

La prévision des risques de décès attribuables à des causes environnementales : l'exemple des expositions chroniques aux particules

Jean Christophe TURLOT

LMA, CNRS-UMR5142 – Université de Pau et des Pays de l'Adour

1. Les enjeux de l'évaluation des risques sanitaires associés à la pollution de l'air

L'exposition à long terme aux niveaux actuels de concentration ambiante des particules peut conduire à une réduction significative de l'espérance de vie. Des différences importantes dans la densité des particules fines en suspension dans l'air sont observées entre les communes peu polluées et les communes fortement polluées des États-Unis. Ces différences sont de l'ordre de $25 \mu\text{g}/\text{m}^3$. On estime à 1,5 années la différence de durée de vie des habitants de ces deux catégories de communes [B.OSTRO, 2004]. En France (chiffres 2002), on estime à 670 le nombre de décès annuels par cancer du poumon et à 3330 le nombre de décès annuels par accident cardio-pulmonaire attribuables à la pollution de l'air.

Il peut donc paraître paradoxal que l'existence d'une relation à long terme entre le niveau de pollution et la mortalité par cancer du poumon ou par accident cardiovasculaire soit encore discutée aujourd'hui. De l'étude de référence de POPE et *al.* [2002], il ressort que le risque relatif de décès associé à une augmentation de $10 \mu\text{g}/\text{m}^3$ des particules fines dans l'air est estimé à 1,06 pour le cancer du poumon et à 1,08 pour les accidents cardiovasculaires (résultats obtenus sur la base de la première étude de Pope 1982-1998). Ces risques relatifs sont peu élevés en regard du tabagisme par exemple. Il convient cependant de remarquer que ce facteur de risque concerne de grandes populations – en fait tous les habitants de zones urbanisées, ce qui fait que son impact n'est pas du tout négligeable comme on l'a rapporté plus haut, mais aussi que les estimations de ces risques ne sont pas très précises, d'où leur faible niveau de signification.

Récemment, l'OMS a fait des préconisations méthodologiques pour le calcul des risques de mortalité attribuables à la pollution de l'air (OMS, Londres 1999 ; B. OSTRO 2004). Elles reposent d'une part sur l'étude épidémiologique de POPE déjà citée et sur des calculs simples [KÜNZLI, 2000]. C'est donc le modèle de POPE que nous décrirons sous l'angle de l'estimation statistique du risque relatif RR ou du risque attribuable RA . Puis, sur la base des résultats reproduits dans l'étude de POPE, nous discuterons sur un exemple les préconisations de l'OMS et de la légitimité de la transposition des résultats de Pope à des populations différentes.

2. Risque relatif, risque attribuable

La modélisation du risque relatif de décès (RR) pour une cause spécifiée (par exemple le cancer du poumon) due à l'exposition à la pollution atmosphérique est à la base des calculs d'estimation de la surmortalité due à ce facteur de risque. Le risque attribuable (RA) mesure la probabilité de survenue de la maladie ou du décès due à l'exposition. L'expression du risque attribuable est fonction du risque relatif et de la prévalence de l'exposition.

La situation la plus simple consiste en deux états : décédé A , non décédé \bar{A} et deux niveaux d'exposition : exposé E , non exposé (ou exposition inférieure à un seuil donné) \bar{E} .

Tableau des probabilités des états	décédé (A)	Non décédé (\bar{A})	marge
Exposé (E)	$P(E, A) = \pi_{11}$	$P(E, \bar{A}) = \pi_{12}$	$P(E)$
Non exposé (\bar{E})	$P(\bar{E}, A) = \pi_{21}$	$P(\bar{E}, \bar{A}) = \pi_{22}$	$P(\bar{E})$
marge	$P(A)$	$P(\bar{A})$	1

On rappelle les définitions utiles :

$$\text{Le risque relatif : } RR = \frac{P(A/E)}{P(A/\bar{E})}$$

où $P(A/E)$ désigne la probabilité de décès en étant exposé et $P(A/\bar{E})$ la probabilité de décès en étant non exposé. Cette définition sous-tend une période d'observation spécifiée.

$$\text{L'excès de risque dû à l'exposition : } P(A/E) - P(A/\bar{E})$$

cela signifie que si n personnes sont exposées, le nombre de décès attribuables au facteur de risque vaut : $n \times [P(A/E) - P(A/\bar{E})]$.

$$\text{Le risque attribuable : } RA = \frac{P(E)[P(A/E) - P(A/\bar{E})]}{P(A)}$$

RA représente la probabilité de décès dû au facteur d'exposition. Le risque attribuable est relié au risque relatif par la relation :

$$RA = \frac{P(E)[RR-1]}{P(E)[RR-1] + 1}$$

où $P(E)$ mesure la prévalence de l'exposition.

Ce modèle est insuffisant puisque toute la population urbaine est exposée ($P(E)=1$) et que l'intensité de la pollution n'est pas prise en compte d'autre part. Le facteur de risque est décrit par une mesure numérique X évaluant la pollution annuelle moyenne par unité géographique (l'unité urbaine en France, des unités plus fines en Suisse par exemple). On peut classer les données d'observation sur une échelle de niveaux d'exposition. Un niveau seuil (où niveau de référence) est nécessaire : il est défini par x_0 qui représente un niveau de pollution de fond dont la valeur préconisée par l'OMS [2004] est de $3,5 \mu\text{g}/\text{m}^3$. Le seuil peut aussi être choisi comme étant le plus bas niveau de pollution relevé dans l'étude (par exemple $7,5 \mu\text{g}/\text{m}^3$ dans l'étude de Pope [2002]). Les personnes vivant dans un environnement dont le niveau de pollution ne dépasse pas ce seuil constituent la population témoin (considérée

comme non exposée) \bar{E} . On note E_1, E_2, \dots, E_K les niveaux successifs gradués d'exposition au risque (qu'il conviendra de préciser).

$$\text{À chaque niveau } E_i \text{ correspond un risque relatif } RR_i = \frac{P(A/E_i)}{P(A/E_0)}$$

Ce rapport n'est pas intrinsèque, puisqu'il fait référence à un état de « non exposé » défini de manière plus ou moins arbitraire par un seuil, et cela se retrouve au niveau du risque attribuable :

$$RA_i = \frac{P(E_i) [P(A/E_i) - P(A/E_0)]}{P(A)}$$

En cumulant les risques attribuables calculés pour chaque niveau d'exposition, on a :

$$[1] \quad RA = \frac{1}{P(A)} \sum_{1 \leq i \leq K} P(E_i) [P(A/E_i) - P(A/E_0)]$$

ce qui s'écrit encore :

$$[1b] \quad RA = \frac{\sum_{1 \leq i \leq K} P(E_i) (RR_i - 1)}{1 + \sum_{1 \leq i \leq K} [P(E_i) RR_i - P(E_i)]}$$

Le facteur de risque étant le niveau de pollution x mesuré par la densité des particules PM_{10} et $PM_{2,5}$, il est de nature continue. Dans cette situation, le risque relatif est décrit par une fonction dose-réponse $RR(x)$.

Les modèles dose-réponse utilisés pour l'évaluation du risque attribuable à la pollution atmosphérique (B.OSTRO, WHO 2004) sont les suivants:

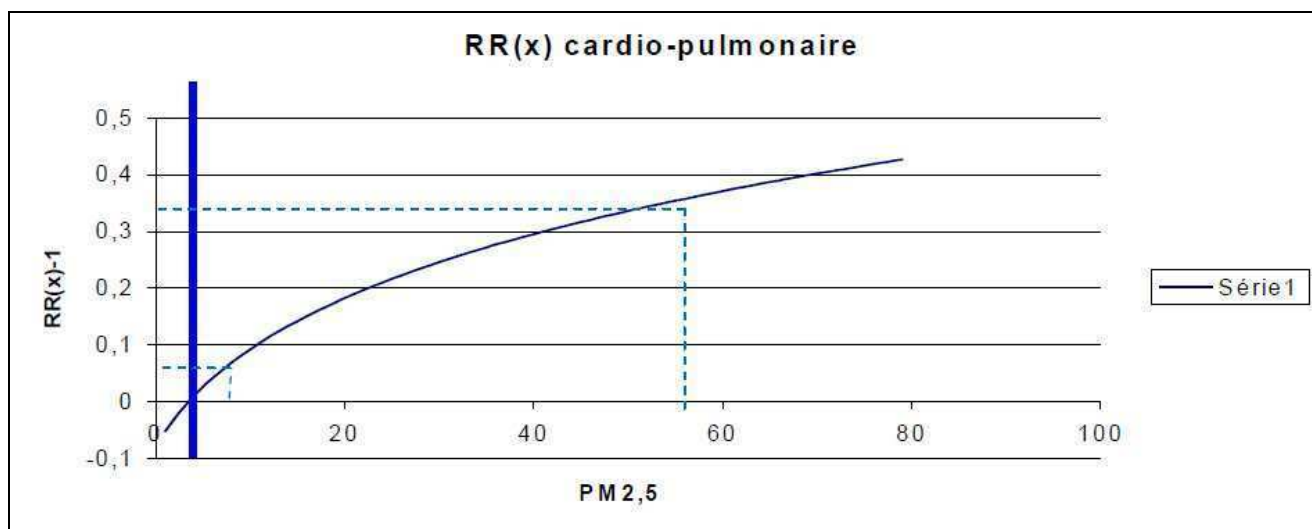
Le modèle log- linéaire :

$$[2] \quad RR(x) = \exp(\alpha + \beta \ln(x+1)) / \exp(\alpha + \beta \ln(x_0 + 1))$$

où x représente la concentration de l'air en $PM_{2,5}$. Cette fonction dose-réponse est recommandée par l'OMS pour modéliser la mortalité cardio-pulmonaire, mais aussi la mortalité par cancer du poumon. Le paramètre d'intérêt est ici β . Le modèle s'écrit encore :

$$[3] \quad \ln RR(x) = \left[\frac{x+1}{x_0+1} \right]^\beta$$

FIGURE 1 : $RR(x)$ EST PROCHE DE 1,05 À LA VALEUR SEUIL DE $7,5\mu\text{G}/\text{M}^3$ POUR ATTEINDRE UNE VALEUR VOISINE DE 1,35 POUR UN NIVEAU DE POLLUTION DE $50\mu\text{G}/\text{M}^3$. LA FONCTION EST DÉTERMINÉE POUR LA VALEUR $\beta = 0,1551$ (POPE, 1995).



Le modèle linéaire a aussi été choisi par certains auteurs [Pope, 2002]:

$$[4] \quad RR(x) = \exp(\alpha + \beta_L x) / \exp(\alpha + \beta x_0) = \exp(\beta_L (x - x_0))$$

Pour l'utilisation de ce modèle, le choix de la valeur $\beta_L = 0,008933$ est recommandé par l'OMS pour l'estimation de la mortalité cardio-pulmonaire.

Les modèles [3] et [4] sont assez proches dans l'intervalle des valeurs concernées. Le modèle [3] est recommandé lorsque certains niveaux de pollution sont vraiment élevés dans certaines unités géographiques ($> 50\mu\text{g}/\text{m}^3$) (B. OSTRO, 2004).

Comme les calculs préconisés par l'OMS sont fondés sur les résultats de l'étude de Pope, l'incertitude dans la prévision du nombre de décès attribuables est directement liée à l'incertitude portant sur l'estimation de β obtenue par l'étude expérimentale menée par Pope qui est la seule étude de référence (avec une étude antérieure du même auteur en 2005). De fait, cette incertitude se reporte sur le calcul du risque attribuable que l'on trouve dans toutes les études s'appuyant sur les résultats de cette seule étude de référence menée dans les grandes villes des États-Unis, par simple extrapolation, c'est à dire sans la prise en compte d'autres facteurs que l'on ne peut pas ignorer a priori.

Deux aspects dans l'utilisation critique du modèle sont à distinguer :

- La qualité du modèle du risque relatif élaboré par Pope pour l'estimation du paramètre d'intérêt β . Pour cela on reprendra sa construction qui en fait un modèle complexe. Sa qualité est mesurée par son niveau d'adéquation aux données expérimentales ; le niveau d'adéquation se mesure par l'amplitude du biais (qui doit être faible) et par l'importance de la variance de l'estimateur de la quantité d'intérêt β (qui doit être petite). Ces caractéristiques se reportent sur $RR(x)$ et RA .
- L'adaptabilité des résultats du modèle de Pope à d'autres populations. En effet, le modèle est utilisé dans le cadre classique de la prévision pour un autre échantillon. Cette démarche est légitime lorsque, hormis les niveaux des facteurs ou variables sous contrôle, les personnes constituant le nouvel échantillon sont en tous points comparables.

Nous allons présenter formellement ces deux points, qui sont essentiellement distincts.

3. L'étude de référence (POPE et al, 2002)

Il s'agit de l'analyse du suivi d'une cohorte (359.000 personnes recrutées en 1982) dont le nombre et les causes de décès par cancer du poumon et par maladie cardio-pulmonaire ont été inventoriés entre 1982 et 1998. La population cible est celle des adultes de plus de 30 ans. Plusieurs indicateurs du niveau de pollution ont été suivis dans le temps, dont le PM_{10} et le $PM_{2,5}$. La composition de la cohorte a été fondée sur les concentrations moyennes journalières des polluants dans l'air observées durant un à deux ans avant la sélection des participants. L'indice d'exposition à long terme retenu est la moyenne de $PM_{2,5}$ calculée sur la période 1979 - 1983. Une seconde étude a été réalisée en 1999 et 2000 sur une nouvelle population.

L'indice calculé est le risque relatif $RR(x)$ ajusté selon plusieurs facteurs de confusion: ce sont l'âge, le sexe, la race, le tabagisme, l'exposition professionnelle, la consommation d'alcool, le niveau d'éducation, le statut marital, le régime alimentaire. Ces informations ont été recueillies au moyen d'un questionnaire confidentiel rempli au départ de l'étude.

3.1. Le modèle statistique

Le cœur du modèle est l'estimation du risque relatif $RR(x)$ fondé sur le modèle de survie de COX standard. On procède en deux étapes :

1. Le modèle de survie de COX – stratifié selon l'âge, le sexe, la race et incluant les covariables d'ajustement qui sont assez nombreuses ainsi que l'indicatrice de l'aire métropolitaine d'habitation (c) – conduit à une estimation du logarithme du risque relatif ajusté pour chacune des aires métropolitaines que l'on note $\log RR_s(c)$.
2. Les valeurs $\log RR_s(c)$ sont mises en relation avec le niveau moyen de pollution mesuré par $PM_{2,5}(c)$, en faisant appel à un modèle linéaire à effet aléatoire (aréolaire) afin de prendre en compte les liaisons spatiales éventuelles entre les données.

La variable temps t entrant dans le modèle de COX est la durée de survie à partir de l'entrée dans la cohorte. Cette durée est censurée à droite chez les personnes non décédées à la date de la fin d'étude. Les niveaux de pollution ont été calculés sur la base des concentrations en $PM_{2,5}$ durant les périodes (1979-1983), puis (1999-2000) et par la moyenne des deux. Les facteurs individuels de risque sont décrits soit par des variables à modalités, soit par des variables continues [POPE et al. 2002, KREWSKI & al, 2000].

3.2. Le modèle général stratifié de COX

Ce modèle permet d'estimer la fonction de survie $S(t) = Prob(T \geq t)$ où T est la variable aléatoire décrivant la durée de vie de la population d'étude. Pour cela on modélise la fonction de hasard $h(t)$ qui est le cœur du modèle :

$$[5] \quad h(t) = \lim_{\delta_i \rightarrow 0} \left\{ \frac{Prob[t \leq T \leq t + \delta_i | T \geq t]}{\delta_i} \right\}$$

La fonction de survie $S(t)$ est reliée à la fonction de hasard (COLLETT, 1999) par la relation :

$$S(t) = \exp\left\{-\int_0^t h(u) du\right\}$$

Considérons une fonction de hasard de référence $h_0(t)$; le modèle des hasards proportionnels consiste à associer à l'individu i une fonction de hasard $h_i(t) = \psi(x_i)h_0(t)$ intégrant ses caractéristiques propres $x_i = (x_{i1}, \dots, x_{ij}, \dots, x_{ip})$ exprimées sous la forme :

$$h_i(t) = \exp(\beta_1 x_{i1} + \dots + \beta_p x_{ip}) h_0(t).$$

Le logarithme du risque relatif instantané $RRI(i)$ s'écrit :

$$[6] \quad \text{Log}\left(\frac{h_i(t)}{h_0(t)}\right) = (\beta_1 x_{i1} + \dots + \beta_p x_{ip})$$

Si deux personnes i et i' ne diffèrent que par le niveau de la pollution atmosphérique mesurée par X_1 , le **rapport des risques instantanés** vaut $\text{Log}\left(\frac{h_i(t)}{h_{i'}(t)}\right) = \beta_1 (x_{i1} - x_{i'1})$.

Cette expression justifie la présentation du risque relatif instantané comme mesure de la réduction du taux instantané de décès dû à une baisse de $10\mu\text{g}/\text{m}^3$ des particules fines $PM_{2.5}$.

Le modèle de COX a été stratifié : la fonction de hasard de référence $h_0(t)$ a été choisie différente selon l'âge, le sexe et le type racial. À la strate j caractérisée par un âge, un sexe et un type racial donnés correspond la fonction de référence $h_{0j}(t)$ et le rapport de risques instantanés s'écrit :

$$[7] \quad \text{Log}\left(\frac{h_{ij}(t)}{h_{0j}(t)}\right) = \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad 1 \leq i \leq n_j \text{ (effectif dans la strate } j), \quad 1 \leq j \leq J$$

(nombre de strates).

Le choix d'utiliser le modèle de COX par strates est nécessité par le fait que la fonction de hasard $h_0(t)$ décrivant le taux instantané de décès (et donc la fonction de survie $S(t)$) dans la classe non exposée évolue selon l'âge. La stratification rend plus réaliste l'hypothèse des risques proportionnels. Cette approche a également été utilisée par ABBEY et al. [1999].

Les variables d'ajustement décrites par le vecteur (x_{i1}, \dots, x_{ip}) sont la commune d'habitation c et les réponses au questionnaire concernant l'usage du tabac, l'exposition particulière à certaines formes de pollution, le niveau d'éducation, etc. [POPE et al, 2002].

Les paramètres d'intérêt (les β_j , $j \geq 1$) sont estimés par la méthode du maximum de vraisemblance avec le modèle [7], l'échantillon étant censuré à droite par la date de la fin du suivi de la cohorte. On en déduit [COLLETT, 1999] un intervalle de confiance asymptotique pour les paramètres fondé sur la matrice d'information $I(\hat{\beta})$ de Fisher :

$$Var(\hat{\beta}) \approx I^{-1}(\hat{\beta}) \quad \text{avec} \quad I(\hat{\beta})_{jl} = - \left(\frac{\partial^2 \text{Log}L(\beta)}{\partial \beta_j \partial \beta_l} \right)_{j|\hat{\beta}}$$

La première étape du modèle de POPE ne prend pas directement en compte la mesure du niveau de pollution $x_1(i) = PM_{2.5}(i)$ associée à chaque personne, mais l'indicatrice $I_c(i)$ de la commune c d'habitation. Pour l'analyse des particules fines, les individus sont issus de 61 communes d'habitation présentant différents niveaux de pollution.

De l'estimation des paramètres $\hat{\beta}_1 = (\hat{\beta}_1^c, 1 \leq c \leq 61)$ pour chaque commune c , on déduit l'estimation de la fonction de survie $\hat{S}_{i,j}^c(t) = \exp\left\{-\int_0^t \hat{h}_{i,j}^c(u) du\right\}$, ajustée par les facteurs individuels de risque $(x_{i1} = c, x_{i2}, \dots, x_{ip})$. On a pour tout t :

$$\hat{S}_{i,j}^c(t) = \left(\hat{S}_{i,j}^0(t)\right)^{\hat{\gamma}_c} \quad \text{avec} \quad \hat{\gamma}_c = \exp(\hat{\beta}_1^c - \hat{\beta}_1^0)$$

où les indices désignent deux individus de la même strate j présentant les mêmes facteurs individuels de risques (x_{i2}, \dots, x_{ip}) , hormis la ville de résidence $x_{i1} = c$ pour l'un et 0 (commune de référence) pour l'autre. On en déduit :

$$[8] \quad \text{Log} \left(\frac{\hat{S}_{i,j}^c(t)}{\hat{S}_{i,j}^0(t)} \right) = (\hat{\gamma}_c - 1) \text{Log} \hat{S}_{i,j}^0(t) \quad \text{pour tout } t,$$

ou de manière équivalente :

$$\text{Log} \left(\hat{S}_{i,j}^c(t) \right) = \hat{\gamma}_c \text{Log} \hat{S}_{i,j}^0(t)$$

Cette formule provient de l'expression du risque relatif instantané RRI estimé :

$$[9] \quad \text{Log} \left(\frac{\hat{h}_{i,j}^c(t)}{\hat{h}_{i,j}^0(t)} \right) = (\hat{\beta}_1^c - \hat{\beta}_1^0)$$

L'estimation du risque relatif pendant la période T d'observation :

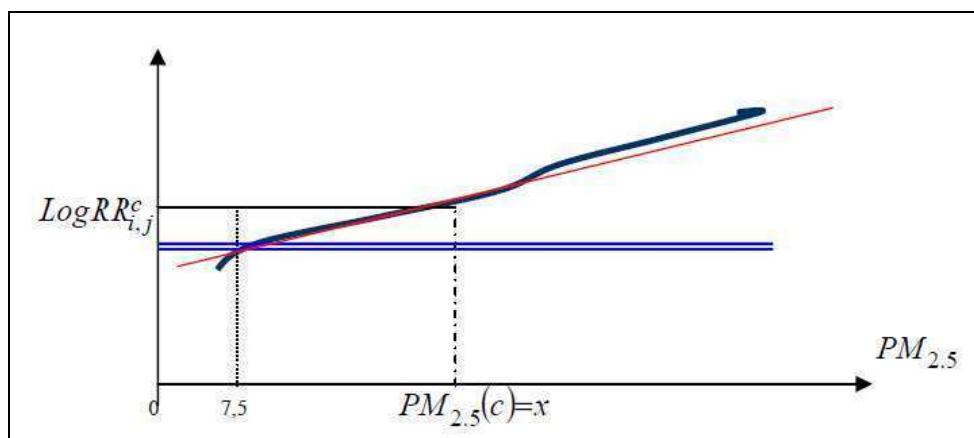
$$[10] \quad RR_{i,j}^c = \frac{1 - \hat{S}_{i,j}^c(T)}{1 - \hat{S}_{i,j}^0(T)}$$

ne s'exprime malheureusement pas sous une forme aussi simple que [4], [8] ou [9].

3.3. La relation entre le risque relatif estimé par commune et la pollution de l'air

Dans une seconde phase, POPE et al mettent en relation le risque relatif ajusté estimé $RR_{i,j}^c$ [10] et le niveau moyen de pollution observé dans les communes c . En ajustant dans chaque strate j la fonction $RR_{i,j}^c$ en fonction du niveau de $PM_{2.5}$ par une régression, on obtient une fonction de réponse de la forme décrite dans la figure 2 :

FIG.2 : L'ÉVOLUTION DU RISQUE RELATIF AJUSTÉ EN FONCTION DU NIVEAU DE $PM_{2.5}(c)$ DANS LA STRATE J .



3.4. La précision des estimations

Dans chaque strate j on peut calculer l'écart type asymptotique de $LogRR_{i,j}^c$ et faire l'ajustement par une fonction dose-réponse de la forme :

$$[11] \quad LogRR_{i,j}(x) = b_{i,j}(x - x_0) + e$$

Dans la seconde phase du modèle de POPE décrite par [11], l'erreur e présente deux composantes : l'une est un effet aléatoire $e_1(c)$ lié à la commune, l'autre $\varepsilon(i)$ est de nature purement aléatoire. La première composante de l'aléa est justifiée par la possibilité d'un effet spécifique associé à chacune des communes c qui soit lié à des considérations environnementales non prises en compte ou mal prises en compte par le modèle ; la seconde composante $\varepsilon(i)$ décrit les fluctuations individuelles.

3.5. Les principaux résultats de l'étude de Pope et al. (2002)

POPE et al. [2002] ne donnent qu'un seul tableau de résultats chiffrés : le risque relatif de mortalité ajusté associé à une augmentation de $10 \mu g / m^3$ des particules fines $PM_{2.5}$ calculé pour la première étude (1982-1998), puis pour la seconde (1999-2000), enfin pour les deux combinées. Les résultats de la première étude donnent :

– pour la mortalité cardio-vasculaire :

$$\hat{\beta} = 1,06 \text{ avec un intervalle de confiance (à 95\%)} \text{ de } [1,02 - 1,10]$$

– pour la mortalité par cancer du poumon :

$$\hat{\beta} = 1,08 \text{ avec un intervalle de confiance (à 95\%)} \text{ de } [1,01 - 1,16]$$

Le fait que les bornes inférieures des intervalles de confiance soient proches de 1 signifie que les pentes ajustées sont *à la limite du seuil de signification*.

Comme résultats complémentaires de leur étude (présentés non pas sous forme chiffrée, mais au moyen de figures) on constate que le facteur d'ajustement majeur est le tabac, que le statut marital et le niveau d'éducation semblent également jouer un rôle dans l'ajustement, mais que ni les autres facteurs, ni l'introduction de l'effet aléatoire ne modifient de manière significative le risque relatif.

Le modèle proposé par l'OMS [OSTRO, 2004] pour le calcul du risque attribuable RRA , est une forme de simplification du modèle de POPE et al. Avant de discuter ce point, voici

comment l'OMS propose de calculer le risque attribuable au niveau d'une métropole en prenant pour exemple Bangkok, Thaïlande.

4. Les calculs préconisés par l'OMS

L'équation donnant le nombre de décès E imputables aux accidents cardio-pulmonaires pour une année donnée s'écrit :

$$E = RA(x) \times I_M \times P$$

où I_M représente le nombre de décès causée par les accidents cardio-pulmonaires au cours de l'année considérée, P la taille de la population d'étude et $RA(x)$ le risque attribuable, qui dépend du niveau de pollution. Seul $RA(x)$ est considéré comme aléatoire, la taille de la population d'étude est supposée parfaitement connue ainsi que le nombre de décès causés par la maladie M (les accidents cardio-vasculaires).

Formule de calcul préconisée par l'OMS : lorsque le niveau x des particules fines en suspension ne dépasse pas $50 \mu\text{g}/\text{m}^3$, L'OMS propose comme fonction dose-réponse pour la mortalité cardio-vasculaire :

$$\text{Log}RR(x) = \beta(x - x_0) \text{ avec } \beta = 0,00893 \text{ et } IC(\beta, 0,95) = [0,00322 - 0,01464]$$

et comme valeur seuil $x_0 = 3 \mu\text{g}/\text{m}^3$ pour les particules fines $PM_{2,5}$.

- La concentration ambiante de $PM_{2,5}$ est évaluée à $35,1 \mu\text{g}/\text{m}^3$ dans la ville de Bangkok, on en déduit le risque relatif estimé : $RR(x) = 1,33$. Puis en utilisant [1b] pour un seul niveau de risque commun à l'ensemble de la population des habitants de plus de 30 ans, on obtient le risque attribuable $RA(x) = [RR(x) - 1] / RR(x) = 0,33/1,33 = 0,248$.
- La taille P de la population des habitants de Bangkok ayant plus de 30 ans est évaluée à 3 millions de personnes.
- Le taux de mortalité I_M pour cause d'accident cardio-pulmonaire est de 2,3 pour 1000.

On en déduit $E = 1710$

De l'intervalle de confiance de β on déduit un intervalle de confiance pour $RR(x)$ puis pour $RA(x)$; celui-ci vaut : $[0,098 - 0,374]$. D'où l'intervalle de confiance à 95% du nombre de décès attribuables : $[675 - 2580]$.

5. Commentaires sur la méthodologie

La démarche proposée par l'OMS suppose connu le taux de mortalité de la maladie étudiée dans la zone de l'étude, ainsi que la population d'étude qui est restreinte aux habitants de plus de 30 ans. Le nombre de cas attribuables est directement proportionnel à ces deux quantités et l'impact d'une sous-évaluation du taux de mortalité par exemple, est directement reporté sur le calcul du nombre de décès attribuables. Deux questions sont stratégiques pour légitimer la méthode d'estimation proposée :

- L'indicateur choisi mesurant le niveau de la pollution atmosphérique est-il pertinent dans la relation entre les décès par cancer du poumon ou par maladie cardio-respiratoire et la pollution atmosphérique ?
- Le modèle statistique retenu reflète-t-il correctement la relation entre la pollution atmosphérique mesurée et l'incidence des risques sanitaires étudiés ?

5.1. Le choix de l'indicateur recouvre plusieurs considérations.

- $PM_{2.5}$: nous ne discuterons pas de la pertinence du choix de la densité de $PM_{2.5}$ dans l'atmosphère comme indicateur de pollution en relation avec les risques sanitaires, car cette question touche au problème de la causalité, encore mal connu actuellement (AFSSE, 2004). Il convient toutefois de constater que cet indicateur a été souvent utilisé dans les études épidémiologiques menées à travers le monde et que la densité des PM a révélé dans la plupart des études (à long terme comme à court terme) une liaison entre mortalité et pollution atmosphérique (KÜNZLI, 2002 ; AFSSE 2004 ; OSTRO, 2004).
- La mesure de $PM_{2.5}$: si l'on considère comme indicateur de pollution la moyenne \bar{x}_c de la densité des particules fines observée dans la commune d'habitation c au cours des dernières années précédant l'estimation du risque, cela signifie que tous les habitants d'une même commune ou d'une même communauté urbaine sont exposés à un même niveau de pollution. C'est le choix préconisé par l'OMS.
- Un niveau d'exposition non individualisé : l'évaluation du risque à long terme sous-tend que l'on fait référence au cumul sur plusieurs années de l'exposition à un certain niveau de pollution. Si une personne réside depuis peu dans une commune fortement polluée, son risque $RR(x)$ n'est pas comparable à celui d'un habitant du même âge né dans cette commune. De même l'exposition au quotidien varie selon l'activité des habitants.

L'individualisation du niveau du risque constitue un élément important de l'étude de POPE (par la prise en compte de facteurs d'ajustement assez nombreux issus des réponses à un questionnaire et les personnes de la cohorte soumises à des risques particuliers sont connues).

Ce souci d'individualisation dans le calcul du risque attribuable n'est pas repris dans le calcul préconisé par l'OMS (cf. exemple Bangkok ci-dessus).

5.2. Le modèle statistique

De la pertinence du modèle statistique retenu dépend la qualité de l'estimation de la quantité recherchée, c'est à dire du risque attribuable. La qualité d'un modèle est mesurée à la fois par l'amplitude du biais d'estimation, qui doit être faible et par sa précision (la variance de l'estimateur du risque attribuable doit être petite ou – de manière équivalente – l'amplitude de l'intervalle de confiance doit être faible).

5.3. Le biais.

L'amplitude du biais ne peut pas être connue (même approximativement) dans ce type de problématique ; on peut cependant identifier des sources possibles de biais du modèle :

- Le biais par sous-estimation de la mesure individuelle x_i^1 de l'amplitude de l'exposition au risque de pollution. Il s'agit d'un biais dû à une évaluation insuffisante du niveau de pollution décrit par la variable X^1 .

Dans l'étude de POPE plusieurs sources de mesures de la pollution ont été agrégées sous la forme d'un indicateur en termes de niveau de $PM_{2.5}$ (en particulier la conversion en $PM_{2.5}$

des particules PM_{10} plus grossières mesurées par certains capteurs). Dans cette même étude, les auteurs ont pris soin d'écartier les communes pour lesquelles les mesures sont insuffisantes ou de mauvaise qualité.

- Le biais dû à la non prise en compte de l'historique des participants avant leur entrée dans la cohorte. Un questionnaire a été rempli par les participants au départ de l'étude, mais nous n'avons pas trouvé dans l'article de Pope de référence à une sélection éventuelle des participants sur des conditions d'ancienneté de résidence dans la commune. Le biais que cela peut entraîner est un biais de choix du modèle : des variables d'ajustement influentes n'ont pas été prises en compte.
- Le modèle de COX, quoique très utilisé, fait appel à une hypothèse forte qui est l'hypothèse des hasards proportionnels. En choisissant de stratifier la fonction de risque de référence selon l'âge en particulier, POPE et al se protègent contre le risque d'une hypothèse illusoire qui serait source d'un biais dû à une erreur de modèle.
- Le modèle de COX retenu intègre des variables d'ajustement sous forme additive ; en particulier, la consommation de tabac et la pollution sont considérées comme ayant des effets additifs sur $\text{Log}RR(x)$. Cependant, il se peut que si l'on dépasse un certain seuil dans la consommation de tabac, le niveau de la pollution urbaine ne modifie plus le risque de décès, auquel cas l'hypothèse d'additivité des effets est mise en défaut. D'où une autre forme de biais dû au choix du modèle.

5.4. La variabilité

La variabilité du risque attribuable exprimé par un nombre de décès peut sembler importante pour le décideur qui doit évaluer l'impact d'une mesure de réduction de la pollution de l'air. L'étendue de l'intervalle de confiance est due à la variance de la pente $\hat{\beta}$ estimée selon le modèle de POPE et retenue par l'OMS. Dans l'exemple de Bangkok, on a vu que le nombre de décès attribuables était, pour les accidents cardio-vasculaires, dans la fourchette $IC(95\%) = [675 - 2580]$. Par un calcul direct on trouve qu'une réduction de $10 \mu g / m^3$ de $PM_{2.5}$ devrait engendrer à terme une réduction de 500 décès environ ($IC = [156 - 1907]$). On voit que les deux intervalles de confiance se recouvrent très largement.

Comme la cohorte de POPE est de taille importante et suivie sur une assez longue période, la variabilité observée est due essentiellement à la « pauvreté » du pouvoir explicatif des prédicteurs retenus. La précision du modèle préconisé ne pourra être revue que par la constitution d'une nouvelle cohorte et par son suivi intégrant :

- les connaissances récentes sur les liens entre la pollution atmosphérique et ses conséquences sanitaires. La composition de $PM_{2.5}$ est hétérogène et sa nocivité peut être variable et peut évoluer au cours de l'étude (AFSSE, 2004).
- le perfectionnement et le déploiement des appareils de surveillance : cela devrait permettre une estimation plus localisée et donc plus précise de la pollution au niveau des quartiers et donc une individualisation plus marquée du niveau d'exposition au risque.
- certaines des covariables intégrées dans le modèle de POPE.

5.5. Approche critique de l'application des résultats de l'étude de Pope [2002] à d'autres populations : la démarche prévisionnelle préconisée par l'OMS.

La démarche classique de la prévision consistant à ajuster un modèle sur un échantillon tiré aléatoirement dans une population, puis à l'appliquer à un nouvel échantillon n'est valable que si le nouvel échantillon est considéré comme étant tiré aléatoirement dans la même population de référence que celui ayant servi à l'ajustement (aux facteurs contrôlés près). Hormis le fait que les études soient restreintes aux personnes de plus de 30 ans, cette hypothèse légitimant l'inférence statistique n'est pas nécessairement valide. Dans l'exemple du calcul concernant Bangkok, on peut penser que le mode de vie et les coutumes font que l'exposition à la pollution atmosphérique n'est pas la même que dans une ville des USA présentant le même indice de pollution ($PM_{2,5} = 35,1 \mu g / m^3$). Rien ne permet d'affirmer non plus que la composition chimique de $PM_{2,5}$ soit la même. Détaillons certains aspects nous écartant des hypothèses nécessaires à la prévision :

- La mesure du niveau de pollution : cette mesure doit être élaborée dans des conditions identiques à celles de l'étude de référence. Si certaines sources localisées de pollution sont mal prises en compte du fait d'un nombre insuffisant de capteurs, il y a risque de sous estimation du niveau de la pollution. De même, comme on l'a dit plus haut, la composition des particules fines peut varier dans la nocivité selon les pays ou la méthode est appliquée.
- L'âge : cette variable influe de manière importante sur l'ajustement du risque relatif (Pope, 2002). En choisissant un modèle de COX stratifié, on fait appel à des fonctions de hasard de référence différentes selon l'âge et le sexe. On peut penser que dans des pays où l'espérance de vie est très différente de celle des États-Unis, les fonctions de hasard devraient être assez différentes (et donc les estimations de pentes aussi). Dans son rapport sur l'impact sanitaire de la pollution atmosphérique urbaine en France métropolitaine en 2002, l'AFSSE (2004) a cru bon de considérer trois classes d'âge. À partir de l'estimation des pentes dans chacune de ces trois classes d'âge (fournies par POPE et al), on ajuste l'estimation du RRA. Cela présuppose que le modèle de POPE est transposable en France.
- Les circonstances d'exposition : elles sont différentes, ainsi que leur durée, selon le mode de vie dans la zone étudiée. Elles évoluent avec le temps également.
- La nature physico-chimique des particules fines et des polluants en suspension : elle peut être différente selon les pays où le modèle est appliqué (normes de rejets des usines, nature des carburants utilisés, etc.).

Il faut donc être critique dans l'interprétation des résultats fournis par la méthodologie préconisée. Il ne serait toutefois pas correct d'omettre de dire qu'à travers les articles sur cette question donnés en bibliographie, les auteurs montrent la plus grande prudence dans l'interprétation de leurs résultats. En témoigne l'expertise jugée nécessaire pour confirmer les résultats de POPE [KREWSKI et al. (2000)] au moyen de diverses analyses de sensibilité (choix du modèle, de la valeur seuil, etc.) et que l'on retrouve dans la plupart des études menées. On conclura par ce propos que l'on peut lire dans l'introduction du document de l'AFSSE(2004) : « Malgré les incertitudes fortes qui demeurent, les enjeux sanitaires principaux de la maîtrise des émissions de polluants dans l'atmosphère sont sans doute représentés par les possibles effets résultants d'une exposition chronique, même modeste, dans une population en général composite. Il est néanmoins important aujourd'hui de fournir aux acteurs concernés des informations utiles à l'éclairage des politiques ou actions qu'ils pourraient mener, en s'appuyant sur les données scientifiques présentes, mêmes imparfaites ».

BIBLIOGRAPHIE

- N. KÜNZLI, R. KAISER, S. MEDINA et al. Public-health impact of outdoor and traffic-related air pollution : a european assessment. *THE LANCET*, Septembre 2000.
- AGENCE FRANÇAISE DE SÉCURITÉ SANITAIRE ENVIRONNEMENTALE (AFSSE). Impact sanitaire de la pollution atmosphérique urbaine. Rapport interne (2004).
- CA. POPE, RT. BURNETT, MJ. THUN et al. Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. *JAMA*, Vol. 287, n° 7, 2002.
- D. COLETT. Modelling survival data in medical research. Chapman & Hall, 1999.
- B. OSTRO. Outdoor air pollution. Assessing the environmental burden of disease at national and local levels. *WHO. Protection of the human environment. Geneve 2004.*
- R. BURNETT, MR JERNETT et al. The spatial association between community air pollution and mortality : a new method of analyzing correlated geographic cohort data. *Environ. Health Perspect.* 109 (suppl. 3) 2001.
- N. KÜNZLI. The public health relevance of air pollution abatement. *Eur. Respir. J.* 2002.
- D. BARD, R. BAROUKI, S. BENHAMOU, J. BENICHOU, J. CLAVEL, E. JOUGLA, G. LAUNOY. Cancer : approche méthodologique du lien avec l'environnement. *Ouvrage collectif INSERM, 2005.*
- S. CASSADOU, C. DECLERCQ, D. EILSTEIN et al. APHEIS Évaluation de l'impact sanitaire de la pollution atmosphérique dans 26 villes européennes. Rapport IVS (2003).
- A. HARIRECHE : Approche géographique de la relation entre pollution atmosphérique et santé. *Colloque démographie et santé, Bordeaux 21-25 mai 2007.*
- DE. ABBEY, N. NISHINO, WF. Mc DONNELL et al. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *Am. J. Respir. Crit. Care*, 1999.
- J. BENICHOU. A review of adjusted estimators of attributable risk. *Stat Meth. Med. Res.* 2001.
- S. MEDINA, E. BOLDO, M. SAKLAD, et al. APHEIS health impact assessment of long-term exposure to $PM_{2.5}$ in 23 european cities. *Eur. J. Epidemiol.* 2006. Et <http://www.apheis.net/>