



**HAL**  
open science

## Multi-pollutant exposure profiles associated with breast cancer risk: A Bayesian profile regression analysis in the French E3N cohort

Camille Giampiccolo, Amina Amadou, Thomas Coudon, Delphine Praud, Lény Grassot, Elodie Faure, Florian Couvidat, Gianluca Severi, Francesca Romana Mancini, Béatrice Fervers, et al.

### ► To cite this version:

Camille Giampiccolo, Amina Amadou, Thomas Coudon, Delphine Praud, Lény Grassot, et al.. Multi-pollutant exposure profiles associated with breast cancer risk: A Bayesian profile regression analysis in the French E3N cohort. *Environment International*, 2024, 190, pp.108943. 10.1016/j.envint.2024.108943 . hal-04673719

**HAL Id: hal-04673719**

**<https://hal.science/hal-04673719>**

Submitted on 20 Aug 2024

**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.



Distributed under a Creative Commons Attribution 4.0 International License



Full length article

## Multi-pollutant exposure profiles associated with breast cancer risk: A Bayesian profile regression analysis in the French E3N cohort

Camille Giampiccolo<sup>a,b,c</sup>, Amina Amadou<sup>a,b,\*</sup>, Thomas Coudon<sup>a,b</sup>, Delphine Praud<sup>a,b</sup>,  
Lény Grassot<sup>a,b</sup>, Elodie Faure<sup>d</sup>, Florian Couvidat<sup>e</sup>, Gianluca Severi<sup>d,f</sup>,  
Francesca Romana Mancini<sup>d</sup>, Béatrice Fervers<sup>a,b,1</sup>, Pascal Roy<sup>c,g,h,1</sup>

<sup>a</sup> Department of Prevention Cancer Environnement, Centre Léon Bérard, Lyon, France

<sup>b</sup> Inserm U1296 Radiations : Défense, Santé, Environnement, Lyon, France

<sup>c</sup> Laboratoire de Biométrie Et Biologie Evolutive, CNRS UMR 5558, Villeurbanne, France

<sup>d</sup> Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, Villejuif, France

<sup>e</sup> National Institute for Industrial Environment and Risks (INERIS), Verneuil-en-Halatte, France

<sup>f</sup> Department of Statistics, Computer Science and Applications (DISIA), University of Florence, Florence, Italy

<sup>g</sup> Université Claude Bernard Lyon 1, Lyon, France

<sup>h</sup> Service de Biostatistique-Bioinformatique, Pole Sante Publique, Hospices Civils de Lyon, Lyon, France

## ARTICLE INFO

Handling Editor: Adrian Covaci

## Keywords:

Breast cancer

Air pollutants

Mixture

Correlated exposures

Bayesian profile regression

Cluster

## ABSTRACT

**Background:** Human exposure to air pollution involves complex mixtures of multiple correlated air pollutants. To date, very few studies have assessed the combined effects of exposure to multiple air pollutants on breast cancer (BC) risk.

**Objectives:** We aimed to assess the association between combined exposures to multiple air pollutants and breast cancer risk.

**Methods:** The study was based on a case-control study nested within the French E3N cohort (5222 incident BC cases/5222 matched controls). For each woman, the average of the mean annual exposure to eight pollutants (benzo(a)pyrene, cadmium, dioxins, polychlorinated biphenyls (PCB153), nitrogen dioxide (NO<sub>2</sub>), ozone, particulate matter and fine particles (PMs)) was estimated from cohort inclusion in 1990 to the index date. We used the Bayesian Profile Regression (BPR) model, which groups individuals according to their exposure and risk levels, and assigns a risk to each cluster identified. The model was adjusted on a combination of matching variables and confounders to better consider the design of the nested case-control study. Odds ratios (OR) and their 95 % credible intervals (CrI) were estimated.

**Results:** Among the 21 clusters identified, the cluster characterised by low exposures to all pollutants, except ozone, was taken as reference. A consistent increase in BC risk compared to the reference cluster was observed for 3 clusters: cluster 9 (OR=1.61; CrI=1.13,2.26), cluster 16 (OR=1.59; CrI=1.10,2.30) and cluster 15 (OR=1.38; CrI=1.00,1.88) characterised by high levels of NO<sub>2</sub>, PMs and PCB153. The other clusters showed no consistent association with BC.

**Discussion:** This is the first study assessing the effect of exposure to a mixture of eight air pollutants on BC risk, using the BPR approach. Overall, results showed evidence of a positive joint effect of exposure to high levels to most pollutants, particularly high for NO<sub>2</sub>, PMs and PCB153, on the risk of BC.

**Abbreviations:** AFP, Age at first full-term pregnancy; BaP, Benzo[a]pyrene; BC, Breast cancer; BMI, Body Mass Index; BPR, Bayesian Profile Regression; CI, Confidence Interval; CrI, Credible Interval; CNIL, National Commission for Data Protection and Privacy; DPM, Dirichlet Process Mixture; E3N, Etude Epidémiologique auprès de femmes de la Mutuelle Générale de l'Éducation Nationale; IARC, International Agency for Research on Cancer; IGN, National Geographic Institute; MCMC, Markov Chain Monte Carlo; MET, Metabolic equivalent task; NO<sub>2</sub>, Nitrogen dioxide; ORs, Odds ratios; O<sub>3</sub>, Ozone; PAM, Partitioning Around Medoids; PCBs, Polychlorinated biphenyls; PCB153, group III of the Wolff's classification of polychlorinated biphenyls; PMs, Particles (including PM<sub>10</sub> and PM<sub>2.5</sub>); PM<sub>10</sub>, Particles (diameter < 10 µm); PM<sub>2.5</sub>, Fine particles (diameter < 2.5 µm); SD, Standard deviation.

\* Corresponding author at: Département Prévention Cancer Environnement, Centre Léon Bérard, Inserm U1296 Radiations : Défense, Santé, Environnement, 28 rue Laënnec 69373, Lyon, France.

E-mail address: [amina.amadou@lyon.unicancer.fr](mailto:amina.amadou@lyon.unicancer.fr) (A. Amadou).

<sup>1</sup> These authors contributed equally to this work and share last authorship.

<https://doi.org/10.1016/j.envint.2024.108943>

Received 25 April 2024; Received in revised form 2 August 2024; Accepted 6 August 2024

Available online 8 August 2024

0160-4120/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Exposure to air pollutants is a major public health concern because of its many adverse effects on human health, including cancer (Turner et al., 2020). Air pollution is a complex mixture of highly correlated pollutants from multiple sources (Billionnet et al., 2012). The International Agency for Research on Cancer (IARC) has classified outdoor air pollution as a whole, and particulate matter (PM) as carcinogenic to humans with sufficient evidence for lung cancer and a positive association for bladder cancer (Loomis et al., 2013). Other chemical contaminants present in the air have been classified as carcinogenic to humans, such as benzo(a)pyrene (BaP), cadmium, dioxins, and polychlorinated biphenyls (PCBs). Exposure to chemical air contaminants, in particular those with endocrine disrupting effects are strongly suspected to play a role in breast cancer (BC) development (Darbre, 2021; Wan et al., 2022; Yilmaz et al., 2020), the most common cancer in women worldwide (2.3 million new cases estimated in 2020) (Sung et al., 2021).

The body of evidence reporting positive associations between exposure to air pollutants and risk of BC is steadily increasing, both for pollutants with endocrine disrupting effects as well as for nitrogen dioxide (NO<sub>2</sub>) and PMs (Gabet et al., 2021; Niehoff et al., 2022; Praud et al., 2023; White et al., 2021, 2018). An increased BC risk has been shown for BaP exposure (Amadou et al., 2021; Large and Wei, 2017), while inconsistent findings have been reported from epidemiological studies for dioxins (Danjou et al., 2019; Fiolet et al., 2022; VoPham et al., 2020; Xu et al., 2016), cadmium (Amadou et al., 2020b; Filippini et al., 2020; Florez-Garcia et al., 2023), as well as PCBs (Deygas et al., 2021; Fiolet et al., 2022). Several studies have shown an increased BC risk related to NO<sub>2</sub> exposure, largely emitted from road traffic (Amadou et al., 2023; Gabet et al., 2021; Praud et al., 2023; Wei et al., 2021). Concerning PM<sub>10</sub> and PM<sub>2.5</sub>, although positive associations were reported in some studies, the findings remained globally inconsistent across studies (Wei et al., 2021; Zhang et al., 2019). With regard to ozone (O<sub>3</sub>), a “secondary” pollutant, the available studies showed no association between BC risk and high exposure to O<sub>3</sub> (Bai et al., 2020; White et al., 2021), but an inverse association has been observed in relation to breast density, a well-established risk factor for BC (Yaghjian et al., 2017).

While the population is exposed simultaneously to multiple air pollutants, previous studies have mainly investigated the adverse effects of exposure to air pollutants considering one pollutant at a time. Traditional epidemiological methods, such as multiple logistic regression, are not adapted to investigate joint health effects of simultaneous exposure to multiple highly correlated pollutants leading to variance inflation of the estimates (Agier et al., 2016; Farrar and Glauber, 1967; Hoerl and Kennard, 1970; MacLehose et al., 2007). Consequently, parameter estimates of the model may not be significantly different from zero in multivariate analyses adjusted on correlated variables, even when an effective relationship between exposure and health exists. There is thus a great interest in assessing the impact of exposure to correlated pollutant mixtures on BC risk, using more appropriate methods (Bellavia, 2023).

An increasing number of statistical methods for considering exposure to pollutant mixtures have been described in the literature to investigate these complex relationships. Among these methods, Bayesian profile regression (BPR) analysis, is a pertinent approach, for identifying at-risk groups of individuals who share similar exposure profiles of correlated pollutants, and assigning a risk to each of these groups (Molitor et al., 2010).

The main objective of this study was to estimate the joint effect of exposure to eight air pollutants (BaP, cadmium, dioxins, PCB153, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub> and PM<sub>2.5</sub>) on the risk of BC in a matched case-control study nested within the French E3N cohort using BPR analysis (Liverani et al., 2015; Molitor et al., 2010).

## 2. Material and methods

### 2.1. The cohort study population

The E3N-Generation prospective study is an ongoing French familial cohort study established as an extension of the E3N cohort of women (Etude Epidémiologique auprès des femmes de la Mutuelle Générale de l'Éducation Nationale) including the E3N women's children, their fathers and, in the future, their grandchildren. The present study focuses only on the E3N part of the cohort that was created in 1990 with the recruitment of 98,995 women that at that time were the E3N cohort of women (Etude Epidémiologique auprès des femmes de la Mutuelle Générale de l'Éducation Nationale) including the E3N women's children, their fathers and, in the future, their grandchildren. The present study focus only on the E3N part of the cohort that was created in 1990 with the recruitment of 98,995 women that at that time were aged 40–65 years, were living in France and were affiliated to the national health insurance covering workers from the French National Education System (Mutuelle Général de l'Éducation Nationale, MGEN) (Clavel-Chapelon and E3N Study Group, 2015). The E3N cohort is still followed-up after more than 30 years with self-administered questionnaires including data on socio-demographic characteristics, lifestyle, reproductive factors, anthropometry, past medical history, and familial history of cancer. The addresses of the cohort participants included in the study were collected at baseline (1990), and at each of the twelve follow-up questionnaires. BC occurrence was self-reported in follow-up questionnaires and a few additional cases were retrieved from health insurance data or from mortality data including causes of death obtained from the National Services on Causes of Deaths (CépiDC- Inserm). Self-reported cases were validated through retrieval of medical records from treating physicians and pathological confirmation was obtained for 93 % of cases. Since the false-positive rate of self-reports was low in the cohort population (<5%), we also included the cases that were not pathologically confirmed. For one fourth of the cohort blood samples were collected between 1995 and 1998 and are stored in liquid nitrogen as aliquots of buffy coat, plasma, serum and erythrocytes. The study was approved by the French National Commission for Data Protection and Privacy (CNIL), and informed consent was obtained from each participant.

### 2.2. The nested case-control study

This study is based on data from a case-control study nested in the E3N cohort described in previous articles (Amadou et al., 2020a). In brief, the study included participants without any cancer at baseline, who had reported their home address at baseline, and lived in metropolitan France during the follow-up period from 1990 to 2011 (Deygas et al., 2021). Of the 6,298 invasive BC cases diagnosed during follow-up, we excluded women with phyllodes tumours which are biologically and clinically different (N=19) (Zhang and Kleer, 2016), those with missing data on matching variables (N=3) and those with more than one missing address (N=1,054 cases) (Amadou et al., 2021; Deygas et al., 2021). For each case, one control free of cancer was randomly selected by incidence density sampling, among cohort participants at risk of BC at the time of case diagnosis (Amadou et al., 2020a). This incidence density sampling is an efficient method of choice for obtaining unbiased results, in which controls are selected without replacement from all persons at risk at the time of case occurrence, excluding the index case itself (Richardson, 2004). Two matchings were done according to the availability of a biological sample (blood or saliva) (Amadou et al., 2020a). This variable was one of the matching variables, as the case-control study was originally designed to also address additional objectives (impact of exposure on DNA methylation and interactions with genetic polymorphisms). The protocol of the matched nested case control study has been published in JMIR protocol (Amadou et al., 2020a). Briefly, cases having blood samples collected before diagnosis were matched to controls with a

blood sample collected before the index date of the matched case, and matched on the department of residence, age ( $\pm 1$  year), date ( $\pm 3$  months), and menopausal status, all at the time of blood collection. Cases without a blood sample were matched to controls on the same criteria, at inclusion, and additionally matched on availability of a saliva sample. The final study population included 5,222 women diagnosed with an invasive BC and 5,222 matched controls. The flow chart is presented in [Supplementary Figure S1](#).

### 2.3. Assessment of long-term exposure to eight atmospheric pollutants

The approach used to estimate the subjects' atmospheric exposure based on the pollutant concentrations at the women's consecutive residential addresses has been applied in our previous studies and described elsewhere ([Amadou et al., 2023, 2021](#)). Briefly, mean annual atmospheric BaP ( $\text{ng}/\text{m}^3$ ), cadmium ( $\text{pg}/\text{m}^3$ ), dioxins ( $\text{fg-TEQ}/\text{m}^3$ ), PCB153 ( $\text{pg}/\text{m}^3$ ),  $\text{NO}_2$  ( $\mu\text{g}/\text{m}^3$ ),  $\text{O}_3$  ( $\text{pp}/\text{m}^3$ ),  $\text{PM}_{10}$  ( $\mu\text{g}/\text{m}^3$ ) and  $\text{PM}_{2.5}$  ( $\mu\text{g}/\text{m}^3$ ) concentrations were estimated at the women's geocoded consecutive residence addresses, from 1990 to their index date using a chemistry-transport model (CHIMERE) with a spatial resolution of  $0.125 \times 0.0625$  (approximately  $7 \times 7$  km) ([Couvidat et al., 2018](#)). This deterministic model simulates pollutant concentrations on large scales using emission inventories and meteorological data as input data ([Couvidat et al., 2018; Guerreiro et al., 2016; Menut et al., 2013](#)). The CHIMERE model has been validated by comparisons with pollutant measurements ([Clément et al., 2017; Couvidat et al., 2018](#)) and has been used to estimate for example BaP exposure and other air pollutants in Europe ([Guerreiro et al., 2016](#)). French departments are administrative divisions of territories ("NUTS-3" in the classification of territorial divisions of the European Union) ([Eurostat, 2024](#)) with surfaces sufficiently large to exhibit heterogeneous exposure profiles with the CHIMERE model ( $7 \times 7$  km spatial resolution). Surface areas range from 105 to 10,000  $\text{km}^2$  (with 25th, 50th, and 75th percentiles of 5147  $\text{km}^2$ , 5954  $\text{km}^2$ , and 6775  $\text{km}^2$  respectively); with population size varying from 76,604 to 2,608,346 persons ([Amadou et al., 2023; INSEE, 2024](#)).

The consecutive residential addresses of the participants were geocoded with the ArcGIS Software (ArcGIS Locator version 10.0, Environmental System Research Institute – ESRI, Redlands, CA, USA), and its reference street network database, BD Adresse® from the National Geographic Institute (IGN) ([Faure et al., 2017](#)).

Weighted annual estimates of pollutant concentrations were assigned to the subsequent geocoded residential addresses of subjects for each year during the 1990–2011 follow-up period. For each woman, mean concentration levels of pollutant exposure at the residential address were calculated, from their entry into the cohort to their index date. Entry into the E3N (1990) cohort is thus the starting point for the exposure assessment in the present nested case-control study. Exposure assessment is performed until the index date (corresponding to the date of BC diagnosis for cases and date of selection for controls). The average of the mean annual concentrations for each pollutant prior to BC diagnosis and used as an estimate of the actual pollutant exposure in the present study ([Amadou et al., 2023, 2021; Deygas et al., 2021](#)).

### 2.4. Statistical analyses

#### 2.4.1. Descriptive analysis

Exposure estimates for each pollutant, socio-demographic characteristics, and other covariates of the participants were described using mean and standard deviation (SD) for continuous variables, and frequency and percentage for categorical variables. The Pearson correlation coefficients were computed to assess the correlations between exposure to different pollutants ([Gallo et al., 2018](#)).

#### 2.4.2. Bayesian profile regression (BPR)

BPR was used to assess the association between the combined exposure to the eight pollutants and the risk of BC by estimating the odds

ratio (OR) between groups of individuals sharing a similar exposure profile. BPR assigns individuals to clusters according to an assignment model. When grouping individuals into clusters, the assignment model considers both the information on joint exposures, provided by an exposure sub-model, and the expected cluster effect on BC estimated from risk provided by a disease sub-model ([Liverani et al., 2015; Molitor et al., 2010](#)).

Details of the BPR model are described in [supplementary material](#). Briefly, the assignment sub-model is based on a Dirichlet Process Mixture (DPM) model, a non-parametric model used for probabilistic clustering of mixture profiles ([Molitor et al., 2010](#)). This sub-model estimates the optimal number of clusters and calculates the probability that an individual belongs to each cluster. The exposure sub-model determines the exposure profile of each cluster ([Belloni et al., 2020](#)). The eight exposure variables, represent the average of the mean annual exposure levels of BaP, cadmium, dioxins, PCB153,  $\text{NO}_2$ ,  $\text{O}_3$ ,  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  respectively during the follow-up period of the woman, and were continuous.

The disease sub-model links clusters of profiles to the outcome, by estimating the risk of BC within each cluster. An unconditional logistic regression model adjusted for confounders was therefore fitted. The binary indicator  $Y_i$  specified the woman's  $i$  status (1 for cases and 0 for controls). This model included specific parameters for each cluster and a vector of parameters shared by all clusters. The assignment model is written by specifying the respective contributions of the exposure and disease sub-models.

The number of clusters and the construction of the mixing weights  $\psi_c$  were defined by a "stick-breaking" process, with the maximum number of clusters  $C_{max}$  specified ([Molitor et al., 2010](#)). From the elements of the likelihood, the posterior probabilities of a given individual belonging to each cluster was calculated using Bayes theorem. According to this, an individual was assigned to the cluster with the highest probability. This was repeated for each individual.

The parameters of the BPR model were estimated by a Bayesian approach using a Markov chain Monte Carlo (MCMC) algorithm, with 30,000 iterations after a burn-in sample of 10,000 iterations. Every iteration of the MCMC algorithm assigned each individual to a cluster therefore partitioning the study population. As the MCMC algorithm finds it harder to split clusters than to merge them, and given the large population of the present study (including 10,444 participants), the initial number of clusters was set to a large number, namely 2,500. During the MCMC simulation, the number of clusters could vary and the label associated with each cluster changed during the iterations, thus postprocessing of the MCMC output was carried out to obtain an optimal partition. First, a score matrix of  $(n \times n)$  ( $n$  being the total number of individuals) size, was generated at each iteration, to determine whether two individuals belonged to the same cluster or not. Second, a probability matrix of  $(n \times n)$  size was created, to indicate the probability that 2 individuals belonged to the same cluster, by averaging each matrix score. Lastly, based on this probability matrix, an optimal partition was obtained by maximising the average silhouette width across various partitions obtained by partitioning Around Medoids (PAM) ([Molitor et al., 2010](#)).

Posterior distribution of cluster parameters was obtained from MCMC output. For each iteration of the MCMC sampler, the risk associated with a given cluster  $k$  is the average of the risk obtained, for all individuals included in the cluster  $k$ . Repeating this for all iterations provides a distribution for the risk for the optimal partition, applied for all cluster parameters ([Molitor et al., 2010](#)).

The MCMC algorithm was run five times with different initial positions randomly assigned and the marginal posterior distributions of the  $\alpha$  parameter of DPM were compared between the five runs to assess convergence. The final partition selected was the one maximising the average silhouette clustering score among the optimal partitions obtained by post-processing ([Lengyel and Botta-Dukát, 2019; Liverani et al., 2015](#)).



Confounders were identified using a directed acyclic graph (Supplementary figure S2). Two minimal sufficient adjustment sets of variables have been proposed. The first includes physical activity (<25.3, 25.3–35.5, 35.6–51.8, and  $\geq 51.8$  METs-h/week), alcohol intake (0, 1–6.7, > 6.7 g/day), body mass index (BMI) (<25, 25–<30, and  $\geq 30$  kg/m<sup>2</sup>), breastfeeding (ever, never), oral contraceptive use (ever, never), menopausal hormone treatment (ever, never), status of birthplace (rural, urban), status of inclusion (rural, urban), smoking status (never, current, and former), parity and age at first full-term pregnancy (AFP) (0, 1–2 children & AFP<30 years, 1–2 children & AFP $\geq$ 30 years, and 3 children), and mammography screening (yes, no). The second includes the level of education, considered as a proxy of socio-economic status. The model with more adjustment variables was selected for the analysis. Confounding variables with less than 5 % missing data were imputed using the median for continuous variables and the mode for categorical variables, based on control population data. Covariates with more than 5 % missing values, specifically alcohol consumption and birthplace, were treated as a separate category (Amadou et al., 2020a).

Analyses were run in R (version 4.0.2; R Core Team 2020). The R PRemium package (Liverani et al., 2015) designed to fit the BPR model includes an unconditional logistic regression disease sub-model but not the conditional regression model more suitable for the matched case-control design of the present study. In consequence, we fitted the unconditional logistic regression model, adjusted on the combination of matching variables (including department of residence, age ( $\pm 1$  year), date ( $\pm 3$  months), menopausal status at blood collection (menopausal or not) and availability of a saliva sample (yes, no), in addition to confounders as proposed by Pearce (Pearce, 2016). To fit mixtures and potentially link covariates with responses, we identified covariates that actively influence the mixture components and those that exhibit characteristics common to all components through variable selection available in PRemium package. Odds ratios (ORs) of BC risk associated with combined exposure to the 8 pollutants in each cluster, and their 95 % credible intervals (CrIs) were estimated (Coker et al., 2020). The average exposure profiles of each cluster also assessed and presented with boxplot and heat map. The geocoded residential locations of all women were graphically displayed by cluster.

Sensitivity analyses were further conducted by defining exposure profiles solely based on the exposure variables, without considering the risk provided by the disease sub-model in the profile regression, using an option of the BPR model in the R PRemium package that excludes the outcome variable from the profile regression model. To assess the model's convergence, we followed the same approach as in the main analyses. Associations between these exposure profiles and BC risk were analysed using unconditional logistic regression models and adjusted on the matching variables and previously defined confounders.

### 3. Results

#### 3.1. Characteristics of the study population

General characteristics of the study population are presented in Supplementary Table S1. Overall, women were included on average at age 49.6 (6.3) years, 54.5% were ever smokers, 62.1 % were born in urban areas, 37 % were in the highest tertile level of alcohol intake (>6.7 g/day), 83 % had normal BMI (<25 kg/m<sup>2</sup>). Fig. 1 shows the Pearson correlation matrix between the average of the annual mean concentrations (1990–2011) for each of the eight pollutants. Overall, moderate to strong positive correlations were observed between the majority of the pollutants. The strongest correlations were observed between PM<sub>2.5</sub> and PM<sub>10</sub> (0.99), between PCB153 and PM (0.93 and 0.94, for PM<sub>10</sub> and PM<sub>2.5</sub>, respectively), between NO<sub>2</sub> and PM (0.87 and 0.86 for PM<sub>10</sub> and PM<sub>2.5</sub>, respectively) and between PCB153 and NO<sub>2</sub> (0.81). In contrast, O<sub>3</sub> was negatively correlated with all other pollutants, with a strong negative correlation for NO<sub>2</sub> (−0.92), PM<sub>10</sub> (−0.90), PM<sub>2.5</sub> (−0.89), and PCB153 (−0.83).

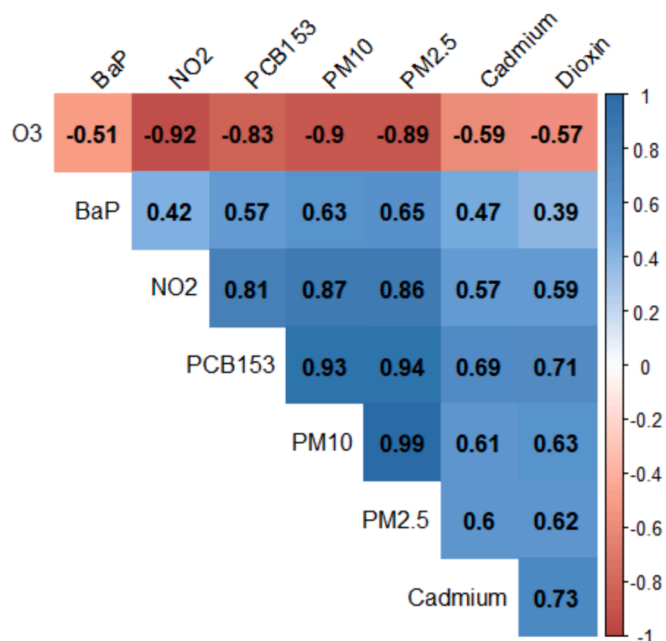


Fig. 1. Pearson's correlation matrix between the average of the annual mean concentrations (1990–2011) for each of the eight pollutants (BaP, dioxins, cadmium, PCB153, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>) in the case-control study nested within the E3N-generation cohort, France, 1990–2011.

#### 3.2. Association between exposure profiles and BC risk

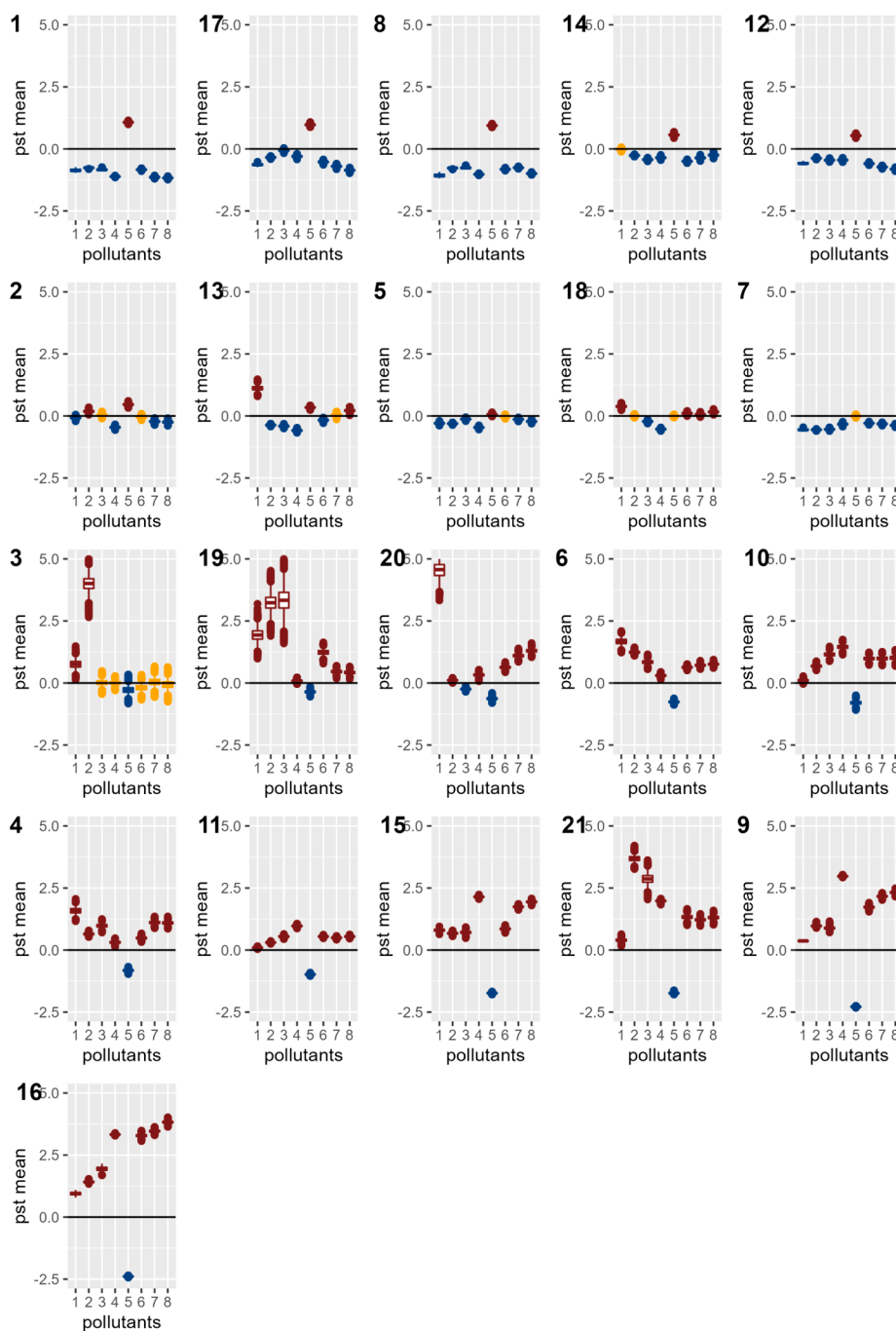
The marginal posterior distributions of the  $\alpha$  parameter (Markov chains), shown in Supplementary Figure S3, overlapped among the five runs of the MCMC algorithm and showed the model convergence. Supplementary Table S2 presents the number of clusters for each partition, ranging from 18 to 26 clusters, and the associated average silhouette score. Partition 2, with 21 clusters, maximising the silhouette average score (0.866), was selected as the final partition. Visual comparison of the five post-processed partitions with the highest silhouette average scores (0.79–0.87), showed similar exposure profiles, supporting the convergence of the model (data not shown).

During the iterations of the MCMC algorithm, the selection probability for each pollutant remained consistently above 0.8 (Supplementary Figure S4).

For a visual overview of cluster exposure profiles, Fig. 2 shows the normalised posterior means of each of the eight pollutants, for the 21 clusters ordered by decreasing levels of O<sub>3</sub> exposure. The Supplementary Figure S5 shows a heat map of the posterior clustering. Average concentration estimates of the pollutants for the 21 resulting clusters are shown in Table 1. The number of subjects in each cluster ranged from 52 (cluster 3) to 1,223 (cluster 11). Table 2 shows the ORs and corresponding 95 % credible intervals for each cluster.

Cluster 1 (1141 women), characterised by low exposure to all pollutants except O<sub>3</sub> was selected as the reference cluster (Fig. 2). Graphical displays of women residential locations by cluster, shows that women in the reference cluster were scattered over a wide area of France (Fig. 3).

The three clusters with the highest ORs were characterised by quite similar pollutant patterns with varying levels of exposures (Table 1, Table 2, Fig. 2 and Supplementary Figure S6): cluster 9 (265 participants; OR=1.61; CrI=1.13,2.26), cluster 16 (210 participants; OR=1.59; CrI=1.10,2.30) and cluster 15 (357 participants; OR=1.38; CrI=1.00,1.88). Cluster 9 corresponded to women with very high NO<sub>2</sub> exposure levels and high exposure to PCB153, PM<sub>10</sub> and PM<sub>2.5</sub>. However, exposure to BaP, cadmium and dioxins were lower than the other pollutants and exposure to O<sub>3</sub> was extremely low. Cluster 16 represented women with very high exposure levels for NO<sub>2</sub>, PCB153, PM<sub>10</sub> and



**Fig. 2.** Characterization of the exposure profiles associated to each cluster using BPR, in the case-control study nested within the E3N-generation cohort, France, 1990–2011. Clusters were numbered from 1 to 21 and ordered by decreasing levels of  $O_3$  exposure. For each cluster, the boxplot of the normalised means distribution of each pollutant were represented from left to right in the following order: BaP (1), cadmium (2), dioxin (3),  $NO_2$  (4),  $O_3$  (5), PCB153 (6),  $PM_{10}$  (7),  $PM_{2.5}$  (8). These distributions came from MCMC algorithm of BPR model. The two different colours, blue and red, correspond to a 95 % credible interval, respectively under or upper 0 which represents low or high exposure level respectively. Orange colour shows no evidence of high or low values of exposures. Pst mean: normalised posterior mean of the exposures within cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$PM_{2.5}$ , lower exposure to BaP, cadmium and dioxins and low exposure to  $O_3$ . Cluster 15 was characterised by high exposure levels for  $NO_2$  and PMs, lower exposures to BaP, cadmium, dioxins and PCB153, and low exposure to  $O_3$  (Table 1, Table 2 and Fig. 2). Women in these three clusters were mainly located in urban metropolitan areas of Paris and Lyon (Fig. 3). Two other clusters with less elevated ORs and CrI limits showed close profiles with lower exposure levels: cluster 11 (1,223 participants; OR=1.21; CrI=0.95,1.55) and cluster 10 (285 participants; OR=1.22; CrI=0.91,1.64).

Seven other clusters essentially characterised by higher exposure levels for cadmium, dioxins and/or BaP, were not associated with BC risk (Table 1, Fig. 2). Of these, six clusters had ORs greater than 1: cluster 3 (52 participants; OR=1.08; CrI=0.64,1.83), cluster 4 (249 participants; OR=1.08; CrI=0.75,1.53), cluster 6 (361 participants; OR=1.11; CrI=0.81,1.52), cluster 19 (148 participants; OR=1.26; CrI=0.87,1.80), cluster 20 (158 participants; OR=1.22; CrI=0.66,2.28) and cluster 21 (54 participants; OR=1.26; CrI=0.72,2.20). One cluster, cluster 13 was associated with an OR below 1 (321 participants; OR=0.74;

**Table 1**

Characterization of the exposure profiles associated to each cluster when fitting a BPR model, in the case-control study nested within the E3N-generation cohort, France, 1990–2011.

Cluster Label	Bap*	Cadmium*	Dioxin*	PCB153*	NO2*	O3*	PM2.5*	PM10*
1	0.09 (0.09,0.10)	71.88 (67.98,73.11)	2.83 (2.51,3.45)	6.40 (6.31,6.49)	2.83 (2.70,2.96)	32.44 (32.26,32.61)	12.00 (11.84,12.16)	14.47 (14.32,14.62)
2	0.18 (0.18,0.19)	178.05 (172.33,185.09)	10.23 (9.74,10.73)	9.65 (9.43,9.86)	8.91 (8.50,9.34)	28.68 (28.33,29.02)	17.48 (17.12,17.85)	20.13 (19.78,20.48)
3	0.28 (0.25,0.31)	608.34 (548.34,660.23)	10.09 (8.57,11.60)	9.06 (8.27,9.84)	13.21 (12.15,14.16)	24.10 (22.82,25.54)	18.58 (16.99,20.09)	21.78 (20.31,23.19)
4	0.37 (0.35,0.39)	230.08 (224.20,236.25)	18.27 (17.25,19.21)	11.80 (11.52,12.07)	16.00 (15.34,16.65)	20.79 (20.42,21.17)	25.42 (24.86,25.98)	28.35 (27.80,28.91)
5	0.16 (0.15,0.16)	122.72 (120.35,124.87)	8.87 (8.61,9.12)	9.71 (9.56,9.85)	8.85 (8.57,9.14)	26.22 (26.02,26.42)	17.64 (17.44,17.85)	20.65 (20.46,20.84)
6	0.38 (0.36,0.40)	296.52 (289.68,305.75)	17.11 (16.11,18.10)	12.40 (12.20,12.61)	15.90 (15.39,16.42)	21.15 (20.87,21.44)	23.43 (23.07,23.80)	25.91 (25.54,26.29)
7	0.13 (0.12,0.14)	95.12 (92.25,96.61)	5.38 (5.12,5.79)	8.63 (8.52,8.73)	10.07 (9.72,10.42)	25.81 (25.64,25.99)	16.70 (16.54,16.86)	19.54 (19.38,19.69)
8	0.07 (0.06,0.08)	70.76 (66.83,72.14)	3.46 (3.15, 4.02)	6.46 (6.36,6.56)	3.67 (3.53,3.82)	31.62 (31.43,31.81)	13.06 (12.93,13.20)	16.84 (16.72,16.96)
9	0.23 (0.23,0.24)	266.46 (262.44,272.62)	17.45 (16.74,18.06)	16.89 (16.62,17.17)	40.58 (40.40,40.78)	11.79 (11.70,11.88)	32.70 (32.33,33.08)	34.86 (34.52,35.20)
10	0.20 (0.20,0.21)	233.93 (225.73,242.40)	19.72 (18.65,20.72)	13.85 (13.41,14.28)	26.55 (25.40,27.71)	20.95 (20.21,21.67)	24.96 (24.15,25.79)	27.61 (26.87,28.36)
11	0.20 (0.20,0.21)	191.42 (189.15,193.90)	14.57 (14.13,14.92)	12.04 (11.93,12.16)	22.08 (21.73,22.42)	19.81 (19.65,19.97)	22.17 (21.97,22.36)	24.56 (24.37,24.74)
12	0.13 (0.12,0.13)	115.25 (112.75,117.47)	6.10 (5.83,6.47)	7.39 (7.29,7.49)	8.99 (8.67,9.31)	29.14 (28.90,29.37)	14.11 (13.93,14.29)	16.99 (16.82,17.16)
13	0.32 (0.30,0.34)	115.97 (113.32,118.51)	6.48 (6.15,6.84)	9.12 (8.94,9.31)	7.71 (7.36,8.09)	27.94 (27.70,28.15)	20.23 (19.85,20.61)	21.73 (21.37,22.09)
14	0.19 (0.19,0.19)	127.83 (125.60,129.94)	6.30 (6.04,6.64)	7.78 (7.65,7.91)	9.86 (9.50,10.23)	29.30 (29.00,29.61)	17.46 (17.13,17.79)	19.29 (18.97,19.61)
15	0.28 (0.27,0.29)	233.15 (228.11,238.57)	15.98 (15.21,16.72)	13.30 (13.04,13.57)	32.93 (32.61,33.25)	15.18 (15.04,15.31)	30.44 (30.11,30.78)	32.26 (31.96,32.55)
16	0.30 (0.29,0.31)	316.04 (312.01,323.39)	26.53 (25.08,27.38)	23.23 (22.88,23.59)	43.85 (43.66,44.03)	11.12 (11.02,11.23)	41.59 (41.12,42.05)	42.87 (42.45,43.29)
17	0.12 (0.12,0.13)	118.94 (116.12,121.50)	9.41 (9.09,9.72)	7.67 (7.50,7.84)	10.34 (9.97,10.72)	31.82 (31.57,32.07)	13.85 (13.60,14.11)	17.20 (16.93,17.47)
18	0.24 (0.23,0.24)	157.88 (154.35,161.46)	8.08 (7.80,8.36)	10.23 (10.06,10.39)	8.20 (7.95,8.46)	25.84 (25.68,26.00)	19.91 (19.64,20.17)	21.89 (21.64,22.14)
19	0.41 (0.37,0.46)	520.90 (462.90,580.59)	38.17 (31.75,44.88)	14.87 (14.18,15.56)	13.91 (13.39,14.43)	23.64 (23.16,24.11)	21.47 (20.88,22.06)	24.35 (23.77,24.94)
20	0.72 (0.65,0.78)	170.05 (165.58,174.05)	7.89 (7.50,8.27)	12.40 (12.09,12.71)	16.14 (15.28,16.91)	21.97 (21.53,22.48)	26.65 (26.00,27.29)	28.31 (27.67,28.93)
21	0.24 (0.23,0.25)	571.19 (551.52,594.23)	34.24 (31.47,36.84)	15.26 (14.75, 15.7)	31.40 (31.07,31.72)	15.16 (15.04,15.29)	26.71 (26.06,27.33)	29.03 (28.47,29.58)
Overall	0.19	157.32	9.96	9.82	13.12	25.84	18.94	21.51

BaP was expressed in ng/m<sup>3</sup>, PCB153 in pg/m<sup>3</sup>, cadmium in pg/m<sup>3</sup>, dioxin in fg-TEQ/m<sup>3</sup>, PM10 in μm<sup>3</sup>, PM2.5 in μg/m<sup>3</sup>, NO2 in μg/m<sup>3</sup>, O3 in pp/m<sup>3</sup>.

\* : estimated mean (95 %CrI).

CrI=0.49,1.17) (Table 2).

Four further clusters associated with exposures relatively close to the overall average for each pollutant were found (Table 1, Fig. 2): cluster 5 (656 participants; OR=1.13; CrI=0.85,1.48), cluster 18 (562 participants; OR=1.15; CrI=0.87,1.54), cluster 7 (792 participants; OR=1.23; CrI=0.97,1.56) and cluster 2 (529 participants; OR=1.20; CrI=0.97,1.49) (Table 2).

The remaining four clusters exhibited high O<sub>3</sub> levels and low exposures to other pollutants (Table 1, Fig. 2): cluster 17 (662 participants; OR=1.24; CrI=0.93,1.66), cluster 8 (767 participants; OR=1.10; CrI=0.84,1.45), cluster 14 (842 participants; OR=1.03; CrI=0.79,1.33); and cluster 12 (810 participants; OR=0.96; CrI=0.77,1.18) (Table 2).

### 3.3. Sensitivity analyses

The results of the sensitivity analyses of pollutant exposure profiles, without considering the risk, are shown in Supplementary Figures S7 and S8, and, Supplementary Tables S3, S4 and S5. The marginal posterior distributions of the parameter (Markov chains) overlapped among the five runs of the MCMC algorithm and showed the model convergence (Supplementary Figure S7). Partition 3, with 18 clusters, maximising the silhouette average score (0.863), was selected as the final partition

(Supplementary Table S3). Cluster S1 (1,085 women), characterised by low exposure to all pollutants except O<sub>3</sub> was selected as the reference cluster (Supplementary Figure S8). The three clusters with the highest ORs of BC risk were cluster S7 (88 participants; OR=1.99; CI=1.15,3.46), cluster S9 (290 participants; OR=1.96; CI=1.29,2.98) and cluster S12 (1,433 participants; OR=1.36; CI=1.03,1.80) (Supplementary Table S4). Cluster S7 and S9 were characterised by quite similar pollutant patterns with women exposed to very high levels of NO<sub>2</sub>, PCB153, PM<sub>10</sub> and PM<sub>2.5</sub> (Supplementary Figure S8). Clusters S7 and S9 displayed similar pollutant profiles than clusters 9, 15 and 16 in the main analyses, but with slightly higher ORs. In contrast, cluster S12 corresponded to women considerably exposed to NO<sub>2</sub> and moderately or slightly exposed to all other pollutants (Supplementary Figure S8). Characteristics of women in cluster S12 did not substantially differ from the overall population (Supplementary Figure S5).

## 4. Discussion

To the best of our knowledge, this is the first study assessing the association between the combined exposure to eight highly correlated air pollutants and BC risk using the BPR model.

Overall, the present study demonstrated a positive association

**Table 2**

Association between BC risk and clusters, using BPRmodel, in the case-control study nested within the E3N-generation cohort, France, 1990–2011.

Clusters Label	Cases/controls	OR (95 % CrI)
1	559/582	Reference
2	282/247	1.20(0.97,1.49)
3	26/26	1.08(0.64,1.83)
4	123/126	1.08 (0.75,1.53)
5	329/327	1.13 (0.85,1.48)
6	174/187	1.11 (0.81,1.52)
7	410/382	1.23 (0.97,1.56)
8	377/390	1.10 (0.84,1.45)
9	151/114	1.61 (1.13,2.26)
10	138/147	1.22 (0.91,1.64)
11	595/628	1.21 (0.95,1.55)
12	384/426	0.96 (0.77,1.18)
13	152/169	0.74 (0.49,1.17)
14	410/432	1.03 (0.79,1.33)
15	186/171	1.38 (1.00,1.88)
16	121/89	1.59 (1.10,2.30)
17	337/325	1.24 (0.93,1.66)
18	287/275	1.15 (0.87,1.54)
19	78/70	1.26 (0.87,1.80)
20	80/78	1.22 (0.66,2.28)
21	23/31	1.26 (0.72,2.20)

BPR model were adjusted for physical activity, alcohol intake, body mass index, breast-feeding, oral contraceptive use, menopausal hormone treatment, status of birthplace, status of inclusion smoking status, parity and age at first full-term pregnancy, mammography screening, department of residence, menopausal status and availability of a saliva sample.

OR: Odds Ratio; CrI: credible interval.

between BC risk and exposure to combinations of the eight pollutants, with significant evidence only for three clusters, characterised mainly by high levels of NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, and PCB153. Women in these three clusters (9, 15 and 16) were mainly located in high-density urban environments of Paris and Lyon. Two other clusters (10 and 11) with close exposure profiles but lower exposure levels, likewise predominantly located in high-density urban areas, were associated with increased ORs of lower magnitude. Moreover, several clusters (3, 4, 6, 19, 20, and 21) driven by high exposure levels to cadmium, dioxins, and / or BaP, were associated with positive ORs, although not consistent. Women in these clusters were mainly located in eastern France, near the border with Germany and Belgium (cluster 4, 6 and 20), and potentially subject to cross-border pollution, or located in highly industrialised areas known to be associated with industrial emissions (clusters 3, 19 and 21) (Jeanjean et al., 2021).

Most studies have investigated the effects of single air pollutants on BC risk, with classical and well-known methods such as multiple linear or logistic regression (Niehoff et al., 2022; White et al., 2021, 2018). The results observed for the three clusters associated with highest risk are consistent with results obtained from single pollutant analyses. Increased risks of BC have been constantly found in previous studies for NO<sub>2</sub> (Gabet et al., 2021; Praud et al., 2023), including a previous study in the same nested case-control study with OR=1.15 (95 % CI= 1.06,1.26) (Amadou et al., 2023). Likewise, there was a statistically significant positive association between cumulative atmospheric exposure to PCB153 and BC risk (OR=1.19; 95 % CI= 1.08,1.31) (Deygas et al., 2021). Yet, divergent results were reported between PM (PM<sub>2.5</sub> and PM<sub>10</sub>) and BC risk (Gabet et al., 2021; Wei et al., 2021; Zhang et al., 2019). However, the ORs for clusters of combined exposures (cluster 9 (OR=1.61; CrI=1.13,2.26), cluster 16 (OR=1.59; CrI=1.10,2.30) and cluster 15 (OR=1.38; CrI=1.00,1.88)) were superior to the ORs observed in the single pollutant studies.

Our study has revealed that exposure profiles with high exposure to cadmium, dioxin, and BaP, showed no consistent increase of BC risk. Similar to these findings, single epidemiological analyses available on the association between cadmium and dioxins and risk of BC showed divergent results (Amadou et al., 2020b; Danjou et al., 2019; Filippini

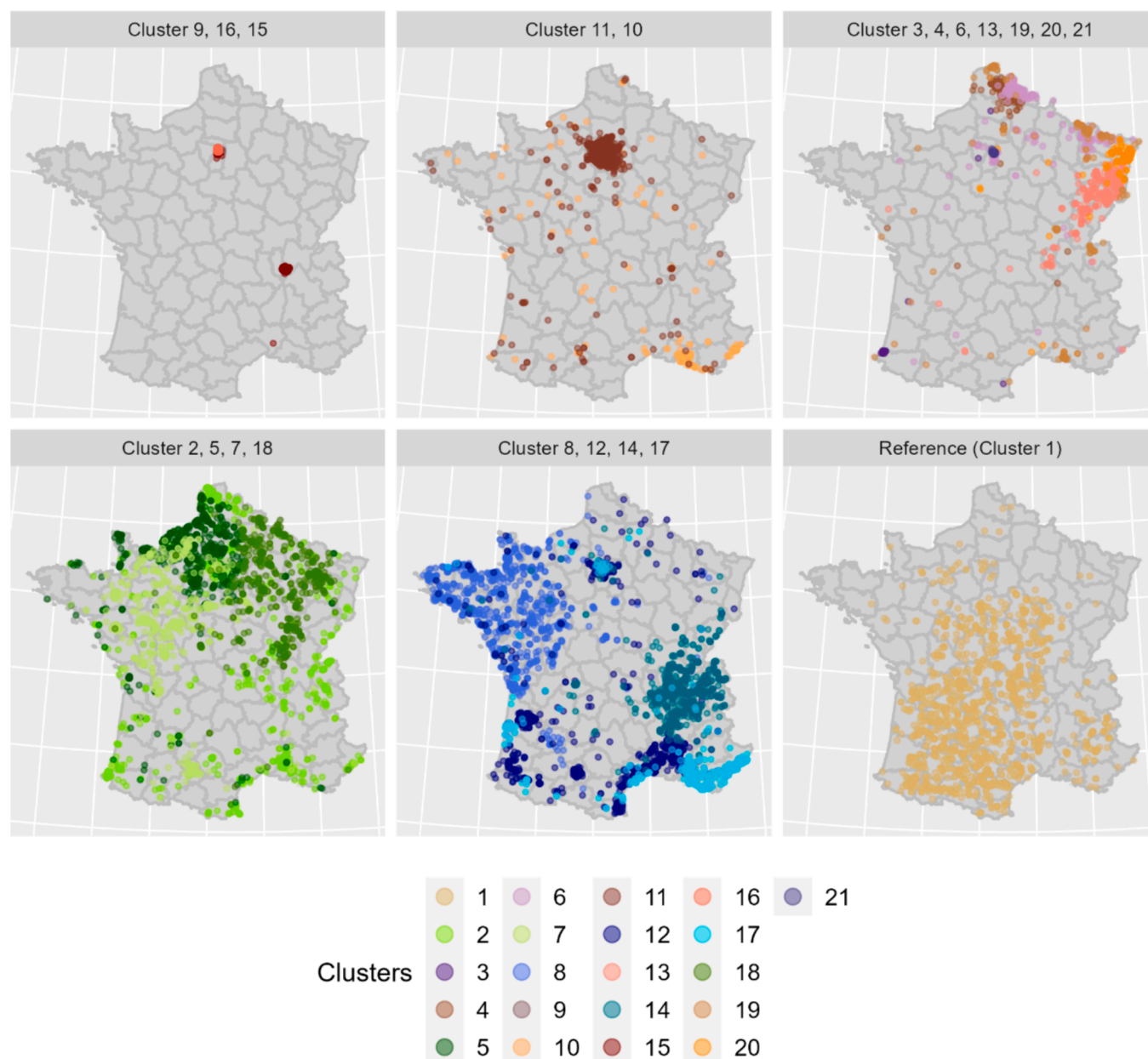
et al., 2020; Fiolet et al., 2022; Florez-Garcia et al., 2023; VoPham et al., 2020; Xu et al., 2016). A recent meta-analysis showed that exposure to higher levels of cadmium (dietary, airborne, and biomarker measures) was associated with an increased risk of BC (Florez-Garcia et al., 2023), while our previous study found no evidence of an association between single airborne exposure to cadmium and overall risk of BC (Amadou et al., 2020b). Regarding dioxins, a study by Danjou et al. showed no increased risk of BC in relation to higher dioxin exposure levels (Danjou et al., 2019). In contrast, evidence of increased risk of BC associated with exposure to BaP was reported in the present nested-case control study by Amadou et al. (OR=1.15; 95 %CI= 1.04,1.27) (Amadou et al., 2021).

Of note, one prospective cohort study has investigated the combined effect of multiple metallic air pollutants exposures (including cadmium) on the risk of BC, using the weighted quantile sum regression (WQS) method. This study showed a statistically significant association between the WQS index (combined exposure to ten air toxic metallics) and postmenopausal BC (OR=1.06; 95 %CI= 1.00–1.13) (White et al., 2019). Regarding air pollutant exposure profile approaches, our findings are notably consistent with the study by Niehoff et al., in which classification and regression tree method has been used to identify patterns and combinations of 29 air toxics related to BC. This exploratory analysis identified combinations of age, methylene chloride, BMI and four other toxic substances (propylene dichloride, ethylene dibromide, ethylene dichloride, styrene) associated with overall BC risk, highlighting the complex association between these pollutants and BC risk (Niehoff et al., 2019). However, no study has employed the BPR model to estimate the joint effect of long-term exposure to several air pollutants on the risk of BC in the context of correlated exposures.

The BPR model was selected for the present study, as it considers the risk of BC in addition to the level of exposure, as well as its interest to jointly analyse positively correlated exposures, and also negatively correlated ones (O<sub>3</sub>). In this study, the MCMC algorithm converged to 21 clusters ( $\pm 3$  clusters depending on the initial positions of the Markov chains), a number slightly higher than those obtained in most previous studies, which often had between five and 15 clusters (Coker et al., 2020, 2023; Hoover et al., 2023; Mattei et al., 2016; Rouanet et al., 2021). For example, a study assessing the association between metal mixtures and Preterm birth among pregnant Indigenous women from the Navajo Birth Cohort Study, based on 417 individuals, has estimated six clusters (Hoover et al., 2023). Another study investigating the effect of occupational exposures to organic solvents on lung cancer risk, based on 5,012 individuals, identified 13 exposure profiles (Mattei et al., 2016). Of note, the observed difference in the number of clusters can be explained by the size of our study population, including 10,444 women, exceeding that of most published studies with between 551 and 5,012 subjects (Coker et al., 2023; Hoover et al., 2023; Mattei et al., 2016). It has been shown that the MCMC algorithm of a DPM tends to converge towards a higher number of clusters with increasing population size (Chaumeny et al., 2022). It should be highlighted that the MCMC algorithm with the PreMiuM R package has a superior performance to merge rather than to split clusters (Liverani et al., 2015). It is then important that the number of clusters selected at the initiation of the MCMC algorithm is substantially higher than the true number of clusters, possibly in the dataset. If the number of clusters at MCMC initiation is too low, the final number of clusters may be overestimated by a DPM as the algorithm might be unable to converge to the optimal number of clusters (Hastie et al., 2015). In response to this, the initial number of clusters at initialization of the MCMC algorithm in the present study was fixed at 2,500. We observed a similar number of final clusters produced by the five runs of the MCMC chain with convergence of the model as shown by the marginal posterior distributions of the  $\alpha$  parameters.

The sensitivity analyses, ignoring the risk provided by the disease sub-model in the clustering process, yielded similar average exposure profiles than the full BPR model, with slightly higher ORs for the two clusters (S7 and S9) characterised by high levels of NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, and PCB153, than clusters 9, 15 and 16 in the main analysis. However,





**Fig. 3.** Geographic location of women in France, based on their addresses. Women's residential address coloured by cluster at all addresses. "Cluster 9, 16, 15": associated with women at highest risk; "Cluster 10, 11": characterised by high exposure to NO<sub>2</sub>, PCB153, PM<sub>2.5</sub> or PM<sub>10</sub>; "Cluster 3, 4, 6, 13, 19, 20, 21": characterised by higher exposure levels for cadmium, dioxins and/or BaP; Cluster 2, 5, 7, 18: characterised by exposures relatively close to the overall average for each pollutant; "Cluster 8, 12, 14, 17": characterised by high O<sub>3</sub> levels and low exposures to other pollutants.

as a difference we observed a profile significantly associated with BC risk (cluster 12) consisting of women with low exposure to most pollutants. This profile was not observed with the full BPR model. Grouping individuals together based on their similarity in terms of both, the exposure profile and probability of outcome, the full BPR model may achieve clustering of more homogenous groups of individuals.

One of the strengths of the present study is the use of the BPR method to estimate the risk of BC in the presence of multiple correlated air pollutants, with concentration levels of the eight pollutants estimated over a long period (up to 22 years). With the use of BPR, the limitations of unstable estimates encountered in traditional methods such as multiple regression in presence of multiple correlated predictors are circumvented. This method could be applied to other multiple exposures and diseases risks. Another strength of this study is the design of the nested case-control study within the large prospective French E3N-

generation cohort providing the availability of information on a wide range of potential confounders. The low percentages of missing data reflect the high quality of the data.

Despite these strengths, it is important to note some limitations. The non-implementation of the likelihood of a conditional logistic regression model in the R premium package, as a disease sub-model, required thorough consideration of applying the BPR model to the matched case-control study. Further adjusting the models for matching variables, allowed us to get correct estimates comparable to that obtained with a conditioned model. Another limitation is that the main pollutants responsible for the association with BC have not been identified. This method is unable to provide any evidence of interactions throughout the mixture. Moreover, BPR tends to estimate a greater number of clusters when dealing with a larger population (Chaumeny et al., 2022), of note several clusters have comparable ORs and exposure profiles, suggesting

that there is redundancy in the partitioning of the study population.

## 5. Conclusion

Overall, this study supports evidence of a positive association between a mixture of air pollutant exposures and risk of BC, with greater risk observed for clusters with high concentrations levels of NO<sub>2</sub>, PMs and PCB153. BPR was an appropriate model for the present research, considering women's levels of risk and air pollutant exposure profiles. Overall, this is the first time that BPR was applied to estimate the joint effects of exposure to eight air pollutants on the risk of BC. This study suggests the importance of considering combinations of air pollutants when estimating BC risk, in order to identify specific subgroups at high risk, to inform public health policies and development of preventive strategies.

## 6. Research ethics and patient consent ethical approval

Our research is based on the existing French national cohort E3N. Informed consent was obtained from all participants and the study was approved by the French National Commission for Data Protection and Privacy (CNIL). The datasets generated and/or analysed during the current study are not publicly available for ethical reasons and in accordance with the specific consent signed by the participants. Data are available from the E3N-Generations team through the corresponding author on reasonable request.

## Funding

This work was supported by ADEME (21ESD0020), the Regional Committee of the French League against Cancer of the Savoie Region (18-383-C). CG is supported by a doctoral fellowship of the Regional Committee of the French League Against Cancer of Rhône. The research was carried out using data from the Inserm (French National Institute for Health and Medical Research) E3N-Generation cohort, which was established and maintained with the support of the Mutuelle Générale de l'Éducation Nationale (MGEN), Gustave Roussy, and the French League against Cancer (LNCC). The E3N-Generation cohort is also supported by the French National Research Agency (ANR) under the Investment for the future Program (PIA; ANR-10-COHO-0006) and by the French Ministry of Higher Education, Research and Innovation (subsidy for public service charges No. 2102 918823, 2103236497, and 2103586016).

## CRedit authorship contribution statement

**Camille Giampiccolo:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Amina Amadou:** Supervision, Resources, Project administration, Conceptualization. **Thomas Coudon:** Writing – review & editing, Supervision, Resources, Conceptualization. **Delphine Praud:** Writing – review & editing, Data curation, Conceptualization. **Lény Grassot:** Writing – review & editing, Resources, Data curation. **Elodie Faure:** Writing – review & editing, Resources, Data curation. **Florian Couvidat:** Writing – review & editing, Resources, Data curation. **Gianluca Severi:** Writing – review & editing, Resources, Data curation, Conceptualization. **Francesca Romana Mancini:** Writing – review & editing, Resources, Data curation. **Béatrice Fervers:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Pascal Roy:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

We sincerely thank John Molitor, Silvia Liverani, Muriel Rabilloud and Sophie Ancelet for their advice on the application of the BPR model. The authors are grateful to the study participants for their continued participation and to medical practitioners for providing pathology reports.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envint.2024.108943>.

## References

- Agier, L., Portengen, L., Chadeau-Hyam, M., Basagaña, X., Giorgis-Allemand, L., Siroux, V., Robinson, O., Vlaanderen, J., González, J.R., Nieuwenhuijsen, M.J., Vineis, P., Vrijheid, M., Slama, R., Vermeulen, R., 2016. A systematic comparison of linear regression-based statistical methods to assess exposure-health associations. *Environ. Health Perspect.* 124, 1848–1856. <https://doi.org/10.1289/EHP172>.
- Amadou, A., Coudon, T., Praud, D., Salizzoni, P., Leffondré, K., Lévêque, E., Boutron-Ruault, M.-C., Danjou, A.M.N., Morelli, X., Le Cornet, C., Perrier, L., Couvidat, F., Bessagnet, B., Caudeville, J., Faure, E., Mancini, F.R., Gulliver, J., Severi, G., Fervers, B., 2020a. Chronic low-dose exposure to xenoestrogen ambient air pollutants and breast cancer risk: XENAIR protocol for a case-control study nested within the French E3N Cohort. *JMIR Res. Protoc.* 9, e15167 <https://doi.org/10.2196/15167>.
- Amadou, A., Praud, D., Coudon, T., Danjou, A.M.N., Faure, E., Leffondré, K., Le Romancer, M., Severi, G., Salizzoni, P., Mancini, F.R., Fervers, B., 2020b. Chronic long-term exposure to cadmium air pollution and breast cancer risk in the French E3N cohort. *Int. J. Cancer* 146, 341–351. <https://doi.org/10.1002/ijc.32257>.
- Amadou, A., Praud, D., Coudon, T., Deygas, F., Grassot, L., Faure, E., Couvidat, F., Caudeville, J., Bessagnet, B., Salizzoni, P., Gulliver, J., Leffondré, K., Severi, G., Mancini, F.R., Fervers, B., 2021. Risk of breast cancer associated with long-term exposure to benzo[a]pyrene (BaP) air pollution: Evidence from the French E3N cohort study. *Environ. Int.* 149, 106399 <https://doi.org/10.1016/j.envint.2021.106399>.
- Amadou, A., Praud, D., Coudon, T., Deygas, F., Grassot, L., Dubuis, M., Faure, E., Couvidat, F., Caudeville, J., Bessagnet, B., Salizzoni, P., Leffondré, K., Gulliver, J., Severi, G., Mancini, F.R., Fervers, B., 2023. Long-term exposure to nitrogen dioxide air pollution and breast cancer risk: A nested case-control within the French E3N cohort study. *Environ. Pollut.* 317, 120719 <https://doi.org/10.1016/j.envpol.2022.120719>.
- Bai, L., Shin, S., Burnett, R.T., Kwong, J.C., Hystad, P., van Donkelaar, A., Goldberg, M. S., Lavigne, E., Weichenthal, S., Martin, R.V., Copes, R., Kopp, A., Chen, H., 2020. Exposure to ambient air pollution and the incidence of lung cancer and breast cancer in the Ontario Population Health and Environment Cohort. *Int. J. Cancer* 146, 2450–2459. <https://doi.org/10.1002/ijc.32575>.
- Bellavia, A., 2023. 1.1 Environmental mixtures | Statistical Methods for Environmental Mixtures.
- Belloni, M., Laurent, O., Guihenneuc, C., Ancelet, S., 2020. Bayesian Profile Regression to Deal With Multiple Highly Correlated Exposures and a Censored Survival Outcome. First Application in Ionizing Radiation Epidemiology. *Frontiers in Public Health* 8.
- Billionnet, C., Sherrill, D., Annesi-Maesano, I., GERIE study, 2012. Estimating the health effects of exposure to multi-pollutant mixture. *Ann Epidemiol* 22, 126–141. Doi: 10.1016/j.annepidem.2011.11.004.
- Chaumeny, Davison, J. van der M.M., Anthony C., Paul D. W. Kirk, 2022. Bayesian nonparametric mixture inconsistency for the number of components: How worried should we be in practice?.
- Clavel-Chapelon, F., E3N Study Group, 2015. Cohort Profile: The French E3N Cohort Study. *Int J Epidemiol* 44, 801–809. Doi: 10.1093/ije/dyu184.
- Clément, F., Xu, X., Donini, C.F., Clément, A., Omarjee, S., Delay, E., Treilleux, I., Fervers, B., Le Romancer, M., Cohen, P.A., Maguer-Satta, V., 2017. Long-term exposure to bisphenol A or benzo(a)pyrene alters the fate of human mammary epithelial stem cells in response to BMP2 and BMP4, by pre-activating BMP signaling. *Cell Death Differ.* 24, 155–166. <https://doi.org/10.1038/cdd.2016.107>.
- Coker, E., Katamba, A., Kizito, S., Eskenazi, B., Davis, J.L., 2020. Household air pollution profiles associated with persistent childhood cough in urban Uganda. *Environ. Int.* 136, 105471 <https://doi.org/10.1016/j.envint.2020.105471>.
- Coker, E.S., Molitor, J., Liverani, S., Martin, J., Maranzano, P., Pontarollo, N., Vergalli, S., 2023. Bayesian profile regression to study the ecologic associations of correlated environmental exposures with excess mortality risk during the first year of the Covid-19 epidemic in Lombardy, Italy. *Environ. Res.* 216, 114484 <https://doi.org/10.1016/j.envres.2022.114484>.

- Couvidat, F., Bessagnet, B., Garcia-Vivanco, M., Real, E., Menut, L., Colette, A., 2018. Development of an inorganic and organic aerosol model (CHIMERE 2017 $\beta$  v1.0): Seasonal and spatial evaluation over Europe. *Geosci. Model Dev.* 11, 165–194. <https://doi.org/10.5194/gmd-11-165-2018>.
- Danjou, A.M.N., Coudon, T., Praud, D., Lévêque, E., Faure, E., Salizzoni, P., Le Romancer, M., Severi, G., Mancini, F.R., Leffondré, K., Dossus, L., Fervers, B., 2019. Long-term airborne dioxin exposure and breast cancer risk in a case-control study nested within the French E3N prospective cohort. *Environ. Int.* 124, 236–248. <https://doi.org/10.1016/j.envint.2019.01.001>.
- Darbre, P.D., 2021. Endocrine disrupting chemicals and breast cancer cells. *Adv. Pharmacol.* 92, 485–520. <https://doi.org/10.1016/bs.apha.2021.04.006>.
- Deygas, F., Amadou, A., Coudon, T., Grassot, L., Couvidat, F., Bessagnet, B., Faure, E., Salizzoni, P., Gulliver, J., Caudeville, J., Severi, G., Mancini, F.R., Leffondré, K., Fervers, B., Praud, D., 2021. Long-term atmospheric exposure to PCB153 and breast cancer risk in a case-control study nested in the French E3N cohort from 1990 to 2011. *Environ. Res.* 195, 110743. <https://doi.org/10.1016/j.envres.2021.110743>.
- Eurostat, 2024. Contexte - NUTS - Nomenclature des unités territoriales statistiques - Eurostat [WWW Document]. URL [https://ec.europa.eu/cas/login?reqsurl=ECAS\\_LR-16888280-M3H0v1bFbSx4npCwalt3EYeABmoVSGOmK3oP8LD1zohWZkzHVziwLbVsXhmGj8nw0v6y5SoBztCSwzQAK5xMzG-ynt0f97TTHqmO1W4k0Hqea-KhMAUiv0cVjR2StB9kuYZ40zkrCXZcQPendWbu74zvAZSLjXD7UqXdk3U8FrueHdYnv4zJxfmLdsREUeazq2s0](https://ec.europa.eu/cas/login?reqsurl=ECAS_LR-16888280-M3H0v1bFbSx4npCwalt3EYeABmoVSGOmK3oP8LD1zohWZkzHVziwLbVsXhmGj8nw0v6y5SoBztCSwzQAK5xMzG-ynt0f97TTHqmO1W4k0Hqea-KhMAUiv0cVjR2StB9kuYZ40zkrCXZcQPendWbu74zvAZSLjXD7UqXdk3U8FrueHdYnv4zJxfmLdsREUeazq2s0) (accessed 8.2.24).
- Farrar, D.E., Glauber, R.R., 1967. Multicollinearity in regression analysis: The problem revisited. *Rev. Econ. Stat.* 49, 92–107. <https://doi.org/10.2307/1937887>.
- Faure, E., Danjou, A.M.N., Clavel-Chapelon, F., Boutron-Ruault, M.-C., Dossus, L., Fervers, B., 2017. Accuracy of two geocoding methods for geographic information system-based exposure assessment in epidemiological studies. *Environ. Health* 16, 15. <https://doi.org/10.1186/s12940-017-0217-5>.
- Filippini, T., Torres, D., Lopes, C., Carvalho, C., Moreira, P., Naska, A., Kasdagli, M.-I., Malavolti, M., Orsini, N., Vinceti, M., 2020. Cadmium exposure and risk of breast cancer: A dose-response meta-analysis of cohort studies. *Environ. Int.* 142, 105879. <https://doi.org/10.1016/j.envint.2020.105879>.
- Fiolet, T., Casagrande, C., Nicolas, G., Horvath, Z., Frenoy, P., Weiderpass, E., Katzke, V., Kaaks, R., Rodriguez-Barranco, M., Panico, S., Sacerdote, C., Manjer, J., Sonestedt, E., Grioni, S., Agudo, A., Rylander, C., Haugdahl Nøst, T., Skeie, G., Tjønneland, A., Raaschou-Nielsen, O., Ardanaz, E., Amiano, P., Dolores Chirlaque López, M., Schulze, M.B., Wennberg, M., Harlid, S., Cairat, M., Kvaskoff, M., Huybrechts, I., Romana Mancini, F., 2022. Dietary intakes of dioxins and polychlorobiphenyls (PCBs) and breast cancer risk in 9 European countries. *Environ Int* 163, 107213. <https://doi.org/10.1016/j.envint.2022.107213>.
- Florez-Garcia, V.A., Guevara-Romero, E.C., Hawkins, M.M., Bautista, L.E., Jenson, T.E., Yu, J., Kalkbrenner, A.E., 2023. Cadmium exposure and risk of breast cancer: A meta-analysis. *Environ. Res.* 219, 115109. <https://doi.org/10.1016/j.envres.2022.115109>.
- Gabet, S., Lemarchand, C., Guénel, P., Slama, R., 2021. Breast cancer risk in association with atmospheric pollution exposure: A meta-analysis of effect estimates followed by a health impact assessment. *Environ. Health Perspect.* 129, 57012. <https://doi.org/10.1289/EHP8419>.
- Gallo, A., Agnese, V., Coronello, C., Raffa, G.M., Bellavia, D., Conaldi, P.G., Pilato, M., Pasta, S., 2018. On the prospect of serum exosomal miRNA profiling and protein biomarkers for the diagnosis of ascending aortic dilatation in patients with bicuspid and tricuspid aortic valve. *Int. J. Cardiol.* 273, 230–236. <https://doi.org/10.1016/j.ijcard.2018.10.005>.
- Guerrero, C.B.B., Horálek, J., de Leeuw, F., Couvidat, F., 2016. Benzo(a)pyrene in Europe: Ambient air concentrations, population exposure and health effects. *Environ. Pollut.* 214, 657–667. <https://doi.org/10.1016/j.envpol.2016.04.081>.
- Hastie, D.I., Liverani, S., Richardson, S., 2015. Sampling from Dirichlet process mixture models with unknown concentration parameter: Mixing issues in large data implementations. *Stat. Comput.* 25, 1023–1037. <https://doi.org/10.1007/s11222-014-9471-3>.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67. <https://doi.org/10.1080/00401706.1970.10488634>.
- Hoover, J.H., Coker, E.S., Erdei, E., Luo, L., Begay, D., MacKenzie, D., NBCS Study Team, Lewis, J., 2023. Preterm Birth and Metal Mixture Exposure among Pregnant Women from the Navajo Birth Cohort Study. *Environmental Health Perspectives* 131, 127014. <https://doi.org/10.1289/EHP10361>.
- INSEE, 2024. Base du comparateur de territoires [WWW Document]. URL <https://www.insee.fr/fr/statistiques/2521169> (accessed 8.2.24).
- Jeanjean, M., Goix, S., Periot, M., Douib, K., Dron, J., Etienne, M.-P., Marchand, P., Austruy, A., Revenko, G., Chamaret, P., 2021. Environmental and dietary exposures near a major industrial harbour (Fos-sur-Mer, France) identified as a significant pathway for PCBs and PCDD/Fs accumulation in residents' blood serum. *Expo. Health* 13, 447–464. <https://doi.org/10.1007/s12403-021-00395-8>.
- Large, C., Wei, Y., 2017. Geographic variations in female breast cancer incidence in relation to ambient air emissions of polycyclic aromatic hydrocarbons. *Environ. Sci. Pollut. Res. Int.* 24, 17874–17880. <https://doi.org/10.1007/s11356-017-9395-5>.
- Lengyel, A., Botta-Dukat, Z., 2019. Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecol. Evol.* 9, 13231–13243. <https://doi.org/10.1002/ece3.5774>.
- Liverani, S., Hastie, D.I., Azizi, L., Papathomas, M., Richardson, S., 2015. PRMiuM: An R package for profile regression mixture models using Dirichlet processes. *J. Stat. Softw.* 64, 1–30. <https://doi.org/10.18637/jss.v064.i07>.
- Loomis, D., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Baan, R., Mattock, H., Straif, K., International Agency for Research on Cancer Monograph Working Group IARC, 2013. The carcinogenicity of outdoor air pollution. *Lancet Oncol* 14, 1262–1263. [https://doi.org/10.1016/s1470-2045\(13\)70487-x](https://doi.org/10.1016/s1470-2045(13)70487-x).
- MacLehose, R.F., Dunson, D.B., Herring, A.H., Hoppin, J.A., 2007. Bayesian methods for highly correlated exposure data. *Epidemiology* 18, 199–207. <https://doi.org/10.1097/01.ede.0000256320.30737.c0>.
- Mattei, F., Liverani, S., Guida, F., Matrat, M., Cenée, S., Azizi, L., Menvielle, G., Sanchez, M., Piloget, C., Lapôtre-Ledoux, B., Luce, D., Richardson, S., Stücker, I., 2016. Multidimensional analysis of the effect of occupational exposure to organic solvents on lung cancer risk: the ICARE study. *Occup. Environ. Med.* 73, 368–377. <https://doi.org/10.1136/oemed-2015-103177>.
- Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquet, S., Valari, M., Vautard, R., Vivanco, M.G., 2013. CHIMERE 2013: A model for regional atmospheric composition modelling. *Geosci. Model Dev.* 6, 981–1028. <https://doi.org/10.5194/gmd-6-981-2013>.
- Molitor, J., Papathomas, M., Jerrett, M., Richardson, S., 2010. Bayesian profile regression with an application to the National survey of children's health. *Biostatistics* 11, 484–498. <https://doi.org/10.1093/biostatistics/kxq013>.
- Niehoff, N.M., Gammon, M.D., Keil, A.P., Nichols, H.B., Engel, L.S., Sandler, D.P., White, A.J., 2019. Airborne mammary carcinogens and breast cancer risk in the Sister Study. *Environ. Int.* 130, 104897. <https://doi.org/10.1016/j.envint.2019.06.007>.
- Niehoff, N.M., Terry, M.B., Bookwalter, D.B., Kaufman, J.D., O'Brien, K.M., Sandler, D.P., White, A.J., 2022. Air pollution and breast cancer: An examination of modification by underlying familial breast cancer risk. *Cancer Epidemiol. Biomark. Prev.* 31, 422–429. <https://doi.org/10.1158/1055-9965.EPI-21-1140>.
- Pearce, N., 2016. Analysis of matched case-control studies. *BMJ* 352, i969. <https://doi.org/10.1136/bmj.i969>.
- Praud, D., Deygas, F., Amadou, A., Bouilly, M., Turati, F., Bravi, F., Xu, T., Grassot, L., Coudon, T., Fervers, B., 2023. Traffic-related air pollution and breast cancer risk: A systematic review and meta-analysis of observational studies. *Cancers (Basel)* 15, 927. <https://doi.org/10.3390/cancers15030927>.
- Richardson, D., 2004. An incidence density sampling program for nested case-control analyses. *Occup. Environ. Med.* 61, e59. <https://doi.org/10.1136/oem.2004.014472>.
- Rouanet, A., Johnson, R., Strauss, M.E., Richardson, S., Tom, B.D., White, S.R., Kirk, P.D. W., 2021. Bayesian profile regression for clustering analysis involving a longitudinal response and explanatory variables. *arXiv:2111.04518 [stat]*.
- Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. <https://doi.org/10.3322/caac.21660>.
- Turner, M.C., Andersen, Z.J., Baccarelli, A., Diver, W.R., Gapstur, S.M., Pope, C.A., Prada, D., Samet, J., Thurston, G., Cohen, A., 2020. Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. *CA Cancer J. Clin.* <https://doi.org/10.3322/caac.21632>.
- Vofham, T., Bertrand, K.A., Jones, R.R., Deziel, N.C., Dupré, N.C., James, P., Liu, Y., Vieira, V.M., Tamimi, R.M., Hart, J.E., Ward, M.H., Laden, F., 2020. Dioxin exposure and breast cancer risk in a prospective cohort study. *Environ. Res.* 186, 109516. <https://doi.org/10.1016/j.envres.2020.109516>.
- Wan, M.L.Y., Co, V.A., El-Nezami, H., 2022. Endocrine disrupting chemicals and breast cancer: A systematic review of epidemiological studies. *Crit. Rev. Food Sci. Nutr.* 62, 6549–6576. <https://doi.org/10.1080/10408398.2021.1903382>.
- Wei, W., Wu, B.-J., Wu, Y., Tong, Z.-T., Zhong, F., Hu, C.-Y., 2021. Association between long-term ambient air pollution exposure and the risk of breast cancer: a systematic review and meta-analysis. *Environ. Sci. Pollut. Res.* 28, 63278–63296. <https://doi.org/10.1007/s11356-021-14903-5>.
- White, A.J., Bradshaw, P.T., Hamra, G.B., 2018. Air pollution and breast cancer: A review. *Curr. Epidemiol. Rep.* 5, 92–100. <https://doi.org/10.1007/s40471-018-0143-2>.
- White, A.J., O'Brien, K.M., Niehoff, N.M., Carroll, R., Sandler, D.P., 2019. Metallic air pollutants and breast cancer risk in a nationwide cohort study. *Epidemiology* 30, 20–28. <https://doi.org/10.1097/EDE.0000000000000917>.
- White, A.J., Gregoire, A.M., Niehoff, N.M., Bertrand, K.A., Palmer, J.R., Coogan, P.F., Bethea, T.N., 2021. Air pollution and breast cancer risk in the Black Women's Health Study. *Environ. Res.* 194, 110651. <https://doi.org/10.1016/j.envres.2020.110651>.
- Xu, J., Ye, Y., Huang, F., Chen, H., Wu, H., Huang, J., Hu, J., Xia, D., Wu, Y., 2016. Association between dioxin and cancer incidence and mortality: A meta-analysis. *Sci. Rep.* 6, 38012. <https://doi.org/10.1038/srep38012>.
- Yaghjian, L., Arao, R., Brokamp, C., O'Meara, E.S., Sprague, B.L., Ghita, G., Ryan, P., 2017. Association between air pollution and mammographic breast density in the Breast Cancer Surveillance Consortium. *Breast Cancer Res.* 19, 36. <https://doi.org/10.1186/s13058-017-0828-3>.
- Yilmaz, B., Terekci, H., Sandal, S., Kelestimur, F., 2020. Endocrine disrupting chemicals: Exposure, effects on human health, mechanism of action, models for testing and strategies for prevention. *Rev. Endocr. Metab. Disord.* 21, 127–147. <https://doi.org/10.1007/s11554-019-09521-z>.
- Zhang, Y., Kleer, C.G., 2016. Phyllodes tumor of the breast: histopathologic features, differential diagnosis, and molecular/genetic updates. *Arch. Pathol. Lab. Med.* 140, 665–671. <https://doi.org/10.5858/arpa.2016-0042-RA>.
- Zhang, Z., Yan, W., Chen, Q., Zhou, N., Xu, Y., 2019. The relationship between exposure to particulate matter and breast cancer incidence and mortality: A meta-analysis. *Medicine (Baltimore)* 98, e18349. <https://doi.org/10.1097/MD.00000000000018349>.