

## Survival analysis to estimate association between short-term mortality and air pollution.

Johanna Lepeule, Virginie Rondeau, Laurent Filleul, Jean-Francois Dartigues

► **To cite this version:**

Johanna Lepeule, Virginie Rondeau, Laurent Filleul, Jean-Francois Dartigues. Survival analysis to estimate association between short-term mortality and air pollution.. Environmental Health Perspectives, National Institute of Environmental Health Sciences, 2006, 114 (2), pp.242-7. inserm-00138548

**HAL Id: inserm-00138548**

**<https://www.hal.inserm.fr/inserm-00138548>**

Submitted on 27 Mar 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Survival Analysis to Estimate Association Between Short-term Mortality and Air Pollution

Lepeule Johanna<sup>1,4</sup>, Rondeau Virginie<sup>1,4</sup>, Filleul Laurent<sup>2</sup>, Dartigues Jean-François<sup>3,4</sup>.

*Affiliation of authors:*

<sup>1</sup> INSERM EMI 0338, Biostatistique, ISPED-Université Victor Segalen Bordeaux2, case11, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.

<sup>2</sup> InVS - Cire Aquitaine

<sup>3</sup> INSERM U593, Université Victor Segalen Bordeaux2, Bordeaux, France.

<sup>4</sup> Institut Fédératif de Recherche en Santé Publique (IFR INSERM 99), Université Victor Segalen Bordeaux2, Bordeaux, France.

*Correspondance Author :*

Johanna Lepeule

UMR 708 Gestion de la santé animale, Ecole Vétérinaire de Nantes / INRA

BP 40706, 44307 Nantes Cedex 03

e-mail : lepeule@vet-nantes.fr

Tel.: (33) 240 687 821; Fax.: (33) 240 687 768

## **Running title**

Survival analysis: mortality and air pollution

## **Key Words**

air pollution, Cox proportional hazards model, distributed lag, mortality, short-term effect

## **Acknowledgements**

This study was granted by Fondation de France, Novartis Pharma, SCOR, Caisse Nationale d'Assurance Maladie, Conseil Général de la Dordogne, Conseil Général de la Gironde, Mutualité Sociale Agricole, Agrica and by a donation from Mrs Bonnement. The authors would like to thank the referees for their helpful comments that led to substantial improvements in the paper.

The authors declare they have no competing financial interests.

## **List of abbreviations**

AIC Akaike's Information Criterion

APPA Association de Prévention de la Pollution Atmosphérique

BS Black Smoke

CI Confidence Interval

ICD-9 International Classification of Diseases, ninth revision

INSERM National Institute of Health and Medical Research

PAQUID Personnes Agées QUID

PM10 Particulate Matter < 10  $\mu$ m aerodynamic diameter

RR Rate Ratio

SAS Statistical Analysis System

SD Standard Deviation

SO<sub>2</sub>-AF Sulfur dioxide

TP-Spline Truncated Power Basis Spline

## **Outline of section headings**

**ABSTRACT**

**INTRODUCTION**

**MATERIALS AND METHODS**

**Study subjects**

**Health data**

**Environmental data**

**Analytical approach**

**Potential modifiers**

**RESULTS**

**Sensitivity**

**DISCUSSION**

**REFERENCES**

**TABLES**

**FIGURE LEGEND**

**FIGURE**

## ABSTRACT

**Background:** To report associations between short-term air pollution and mortality, ecological studies are commonly used. In such studies, the unit of observation is the day, rather than the individual. Moreover, individual data on the subjects are rarely available, which limits the assessment of individual risk factors. These associations can also be investigated using case-crossover studies. However, by definition, individual risk factors are not studied and such studies only analyse dead subjects, which limits the statistical power.

**Objective:** We suggest that the survival analysis is more suitable when cohorts are examined with a time-dependent ecological exposure. To our knowledge, this type of analysis has never been proposed to date.

**Design, participants, measurements:** The present study uses a Cox proportional hazards model to investigate the distribution over time of the short-term effect of black smoke and sulfur dioxide in 439 non-accidental and 158 cardio-respiratory deaths among the 1,469 subjects of the PAQUID cohort in Bordeaux, France. The model has a delayed entry and a polynomial distributed lag from zero to 5 days. Results are adjusted for individual risk factors, temperature, relative humidity, weekday, season, influenza epidemics, and for a time function to control temporal trends.

**Results:** We identify a positive and significant association between cardio-respiratory mortality and black smoke, with a 24% increase in deaths three days after a 10  $\mu\text{g}/\text{m}^3$  increase in black smoke (95% CI : 4-47).

**Conclusions:** We conclude that, the Cox proportional hazards model with time-dependent covariates is very suitable to investigate simultaneously the short-term effect of air pollution on health and the effect of individual risk factors on a cohort study.

## INTRODUCTION

Several epidemiological study designs are used to investigate air pollution and health. They apply different methods for estimating health risks associated with variations in exposure across spatial and temporal gradients. Studies are often classified according to the type of data: individual- or aggregate-level data on exposure, health, and confounding factors. Among study designs assessing the association of short-term variations in pollution and health outcomes, the most widely used are time-series studies and case-crossover studies. Various studies have shown that concentrations of ambient air particles are associated with an increase, the same day and the day after, in all-cause mortality (Bremner et al. 1999; Peters et al. 2000; Rossi et al. 1999), respiratory mortality (Bremner et al. 1999; Rossi et al. 1999) and cardio-vascular mortality (Braga et al. 2000; Bremner et al. 1999). Regression models, such as the generalized additive models with nonparametric splines or the generalized linear models with parametric regression splines, are commonly used in time-series analysis to estimate the increase in risk for a health outcome such as mortality, associated with a unit increase in ambient air pollution levels on a short-term basis (Filleul et al. 2004a). Such models make it possible to include smooth functions of time and temperature to adjust for seasonal variations, long-term trends and temporal changes in factors that might bias the estimation of the health risk. Studies using this approach are named ecological studies because data are aggregated and the daily number of deaths is investigated. The unit of observation is the day rather than the individual. Moreover, individual data on the subjects are rarely available, which limits the assessment of individual risk factors. In the case-crossover design, each subject is his own control and air pollution levels on the dates of death (case period) are compared with those one week before or after death (control period) (Bateson and Schwartz 2001; Navidi and Weinhandl 2002). Consequently, all individual risk factors are inherently controlled and their effect can not be assessed. Individual risk factors have to be

taken into account to investigate air pollution and health because they can explain variations in the susceptibility or resistance among individuals to variations in air pollution concentrations. Moreover, in the case-crossover design only subjects who have died are included in the analysis, thereby involving a loss of power whenever a cohort is available (because information about alive subjects is not included in the analysis). Given that the effects on mortality associated with short-term increases in particulate air pollution are relatively slight, this loss of information can not be neglected.

The purpose of our new approach is to treat simultaneously daily exposure to air pollution and individual risk factors, without aggregating over subjects or time. We propose to use the Cox proportional hazards model (Cox 1972) to analyse the effect of air pollution on the short-term mortality, which has never been proposed to date. This new analysis combined the advantages of the cohort and time-series methods. The key advantage of the cohort approach is its ability to assess and to adjust for individual risk factors of susceptibility such as smoking habits, gender and occupation, which have previously been used only to study the long-term associations between air pollution exposure and health outcomes (Dockery et al. 1993; Pope et al. 1995). The key advantage of the time-series approach is to adjust for seasonal variations, long-term trends and temporal changes for factors such as temperature, humidity and day of the week. In this way, insights may be gained into the exposure-response relationship by allowing for simultaneous examination of the impact of both subject-specific and time-related factors on mortality. Furthermore, the power of the survival analysis is increased compared to the case-crossover approach, simply because all the subjects are studied. We use cohort data with the Cox proportional hazards model, in which exposure to air pollution is considered as a time-dependent covariate. We analyse the distribution over time of the short-term effect of air pollution concentrations on mortality. These data were previously analysed using the case-crossover method with a semi-symmetric bidirectional design (Filleul et al. 2004b) but this

analysis was not completely satisfactory simply because only deaths could be studied, so a large part of the cohort was not analysed. Moreover, with the case-crossover design, we can not identify and assess the effects of the individual risk factors because there are inherently controlled .

A survival method taking into account all the information about the cohort seems more appropriate. In this approach, survival times are not aggregated and the Cox proportional hazards model takes into account for individual factors and quantifies their effects on the association between mortality and air pollution concentration, which is impossible with the case-crossover design.

## **MATERIALS AND METHODS**

### **Study subjects**

All the subjects of the PAQUID (Personnes Agées QUID) cohort living in the urban area of Bordeaux in southwestern France were included. Data on air pollution were only available for this area. This cohort was designed to prospectively study cerebral and functional factors of aging on a representative sample of 3,777 people. Subjects were 65 years of age and older at inclusion and living at home in the administrative areas of Gironde and Dordogne. They were randomly selected from the general electoral lists of the administrative areas after stratification by age, sex and urban unit. An informed consent was obtained from each participant before the study embarked. Trained psychologists interviewed the subjects at home at inclusion in 1988. Interviews made it possible to fill in a detailed questionnaire on socio-demographic characteristics and health status. The general methodology of PAQUID has been previously published (Dartigues et al. 1992). The studied sample composed of 1,469 subjects, from which 543 died between 1988 and 1997.



## **Health data**

Mortality data were provided by the National Institute of Health and Medical Research (INSERM), which carries out the coding of the medical causes of death according to the International Classification of Diseases, ninth revision (ICD-9). Causes of death corresponded to the principal cause recorded in the death register. Most studies investigate mortality from all non-accidental causes or from broad categories of illness such as cardiovascular and respiratory diseases. To ensure a sufficient sample size, we considered only two causes of deaths: cardio-respiratory causes (ICD-9 codes 460-519 and codes 390-459) and all-cause deaths except those from accidental causes (ICD-9 < 800). The residence area at the time of death was recorded. Influenza data were provided by the teleprocessing national network of monitoring and information on transmissible diseases (Valleron and Garnerin 1992).

## **Environmental data**

The main sources of air pollution in the urban area of Bordeaux from 1988 to 1997 was motor vehicle emissions. Air pollution data were obtained from the APPA (Association de Prévention de la Pollution Atmosphérique), which operated a local monitoring network from 1981 to 1997. Stations were selected to represent background inner-city air quality levels (i.e., stations not directly influenced by industrial or road traffic sources of pollution). The ambient urban stations measures had to be sufficiently correlated (i.e., correlation > 0.70) and to have sufficiently similar mean levels of pollution. Four stations corresponded to these criteria for BS, measured by reflectometry, and six for SO<sub>2</sub>-AF measured by the acidimetric method. We constructed exposure indicators by calculating the arithmetic mean of daily concentrations recorded in the selected ambient urban stations. Meteorological data (daily temperature and daily relative humidity) were provided by Météo-France.

## Analytical Approach

The Cox proportional hazards model is widely used for statistical analysis in epidemiology studies, particularly owing to its simple calculation and clear interpretation. It provides a parametric relation between the risk factors included in the model and the survival distribution without imposing a parametric form. In failure time analysis, continuous time-dependent covariates are rarely used. In this paper, the measurements of air pollutant concentrations were time-dependent covariates with 3,653 different values over the 10 years of follow-up. For each cause of death and each pollutant, a time-dependent Cox proportional hazards model was fitted, modeling the relative risk of death for a  $10 \mu\text{g}/\text{m}^3$  increase in pollutant concentration. Age was chosen as the basic time scale for two major reasons. First, it makes possible to study age, an important risk factor of death, without making parametric assumptions about the effect of this variable. Second, the effect of air pollution is not identifiable when calendar time is used. Using age instead of calendar time solves the problem. We then used a time-dependent Cox proportional hazards model with delayed entry (Klein and Moeschberger 1997) represented by the risk function at age  $a$  for a subject  $i$  :

$$h_i(a) = h_0(a) \exp[\beta_1 \mathbf{X}_i \beta_2 \mathbf{Z}_i(a)] \quad (1)$$

where  $a$  was the age,  $h_0(a)$  was the unspecified baseline hazard function,  $\mathbf{X}_i$  was the vector of time-independent explicative variables (gender, occupational exposure, and cigarette smoke exposure) and  $\mathbf{Z}_i(a)$ , the vector of time-dependent explicative variables (air pollution, temperature, humidity, influenza epidemics, season, and day of the week),  $\beta_j$ ,  $j=1,2$  being the vector of the unknown model parameters. The risk set was calculated for each age of death with  $a_1, a_2, \dots, a_{De}$ , corresponding to the ordered ages of death observed in the sample. Because we used a Cox proportional hazards model, a first condition to be included in the risk

set at age  $a_i$ , was to be alive until this age. The second condition to be included in the risk set at age  $a_i$ , was to be younger than  $a_i$  at inclusion. This ensure that we know air pollution exposure of all the subjects included in the risk set at age  $a_i$ .

It is now admitted that the death on a given day is not only a function of the same day exposures to pollutant but is also affected by exposures during a certain lag period (few days). Therefore, a distributed lag model was used. Distributed lag models have been used for decades in social sciences (Judge et al. 1980) and Pope and Schwartz (Pope et al. 1992; Pope and Schwartz 1996; Schwartz 2000), Zanobetti et al. (2000), and Rondeau et al. (2005) recently described the use of this approach in epidemiology for generalized additive models or for generalized linear models. We have adapted the distributed lag structure on the survival models:

$$h_i(a) = h_0(a) \exp \left[ \beta_1 X_i + \beta_2 Z_i(a) + \sum_{\ell=0}^L \alpha_\ell P_{i\ell}(a) \right] \quad (2)$$

where air pollution concentrations were represented by  $P_{i\ell}(a)$ , and  $\alpha_\ell$  indicated the magnitude of the effect of air pollution at day  $\ell$  with  $\ell=1, \dots, L$  and  $L$  being the number of lag days for the pollutant covariates. The length of the lag was chosen to minimize the Akaike's Information Criterion (AIC) (Akaike 1973) for each of the four models. AIC represents improvement in model fit versus the number of degrees of freedom that produced that improvement. Given that exposure is serially correlated on several subsequent days, the above unconstrained distributed lag model will have a high degree of multi-collinearity (Schwartz 2000). The collinearity among the lagged coefficients will lead to unstable estimation of the  $\alpha_\ell$  values with an increase in the variance of each estimator. Following the work of Almon in 1965, we assumed that the effects  $\alpha_\ell$  of the lagged exposure in the

previous model (2) followed a polynomial of sufficient degree  $D$ , i.e.  $\alpha_\ell = \sum_{d=0}^D \gamma_d \ell^d$ . A constrained model was used. This implied that the effects of air pollution were distributed over the previous  $L$  days following a polynomial function of degree  $D$ . The quantity  $\alpha_\ell$  was then interpretable as the polynomially smoothed estimate of the effect of air pollution on lagged  $\ell$  days, and their sum,  $\sum_{\ell=0}^L \alpha_\ell$ , was the overall effect of air pollution over the entire lag period. Although we know that the effect of air pollution concentrations is distributed over time, no assumption is made about the form of the effect on days close together which allows for a wide variety of lag structures. The polynomial distributed lag model allows to adjust for non-linear effects which are more in line with reality. This approach has the advantage of reducing both the colinearity and the number of parameters to be estimated via the assumed polynomial structure. In our case, the number of parameters was small, but when the lag is longer this advantage can not be neglected. The explanatory power of air pollution on daily deaths being modest, parsimony in the degree of the polynomial was necessary. Polynomial degree can be chosen arbitrarily or by AIC, but should not exceed three in most cases (Schwartz 2000; Zanobetti et al. 2000). In fact, a 2<sup>nd</sup> or 3<sup>rd</sup> degree offers sufficient flexibility in most cases of distributed constrained lag models (Pope and Schwartz 1996). For these reasons and in order not to multiply the number of tests, we chose a 2<sup>nd</sup> degree polynomial. .

### **Individual risk factors and potential modifiers**

Results were obtained after adjusting for individual risk factors (gender, cigarette smoke exposure and occupational exposure) and for time-dependent confounders (temperature, humidity, influenza epidemics, season, and day of the week). An ascending method was applied to include individual risk factors in each of the four models according to

the level of significance (<25 per cent). Individual factors were considered constant on average during the 10-year study period because subjects were aged over 65 years old so they did not work anymore and their occupational exposure could not change. Three classes were defined for occupational exposure: never worked, white collar and blue collar. This factor represents direct effect of occupational exposure on mortality but also indirect effect such as the way of life and life habits strongly depend on the occupational category. Status about cigarette smoke exposure was defined at the inclusion in the study such as non-smoker, ex-smoker or current smoker. This factor was considered constant during the ten years of follow-up because there are few changes in smoking habits over 65 years old. Generally, changes occurring over 65 years old concern rather subjects who stop smoking but maintain a risk due to their former smoking exposure than subjects who start smoking. This assumption of constant individual factors could lead to possible bias, but with a limited impact since we assumed here a long-term effect of the smoking. If the study period had been longer or the subjects younger, it would have been necessary to take into account the evolution of these factors over time. We took into account the mobility of the subjects in the analyses. Subjects were censored at the exact date of moving house. When we did not have precise information on the moving, they were censored at their last follow-up before moving. Mortality and pollution indicators undergo temporal variations due to several factors known or unknown. Such changes may appear in the long (annual variations) or medium term (seasonal variations, weekly). An ascending method was applied to include potential time-dependent confounders in each of the four models according to the level of variation of the other estimated parameters (25 per cent at least) already included in the model. Minimal temperature and relative humidity were included using the average on the selected lag for each model. Seasons were represented by a binary variable: summer-spring (21<sup>st</sup> March-20<sup>th</sup> September) versus fall-winter (21<sup>st</sup> September-20<sup>th</sup> March). Influenza epidemics were defined on the basis of

HAL author manuscript inserm-00138548, version 1

their graphical description when more than 300 cases occurred in a week and were treated as a dichotomous variable. In addition, we introduced in each model an unspecified function of time to take into account the long-term time trends in the data. It was estimated by a Truncated Power Basis Spline (Heuer 1997) usually written TP-Spline defined by  $m$  knots. We used 5 equidistant inner knots during the 10-year study period. We also tested other combinations of number and position of knots in order to assess the sensitivity of the results. Analyses were conducted with the SAS software (The Statistical Analysis System (version 8.2), SAS Institute, Inc., Cary, North Carolina).

## RESULTS

During the study period, the mean level of BS was  $17.0 \mu\text{g}/\text{m}^3$  (Standard Deviation =  $10.6 \mu\text{g}/\text{m}^3$ ) with a minimum of  $1.8 \mu\text{g}/\text{m}^3$  and a maximum of  $99.0 \mu\text{g}/\text{m}^3$ . For SO<sub>2</sub>-AF, the mean level was  $10.3 \mu\text{g}/\text{m}^3$  (SD =  $6.6 \mu\text{g}/\text{m}^3$ ) varying between 0.0 and  $64.6 \mu\text{g}/\text{m}^3$ . The correlation between daily levels of BS and SO<sub>2</sub>-AF was 0.67. During the same period, mean for minimal temperature was  $9.3^\circ\text{C}$  and mean relative humidity was 59.9 %. Among the 543 deceased subjects (248 women and 295 men), we studied 439 deaths from non-accidental causes and 158 from cardio-respiratory causes (127 cardiac and 31 respiratory). For all non-accidental causes, 50 % of deceased subjects were aged over 83 years. Table 1 describes the main characteristics of the 1,469 subjects included in the analysis.

By univariate analysis, the data did not demonstrate a significant association between mortality and educational level. Mortality differed between occupational exposure categories. After adjustment for gender, smoke exposure or air pollution concentration, the estimated risk remained higher but not statistically significant for blue collar workers. Because its confounding role is recognized in the literature, we forced occupational exposure into the models.

HAL author manuscript inserm-00138548, version 1

Whatever the causes of mortality or the pollutant studied, the Cox proportional hazards model with the polynomial distributed lag period selected by the AIC did not demonstrate any significant cumulative effect after adjustment for individual factors and temporal confounders. Figure 1 shows the estimated cumulative effect and the estimated effect of each single day's exposure to BS and SO<sub>2</sub>-AF across five days and for different causes of death. For all non-accidental deaths and according to the AIC, lag 3 and 4 were selected to represent the association with BS and SO<sub>2</sub>-AF, respectively (Table 2). Table 2 summarizes the results of all non-accidental death analyses adjusted for individual factors and temporal confounders. As expected, women had a lower risk of death than men (RR = 0.61, 95% CI: 0.46, 0.79), and smokers and ex-smokers had a risk of death about 50 % greater than non-smokers. For single day exposure, there was a greater risk of death for the 3<sup>rd</sup> day after exposure to BS (RR = 1.12, 95% CI: 0.99, 1.26) and for the 4<sup>th</sup> day after exposure to SO<sub>2</sub>-AF (RR = 1.17, 95% CI: 0.99, 1.39), but these associations were not significant at 5 %. A protective effect for the 1<sup>st</sup> and 2<sup>nd</sup> days after exposure to BS was found, which is very surprising and probably due to an accentuation of the polynomial structure. The estimated effects of the 1<sup>st</sup> and 2<sup>nd</sup> days' exposure to SO<sub>2</sub>-AF were found to be negative.

A lag 3 for BS and a lag 5 for SO<sub>2</sub>-AF were selected to examine the association of these two pollutants with deaths from cardio-respiratory causes (Table 3). Concerning individual characteristics, the same estimated risk of death was observed for women versus men as with all non-accidental deaths. The estimated effect of smoking was greater for cardio-respiratory deaths than for all non-accidental deaths. The insignificant effect for current smokers was probably due to the fact that there were only 17 subjects in this class. When adjusted for individual factors and temporal confounders, results according to single day exposure showed that a 10 µg/m<sup>3</sup> increase in BS was associated with an estimated 24 % (RR = 1.24, 95% CI : 1.04, 1.47) increase in cardio-respiratory mortality three days after. The same increase in

SO<sub>2</sub>-AF was associated with an estimated 19 % excess of deaths on the 2<sup>nd</sup> and 3<sup>rd</sup> days after exposure (RR = 1.19, 95% CI: 1.03, 1.37).

### **Sensitivity**

According to the AIC, introduction of the unspecified function of time improved the fit of the four models. To check the stability of our results, we tested on the one hand a 3<sup>rd</sup> degree function and on the other hand, an increase until 16 inner knots for one of the four models. No appreciable differences in the estimated associations were observed. We also explored the influence of the degree of the polynomial lag structure and then replaced the 2<sup>nd</sup> degree of the polynomial lag structure by a 3<sup>rd</sup> degree. The estimated associations between air pollution and mortality were unchanged for all the models except for all non-accidental deaths and BS which had a rate ratio lower than one. For this association, 1<sup>st</sup> and 2<sup>nd</sup> day exposures became insignificant with the 3<sup>rd</sup> degree whereas they were significantly associated with a 2<sup>nd</sup> degree polynomial lag structure (cf. Table 2).

To validate the assumption that the effects of air pollution on mortality were not distributed beyond 5 days, we tested a lag period of 10 and 15 days for the association between cardio-respiratory mortality and SO<sub>2</sub>-AF. The models were no longer statistically satisfactory. The cumulative effects were very similar for 5, 10 and 15 days but the confidence intervals increased with the number of days. By increasing the lag period to 10 or 15 days, the individual effects of each single day were no more significant and the curve of the distributed lag effects tended to be smoother. Moreover, this effect reached zero after the 5<sup>th</sup> day of the lag period and became negative, as in Schwartz (2000) and in the report of the InVS (2001). The proportional hazard assumption was tested by using an interaction between age and different variables, according to the log likelihood ratio test. This assumption was valid for all



fixed variables (time-independent) of all models except for cigarette smoke exposure ( $p=0.05$ ) in the two models concerning all non-accidental mortality.

## DISCUSSION

We found a positive association between short-term variations in BS and SO<sub>2</sub>-AF levels and mortality among persons aged of 65 years and older. Several studies have found some effects of particulate air pollution on mortality outcomes. Schwartz (2000) observed an association between daily deaths of persons 65 years of age and older for all causes with particulate matter  $< 10 \mu\text{m}$  aerodynamic diameter (PM<sub>10</sub>), a result confirmed by Katsouyanni et al. (2001) in Europe with BS in the study entitled “Air Pollution and Health: a European Approach” (APHEA-2). In London, Bremner et al. (1999) showed a significant association between BS and respiratory and cardiovascular mortality. They also showed that SO<sub>2</sub>-AF pollution was significantly associated with respiratory mortality among the elderly. Our findings point to a significant association between cardio-respiratory mortality and air pollution among the elderly only for the 3<sup>rd</sup> lag day with BS and for the 2<sup>nd</sup> and 3<sup>rd</sup> lags with SO<sub>2</sub>-AF. In France, a study (InVS 2001) analysed the time series of deaths due to cardiovascular and respiratory diseases in nine French cities with a 3<sup>rd</sup> degree polynomial lag structure from 0 to 5 days. They found that BS association with cardiovascular deaths was strongest for lag 3 but there was not significant effect with respiratory mortality. The association between SO<sub>2</sub>-AF and cardiovascular deaths was strongest for lags 1 and 2 and no significant association was observed with respiratory mortality. Bremner et al. (1999) explored different lag periods in the relationship between air pollution and mortality for cardiovascular and respiratory causes without polynomial structure. They found a greater risk of death associated with BS for cardiovascular causes for lag 1 and for respiratory causes for lag 3. Concerning SO<sub>2</sub>-AF, the three first days of the lag were positively associated with

respiratory mortality but no association was observed with cardiovascular deaths. Moreover, a study by Téllez-Rojo et al. (2000) also found a greater risk of death for respiratory causes outside medical units for lag 3 of PM10 among elderly subjects in Mexico City.

Concerning all non-accidental deaths, our analysis found positive associations especially for lag 3 for BS and lag 4 for SO<sub>2</sub>-AF, which were at the limit of significance, as opposed to the negative effect of lags 1 and 2. This result can-not be explained in terms of the French health care system delaying the deaths. We also tested an unconstrained model (by adjusting directly on each lag day) and the effects were negative but not significant. We believe that the significance of this result could be due to the polynomial structure of the effect of the pollutant. Bremner et al. (1999) found insignificant results with all-cause mortality. However, the InVS study (2001) found a significant association between the four first days of the lag for BS and SO<sub>2</sub>-AF. On the contrary, we did not find any significant cumulative effects, but the risks of death of a single day's exposure that we observed were greater than in the other studies. Filleul et al. (2004a) analysed the time series of deaths due to respiratory, cardiovascular and all on-accidental causes among the elderly over the period 1988-1997 in Bordeaux City. They used generalized additive models with a cumulated lag period of 5 days and found concordant results with ours. The increase in respiratory mortality cumulated over five day's lag was 9.2% (95% CI= 3.4 %-15.3 %) for a 10 µg/m<sup>3</sup> increase in BS. Our study showed an increase in cardio-respiratory mortality of 23.7 % (95% CI= 3.9 %-47.2 %) for lag 3. Concerning exposure to SO<sub>2</sub>-AF, Filleul et al. (2004a) showed a 20.6 % (95% CI= 9.3 %-33.2 %) excess of respiratory mortality and we found a 19.0 % (95% CI= 3.1 %-37.4 %) excess of cardio-respiratory mortality for the 2<sup>nd</sup> and 3<sup>rd</sup> days after exposure.

The results obtained with the survival analysis are very similar to those obtained by the case-crossover method for exposure to BS (Filleul et al. 2004b) in the same population. In the case-crossover analysis, a restricted distributed lag model with a polynomial effect of the pollutant was used. According to the AIC, a 2<sup>nd</sup> degree was chosen for the polynomial. An a priori lag period of 0 to 3 days before the event was chosen based on the literature. The data did not demonstrate any cumulative effect after adjustment for meteorological data (daily temperature and relative humidity). Odds ratios for the cumulative effect and for all non-accidental and cardio-respiratory mortality were respectively 0.79 (95% CI: 0.62, 1.02) and 0.89 (95% CI: 0.59, 1.34) for a 10 µg/m<sup>3</sup> increase in BS. Nevertheless, an association was found between the third lag day and all non-accidental mortality (OR = 1.19, 95% CI: 0.99, 1.43), which was significant for cardio-respiratory mortality (OR = 1.30, 95% CI: 1.01, 1.68). Therefore, the present results using survival analysis are concordant with those obtained with the case-crossover analysis for the lag period, for the degree of the polynomial and for the level of the risks. The confidence intervals are more restricted with the Cox proportional hazards model. Therefore, we consider that the Cox proportional hazards model should be applied when a cohort is available because survival analysis exploits all available information and increases the power of the study. Moreover, it makes it possible to identify and adjust for individual risk factors. By using the Cox proportional hazards model where age is the time scale, it is possible to adjust non-parametrically for age. This method should prove particularly useful in the future to simultaneously analyse the short- and long-term effects of air pollution.

## REFERENCES

Akaike H. 1973. Information theory and an extension of the maximum likelihood principal. In: Second International Symposium on Information Theory (Petrov BN, Csaki F, eds). Budapest: Akademiai Kiado, 267-281.

Almon S. 1965. The distributed lag between capital appropriations and expenditures. *Econometrica* 33:178-196.

Bateson TF, Schwartz J. 2001. Selection bias and confounding in case-crossover analyses of environmental time-series data. *Epidemiology* 12:654-661.

Braga ALF, Zanobetti A, Schwartz J. 2000. Do respiratory epidemics confound Associations between mortality and air pollution in central Europe. *Environ Health Perspect* 108:283-287.

Bremner SA, Anderson HR, McMichael AJ, Strachan DP, Bland JM, Bower JS. 1999. Short term associations between outdoor air pollution and mortality in London 1992-4. *Occup Environ Med* 56:237-244.

Cox DR. 1972. Regression models and life-tables. *J R Stat Soc* 34:187-220.

Dartigues JF, Gagnon M, Baberger-Gateau P, Letenneur L, Commenges D, Sauvel C, et al. 1992. The paquid epidemiological program on brain ageing. *Neuroepidemiology* 11:14-18.

Dockery DW, Pope III CA, Xu X, Spengler JD, Ware JH, Fay ME, et al. 1993. An association between air pollution and mortality in six US cities. *N Engl J Med* 329:1753-1759.

Filleul L, Le Tertre A, Baldi I, Tessier JF. 2004a. Difference in the relation between daily mortality and air pollution among elderly and all-ages populations in southwestern France. *Environ Res* 94:249-253.

Filleul L, Rondeau V, Cantagrel A, Dartigues JF, Tessier JF. 2004b. Do subject characteristics modify the effects of particulate air pollution on daily mortality among elderly. *J Occup Environ Med* 46:1115-1122.

Heuer C. 1997. Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics* 53:161-177.

InVS (Institut de Veille Sanitaire). 2001. Surveillance des effets sur la santé liés à la pollution atmosphérique en milieu urbain—phase II: rapport de l'étude. Saint Maurice.

Judge GG, Griffiths WE, Hill RC, Lutkepohl H, Lee TC. 1980. *The theory and practice of econometrics*. New York:Wiley and Sons.

Katsouyanni K, Touloumi G, Samoli E, Gryparis A, Le Tertre A, Monopolis Y, et al. 2001. Confounding effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA-2 project. *Epidemiology* 12:521-531.

Klein JP, Moeschberger ML. 1997. *Survival analysis: Techniques for censored and truncated data*. New-York:Springer-Verlag.

Navidi W, Weinhandl E. 2002. Risk set sampling for case-crossover designs. *Epidemiology* 13:100-105.

Peters A, Skorkovsky J, Kotesovec F, Brynda J, Spix C, Wichmann HE, et al. 2000. Associations between mortality and air pollution in central Europe. *Environ Health Perspect* 108:283-287.

Pope CA, Schwartz J. 1996. Time series for the analysis of pulmonary health data. *Am J Respir Crit Care Med* 154:S229-S233.

Pope CA, Schwartz J, Ransom M. 1992. Daily mortality and PM10 pollution in Utah Valley. *Arch Environ Health* 42:211-217.

Pope CA, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, et al. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. *Am J Respir Crit Care Med* 151:669-79.

Rondeau V, Berhane K, Thomas DC. 2005. A three level model for binary time-series data: the effects of air pollution on school absences in the southern California Children's Cohort Health Study. *Stat Med* 24:1103-1115.

Rossi G, Vigotti MA, Zanobetti A, Repetto F, Gianelle V, Schwartz J. 1999. Air pollution and cause specific mortality in Milan, Italy, 1980-1989. *Arch Environ Health* 54:158-164.

Schwartz J. 2000. The distributed lag between air pollution and daily deaths. *Epidemiology* 11:320-326.

Télez-Rojo MM, Romieu I, RuizVelasco S, Lezana MA, HernandezAvila MM. 2000. Daily respiratory mortality and PM10 pollution in Mexico City: importance of considering place of death. *Eur Respir J* 16:391-396.

Valleron AJ, Garnerin P. 1992. Computer networking as a tool for public health surveillance: The French experiment. *Morb Mortal Wkly Rep* 41 :101-110.

Zanobetti A, Wand MP, Schwartz J, Ryan LM. 2000. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* 1:279-292.

**Table 1: Characteristics of the PAQUID cohort living in the urban area of Bordeaux, 1988-1997.**

	Deaths for all non-accidental causes N=439	Deaths for cardio-respiratory causes N=158	All subjects N=1469
Age of death (years, median (Min., Max.))	83.1 (66.1;106.1)	84.2 (67.3;102.9)	
Gender (%)			
Men	48.5	51.9	38.3
Women	51.5	48.1	61.7
Family (%)			
Living alone	49.9	46.2	47.0
Living in couple	50.1	53.8	53.0
Educational levels (%)			
Without primary school diploma	28.9	27.2	25.5
Primary school diploma or secondary not validated	57.4	59.5	59.9
Secondary validated or higher	13.7	13.3	14.6
Occupational exposure (%)			
Never worked	11.8	7.0	11.1
White collar	40.8	44.9	45.2
Blue collar	47.4	48.1	43.7
Smoking habits (%)			
Non-smoker	53.3	49.4	62.3
Ex-smoker	34.6	39.2	26.8
Current smoker	11.8	10.8	10.5



**Table 2: Adjusted all non-accidental mortality Rate Ratio estimates from Cox proportional hazards models with a polynomial distributed lag effect for a 10  $\mu\text{g}/\text{m}^3$  increase in air pollution (BS and SO<sub>2</sub>-AF), Bordeaux, France, 1988-1997.**

	All non-accidental deaths and			
	Black smoke <sup>a</sup>		SO <sub>2</sub> -AF <sup>b</sup>	
	RR	95% CI	RR	95% CI
Women vs. men	0.61*	0.46-0.79	0.61*	0.47-0.79
Occupational exposure vs. never worked				
White collar	0.77	0.55-1.08	0.77	0.55-1.08
Blue collar	0.97	0.70-1.34	0.97	0.70-1.34
Smoking habits vs. non-smoker				
Ex-smoker	1.50*	1.14-1.97	1.50*	1.14-1.97
Current smoker	1.65*	1.17-2.32	1.65*	1.17-2.32
Distributed effect of air pollution				
Lag 0	1.11	0.98-1.25	1.03	0.86-1.24
Lag 1	0.90*	0.82-0.98	0.96	0.88-1.06
Lag 2	0.90*	0.82-0.99	0.96	0.85-1.09
Lag 3	1.12	0.99-1.26	1.03	0.94-1.12
Lag 4			1.17	0.99-1.39
Cumulative effect	1.00	0.87-1.16	1.16	0.86-1.55

<sup>a</sup> Adjusted for temperature, day of week and function of time.

<sup>b</sup> Adjusted for temperature, humidity, day of week and function of time.

\*  $p < 0.05$

**Table 3: Adjusted cardio-respiratory mortality Rate Ratio estimates from Cox proportional hazards models with a polynomial distributed lag effect for a 10 µg/m<sup>3</sup> increase in air pollution (BS and SO<sub>2</sub>-AF), Bordeaux, France, 1988-1997.**

	Cardio-respiratory deaths and			
	Black smoke <sup>a</sup>		SO <sub>2</sub> -AF <sup>b</sup>	
	RR	95% CI	RR	95%CI
Women vs. men	0.65	0.42-1.01	0.65	0.42-1.01
Occupational exposure vs. never worked				
White collar	1.34	0.68-2.63	1.34	0.68-2.62
Blue collar	1.55	0.80-2.98	1.53	0.80-2.95
Smoking habits vs. non-smoker				
Ex-smoker	1.85*	1.18-2.89	1.84*	1.18-2.88
Current smoker	1.75	0.97-3.16	1.75	0.97-3.16
Distributed effect of air pollution				
Lag 0	1.09	0.91-1.32	0.84	0.65-1.10
Lag 1	0.92	0.80-1.05	1.06	0.94-1.19
Lag 2	0.96	0.84-1.10	1.19*	1.03-1.37
Lag 3	1.24*	1.04-1.47	1.19*	1.03-1.37
Lag 4			1.07	0.95-1.19
Lag 5			0.85	0.66-1.10
Cumulative effect	1.19	0.95-1.47	1.15	0.75-1.77

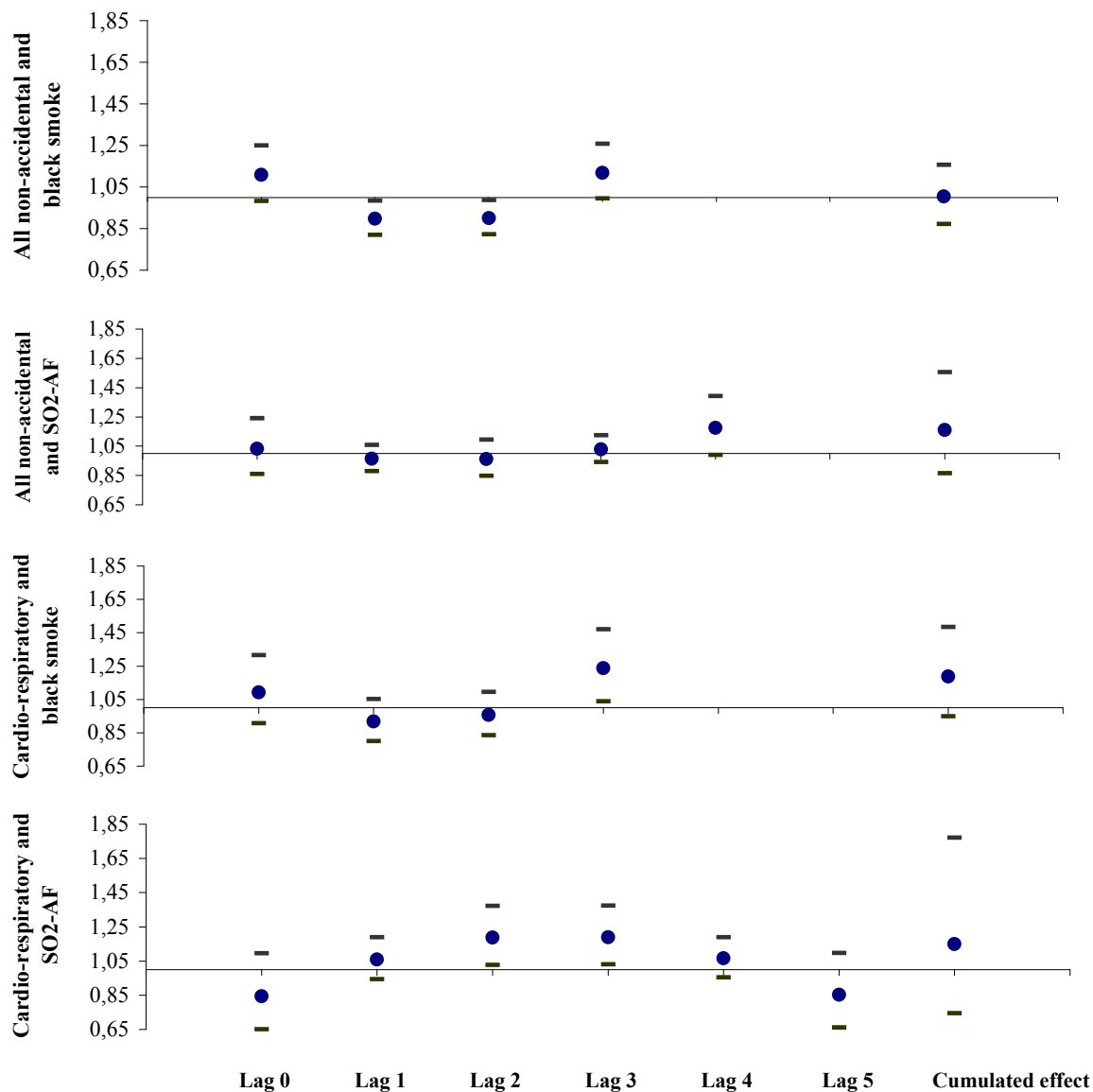
<sup>a</sup> Adjusted for temperature, day of week and function of time.

<sup>b</sup> Adjusted for temperature and function of time.

\* p < 0.05

## FIGURE LEGEND

Figure 1: Effect of a  $10 \mu\text{g}/\text{m}^3$  increase in air pollution on death using a Cox proportional hazards model with a second degree polynomial distributed lag model.



<sup>a</sup> adjusted for meteorological variables

**Figure 1:** Effect of a 10 µg/m<sup>3</sup> increase in air pollution on death using a Cox proportional hazards model with a second degree polynomial distributed lag model<sup>a</sup>.