

MODÉLISATION STATISTIQUE DU LIEN ENTRE L'EXPOSITION INDIRECTE
AUX PRODUITS PHYTOSANITAIRES AGRICOLE ET LE RISQUE DE SURVENU
D'UNE HÉMOPATHIE MALIGNE EN FRANCE

C. C. AGUIDA^{1,2} H-P. Dang^{3,4} A. Monnereau^{1,2}, B. Vacquier², S. Orazio^{1,2}

¹Registre des Hémopathies Malignes de la Gironde, Institut Bergonié

²Epicene team, University of Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219

³*Univ. de Bretagne Occidentale, LaTIM, INSERM - UMR 1101*

⁴*Ensaï, CREST - UMR 9194*

14 mars 2022

PLAN DE PRÉSENTATION

- 1 CONTEXTE ET PROBLÉMATIQUE
- 2 OBJECTIFS
- 3 ÉTUDE ÉCOLOGIQUE
- 4 MÉTHODE D'ANALYSE
- 5 RÉSULTATS
- 6 DISCUSSIONS

CONTEXTE ET PROBLÉMATIQUE

CONTEXTE

- ☞ France, premier pays agricole européen en termes de Surface Agricole Utile (SAU),
- ☞ un des premiers consommateurs de pesticides agricoles en Europe.

PROBLÉMATIQUE

- ☞ les pesticides épanchés se propagent bien au-delà de la zone sur laquelle ils sont appliqués,
- ☞ plusieurs pesticides sont classés cancérigènes certains ou probables pour l'homme.

Les populations riveraines des zones agricoles sont donc exposées.

PROBLÉMATIQUE

- ⇒ peu d'études épidémiologiques sur l'impact sanitaire des pesticides lié au voisinage d'activités agricoles.
- ⇒ projet GEO-K-PHYTO financé par le plan ECOPHYTO II+.



GEO-K-PHYTO

PLAN DE PRÉSENTATION

- 1 CONTEXTE ET PROBLÉMATIQUE
- 2 OBJECTIFS**
- 3 ÉTUDE ÉCOLOGIQUE
- 4 MÉTHODE D'ANALYSE
- 5 RÉSULTATS
- 6 DISCUSSIONS

GEO-K-PHYTO

OBJECTIF : mettre en place un dispositif de surveillance épidémiologique des **cancers de l'adulte** en lien avec les **expositions environnementales** aux phytopharmaceutiques.

OBJECTIF SPÉCIFIQUE

OBJECTIF : Modélisation statistique du lien entre l'**exposition indirecte aux produits phytosanitaires agricole** et le **risque de survenue d'une Hémopathie Maligne** en France



Réalisation d'une étude épidémiologique du type écologique

ENJEUX

- ① Indicateur d'exposition indirecte des riverains proche de l'exposition individuelle ;
- ② Hétérogénéité des taux d'incidence des hémopathies malignes ;
- ③ Spatialisation des indicateurs ;
- ④ Estimation des effets.

OBJECTIF DE LA COMMUNICATION

- ★ décrire les choix méthodologiques effectués dans cette phase d'étude écologique du projet.

PLAN DE PRÉSENTATION

- 1 CONTEXTE ET PROBLÉMATIQUE
- 2 OBJECTIFS
- 3 ÉTUDE ÉCOLOGIQUE**
- 4 MÉTHODE D'ANALYSE
- 5 RÉSULTATS
- 6 DISCUSSIONS

ÉTUDE ÉCOLOGIQUE

- Approche descriptive ==> hypothèses étiologiques individuelles
- futures investigations plus coûteuses

PATHOLOGIES D'INTÉRÊTS

Hémopathies Malignes (HM) : ensemble hétérogène de cancers des cellules sanguines et de leurs précurseurs.

UNITÉ SPATIALE

- ☞ Commune ✓
- ☞ Iris

Période d'étude : 01/01/2006 au 31/12/2017.

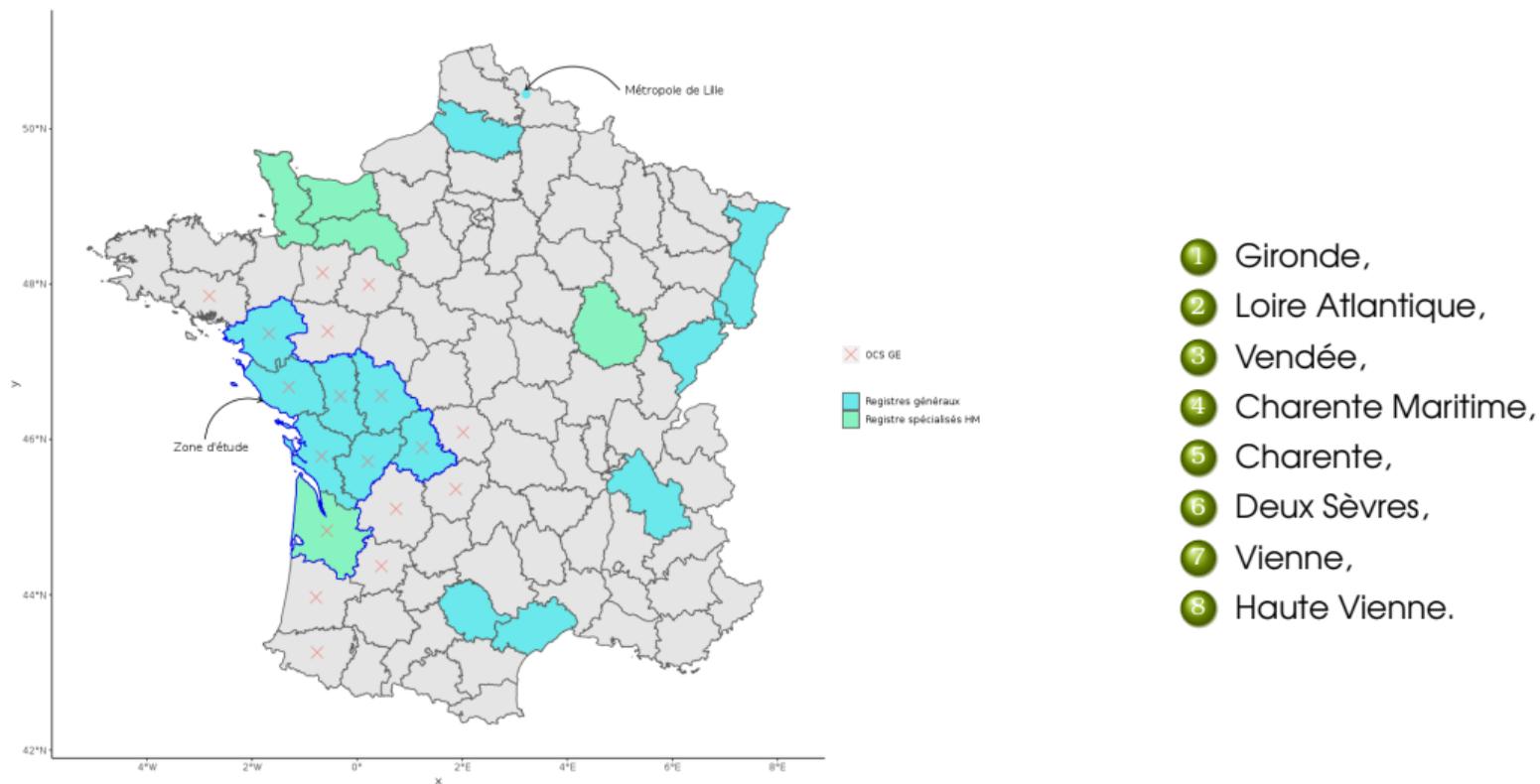


FIGURE 1 – Carte des Registres et de disponibilité des données OCS-GE

RÉSEAU FRANCIM

Données sanitaires : les cas diagnostiqués

- Département
- Sexe
- Topographie
- Morphologie
- Année de diagnostic
- Classe d'âge
- Commune

IGN : DONNÉES OCS GE

Caractérisation de l'exposition :

Surface agricole utile(SAU)

- Département
- Code de la commune
- Libellés de Commune
- Surface en m² par culture
 - ☞ Verger ✓
 - ☞ Terre arable ✓
 - ☞ Vigne ✓
 - ☞ Friche
 - ☞ Prairie
- Surface en m² de la commune

FACTEURS DE CONFUSION POTENTIELS

- Densité de la commune ==> INSEE
- Zonage en Aires Urbaines(ZAU) ==> INSEE
- Niveaux socio-économiques : EDI de la commune ==> plateforme MapInMed
- Nombre d'industrie polluants ==> IREP
- Radon ==> Institut de Radioprotection et de Sûreté Nucléaire (IRSN)

DÉFINITION DE L'INDICATEUR SANITAIRE :

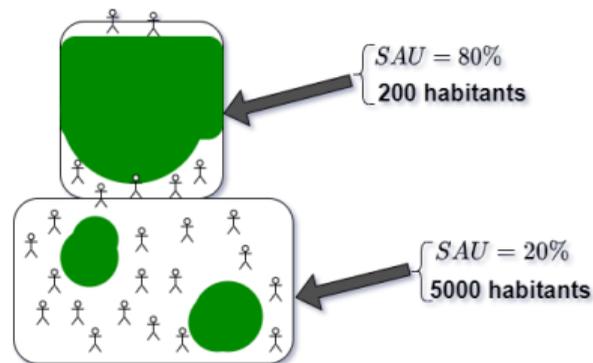
$$SIR = \frac{O}{E} = \frac{\text{nombre de cas observés}}{\text{nombre de cas attendus}}$$

INDICATEUR D'EXPOSITION INDIRECTE

- Transformation proposée :

$$SAU_{net} = \%SAU \times \frac{\text{Effectif de la commune}}{\text{Superficie} \times (1 - \%SAU)}$$

☛ Exposition indirecte par habitant au m^2 habitable



PLAN DE PRÉSENTATION

- 1 CONTEXTE ET PROBLÉMATIQUE
- 2 OBJECTIFS
- 3 ÉTUDE ÉCOLOGIQUE
- 4 MÉTHODE D'ANALYSE**
- 5 RÉSULTATS
- 6 DISCUSSIONS

TYPE DE DONNÉES

- données de comptage : nombre de cas par HM par unité géographique
- rareté de la maladie + la durée limitée de la période de déclaration ==> dispersion dans l'espace et dans le temps : excès de zéros
- informations géographiques

DESCRIPTION DU MODÈLE BYM(BESAG YORK MOLLIE) [1]

Le modèle s'écrit :

$$\log(\mathbf{E}(O)) - \log(E) = \mu + X^T \beta + \phi + \theta. \quad (1)$$

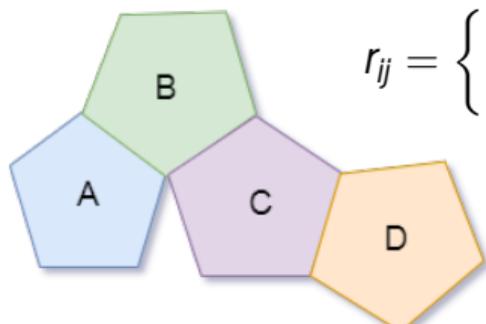
avec μ l'intercept, β les coefficients des covariables et $X (\in \mathcal{X} \subseteq \mathbb{R}^p)$ les variables explicatives. ϕ est une composante spatiale ICAR et θ une composante à effets aléatoires ordinaires pour l'hétérogénéité non spatiale.

CAR

- La dépendance spatiale est le plus souvent modélisée par le modèle autorégressif conditionnel (CAR) ,
- Soit n le nombre de région, D une matrice diagonale $n \times n$, et W la matrice d'adjacence $n \times n$, l'auto-corrélation spatiale est modélisée par une matrice de voisinage $R = D(I - \alpha W)$.
- ICAR est un cas particulier du CAR avec $\alpha = 1$.

Dans le cas ICAR, chaque élément de $R = D - W$ est :

$$r_{ij} = \begin{cases} n_i & i = j, n_i : \text{le nombre de voisin de } i \\ -1(i \sim j) & i \neq j, \text{i.e -1 si } i \text{ et } j \text{ sont voisins} \end{cases}$$



$$R = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

$$p(\phi_i | \phi_{i \sim j}) = N \left(\sum_{i \sim j} \frac{\phi_i}{n_i}, \frac{\kappa_j^2}{n_i} \right), \quad \kappa \text{ inconnue}$$

$$p(\phi) \propto \exp \left\{ -\frac{1}{2\kappa} \sum_{i \sim j} (\phi_i - \phi_j)^2 \right\}, \quad \phi_i \in \mathcal{R}$$

$$\theta = N(0, \lambda) \quad \text{et} \quad \kappa, \lambda \sim \text{Inverse-Gamma}(a, b)$$

ϕ ET θ

- ϕ et θ sont utiles pour tenir compte de la surdispersion non modélisée par les variables de Poisson,
- Lorsque la variance observée n'est pas entièrement expliquée par la structure spatiale des données, les termes d'erreur indépendants expliquent le reste.

ZIP : IDÉE

- prise en compte de l'excès de zéro
- forte hétérogénéité du nombre de cas entre les communes à faible densité et celles densément peuplées

$Y_i \sim \text{ZIP}(h, p)$, le modèle peut s'écrire :

$$\begin{cases} P(y_i) = p + (1 - p)\exp(-E_i h) & , y_i = 0 \\ P(y_i) = (1 - p) \frac{\exp(-E_i h)(E_i h)^{y_i}}{y_i!} & , y_i > 0 \end{cases}$$

IDÉE

- capter la dynamique des risques en fonction des niveaux d'exposition mesurés en continue,
- les splines naturelles sont des cas particuliers des splines cubiques contraintes à être linéaires dans les queues.

$$f(x) = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1} \quad (2)$$

où $x_1 = x$ et pour tout $i = 1, \dots, k-2$,

$$x_{i+1} = (x - s_i)_+^3 - \frac{s_k - s_i}{s_k - s_{k-1}} (x - s_{k-1})_+^3 + \frac{s_{k-1} - s_i}{s_k - s_{k-1}} (x - s_k)_+^3.$$

PLAN DE PRÉSENTATION

- 1 CONTEXTE ET PROBLÉMATIQUE
- 2 OBJECTIFS
- 3 ÉTUDE ÉCOLOGIQUE
- 4 MÉTHODE D'ANALYSE
- 5 RÉSULTATS**
- 6 DISCUSSIONS

⚠ CONFIDENTIALITÉ

seront présentés dans la suite du document les résultats nécessaires à la compréhension des choix méthodologiques qui ont été réalisés.

RÉSUMÉ

- 👉 **Type d'étude** : étude écologique
- 👉 **Unité géographique d'observation** : la commune (**2546** communes)
- 👉 **Période d'étude** : 01/01/2006 au 31/12/2017
- 👉 **Population d'étude** : patients atteints de certains sous type d'HM de l'adulte (plus de 15 ans révolus)
- 👉 **Zone d'étude** : 8 départements
- 👉 **Nombre de cas recensés** : 40659 cas

RÉSULTATS : DONNEES SANITAIRES

Hémopathies malignes	Sous types	Nombre de Cas	Nombre de Commune
Lymphome de Hodgkin	Lymphome de Hodgkin	1945	791
Lymphome non hodgkinien	Leucémie lymphoïde chronique / Lymphome lymphocytaire	5363	1384
	Myélome multiple et plasmocytome	5092	1341
	Lymphome diffus à grandes cellules B	4837	1278
	Lymphome folliculaire	2807	947
	Lymphome de la zone marginale	2319	904
	Lymphome lymphoplasmocytaire / Macroglobulinémie de Waldenström	1962	795
	Lymphome T/NK à cellules matures (LNH-T)	1792	755
	Lymphome à cellules du manteau	770	434
	Leucémie / Lymphome lymphoblastique à cellules précurseurs (B, T ou SAI)	520	327
	Leucémie à tricholeucocytes	313	227
	Lymphome / Leucémie de Burkitt	170	130

TABLE 1 – Nombre de cas/communes par sous types d'hémopathies malignes, **2546 communes**

RÉSULTATS : NOMBRE DE CAS VS EXPOSITION

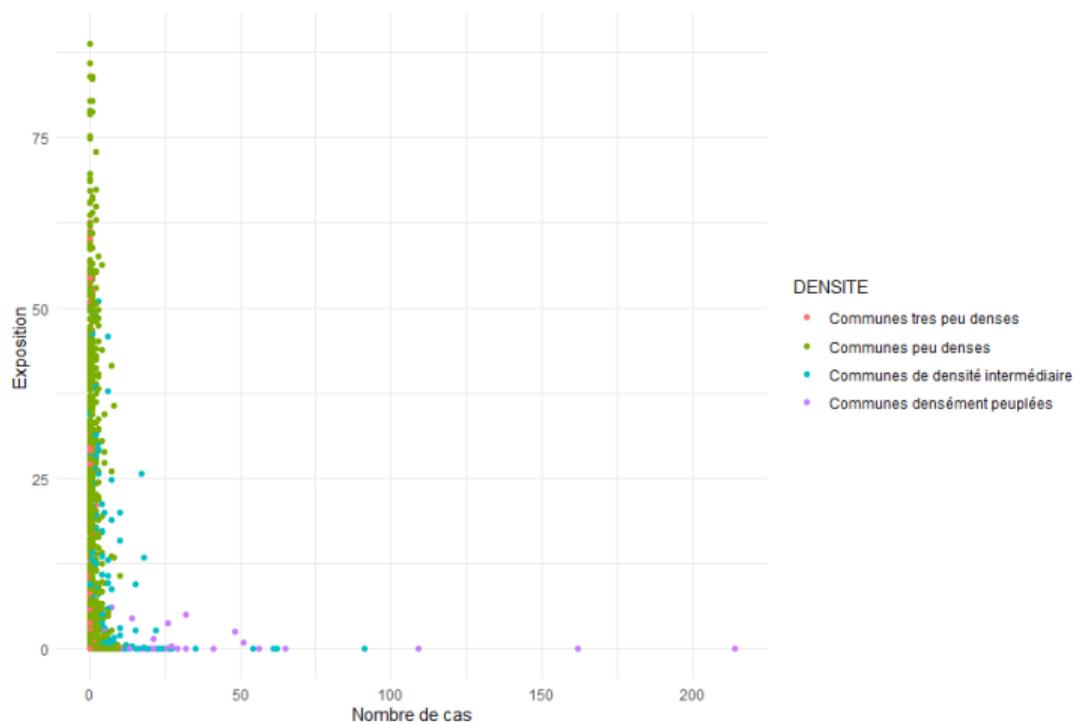
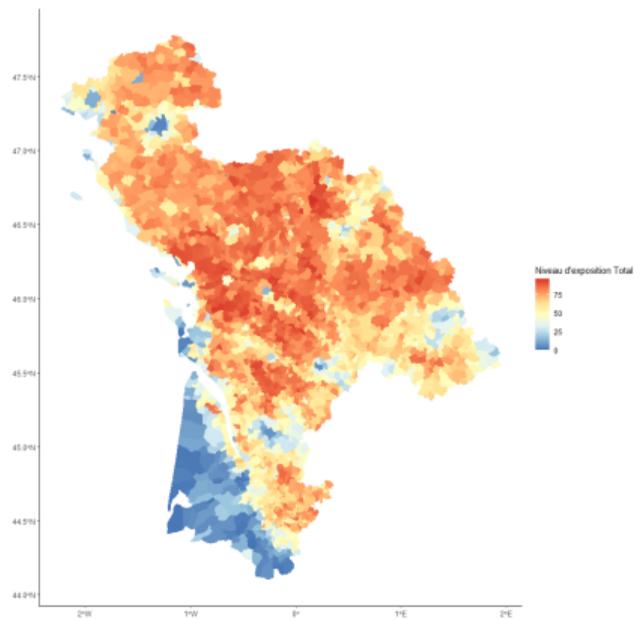
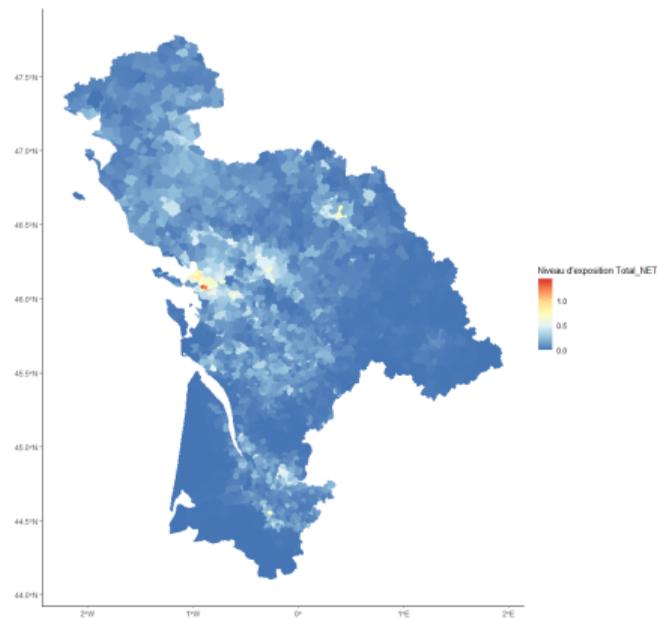


FIGURE 2 – Distribution du nombre de cas observés de Myélome Multiple en fonction de la surface agricole.

RÉSULTATS : INDICATEUR D'EXPOSITION



(A) Surface agricole utile brut



(B) Surface agricole utile nette

FIGURE 3 – Cartographie des indicateurs d'expositions (Pourcentage de SAU brute vs pourcentage de SAU nette) résumé sur la période 2006-2017 et calculé à l'échelle de la commune.

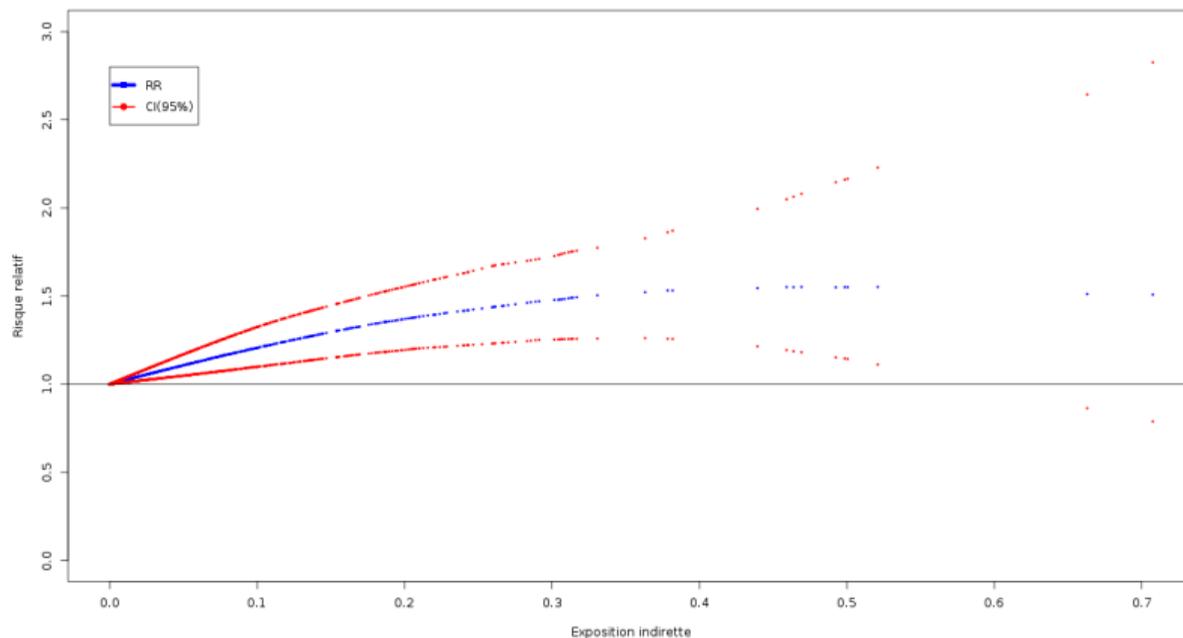


FIGURE 4 – Exemple de dynamique du risque relatif avec son intervalle de crédibilité

PLAN DE PRÉSENTATION

- 1 CONTEXTE ET PROBLÉMATIQUE
- 2 OBJECTIFS
- 3 ÉTUDE ÉCOLOGIQUE
- 4 MÉTHODE D'ANALYSE
- 5 RÉSULTATS
- 6 DISCUSSIONS

⚠ ETUDE ÉCOLOGIQUE

- simple, rapide à mettre en œuvre,
- peu coûteuse en raison de la disponibilité des données exploitées,
- n'offre qu'une fenêtre d'observation et de ce fait ne pourra pas servir à établir avec certitude un quelconque lien causal.

LES DONNÉES

- ① Existence des grands pôles urbains, une concentration des cas → travailler à l'échelle de l'IRIS
- ② Amélioration de l'indicateur d'exposition indirect
- ③ intégrer d'autres facteurs individuels dont les effets cancérogènes sont bien établis.
- ④ s'affranchir du calcul des attendus → modèle Binomial Bêta à composante spatiale