the département scale does not imply that there is no spatial heterogeneity in incidence at all scales. In this study population and in others, suspected risk factors have been detected on a finer scale than the département scale and found to be linked to child hood AL, in particular population mixing (Rudant *et al.*, 2006; Bellec *et al.*, 2008; McNally *et al.*, 2009) and, possibly, domestic radon exposure (Laurier *et al.*, 2001; Evrard *et al.*, 2005).

Other studies have investigated links between childhood AL incidence and localized environmental sources of exposure. For example, there was no evidence of increased risk of CL around French nuclear sites (White-Koning *et al.*, 2004). However, exposure to benzene (related to road traffic) at levels greater than $10\,\mu g/m^3$ was found to be related to a four-fold increase in the risk of CL (Crosignani *et al.*, 2004) and a review of nine case–control studies on extremely low-frequency electromagnetic fields found a two-fold increase in the risk of leukaemia for children exposed to at least $0.4\,\mu T$ (Ahlbom *et al.*, 2000). Moreover, some individual risk factors for childhood AL have been identified, such as high-dose ionizing radiations, iatrogenic factors (chemotherapy) and genetic factors such as trisomy 21. However, these risk factors are not likely to create spatial heterogeneity on the département scale.

In conclusion, this study conducted using various complementary Bayesian smoothing methods did not find evidence of any consistent geographical variations in the incidence of childhood AL in France at the département scale over the period 1990–2004. This result neither excludes that there is spatial heterogeneity in incidence at finer scales nor that CL incidence could correlate spatially with some population characteristics that would exhibit interdépartement variability. It does imply, however, that there does not seem to be any reason to adjust for overall between-département variations when examining spatial variation in the incidence of CL.

Acknowledgements

The authors are grateful to Peter Green for kindly providing us the computer code of the MIX model and valuable advice. The authors also express their gratitude to Andrew Mullarky for his skilful revision of the manuscript. This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), the Agence Nationale de la Recherche (ANR) and the Institut de Veille Sanitaire (InVS).

References

Ahlbom A, Day N, Feychting M, Roman E, Skinner J, Dockerty J, *et al.* (2000). A pooled analysis of magnetic fields and childhood leukaemia. *Br J Cancer* **83**:692–698.

Alexander FE, Chan LC, Lam TH, Yuen P, Leung NK, Ha SY, *et al.* (1997). Clustering of childhood leukaemia in Hong Kong: association with the childhood peak and common acute lymphoblastic leukaemia and with population mixing. *Br J Cancer* **75**:457–463.

Alexander FE, Boyle P, Carli PM, Coebergh JW, Draper GJ, Ekbom A, *et al.* (1998). Spatial clustering of childhood leukaemia: summary results from the EUROCLUS project. *Br J Cancer* **77**:818–824.

Bellec S, Hémon D, Rudant J, Goubin A, Clavel J (2006). Spatial and space-time clustering of childhood acute leukaemia in France from 1990 to 2000: a nationwide study. *Br J Cancer* **94**:763–770.

Bellec S, Baccaini B, Goubin A, Rudant J, Ripert M, Hémon D, *et al.* (2008). Childhood leukaemia and population movements in France, 1990–2003. *Br J Cancer* **98**:225–231.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc* **57**:289–300.

Besag J, York J, Mollié A (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann Institute of Stat Math* **43**:1–59.

Best N, Richardson S, Thomson A (2005). A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* **14**:35–59.

Clavel J, Goubin A, Auclerc MF, Auvrignon A, Waterkeyn C, Patte C, *et al.* (2004). Incidence of childhood leukaemia and non-Hodgkin's lymphoma in France: national registry of childhood leukaemia and lymphoma, 1990–1999. *Eur J Cancer Prev* **13**:97–103.

Cliff A, Ord J (1981). *Spatial processes. Models and applications.* Pion: London.

Clifford P, Richardson S, Hémon D (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics* **45**:123–134.

Cressie N (1993). *Statistics for spatial data, revised edition*. NY: John Wiley & Sons.
Crosignani P, Tittarelli A, Borgini A, Codazzi T, Rovelli A, Porro E, *et al.* (2004). Childhood leukemia and road traffic: a population-based case-control study. *Int J Cancer* **108**:596–599.
Estève J, Benhamou E, Raymond L (1993). *Méthodes statistiques en épidémiologie descriptive*. Les Editions INSERM Paris.
Evrard AS, Hémon D, Billon S, Laurier D, Jougla E, Tirmarche M, *et al.* (2005). Ecological association between indoor radon concentration and childhood leukaemia incidence in France, 1990–1998. *Eur J Cancer Prev* **14**:147–157.
Green P (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**:711–732.
Green P, Richardson S (2002). Hidden Markov models and disease mapping. *J Am Stat Assoc* **97**:1055–1070.
Knorr-Held L, Rasser G (1999). *Bayesian detection of clusters and discontinuities in disease maps: simulations*, discussion paper no. 142, Sonderforschungsbereich 386 der Ludwig–Maximilians–Universität München, available at http: //epub.ub.uni-muenchen.de/1531/
Knorr-Held L, Rasser G (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56**:13–21.
Laurier D, Valenty M, Tirmarche M (2001). Radon exposure and the risk of leukemia: a review of epidemiological studies. *Health Phys* **81**:272–288.
Lunn D, Thomas A, Best N, Spiegelhalter D (2000). Winbugs–a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* **10**:325–337.
McNally R, Eden T (2004). An infectious aetiology for childhood acute leukaemia: a review of the evidence. *Br J Haematol* **127**:243–263.
McNally RJQ, Alexander FE, Vincent TJ, Murphy MFG (2009). Spatial clustering of childhood cancer in Great Britain during the period 1969–1993. *Int J Cancer* **124**:932–936.
Muirhead CR (1995). Childhood leukemia in metropolitan regions in the United States: a possible relation to population density? *Cancer Causes Control* **6**:383–388.
Muirhead C, Butland B (1996). Testing for over-dispersion using an adapted form of the Potthoff-Whittinghill method. In: Alexander F, Boyle P, editors. *Methods for investigating localized clustering of disease*. Lyon, France: IARC Scientific Publications No. 135; pp. 40–52.
Petridou E, Alexander FE, Trichopoulos D, Revinthi K, Dessypris N, Wray N, *et al.* (1997). Aggregation of childhood leukemia in geographic areas of Greece. *Cancer Causes Control* **8**:239–245.
Potthoff R, Whittinghill M (1966a). Testing for homogeneity. i. The binomial and multinomial distributions. *Biometrika* **53**:167–182.
Potthoff R, Whittinghill M (1966b). Testing for homogeneity. ii. The Poisson distribution. *Biometrika* **53**:183–190.
Richardson S, Thomson A, Best N, Elliott P (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* **112**:1016–1025.
Rudant J, Baccaïni B, Ripert M, Goubin A, Bellec S, Hémon D, *et al.* (2006). Population-mixing at the place of residence at the time of birth and incidence of childhood leukaemia in France. *Eur J Cancer* **42**:927–933.
Spiegelhalter D, Best N, Carlin B, van der Linde A (2002). Bayesian measures of model complexity and fit. *J R Statist Soc* **64**:583–639.
Stern HS, Cressie N (2000). Posterior predictive model checks for disease mapping models. *Stat Med* **19**:2377–2397.
Stiller CA, Marcos-Gragera R, Ardanaz E, Pannelli F, Almar Marqués E, Cañada Martínez A, *et al.* (2006). Geographical patterns of childhood cancer incidence in Europe, 1988–1997. Report from the Automated Childhood Cancer Information System project. *Eur J Cancer* **42**:1952–1960.
Wakefield J, Best N, Waller L (2000). Bayesian approaches to disease mapping. In: Elliott P, Wakefield J, Best N, Briggs D, editors. *Spatial epidemiology. Methods and applications*. Oxford: Oxford University Press; pp. 104–127.
White-Koning ML, Hémon D, Laurier D, Tirmarche M, Jougla E, Goubin A, *et al.* (2004). Incidence of childhood leukaemia in the vicinity of nuclear sites in France, 1990–1998. *Br J Cancer* **91**:916–922.
Williams JR, Alexander FE, Cartwright RA, McNally RJQ (2001). Methods for eliciting aetiological clues from geographically clustered cases of disease, with application to leukaemia-lymphoma data. *J R Statist Soc A* **164**:49–60.

uicc
global cancer control

IJC
International Journal of Cancer

Childhood leukemia around French nuclear power plants—The Geocap study, 2002–2007

Claire Sermage-Faure[1,2], Dominique Laurier[4], Stéphanie Goujon-Bellec[1,2,3], Michel Chartier[5], Aurélie Guyot-Goubin[1,2,3], Jérémie Rudant[1,2,3], Denis Hémon[1,2] and Jacqueline Clavel[1,2,3]

[1] INSERM Unit8, CESP Center for research in Epidemiology and Population Health, Environmental Epidemiology of Cancer Team, Villejuif F-94807, Paris, France
[2] Université Paris-Sud, UMRS1018, F-94807, Villejuif, Paris, France
[3] French National Registry of Childhood Hematological malignancies (NRCH), Villejuif F-94807, Paris, France
[4] Institute of Radiological Protection and Nuclear Safety, IRSN/DRPH/SRBE, Fontenay-aux-Roses F-92262, Cedex, France
[5] Institute of Radiological Protection and Nuclear Safety, IRSN/DRPH/SER, F-92262, Fontenay-aux-Roses, France

The aim of this work is to study the risk of childhood acute leukemia (AL) around French nuclear power plants (NPPs). The nationwide Geocap case–control study included the 2,753 cases diagnosed in mainland France over 2002–2007 and 30,000 contemporaneous population controls. The last addresses were geocoded and located around the 19 NPPs. The study used distance to NPPs and a dose-based geographic zoning (DBGZ), based on the estimated dose to bone marrow related to NPP gaseous discharges. An odds ratio (OR) of 1.9 [1.0–3.3], based on 14 cases, was evidenced for children living within 5 km of NPPs compared to those living 20 km or further away, and a very similar association was observed in the concomitant incidence study (standardized incidence ratio (SIR) = 1.9 [1.0–3.2]). These results were similar for all the 5-year-age groups. They persisted after stratification for several contextual characteristics of the municipalities of residence. Conversely, using the DBGZ resulted in OR and SIR close to one in all of the dose categories. There was no increase in AL incidence over 1990–2001 and over the entire 1990–2007 period. The results suggest a possible excess risk of AL in the close vicinity of French NPPs in 2002–2007. The absence of any association with the DBGZ may indicate that the association is not explained by NPP gaseous discharges. Overall, the findings call for investigation for potential risk factors related to the vicinity of NPP and collaborative analysis of multisite studies conducted in various countries.

Key words: childhood leukemia, epidemiology, incidence, case–control, nuclear reactors

Abbreviations: AL: acute leukemia; ALL: acute lymphoblastic leukemia; AML: acute myeloblastic leukemia; DBGZ: dose-based geographic zoning; NPP: nuclear power plant; NRCH: French National Registry of Childhood Hematopoietic Malignancies; O (resp. E): observed (resp. expected) number of cases; OR: odds ratio; SIR: standardized incidence ratio

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Institut de Veille Sanitaire (InVS), Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES), Association pour la Recherche sur le Cancer (ARC), Fondation Pfizer, Institut National du Cancer (INCa), Agence Nationale de la Recherche (ANR)
DOI: 10.1002/ijc.27425
History: Received 24 Aug 2011; Accepted 20 Dec 2011; Online 5 Jan 2012

Correspondence to: Jacqueline Clavel, CESP Equipe 6, INSERM U1018, 16 av. Paul Vaillant-Couturier, F-94807 Villejuif, Cedex, France, Tel: +33-1-45-59-5037, Fax: +33-1-45-59-5151, E-mail: jacqueline.clavel@inserm.fr

The risk of childhood leukemia around nuclear power plants (NPPs) has given rise to considerable debate. Several epidemiological studies have analyzed the incidence of childhood leukemia around nuclear sites[1–5] and, more specifically, NPPs.[6–13] In general, no excess risk has been evidenced by the multisite studies although persistent localized excesses of childhood acute leukemia (AL) have been reported around a few specific sites: the Sellafield[14–16] and Dounreay[17,18] nuclear fuel reprocessing plants in the United Kingdom and the Krummel NPP in Germany.[10,19,20] Given the low radiation levels measured near the sites, the hypothesis that local population mixing related to the installation of the sites might ease the spread of a leukemogenic agent was raised[21–23] and appeared consistent with some further observations in the village of Seascale close to the Sellafield nuclear site.[24] Seascale underwent a period of intensive population mixing, and most of the observed excess appeared primarily related to this factor. Recently, the German KiKK population-based case–control study showed an association between leukemia in children less than 5 years old and living less than 5 km from a NPP.[25–28] In France, a national geographic study of the period 1990–1998 was carried out[12,29]; however, no

association was found between the incidence of childhood AL and distance from nuclear sites. No association was demonstrated by an extended 1990–2001 study that used geographic zoning based on gaseous discharge dose estimates to assess radiation exposure in the municipalities near the sites.[30]

The aim of our study was to provide updated results on the risk of childhood leukemia near French NPPs, using a case–control design in addition to the usual geographic incidence approach. All the cases of childhood AL for the period 2002–2007 and a set of contemporaneous population controls were individually located and categorized in terms of their distance from the 19 French NPPs and dose-based geographic zoning (DBGZ). The updated incidence study (1990–2007) is also presented.

Material and Methods

The Geocap case–control study, 2002–2007

The case–control study included all the 2,753 French childhood leukemia cases aged up to 15 years at the end of the year of diagnosis, diagnosed between 2002 and 2007, and residing in metropolitan France. The cases were obtained from the French National Registry of Childhood Hematopoietic Malignancies (NRCH).[31]

A set of 30,000 control addresses, 5,000 each year for the period 2002–2007, representative of the French pediatric population for those years, was randomly sampled from the pediatric population of mainland France by the National Institute for Statistics and Economic Studies (INSEE, Paris/France) using the income and council tax databases. The sample was stratified on the *Département* administrative unit. The control sample was closely representative of its source population in terms of age and number of children in the household and in terms of contextual socioeconomic and demographic variables: size of the urban unit, median income, proportion of blue-collar workers, proportion of subjects who successfully completed high school and proportion of homeowners in the municipality of residence (Table 1).

The age available in the INSEE databases was the age at the end of the year (*i.e.*, based on the year of birth), and therefore, this age variable was also used for the cases. The INSEE database did not specify child gender.

The age distribution of the cases included in the study showed the expected peak of incidence between 2 and 4 years old. The cases were closely comparable to the controls in terms of the contextual sociodemographic variables (Table 1).

Incidence studies: 1990–2001, 2002–2007 and 1990–2007

The incidence studies performed at the Commune level (the smallest French administrative unit) for the same period 2002–2007 as the case–control study, for the preceding period 1990–2001 and for the complete period 1990–2007. The cases consisted in all children diagnosed with AL before the age of 15 years and registered in the NRCH during the periods. The annual national incidence rates of AL estimated by the NRCH were taken as reference rates.

The estimates of Commune populations by year of age were directly provided by the INSEE for the census years: 1990, 1999, 2006 and 2007. For the other years, the estimates were interpolated from the census data and the yearly estimates provided by the INSEE for the 96 French mainland *Départements*. Person-years and expected numbers of AL cases were then computed for each year of the 1990–2007 period by Commune and 5-year-age group.

Geocoding of addresses, 2002–2007

The addresses of the cases and controls were geocoded by the GEOCIBLE Company (Saint-Maurice/France) using the geographic information system MAPINFO (NY/USA), the NAVTEQ (Paris/France) street databases and detailed vectorized maps from the National Geographic Institute (Saint-Mandé/France). The process resulted in the location of the addresses with an uncertainty of at most 100 m for 92% of the cases and 96% of the controls and with an uncertainty of 15 m for 67% of the cases and 76% of the controls. Only 3% of the cases and 1% of the controls were located by their Commune of residence and were thus allocated the coordinates of the town hall of their Commune.

The geocoding uncertainties were small compared to the range of distances considered in the analysis. All the cases and controls were located in the five categories of distance (0–5, 5–10, 10–15, 15–20 and $\geq$20 km) from NPPs without any uncertainty.

NPP characteristics

There are 19 NPPs in France (Supporting Information Table S1), all but two (*Chooz* and *Civaux*) having been commissioned before 1990. The majority of the NPPs are located near a river, but four of them are coastal. The NPPs *Cattenom*, *Chooz* and *Fessenheim* are located less than 20 km from the border with Luxembourg, Belgium and Germany, respectively. The nuclear electrical power generated ranges from 1,800 to 5,400 MW.

Distance from the nearest NPP

In the Geocap case–control study, the distance between the residence and the nearest NPP was derived from the geocoding. The 32,753 subjects included in the case–control study were located in 5-km wide rings within 20 km of the NPP or outside of the area. In the incidence studies, the Communes were positioned by the coordinates of the town hall and assigned to the areas defined above around the nearest NPP.

Dose-based geographic zoning

DBGZ had been developed by the National Institute of Radiological Protection and Nuclear Safety for the authors' previous analysis.[32] It used estimated bone marrow doses associated with gaseous radioactive discharges from the NPPs to classify the exposures at the town halls of the Communes

Table 1. Comparability of the Geocap controls with their source population by individual and contextual socioeconomic and demographic variables, 2002–2007, mainland France and distribution of 2002–2007 cases by the same individual or contextual variables

	Source population (%)	Geocap controls (%)	Geocap cases (%)
Individual characteristics			
Age at the end of the year			
0	6.2	6.1	2.5
1	6.6	6.3	6.4
2	6.7	6.9	12.3
3	6.7	6.7	14.1
4	6.7	6.4	11.6
5	6.7	6.9	8.5
6	6.7	6.7	6.9
7	6.7	6.8	6.1
8	6.6	6.5	5.1
9	6.6	6.9	4.9
10	6.6	6.6	4.4
11	6.7	6.7	4.2
12	6.7	6.8	4.1
13	6.8	6.9	4.1
14	6.9	7.0	5.0
Number of children in the household			
1	29.7	29.4	NA
2	42.9	43.1	NA
3	19.8	20.0	NA
≥4	7.7	7.5	NA
Contextual variables			
Size of the urban unit[1] of the Commune of residence			
Rural unit	26.1	25.8	26.7
<5,000 inhabitants	6.3	6.3	5.9
5,000–9,999 inhabitants	5.5	5.5	5.4
10,000–19,999 inhabitants	5.0	5.0	5.0
20,000–49,999 inhabitants	6.0	6.0	6.1
50,000–99,999 inhabitants	6.6	6.5	6.5
100,000–199,999 inhabitants	5.0	5.1	5.5
200,000–1,999,999 inhabitants	21.7	21.9	22.6
Paris urban unit	17.9	17.9	16.4
Median annual income of the Commune (€)			
<12,864	25.0	24.4	24.3
12,864–14,332	25.0	25.6	24.9
14,333–16,429	25.0	24.8	26.1
>16,429	25.0	25.2	24.7
Proportion of blue-collar workers in the Commune (%)			
<20	25.0	25.3	25.3
20–26.5	25.0	25.1	25.2
26.6–34.2	25.0	24.7	24.6
>34.2	25.0	24.9	24.9

Table 1. Comparability of the Geocap controls with their source population by individual and contextual socioeconomic and demographic variables, 2002–2007, mainland France and distribution of 2002–2007 cases by the same individual or contextual variables (Continued)

	Source population (%)	Geocap controls (%)	Geocap cases (%)
Proportion of baccalaureate holders in the Commune (%)			
<19.2	25.0	24.7	24.6
19.2–24.5	25.0	25.3	25.8
24.6–30.6	25.0	24.8	25.3
>30.6	25.0	25.2	24.3
Proportion of homeowners in the Commune (%)			
<42.8	25.0	25.2	23.7
42.8–60.1	25.0	24.7	25.5
60.2–76.0	25.0	24.9	25.6
>76.0	25.0	25.2	25.2

[1]The urban unit is defined by the National Institute of Statistics and Economic Studies (INSEE) as a group of Communes in which the distance between dwellings is not more than 200 m.

located less than 20 km from the nuclear sites. The Communes close to more than one NPP or to another nuclear site included in the previous analysis were allocated the sum of the doses estimated for each of the neighboring installations. This was the case for the *Flamanville* NPP, which is close to the *La Hague* nuclear site, and for the *Cruas* and *Tricastin* NPPs, which are close to the *Pierrelatte* nuclear site. For the analyses stratified by NPP, the Communes close to more than one NPP were assigned to the NPP for which the estimated dose was the highest.

In the case–control study, the cases and controls were allocated the dose estimated at the town hall of the Commune of residence. The cutoffs were chosen so as to obtain the same distribution of expected cases as that obtained by dividing the area around the NPPs into rings 5 km wide, that is, approximately 10, 20, 30 and 40% of the expected cases.

Statistical analysis

All the statistical analyses were performed using the SAS software package (version 9; SAS Institute, Cary, NC). The analyses were conducted on all the NPPs and all the cases, and by age group (0–4, 5–9 and 10–14 years), AL subtype: lymphoblastic (acute lymphoblastic leukemia [ALL]) and myeloblastic (acute myeloblastic leukemia [AML]), NPP and year of study.

In the case–control study, the odds ratios (ORs) and their 95% confidence intervals (CIs) were estimated by unconditional logistic regression adjusted for age (5-year-age groups) and *Département*.

In the incidence studies, the relative risk of AL was estimated by the standardized incidence ratio (SIR), defined as the ratio of the observed (O) and expected (E) numbers of cases. The 95% CI was calculated using Byar's approximation.[33] The statistical distribution of the number of cases observed yearly around each NPP was compared to that obtained under the hypothesis of a Poisson distribution, and no statistically or quantitatively significant departure was observed. Additionally, our previous analyses evidenced no quantitatively significant overdispersion at the Commune[34] or *Département*[35] scale in France.

The heterogeneity of the SIRs by year or NPP was tested using Pearson's χ^2 statistic, with external (national rates) and internal (rates within 20 km around NPPs) references. As small numbers might have impaired the validity of the tests, the statistical significance levels of the tests were estimated by simulation. In all, 50,000 distributions of the cases in the Communes were generated under the null hypothesis of a Poisson distribution with the corresponding expected number of cases, and the observed statistics were compared to the distribution of the 50,000 simulated statistics.

Analyses were also performed after excluding the subjects living further than 50 km from a NPP in the case–control study and computing the incidence between 20 and 50 km. Stratified analyses were conducted by NPP characteristic, that is, electrical power, coastal location and border location in the 2002–2007 case–control study and in the 1990–2007 incidence study. To account for potential confounders, the analyses were repeated after stratification or adjustment for several contextual variables in the models. The contextual variables were extracted from the 1999 census data and consisted in the urban status of the Commune (rural, semiurban or urban), the median income of the households, the proportion of blue-collar workers and the proportion of baccalaureate holders. The latter variables were used separately as well as jointly to take spatial socioeconomic heterogeneity into account. The data were also analyzed after exclusion of the cases and controls who lived less than 200 m from a high-voltage power line and those who lived less than 600 m from a power line, as an association between the proximity of power lines and the risk of AL has been suggested.[36,37]

Table 2. Distance to the nearest French nuclear power plant (NPP) and childhood acute leukemia: Geocap case–control study and incidence studies

	Geocap study, 2002–2007[1] (distance: residence-nearest NPP)			Incidence study, 2002–2007[2] (distance: municipality-nearest NPP)			Incidence study, 1990–2001[3] (distance: municipality-nearest NPP)			Incidence study, 1990–2007[2] (distance: municipality-nearest NPP)		
	Cases	Controls	OR [95% CI]	O	E	SIR [95% CI]	O	E	SIR [95% CI]	O	E	SIR [95% CI]
Age <15 years[3]												
0–4.99 km	14	80	1.9 [1.0–3.3]*	14	7.4	1.9 [1.0–3.2]*	10	13.6	0.7 [0.4–1.4]	24	21.0	1.1 [0.7–1.7]
5–9.99 km	17	213	0.9 [0.5–1.5]	19	20.6	0.9 [0.6–1.4]	40	39.2	1.0 [0.7–1.4]	59	59.8	1.0 [0.8–1.3]
10–14.99 km	27	320	0.9 [0.6–1.4]	30	25.4	1.2 [0.8–1.7]	50	48.5	1.0 [0.8–1.4]	80	73.9	1.1 [0.9–1.4]
15–19.99 km	41	447	1.0 [0.7–1.4]	36	42.4	0.9 [0.6–1.2]	73	81.5	0.9 [0.7–1.1]	109	124.0	0.9 [0.7–1.1]
≥20 km	2,654	28,940	1.0									
<20 km	99	1,060	1.0 [0.8–1.3]	99	95.7	1.0 [0.8–1.3]	173	182.9	0.9 [0.8–1.1]	272	278.6	1.0 [0.9–1.1]
≥20 km	2,654	28,940	1.0									
Age <5 years[3]												
0–4.99 km	6	27	1.6 [0.7–4.1]	8	3.6	2.2 [1.0–4.4]*	6	6.6	0.9 [0.3–2.0]	14	10.2	1.4 [0.8–2.3]
5–9.99 km	7	55	1.0 [0.5–2.3]	10	10.2	1.0 [0.5–1.8]	21	19.3	1.1 [0.7–1.7]	31	29.5	1.1 [0.7–1.5]
10–14.99 km	11	103	0.8 [0.4–1.4]	11	12.6	0.9 [0.4–1.6]	18	23.5	0.8 [0.5–1.2]	29	36.1	0.8 [0.5–1.2]
15–19.99 km	17	136	1.0 [0.6–1.7]	18	20.8	0.9 [0.5–1.4]	39	39.7	1.0 [0.7–1.3]	57	60.6	0.9 [0.7–1.2]
≥20 km	1,248	9,396	1.0									
Total <20 km	41	321	1.0 [0.7–1.4]	47	47.3	1.0 [0.7–1.3]	84	89.1	0.9 [0.8–1.2]	131	136.4	1.0 [0.8–1.1]
≥20 km	1,248	9,396	1.0									

[1]Odds ratios (ORs) and 95% confidence intervals (95% CIs) estimated by logistic regression adjusted for age at the end of the year (5-year-age groups for the 0- to 14-year-old children and 1-year-age groups for the 0- to 4-year-old children) and *Département* of residence. [2]Standardized incidence ratio (SIR) calculated as the ratio of the observed (O) to the expected (E) number of cases with Byar's approximation of 95% CI. [3]Age at the end of the year in the Geocap study; exact age in the incidence studies. *$p_{two\text{-}sided} < 0.05$.

Table 3. Description of the registered cases within 20 km of the nuclear power plants (NPP) over the period, 2002–2007

	Distance: residence-nearest NPP (km)					
	0–4.99 *n* (%)	5–9.99 *n* (%)	10–14.99 *n* (%)	15–19.99 *n* (%)	<20 *n* (%)	≥20 *n* (%)
Gender	14	17	27	41	99	2,654
Female	7 (50.0)	8 (47.1)	13 (48.1)	20 (48.8)	48 (48.5)	1,202 (45.3)
Male	7 (50.0)	9 (52.9)	14 (51.9)	21 (51.2)	51 (51.5)	1,452 (54.7)
Age						
<5 years	6 (42.9)	7 (41.2)	11 (40.7)	17 (41.5)	41 (41.4)	1,248 (47.0)
5–9 years	5 (35.7)	6 (33.3)	11 (40.7)	14 (34.1)	36 (36.4)	829 (31.2)
10–14 years	3 (21.4)	4 (23.5)	5 (18.6)	10 (24.4)	22 (22.2)	577 (21.7)
Down's syndrome	0 (0)	0 (0)	0 (0)	1 (2.4)	1 (1.0)	41 (1.5)
Acute leukemia type						
ALL	14 (100.0)	13 (76.5)	23 (85.2)	30 (73.2)	80 (80.8)	2,179 (82.1)
B-cell precursor ALL	11 (78.6)	10 (58.8)	19 (70.4)	26 (63.4)	66 (66.7)	1,810 (68.2)
T-cell ALL	2 (14.3)	3 (17.6)	3 (11.1)	4 (9.8)	12 (12.1)	249 (9.4)
Other ALL	1 (7.1)	0 (0.0)	1 (3.7)	0 (0.0)	2 (2.0)	120 (4.5)
AML	0 (0.0)	4 (23.5)	3 (11.1)	10 (24.0)	17 (17.2)	407 (15.3)
Other AL	0 (0.0)	0 (0.0)	1 (3.7)	1 (2.0)	2 (2.0)	68 (2.6)

Abbreviations: AL: acute leukemia; ALL: acute lymphoblastic leukemia; AML: acute myeloblastic leukemia.

Sensitivity analyses were performed by excluding, in turn, each year or each NPP or by using a 6-year sliding window over the 1990–2007 period. The analyses with dose estimates were replicated using cutoffs based on the quartiles of the distribution in the controls.

To test for the existence of a trend in the incidence of AL with the distance from a NPP, the inverse distance function was considered. A test for linearity based on a log-likelihood ratio statistic was performed, considering the categorical variable derived from the 2-km-wide rings. The inverse distance was then included in the regression model as a continuous independent variable. The significance of the regression parameters was evaluated using 10,000 replications of the datasets. For the case–control analysis, 10,000 permutations of the case–control status of all the children were made by *Département* and independently of the distance from the NPPs. In the incidence studies, the 10,000 replicated datasets were built under the hypothesis of a Poisson distribution of the cases in the Communes.

Power of the studies

In the Geocap study using a one-sided test at the 5% level, the power to detect an OR of two for living less than 5 km from a NPP (compared to living more than 20 km) was equal to 43% for the 0- to 4-year-age group and 70% for the 0- to 14-year-age group. For the incidence studies, using an exact one-sided Poisson test at the 5% level and the approximation suggested by Breslow and Day,[33] the power to detect an SIR of two for living less than 5 km from a NPP (compared to living more than 20 km) was very close to that of the Geocap study for the period 2002–2007 and greater than 80% for the 1990–2001 and 1990–2007 periods.

Results

Childhood AL risk and proximity of NPP

Geocap case–control study, 2002–2007. Among the 2,753 cases included in the case–control study, 99 were living less than 20 km from a NPP. AL was significantly associated with living less than 5 km from a NPP (OR = 1.9 [1.0–3.2]), and ORs close to unity were observed for all the areas farther from the sites (Table 2). When the cases and controls were located in 2-km-wide rings instead of 5-km-wide rings, the logarithm of the OR tended to increase slightly with the inverse of the distance from the nearest NPP (Supporting Information Fig. S1), although this trend was not statistically significant when the inverse distance was considered a continuous variable in the regression model ($p_{\text{one-sided}} = 0.18$). For children less than 5 years old, the OR observed in the closest area was of the same order of magnitude as that for the whole group, although not significant (OR = 1.6 [0.7–4.1]). Very similar patterns were obtained for the 5- to 9-year-age and 10- to 14-year-age groups (data not shown).

The cases living in each of the 5-km rings around the NPPs presented with the usual age, gender and neoplastic cell characteristics (Table 3). The number of AML cases was very small (17 cases within 20 km, none within 5 km). The results for ALL were very similar to those for all AL, with the OR associated with living less than 5 km from a NPP being 2.4 [1.3–4.2] (Supporting Information Table S2).

Table 4. Distance to the nearest French nuclear power plant (NPP) and childhood acute leukemia: Geocap and Incidence studies, 2002–2007, different stratified analyses

	Geocap case–control study[1] (residence address)						Incidence study, 2002–2007[2] (town hall address)					
	<5 km *vs.* ≥20 km			<20 km *vs.* ≥20 km			<5 km			<20 km		
	Cases	Controls	OR [95% CI]	Cases	Controls	OR [95% CI]	O	E	SIR [95% CI]	O	E	SIR [95% CI]
Electrical power												
5,200–5,400 MW	3	22	1.4 [0.4–4.9]	28	303	0.9 [0.6–1.4]	3	1.8	1.7 [0.3–4.8]	27	26.2	1.0 [0.7–1.5]
3,600 MW	4	22	2.1 [0.7–6.3]*	28	325	1.0 [0.6–1.4]	3	1.9	1.6 [0.3–4.5]	27	31.6	0.9 [0.6–1.2]
2,600–2,800 MW	6	29	2.1 [0.8–5.1]#	33	345	1.1 [0.7–1.5]	7	3.0	2.3 [0.9–4.8]*	34	30.8	1.1 [0.8–1.5]
1,800 MW	1	7	1.4 [0.2–11.6]	10	87	1.3 [0.6–2.6]	1	0.6	1.7 [0.0–9.3]	11	7.1	1.5 [0.8–2.8]
Year												
2002	0	12	0.0	15	180	0.8 [0.5–1.5]	1	1.2	0.8 [0.0–4.5]	15	16.2	0.9 [0.5–1.5]
2003	3	15	2.1 [0.6–7.5]	18	182	1.2 [0.7–2.0]	3	1.2	2.4 [0.5–7.1]	17	16.2	1.1 [0.6–1.7]
2004	1	13	0.8 [0.1–6.4]	19	168	1.2 [0.7–2.0]	1	1.3	0.8 [0.0–4.4]	18	16.3	1.1 [0.7–1.7]
2005	1	13	0.8 [0.1–6.6]	14	161	1.1 [0.6–1.9]	0	1.2	0.0 [0.0–3.1]	17	15.5	1.1 [0.6–1.8]
2006	5	13	4.9 [1.6–14.8]*	18	175	1.1 [0.6–1.9]	5	1.2	4.1 [1.3–9.6]*	17	15.7	1.1 [0.6–1.7]
2007	4	14	3.9 [1.2–12.9]*	15	195	0.8 [0.4–1.4]	4	1.2	3.3 [0.9–8.4]*	15	15.7	1.0 [0.5–1.6]
Coastal location												
Yes	3	18	1.7 [0.5–6.0]	26	258	1.1 [0.7–1.3]	3	1.5	2.0 [0.4–5.9]	29	24.7	1.2 [0.8–1.7]
No	11	62	1.9 [1.0–3.6]*	73	802	1.0 [0.8–1.3]	11	5.9	1.9 [0.9–3.4]*	70	71.1	1.0 [0.8–1.2]
Border location												
Yes	4	19	1.9 [0.6–5.9]	17	187	0.9 [0.5–1.5]	4	1.3	3.0 [0.8–7.6]*	19	16.4	1.2 [0.7–1.8]
No	10	61	1.8 [0.9–3.6]	82	873	1.1 [0.8–1.3]	10	6.0	1.7 [0.8–3.1]#	80	79.3	1.0 [0.8–1.3]
Status of the Commune												
Rural	9	49	2.2 [1.0–4.7]*	49	445	1.2 [0.8–1.6]	9	3.9	2.3 [1.1–4.4]*	51	41.8	1.2 [0.9–1.6]#
Semiurban	3	16	2.2 [0.6–7.9]	26	388	0.8 [0.5–1.3]	3	2.5	1.2 [0.2–3.5]	26	34.6	0.8 [0.5–1.1]
Urban	2	15	1.5 [0.3–6.6]	24	227	1.3 [0.8–2.1]	2	0.9	2.1 [0.2–7.7]	22	19.4	1.1 [0.7–1.7]
Median income[3] of the Commune												
< Median	11	51	2.3 [1.2–4.4]*	63	703	1.0 [0.8–1.3]	11	4.6	2.4 [1.2–4.3]*	63	64.7	1.0 [0.8–1.3]
≥ Median	3	27	1.1 [0.3–3.6]	35	352	1.1 [0.7–1.6]	3	2.8	1.1 [0.2–3.2]	36	31.0	1.2 [0.8–1.6]
Proportion of blue-collar workers in the Commune												
< Median	2	15	1.6 [0.4–7.3]	18	175	1.2 [0.7–2.1]	2	1.4	1.5 [0.2–5.2]	18	15.8	1.1 [0.7–1.8]
≥ Median	12	65	1.9 [1.0–3.5]*	81	885	1.0 [0.7–1.2]	12	6.0	2.0 [1.0–3.5]*	81	79.9	1.0 [0.8–1.3]
Proportion of baccalaureate holders in the Commune												
< Median	13	70	1.9 [1.0–3.6]*	79	819	1.0 [0.8–1.3]	13	6.0	2.2 [1.2–3.7]*	78	71.5	1.0 [0.9–1.4]
≥ Median	1	10	1.4 [0.2–11.6]	20	241	1.0 [0.6–1.6]	1	1.4	0.7 [0.0–4.1]	21	21.3	0.9 [0.5–1.3]

[1]Odds ratios (ORs) and 95% confidence intervals (95% CIs) estimated by logistic regression adjusted for age at the end of the year (5-year age groups for the 0-14-year-old children, 1-year age groups for the 0-4-year-old children) and *Département* of residence. [2]Standardized incidence ratio (SIR) calculated as the ratio of the observed (O) to the expected (E) number of cases with Byar's approximation of 95% CI. [3]Not available for Communes of less than 11 taxed households and/or where a taxpayer represents more than 80% of income.
*$p_{one\text{-}sided} < 0.05$; #$p_{one\text{-}sided} < 0.10$.

Table 5. Cross-classification of the 30,000 controls (2002–2007) by the distance to the nearest French NPP and by the dose based geographic zoning (DBGZ)

Distance to the nearest NPP (km)	DBGZ[1]					Total
	Reference	>0 μSv and ≤0.093 μSv	0.094–0.20 μSv	0.21–0.72 μSv	>0.72 μSv	
0–4.99	0	5	5	38	39	87
5–9.99	0	33	50	93	43	219
10–14.99	0	119	104	62	11	296
15–19.99	1	260	158	20	4	443
≥20	28,955	0	0	0	0	28,955
Total	28,956	417	317	213	97	30,000

[1]The "DBGZ" is based on the estimated bone marrow radiation dose related to NPP gaseous discharge at the location of the Commune town hall and expressed in μSv. The "Reference" geographic zone of the DBGZ includes all the subjects who have a null estimated dose, considering that subjects of Communes, which town hall is located 20 km or farther away from any NPP, have a null estimated dose.

Incidence study, 2002–2007. Over the same period, 2002–2007, the incidence study included 2,831 AL cases less than 15 years old. The SIRs were very close to the OR of the contemporaneous case–control study for all the age groups and distance categories (Table 2). The logarithm of the SIR increased slightly with the inverse distance from the nearest NPP (Supporting Information Fig. 1); however, this trend was not statistically significant in the continuous model ($p_{\text{one-sided}} = 0.25$).

For the 1,159 subjects who lived within 20 km of the nearest NPP, the distances from NPP based on the coordinates of the Commune town hall used in the incidence study and on the individual coordinates used in the case–control study were highly correlated ($r = +0.97$). This finding is connected to the fact that most French NPPs are located in rural areas in which most of the dwellings are located close to the town hall.

Incidence study, 1990–2001. In the preceding period, 1990–2001, already covered by a previous analysis,[30] 5,356 AL cases less than 15 years old were registered in the NRCH. Among the cases, 173 lived less than 20 km from a NPP (Table 2). The SIR did not differ from one for any of the four 5-km rings around the NPPs.

Incidence study, 1990–2007. Over the whole period, 1990–2007, 272 of the 8,187 cases registered by the NRCH lived less than 20 km and 24 cases less than 5 km from a NPP. The SIR were 1.0 [0.9–1.1] and 1.1 [0.7–1.7] (Table 2). The SIR for the closest area was slightly but not significantly higher for the 0- to 4-year-age group (1.4 [0.8–2.3]; $p_{\text{one-sided}} = 0.15$).

Subgroup analyses, control of confounders and sensitivity analyses

The results were unchanged when the subjects living further than 50 km from a NPP were excluded from the case–control analysis. The SIR was very close to one (SIR = 1.05 [0.96–1.15]) for residences between 20 and 50 km from a NPP. The stratified analyses showed that the association between AL and living within 5 km of a NPP did not vary substantially with the power of the NPP, with location on a coast or border or with the urban/rural status of the Commune (Table 4). The increased risk with living less than 5 km from a NPP appeared more marked, although not significant, in Communes with the lowest median income or lowest proportion of baccalaureate holders than in other Communes.

Adjustments for the contextual variables either separately and jointly and for the deprivation index did not change the estimates. No case and five controls lived less than 5 km from a NPP and less than 200 m from a high-voltage power line, and excluding them did not substantially modify the association with proximity to NPP (OR = 2.0 [1.1–3.6]). Considering children living less than 600 m from the lines did not change the association with NPP either (OR = 2.1 [0.7–6.4] < 600 m and OR = 1.8 [0.9–3.5] further). Using the distance between the municipality and NPP rather than individual distances led to very similar OR and SIR.

The small numbers hampered the detailed analyses by NPP or year. No specific association with living less than 5 km from a given NPP was evidenced. When each NPP was excluded, in turn, from the case–control and incidence analyses, the OR/SIR estimated on the 18 remaining NPPs was very similar to that for the 19 NPPs (Supporting Information Table S3). There was no heterogeneity of the SIRs estimated by NPP ($p = 0.13$), and the test of the hypothesis that they were all equal to one was on the borderline of statistical significance ($p = 0.07$).

Over the period, 2002–2007, AL cases living less than 5 km from a NPP were mostly diagnosed in 2003, 2006 and 2007 (Table 4). However, the results remained similar when each year of observation was excluded, in turn, from the case–control and incidence analyses. For the whole period, 1990–2007, no significant heterogeneity of the annual SIRs was evidenced ($p = 0.12$); however, the test of the hypothesis that the SIRs were all equal to one was on the borderline of significance ($p = 0.06$). When the incidence analyses were restricted to successive 6-year sliding windows over the whole

Table 6. Association between DBGZ[1] around the French NPPs and childhood acute leukemia: Geocap case–control study and incidence studies

	Case–control study, 2002–2007[2] (distance: municipality-nearest NPP)			Incidence study, 2002–2007[3] (distance: municipality-nearest NPP)			Incidence study, 1990–2001[3] (distance: municipality-nearest NPP)			Incidence study 1990–2007[3] (distance: municipality-nearest NPP)		
	Cases	Controls	OR [95% CI]	O	E	SIR [95% CI]	O	E	SIR [95% CI]	O	E	SIR [95% CI]
Age < 15 years[4]												
>0.72 μSv[1]	8	97	1.0 [0.5–2.1]	8	8.3	1.0 [0.4–1.9]	13	16.4	0.8 [0.4–1.4]	21	24.7	0.9 [0.5–1.3]
0.21–0.71 μSv	19	213	1.0 [0.6–1.6]	20	18.5	1.1 [0.7–1.7]	42	36.7	1.1 [0.8–1.6]	62	55.2	1.1 [0.9–1.4]
0.094–0.20 μSv	29	317	1.0 [0.7–1.5]	31	30.0	1.0 [0.7–1.5]	54	56.8	1.0 [0.7–1.2]	85	86.7	1.0 [0.8–1.2]
>0 μSv and ≤ 0.093 μSv	40	417	1.0 [0.7–1.4]	40	39.0	1.0 [0.7–1.4]	64	73.1	0.9 [0.7–1.1]	104	112.0	0.9 [0.8–1.1]
Reference	2,657	28,956	1.0									
<20 km and >0 μSv	96	1,044	1.0 [0.8–1.3]	99	95.7	1.0 [0.8–1.3]	173	182.9	1.0 [0.8–1.1]	272	278.6	1.0 [0.9–1.1]
Reference	2,657	28,956	1.0									
Age < 5 years[4]												
> 0.72 μSv[1]	4	28	1.1 [0.4–3.2]	5	4.0	1.2 [0.4–2.9]	7	8.0	0.9 [0.4–1.8]	12	12.0	1.0 [0.5–1.8]
0.21–0.71 μSv	6	59	0.9 [0.4–2.0]	8	9.2	0.9 [0.4–1.7]	23	17.9	1.3 [0.8–1.9]	31	27.1	1.1 [0.8–1.6]
0.094–0.20 μSv	15	88	1.3 [0.7–2.3]	16	14.9	1.1 [0.6–1.7]	29	27.9	1.0 [0.7–1.5]	45	42.8	1.1 [0.8–1.4]
>0 μSv and ≤0.093 μSv	16	137	0.9 [0.5–1.5]	18	19.1	0.9 [0.6–1.5]	25	35.3	0.7 [0.5–1.1]	43	54.5	0.8 [0.6–1.1]
Reference	1,248	9,396	1.0									
< 20 km and >0 μSv	41	312	1.0 [0.7–1.4]	47	47.3	1.0 [0.7–1.3]	84	89.1	1.0 [0.8–1.1]	131	136.4	1.0 [0.8–1.1]
Reference	1,248	9,405	1.0									

[1]The "DBGZ" is based on the estimated bone marrow radiation dose related to NPP gaseous discharge at the location of the Commune town hall and expressed in μSv. The "Reference" geographic zone of the DBGZ includes all the subjects who have a null estimated dose, considering that subjects of Communes, which town hall is located 20 km or farther away from any NPP, have a null estimated dose. [2]Odds ratios (ORs) and 95% confidence intervals (95% CIs) estimated by logistic regression adjusted for age at the end of the year (5-year-age groups for the 0- to 14-year-old children and 1-year-age groups for the 0- to 4-year-old children) and *Département* of residence. [3]Standardized incidence ratio (SIR) calculated as the radio of the observed (O) to the expected (E) number of cases with Byar's approximation of the 95% CI. [4]Age at the end of the year in the Geocap study; exact age in the incidence studies.

1990–2007 period, 2002–2007 was the only period for which an association with living less than 5 km from a NPP was observed. The use of simulations resulted in similar values of the statistics and the same conclusions.

Childhood AL and DBGZ

The estimated bone marrow doses related to radioactive gaseous discharge did not result in the same categorization of the cases and controls (Table 5): 40% of the controls in the highest exposure category (>0.72 μSv) were in the 0- to 4.99-km ring, 44% in the 5- to 9.99-km ring; 11% in the 10- to 14.99-km ring and 4% in the 15- to 19.99-km ring.

No association between AL and DBGZ was observed in the case–control study or in the 2002–2007 incidence study (Table 6). The SIRs and ORs were close to one for all the DBGZ categories. The results were the same when the DBGZ categories were based on the quartiles of the control exposure distribution (not shown). Exclusion of the NPPs close to other nuclear sites did not change the results. There was no association between AL and DBGZ for the whole period, 1990–2007 (Table 6).

Discussion

The Geocap case–control study evidenced an association between childhood AL and living less than 5 km from a NPP for the 2002–2007 period. The association was also observed in the contemporaneous incidence study, but not for the previous period, 1990–2001. The use of DBGZ yielded very different results, with SIRs and ORs close to one for all the DBGZ categories. The association observed for 2002–2007 was not specific to any age group, NPP or year.

One strength of the study resides in the fact that the cases were identified by the NRCH, which has covered the entire country since 1990. The NRCH relies on about three independent notifications per case on average. Its exhaustiveness has been estimated to be 99.4%.[38] A further strength consists in the fact that the controls were selected from the nearly exhaustive database of taxpayer households with children. As illustrated by Table 1, the controls in the Geocap study were highly representative of the source population. Another strength of the Geocap study is the precise geocoding of the residences of cases and controls. The two complementary approaches used are sensitive to different potential sources of biases, selection of controls for the case–control study and

estimates of expected number of cases for the ecological study based on incidence data aggregated at the Commune level. The fact that both approaches resulted in almost identical findings is a strong argument in favor of the validity of the ORs and SIRs.

Adjustment for, and stratification on, age and the socioeconomic and demographic characteristics of the Commune of residence were assumed to cancel out a number of potential contextual confounders.

A limitation of our study is that the data did not enable adjustment for individual potential risk factors such as birth order, breastfeeding, day-care attendance or pesticide exposure. However, there is no obvious reason for these factors to differ within and outside the 5-km rings close to the NPPs conditionally on rural/urban status or other socioeconomic or demographic variables that were taken into account. Data on parental employment at the NPPs were not available in the Geocap study; however, an impact of parental employment on childhood leukemia is not supported by the literature.[39] Also, population mixing could not be evaluated in this study. However, the Geocap study allowed to account for adjustment on the proximity of high-voltage power lines.

Like most studies of childhood leukemia in the neighborhood of NPPs, the Geocap study did not have access to complete residential histories, which is an important limitation for the evaluation of the true exposure to radiation or any factor related to the proximity of NPPs. However, neither the Finnish study that collected complete residential history and computed the distance from a NPP weighted by the time spent in the house[9] nor the Swiss study that used the addresses at birth and diagnosis[6] revealed an association with past or cumulative proximity to NPP.

Restricting the analyses to children aged less than 5 years, who are less likely to have moved house since birth than older children, resulted in similar ORs and SIRs. Three NPPs are located close to a French border, but only the French part of the 20-km radius was considered. However, excluding the three NPPs did not markedly change the results. None of the NPPs in neighboring countries was less than 20 km from the French border.

The method used to estimate the dose of radiation delivered to bone marrow by the NPPs was based on the average annual gaseous discharge levels, discharge composition and local meteorological parameters especially prevailing winds that influence the dispersion of radionuclides. Because of the lack of data on real discharge rates, the discharge rates for carbon 14 were based on the regulatory limits. Therefore, the contribution of carbon 14 to the radiation dose may have been overestimated. However, the method was the same for all the NPPs. In consequence, the Communes are likely to have been correctly ranked within a given NPP area in terms of exposure despite the potential errors in dose estimates. Contrary to a classic assumption, Table 5 clearly illustrates that the distribution of the population in terms of the dispersion of radionuclides released into the atmosphere cannot be represented by a simple function of the distance from the NPP. The use of the innovative DBGZ appears to be one step toward overcoming the limitations of studies using conventional circular zoning.[40] The model did not take into account liquid discharges, although their contribution to the total radiation exposure related to NPP activity may have been of the same order of magnitude as that of the gaseous discharges. The radiation doses due to liquid discharges are mainly determined by individual behaviors (food consumption and water use) and are therefore not expected to comply with geographic zoning around NPPs. Furthermore, in many situations, the doses are not expected to decrease smoothly with distance from the NPP. In addition, although the NPPs located on a coast or river may be expected to differ with regard to doses from liquid discharges, in the stratified analyses, the associations with AL were similar. Overall, the estimated doses due to NPPs were very low compared to the doses due to natural radiation sources. Such doses are not expected to result in an observable excess risk on the basis of the available evidence.[41] Compared to other studies, the use of DBGZ constitutes in the authors' opinion, a major improvement. The approach used to derive DBGZ is already quite elaborated (consideration of a broad spectrum of 12 radionuclides, use of real average discharge data and local climate data, calculation of the dose to the pertinent organ, *etc.*). In future developments, DBGZ could be refined by including real discharge data for carbon 14 in the models for gaseous discharges and by accounting for the impact of liquid discharges in the determination of dose levels. Also, DBGZ could provide estimates at the place of residence rather than at the town hall. However, although DBGZ could gain in accuracy and precision, the estimates are not expected to increase to the dose ranges that could exert a noticeable effect on leukemia risk according to predictive models.

In the authors' previous multisite incidence studies,[29,30] no association between proximity to NPPs and AL was observed. This was in line with most multisite studies[1,2,8,12] and is also in line with the results of the authors' incidence analysis over the whole period, 1990–2007. In Germany, the KiKK case–control study of children aged less than 5 years evidenced an association between AL and NPP proximity.[25] However, there are several important differences between those results and the results of our study. The German incidence study showed that incidence rates were higher specifically in the 0- to 4-year-age group,[42] which was not the case in this study. In the German study, the risk estimates obtained in the incidence analysis[26] also appeared to be lower than those obtained with the case–control approach[25]; however, in our study, the estimates obtained with the two approaches were very similar. In the German study, the estimated risk in the 5-km ring was highly sensitive to whether or not the Krummel NPP was included,[8] whereas no noteworthy difference between the NPPs was observed in our study. A reassessment of the KiKK results showed the marked impact of the urban/rural status of the residence area

on the estimated risk,[43] whereas no noteworthy difference was observed in our study. In the KiKK study, an increasing trend with the inverse distance from the sites, considered as a continuous variable, was reported; the trend was not detected when the distance was categorical.[25] In the Geocap study, a slight but nonsignificant increasing trend of the OR and SIR with inverse distance was observed.

Overall, the results suggest a possible excess risk of AL in the close vicinity of French NPPs in 2002–2007. The increased incidence observed at less than 5 km from the NPPs in the Geocap study only partially supports the recent German findings as the increase was limited to recent years and was not specific to the youngest children. The absence of any association with DBGZ, which is assumed to reflect the distribution of gaseous radiation discharged from NPPs, may indicate that the association observed with distance <5 km over 2002–2007 and particularly in 2006–2007 is not explained by NPP gaseous discharges. Overall, the results suggest a potential excess risk over 2002–2007 that may be due to unknown factors related to the proximity of NPPs. Among the potential factors are population mixing and exposures to physical agents, including natural or man-made exposures to radiation not modeled by the DBGZ. Overall, the findings call for investigation for potential risk factors related to the vicinity of NPP and for collaborative analysis of all the evidence available from multisite studies conducted in various countries.

Acknowledgements

The authors are particularly grateful to Olivier Lamy and Mathieu Carrère (GEOCIBLE), who carefully ensured all the geocoding, and the Institut Géographique National (IGN), which made precise maps available for the whole country. The authors are also grateful to Magda Tomasini and Laurent Auzet (INSEE), who conducted the control sampling, and Aline Morin (IRSN), who conducted the DBGZ, all the NRCH research assistants who collected the cases' addresses and the pediatric oncology teams for their help in data collection.

References

1. Bithell JF, Keegan TJ, Kroll ME, Murphy MFG, Vincent TJ. Childhood leukaemia near British nuclear installations: methodological issues and recent results. *Radiat Prot Dosimetry* 2008;132:191–7.
2. Laurier D, Jacob S, Bernier MO, Leuraud K, Metz C, Samson E, Laloi P. Epidemiological studies of leukaemia in children and young adults around nuclear facilities: a critical review. *Radiat Prot Dosimetry* 2008;132:182–90.
3. Committee on Medical Aspects of Radiation in the Environment (COMARE). Tenth report. The incidence of childhood cancer around nuclear installations in Great Britain. *J Radiol Prot* 2005;25: 335–6.
4. Michaelis J, Keller B, Haaf G, Kaatsch P. Incidence of childhood malignancies in the vicinity of west German nuclear power plants. *Cancer Causes Control* 1992;3: 255–63.
5. McLaughlin JR, Clarke EA, Nishri ED, Anderson TW. Childhood leukemia in the vicinity of Canadian nuclear facilities. *Cancer Causes Control* 1993;4:51–8.
6. Spycher BD, Feller M, Zwahlen M, Röösli M, von der Weid NX, Hengartner H, Egger M, Kuehni CE. Childhood cancer and nuclear power plants in Switzerland: a census-based cohort study. *Int J Epidemiol* 2011;40:1247–60.
7. Jablon S, Hrubec Z, Boice JDJ. Cancer in populations living near nuclear facilities. A survey of mortality nationwide and incidence in two states. *JAMA* 1991;265: 1403–8.
8. Committee on Medical Aspects of Radiation in the Environment (COMARE). Fourteenth report. available on http://www.comare.org.uk/press_releases/documents/COMARE14report.pdf. Further consideration of the incidence of childhood leukaemia around nuclear power plants in Great Britain. Health Protection Agency, Chilton 2011.
9. Heinävaara S, Toikkanen S, Pasanen K, Verkasalo PK, Kurttio P, Auvinen A. Cancer incidence in the vicinity of Finnish nuclear power plants: an emphasis on childhood leukemia. *Cancer Causes Control* 2010;21:587–95.
10. Kaatsch P, Kaletsch U, Meinert R, Michaelis J. An extended study on childhood malignancies in the vicinity of German nuclear power plants. *Cancer Causes Control* 1998;9:529–33.
11. Mangano JJ, Sherman J, Chang C, Dave A, Feinberg E, Frimer M. Elevated childhood cancer incidence proximate to U.S. nuclear power plants. *Arch Environ Health* 2003; 58:74–82.
12. Laurier D, Hémon D, Clavel J. Childhood leukaemia incidence below the age of 5 years near French nuclear power plants. *J Radiol Prot* 2008;28:401–3.
13. Waller LA, Turnbull BW, Gustafsson G, Hjalmars U, Andersson B. Detection and assessment of clusters of disease: an application to nuclear power plant facilities and childhood leukaemia in Sweden. *Stat Med* 1995;14:3–16.
14. Goldsmith JR. Nuclear installations and childhood cancer in the UK: mortality and incidence for 0–9-year-old children, 1971–1980. *Sci Total Environ* 1992;127:13–35, 43–55.
15. Sharp L, Black RJ, Harkness EF, McKinney PA. Incidence of childhood leukaemia and non-Hodgkin's lymphoma in the vicinity of nuclear sites in Scotland, 1968–93. *Occup Environ Med* 1996;53:823–31.
16. Bithell JF, Dutton SJ, Draper GJ, Neary NM. Distribution of childhood leukaemias and non-Hodgkin's lymphomas near nuclear installations in England and Wales. *BMJ* 1994;309:501–5.
17. Heasman M, Kemp I, Urquhart J, Black R. Childhood leukaemia in northern Scotland. *Lancet* 1986;1:266.
18. Black RJ, Sharp L, Harkness EF, McKinney PA. Leukaemia and non-Hodgkin's lymphoma: incidence in children and young adults resident in the Dounreay area of Caithness, Scotland in 1968–91. *J Epidemiol Community Health* 1994;48:232–6.
19. Hoffmann W, Dieckmann H, Dieckmann H, Schmitz-Feuerhake I. A cluster of childhood leukemia near a nuclear reactor in northern Germany. *Arch Environ Health* 1997;52:275–80.
20. Hoffmann W, Terschueren C, Richardson DB. Childhood leukemia in the vicinity of the Geesthacht nuclear establishments near Hamburg, Germany. *Environ Health Perspect* 2007;115:947–52.
21. Kinlen L. Childhood leukaemia, nuclear sites, and population mixing. *Br J Cancer* 2011;104:12–18.
22. Kinlen L. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* 1995;71:1–5.
23. Kinlen LJ, Dickson M, Stiller CA. Childhood leukaemia and non-Hodgkin's lymphoma near large rural construction sites, with a comparison with Sellafield nuclear site. *BMJ* 1995;310:763–8.
24. Dickinson HO, Parker L. Quantifying the effect of population mixing on childhood leukaemia risk: the Seascale cluster. *Br J Cancer* 1999;81:144–51.
25. Kaatsch P, Spix C, Schulze-Rath R, Schmiedel S, Blettner M. Leukaemia in

young children living in the vicinity of German nuclear power plants. *Int J Cancer* 2008;122:721–6.
26. Kaatsch P, Spix C, Jung I, Blettner M. Childhood leukemia in the vicinity of nuclear power plants in Germany. *Dtsch Arztebl Int* 2008;105:725–32.
27. Spix C, Schmiedel S, Kaatsch P, Schulze-Rath R, Blettner M. Case-control study on childhood cancer in the vicinity of nuclear power plants in Germany 1980–2003. *Eur J Cancer* 2008;44:275–84.
28. Kinlen L. A German storm affecting Britain: childhood leukaemia and nuclear power plants. *J Radiol Prot* 2011;31: 279–84.
29. White-Koning ML, Hémon D, Laurier D, Tirmarche M, Jougla E, Goubin A, Clavel J. Incidence of childhood leukaemia in the vicinity of nuclear sites in France, 1990–1998. *Br J Cancer* 2004;91:916–22.
30. Evrard A, Hémon D, Morin A, Laurier D, Tirmarche M, Backe J, Chartier M, Clavel J. Childhood leukaemia incidence around French nuclear installations using geographic zoning based on gaseous discharge dose estimates.. *Br J Cancer* 2006;94:1342–7.
31. Lacour B, Guyot-Goubin A, Guissou S, Bellec S, Désandes E, Clavel J. Incidence of childhood cancer in France: national children cancer registries, 2000–2004. *Eur J Cancer Prev* 2010;19:173–81.
32. Morin A, Backe J. Programme environnement et santé 1999. Une estimation de l'exposition du public due aux rejets radioactifs des centrales nucléaires. *Note Technique SEGR/SAER/02–51 Indice 1. Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, Juillet 2002.*
33. Breslow NE, Day NE. Statistical methods in cancer research. Volume II: The design and analysis of cohort studies. *IARC Sci Publ* 1987;2:1–406.
34. Bellec S, Baccaïni B, Goubin A, Rudant J, Ripert M, Hémon D, Clavel J. Childhood leukaemia and population movements in France, 1990–2003. *Br J Cancer* 2008;98: 225–31.
35. Faure C, Mollié A, Bellec S, Guyot-Goubin A, Clavel J, Hémon D. Geographical variations in the incidence of childhood acute leukaemia in France over the period 1990–2004. *Eur J Cancer Prev* 2009;18: 267–79.
36. Greenland S, Sheppard AR, Kaune WT, Poole C, Kelsh MA. A pooled analysis of magnetic fields, wire codes, and childhood leukemia. childhood leukemia-EMF study group. *Epidemiology* 2000;11:624–34.
37. Ahlbom A, Day N, Feychting M, Roman E, Skinner J, Dockerty J, Linet M, McBride M, Michaelis J, Olsen JH, Tynes T, Verkasalo PK. A pooled analysis of magnetic fields and childhood leukaemia. *Br J Cancer* 2000;83:692–98.
38. Clavel J, Goubin A, Auclerc MF, Auvrignon A, Waterkeyn C, Patte C, Barrachel A, Leverger G, Nelken B, Philippe N, Sommelet D, Vilmer E, et al. Incidence of childhood leukaemia and non-Hodgkin's lymphoma in France: national registry of childhood leukaemia and lymphoma, 1990–1999. *Eur J Cancer Prev* 2004;13: 97–103.
39. Committee on Medical Aspects of Radiation in the Environment (COMARE). Seventh report. available on http://www.comare.org.uk/reports/comare7threport.pdf. Parents occupationally exposed to radiation prior to the conception of their children. A review of the evidence concerning the incidence of cancer in their children (Chairman: Professor B. A. Bridges Obe). National Radiological Protection Board, 2002.
40. Wing S, Richardson DB, Hoffmann W. Cancer risks near nuclear facilities: the importance of research design and explicit study hypotheses. *Environ Health Perspect* 2011;119:417–21.
41. United Nations Scientific Committee on the Effects of Atomic Radiation UNSCEAR Report 2000, Sources and effects of ionizing radiation, vol. I Sources, Annex B Exposure from natural radiation sources. United Nations, 2000;83–156.
42. Grosche B. The 'kinderkrebs in der umgebung von kernkraftwerken' study: results put into perspective. *Radiat Prot Dosimetry* 2008;132:198–201.
43. Strahlenschutzkommission (SSK). Bewertung der epidemiologischen studie zu kinderkrebs in der umgebung von kernkraftwerken (KiKK study). *Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, Heft* 2009;58. available on http://www.ssk.de/de/werke/2008/volltext/ssk0815.pdf.

Résumé

Le rôle de l'environnement dans l'étiologie des leucémies aigües de l'enfant (LA) fait aujourd'hui l'objet de recherches intenses. Dans ce contexte, le présent travail a pour objectif d'étudier la relation entre l'incidence de LA et la proximité des centrales nucléaires de production d'électricité (CNPE) et des lignes à haute tension (LHT). Avant cette analyse fine, un premier travail a consisté à étudier les variations départementales de l'incidence de LA.

Les cas inclus dans ces études sont toutes les LA du Registre National des Hémopathies malignes de l'Enfant sur la période étudiée : 1990-2004 pour l'étude de l'incidence départementale et 2002-2007 pour les études de l'association avec les facteurs d'exposition environnementale. Dans l'approche cas-témoins principalement utilisée pour ces dernières, les 30 000 sujets témoins constitue un échantillon représentatif de la population pédiatrique française sur la période d''intérêt. D'autre part, la géolocalisation des adresses des sujets et des sources d'exposition permet de définir des critères de proximité en relation avec la probabilité et/ou l'intensité d'exposition aux facteurs d'intérêt.

• L'étude des LA par département n'a pas mis en évidence de tendance ni de structure spatiale dans l'incidence à ce niveau géographique : que ce soit globalement, par classe d'âge, par sexe ou par sous-type de leucémie.

• Sur la période 2002-2007 contrairement aux périodes précédentes, un quasi-doublement de l'incidence des LA à moins de 5 km des CNPE a été mis en évidence, avec une approche cas-témoin comme avec l'étude d'incidence. Ce résultat n'était pas spécifique d'une CNPE ou d'un type de CNPE et non lié à la cartographie des émissions aériennes de radioactivité par les CNPE.

• L'association trouvée entre l'incidence de LA et la proximité aux LHT de plus de 225 kV (<50 m) semble restreinte aux enfants de moins de 5 ans ou n'habitant en milieu urbain ; aucune association n'a été trouvée avec la proximité aux LHT de moins de 150 kV.

Abstract

The role of the environment in the etiology of childhood acute leukemia (AL) is currently investigated. In this context, the aim of the present work is to study the association between the incidence of AL and the proximity no nuclear power plants (NPP) and to high voltage overhead power lines (HV OLs). At first, the geographical variations of AL have been studied at the *Département* level.

The cases included in the studies are all cases of AL of the French National Registry of Childhood Haemopatopoietic Malignancies on the studied periods: 1990-2004 for the study of incidence on *Départements* and 2002-2007 for the studies of association between incidence of AL and environmental exposure factors. Concerning those latter studies, a case-control approach has been used. The control sample, representative of the French pediatric population, contains 30,000 subjects and has been drawn by the INSEE. The precise localization of addresses of subjects and of exposure sources in relation with the type of sources is essential to build indicators of exposure reflecting the probability and intensity of exposure.

• The study of AL by *Département* has highlighted neither trend nor spatial structure in the incidence at this geographical level globally as well as by age, gender and subtype of leukemia.

• On 2002-2007, on the contrary of on previous periods, the incidence of AL at less than 5 km from a NPP was nearly twice higher than expected, with the case-control study as well as with the incidence approach. This result was not specific to any age group, NPP, a type of NPP and was not associated with the geographic zoning of gaseous discharges of NPPs.

• The study of the proximity to HV OLs highlighted an association between the incidence of AL and the close proximity (< 50 m) of lines of more than 225 kV, association which was restricted to children of less than 5 y.o. or living in non-urban areas; but not with the proximity to lines of less than 150 kV.

Laboratoire d'accueil :

INSERM CESP (Centre de: recherche en Epidémiologie et Santé des Populations) U1018
EQ6 – Epidémiologie environnementale des cancers
16 avenue Paul Vaillant Couturier
94807 Villejuif Cedex