

Guide méthodologique

Les études locales autour d'un point source

Les différentes méthodes statistiques,
leurs avantages et leurs inconvénients

Note méthodologique

Sommaire

Introduction	2
1. L'approche par ratio de morbidité/mortalité standardisé	3
1.1 Les zone et période d'étude	3
1.2 L'estimation des effectifs de population	3
1.3 Le calcul du SMR	4
1.4 Le choix des taux de référence	4
1.5 L'intervalle de confiance du SMR	4
1.6 Le test du chi2	4
1.7 Les limites de l'approche par SMR	4
1.8 Conclusion sur ces études	5
2. Les méthodes alternatives	5
2.1 Les tests focalisés	5
2.2 Les tests de détection de cluster : la statistique de scan spatiale	5
2.3 La régression de Poisson	5
2.4 Le lissage cartographique des SMR	5
3. Le problème des tests multiples	6
4. Conclusion : "stratégie" pour une étude locale	6
Références bibliographiques	7

Les études locales autour d'un point source

Les différentes méthodes statistiques,
leurs avantages et leurs inconvénients

Note méthodologique

Rédacteurs

Sarah Goria, Alain Le Tertre
Institut de veille sanitaire, Département santé environnement

Préambule

Cette note ne s'intéresse qu'aux aspects méthodologiques dans les études locales. Ces aspects ne sont pas les seuls à devoir être pris en compte dans l'opportunité de mise en œuvre de telles études.

Introduction

Les études descriptives effectuées autour de sites particuliers cherchent à mettre en évidence une augmentation du taux d'incidence de pathologies ou de taux de mortalité spécifiques de populations vivant à proximité d'une zone qualifiée de polluée. Ces études sont menées pour répondre à l'interrogation des populations locales à la proximité de ces points sources. Quand l'étude est menée en réponse à une inquiétude de la population relative à une surincidence possible, l'interprétation des résultats est plus difficile puisque l'hypothèse est déterminée *a posteriori* et non *a priori*. Ce sont des études de type "géographique" qui sont mises en jeu. Ce type d'études est utilisé pour étudier les risques à proximité de sites nucléaires, chimiques ou des sites de dépôt ou d'incinération d'ordures [1]. "Ces études posent des problèmes quant à la méthodologie d'analyse ainsi qu'au niveau de l'interprétation des résultats" [1].

Cette note a pour objectif de décrire ce type d'études, les méthodologies en jeu, les limites afin de mieux cerner ces difficultés constatées par l'Institut national de la radioprotection et de la sûreté nucléaire (IRSN) [1].

Ces études se rapprochent, dans leurs méthodologies, des enquêtes "exposés-non exposés". La population vivant à proximité d'une source polluante est supposée exposée et est comparée à une population de référence supposée non exposée ou à tout le moins, moins exposée. L'objectif est de comparer l'état de santé de ces populations. Bien souvent, les différences observées sont mises en

relation avec le facteur environnemental alors que ce type d'étude ne permet pas de telles inférences.

Ces études se retrouvent sous différentes formes selon la zone de référence choisie. Il peut s'agir d'une comparaison :

- zone locale *versus* zone de référence (en général, la zone de référence est la France entière) ;
- zone locale au sein de zones voisines de tailles équivalentes ;
- zone locale avec un gradient d'exposition.

Les deux dernières approches sont en fait similaires. Seule la qualification de l'exposition varie.

Dans chacune des zones, le nombre de cas survenus sur une période donnée est comptabilisé, le nombre de cas attendus est estimé et l'existence d'un excès ou d'un déficit de cas est alors testé en comparant le nombre de cas observés au nombre de cas attendus, standardisés sur des facteurs démographiques, l'âge et le sexe, principalement. Ce ratio d'incidence/mortalité standardisé (SIR/SMR) représente le déficit ou l'excès d'incidence/mortalité par rapport à une population type.

L'objectif est de conclure ou ne pas conclure à une différence entre la population exposée et la population de référence, et répondre à la question sur la nécessité ou la non-nécessité de faire d'autres études.

1. L'approche par ratio de morbidité/mortalité standardisé (SMR)

On décrit ici l'approche classiquement utilisée pour l'analyse du risque à proximité d'un point source (calcul du SMR et de son intervalle de confiance et test du χ^2). Dans la suite du document, nous utiliserons l'acronyme SMR pour désigner indistinctement un SMR ou un SIR.

Pour réaliser ces études, il est nécessaire de définir : les pathologies d'intérêt et les indicateurs sanitaires pertinents, la zone d'étude et la période d'étude, la population potentiellement exposée et la population de référence.

Les indicateurs sanitaires retenus sont basés en général sur les données recueillies en routine telles que les données de mortalité issues de la surveillance pérenne des causes médicales de décès ou sur des données d'incidence. Pour les cancers, ces données d'incidence peuvent être issues de la surveillance pérenne des registres des cancers (registres généraux, registres spécifiques). Ce travail peut également s'appuyer sur des indicateurs d'incidence de type administratif tels que les nouvelles mises en affection longue durée (ALD) et les données du Programme de médicalisation des systèmes d'information (PMSI).

1.1 LES ZONE ET PÉRIODE D'ÉTUDE

La **zone d'étude** est définie en général sur la base d'un découpage géographique administratif, lié aux données disponibles. Ce découpage essaie d'être au plus proche de la zone d'exposition supposée au point source. Il est le reflet d'une meilleure discrimination de l'exposition mais aussi de la nécessité de disposer d'une base populationnelle suffisamment large pour les données sanitaires. Ce compromis génère une dilution du risque. Par ailleurs, ce découpage administratif est arbitraire et les résultats attendus peuvent être sensibles à sa redéfinition. Quel que soit le découpage, aucune estimation de l'exposition n'est en général utilisée puisque, dans la très grande majorité des cas, dans ce type d'étude, seule la distance sera retenue comme un indicateur *proxy* de l'exposition.

La **période d'étude** repose habituellement sur les données les plus récentes. En fonction des pathologies étudiées et de la latence de leur survenue par rapport à l'exposition, elle peut refléter une exposition antérieure, allant de plusieurs années à plusieurs dizaines d'années. Cette exposition peut ne plus refléter la situation actuelle et donc être éloignée des préoccupations de la population.

Tout comme pour la zone d'étude, afin de pallier le manque d'effectif, plusieurs années de données sont collectées. Ces années sont en général agrégées dans le but de gommer les fluctuations aléatoires des indicateurs sanitaires. Cette agrégation est source de biais dans l'estimation du risque [2].

1.2 L'ESTIMATION DES EFFECTIFS DE POPULATION

Les données de population par sexe et par âge sont essentielles pour calculer les SMR. Ces données sont obtenues par recensement et sont disponibles auprès de l'Institut national de la statistique et des études économiques (Insee).

Pour les études locales, l'unité statistique est en général la plus petite unité pour laquelle les données des recensements sont disponibles : soit la commune, soit les îlots regroupés pour l'information statistique (Iris). Ces données ne sont disponibles que pour les années des recensements. Les données de population (population par sexe et par âge) ne sont bien estimées, pour les années intercensitaires, qu'au niveau national. L'Insee ne fournit pas d'estimations en dessous du niveau départemental car elle les juge peu fiables. La méthode de l'interpolation diagonale peut être utilisée pour obtenir des estimations pour les années intercensitaires au niveau de la commune ou de l'Iris [3,4]. La démarche classique est la méthode de l'interpolation diagonale simple entre deux recensements. Il s'agit d'une méthode simplificatrice : on ne tient pas compte d'arrivées ou de départs massifs intervenant sur une période de temps limitée. Pour une présentation et une discussion des méthodes, voir les publications de Briggs *et al.* [5] et d'Arnold *et al.* [6].

Les erreurs d'estimation démographique ne sont sans doute pas uniformes dans la population française [7]. Les facteurs d'erreur varient avec la situation géographique, la classe d'âge et la résolution spatiale : à l'échelle de la France, les estimations sont relativement précises, à l'échelle de la commune, on peut imaginer de très fortes erreurs (mouvements massifs de population par exemple). Prou *et al.* [7] ont affecté les effectifs de population des départements, d'erreurs de mesure pour étudier le comportement des tests de surdispersion et l'analyse spatiale du risque de leucémies de l'enfant en France en fonction de l'importance de l'erreur. Les résultats n'étaient significativement perturbés qu'à partir de fortes erreurs au niveau départemental, lesquelles erreurs étaient bien plus élevées que celles résultant de la précision assurée par l'Insee dans ses estimations. À une échelle géographique plus fine, les erreurs de population sont probablement plus grandes et les perturbations des tests sont plus significatives.

Best and Wakefield [8] ont étudié les effets d'erreurs présentes dans les données de santé et/ou de population. Pour les données de santé, les auteurs ont considéré l'effet d'erreurs dues aux méthodes d'enregistrement utilisées par les registres. Pour les données de cancer en particulier, le sous-enregistrement est un problème majeur. Les données de population, en général traitées comme des quantités connues, proviennent des recensements et sont sujettes à erreurs (souvent sous-estimées). Best and Wakefield [8] proposent différents modèles pour prendre en compte ces erreurs dans les études de représentation cartographique des cancers, afin de ne pas surestimer les risques.

1.3 LE CALCUL DU SMR

Le nombre de cas observés (O) dans une unité géographique est comparé au nombre de cas attendus (E). Le nombre de cas attendus est calculé à partir de taux d'incidence de référence (taux "France entière" par exemple) appliqués à la population de l'unité correspondante. Ce rapport du nombre de cas observés sur le nombre de cas attendus définit un SMR (Standardized Mortality/Morbidity Ratio) :

$$SMR = \frac{O}{E}$$

Le SMR mesure le taux de mortalité ou d'incidence relatif d'une zone par rapport à celui de la zone de référence (France entière par exemple).

La standardisation indirecte utilise l'incidence ou la mortalité d'une population de référence. Le nombre de cas attendus est obtenu en appliquant les taux de référence (r_j) à la population de la zone d'étude (P_{ij}) :

$$E_i = \sum_j r_j P_{ij}$$

j indique généralement la classe d'âge.

1.4 LE CHOIX DES TAUX DE RÉFÉRENCE

On peut utiliser, par exemple :

- les taux d'incidence estimés par Francim pour la "France entière" ;
- les taux d'incidence observés dans un registre de cancer, si l'étude se trouve dans un département couvert par un registre de cancer ;
- les taux d'incidence observés dans les différents registres de cancer (ou dans un registre national, s'il existe), si le cancer étudié est très rare ;
- les taux de mortalité pour la "France entière" ;
- les taux de mortalité pour la région.

Plus la population de la zone d'étude est faible, plus les SMR et les résultats des tests seront sensibles au choix des taux de référence. Il est commun de prendre, comme taux de comparaison, le taux moyen observé dans la zone de référence. Cette approche n'intègre qu'une information partielle sur la distribution des taux et réduit artificiellement l'incertitude sur le SMR, liée à ce taux de référence. Elle augmente la probabilité de rejeter à tort l'hypothèse nulle d'égalité. Il est utile de faire une analyse de sensibilité pour voir l'influence des taux de référence sur les SMR (par la comparaison des résultats obtenus avec différents taux de référence).

Jarup *et al.* [9] ont obtenu les taux de référence par modélisation. Ceci était fait pour ne pas avoir des taux instables à cause du petit effectif de la population de référence. Bithell *et al.* [10] estiment le nombre de cas attendus par une régression de Poisson en ajustant sur un score socio-économique et un effet région.

1.5 L'INTERVALLE DE CONFIANCE DU SMR

Le rapport standardisé de mortalité/morbidité permet de quantifier un excès ou un moindre risque dans la zone exposée par rapport à la

zone de référence. La confirmation d'un excès dans la zone d'étude repose sur une valeur du SMR significativement supérieure à 1. La variabilité de cette estimation du risque relatif est quantifiée par l'intervalle de confiance du SMR.

Pour calculer l'intervalle de confiance du SMR, on peut utiliser les tables de la loi de Poisson¹. Si le nombre de cas observés O est grand (O>100), plusieurs approximations sont disponibles pour obtenir un intervalle de confiance du SMR. L'approximation obtenue en supposant que le log(SMR) suit une loi normale donne l'intervalle de confiance à 95 % suivant :

$$\left[\frac{SMR}{\exp\left(\frac{1,96}{\sqrt{O}}\right)} ; SMR \times \exp\left(\frac{1,96}{\sqrt{O}}\right) \right]$$

Lorsque différents taux de référence sont disponibles, avec par exemple les taux de mortalité dans l'ensemble des départements français, une analyse par rééchantillonnage (bootstrap), paramétrique ou non paramétrique selon la population de taux disponibles, permet une meilleure prise en compte de l'incertitude. L'intervalle de confiance est ici exprimé directement à partir des estimateurs calculés sur l'ensemble des échantillons.

1.6 LE TEST DU CHI2

Afin de décider si la valeur du risque est significativement différente de 1, le test du chi2, qui résulte d'une approximation de la loi de Poisson à la loi normale², peut-être aussi utilisé [11].

Pour tester une surincidence, un test unilatéral devrait être utilisé (l'hypothèse nulle est : SMR égal à 1 ; l'hypothèse alternative est : SMR supérieur à 1).

Dans les études locales autour d'un point source, le test bilatéral est en général utilisé : c'est le cas où il n'y a pas de connaissances ou *a priori* sur le niveau supérieur ou inférieur de l'incidence de cancer de la population étudiée par rapport à celui de la population de référence (remarque : en gardant le seuil de significativité à 5 %, par exemple, le test bilatéral est plus conservateur que le test unilatéral puisqu'il acceptera plus facilement l'hypothèse d'égalité).

1.7 LES LIMITES DE L'APPROCHE PAR SMR

Les problèmes liés à cette approche sont nombreux [10,12]. Ces taux ou rapports sont instables lorsqu'ils sont calculés à partir d'un faible nombre de cas attendus. Les études portant généralement sur de faibles effectifs observés sur des zones de surface réduite, ceci est fréquemment le cas. Cette instabilité est aussi exprimée par la variance des SMR, inversement proportionnelle à leur espérance. Ainsi, des SMR élevés peuvent être observés quand le nombre de cas attendus est proche ou inférieur à l'unité [1]. Les autres faiblesses de cette approche [10,12] sont liées au choix critique de la taille de la région d'étude dite "à proximité" du site, à la possibilité qu'une variabilité régionale de l'incidence non liée au site puisse "invalider" les valeurs

¹ Le nombre de cas observés O est supposé suivre une loi de Poisson de moyenne E θ (θ : risque relatif). L'estimateur du maximum de vraisemblance de θ est le SMR.

² Sous l'hypothèse nulle (SMR=1), le nombre de cas observés suit une loi de Poisson de moyenne et variance égale au nombre de cas attendus E. Lorsque le nombre de cas attendus est assez grand, la loi de Poisson tend vers une loi normale.

attendues utilisées, à la perte de puissance dans la détection d'effets localisés quand on utilise des zones trop larges par rapport aux effets à détecter. Les résultats sont sensibles à la détermination de la population et au choix des taux de référence utilisés pour le calcul du nombre de cas attendus. De plus, cette analyse d'incidence n'utilise aucune information spatiale. D'autres tests ont été développés pour prendre en compte la relation spatiale des cas au site.

Par ailleurs, tout le schéma de raisonnement repose sur la comparaison du nombre de cas observés au nombre de cas attendus sous l'hypothèse d'une distribution de loi de Poisson. Or, si cette distribution peut légitimement être postulée, elle n'est réaliste qu'après avoir ajusté les taux sur les différents facteurs de confusion. Sans cet ajustement, une surdispersion des données est en général observée, conduisant de nouveau à rejeter à tort l'hypothèse nulle.

1.8 CONCLUSION SUR CES ÉTUDES

Les avantages de ces études sont :

- la simplicité de réalisation ;
- le faible coût (si des données préexistantes sont utilisées) ;
- la rapidité.

Ces études présentent, cependant, des limites majeures :

- aucun ajustement sur les facteurs de confusion autres que l'âge et le sexe en général ;
- aucune information individuelle n'est utilisée (histoire résidentielle, par exemple) ;
- résultats sensibles au choix du découpage géographique ;
- le plus souvent, aucune estimation de l'exposition n'est utilisée (la distance est en général utilisée comme *proxy* de l'exposition).

2. Les méthodes alternatives

2.1 LES TESTS FOCALISÉS

Les cancers étant rares, on peut s'attendre à ne pas trouver une surincidence significative dans la zone à proximité du site. C'est pour cela que des tests adaptés ont été développés, en particulier des tests plus sensibles à une relation entre le risque et la distance au site étudié : test de Stone du maximum de Poisson, test de Stone du rapport du maximum de vraisemblance et test du score de risque linéaire de Bithell entre autres [10,12,13]. Ces tests permettent également de prendre en compte des covariables.

Mais l'utilisation des tests d'hypothèses pour l'analyse de risque autour d'un point source a quelques limites, notamment la définition de l'hypothèse alternative et le choix des relations distance-risque, la multiplicité des tests et son implication sur le risque de première espèce. De plus, la puissance de ces tests reste faible pour détecter des petites augmentations de risque [14].

2.2 LES TESTS DE DÉTECTION DE CLUSTER : LA STATISTIQUE DE SCAN SPATIALE

Ces tests [15-17] n'ont pas été développés pour répondre à cette question mais, si les données sont disponibles (les données d'un

département, par exemple), il peut être intéressant de voir quel est le cluster le plus probable, s'il y en a un, et s'il est à proximité du site étudié. Ces tests permettent de prendre en compte des covariables (exprimées sous forme de classes).

2.3 LA RÉGRESSION DE POISSON

La régression de Poisson, avec la distance par exemple, comme *proxy* de l'exposition, peut être utilisée dans l'estimation du risque en relation à un point [18]. Cette méthode permet aussi de prendre en compte des facteurs de confusion tels que la densité de population et le niveau socio-économique.

2.4 LE LISSAGE CARTOGRAPHIQUE DES SMR

Dans les études locales liées à un point source, une approche possible est de calculer les SMR, selon les mêmes hypothèses que précédemment, pour l'ensemble des Iris d'un département ou d'une région. Une représentation cartographique permet d'en étudier la distribution spatiale et notamment d'observer si les Iris exhibant les valeurs les plus extrêmes sont situées autour du point source. L'interprétation des résultats de ce type de représentation graphique est parfois délicate et peut conduire à des erreurs, en particulier lorsque, dans certaines unités géographiques, les nombres de cas sont faibles et quand les unités géographiques présentent des différences substantielles dans la taille des populations [19]. Dans ces cas de figure, les taux estimés dans les différentes unités géographiques sont en effet d'un niveau de précision inégal et rendent de ce fait difficile le distinguo entre une fluctuation aléatoire et une vraie variation du niveau de risque entre les zones géographiques étudiées. Ce problème se pose d'autant plus, qu'en général, les estimations les plus extrêmes sont souvent aussi celles pour lesquelles le niveau de précision est le plus faible [19].

Pour obtenir des données plus informatives et vraisemblables pour la cartographie des SMR, on procède à un lissage de type bayésien [20]. Les techniques de lissage bayésien permettent de stabiliser ces estimations basées sur des petits effectifs. Les modèles hiérarchiques bayésiens prennent en compte deux types d'information : les cas observés dans chaque zone géographique, mais également une information *a priori* sur la similarité des risques entre ces zones [19,20]. Les plus utilisés sont le modèle Poisson-gamma [21] et le modèle "somme" de Besag, York et Mollié [22]. Le modèle Poisson-gamma [21] suppose que les risques relatifs sont *a priori* indépendants (pas d'autocorrélation spatiale) et identiquement distribués ; il s'agit d'un lissage global. Ce modèle donne une estimation des risques relatifs qui est un compromis entre le risque moyen et le risque observé (SMR) de l'unité géographique considérée : plus le nombre attendu est élevé plus l'estimation est proche du SMR. Il s'agit d'un lissage global. Le modèle "somme" [22] permet de prendre en compte une éventuelle autocorrélation spatiale des données. La variabilité des risques relatifs est décomposée en une composante spatiale et une composante non structurée spatialement, ou composante d'hétérogénéité. Ce modèle donne un lissage "mixte". Lorsque la dépendance spatiale est peu prononcée, les estimations obtenues sont similaires à celles issues d'un lissage global. Lorsque les risques ont une structure présentant une dépendance spatiale, il s'agit alors d'un lissage local aboutissant à une estimation tendant vers le risque moyen des unités voisines.

3. Le problème des tests multiples

Bien souvent, les études locales menées étudient l'incidence de plusieurs cancers. Or, plus on réalise de tests sur une même population et plus on a de chance de rejeter à tort l'hypothèse nulle d'égalité des ratios au seuil α . Le risque de première espèce α étant généralement fixé à 5%, on accepte de rejeter à tort l'hypothèse nulle 5 fois sur 100. Plus on augmente le nombre de tests réalisés, plus la probabilité d'obtenir un faux positif augmente.

Pour pallier à ce problème, des méthodes de correction des tests multiples existent qui corrigent cette dérive. Les *p-values* sont recalculées pour prendre en compte le nombre de tests réalisés. Une méthode de correction largement répandue est celle de Bonferroni [23] (de α on passe à α/n , n étant le nombre de tests effectués). Cette méthode a pour objectif de s'assurer que le rejet de l'hypothèse nulle est valide au seuil initialement fixé quel que soit le nombre de tests effectués. En d'autres termes, il s'agit d'éviter des faux positifs. D'autres méthodes de correction [24,25], dérivées de celle de Bonferroni, sont plus performantes. À vouloir trop éviter les faux positifs, ces méthodes augmentent le risque d'obtenir des faux-négatifs. Benjamini and Hochberg [26] ont proposé une méthodologie différente qui repose sur un taux attendu de fausses découvertes. Ce taux est le rapport $\left(\frac{FP}{VP+FP}\right)$

positifs et le nombre total de tests significatifs, vrais positifs et faux positifs. Cette méthode suppose de connaître le taux de fausses découvertes. Or, celui-ci n'est jamais connu et doit donc être estimé. Différentes stratégies de contrôle du taux de fausses découvertes sont disponibles, mais toutes ont en commun de se montrer moins conservatives que les approches précédentes, axées sur le contrôle de l'erreur de type I. Avec cette approche, il est également possible d'obtenir une estimation de la probabilité pour chaque test d'être un faux positif [27].

4. Conclusion : “stratégie” pour une étude locale

Les études locales autour d'un point source souffrent :

- d'être mises en place pour répondre à une perception déjà existante d'une surincidence (décision *a posteriori*);
- de leur construction intrinsèque basée sur une analyse écologique géographique (biais écologiques);
- d'une absence quasi-systématique de contrôle des facteurs de confusion;
- de leur caractère unique, les rendant potentiellement sujettes aux variations aléatoires.

Ces défauts rendent difficilement crédible toute analyse locale basée sur cette approche sauf dans le cas exceptionnel où le risque serait tel qu'il ne pourrait *a priori* pas être lié à un facteur de confusion ou au hasard.

Les solutions techniques (bootstrap, tests multiples, étude par rapport à la distance) présentées dans ce guide, limitent mais n'éliminent pas ces défauts structurels.

Face à la demande d'investigation autour de point source, une solution, même imparfaite, serait de réaliser une étude multicentrique autour de sites présentant les mêmes caractéristiques d'émission [4,9,28-31]. Une telle étude, prenant en compte les facteurs de confusion potentiels, permettrait de générer une hypothèse sur une possible sur- ou sous-incidence de pathologie. Elle resterait soumise au biais écologique mais on peut espérer que l'effet du facteur de confusion, non mesuré à l'origine de ce biais, décroisse avec le nombre de sites.

Références bibliographiques

- [1] IRSN. Les études épidémiologiques des leucémies autour des installations nucléaires chez l'enfant et le jeune adulte : revue critique, 2008.
- [2] Ocana-Riola R. The misuse of count data aggregated over time for disease mapping. *Stat Med* 2007;26:4489-504.
- [3] Benhamou E, Laplanche A. Estimate of the population at risk between 2 census records for the calculation of incidence level or mortality of cancer: comparison of 4 methods. *Rev Epidemiol Sante Publique* 1991;39:71-7.
- [4] White-Koning ML, Hemon D, Laurier D, Tirmarche M, Jouglu 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.
- [5] Briggs D, Fecht D, De Hoogh K. Census data issues for epidemiology and health risk assessment: experiences from the small area health statistics unit. *J R Statist Soc A* 2007;170:355-78.
- [6] Arnold RA, Diamond ID, Wakefield JC. The use of population data in spatial epidemiology. *Spatial Epidemiology: methods and applications*. In Elliott P *et al.* ed. Oxford: Oxford University Press; 2000.
- [7] Prou M, Lajaunie C, Wackernagel H, Robelin D, Clavel J, Hémon D. Spatial analysis of risk and demographic errors: childhood leukaemia in France, 1990-1997. *Geography and health* 2003;209-25.
- [8] Best NG, Wakefield J. Accounting for inaccuracies in population counts and case registration in cancer mapping studies. *J Royal Statistical Society A* 1999;162:363-82.
- [9] Jarup L, Briggs D, de HC, Morris S, Hurt C, Lewin A, Maitland I, Richardson S, Wakefield J, Elliott P. Cancer risks in populations living near landfill sites in Great Britain. *Br J Cancer* 2002;86:1732-6.
- [10] Bithell JF, Keegan TJ, Kroll ME, Murphy MF, Vincent TJ. Childhood leukaemia near British nuclear installations: methodological issues and recent results. *Radiat Prot Dosimetry* 2008;132:191-7.
- [11] Bouyer J, Hémon D, Cordier S, Derriennic F, Stücker I, Stengel B *et al.* *Épidémiologie : principes et méthodes quantitatives*. Paris: Inserm; 1995.
- [12] Bithell JF, Stone RA. On statistical methods for analysing the geographical distribution of cancer cases near nuclear installations. *J Epidemiol Community Health* 1989;43:79-85.
- [13] Bithell JF. The choice of test for detecting raised disease risk near a point source. *Stat Med* 1995;14:2309-22.
- [14] Lawson AB, Biggeri A, Williams FLR. A review of modelling approaches in health risk assessment around putative sources. In: Wiley, editor. *Disease mapping and risk assessment for public health*. Chichester;1999.
- [15] Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med* 1995;14:799-810.
- [16] Kulldorff M. A spatial scan statistic. *Commun Statist Theory Meth* 1997;26:1481-96.
- [17] Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med* 2006;25:3929-43.
- [18] Morris SE, Wakefield JC. Assessment of disease risk in relation to a pre-specified source. In: Oxford University Press, editor. *Spatial Epidemiology: methods and applications*. Oxford;2000.
- [19] David S, Remontet L, Bouvier AM, Faivre J, Colonna M, Esteve J. How to choose in practice the model of spatial variation of cancer incidence? Examples of digestive cancers from Cote-d'Or "department" – France. *Rev Epidemiol Sante Publique* 2002;50:413-25.
- [20] Wakefield JC, Best NG, Waller LA. Bayesian approaches to disease mapping. *Spatial Epidemiology: methods and applications*. In Elliott P *et al.* ed. Oxford: Oxford University Press;2000.
- [21] Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987;43:671-81.

- [22] Besag J, York JC, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991;43:1-59.
- [23] Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilit. *A Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936;8:3-62.
- [24] Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979;6:65-70.
- [25] Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 1988;75:383-6.
- [26] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 1995;57:289-300.
- [27] Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* 2002;64:479-98.
- [28] 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.
- [29] Elliott P, Shaddick G, Kleinschmidt I, Jolley D, Walls P, Beresford J, Grundy C. Cancer incidence near municipal solid waste incinerators in Great Britain. *Br J Cancer* 1996;73:702-10.
- [30] Elliott P, Richardson S, Abellan JJ, Thomson A, de HC, Jarup L, Briggs DJ. Geographic density of landfill sites and risk of congenital anomalies in England. *Occup Environ Med* 2009;66:81-9.
- [31] Evrard AS, Hemon D, Morin A, Laurier D, Tirmarche M, Backe JC, 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.

Mots clés : SMR, ratio de mortalité standardisé, point source, études locales

Citation suggérée :

Goria S, Le Tertre A. Les études locales autour d'un point source – Les différentes méthodes statistiques, leurs avantages et leurs inconvénients – Note méthodologique. Saint-Maurice (Fra) : Institut de veille sanitaire, juillet 2010, 8 p. Disponible sur : www.invs.sante.fr