

# Utilisation de données en libre accès pour caractériser le lien entre insuffisance rénale chronique terminale et environnement

Martin PAUMELLE, Florent OCCELLI, Caroline LANIER,  
Damien CUNY, Annabelle DERAM

## Mots clés

environnement – santé – insuffisance rénale – clusters spatiaux – étiologie

Certaines pathologies chroniques présentent une étiologie multifactorielle encore en partie indéterminée pour laquelle un lien avec l'environnement est suspecté. C'est le cas de l'insuffisance rénale chronique terminale (IRCT) pour laquelle un registre de santé (Néphronor) est tenu à jour depuis 2005 dans la région Hauts-de-France. Ces données recueillies par les professionnels de santé ont permis de calculer l'incidence de l'IRCT dans chaque commune et ainsi d'étudier son hétérogénéité dans l'espace. Grâce à des approches épidémiologiques géographiques, des clusters de sur-incidence et de sous-incidence<sup>1</sup> ont été détectés.

L'objectif de ces travaux est de contribuer à identifier de nouvelles pistes étiologiques en lien avec l'environnement. La démarche générale consiste à mettre en lien les données de santé avec des données environnementales<sup>2</sup> disponibles en libre accès. Après avoir défini les limites du sujet et les critères d'inclusion de nouvelles variables, une base de données environnementale a été construite. Un état des lieux a permis de retenir des données en *open data* couvrant des thématiques aussi diverses que l'occupation des sols, la pollution des milieux, la défaveur sociale, l'accès aux soins, etc.

Pour exploiter cette base de données et formuler des hypothèses sur l'étiologie de l'IRCT, deux méthodes de caractérisation de l'environnement ont été mises en œuvre. Une première consiste à utiliser des indices spatiaux composites : des indices de vulnérabilité et de résilience ont permis de quantifier le cumul d'effets délétères ou d'effets bénéfiques de l'environnement sur la santé, afin de tester le poids global de l'environnement sur la répartition spatiale d'incidence de l'IRCT. La seconde approche est une méthode de partitionnement du territoire qui permet de construire une typologie des communes en regroupant celles qui présentent un profil environnemental similaire. La méthode repose sur deux étapes successives : une analyse en composantes principales suivie d'une classification ascendante hiérarchique, afin de déterminer si certains profils environnementaux sont surreprésentés ou sous-représentés au sein des clusters.

---

1 Un cluster de sur-incidence est une zone géographique où l'incidence de la maladie est anormalement élevée par rapport à la zone d'étude. Au contraire, un cluster de sous-incidence présente une incidence anormalement faible.

2 L'environnement est compris au sens de milieu de vie des populations (physique, social, économique) en excluant ce qui relève du comportement individuel.

Dans ce contexte de recherche, l'avantage indéniable de l'*open data* est la possibilité de générer des hypothèses étiologiques à moindre coût. Avant d'envisager des études épidémiologiques individuelles beaucoup plus coûteuses, il permet de tester ces hypothèses à l'échelle du territoire grâce à une approche géographique et d'identifier des pistes qui pourront ensuite être mises à l'épreuve à l'échelle individuelle. Cependant son utilisation présente certaines limites dont une première assez évidente : la dépendance du travail de recherche vis-à-vis des données disponibles, qui doivent être fiables, pertinentes, exhaustives et disponibles à l'échelle considérée. D'autre part, il est crucial de garder un esprit critique sur la qualité des données utilisées et sur les résultats qui en découlent. Par exemple, pour étudier la contamination des milieux, il est assez rare de disposer de données directes mesurées dans l'environnement. Cela impose d'utiliser des proxys, c'est-à-dire d'évaluer indirectement le phénomène en considérant par exemple les sources d'émissions (avec toutes les limites que cela comporte). Un autre point à garder à l'esprit concerne la temporalité des données. Les clusters IRCT ont par exemple été détectés sur la période 2005-2018. Or il est généralement compliqué de remonter aussi loin dans le temps avec l'*open data*, sachant qu'idéalement il faudrait disposer de données encore plus anciennes pour prendre en compte le temps de latence potentiel entre une exposition à un environnement dégradé et le déclenchement d'une pathologie.

## Les auteur·es

### **Martin PAUMELLE**

Université de Lille – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)  
[martin.paumelle.etu@univ-lille.fr](mailto:martin.paumelle.etu@univ-lille.fr)

### **Florent OCCELLI**

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)  
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)  
[florent.occelli@univ-lille.fr](mailto:florent.occelli@univ-lille.fr)

### **Caroline LANIER**

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)  
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)  
[caroline.lanier@univ-lille.fr](mailto:caroline.lanier@univ-lille.fr)

### **Damien CUNY**

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)  
Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)  
[damien.cuny@univ-lille.fr](mailto:damien.cuny@univ-lille.fr)

### **Annabelle DERAM**

Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France – ULR4515 Laboratoire de génie civil et géo-environnement (LGCgE)  
Faculté d'ingénierie et management de la santé (ILIS), Faculté de pharmacie de Lille – Laboratoire de sciences végétales et fongiques (LSVF)  
[annabelle.deram@univ-lille.fr](mailto:annabelle.deram@univ-lille.fr)