LEUCEMIES DE L'ENFANT ET ENVIRONNEMENT – programme GEOCAP

Jacqueline Clavel[1], Claire Faure-Sermage et Denis Hémon
INSERM, unité d'Épidémiologie environnementale des cancers, UMR-S 754, Université Paris Sud 11
Registre National des Hémopathies Malignes de l'Enfant
16 avenue Paul Vaillant-Couturier – F-94807 VILLEJUIF cedex

1. Objectifs et situation du sujet,

L'objectif du programme Geocap est d'étudier l'influence de plusieurs expositions environnementales sur le risque de cancer de l'enfant.

Le projet porte dans un premier temps sur les leucémies de l'enfant et cible plus particulièrement :
- les lignes à haute tension, à l'origine d'expositions aux champs électromagnétiques à extrêmement basses fréquences ;
- le trafic routier, et la pollution atmosphérique au benzène et aux hydrocarbures polycycliques aromatiques ;
- la proximité des sites nucléaires ;
- la concentration locale en radon ;
- la proximité de sites industriels classés Seveso ;
- la proximité des stations-service.

Geocap s'intègre dans le programme de recherche sur les cancers de l'enfant de l'U754-Inserm

2. Matériel et méthodes,

2.1 Echantillonnage

- **Echantillon des cas**

GEOCAP est une étude cas-témoin qui couvre la France métropolitaine et utilise les informations relatives aux 12 000 cas enregistrés par le Registre National des Hémopathies malignes de l'Enfant (RNHE) depuis 1990 et par le Registre National des Tumeurs Solides de l'Enfant (RNTSE) depuis 2000. Ces 2 registres, qualifiés par le Comité National des Registres, fonctionnent de façon conjointe et sont soutenus par l'InVS et l'INCa.

Dans un premier temps, nous travaillons sur les leucémies aiguës. Les cas, issus du Registre National des Hémopathies malignes de l'Enfant (RNHE), sont tous les enfants de moins de 15 ans résidant en France métropolitaine et atteints d'une leucémie depuis le 1/1/1990, soit 6 640 cas sur 1990-2004 dont 1 400 sur la période 2002-2004. L'inclusion des cas de la période 2005-2007 est en cours. L'exhaustivité du RNHE avait été évaluée à 99% par capture recapture sur la période 1990-1996 et un bais de sélection à l'identification des cas est donc très peu probable[2].

- **Echantillon des témoins**

Les témoins sont 15 000 enfants de moins de 15 ans échantillonnés en 2002, 2003 et 2004 par l'Insee, selon une procédure assurant un total anonymat des enfants inclus. Malheureusement, l'Insee ne dispose pas des données permettant de constituer de tels échantillons sur les années 1990-2001, c'est à dire depuis le début de l'enregistrement des cas. La constitution d'un nouvel échantillon de 15 000 enfants sur la période 2005-2007 est en cours pour renforcer la puissance de nos investigations. Ces échantillons sont représentatifs de la population générale pédiatrique

---

[1] jacqueline.clavel@inserm.fr
[2] 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;13:97-103.

métropolitaine en termes d'âge, de nombre d'enfants de moins de 15 ans vivant au foyer et de degré d'urbanisation de la commune de résidence (tableau 1). Cette représentativité est respectée au niveau départemental.

Tableau 1 : Répartition de la population source et des témoins selon, l'âge atteint au 31 décembre, le nombre d'enfant du ménage et le type d'unité urbaine de la commune de résidence

	Population source		Echantillon Geocap	
	n	%	n	%
Age				
0-2 ans	6508047	19,6%	2842	19,0%
3-5 ans	6676939	20,2%	3000	20,0%
6-8 ans	6532286	19,7%	2965	19,8%
9-11 ans	6521195	19,7%	3015	20,1%
12-14 ans	6901748	20,8%	3178	21,2%
Nombre d'enfants du ménage				
1	9470170	28,6%	4245	28,3%
2	14220030	42,9%	6527	43,5%
3	6590544	19,9%	3012	20,1%
≥ 4	2859476	8,6%	1216	8,1%
Type d'unité urbaine de la commune de résidence				
Rural	8424156	25,4%	3775	25,2%
< 5000 hab	2074149	6,3%	895	6,0%
5 000-9 999 hab	1800015	5,4%	827	5,5%
10 000-19 999 hab	1670255	5,0%	743	5,0%
20 000-49 999 hab	2001496	6,0%	927	6,2%
50 000-99 999 hab	2216141	6,7%	958	6,4%
100 000-199 999 hab	1669872	5,0%	795	5,3%
200 000-1 999 999 hab	7274607	22,0%	3367	22,5%
Unité urbaine de Paris	6009529	18,1%	2713	18,1%
TOTAL	33140220	100,0%	15000	100,0%

La représentativité de l'échantillon a également été vérifiée pour certaines variables socio-économiques des communes de résidence des sujets de l'échantillon : revenu médian, part de résidents propriétaires, part de personnes vivant en HLM, nombre moyen de personnes par pièce, part d'ouvriers dans la population active, part de bacheliers dans la population active et le taux de chômage. Dans tous les cas l'échantillon témoin constitué s'est avéré étroitement comparable à sa population source.

2.2 Recueil de données

Les données disponibles à la fois pour les cas et les témoins sont l'âge, l'adresse exacte, le nombre d'enfants au foyer, et les variables INSEE dérivées de la commune de résidence. Ces données sont obtenues à la date du recrutement (date du diagnostic pour les cas).

- **Géocodage des adresses des cas et des témoins**

Afin d'éviter tout biais différentiel de classement, l'exposition des sujets est estimée à l'aveugle de leur statut cas/témoins.

Le géocodage et une partie du traitement des données sont assurés par la société Géocible. Le géocodage des adresses comporte une procédure automatique, suffisante pour les adresses bien référencées dans la base Navteq. La collaboration de l'Institut Géographique National (IGN) a permis d'améliorer la précision de codage d'environ 35% des adresses de notre échantillon, situées dans des rues référencées de façon insuffisamment précise (numéro de rue absent de la base mais segment de rue géolocalisé, notamment). Les adresses qui ne sont pas retrouvées automatiquement sont traitées manuellement : recherche de rues auprès des mairies, positionnement manuel sur les cartes au 25 000ème de lieux-dits non référencés.

- **Expositions environnementales**

Les sources d'exposition environnementale provenant de bases de données validées ont été géocodées et les expositions sont quantifiées en collaboration avec les spécialistes du domaine.

- *Sites nucléaires*

La publication des travaux allemands rapportant un lien entre l'incidence des leucémies de l'enfant de moins de 5 ans et la proximité des sites nucléaires (Kaatsch 2007), nous a conduit à reprendre cette question en priorité.

Nous travaillons donc tout d'abord sur l'incidence des leucémies dans les communes proches des sites nucléaires français sur la période 1990-2004. Plusieurs études d'incidence avaient déjà été faites par notre équipe sur les périodes 1990-1998 [White-Koning, 2004] et 1990-2001 [Evrard, 2006], utilisant la localisation des communes par rapport aux sites (zonages circulaires concentriques de 5 à 30 km de rayon) et l'exposition des communes aux radiations des rejets gazeux modélisées par l'IRSN en fonction des caractéristiques des sites et des vents dominants. Ces méthodologies sont utilisées à nouveau dans l'étude cas-témoins GEOCAP sur la période 1990-2004 : pour la distance de la résidence des sujets au moment de l'enquête (au diagnostic pour les cas) aux barycentres des sites, et pour déterminer l'exposition de leur résidence aux rejets estimés par modélisation. La définition des sites nucléaires retenus dans cette nouvelle analyse est en cours dans le cadre d'un groupe de travail réuni sous l'égide de l'Autorité de Sûreté Nucléaire. Elle sera vraisemblablement un peu plus large que celle que nous avions retenue pour nos analyses précédentes.

- *Lignes à haute tension et CEM-EBF*

Les CEM-EBF dépendent du nombre, du type et de la configuration des lignes à proximité des habitations étudiées, de l'intensité du courant qui parcourt la ligne ainsi que des distances de l'habitation à la ligne et au pylône le plus proche. Des zones d'influence globales des lignes ont été fixées en fonction du voltage des lignes, l'exposition due à la ligne étant considérée comme négligeable quand elle est équivalente au bruit de fond, c'est-à-dire inférieure à 0,1 µT. Nous travaillons en collaboration avec le Réseau de Transport d'Electricité (RTE) dans la prise en compte de ces paramètres et la détermination la plus précise possible de l'exposition de chaque sujet.

- *Trafic et pollution atmosphérique*

Nous avons recueilli des données de trafic moyen journalier annualisé pour les autoroutes et routes nationales (30 000 km de voies) et de densité de trafic communale pour les routes départementales (365 000 km de voies) en 2000. Nous avons également recueilli des concentrations de polluants dans l'air entre 1999 et 2004 auprès de l'Agence de l'Environnement et de la Maîtrise de l'Energie. Par ailleurs, nous travaillons avec AIRPARIF (association pour la surveillance de la qualité de l'air en Ile de France) pour caractériser plus finement l'exposition d'abord en Ile de France puis sur l'ensemble de la métropole.

- *Radon*

Les données concernant l'exposition à la radioactivité d'origine naturelle, radon et rayonnement gamma tellurique, ont été produites par l'IRSN (Institut de Radioprotection et de Sûreté Nucléaire) lors de ses campagnes de mesures nationales (complétées pour les rayonnement gamma, par un réseau de capteurs permanents) et cartographiées à l'échelle du département [Billon et al. 2005] . Nous avons analysé ces données dans le cadre de nos précédentes études épidémiologiques [Evrard et al. 2005] [Evrard et al. 2006]. Nous ajouterons à ces données celles d'une cartographie du potentiel radon des formations géologiques à l'échelle 1/1 000 000, en cours de réalisation à l'IRSN.

- *Stations-service*

Les sources de données CATALIST, BASIAS, APUR ne sont pas exhaustives et ne comportent généralement pas d'historique. Des recoupements sont en cours pour exploiter au mieux ces bases de données. L'estimation de l'exposition au benzène autour des stations service se fera en premier lieu à l'aide de la distance des habitations aux stations, du degré d'urbanisation de la commune et des vents dominants. Une étude visant à évaluer le taux de benzène autour de stations service est en cours de réalisation par l'AFSSET (Agence Française de Sécurité Sanitaire de l'Environnement et du Travail). Les résultats de cette étude pourront nous aider à mieux estimer par la suite cette exposition.

- *Sites industriels*

Cette partie du projet est encore embryonnaire. La localisation des sites industriels rejetant de la dioxine et des PCB est réalisée en lien avec le ministère du développement durable et l'iREP (Registre français des Emissions Polluantes).

- **Evaluation des expositions environnementales**

Deux approches complémentaires sont prévues dans le programme de recherche GEOCAP pour l'estimation des expositions : (1) Une approche semi-quantitative exclusivement fondée sur la distance des sujets aux sources d'exposition et sur les caractéristiques évidentes et objectives des sources (importance des routes, voltage des lignes à haute tension) (2) Une approche quantitative fondée sur l'estimation de l'exposition par modélisation des termes source et des modalités de diffusion des polluants.

Les termes source peuvent en effet être différents selon les lieux, les expositions ne se répartissant généralement ni de façon isotrope, ni de façon homogène autour de toutes les sources d'un même territoire (compte tenu, par exemple, des vents dominants, du caractère rural ou urbain des lieux d'habitation, …). L'étude des concentrations de benzène autour des voies de circulation a, par exemple, bien montré qu'elles étaient différentes suivant le type de bâti encadrant les voies et en fonction des vents dominants.

L'estimation quantitative des expositions tenant compte des caractéristiques locales des expositions a commencé pour (1) les CEM-EBF (collaboration avec RTE – prise en compte des types de ligne et des niveaux de courant) ; (2) la pollution atmosphérique liée au trafic (Collaboration avec l'ADEME et AIRPARIF); (3) les radiations ionisantes liées au radon d'une part et à la proximité des sites nucléaires d'autre part (Collaboration avec l'IRSN).

2.3 Analyse statistique

Le but principal de ce programme est de rechercher si l'incidence des cancers de l'enfant est "normale" ou éventuellement "plus élevée que la normale" à proximité des sources d'exposition et/ou dans les zones les plus exposées. Pour répondre à cette question, l'analyse statistique des informations recueillies sera conduite en plusieurs étapes.

- *Etude des déterminants des estimations quantitatives des expositions*

L'analyse des relations entre les caractéristiques générales des sources et témoins et le niveau d'exposition des témoins fournira des informations permettant de quantifier les expositions en utilisant les paramètres semi-quantitatifs. Il s'agira à la fois d'identifier ceux de ces paramètres qui sont les plus pertinents et de définir leur influence sur le niveau des expositions estimées de façon quantitative.

- *Distribution statistique des expositions dans la population générale d'âge pédiatrique et ses facteurs de variation*

Les informations recueillies sur nos échantillons témoins nous permettront de décrire la distribution statistique de l'exposition dans la population d'âge pédiatrique et ses variations en fonction de caractéristiques des individus (âge, taille du foyer) et de leurs communes d'habitation (type d'unité urbaine, caractéristiques socio-économiques, ...).

- *Association entre les caractéristiques des expositions et le risque de leucémie de l'enfant*

Les analyses seront réalisées par régression logistique ajustée sur l'âge et prendront en compte les caractéristiques sociodémographiques des communes de résidence.

Les données seront également analysées par sous-type cytologique et immunitaire de leucémie, et par strate d'âge.

- *Prise en compte des autres facteurs de risque établis ou suspectés*

Les facteurs de risque établis ou suspectés des cancers de l'enfant seront pris en compte :

Toutes les comparaisons cas-témoins seront ajustées sur l'âge, qui est un facteur de variation de l'incidence majeur. Nous étudierons la stabilité de chaque association conditionnellement aux autres expositions et l'exposition conjointe aux différentes expositions environnementales. L'influence d'autres facteurs de variation, comme les mouvements de population, sera également appréciée.

Pour apprécier la stabilité des résultats en fonction d'autres facteurs individuels nous utiliserons deux études cas-témoins nationales en population générale réalisées par l'U754 en 1998 et en 2003-2004. Les adresses des enfants au moment du diagnostic des cas ont été géocodées. Ces études disposent de l'historique des communes de résidence depuis la naissance.

3. Etat d'avancement du projet

3.1 Echantillonnage et géocodage des adresses

Les étapes suivantes ont été réalisées :

- Mise au point des procédures de transmission et d'anonymisation des informations recueillies pour les 15 000 témoins (collaboration technique et juridique avec l'INSEE)
- Recrutement des témoins et contrôle de leur représentativité de la population d'âge pédiatrique en termes d'âge, de tranche d'unité urbaine des communes et de taille du foyer (collaboration avec l'INSEE)
- Géocodage individuel des cas et des témoins (collaboration avec Geocible et avec l'IGN). La précision est de moins de 100 mètres pour 97% des cas et 99% des témoins, et de l'ordre de 15 m pour environ 80% d'entre eux. Des améliorations de la précision au cas par cas peuvent être apportées par rapport à chaque source d'exposition. Des études de sensibilité seront réalisées pour évaluer l'impact des imprécisions de codage sur les résultats.

3.2 Répartition de la population par rapport aux sources d'exposition

L'estimation des distances entre les résidences des cas et des témoins et les sources d'exposition est en cours. Elle est pratiquement achevée pour les lignes à haute tension, le réseau routier et les sites nucléaires.

Les tableaux ci-dessous montrent que 3,6% des témoins résident à moins de 20 km d'une centrale nucléaire de production d'électricité (tableau 2), 2,3% à moins de 200 m d'une ligne aérienne à haute tension (tableau 3), et 2,8% à moins de 500 m d'une route à grande circulation (tableau 4). Ces pourcentages faibles justifient l'extension de l'enquête à 3 années supplémentaires pour atteindre une puissance suffisante. Cette extension devrait être réalisée d'ici juin 2010, grâce à un financement complémentaire de l'InVS et de l'Afsset.

Tableau 2 : Distance des témoins à la centrale nucléaire de production d'électricité la plus proche

Distance	**0-5 km**	**5-10 km**	**10-15 km**	**0-20 km**
N	40	115	150	540
%	0,3%	0,8%	1,0%	3,6%

Tableau 3 : Distance des témoins à la ligne à haute tension (63, 90, 150, 225, 400 kV) la plus proche

Distance	**0-50 m**			**50-100 m**			**100-200 m**			**0-200 m**		
kV	< 225	225	400	< 225	225	400	< 225	225	400	< 225	225	400
N	77	24	9	78	38	4	51	46	15	206	108	28
		110			120			112			342	
%		0,7%			0,8%			0,8%			2,3%	

Tableau 4 : Proximité des témoins à la voie de circulation de classe 1, 2 ou 3 la plus proche

	Classe 1[1] ou 2[2] < 500 m	**Classe 3[3] < 500 m**	**Pas de classe <4 < 500 m**
N	419	9 220	5 361
%	2,8%	61,5%	35,7%

[1] Autoroute, rocade; [2] Route nationale; [3] Route départementale ou gros axe urbain

4. Principales publications obtenues,

- *Article publié*

Etude préliminaire de la répartition des cas de leucémie de l'enfant sur le territoire.
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(4):267-79.

- *Articles en préparation*

Association entre les leucémies aigues de l'enfant et la proximité des voies de circulation – enquête cas-témoins Escale (utilisation des données Geocap dans l'enquête cas-témoins 2003-2004)

Méthode d'estimation de l'exposition aux CEM-EBF en fonction des caractéristiques des lignes de transport d'électricité à haute tension et de la position par rapport à ces lignes.

5. Faits marquants, retombées prévisibles et perspectives de valorisation

Le programme GEOCAP tire bénéfice de l'investissement majeur réalisé antérieurement avec l'enregistrement annuel des leucémies de l'enfant depuis 1990 et des tumeurs solides depuis 2000. A partir de cette source d'information de grande envergure (près de 12 000 cas de leucémies et tumeurs solides de l'enfant avec une exhaustivité, une non redondance et une qualité d'informations recueillies validées) le programme GEOCAP a pu être amorcé fin 2005 par l'U754 – Inserm grâce à un financement d'EDF-Appel d'offres Epidémiologie et poursuivi grâce au financement de l'ANR – SEST. Un financement complémentaire a été apporté par la fondation Wyeth et par l'Institut de Veille Sanitaire. Le projet bénéficiera également en 2010 d'une allocation de recherche de l'ARC et d'une subvention de l'AFSSET.

GEOCAP est un projet de grande envergure, amené à se prolonger sur le long terme. L'exhaustivité du recrutement des cas, la représentativité de l'échantillon témoin, l'objectivité et la précision des

données sur les expositions environnementales en font un outil innovant et évolutif de recherche en santé publique. A cette étape, le soutien financier de l'ANR SEST 2005 a largement contribué :

- à la définition de procédures valides de constitution d'échantillons témoins géolocalisés,
- à l'élaboration de méthodes d'estimation des expositions environnementales liées à la proximité des lignes à haute tension (CEM-EBF) et aux voies de circulation (benzène aérien)
- à l'estimation de la proportion de la population exposée et des facteurs démographiques associés

L'extension à 3 années supplémentaires, qui pourra être réalisée d'ici quelques mois, permettra d'aborder dans de meilleurs conditions de puissance la suite de notre programme, et de ainsi de contribuer de façon originale à mieux cerner les facteurs de risque environnementaux des leucémies.