

## **Pollution de l'air et surmortalité :**

### **Quand les particules fines nous enfument...**

Les études de surmortalité sont un outil statistique relativement récent, qui a permis des avancées significatives en matière d'épidémiologie : elles ont par exemple permis de réviser à la hausse les conséquences de la grippe dans les pays développés, et de démontrer que le chikungunya, maladie connue depuis longtemps, mais supposée non mortelle, avait en fait un impact bien réel sur la mortalité des populations concernées. Dans ce type d'étude, les épidémiologistes suivent les variations de la mortalité des populations concernées, le plus souvent à l'échelle journalière, pour quantifier l'effet du facteur étudié (par exemple une épidémie, ou un risque environnemental temporaire). Comme toute méthode statistique, il s'agit de techniques complexes, car elles doivent tenir compte de l'effet de nombreuses causes de variations de mortalité sans intérêt épidémiologique (par exemple des variations en fonction du jour de la semaine ou de la saison), et le plus souvent d'autres facteurs simultanés dont l'effet peut brouiller celui du facteur que l'on veut étudier : ce qu'on appelle les facteurs de confusion. Par exemple, les épidémies de grippe se produisant en hiver, il y a un risque de confondre l'effet de la grippe avec celui du froid. Pour éviter ce biais, on va donc construire en fait un modèle multivarié, dans lequel on va prendre en compte non seulement le facteur étudié, mais aussi l'ensemble des facteurs de confusion potentiels identifiés, afin de faire la part de chaque facteur.

Les modèles statistiques multivariés utilisés dans ce type d'étude doivent être employés avec précautions. Comme tout modèle statistique, leur qualité dépend fortement de la qualité des données d'entrée utilisées, et donc de la représentativité de la population sur laquelle on a travaillé. Il est déjà difficile d'obtenir ou de sélectionner des données pertinentes pour modéliser l'effet d'un facteur environnemental isolé. C'est encore plus compliqué, voire impossible, d'avoir des données permettant de modéliser correctement à la fois l'effet de ce facteur, et de l'ensemble de ses facteurs de confusion potentiels. Reprenons l'exemple de la grippe et du froid. Pour modéliser l'effet de la grippe sur la mortalité en France, on a intérêt à sélectionner une série de données récentes, ce qui permet de limiter les biais dus à l'évolution des méthodes de diagnostic et de prise en charge de la grippe. Par contre, des données récentes ne seront pas très pertinentes pour mesurer l'effet du froid sur la mortalité, car nous n'avons pas connu d'hiver très froid ces dernières années : pour cela, il serait préférable de choisir des données remontant jusqu'aux années 80, qui avaient connu plusieurs hivers très rigoureux. Un modèle obtenu avec des données récentes sur la grippe risque donc de n'attribuer aucun effet au froid, ce qui ne serait pas forcément le cas avec un modèle incluant des données plus anciennes. Et si le modèle multivarié sous-estime l'effet d'un facteur de confusion, cet effet sous-estimé risque d'être affecté par erreur à la variable principale.

Ces problèmes de modélisation des facteurs de confusion sont inévitables, et ne doivent pas discréditer ce type d'études. Par contre, ils devraient inciter les auteurs de ces études à prendre trois précautions élémentaires :

- Vérifier si les résultats obtenus avec ce modèle multivarié sont compatibles avec ce qu'on connaît de l'effet individuel de chaque facteur

- Comparer l'ordre de grandeur de l'effet de la variable d'intérêt (par exemple la grippe) à celui du facteur de confusion (par exemple le froid). Si la variable de confusion a un effet moyen beaucoup plus important que la variable étudiée, l'erreur sur la modélisation du facteur de confusion risque en effet d'être aussi forte que l'effet de la variable d'intérêt !
- Chaque fois que c'est possible, comparer l'effet de la variable d'intérêt sur deux sous-échantillons de la population de départ, l'un peu affecté par le facteur de confusion, l'autre fortement affecté. Cela permet de vérifier la validité du modèle, sur une sous population permettant d'exprimer au mieux l'effet de la variable d'intérêt

Ces précautions peuvent sembler évidentes, mais une étude récente publiée par l'INVS nous montre un bel exemple de publication où elles ont été allègrement foulées au pied. L'étude en question :

([http://www.invs.sante.fr/beh/2015/1-2/pdf/2015\\_1-2\\_3.pdf](http://www.invs.sante.fr/beh/2015/1-2/pdf/2015_1-2_3.pdf))

concerne l'effet des particules fines (PM10) sur la mortalité non accidentelle pendant les pics de pollution. L'INVS a analysé les courbes de mortalité journalière dans 17 villes françaises, et les a mises en relation avec la concentration de PM 10 dans l'air. Il en conclut que la pollution a un effet significatif sur la mortalité non accidentelle dans ces villes, en particulier pendant les jours 2 à 5 après le début du pic de pollution, sauf en été où un effet particulièrement élevé s'observe dès le jour 1. Ces conclusions sont particulièrement alarmantes, puisque l'INVS estime que cet effet s'observe même à des doses inférieures aux seuils d'alerte actuels. Logiquement, il faudrait donc en conclure que ces seuils devraient être abaissés, ce qui ne simplifiera pas la vie des citoyens dépendants de la voiture.

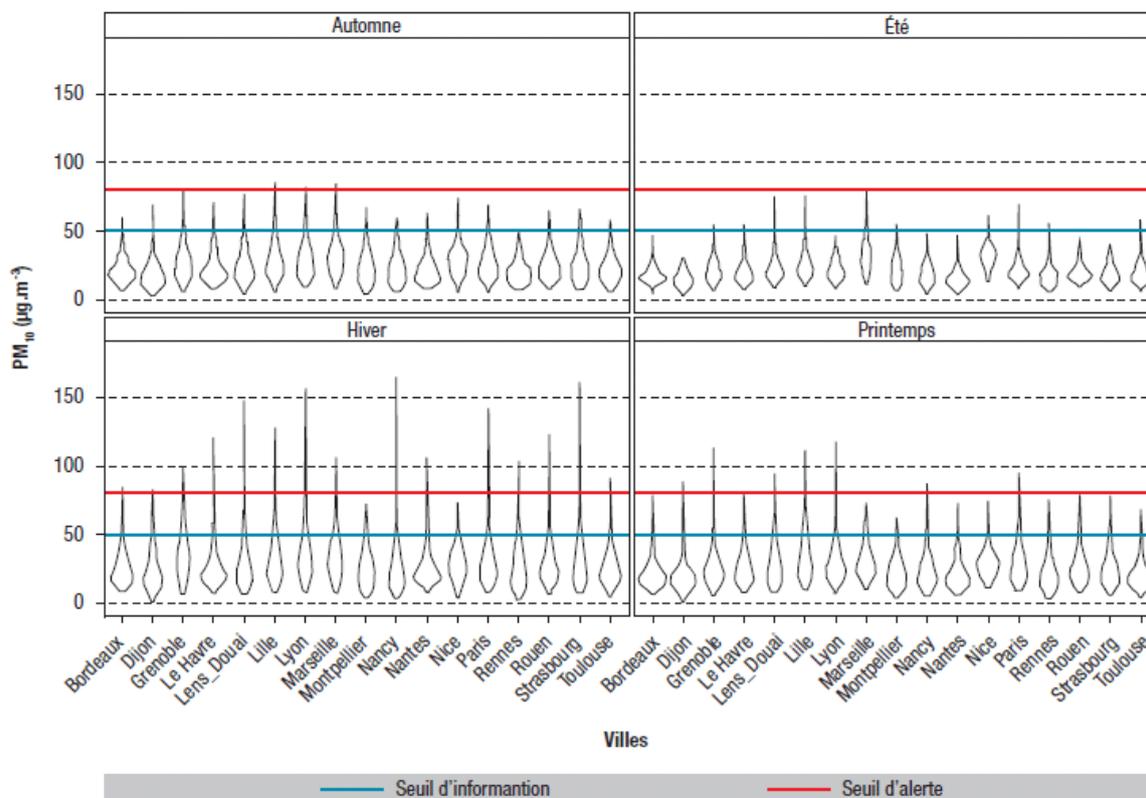
Quel rapport avec notre introduction ? C'est que ces résultats ont été obtenus avec un modèle intégrant la température comme facteur de confusion potentiel. En soi, c'est tout-à-fait logique, car les pics de pollution se produisent par temps anticyclonique. Ils sont donc souvent associés en hiver à des températures particulièrement froides, et en été aux périodes de canicule. L'INVS a donc travaillé avec un modèle multivarié prenant en compte non seulement les PM10, mais aussi la température, plus quelques autres facteurs de confusion possibles. Ce modèle devrait permettre de dissocier l'effet des PM10 et celui de la température. C'était bien sûr la démarche à suivre, mais la façon dont la température a été traitée est plutôt surprenante... surtout pour des chercheurs de l'INVS !

### **Quand l'INVS ignore les résultats de l'INVS**

Pour le comprendre, revenons d'abord aux résultats de l'étude. Celle-ci confirme un résultat déjà observé depuis longtemps, comme le rappellent les auteurs dans leur bibliographie : la corrélation entre taux de particules fines et mortalité est nette en été, mais beaucoup moins lors des autres saisons. Certes, dans leur tableau 3, les auteurs montrent qu'ils trouvent un effet significatif des particules PM10 sur l'ensemble de l'année. Mais si on examine le tableau 4 qui détaille les résultats en fonction des saisons, on constate que cet effet ne s'exerce en fait de façon significative qu'en été. Cela soulève déjà un paradoxe plutôt surprenant, quand on confronte ce résultat à la figure 1 de la page 18 : dans l'ensemble des villes étudiées, les concentrations en PM10 sont plus élevées en hiver qu'en été, en valeur moyenne et surtout en importance des pics de pollution. Pourtant, c'est seulement en été que les PM10 auraient un effet significatif sur la mortalité !

Figure

Distribution des concentrations de  $PM_{10}$  par saison et par ville selon le seuil d'information et d'alerte (17 villes, France métropolitaine, 2007-2010)



Source : Associations agréées de surveillance de la qualité de l'air (AASQA).

Figure p. 18 de l'étude INVS : on constate que dans toutes les villes de l'étude, les concentrations en  $PM_{10}$  sont plus faibles en été qu'en hiver. C'est vrai pour les valeurs moyennes, mais encore plus pour les valeurs des pics de pollution, aucune des villes n'ayant connu un dépassement du seuil d'alerte en été, alors que la plupart en ont connu en hiver.

Tableau 4

Excès de risque relatif (ERR) pour  $10 \mu g.m^{-3}$  de  $PM_{10}$  par cause de mortalité et saison (17 villes, France métropolitaine, 2007-2010)

	Mortalité non accidentelle			
	Printemps	Été	Automne	Hiver
Lag 0-1	0,36 [-0,26;1,00]	1,30 [0,06;2,56]*	-0,08 [-0,80;0,64]	-0,10 [-0,56;0,37]
Lag 2-5	0,35 [-0,49;1,20]	0,08 [-1,32;1,51]	-0,23 [-1,22;0,78]	0,28 [-0,26;0,82]
Lag 0-1 et 2-5	0,72 [-0,09;1,53]	1,38 [-0,06;2,85]	-0,31 [-1,45;0,85]	0,18 [-0,41;0,78]
	Mortalité cardiovasculaire			
	Printemps	Été	Automne	Hiver
Lag 0-1	-0,02 [-1,19;1,16]	2,06 [-0,34;4,52]	-0,86 [-2,25;0,55]	0,09 [-0,81; 0,99]
Lag 2-5	1,08 [-0,24;2,42]	0,92 [-1,80;3,71]	0,07 [-1,65;1,81]	0,07 [-1,15;1,31]
Lag 0-1 et 2-5	1,06 [-0,29;2,41]	3,00 [0,28;5,79]*	-0,79 [-2,74;1,19]	0,16 [-1,17;1,51]

\* Significativité.

IC95% : intervalle de confiance à 95%

**Tableau 4 de l'étude INVS : les seuls effets significatifs des PM 10 sur la mortalité seraient observés en été (saison où les concentrations sont pourtant les plus faibles). En hiver, elles n'ont aucun effet sur la mortalité, alors que leur concentration est nettement plus élevée. Un paradoxe qui n'est discuté nulle part par les auteurs.**

Aurions-nous donc affaire à un ennemi particulièrement fourbe, inoffensif à forte dose, mais mortel à dose plus faible ? Plus sérieusement, on peut bien sûr envisager une hypothèse non négligeable : les PM10 pourraient n'avoir aucun effet aux températures froides et moyennes, mais aggraver l'effet sur la santé des températures élevées, dont la canicule de 2003 a révélé tragiquement l'importance passée inaperçue jusqu'alors. Ou bien, inversement, les températures élevées incitent à aérer davantage les habitations, ce qui aggraverait l'exposition de leurs occupants, à concentration égale dans l'air extérieur.

Quoi qu'il en soit, c'est bien ici la température (et plus précisément les températures élevées) le facteur principal, et la concentration en PM10 le facteur secondaire. Du coup, nous sommes typiquement dans la situation que nous évoquions plus haut, où l'erreur résiduelle sur le redressement du facteur de confusion (la température) risque d'être plus élevée que l'effet du facteur étudié (la concentration en PM10). Il va donc falloir être vigilant sur la qualité de la prise en compte de la température dans le modèle.

A ce stade de la réflexion, nous avons déjà un premier souci : les auteurs ne nous indiquent nulle part l'effet de la température sur la mortalité, saison par saison. Du coup, nous n'avons aucun élément pour comparer son ordre de grandeur avec celui des PM10. Mais surtout, le mode de prise en compte de la température est pour le moins étrange, surtout pour une étude de l'INVS. En effet, les auteurs n'ont introduit dans leur modèle que la température moyenne. Or chacun sait grâce aux alertes canicules (dont les critères ont été définis précisément par l'INVS !) que la température moyenne à elle seule n'est pas l'indicateur le plus déterminant des risques pour la santé en cas de forte chaleur : les températures extrêmes de la journée, et en particulier la température minimale, sont un bien meilleur indicateur du risque pour la santé. (cf par exemple [http://www.invs.sante.fr/publications/2004/rapport\\_annuel\\_2003/alertes\\_canicule\\_2003.pdf](http://www.invs.sante.fr/publications/2004/rapport_annuel_2003/alertes_canicule_2003.pdf) ; depuis cette publication initiale, les seuils de risque ont été révisés plusieurs fois, mais les indicateurs calculés sont restés les mêmes). De plus, les seuils de risque employés varient considérablement selon les régions. Bien que reposant sur les mêmes techniques statistiques, le modèle employé pour définir les alertes canicule est donc beaucoup plus complexe.

Par conséquent, réfléchissons à ce qui va se passer dans le modèle utilisé pour cette étude PM10, pour les trois périodes suivantes :

	Température minimale (moyenne sur 3 jours)	Température moyenne sur 3 jours	Température maximale (moyenne sur 3 jours)	Alerte canicule (selon les seuils de Paris : Tmin >21°C et Tmax>31°C)
Période 1	20 ° C	27°C	34°C	Non
Période 2	22° C	27°C	32°C	Oui
Période 3	24°C	27°C	30°C	Non

*Tab. 1 : Selon les travaux de l'INVS sur l'effet de la canicule, ces trois périodes ne présentent pas les mêmes risques pour la santé : bien qu'elles aient la même température moyenne, seule la période 2 présente un risque*

*de surmortalité, car elle est la seule où le seuil critique soit dépassé à la fois pour la température minimale et la température maximale.*

Comme le modèle utilisé par l'INVS pour les PM10 ne prend en compte que la température moyenne, il va attribuer à la température de 27°C la surmortalité moyenne de ces 3 périodes. Il va donc considérer qu'une température moyenne de 27°C ne génère qu'une faible mortalité. Dans ce modèle, la surmortalité propre à la période 2 (la seule vraie période de canicule) ne va donc être redressée que de cet effet moyen. L'excès de mortalité restant va-t-il être considéré comme un effet résiduel aléatoire ? Sans doute pas, car ces périodes de canicule coïncident avec un temps calme, pendant lequel la pollution de l'air s'aggrave progressivement faute de vent. L'effet canicule va donc être attribué de manière erronée à un accroissement des PM10, même si leur concentration n'est pas particulièrement élevée à ce moment : il suffit qu'elle soit croissante. Il est donc évident que le modèle choisi par l'INVS génère un artefact qui présente exactement les mêmes caractéristiques troublantes que l'effet qu'il prête aux PM10 : il ne s'exerce qu'en été, et ne dépend pas de la concentration de PM10, puisqu'il est en fait dû aux températures minimales de la nuit !

Si on regarde les résultats avec un minimum d'objectivité, on peut donc craindre que l'effet estival des PM10 ne soit qu'un artefact du mode de redressement aberrant de la température choisi pour cette étude. Y a-t-il moyen de le vérifier ? Ce serait facile au vu des données présentées dans la Figure 1 de l'étude : plusieurs villes (Dijon, Lyon, Strasbourg) n'ont jamais connu de pic de PM10 estival, alors qu'elles connaissent souvent des températures caniculaires en été. A l'inverse, Lens-Douai, Lille, Marseille et Paris ont frôlé plusieurs fois le seuil d'alerte en été, et ont fréquemment dépassé le seuil d'information. Pour trancher le débat, il aurait suffi de comparer l'effet de la température dans ces deux groupes de villes, pour vérifier si les PM10 ont réellement un effet aggravant celui de la température. Pourquoi l'INVS n'a-t-il pas fait cette comparaison, qui serait le seul moyen de justifier son hypothèse pour le moins surprenante ?