

Apport de la méthode cas-croisé à l'analyse des effets sanitaires à court terme de la pollution atmosphérique

Réanalyse de données du programme Erpurs

Benoît Chardon (ORS), Sabine Host (ORS), Gaëlle Pedrono (Sépia-Santé), Isabelle Grémy (ORS)

Les effets sur la santé de la pollution atmosphérique ont été observés pour la première fois lors d'épisodes aigus de pollution dans les années 1930 et 1950. Par la suite, ont été développés des outils épidémiologiques permettant de mettre en évidence ces effets pour de faibles concentrations de polluant. Il s'agit des analyses écologiques temporelles rétrospectives fondées aujourd'hui sur des modèles performants, permettant notamment de prendre en compte les tendances à long et moyen termes. Le principe de ces analyses est de relier les variations temporelles à court terme (d'un jour à l'autre) d'un indicateur de l'état de santé d'une population à celles des indicateurs d'exposition de cette population à la pollution atmosphérique.

Le programme Erpurs (Evaluation des risques de la pollution urbaine sur la santé) mis en place à l'Observatoire régional de santé d'Île-de-France en 1990 a pour objectif de quantifier les liens existant à court terme entre les niveaux de pollution atmosphérique couramment rencontrés dans la région et l'état sanitaire de la population. Jusqu'à présent cette quantification était effectuée au moyen d'analyses de séries temporelles rétrospectives.

Récemment, une méthode alternative a été appliquée à l'étude des effets à court terme de la pollution atmosphérique, il s'agit de la méthode cas-croisé. Ces dernières années, cette méthode a été de plus en plus utilisée (Lu & Zeger, 2006).

L'objet de ce document est de décrire cette méthode et de faire une revue bibliographique des études y faisant appel. Les différences entre la méthode cas-croisé et les analyses de séries temporelles appliquées dans le champ de la pollution atmosphérique seront également discutées. Enfin, les données utilisées dans le cadre d'Erpurs seront réanalysées en utilisant la méthode cas-croisé afin de comparer les résultats obtenus avec les analyses de séries temporelles.

La méthode cas-croisé (case crossover)

Cette méthode a été introduite par Maclure en 1991 pour l'analyse des déterminants d'un infarctus du myocarde (Maclure, 1991). Elle permet en effet d'analyser les effets d'une exposition ponctuelle sur le risque de survenue d'un événement aigu.

Cette méthode est basée sur le modèle d'une étude cas-témoin où chaque cas est son propre témoin. Les sujets présentant la pathologie d'intérêt sont inclus dans l'étude. L'exposition d'un sujet dans un délai proche de l'événement sanitaire (période cas) est comparée avec celle d'une période pendant laquelle ce même sujet ne présentait pas l'événement (période témoin).

Depuis ces dernières années, cette méthode est de plus en plus utilisée dans le domaine de la pollution atmosphérique.

Les données nécessaires à la réalisation de cette étude ont été fournies par :

la Délégation à l'information médicale et à l'épidémiologie de l'Assistance publique des hôpitaux de Paris ; SOS-médecins Paris ; Météo-France ; le Groupe régional d'observation de la grippe (Grog, Open Rome) ; le Réseau national de surveillance aérobiologique (RNSA) ; le réseau de surveillance de la qualité de l'air en Île-de-France (Airparif). Que tous soient ici vivement remerciés pour leur concours.

Méthodologie cas-croisé appliquée au champ de la pollution atmosphérique

L'application de cette méthode au champ de la pollution atmosphérique exige de prendre en compte ses spécificités. En effet, il faudra tenir compte des variations temporelles des niveaux de pollution qui sont à la fois quotidiennes et saisonnières. Par ailleurs, le choix des périodes témoins sera influencé par l'interdépendance des niveaux de pollution d'un jour à l'autre. En effet, les niveaux de pollution d'un jour donné sont corrélés aux niveaux des jours précédents (jusqu'à 7 jours) (Levy et al., 2001).

Principe

Dans une étude de type cas-croisé chaque cas est son propre témoin. Tous les sujets inclus dans l'analyse ont connu l'événement sanitaire. L'exposition d'un sujet au moment de l'événement sanitaire, soit la période cas, est comparée à celle de ce même sujet à un moment différent, soit la période témoin, où *a priori* le sujet ne présentait pas l'événement sanitaire. Cette dernière hypothèse est difficilement vérifiable dans la mesure où il est rare de pouvoir relier entre eux les événements sanitaires d'un même sujet.

L'analyse cas-croisé repose sur l'hypothèse selon laquelle les niveaux d'exposition mesurés à la date de l'événement sanitaire sont supérieurs à ceux enregistrés aux dates témoins. Par conséquent, seuls les sujets pour lesquels une variation de l'exposition entre la période cas et la période témoin a été observée sont informatifs.

Par ailleurs, comme dans les analyses de séries temporelles, un délai entre l'exposition et l'éventuel effet sanitaire peut être pris en compte : ainsi peuvent être considérés le niveau d'exposition du jour de l'événement sanitaire (ou du jour témoin), le niveau d'exposition cumulée du jour et des jours précédents, le niveau d'exposition décalé dans le temps par rapport à l'apparition de l'événement sanitaire (Jaakkola, 2003).

Prise en compte des facteurs de confusion

Chaque sujet étant son propre témoin, les études

cas-croisé permettent d'éviter les biais liés aux caractéristiques individuelles des sujets. Il est également possible, dans ce type d'analyse, de contrôler les facteurs de confusion connus qui interviennent dans la mesure de l'association entre l'exposition à la pollution atmosphérique et l'événement sanitaire : les facteurs météorologiques, les épidémies de grippe, les pollens, les jours fériés ainsi que les vacances scolaires.

Choix des périodes témoins

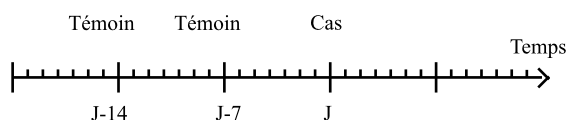
C'est le choix des périodes témoins qui conditionne la prise en compte des variations temporelles à court et long terme de l'indicateur d'exposition et de l'indicateur sanitaire. La sélection des périodes témoins est donc cruciale. Des périodes témoins trop distantes de l'événement sanitaire ne permettent pas de prendre en compte la saisonnalité de l'indicateur d'exposition et de l'indicateur sanitaire, des périodes témoins trop proches de l'événement sanitaire peuvent être à l'origine d'un biais dû à l'existence de fortes autocorrélations dans les données (Fung et al., 2003). Plusieurs stratégies d'appariement ont déjà été utilisées dans les études cas-croisé : l'approche unidirectionnelle, l'approche bidirectionnelle, et celle du *time stratified design* (Figure 1).

Approche unidirectionnelle

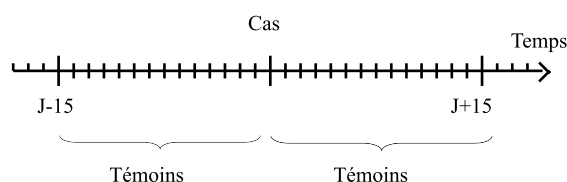
C'est la méthode d'appariement la plus ancienne (Maclure, 1991). Il s'agit de prendre comme jour témoin un jour antérieur à l'événement sanitaire et correspondant au même jour de la semaine que le jour cas. Si la période témoin est assez proche de la période cas (j-7 par exemple) cette approche permet de contrôler la variabilité temporelle à court terme (saisonnalité et jour de la semaine). En revanche, comme la période cas suit toujours la période témoin, l'approche unidirectionnelle ne permet pas de prendre en compte l'existence d'une éventuelle tendance à long terme dans les données, ce qui est un inconvénient majeur pour l'étude des effets sanitaires de la pollution atmosphérique.

Figure 1 : Quelques exemples de sélection des périodes témoins

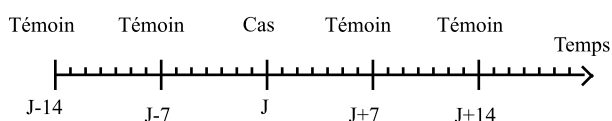
a. Approche unidirectionnelle : exemple où les jours témoins sont choisis 7 et 14 jours avant le jour cas



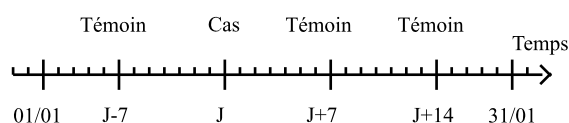
b. Approche bidirectionnelle *full stratum* : exemple où les jours témoins correspondent à l'ensemble des jours situés 15 jours avant et après le jour cas



c. Approche bidirectionnelle symétrique : exemple où les jours témoins sont sélectionnés 7 et 14 jours avant et après le jour cas



d. *Time stratified design* : exemple sur une période d'un mois



Approche bidirectionnelle

Pour contrôler la tendance à long terme, Navidi (1998) propose de choisir une période témoin avant et après l'événement sanitaire : c'est l'approche bidirectionnelle. Il faut alors vérifier que les expositions futures ne sont pas influencées par l'événement sanitaire, ce qui est généralement le cas dans le domaine de la pollution atmosphérique.

En pratique, l'approche bidirectionnelle n'est valide que si les cas sont toujours exposés au risque après l'événement sanitaire ce qui pose problème lorsque l'événement sanitaire étudié est le décès. Cependant, même dans ce cas, Lumley et Levy (1999) ont montré, qu'en présence d'une tendance à long terme dans les données, le biais associé au choix d'une période témoin après l'événement sanitaire était moins important que celui associé à l'utilisation d'une approche unidirectionnelle.

Il existe plusieurs techniques d'appariement possibles avec l'approche bidirectionnelle. Navidi (1998) suggère d'utiliser comme période témoin l'ensemble des jours avant et après le jour cas (*full stratum bidirectional design*), cependant cette technique ne permet pas de contrôler les variations saisonnières (Janes et al., 2004). Bateson et Schwartz (1999) proposent d'utiliser des périodes témoins proches et équidistantes de la période cas (symétrique bidirectionnelle). Cette technique, fréquemment utilisée, permet de prendre en compte

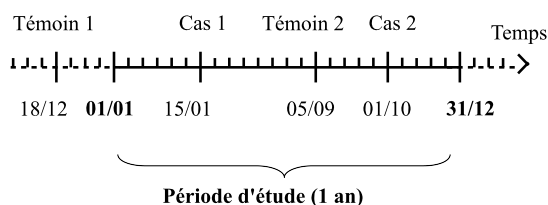
les variations saisonnières uniquement dans le cas où celles-ci sont symétriques (Lee et al., 2000). Enfin, Navidi et Weinhand (2002) proposent pour chaque cas de choisir au hasard une période témoin avant ou après (semi-symétrique bidirectionnelle).

Approche time stratified sur un mois

Avec cette approche, les périodes témoins correspondent au même jour de la semaine que le jour cas, pour le même mois et la même année (Janes et al., 2005, Levy et al., 2001). Il y a donc 3 à 4 témoins pour un cas. Cette approche offre une solution adéquate pour le contrôle des variations temporelles à court et long termes.

Le choix des périodes témoins joue sur la puissance de l'analyse : elle augmente avec le nombre de jours témoins (Levy et al., 2001), toutefois, plus la période témoin est éloignée de la période cas, plus la perte d'information est importante (Lee et al., 2000).

Figure 2 : Perte d'information liée à l'éloignement de la période témoin : cas d'une approche unidirectionnelle avec 4 semaines de décalage entre le jour cas et le jour témoin



Par exemple, sur une période d'analyse d'un an, dans le cas d'une approche unidirectionnelle, si la période témoin se situe quatre semaines avant la période cas, on observe une perte d'information de 7,7% (l'équivalent de quatre semaines sur un an) car les données ne pourront être analysées qu'à partir du deuxième mois de la période (Figure 2). Le choix de l'approche time stratified évite cette perte d'information.

Modélisation et estimation du risque sanitaire

Par analogie avec les études cas-témoins, le modèle statistique utilisé est une régression logistique conditionnelle. Le risque de déclarer l'événement sanitaire est estimé par l'odds-ratio qui est ici identique à un risque relatif (cf. encadré ci-dessous). Des analyses stratifiées peuvent être employées dans le but de déterminer les facteurs modificateurs de l'effet ou encore les sous-populations à risque.

La modélisation en pratique : utilisation du logiciel SAS® (cf. Annexe pour l'application)

En pratique, dans les logiciels statistiques comme SAS®, la régression logistique conditionnelle n'est pas directement implémentée. Cependant, la vraisemblance conditionnelle d'un modèle de régression logistique conditionnelle peut s'écrire comme la vraisemblance d'un modèle à risques proportionnels de Cox discret stratifié. Ceci explique l'utilisation de la procédure PHREG de SAS® par la majorité des auteurs d'articles sur les études cas-croisé. Dans le modèle de Cox classique, les effets des covariables sont exprimés en risques relatifs de subir l'événement. Dans le cas des études cas-croisés, le risque relatif (RR) du modèle de Cox devient un estimateur sans biais de l'odds-ratio (OR) classiquement obtenu dans les régressions logistiques. C'est pourquoi les résultats reportés dans la littérature sont tantôt exprimés en OR tantôt en RR. Dans les séries temporelles, les effets s'interprètent en termes de RR, il est donc préférable de parler également de RR pour le modèle cas-croisé quand on souhaite comparer les deux méthodes.

Revue des études appliquant la méthode cas-croisé pour l'analyse des effets à court terme de la pollution atmosphérique.

Si les études de type cas-croisé sont le plus souvent utilisées avec des données recueillies à l'échelle individuelle, il est possible également d'utiliser cette approche dans le cas de données agrégées. Ainsi, les premières études dans le champ de la pollution atmosphérique utilisant cette méthode ont été publiées en 1999 avec la réanalyse d'une précédente étude utilisant la méthode des séries temporelles sur des données agrégées (Lee & Schwartz, 1999). Cette étude a mis en évidence une association significative entre les niveaux de SO₂ et la mortalité, soit un excès de risque relatif de 5 % (IC 95 % [1,6-8,4]) pour une augmentation de 50 ppb du niveau de SO₂, ces résultats étant concordants avec ceux issus de la première analyse. Cependant, les auteurs démontrent de manière empirique que le choix de la période témoin influence grandement l'estimation du risque relatif. Depuis, de nombreuses études font appel à cette méthode, ainsi une cinquantaine d'articles scientifiques ont été

recensés dans la littérature. Une revue systématique a été réalisée afin de décrire le contexte d'utilisation de cette méthode ainsi que les recommandations des auteurs.

Indicateurs sanitaires et indicateurs de pollution

En terme d'indicateurs sanitaires, les études relevées s'intéressent aussi bien à des indicateurs de mortalité, toutes causes ou causes spécifiques, que de morbidité. Les indicateurs de morbidité sont généralement les admissions hospitalières pour causes cardio-vasculaires ou bien respiratoires, avec des pathologies spécifiques telles qu'insuffisances cardiaques (Symons et al., 2006), arrêts cardiaques (Zanobetti & Schwartz, 2005) ou encore crises d'asthme (Barnett et al., 2005). Les indicateurs de pollution atmosphérique utilisés dans ces études sont ceux classiquement utilisés dans les études de séries temporelles tels que particules (PM_{2,5} et PM₁₀), SO₂, NO₂, CO et O₃.

Le nombre de cas étudiés est variable d'une étude à l'autre, allant de quelques centaines à plusieurs centaines de milliers d'individus (Peters et al., 2001 ; Sunyer et al., 2002 ; Wellenius et al., 2005 ; Neas et al., 1999 ; Schwartz, 2004). Il en est de même pour les périodes d'études, allant d'un an (Boutin-Forzano et al., 2004) à 15 ans (Zanobetti & Schwartz, 2005).

Choix de la période « témoin »

La plupart des études appliquent l'approche bidirectionnelle. Lorsque les auteurs ont mis en œuvre une approche unidirectionnelle, qu'elle soit prospective ou rétrospective, ils soulignent un manque de précision des résultats, avec une surestimation ou bien une sous-estimation du risque relatif (Kan & Chen, 2003 ; Lee & Schwartz, 1999). L'approche bidirectionnelle se décline selon plusieurs méthodes, avec des périodes témoins symétriques de part et d'autre de la période cas (Bateson & Schwartz, 2004), ou bien choisies de manière aléatoire ou encore *time stratified*.

Le nombre de jours témoins est variable, entre 2 et 18 (Bateson & Schwartz, 2004). Ces jours témoins sont choisis entre une semaine et trois semaines avant et après le jour cas. Certains auteurs reproduisent les analyses avec différentes méthodes. Kan et Chen (2003) trouvent ainsi une association plus forte en augmentant le nombre de périodes témoins. Cela confirme que le choix de la période témoin est susceptible d'introduire une variabilité dans les résultats. Cependant, des recommandations peuvent être fournies. Par exemple, Lee et Schwartz (1999) mentionnent que la période témoin prise au-delà de deux semaines n'assure plus le contrôle de la saisonnalité. Schwartz (2004) précise également que la présence d'une auto-corrélation dans les données implique de ne pas choisir la période témoin trop proche de la période cas.

La méthode proposée par Levy (*time stratified*) (2001) permet de contrôler ces différents biais, de nombreux auteurs l'ont ainsi adoptée dans des récentes publications (Sunyer et al., 2002 ; Rich et al., 2005 ; Schwartz, 2004 ; Ruidavets et al., 2005 ; Wellenius et al., 2005).

Modélisation et ajustement sur les facteurs de confusion

Les résultats sont estimés par régression logistique conditionnelle avec ajustement sur les co-facteurs, notamment les facteurs météorologiques (température, humidité, pression atmosphérique), les pollens, les épidémies de grippe, les jours fériés et les vacances scolaires. De plus, dans le cas où ce n'est pas déjà pris en compte dans le choix de la période témoin, un ajustement sur le jour de la semaine est appliqué (Bateson & Schwartz, 2004).

Présentation des résultats

Les résultats sont exprimés à l'aide d'odds-ratio ou de risques relatifs et pour des différentiels d'augmentation des niveaux des indicateurs de pollution variables, allant de 5 $\mu\text{g}/\text{m}^3$ (Ruidavets et al., 2005) à 100 $\mu\text{g}/\text{m}^3$ (Neas et al., 1999), l'augmentation de 10 $\mu\text{g}/\text{m}^3$ étant la plus répandue (Boutin-Forzano et al., 2004 ; Zanobetti & Schwartz, 2005 ; D'Ippoliti et al., 2003 ; Kan & Chen, 2003). Ce différentiel est également parfois exprimé en écart interquartile (Sunyer et al., 2002 ; Luginaah et al., 2005 ; Wellenius et al., 2005 ; Rich et al., 2005 ; Tsai et al., 2003).

Utilisation de données individuelles

Cette méthode permet de tester également des interactions possibles avec certains facteurs individuels tels que les antécédents médicaux ou les caractéristiques socio-économiques. Ainsi, de nombreuses études ont tiré avantage de cette possibilité (Sullivan et al., 2003 ; Ruidavets et al., 2005 ; Zanobetti & Schwartz, 2005 ; Wellenius et al., 2005 ; Bateson & Schwartz, 2004). A titre d'exemple, D'Ippoliti et al. (2003) montrent que l'effet de la pollution atmosphérique sur les admissions à l'hôpital pour crise cardiaque est accru chez les patients atteints d'affections cardiaques telles que des troubles de la conduction. Des études ont pu aussi cibler des sous-populations à risque. A ce titre, Sunyer et al. (2002) étudient le risque de décès chez les individus précédemment hospitalisés pour asthme. Symons et al. (2006) étudient le risque d'admission hospitalière chez des sujets atteints d'insuffisance cardiaque congestive.

Cas-croisé et séries temporelles

Les analyses de séries temporelles sont couramment utilisées pour l'estimation des effets sanitaires de la pollution atmosphérique (Katsouyanni et al., 1995 ; Daniels et al., 2004 ; Filleul et al., 2006). Le principe de ces analyses est de relier les variations d'un jour à l'autre des niveaux de pollution aux variations d'un jour à l'autre d'un indicateur de l'état de santé d'une population. L'unité d'observation est donc le jour. Le compte journalier d'événements sanitaires est supposé suivre une loi de Poisson. Au cours de la dernière décennie, des modèles additifs généralisés (GAM) (Hastie & Tibshirani, 1990) ont régulièrement été utilisés.

Les études de type cas-croisé représentent une alternative récente aux analyses de séries temporelles. Les spécificités de ces deux méthodes sont détaillées dans le tableau suivant :

	Séries temporelles	Cas-croisé
Mise en œuvre de la méthode	Modèles robustes, avec prise en compte de la sur-dispersion ⁽¹⁾ , des facteurs de confusion potentiels (grippe, facteurs météorologiques, saisonnalité, pollens) et des éventuels effets non-linéaires de ces variables par l'utilisation de fonctions de lissage (Wood, 2003), mais mise en œuvre complexe. Résultats influencés par le choix des paramètres de lissage pour le contrôle de la saisonnalité.	Facilité de mise en œuvre avec contrôle de la saisonnalité via le protocole (choix des périodes témoins), mais non prise en compte de la sur-dispersion des données (Lu & Zeger, 2006). Résultats influencés par le choix des périodes témoins et difficulté de vérifier l'hypothèse selon laquelle un sujet ne présente pas l'événement sanitaire lors d'un jour témoin.
Données agrégées vs données individuelles	Recours à des données agrégées : données aisément disponibles par le biais de systèmes d'information existants. Facteurs de confusion individuels non pris en compte directement dans les analyses. Ils sont supposés constants sur une courte période de temps.	Possibilité d'inclure dans l'analyse les caractéristiques personnelles des sujets telles que les antécédents de santé, les facteurs de risque... (Jaakkola, 2003). Contrôle des facteurs de confusion individuels, chaque cas étant son propre témoin.
Puissance (effectifs et période d'étude)	Pour une bonne précision des résultats, nécessité d'avoir un nombre de cas journaliers le plus élevé possible (moyenne journalière supérieure à 5) et une période d'étude au moins égale à 4 ans permettant de bien prendre en compte la saisonnalité.	Pas de contrainte de durée de la période d'étude (un an peut suffire), avec un effectif de quelques centaines d'individus.

(1) Une variable aléatoire qui suit une loi de Poisson a notamment pour propriété d'avoir l'espérance égale à la variance. En pratique, il peut arriver que la variance soit supérieure à l'espérance, on parle alors de sur-dispersion.

Réanalyse de données du programme Erpurs en appliquant la méthode cas-croisé

Les dernières études du programme Erpurs ont permis d'estimer le lien entre, d'une part, les niveaux de pollution atmosphérique couramment rencontrés à Paris et sa proche couronne et, d'autre part, le nombre d'appels reçus par SOS-médecins (Chardon et al., 2005), et le nombre journalier d'hospitalisations entre 2000 et 2003 (Granados et al., 2006). Pour mettre en évidence ces associations,

des analyses de séries temporelles avaient été utilisées. Cette partie a pour objectif de réanalyser ces données recueillies à l'échelle de la population en utilisant une analyse de type cas-croisé et de comparer les résultats obtenus.

Zone d'étude

Pour des raisons d'homogénéité des niveaux de

pollution, la zone d'étude concerne Paris et les trois départements de la proche couronne : Seine-Saint-Denis, Hauts-de-Seine, et Val-de-Marne. Dans cette zone comptabilisant 56 % des habitants de la région, le trafic routier est le principal émetteur d'oxydes d'azote et de particules (DRIRE, 2006). Cette zone d'étude représente donc le cœur dense et urbanisé de l'Île-de-France.

Données recueillies

Indicateurs sanitaires

Quatre indicateurs sanitaires ont été construits : il s'agit du nombre journalier d'hospitalisations pour causes respiratoires des 15 ans et plus, et du nombre journalier d'appels reçus par SOS-médecins pour des affections des voies respiratoires supérieures (VRS), inférieures (VRI) et pour asthme. Les données correspondantes ont été fournies par l'association SOS-médecins Paris et par la délégation à l'information médicale et à l'épidémiologie de l'Assistance publique des hôpitaux de Paris (AP-HP). Le détail de la construction de ces indicateurs sanitaires est précisé dans les rapports correspondants (Chardon et al., 2005 ; Granados et al., 2006).

Indicateurs d'exposition

Le dioxyde d'azote (NO₂), et les particules en suspension dont le diamètre est inférieur à 10 micromètres (PM10) ont été pris en compte en tant qu'indicateurs d'exposition à la pollution atmosphérique. Une analyse descriptive de ces polluants entre 2000 et 2003 a permis de montrer l'homogénéité de leurs niveaux de fond à l'intérieur de la zone d'étude (Chardon et al., 2005). Les niveaux moyens du jour de l'événement sanitaire et des trois jours précédents ont été pris en compte comme indicateurs d'exposition.

Cofacteurs pris en compte

Il s'agit des cofacteurs biaisant l'estimation de la relation pouvant exister entre les niveaux de pollution atmosphérique et l'indicateur sanitaire. Les principaux cofacteurs sont les conditions météorologiques (températures, humidité) mesurées

quotidiennement par Météo-France, les épidémies de grippe définies par le Groupe régional d'observation de la grippe (Grog) et les périodes de présence des principaux pollens allergisants recensées par le Réseau national de surveillance aérobiologique (RNSA). Afin de rendre compte des variations temporelles des indicateurs sanitaires et des indicateurs d'exposition, la tendance à long et moyen termes des indicateurs (fonction de lissage avec 5 degrés de liberté par an) ainsi que les jours fériés, les jours de la semaine et les vacances scolaires ont été utilisés. Concernant la méthode cas-croisé, les variations temporelles sont prises en compte par le choix des jours témoins. En moyenne, les témoins se situent à environ 13 jours du cas.

La méthode statistique

Une analyse de type cas-croisé a été menée. La technique d'appariement *time stratified*, qui semble la plus adaptée à ce jour pour la modélisation des effets sanitaires de la pollution atmosphérique, a été utilisée. Afin de pouvoir comparer les résultats avec ceux obtenus dans le cadre du programme Erpurs au moyen d'analyses de séries temporelles, les mêmes variables d'ajustement ont été introduites dans le modèle. La température et l'humidité ont été introduites dans le modèle sous la forme de *natural splines* à 3 degrés de liberté. Les résultats sont présentés sous forme de pourcentage d'excès de risque relatif. L'analyse de séries temporelles a été réalisée à l'aide du logiciel R® version 1.8, l'analyse de type cas-croisé à l'aide du logiciel SAS® version 8.

Résultats

Les résultats obtenus dans les précédentes études Erpurs au moyen d'analyses de séries temporelles montraient :

- l'existence d'un lien significatif entre les niveaux de pollution atmosphérique (NO₂ et PM10) et le risque d'hospitalisation pour causes respiratoires des personnes âgées de 15 ans et plus ;
- l'existence de liens significatifs entre les niveaux de pollution particulaire (PM10) et le nombre de visites médicales à domicile effectuées par

Figure 3 : Excès de risque relatif (%) et intervalle de confiance à 95 % d'admissions hospitalières pour causes respiratoires de personnes âgées de 15 ans et plus, pour une augmentation de 10µg/m³ des niveaux de pollution (2000-2003)

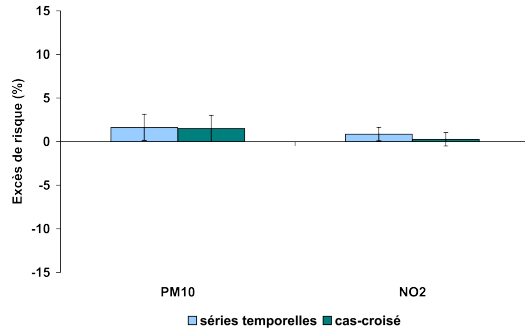


Figure 4 : Excès de risque relatif (%) et intervalle de confiance à 95 % d'appels reçus par SOS-médecins pour des affections des voies respiratoires supérieures, pour une augmentation de 10µg/m³ des niveaux de pollution (2000-2003)

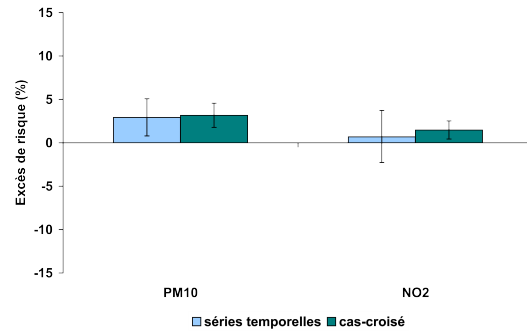


Figure 5 : Excès de risque relatif (%) et intervalle de confiance à 95 % d'appels reçus par SOS-médecins pour des affections des voies respiratoires inférieures, pour une augmentation de 10µg/m³ des niveaux de pollution (2000-2003)

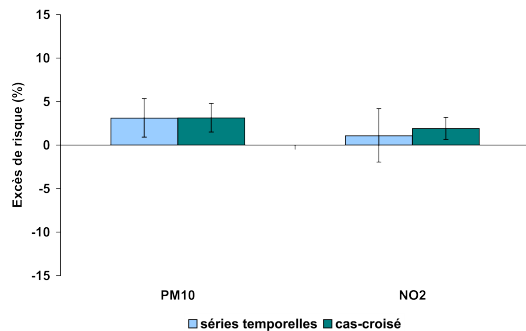
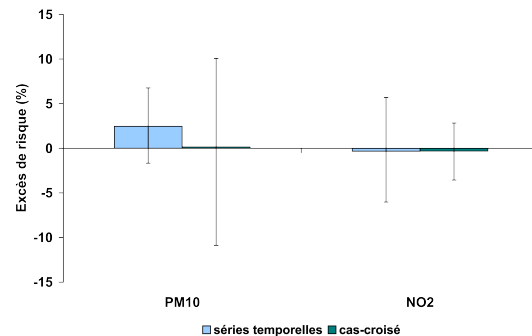


Figure 6 : Excès de risque relatif (%) et intervalle de confiance à 95 % d'appels reçus par SOS-médecins pour asthme, pour une augmentation de 10µg/m³ des niveaux de pollution (2000-2003)



SOS-médecins pour des affections des voies respiratoires supérieures ou inférieures ;

- aucun lien significatif entre les niveaux de NO₂ et le nombre d'appels journaliers reçus par SOS-médecins pour les pathologies étudiées ;

- aucun lien significatif entre les appels ayant pour motif l'asthme, et la pollution atmosphérique.

Les résultats obtenus ici au moyen d'une étude de type cas-croisé sont similaires (Figures 3 à 6) concernant la pollution particulaire. En revanche, des différences sont observées au sujet du NO₂ : contrairement aux résultats des études précédentes, il n'a pas été observé de liens significatifs entre cet indicateur d'exposition et le nombre d'hospitalisations pour causes respiratoires chez les 15 ans et plus. De même une association

significative a été observée entre les niveaux de NO₂ et le nombre d'appels reçus par SOS-médecins pour des affections des voies respiratoires supérieures et inférieures ce qui est en contradiction avec les résultats des études précédentes.

Discussion

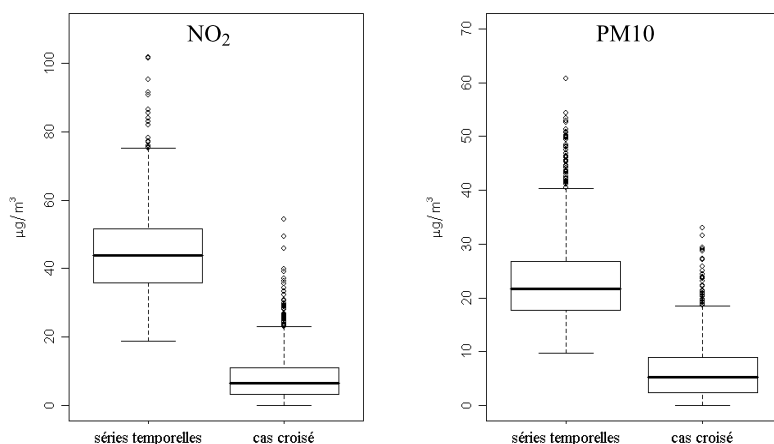
La non prise en compte de la sur-dispersion des données dans l'analyse cas-croisé pour estimer l'écart-type peut expliquer les différences observées entre les deux méthodes d'analyse au niveau des intervalles de confiance.

Une autre explication peut venir de la distribution de l'exposition qui n'est pas la même suivant que l'on soit dans une analyse de séries temporelles ou dans une analyse cas-croisé. En effet, dans le

premier cas l'exposition d'intérêt correspond aux niveaux journaliers du polluant alors que dans le second cas elle correspond à la différence entre le niveau du polluant le jour cas et celui du jour témoin (Kunzli & Schindler, 2005). Une faible variabilité

de l'exposition peut limiter la puissance de l'analyse à détecter son effet sur la santé. La figure ci-dessous montre une variabilité plus importante pour les niveaux journaliers du polluant utilisés dans l'analyse de séries temporelles.

Figure 7 : Distributions comparées des niveaux d'exposition au NO₂ et PM10 (en µg/m³) dans l'analyse de séries temporelles (a) et dans l'analyse cas-croisé (b) pour la période 2000-2003



(a) distributions des niveaux journaliers du polluant (niveaux moyens du jour et des 3 jours précédant l'événement sanitaire)

(b) distributions de la valeur absolue des différences entre les niveaux journaliers du polluant du jour cas (niveaux moyens du jour et des 3 jours précédents) et la moyenne des concentrations des jours témoins (niveaux moyens du jour et des 3 jours précédents)

Conclusion

Ces dernières années des études de type cas-croisé plus faciles d'utilisation ont régulièrement été menées pour déterminer les effets sanitaires de la pollution atmosphérique. Dans ce type d'analyse chaque cas est son propre témoin ce qui permet de s'affranchir des problèmes liés au contrôle des facteurs individuels à la différence des analyses de séries temporelles qui ne prennent pas en compte directement ces facteurs de confusion.

Dans une étude cas-croisé, la sélection des jours témoins est très importante : c'est elle qui va permettre d'ajuster sur les variations à court et moyen termes des indicateurs sanitaires et d'exposition. A ce jour, la technique d'appariement *time stratified* semble la plus adaptée pour la modélisation des effets sanitaires de la pollution atmosphérique.

Les résultats obtenus dans la littérature et dans le cadre de ce travail à partir de ce type d'analyse confirment l'existence de liens significatifs entre les niveaux de pollution atmosphérique couramment rencontrés en agglomération urbaine et divers indicateurs de mortalité ou de morbidité. Les différences de résultats observées dans ce travail suivant la méthode d'analyse employée (séries temporelles ou cas-croisé) peuvent s'expliquer par le fait que la sur-dispersion des données sanitaires (fortement présente notamment concernant les données de SOS-médecins) n'est pas prise en compte dans le cadre des études cas-croisé et par une variabilité de l'exposition moins importante dans le cas de l'analyse cas-croisé. C'est pourquoi, l'utilisation des analyses de séries temporelles est conseillée dans le cadre de données agrégées si les comptes journaliers sont suffisamment importants ; cependant, la mise en œuvre des modèles nécessite une bonne expérience de la méthode. En revanche, lorsque des données à l'échelle individuelle sont disponibles, l'utilisation d'une étude de type cas-croisé peut permettre la réalisation d'analyses plus fines ou stratifiées.

Annexe : Utilisation de la procédure PHREG du logiciel SAS®

I. La procédure PHREG

L'analyse des données recueillies dans le cadre d'une étude de type cas-croisé peut être effectuée sous le logiciel SAS® à l'aide de la procédure PHREG (proportional hazard regression).

```
PROC PHREG DATA=<nom du fichier de données> ;
```

```
MODEL <écriture du modèle> ;
```

```
<Instructions supplémentaires optionnelles>
```

```
RUN ;
```

L'utilisation de cette procédure nécessite que le fichier de données soit structuré de manière adéquate. Lorsque pour un « cas » des valeurs relatives aux variables explicatives sont manquantes et qu'il est impossible de les remplacer, il est conseillé de supprimer le « cas » et ses « témoins » correspondants.

II. Présentation du fichier de données (exemple)

Le fichier de données peut se présenter de deux manières différentes suivant que l'on possède des informations à l'échelle individuelle ou agrégée.

1. Cas de données individuelles

Dans le cas de données individuelles, chaque sujet est caractérisé par plusieurs lignes : une ligne correspondant au jour cas (le jour où le sujet a subi l'événement) et une ou plusieurs lignes (suivant la stratégie d'appariement choisie) correspondant aux différents jours témoins (jours pour lesquels le sujet n'a pas subi l'événement). Pour chaque ligne du fichier, plusieurs variables doivent être définies :

- une variable définissant le jour cas (variable « cas »), elle correspond à la variable de censure dans le modèle de Cox. Les cas ont tous connu l'événement et les témoins sont tous censurés ;
- une variable permettant d'identifier le groupe composé du cas et de ses témoins appariés pour chacun des sujets (variable « strate ») ; il y a donc autant de strates que de sujets ;
- une variable introduisant de manière fictive la notion de temps discret nécessaire à la mise en œuvre du modèle de Cox discret (variable « temps »). Ainsi les cas subissent l'événement au temps 1 et les témoins sont censurés au temps 2 ;
- une ou plusieurs variables indépendantes (qualitatives ou quantitatives).

Tableau 1 : Première présentation possible des données (témoins choisis selon la méthode *time stratified*)

sujet	date	strate	cas	temps	férié	vacances	NO ₂
sujet 1	01/01/2000	1	1	1	1	1	39
	08/01/2000	1	0	2	0	0	50
	15/01/2000	1	0	2	0	0	55
	22/01/2000	1	0	2	0	0	37
	29/01/2000	1	0	2	0	0	46
sujet 2	01/01/2000	2	1	1	1	1	40
	08/01/2000	2	0	2	0	0	53
	15/01/2000	2	0	2	0	0	58
	22/01/2000	2	0	2	0	0	32
	29/01/2000	2	0	2	0	0	21
...

Remarque : dans l'exemple, on suppose que les concentrations de pollution sont mesurées par différents capteurs car les niveaux de NO₂ du sujet 1 sont différents de ceux du sujet 2 alors qu'ils ont subi l'événement le même jour.

Dans cet exemple, les variables « férié », « vacances », et « NO₂ » sont des variables indépendantes. Pour mesurer l'effet de la variable NO₂ sur l'événement sanitaire en ajustant sur les jours fériés et les vacances scolaires, le programme s'écrira de la manière suivante :

```
PROC PHREG DATA=tableau1 NOSUMMARY;
MODEL temps*cas(0)=vacances férié NO2 / RL;
SRATA strate;
RUN;
```

L'instruction STRATA permet de stratifier le modèle sur chacun des sujets. La stratification permet à chaque sujet d'avoir une fonction de risque de base différente mais estime des effets communs à l'ensemble des sujets pour les variables indépendantes.

2. Cas de données agrégées

Lorsque les informations sont recueillies à l'échelle de la population (pour chaque jour, on possède par exemple une valeur unique d'exposition pour l'ensemble de l'échantillon), il est possible de regrouper les cas qui surviennent le même jour sur une même ligne. L'instruction FREQ dans la procédure PHREG permet de préciser le nombre de cas survenus un jour donné.

Tableau 2 : Deuxième présentation possible des données (témoins choisis selon la méthode *time stratified*)

date	strate	cas	temps	nombre	férié	vacances	NO ₂
01/01/2000	1	1	1	9	1	1	39
08/01/2000	1	0	2	9	0	0	50
15/01/2000	1	0	2	9	0	0	55
22/01/2000	1	0	2	9	0	0	37
29/01/2000	1	0	2	9	0	0	46
02/01/2000	2	1	1	5	1	1	38
09/01/2000	2	0	2	5	0	0	42
16/01/2000	2	0	2	5	0	0	45
23/01/2000	2	0	2	5	0	0	32
30/01/2000	2	0	2	5	0	0	22
...

Dans ce tableau, la variable « nombre » correspond au nombre de cas survenus le "jour cas" (de même pour les jours témoins, la variable « nombre » est le nombre d'événements survenus le "jour cas"). Par exemple, le 01/01/2000, 9 individus ont déclaré l'événement sanitaire. Concrètement, l'instruction FREQ permet de répéter une ligne autant de fois qu'il y a de cas. Par exemple, pour le sujet 1, FREQ permet de répéter 9 fois les cas et 9 fois les témoins, cela revient en quelque sorte à une pondération par le nombre de cas journalier.

Le programme permettant d'estimer l'effet de la variable NO₂ sur l'événement sanitaire en ajustant sur les jours fériés et les vacances scolaires est le suivant :

```
PROC PHREG DATA =tableau2 NOSUMMARY;
MODEL temps*cas(0)=vacances férié NO2 / RL ;
SRATA strate;
FREQ nombre;
RUN;
```

NB : l'option NOSUMMARY évite l'édition des résultats par strate dans les sorties, l'option RL permet d'obtenir les intervalles de confiance à 95% des risques relatifs.

Bibliographie

- Barnett A.G., Williams G.M., Schwartz J., Neller A.H., Best T.L., Petroeschevsky A.L. & Simpson R.W. Air pollution and child respiratory health: a case-crossover study in Australia and New Zealand. *Am J Respir Crit Care Med.* 2005, n°11, pp.1272-8.
- Bateson T.F. & Schwartz J. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiology.* 1999, n°5, pp.539-44.
- Bateson T.F. & Schwartz J. Who is sensitive to the effects of particulate air pollution on mortality? A case-crossover analysis of effect modifiers. *Epidemiology.* 2004, n°2, pp.143-9.
- Boutin-Forzano S., Adel N., Gratecos L., Jullian H., Garnier J.M., Ramadou M. et al. Visits to the emergency room for asthma attacks and short-term variations in air pollution. A case-crossover study. *Respiration.* 2004, n°2, pp.134-7.
- Chardon B., Lefranc A., Granados D. & Grémy I. Evaluation des risques de la pollution urbaine sur la santé : analyse des liens à court terme entre niveaux de pollution atmosphérique et visites médicales à domicile (2000-2003) pour pathologies respiratoires. 2005.
- Daniels, M. J., Dominici, F., Zeger, S. L. & Samet, J.M. The National Morbidity, Mortality, and Air Pollution Study. Part III: PM10 concentration-response curves and thresholds for the 20 largest US cities. *Res Rep Health Eff Inst.* 2004, n°94 Pt 3, pp.1-21; discussion 23-30.
- D'ippoliti D., Forastiere F., Ancona C., Agabiti N., Fusco D., Michelozzi P. et al. Air pollution and myocardial infarction in Rome: a case-crossover analysis. *Epidemiology.* 2003, n°5, pp.528-35.
- DRIRE. Plan de protection de l'atmosphère pour l'Île-de-France. 2006.
- Filleul, L., Cassadou, S., Medina, S., Fabres, P., Lefranc, A., Eilstein, D. et al. The relation between temperature, ozone, and mortality in nine French cities during the heat wave of 2003. *Environ Health Perspect.* 2006, n°9, pp.1344-7.
- Fung K. Y., Krewski D., Chen Y., Burnett R. & Cakmak S. Comparison of time series and case-crossover analyses of air pollution and hospital admission data. *Int J Epidemiol.* 2003, n°6, pp.1064-70.
- Granados D., Chardon B., Lefranc A. & Grémy I. Evaluation des risques de la pollution urbaine sur la santé : exploration de la différence de sensibilité à la pollution atmosphérique chez les hommes et les femmes. 2006.
- Hastie, T. & Tibshirani, R. Chapman and Hall (Ed.), 1990. Generalized additive models.
- Jaakkola J.J.K. Case-crossover design in air pollution epidemiology. *Eur Respir J Suppl.* 2003, pp.81s-85s.
- Janes H., Sheppard L., Lumley T. Referent Selection strategies in case-crossover analyses of air pollution exposure data : implication for bias. 2004.
- Janes H., Sheppard L. & Lumley T. Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias. *Epidemiology.* 2005, n°6, pp.717-26.
- Kan H. & Chen B. A case-crossover analysis of air pollution and daily mortality in Shanghai. *J Occup Health.* 2003, n°2, pp.119-24.
- Katsouyanni, K., Zmirou, D., Spix, C., Sunyer, J., Schouten, J. P., Ponka et al. Short-term effects of air pollution on health: a European approach using epidemiological time-series data. The APHEA project: background, objectives, design. *Eur Respir J.* 1995, n°6, pp.1030-8.
- Künzli N & Schindler C. A call for reporting the relevant exposure term in air pollution case-crossover studies. *J Epidemiol Community Health.* 2005, n°59, pp. 527-30.
- Lee J.T. & Schwartz J. Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: A case-crossover design. *Environ Health Perspect.* 1999, n°8, pp.633-6.
- Lee, J. T., Kim, H. & Schwartz, J. Bidirectional case-crossover studies of air pollution: bias from skewed and incomplete waves. *Environ Health Perspect.* 2000, n°12, pp.1107-11.
- Levy, D., Lumle, T., Sheppard L., Kaufman J. & Checkoway H. Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology.* 2001, n°2, pp.186-92.
- Lu Y. & Zeger S. On the equivalence of case-crossover and time series methods in environmental epidemiology. *Biostatistics.* 2006.
- Luginaah I.N., Fung K.Y., Gorey K.M., Webster G. & Wills C. Association of ambient air pollution with respiratory hospitalization in a government-designated "area of concern": the case of Windsor, Ontario. *Environ Health Perspect.* 2005, n°3, pp.290-6.
- Lumley T. & Levy D. Bias in the case-crossover design: implications for studies of air pollution. 1999.
- Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol.* 1991, n°2, pp.144-53.
- Navidi W. Bidirectional case-crossover designs for exposures with time trends. *Biometrics.* 1998, n°2, pp.596-605.
- Navidi W. & Weinhandl E. Risk set sampling for case-crossover designs. *Epidemiology.* 2002, n°1, pp.100-5.
- Neas L. M., Schwartz J. & Dockery D. A case-crossover analysis of air pollution and mortality in Philadelphia. *Environ Health Perspect.* 1999, n°8, pp.629-31.
- Peters A., Dockery D.W., Muller J.E. & Mittleman M.A. Increased particulate air pollution and the triggering of myocardial infarction. *Circulation.* 2001, n°23, pp.2810-5.
- Rich D. Q., Schwartz J., Mittleman M.A., Link M., Luttmann-Gibson H., Catalano P.J. et al. Association of short-term ambient air pollution concentrations and ventricular arrhythmias. *Am J Epidemiol.* 2005, n°12, pp.1123-32.
- Ruidavets J., Coumot M., Cassadou S., Giroux M., Meybeck M. & Ferrieres J. Ozone air pollution is associated with acute myocardial infarction. *Circulation.* 2005, n°5, pp.563-9.
- Schwartz J. The effects of particulate air pollution on daily deaths: a multi-city case crossover analysis. *Occup Environ Med.* 2004, n°12, pp.956-61.
- Sullivan J., Ishikawa N., Sheppard L., Siscovick D., Checkoway H. & Kaufman J. Exposure to ambient fine particulate matter and primary cardiac arrest among persons with and without clinically recognized heart disease. *Am J Epidemiol.* 2003, n°6, pp.501-9.
- Sunyer J., Basagana X., Belmonte J. & Anto J.M. Effect of nitrogen dioxide and ozone on the risk of dying in patients with severe asthma. *Thorax.* 2002, n°8, pp.687-93.
- Symons J.M., Wang L., Guallar E., Howell E., Dominici F., Schwab M. et al. A case-crossover study of fine particulate matter air pollution and onset of congestive heart failure symptom exacerbation leading to hospitalization. *Am J Epidemiol.* 2006, n°5, pp.421-33.
- Tsai S., Goggins W.B., Chiu H. & Yang C. Evidence for an association between air pollution and daily stroke admissions in Kaohsiung, Taiwan. *Stroke.* 2003, n°11, pp.2612-6.
- Wellenius G. A., Bateson T. F., Mittleman M.A. & Schwartz J. Particulate air pollution and the rate of hospitalization for congestive heart failure among medicare beneficiaries in Pittsburgh, Pennsylvania. *Am J Epidemiol.* 2005, n°11, pp.1030-6.
- Wood, S. Thin-plate regression splines. *J R Stat Soc [Ser B].* 2003, n°, pp.95-114.
- Zanobetti A. & Schwartz J. The effect of particulate air pollution on emergency admissions for myocardial infarction: a multicity case-crossover analysis. *Environ Health Perspect.* 2005, n°8, pp.978-82.