



Profile of exposures and lung function in adults with asthma: An exposome approach in the EGEA study

Alicia Guillien, Johanna Lepeule, Emie Seyve, Nicole Le Moual, Isabelle Pin, Bruno Degano, Judith Garcia-Aymerich, Jean-Louis Pépin, Christophe Pison, Orianne Dumas, et al.

► To cite this version:

Alicia Guillien, Johanna Lepeule, Emie Seyve, Nicole Le Moual, Isabelle Pin, et al.. Profile of exposures and lung function in adults with asthma: An exposome approach in the EGEA study. *Environmental Research*, 2020, 196, pp.110422. 10.1016/j.envres.2020.110422 . inserm-03156586

HAL Id: inserm-03156586

<https://inserm.hal.science/inserm-03156586>

Submitted on 2 Mar 2021

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.

Profile of exposures and lung function in adults with asthma: an exposome approach in the EGEA study

Alicia Guillien¹, Johanna Lepeule¹, Emie Seyve¹, Nicole Le Moual^{2,3}, Isabelle Pin^{1,4}, Bruno Degano⁵, Judith Garcia-Aymerich^{6,7,8}, Jean-Louis Pépin⁵, Christophe Pison⁹, Orianne Dumas^{2,3}, Raphaëlle Varraso^{2,3}, Valérie Siroux¹

¹Univ. Grenoble Alpes, Inserm, CNRS, Team of environmental epidemiology applied to Reproduction and Respiratory health, IAB, 38000 Grenoble, France

²INSERM, U1168, VIMA: Aging and Chronic Diseases, Epidemiological and Public Health Approaches, Villejuif, France

³University Versailles St-Quentin-en-Yvelines, UMR-S 1168, Montigny le Bretonneux, France

⁴Pediatric Department, CHU Grenoble Alpes, Grenoble, France

⁵Univ. Grenoble Alpes, Inserm, CHU Grenoble Alpes, HP2, Grenoble, France

⁶ISGlobal, Barcelona, Spain

⁷Universitat Pompeu Fabra (UPF), Barcelona, Spain

⁸CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

⁹Univ. Grenoble Alpes, Inserm, CHU Grenoble Alpes, LBFA, Grenoble, France

Correspondence:

Alicia Guillien

Institut pour l'Avancée des Biosciences

Equipe d'épidémiologie environnementale appliquée à la reproduction et à la santé
respiratoire

Centre de Recherche UGA / Inserm U 1209 / CNRS UMR 5309

Site Santé – Allée des Alpes

38700 La Tronche

Tel: +33 4 76 54 94 69

Email: alicia.guillien@univ-grenoble-alpes.fr

Funding: This work was supported in part by the Hospital Program of Clinical Research (PHRC)–Paris, PHRC-Grenoble, National PHRC 2012, the scientific committee “AGIR pour les Maladies Chroniques” Merck Sharp & Dohme (MSD), the GA2LEN project (Global Allergy and Asthma European Network) and ANR “Investissements d’avenir” (ANR-15-IDEX-02). ISGlobal is a member of CERCA Programme/Generalitat de Catalunya. Funding sources had no involvement in study design, the collection, analysis and interpretation of data; in the writing of the report; and the decision to submit the article for publication.

Ethics approval: Ethical approval was received from the local ethics committee (Comité de Protection des Personnes participant à la Recherche Biomédicale (CPPRB) of Paris-Necker on 31 March 2003, n°01-07-07)

Consent to participate and for publication: Written consent was obtained from all participants

Abstract

Background: Environmental research on multifactorial health outcomes calls for exposome approaches able to assess the joint effect of multiple exposures.

Objective: Our aim was to identify profiles of exposure to lifestyle/environmental factors associated with lung function in adults with asthma using a cluster-based approach.

Methods: We used data from 599 adults of the Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy (EGEA) (mean age 39.0 years, 52% men) who ever had asthma. Exposures to 53 lifestyle/environmental factors were assessed by questionnaires or geographic information systems-based models. A two-step approach was developed: 1) exposome dimension reduction by selecting factors showing association with forced expiratory volume in 1 second (FEV₁) ($p < 0.20$) in an exposome-wide association study (ExWAS), 2) clustering analysis using the supervised Bayesian Profile Regression (sBPR) to group individuals according to FEV₁ level and to their profile of exposure to a reduced set of uncorrelated exposures (each paired correlation < 0.70) identified in step 1.

Results: The ExWAS identified 21 factors showing suggestive association with FEV₁ (none significant when controlling for multiple tests). The sBPR conducted on 15 uncorrelated exposures identified in step 1, revealed 3 clusters composed of 30, 115 and 454 individuals with a mean $\pm$ SD FEV₁(%pred) of 79% $\pm$ 21, 90% $\pm$ 19 and 93% $\pm$ 16, respectively. Cluster 1 was composed of individuals with heavy smoking, poor diet, higher outdoor humidity and proximity to traffic, while cluster 2 and 3 included individuals with moderate/low levels of exposure to these factors.

Discussion: This exposome study identified a specific profile of joint lifestyle and environmental factors, associated with a low FEV₁ in adults with asthma. None of the exposures revealed significant association when considered independently.

Words count: 272

Key words: Lung function, asthma, exposome, Bayesian Profile Regression

1 **Introduction**

2 Over decades, several studies reported associations between exposure to various environmental
3 factors and asthma related outcomes. The most consistent associations are observed for tobacco
4 smoke (Strachan and Cook, 1998; Vignoud et al., 2011), allergens and sensitizers (house-dust
5 mites, moulds, domestic pets, ...), (Wright and Phipatanakul, 2014) traffic-related air pollution
6 (Guarnieri and Balmes, 2014) and occupational exposures to asthmagens (Tarlo and Lemiere,
7 2014). Other exposures are suspected to be associated with asthma, such as use of cleaning
8 products at home (Zock et al., 2007) and climate factors (Correia Junior et al., 2017). Finally,
9 there are emerging concerns for other exposures including greenness or lifestyle factors such
10 as diet and physical activity (Fuentes et al., 2016; Sbihi et al., 2017; Varraso and Camargo,
11 2016). Although asthma is a multifactorial disease in which environmental and lifestyle factors
12 are suspected to interact, most of previous environmental studies are limited by the fact that
13 they focused on a single exposure or family of exposures (Vineis, 2018).

14 The concept of exposome was first proposed by C. Wild in 2005 to encompass life-course
15 environmental exposures (including lifestyle factors), from the prenatal period onwards (Wild,
16 2005). It was developed to highlight the need for more comprehensive approaches to address
17 the role of the environment in complex diseases (Schwartz and Collins, 2007) carrying the
18 expectation that the use of holistic and data-driven approaches can expand our understanding
19 of the aetiology and long-term trajectories of complex diseases.

20 The exposome includes a large number of covariates that are inter-related and aggregated, in
21 particular within exposure families (Tamayo-Uria et al., 2019). Among the different statistical
22 approaches to deal with such exposome data (Santos et al., 2020), a comprehensive
23 methodology based on clusters representing joint covariate profiles as opposed to individual
24 risk factors offers new avenues in environmental research on multifactorial diseases (Molitor
25 et al., 2010). This type of clustering approach has already been applied in the field of asthma to

disentangle the asthma phenotypic heterogeneity by simultaneously considering several clinical and individual characteristics (Boudier et al., 2013; Halder et al., 2008; Siroux et al., 2011). In the context of the exposome-respiratory health association, although a recent study from Agier *et al.* considered a large number of exposures using an exposome approach to evaluate their association with lung function in European children, their work did not consider cluster-based analyses to account for combined exposures (Agier et al., 2019). A recent study applied a cluster-based analysis to account for joint prenatal exposures to chemicals in asthma and lung function in 319 children, but the study was limited to a single family of exposures (phenols and phthalates) (Berger et al., 2020). Because some of these risk factors for respiratory health might have a stronger effect among individuals with asthma than among individuals without asthma, as clearly demonstrated for smoking (Polosa and Thomson, 2013), investigating the exposome-lung function association in individuals with asthma might be more powerful than in the general population.

In the frame of the Epidemiological Study on the Genetics and Environment of Asthma (EGEA) and based on a detailed exposome in adulthood, we aimed to identify the joint effect of lifestyle and environmental exposures on lung function in adults with asthma using a supervised clustering approach.

Methods

Study population

EGEA is a French cohort based on three surveys over 20 years. At EGEA1 in early 1990s, 388 cases with asthma were recruited in chest clinics, so as their first-degree relatives (n=1244) and 415 population-based controls. All participants were born in France as well as their two parents. The population was invited to participate to the 12-year and 20-year follow-ups (EGEA2 and

EGEA3). At each survey, individuals responded to a detailed questionnaire on respiratory health and environment, based on international standardised tools (Kauffmann and Dizier, 1995). The present study focuses on adults with ever asthma at EGEA2 (2003-2007) who performed a spirometry. Ever-asthma was defined at EGEA1 by being recruited as an asthma case in chest clinics or, for first degree relatives, by a positive answer to either “Have you ever had attacks of breathlessness at rest with wheezing?” or “Have you ever had asthma attacks?”.

Spirometry testing

Spirometry was performed using a flow-volume spirometer (SpiroDyn'R; SAS Dyn'R, Aix-en-Provence, France) according to the American Thoracic Society and European Respiratory Society guidelines (Miller et al., 2005). Pre-bronchodilator forced expiratory volume in 1 s (FEV₁) was expressed in percentage of the predicted value using the Global Lung Function Initiative reference equations, thereafter referred as FEV₁% (Quanjer et al., 2012) to account for the major effects of age, height and gender in FEV₁.

Characterisation of the exposome

Data regarding lifestyle factors were assessed mostly through self-completed questionnaires, which provided information about 9 exposure families: indoor air (3 variables), tobacco exposure (4 variables), season of examination (season at which individual performed the spirometry, 1 variable), pets (2 variables), occupational exposures (2 variables, assessed through a job-exposure matrix), household cleaning products (2 variables), rurality (2 variables), socio-economic status (SES) (assessed at the contextual level by the French Deprivation Index (FDep) from current geocoded residential address, built to capture health inequalities at ecological level (Temam et al., 2017), 1 variable), and physical activity and diet (2 variables, assessed through the Metabolic Equivalent Tasks (METs) and the Alternative Healthy Eating Index-2010 (AHEI-2010) diet score, respectively). Moreover, a broad range of environmental exposures was assessed through geographic information system (GIS) models

that allow allocating a spatio-temporal estimate of exposure from geographical residential coordinates. These GIS-based exposures covered 8 families: altitude (1 variable), building and population densities (3 variables), atmospheric pollutants (4 variables), natural spaces (8 exposures), weather (6 variables), roads/cycle paths (3 variables), ambient ultra violet radiation (UV) (3 variables) and health services (6 variables). The GIS-based models are resting on National Geographic Institute (IGN) maps, the French National Institute for Statistics and Economic Research (INSEE) data, Météo-France data, geospatial models, land cover databases and satellite data. All these exposures were calculated from the current geocoded residential addresses. Most of these exposures was assessed during the last 12 months before the spirometry examination, except for some exposures assessed by GIS-based models for which models of exposure were not available for the relevant time window (see Exposome assessment method in the Supplementals).

In total, 53 exposure variables (19 lifestyle and 34 environmental), a priori candidate risks factors in respiratory health, were assessed at EGEA2.

Statistical analysis

Figure 1 shows the statistical workflow. Exposure variables were transformed (to approach normality) or dichotomized and missing values were imputed using the method of chained equations (White et al., 2011) and 5 complete datasets were generated. All continuous exposures variables were standardized by the inter-quartile range (IQR) to allow between-covariate comparison of regression estimates. Correlations between exposure variables were presented through heatmaps. Correlations were assessed through Pearson coefficient for two quantitative variables, polychoric coefficient for two qualitative variables and polyserial coefficient for a quantitative and a qualitative variable. A second heat map assessed absolute within (diagonal) and between (off-diagonal) families of exposure median correlations.

99 We proposed a two-step approach to evaluate the impact of the exposome on lung function.
100 The first step of the association study consisted in the exposome dimension reduction to allow
101 a better stability of the clustering model applied in step 2. Exposome-Wide Association
102 Analysis (ExWAS) was performed for FEV₁% by fitting independent mixed linear regression
103 models for each exposure variable, with adjustment for the study centre, gender, age, height
104 and socio-professional category as fixed effects and family as random effect (Patel et al., 2010).
105 Factors showing association with FEV₁% (p-value<0.2) were selected in the reduced set of
106 exposures.

107 The second step consisted in identifying profiles of co-exposures associated with low lung
108 function. A supervised cluster analysis, *i.e.* simultaneously considering in the model the
109 outcome and exposures, was conducted using the supervised Bayesian Profile Regression
110 (sBPR) clustering method that relies on Markov Chain Monte Carlo sampling methods
111 (200,000 iterations) (Molitor et al., 2010). The sBPR, that requires to *a priori* select a unique
112 set of variables, was applied in the reduced set of uncorrelated (<0.7) exposures identified in
113 step 1 by the ExWAS, FEV₁% and included confounders used in the ExWAS analysis (age,
114 gender, height, centre of inclusion, number of smoked pack-years and socio-professional
115 category) as covariates. When the absolute correlation between two of the exposures identified
116 in step 1 was above 0.70, we retained the variable with the lowest number of missing data, and
117 when these two numbers were similar we retained the variable with the lowest p-value in the
118 ExWAS analysis. Socio-demographic characteristics and exposure factors were compared
119 between clusters to identify the factors mainly driving the clustering. This analysis was also
120 performed after adjustment for age, gender and centre to identify exposures associated with
121 clusters independently of these 3 variables. Moreover, phenotypic characteristics at EGEA2,
122 but also at EGEA3 about 7 years later, were compared between clusters.

Because this clustering method relies on a stochastic process, we tested the stability of our final model by running the model with 10 different seeds (all with 200,000 iterations). Each model identified 3 clusters and the median [interquartile range] number of subjects belonging to the same cluster whatever the seed used was 0.86 [0.73-0.94] (data not shown).

All analyses were performed using R software (version 3.5). We used the software package *mice* for multiple imputation, *polycor* for correlation coefficients, *lme4* for mixed models and *PReMiuM* to fit BPR models.

Results

Description of the population

Among 1,903 individuals included in the EGEA2 study, 698 individuals ever had asthma. Among them, 99 were not analysed further because they had not performed spirometry or had missing data for their residential address or did not complete the environmental questionnaire (Figure 2). Excluded individuals had similar characteristics than the 599 included individuals, except that they were younger (mean $\pm$ SD age 39 $\pm$ 6 vs. 34 $\pm$ 18 years, $p=0.01$, for included vs. excluded individuals, respectively, see supplemental Table E1). Main characteristics of the studied population are presented in Table 1. The 599 included individuals belonged to 356 families, and half of them were men and had never smoked. Two thirds of the population were executives or technicians. The studied population had a mean $\pm$ SD FEV₁% of 92% $\pm$ 17% and 88% had either at least one asthma attack or used asthma treatment in the past 12 months.

Description of the exposome

Descriptive statistics, transformation used and number of missing values for each exposure variable are presented in Table 2. Most of the environmental exposures showed high between-centre variability (see supplemental Table E2). Overall, correlations between exposures were higher within-families of exposure (median [IQR] absolute correlation = 0.46 [0.28-0.76]) than

between-families of exposures (median [IQR] absolute correlation = 0.08 [0.04-0.13]) (Figures 3 and 4). As an example, the three variables belonging to the building and population densities family had a median absolute 2-way correlations of 0.78 while the median of the absolute 2-way correlations between any of these three variables and the two variables belonging to the rurality family was 0.26.

Exposome dimension reduction from ExWAS results

Among the 53 exposure variables, the ExWAS analysis identified 21 exposures from 10 families of exposures associated with FEV₁% with a p-value<0.20 (Supplemental Table E3 and Figure E1). After the examination of the correlation matrix of these 21 variables, 15 uncorrelated exposure variables from 9 exposure families were retained for the inclusion in the sBPR analysis. Among them, 7 had a positive association with FEV₁% (distance to nearest green space and major road, humidity rate, population density, AHEI-2010 score, the second quintile of the French Deprivation Index (FDep) score (vs. the first quintile), and having at least one pharmacy in their surrounding space), while 8 had negative association with FEV₁% (≥ 1 major road and/or one cycle path in a 100m buffer, be born in a rural municipality, mean annual wind ≥ 2 m/s, distance to nearest blue space, mean annual temperature, season of examination (reference=spring) and intermediate exposure to passive smoking (vs. no exposure)). None of the association tested in the ExWAS was significantly associated with FEV₁% when correcting for multiple comparison (Supplemental Figure E1).

Identification of profiles of joint exposures associated with FEV₁% using the sBPR

The sBPR, conducted on FEV₁ and on the 15 uncorrelated exposures identified in the previous step, identified 3 clusters composed of 30, 115 and 454 individuals, respectively. Lifestyle and environmental characteristics of the three clusters are described in Table 3. Individuals belonging to cluster 1 (n=30), that shows the lowest mean FEV₁% (mean $\pm$ SD 79% $\pm$ 21%), were more often heavy smokers and had more often an unhealthy diet, ≥ 1 major road in a 100m

buffer surrounding their home, a higher outdoor humidity rate and less pharmacy surrounding their residential address. Regarding individuals assigned in cluster 2 ($n=115$, $\text{mean} \pm \text{SD FEV}_1\% = 90\% \pm 19\%$), they had the same mean age and mean pack-years than individuals from cluster 1, but had a healthier diet and were exposed to lower rate of humidity than individuals belonging to other clusters. Finally, individuals belonging to cluster 3 ($n=454$, $\text{mean} \pm \text{SD FEV}_1\% = 93\% \pm 16\%$) were younger, more often never-smokers but more often exposed to passive smoking, and lived closer to blue spaces. After adjustment on age, gender and centre, comparison of exposures between the 3 clusters remained overall similar, except for the distance to nearest green space that became significantly lower in cluster 1 as compared to the other clusters.

Phenotypic characterization of the clusters at EGEA2 and about 7 years later at EGEA3

Individuals from cluster 1 had a statistically significant lower prevalence of allergic sensitization and higher rate of neutrophils at EGEA2 than individuals from other clusters (Table 4). Although not statistically significant, individuals belonging to cluster 1 more often reported use of inhaled steroids in the past 12 months as compared to those from cluster 2 and 3 (57%, 36% and 40% respectively). No difference was observed for the number of asthma symptoms and the total Asthma Quality of Life Questionnaire score. Comparison of phenotypic characteristics observed about 7 years later at EGEA3 ($n=502$, Supplemental Table E4) between the three clusters did not show statistically significant differences, although individuals belonging to cluster 1 tended to have more often asthma exacerbation over the past 12 months than those belonging to cluster 2 and 3 (35% vs 18% and 21%, respectively, Table 4).

Discussion

To our knowledge, this is the first study that addresses the impact of the exposome on lung function in adults with asthma by considering simultaneously a broad range of lifestyle factors and social, built and natural environment. While none of the factors revealed significant association with FEV₁ when considered independently in an ExWAS approach, the profile regression method revealed 3 profiles of joint exposures that show strong differences in terms of FEV₁% level and exposure profiles including lifestyle (smoking, diet), social (at the area and individual levels), built (proximity to major road, pharmacy) and natural (blue space, humidity) environments.

Methodological considerations

A major novelty in our study relies on the comprehensive analytical approach used, allowing considering the outcome and a large set of exposures simultaneously to cluster individuals. Indeed, until now most exposome studies considered exposure variables individually while, in their daily life, individuals are simultaneously exposed to a large number of various factors, which may jointly affect health. Thus, joint exposures were considered to identify clusters of adults with asthma sharing the same exposure profiles. This method has the advantage to deal with inter-correlated variables, as it is the case in exposome data. Nevertheless, this method did not allow us to consider the family structure of our data. Thus, we cannot rule out that the family design, although considered in the first step based on ExWAS, led to confounding issues through shared environment and genetic factors within families. In addition, the BPR method relies on Markov chain Monte Carol sampling methods which might lead to instability issue, in particular with increased number of variables. Although we attempted to limit the model instability by reducing the exposome dimension, avoiding very strong correlation between exposures and running the algorithm for 200,000 iterations, this remains a limitation. (Molitor et al., 2010). Finally, clustering methods being based on the entire set of exposure variables that are considered simultaneously to identify clusters, the model is highly dependent of the set of

variables including in the model and the addition or the deletion of one of these variables could lead to substantial changes in results. Strong *a priori* hypotheses on the input variables are therefore needed and sensitivity analyses on different set of variables should be avoided.

Results interpretation

Interestingly, the sBPR method allowed identifying joint exposures that formed subpopulations at risk of higher or lower FEV₁%, while the ExWAS resting on traditional regression models revealed no statistically significant association after considering multiple comparisons. Mixture of exposures is expected to have higher impaired health effects than each individual exposure because the effects of each factor are superimposed (Kortenkamp and Faust, 2018). Accordingly, cluster 1, that was associated with the lowest FEV₁, was characterized by smoking, unhealthy diet and exposure to traffic that are susceptible to be involved in similar biological pathways, such as for example inflammation and oxidative stress. A synergistic effect cannot be dismissed, as suggested by a study indicating a modifier effect of antioxidants on lung function in Mexican children exposed to high levels of air pollutants (Romieu et al., 2002), although environment-by-environment interactions have been poorly investigated.

As previously mentioned, our study is the first to address the impact of joint exposures on respiratory health, and therefore, results are difficult to compare with previous literature. Among the wide range of lifestyle, social, built and natural environment factors that we took into account in our exposome approach, only some of them play a major role in the identification of the clusters. Most of these factors, taken individually, have previously been identified, or were suggested to be involved, in asthma characteristics. For example, previous studies showed that smoking reduces lung function, in both individuals with and without asthma (Lange et al., 1998). A greater AHEI-2010, an overall measure of diet quality, has been linked with improved or fewer asthma symptoms and greater asthma control (Andrianasolo et al., 2018; Chiuve et al., 2012; Li et al., 2017). The presence of at least 1 major road in a 100m

buffer surrounding the house, a proxy of traffic-related air pollution, was associated with lower FEV₁ and exacerbations in asthma patients (Balmes et al., 2009; Guarnieri and Balmes, 2014). For other factors (distance from pharmacies, annual wind and outdoor humidity) results are more novel and warrant further investigation. Besides these factors identified in our clusters, some factors suggested to be associated with asthma characteristics or lung function decline in the literature, such as molds exposure (Pekkanen et al., 2007; *WHO Guidelines for Indoor Air Quality*, 2009) or domestic exposure to cleaning products (Dumas et al., 2017; Le Moual et al., 2012), were not identified as major determinants by our study. Overall, our findings underline the interest of the exposome, as a complementary approach in environmental epidemiology to better understand how co-exposures may impact respiratory health.

Regarding the phenotypic differences between the three clusters we found that cluster 1, showing the lowest FEV₁% value, included more subjects without allergic sensitisation and with higher level of neutrophils as well as a higher subsequent risk for asthma exacerbations, indicating that it might be a group characterized by a more severe form of the disease.

Strengths and weaknesses

Our study has a number of strengths. The main one, as previously discussed, lies in the fact that the comprehensive method applied allows for the first time to investigate the effect of multiple exposures together on lung function and does not simply adjust for co-exposures. By allowing to consider all possible interactions between the set of exposures, the BPR approach used allows to make a step forward in addressing interactions, usually restricted to approach considering interactions between pairs of exposures. In addition, the analysis was conducted in a well-characterized cohort of adults with asthma with strong attention paid to the quality of the objective lung function measures (Siroux et al., 2016). All exposure data were collected through validated questionnaires or state of the art GIS-based models.

We acknowledge that our study also suffers from several weaknesses. First, although our study is one of the first attempts to integrate multiple families of exposures, our exposome does not include all exposures relevant for lung function. For example, some chemical exposures (pesticides, phenols, phthalates, perfluorinated compounds, ...) for which there is emerging concern regarding a potential impact on respiratory health (Agier et al., 2019; Hoppin et al., 2009), concentration of aeroallergens, such as pollens, and early life exposures were not available. Secondly, all considered exposures were either self-reported in questionnaires or derived from GIS-based models, which could imply a lack of accuracy in the exposure assessment. Nevertheless, GIS used in this study were based on high spatial resolution models, below 1km x 1km for most of them. Third, although we collected through questionnaire phenotypic characteristics of individuals 7 years after their participation to the EGEA2, the current analysis is mainly a cross-sectional study and longitudinal studies are warranted to assess the impact of exposome on lung function trajectories. Fourth, the sBPR method, that identifies the best partition of a population without controlling for the size of the clusters, resulted in a cluster with few individuals ($n = 30$, 5% of the population), consistent with previous studies that applied this method (Coker et al., 2017). Interestingly, this specific cluster identified a subpopulation with a specific co-exposure profile associated with a mean FEV₁% 10 points lower in average than individuals from the other clusters, therefore a restricted population that could benefit from a targeted reinforced surveillance. Fifth, our population was essentially composed of Caucasian individuals living in France which is a developed country. Therefore, our results cannot be generalized to individuals from developing countries or individuals of other ethnic groups. Finally, we acknowledge that a replication study in an independent cohort would be particularly relevant, but we were unable to identify a cohort of adults with asthma, in which both similar information on a large set of exposures and the measure of lung function were available.

Clinical and public health perspectives

While previous studies identified factors individually associated with lung function in adults with asthma, here we demonstrated that identifying comprehensive profiles of social, lifestyle and environmental exposure, whose effects on lung function are likely to interact, can help to get better insight into the etiology of multifactorial health outcomes. In terms of public health, our results support that complementary approaches to simple public health messages focusing on a unique exposure to prevent health issues at the general population level (no or quit smoking, daily consumption of 5 fruits or vegetables, ...) are warranted. Approaches aimed at identifying restricted subpopulations at very high risk of disease development (or worsening) given their exposure profile could provide further policy relevant information. From clinical perspectives, methods for more systematic assessment of exposome, including questionnaires and digital health tools (apps) need to be developed and validated to be deployed in routine asthma care (Bédard et al., 2019). Then, practitioner could identify patients belonging to co-exposure profile associated with higher risk of low lung function and adapt accordingly healthcare management. Secondary and tertiary preventions in this particular subpopulation could help to limit costs related to their disease.

Conclusion

This exposome study identified a subpopulation of adults with asthma with a specific profile of joint environmental and lifestyle factors associated with a strong decrease of FEV₁. By characterizing the joint effect of multiple lifestyle and environmental risk factors in lung function, the present study goes a step further in the understanding of the complex role of environment in asthma and led to the identification of a specific asthma subgroup that could benefit from a targeted reinforced surveillance. A such comprehensive exposome approach, easily transposable to any multifactorial disease, is expected to help unravelling the impact of

environment on health.

Acknowledgments: The authors thank all those who participated in the setting of the EGEA study and on the examinations of the individuals.

We thank the EGEA cooperative group members: Coordination: V Siroux (epidemiology, PI since 2013); F Demenais (genetics); I Pin (clinical aspects); R Nadif (biology); F Kauffmann (PI 1992-2012). Respiratory epidemiology: Inserm ex-U 700, Paris: M Korobaëff (Egea1), F Neukirch (Egea1); Inserm ex-U 707, Paris: I Annesi- Maesano (Egea1-2); Inserm ex-U 1018, Villejuif: F Kauffmann, MP Oryszczyn (Egea1-2); Inserm U 10188, Villejuif: N Le Moual, R Nadif, R Varraso (ex-U1168); Inserm U 1209 Grenoble: V Siroux. Genetics: Inserm ex-U 393, Paris: J Feingold; Inserm U 946, Paris: E Bouzigon, F Demenais, MH Dizier; CNG, Evry: I Gut (now CNAG, Barcelona, Spain), M Lathrop (now Univ McGill, Montreal, Canada). Clinical centers: Grenoble: I Pin, C Pison; Lyon: D Ecochard (Egea1), F Gormand, Y Pacheco; Marseille: D Charpin (Egea1), D Vervloet (Egea1 2); Montpellier: J Bousquet; Paris Cochin: A Lockhart (Egea1), R Matran (now in Lille); Paris Necker: E Paty (Egea1-2), P Scheinmann (Egea1-2); Paris-Trousseau: A Grimfeld (Egea1-2), J Just. Data and quality management: Inserm ex-U155 (Egea1): J Hochez; Inserm U 10188, Villejuif: N Le Moual, L Orsi; Inserm ex-U780: C Ravault (Egea1-2); Inserm ex-U794: N Chateigner. (Egea1-2); Grenoble: J Quentin (Egea1-2). The authors thank all those who participated to the setting of the study and on the various aspects of the examinations involved: interviewers, technicians for lung function testing and skin prick tests, blood sampling, IgE determinations, coders, those involved in quality control, data and sample management and all those who supervised the study in all centers. The authors are grateful to the three CIC-Inserm of Necker, Grenoble and Marseille who supported the study and in which participants were examined. They are also grateful to the biobanks in

344 Lille (CIC Inserm), and at Annemasse (Etablissement français du sang) where biological
345 samples are stored. They are indebted to all the individuals who participated, without whom the
346 study would not have been possible.

347 We thank Météo-France for providing humidity and wind speed data from the SAFRAN model.

348 We thank Itai Kloog and Ian Hough for providing the temperature data and Ian Hough for the

349 NDVI data.

- Agier, L., Basagaña, X., Maitre, L., Granum, B., Bird, P.K., Casas, M., Oftedal, B., Wright, J., Andrusaityte, S., de Castro, M., Cequier, E., Chatzi, L., Donaire-Gonzalez, D., Grazuleviciene, R., Haug, L.S., Sakhi, A.K., Leventakou, V., McEachan, R., Nieuwenhuijsen, M., Petraviciene, I., Robinson, O., Roumeliotaki, T., Sunyer, J., Tamayo-Uria, I., Thomsen, C., Urquiza, J., Valentin, A., Slama, R., Vrijheid, M., Siroux, V., 2019. Early-life exposome and lung function in children in Europe: an analysis of data from the longitudinal, population-based HELIX cohort. *Lancet Planet Health* 3, e81–e92. [https://doi.org/10.1016/S2542-5196\(19\)30010-5](https://doi.org/10.1016/S2542-5196(19)30010-5)
- Andrianasolo, R.M., Kesse-Guyot, E., Adjibade, M., Hercberg, S., Galan, P., Varraso, R., 2018. Associations between dietary scores with asthma symptoms and asthma control in adults. *Eur Respir J* 52, 1702572. <https://doi.org/10.1183/13993003.02572-2017>
- Balmes, J.R., Earnest, G., Katz, P.P., Yelin, E.H., Eisner, M.D., Chen, H., Trupin, L., Lurmann, F., Blanc, P.D., 2009. Exposure to traffic: Lung function and health status in adults with asthma. *J Allergy Clin Immunol* 123, 626–631. <https://doi.org/10.1016/j.jaci.2008.10.062>
- Bédard, A., Basagaña, X., Anto, J.M., Garcia-Aymerich, J., Devillier, P., Arnavielhe, S., Bedbrook, A., Onorato, G.L., Czarlewski, W., Murray, R., Almeida, R., Fonseca, J., Costa, E., Malva, J., Morais-Almeida, M., Pereira, A.M., Todo-Bom, A., Menditto, E., Stellato, C., Ventura, M.T., Cruz, A.A., Stelmach, R., da Silva, J., Larenas-Linnemann, D., Fuentes-Pérez, J.M., Huerta-Villalobos, Y.R., Emuzyte, R., Kvedariene, V., Valiulis, A., Kuna, P., Samolinski, B., Klimek, L., Mösges, R., Pfaar, O., Shamaï, S., Annesi-Maesano, I., Bosse, I., Demoly, P., Fontaine, J.-F., Cardona, V., Mullol, J., Valero, A., Roller-Wirnsberger, R.E., Tomazic, P.V., Chavannes, N.H., Fokkens, W.J., Reitsma, S., Bewick, M., Ryan, D., Sheikh, A., Haahtela, T., Toppila-Salmi, S., Valovirta, E., Makris, M., Papadopoulos, N.G., Prokopakis, E.P., Psarros, F., Cingi, C., Gemicioglu, B., Yorgancioglu, A., Bosnic-Anticevich, S., O’Hehir, R.E., Bachert, C., Hellings, P.W., Pugin, B., Bindeslev-Jensen, C., Eller, E., Kull, I., Melén, E., Wickman, M., De Vries, G., van Eerd, M., Agache, I., Ansotegui, I.J., Dykewicz, M.S., Casale, T., Wallace, D., Wasserman, S., Laune, D., Bousquet, J., 2019. Mobile technology offers novel insights into the control and treatment of allergic rhinitis: The MASK study. *J Allergy Clin Immunol* 144, 135-143.e6. <https://doi.org/10.1016/j.jaci.2019.01.053>
- Berger, K., Coker, E., Rauch, S., Eskenazi, B., Balmes, J., Kogut, K., Holland, N., Calafat, A.M., Harley, K., 2020. Prenatal phthalate, paraben, and phenol exposure and childhood allergic and respiratory outcomes: Evaluating exposure to chemical mixtures. *Sci Total Environ* 725, 138418. <https://doi.org/10.1016/j.scitotenv.2020.138418>
- Boudier, A., Curjuric, I., Basagaña, X., Hazgui, H., Anto, J.M., Bousquet, J., Bridevaux, P.O., Dupuis-Lozeron, E., Garcia-Aymerich, J., Heinrich, J., Janson, C., Künzli, N., Leynaert, B., de Marco, R., Rochat, T., Schindler, C., Varraso, R., Pin, I., Probst-Hensch, N., Sunyer, J., Kauffmann, F., Siroux, V., 2013. Ten-Year Follow-up of Cluster-based Asthma Phenotypes in Adults. A Pooled Analysis of Three Cohorts. *Am J Respir Crit Care Med* 188, 550–560. <https://doi.org/10.1164/rccm.201301-0156OC>
- Chiuve, S.E., Fung, T.T., Rimm, E.B., Hu, F.B., McCullough, M.L., Wang, M., Stampfer, M.J., Willett, W.C., 2012. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. *J Nutr* 142, 1009–1018. <https://doi.org/10.3945/jn.111.157222>
- Coker, E., Gunier, R., Bradman, A., Harley, K., Kogut, K., Molitor, J., Eskenazi, B., 2017. Association between Pesticide Profiles Used on Agricultural Fields near Maternal Residences

- during Pregnancy and IQ at Age 7 Years. *Int J Environ Res Public Health* 14, 506. <https://doi.org/10.3390/ijerph14050506>
- Correia Junior, M.A. de V., Sarinho, E.S.C., Rizzo, J.A., Sarinho, S.W., 2017. Lower prevalence and greater severity of asthma in hot and dry climate. *J Pediatr (Rio J)* 93, 148–155. <https://doi.org/10.1016/j.jped.2016.05.006>
- Dumas, O., Wiley, A.S., Quinot, C., Varraso, R., Zock, J.-P., Henneberger, P.K., Speizer, F.E., Le Moual, N., Camargo, C.A., 2017. Occupational exposure to disinfectants and asthma control in US nurses. *Eur Respir J* 50, 1700237. <https://doi.org/10.1183/13993003.00237-2017>
- Fuertes, E., Markevych, I., Bowatte, G., Gruzieva, O., Gehring, U., Becker, A., Berdel, D., von Berg, A., Bergström, A., Brauer, M., Brunekreef, B., Brüske, I., Carlsten, C., Chan-Yeung, M., Dharmage, S.C., Hoffmann, B., Klümper, C., Koppelman, G.H., Kozyrskyj, A., Korek, M., Kull, I., Lodge, C., Lowe, A., MacIntyre, E., Pershagen, G., Standl, M., Sugiri, D., Wijga, A., MACS, Heinrich, J., 2016. Residential greenness is differentially associated with childhood allergic rhinitis and aeroallergen sensitization in seven birth cohorts. *Allergy* 71, 1461–1471. <https://doi.org/10.1111/all.12915>
- Guarnieri, M., Balmes, J.R., 2014. Outdoor air pollution and asthma. *The Lancet* 383, 1581–1592. [https://doi.org/10.1016/S0140-6736\(14\)60617-6](https://doi.org/10.1016/S0140-6736(14)60617-6)
- Haldar, P., Pavord, I.D., Shaw, D.E., Berry, M.A., Thomas, M., Brightling, C.E., Wardlaw, A.J., Green, R.H., 2008. Cluster analysis and clinical asthma phenotypes. *Am. J. Respir. Crit. Care Med.* 178, 218–224. <https://doi.org/10.1164/rccm.200711-1754OC>
- Hoppin, J.A., Umbach, D.M., London, S.J., Henneberger, P.K., Kullman, G.J., Coble, J., Alavanja, M.C.R., Beane Freeman, L.E., Sandler, D.P., 2009. Pesticide use and adult-onset asthma among male farmers in the Agricultural Health Study. *Eur Respir J* 34, 1296–1303. <https://doi.org/10.1183/09031936.00005509>
- Kauffmann, F., Dizier, M.H., 1995. EGEA (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy)--design issues. EGEA Co-operative Group. *Clin. Exp. Allergy* 25 Suppl 2, 19–22. <https://doi.org/10.1111/j.1365-2222.1995.tb00413.x>
- Kortenkamp, A., Faust, M., 2018. Regulate to reduce chemical mixture risk. *Science* 361, 224–226. <https://doi.org/10.1126/science.aat9219>
- Lange, P., Parner, J., Vestbo, J., Schnohr, P., Jensen, G., 1998. A 15-year follow-up study of ventilatory function in adults with asthma. *N. Engl. J. Med.* 339, 1194–1200. <https://doi.org/10.1056/NEJM199810223391703>
- Le Moual, N., Varraso, R., Siroux, V., Dumas, O., Nadif, R., Pin, I., Zock, J.-P., Kauffmann, F., 2012. Domestic use of cleaning sprays and asthma activity in females. *Eur Respir J* 40, 1381–1389. <https://doi.org/10.1183/09031936.00197611>
- Li, Z., Kesse-Guyot, E., Dumas, O., Garcia-Aymerich, J., Leynaert, B., Pison, C., Le Moual, N., Romieu, I., Siroux, V., Camargo, C.A., Nadif, R., Varraso, R., 2017. Longitudinal study of diet quality and change in asthma symptoms in adults, according to smoking status. *Br J Nutr* 117,

562–571. <https://doi.org/10.1017/S0007114517000368>

Miller, M.R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Crapo, R., Enright, P., van der Grinten, C.P.M., Gustafsson, P., Jensen, R., Johnson, D.C., MacIntyre, N., McKay, R., Navajas, D., Pedersen, O.F., Pellegrino, R., Viegi, G., Wanger, J., ATS/ERS Task Force, 2005. Standardisation of spirometry. *Eur. Respir. J.* 26, 319–338. <https://doi.org/10.1183/09031936.05.00034805>

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>

Patel, C.J., Bhattacharya, J., Butte, A.J., 2010. An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *PLoS One* 5, e10746. <https://doi.org/10.1371/journal.pone.0010746>

Pekkanen, J., Hyvarinen, A., Haverinen-Shaughnessy, U., Korppi, M., Putus, T., Nevalainen, A., 2007. Moisture damage and childhood asthma: a population-based incident case-control study. *Eur Respir J* 29, 509–515. <https://doi.org/10.1183/09031936.00040806>

Polosa, R., Thomson, N.C., 2013. Smoking and asthma: dangerous liaisons. *Eur Respir J* 41, 716–726. <https://doi.org/10.1183/09031936.00073312>

Quanjer, P.H., Stanojevic, S., Cole, T.J., Baur, X., Hall, G.L., Culver, B.H., Enright, P.L., Hankinson, J.L., Ip, M.S.M., Zheng, J., Stocks, J., the ERS Global Lung Function Initiative, 2012. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 40, 1324–1343. <https://doi.org/10.1183/09031936.00080312>

Romieu, I., Sienra-Monge, J.J., Ramírez-Aguilar, M., Téllez-Rojo, M.M., Moreno-Macías, H., Reyes-Ruiz, N.I., del Río-Navarro, B.E., Ruiz-Navarro, M.X., Hatch, G., Slade, R., Hernández-Avila, M., 2002. Antioxidant supplementation and lung functions among children with asthma exposed to high levels of air pollutants. *Am. J. Respir. Crit. Care Med.* 166, 703–709. <https://doi.org/10.1164/rccm.2112074>

Santos, S., Maitre, L., Warembourg, C., Agier, L., Richiardi, L., Basagaña, X., Vrijheid, M., 2020. Applying the exposome concept in birth cohort research: a review of statistical approaches. *Eur J Epidemiol* 35, 193–204. <https://doi.org/10.1007/s10654-020-00625-4>

Sbihi, H., Koehoorn, M., Tamburic, L., Brauer, M., 2017. Asthma Trajectories in a Population-based Birth Cohort. Impacts of Air Pollution and Greenness. *Am J Respir Crit Care Med* 195, 607–613. <https://doi.org/10.1164/rccm.201601-0164OC>

Schwartz, D., Collins, F., 2007. MEDICINE: Environmental Biology and Human Disease. *Science* 316, 695–696. <https://doi.org/10.1126/science.1141331>

Siroux, V., Basagana, X., Boudier, A., Pin, I., Garcia-Aymerich, J., Vesin, A., Slama, R., Jarvis, D., Anto, J.M., Kauffmann, F., Sunyer, J., 2011. Identifying adult asthma phenotypes using a clustering approach. *European Respiratory Journal* 38, 310–317. <https://doi.org/10.1183/09031936.00120810>

- Siroux, V., Boudier, A., Dolgopoff, M., Chanoine, S., Bousquet, J., Gormand, F., Just, J., Le Moual, N., Nadif, R., Pison, C., Varraso, R., Matran, R., Pin, I., 2016. Forced midexpiratory flow between 25% and 75% of forced vital capacity is associated with long-term persistence of asthma and poor asthma outcomes. *J Allergy Clin Immunol* 137, 1709-1716.e6. <https://doi.org/10.1016/j.jaci.2015.10.029>
- Strachan, D.P., Cook, D.G., 1998. Health effects of passive smoking. 6. Parental smoking and childhood asthma: longitudinal and case-control studies. *Thorax* 53, 204–212.
- Tamayo-Uria, I., Maitre, L., Thomsen, C., Nieuwenhuijsen, M.J., Chatzi, L., Siroux, V., Aasvang, G.M., Agier, L., Andrusaityte, S., Casas, M., de Castro, M., Dedele, A., Haug, L.S., Heude, B., Grazuleviciene, R., Gutzkow, K.B., Krog, N.H., Mason, D., McEachan, R.R.C., Meltzer, H.M., Petravičienė, I., Robinson, O., Roumeliotaki, T., Sakhi, A.K., Urquiza, J., Vafeiadi, M., Waiblinger, D., Warembourg, C., Wright, J., Slama, R., Vrijheid, M., Basagaña, X., 2019. The early-life exposome: Description and patterns in six European countries. *Environ Int* 123, 189–200. <https://doi.org/10.1016/j.envint.2018.11.067>
- Tarlo, S.M., Lemiere, C., 2014. Occupational Asthma. *N Engl J Med*, zock 370, 640–649. <https://doi.org/10.1056/NEJMra1301758>
- Temam, S., Varraso, R., Pornet, C., Sanchez, M., Affret, A., Jacquemin, B., Clavel-Chapelon, F., Rey, G., Rican, S., Le Moual, N., 2017. Ability of ecological deprivation indices to measure social inequalities in a French cohort. *BMC Public Health* 17, 956. <https://doi.org/10.1186/s12889-017-4967-3>
- Varraso, R., Camargo, C.A., 2016. Diet and asthma: need to account for asthma type and level of prevention. *Expert Rev Respir Med* 10, 1147–1150. <https://doi.org/10.1080/17476348.2016.1240033>
- Vignoud, L., Pin, I., Boudier, A., Pison, C., Nadif, R., Le Moual, N., Slama, R., Makao, M.N., Kauffmann, F., Siroux, V., 2011. Smoking and asthma: Disentangling their mutual influences using a longitudinal approach. *Respir Med* 105, 1805–1814. <https://doi.org/10.1016/j.rmed.2011.07.005>
- Vineis, P., 2018. From John Snow to omics: the long journey of environmental epidemiology. *Eur J Epidemiol* 33, 355–363. <https://doi.org/10.1007/s10654-018-0398-4>
- White, I.R., Royston, P., Wood, A.M., 2011. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 30, 377–399. <https://doi.org/10.1002/sim.4067>
- WHO Guidelines for Indoor Air Quality: Dampness and Mould, 2009. , WHO Guidelines Approved by the Guidelines Review Committee. World Health Organization, Geneva.
- Wild, C.P., 2005. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol Biomarkers Prev* 14, 1847–1850. <https://doi.org/10.1158/1055-9965.EPI-05-0456>
- Wright, L.S., Phipatanakul, W., 2014. Environmental remediation in the treatment of allergy and asthma: latest updates. *Curr Allergy Asthma Rep* 14, 419. <https://doi.org/10.1007/s11882-014-0419-7>

Zock, J.-P., Plana, E., Jarvis, D., Antó, J.M., Kromhout, H., Kennedy, S.M., Künzli, N., Villani, S., Olivieri, M., Torén, K., Radon, K., Sunyer, J., Dahlman-Hoglund, A., Norbäck, D., Kogevinas, M., 2007. The Use of Household Cleaning Sprays and Adult Asthma: An International Longitudinal Study. *Am J Respir Crit Care Med* 176, 735–741. <https://doi.org/10.1164/rccm.200612-1793OC>

Figure captions

Fig. 1 Workflow of the study

Fig. 2 Flowchart of the population

Fig. 3 Heatmap of the between-exposure correlations

Fig. 4 Heatmap of the absolute within- and between-families of exposure correlations

The diagonal of the heat map assesses the median of the 2-way absolute correlations between exposures variables belonging to the same family (within-family correlations) while the off-diagonal assesses the median of the absolute correlation between each exposure variable of two different family of exposures (between-family correlations).

Table 1. Main characteristics of the studied population (n=599)

Variable	Number of individuals with missing values	Range (min - max)	Mean $\pm$ SD or n (%)
Age (years)	0	16 - 79	39 $\pm$ 16
Height (cm)	0	145 - 192	172 $\pm$ 9
Gender	0		
Female, n (%)			288 (48)
Male, n (%)			311 (52)
Socio-professional category	3		
No profession, n (%)			84 (14)
Executive, n (%)			189 (32)
Technician, n (%)			254 (42)
Farmer/Labourer, n (%)			72 (12)
Centre	0		
Paris, n (%)			184 (31)
Lyon, n (%)			94 (16)
Marseille, n (%)			89 (15)
Montpellier, n (%)			47 (8)
Grenoble, n (%)			185 (31)

SD: standard deviation

Table 2. Description of the lifestyle, social and environmental factors of the studied population.

Family of exposures	Exposure	Number of individuals with missing data	Transformation	Range (min-max)	Mean $\pm$ SD	n (%)
Indoor air	Mold in the bedroom/living-room/bathroom	16	None			45 (8)
	Water damage in the last 12 months	14	None			59 (10)
	Use of gaz for cooking	13	None			397 (68)
Tobacco	Smoking status	0	None			
	Never smoker					301 (50)
	Former smoker					151 (25)
	Active smoker					147 (25)
	Number of pack-years	3	None	0.0 - 111.7	5.1 $\pm$ 10.9	
	0					301 (50)
	< 10					198 (33)
	10 - 20					51 (9)
	>20					46 (8)
	Passive smoking	1	None			
Season	None					257 (43)
	Intermediate					162 (27)
	High					179 (30)
	Season of examination	0	None			
	Summer					138 (23)
Pets	Autumn					168 (28)
	Winter					139 (23)
	Spring					154 (26)
Pets	Having cat/dog currently	4	None			244 (41)
	Having cat/dog in childhood	0	None			382(64)

Occupational exposure to asthmagens	Occupational exposure to asthmagens ever	45	None		
	None				271 (49)
	Low				116 (21)
	High				167 (30)
	Occupational exposure in the last job	46	None		
	None				400 (72)
	Low				73 (13)
	High				80 (14)
Cleaning products	Use of bleach	9	None		
	<1 day/week				458 (78)
	1 to 3 days/week				96 (16)
	4 to 7 days/week				36 (6)
	Use of sprays	139	None		
	None				288 (63)
	1 spray/week				91 (20)
	≥2 sprays/week				81 (17)
Rurality	Born in a rural municipality	9	None		132 (22)
	Ever-living in a rural municipality	10	None		327 (56)
Social	Social Deprivation Index	6	None		
	1st quintile				297 (50)
	2nd quintile				122 (21)
	3rd quintile				78 (13)
	4th quintile				64 (11)
	5th quintile				32 (5)
Physical activity and diet	Physical activity (METs)	67	Dichotomized	66 - 236	128 ± 40
	Diet (AHEI-2010 score)	99	None	24 - 79	50 ± 10
Altitude	Altitude, m	0	Log ₂	1 - 1436	169 ± 159

Building and population densities	Building density in a 100m buffer	0	Square root	0.00 - 1.00	0.29 ± 0.20
	Building density in a 300m buffer	0	Square root	0.00 - 0.68	0.22 ± 0.15
	Population density (in a 200x200m area)	15	Log ₂	1 - 2718	382 ± 475
Pollution*	NO ₂ , µg/m ³	0	None	3.8 - 50.1	29.1 ± 9.7
	PM _{2.5} , µg/m ³	0	None	3.6 - 27.4	18.1 ± 4.5
	PM ₁₀ , µg/m ³	0	None	9.7 - 44.1	27.8 ± 5.6
	O ₃ , µg/m ³	0	None	74.0 - 127.7	102.0 ± 11.7
Natural spaces	NDVI in a 100m buffer	0	None	0.02 - 0.75	0.37 ± 0.13
	NDVI in a 300m buffer	0	None	0.06 - 0.73	0.39 ± 0.13
	NDVI in a 500m buffer	0	None	0.05 - 0.73	0.40 ± 0.13
	Number of green spaces in a 300m buffer	0	0 / ≥1	0 - 30	4.3 ± 7.2
	Number of blue spaces in a 300m buffer	0	0 / ≥1	0 - 13	0.22 ± 1.3
	Distance to nearest green space, m	0	Log ₂	7 - 5227	876 ± 1110
	Distance to nearest blue space, m	0	Log ₂	57 - 37633	4470 ± 5055
	Shannon Evenness Index [†] in a 300m buffer	0	None	0.00 - 0.37	0.11 ± 0.09
Weather*	Wind, m/s	0	Dichotomized	0.8 - 4.8	2.5 ± 1.0
	Relative humidity, %	0	None	62 - 87	73 ± 5
	Absolute humidity, g/m ³	0	None	4.1 - 9.9	8.2 ± 0.7
	Minimal temperature, °C	0	None	1.0 - 13.9	8.5 ± 1.7
	Mean temperature, °C	0	None	5.4 - 16.7	12.3 ± 1.6
	Maximal temperature, °C	0	None	11.6 - 22.3	18.0 ± 1.8
Roads	Distance to nearest major road, m	0	Log ₂	6 - 44594	2530 ± 4954
	Number of major roads in a 100m buffer	0	0 / ≥1	0 - 8	0.1 ± 0.5
	Number of cycle paths in a 100m buffer	0	0 / ≥1	0 - 6	0.2 ± 0.9
UV*	Erythral UV dose, kJ/m ²	0	None	1.9 - 2.4	2.2 ± 0.2
	Vitamin-D UV dose, kJ/m ²	0	None	3.2 - 4.2	3.8 ± 0.3
	DNA-damage UV dose, kJ/m ²	0	None	0.8 - 1.1	0.9 ± 0.1

Health services [‡]	Number of general practitioners	0	Dichotomized	0 - 42	2.4 ± 4.5
	Number of pulmonologists	0	Dichotomized	0 - 5	0.1 ± 0.5
	Number of other physicians	0	Dichotomized	0 - 77	4.1 ± 8.3
	Number of paramedics	0	Dichotomized	0 - 98	10.1 ± 12.2
	Number of pharmacies	0	Dichotomized	0 - 8	0.8 ± 1.2
	Number of other health services	0	Dichotomized	0 - 12	0.8 ± 1.5

SD: standard deviation ; AHEI: Alternative Healthy Eating Index; METs: Metabolic Equivalent of Tasks ; NO₂: nitrogen dioxide ; PM_{2.5}: particles matters with aerodynamical diameter less than 2.5 µm ; PM₁₀: particles matters with aerodynamical diameter less than 10 µm; O₃: ozone; NDVI: normalized difference vegetation index; UV: ultraviolet

*: Daily values averaged on the 365 days before performing the spirometry; †: index of area composition and richness; ‡: number of health services in the IRIS (“Ilots Regroupés pour l’Information Statistique”) where an IRIS represents an area covering between 1,800 and 5,000 inhabitants.

Table 3. Sociodemographic characteristics and exposures of the 3 clusters obtained by Bayesian Profile Regression supervised on FEV₁ (%predicted).

	Cluster 1	Cluster 2	Cluster 3		
	n = 30	n = 115	n = 454	p	p adjusted [‡]
Age (years)	51 ± 14 [†]	48 ± 13 [†]	36 ± 16	< 0.001	-
Height (cm)	170 ± 10	169 ± 8	170 ± 9	0.52	0.68
Gender				0.12	-
Female, n(%)	9 (30)	55 (48)	224 (49)		
Male, n(%)	21 (70)	60 (52)	230 (52)		
Socio-professional category				< 0.001	0.22
No profession, n(%)	0 (0)	2 (2) [†]	82 (18)		
Executive, n(%)	10 (33)	43 (37) [†]	136 (30)		
Technician, n(%)	17 (57)	49 (43) [†]	188 (41)		
Farmer/Labourer, n(%)	3 (10)	21 (18) [†]	48 (11)		
Centre				< 0.001	-
Paris, n(%)	16 (53) ^{*†}	20 (17) [†]	148 (33)		
Lyon, n(%)	8 (27) ^{*†}	25 (22) [†]	61 (13)		
Marseille, n(%)	4 (13) ^{*†}	22 (19) [†]	63 (14)		
Montpellier, n(%)	1 (3) ^{*†}	12 (10) [†]	34 (7)		
Grenoble, n(%)	1 (3) ^{*†}	36 (31) [†]	148 (33)		
Season of examination				0.23	0.17
Spring, n(%)	13 (43)	33 (29)	108 (24)		
Summer, n(%)	6 (20)	25 (22)	107 (24)		
Autumn, n(%)	8 (27)	28 (24)	132 (29)		
Winter, n(%)	3 (10)	29 (25)	107 (24)		
Number of pack-years	23 ± 28 [†]	17 ± 10 [†]	1 ± 2	< 0.001	< 0.001
Passive smoking				0.005	0.17
None, n(%)	18 (60) [†]	45 (39) [†]	194 (43)		

Intermediate, n(%)	9 (30) [†]	43 (37) [†]	111 (24)		
High, n(%)	3 (10) [†]	27 (23) [†]	149 (33)		
Born in a rural municipality, n(%)	7 (23)	29 (25)	97 (21)	0.67	0.98
AHEI-2010 score	45 ± 11 [*]	53 ± 10 [†]	49 ± 9	< 0.001	< 0.001
FDep				0.13	0.09
1st quintile, n(%)	14 (47)	50 (43)	235 (52)		
2nd quintile, n(%)	8 (27)	19 (17)	96 (21)		
3rd quintile, n(%)	6 (20)	19 (17)	54 (12)		
4th quintile, n(%)	2 (7)	16 (14)	47 (10)		
5th quintile, n(%)	0 (0)	11 (10)	22 (5)		
Mean annual humidity (%)	75 ± 5 [*]	72 ± 4 [†]	73 ± 5	0.003	0.07
Mean annual wind ≥ 2 m/s, n(%)	29 (97) ^{*†}	75 (65)	322 (71)	0.003	0.11
Mean annual temperature (°C)	12.0 ± 1.3	12.2 ± 2.0	12.3 ± 1.5	0.63	0.01
≥1 pharmacy, n(%)	3 (10) ^{*†}	58 (50)	234 (52)	< 0.001	0.002
Distance to the nearest blue space (km)	7.3 ± 5.8 [†]	5.3 ± 6.4 [†]	4.1 ± 4.5	< 0.001	< 0.001
Distance to the nearest green space (km)	0.7 ± 1.0	0.8 ± 1.1	0.9 ± 1.1	0.56	0.04
Population density in a 200*200m area	330 ± 605	337 ± 409	388 ± 478	0.51	0.11
Distance to the nearest major road (km)	3.5 ± 6.7	4.5 ± 9.1 [†]	2.0 ± 2.7	< 0.001	< 0.001
≥1 major road in a 100m buffer, n(%)	7 (23) ^{*†}	0 (0)	2 (0)	< 0.001	< 0.001
≥1 cycle path in a 100m buffer, n(%)	1 (3)	13 (11)	35 (8)	0.33	0.43

FEV1: Forced expiratory volume in 1s ; FVC: forced vital capacity ; AHEI: Alternative Healthy Eating Index ; FDep: French Deprivation Index.

*: p < 0.05 vs cluster 2 ; [†]: p < 0.05 vs cluster 3 ; [‡]: p-value adjusted on age, gender and centre

Table 4. Phenotypic characteristics at EGEA2 and EGEA3 according to the 3 environmental clusters obtained by the supervised Bayesian Profile Regression

Phenotypic characteristics	EGEA2					EGEA3				
	n	Cluster 1 n = 30	Cluster 2 n = 115	Cluster 3 n = 454	p	n	Cluster 1 n = 22	Cluster 2 n = 97	Cluster 3 n = 383	p
Use of inhaled corticosteroids in the last 12 months, n (%)	59	17	41	181	0.1	40	8	36	143	0.
	4	(57%)	(36%)	(40%)	1	1	(50%)	(47%)	(46%)	95
Asthma exacerbation [‡] in the last 12 months, n (%)	55	1 (3%)	14	62	0.2	38	6	14	62	0.
	7		(13%)	(15%)	2	8	(35%)	(18%)	(21%)	30
Current asthma [§] , n (%)	55	26	95	371	0.7	45	13	55	221	0.
	6	(87%)	(90%)	(88%)	6	6	(65%)	(60%)	(64%)	81
Asthma symptoms score in the past 12 months , m±SD	59	1.7 ±	1.9 ±	2.0 ±	0.7	46	1.5 ±	1.8 ±	1.6 ±	0.
	4	1.4	1.5	1.6	3	6	1.3	1.5	1.5	50
Total Asthma Quality of Life Questionnaire score, m±SD	53	6.0 ±	6.2 ±	6.2 ±	0.2	35	6.0 ±	6.3 ±	6.3 ±	0.
	5	0.9	0.9	0.8	8	1	0.7	0.8	0.8	46
Allergic sensitization ^{**} , n (%)	56	18	75	344	0.0					
	0	(60%) [†]	(69%) [†]	(82%)	01					
Blood neutrophils (%), m±SD	59	62.6 ±	58.6 ±	58.1 ±	0.0					
	1	9.8 [†]	8.4	8.9	3					
Blood eosinophils (%), m±SD	59	3.3 ±	3.2 ±	3.8 ±	0.1					
	1	2.5	2.5	2.8	1					

*: p < 0.05 vs cluster 2; †: p < 0.05 vs cluster 3; ‡: exacerbation was defined by use of oral steroids or hospitalization or emergency visit for asthma in the last 12 months; §: current asthma defined by the presence of ≥1 asthma attacks or use of asthma treatment in the past 12 months; ||: Asthma symptoms score as defined by Sunyer et al [38] by the sum of five symptoms over the past 12 months (wheezing or whistling in the chest, woken up with a feeling of chest tightness, attack of shortness of breath at rest, attack of shortness of breath that came on following strenuous activity, woken by an attack of shortness of breath; **: allergic sensitization was defined by >1 positive skin prick test response to any of the 12 allergens (cat, *Dermatophagoides pteronyssinus*, *Blattella germanica*, olive, birch, *Parietaria judaica*, timothy grass, ragweed pollen, *Aspergillus species*, *Cladosporium herbarum*, *Alternaria tenuis*, and cypress)

Figure 1.

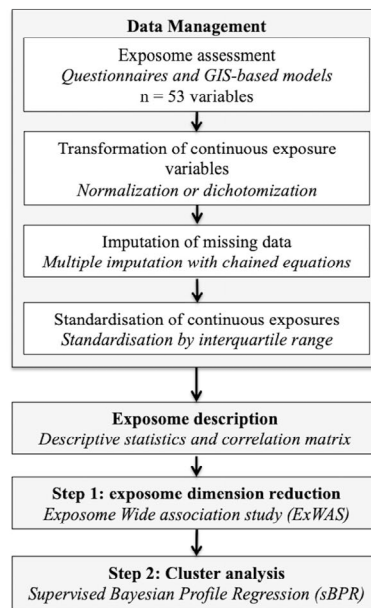


Figure 2.

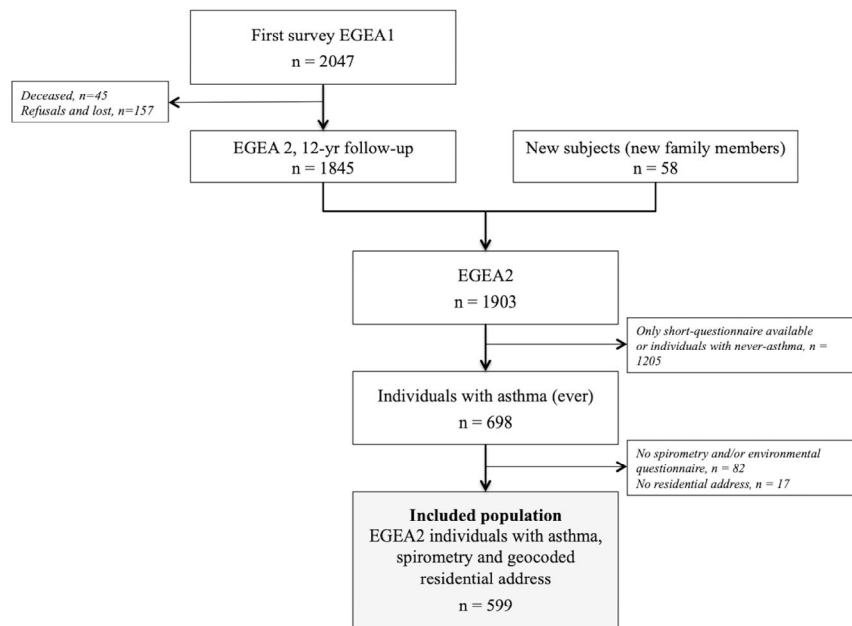


Figure 3.

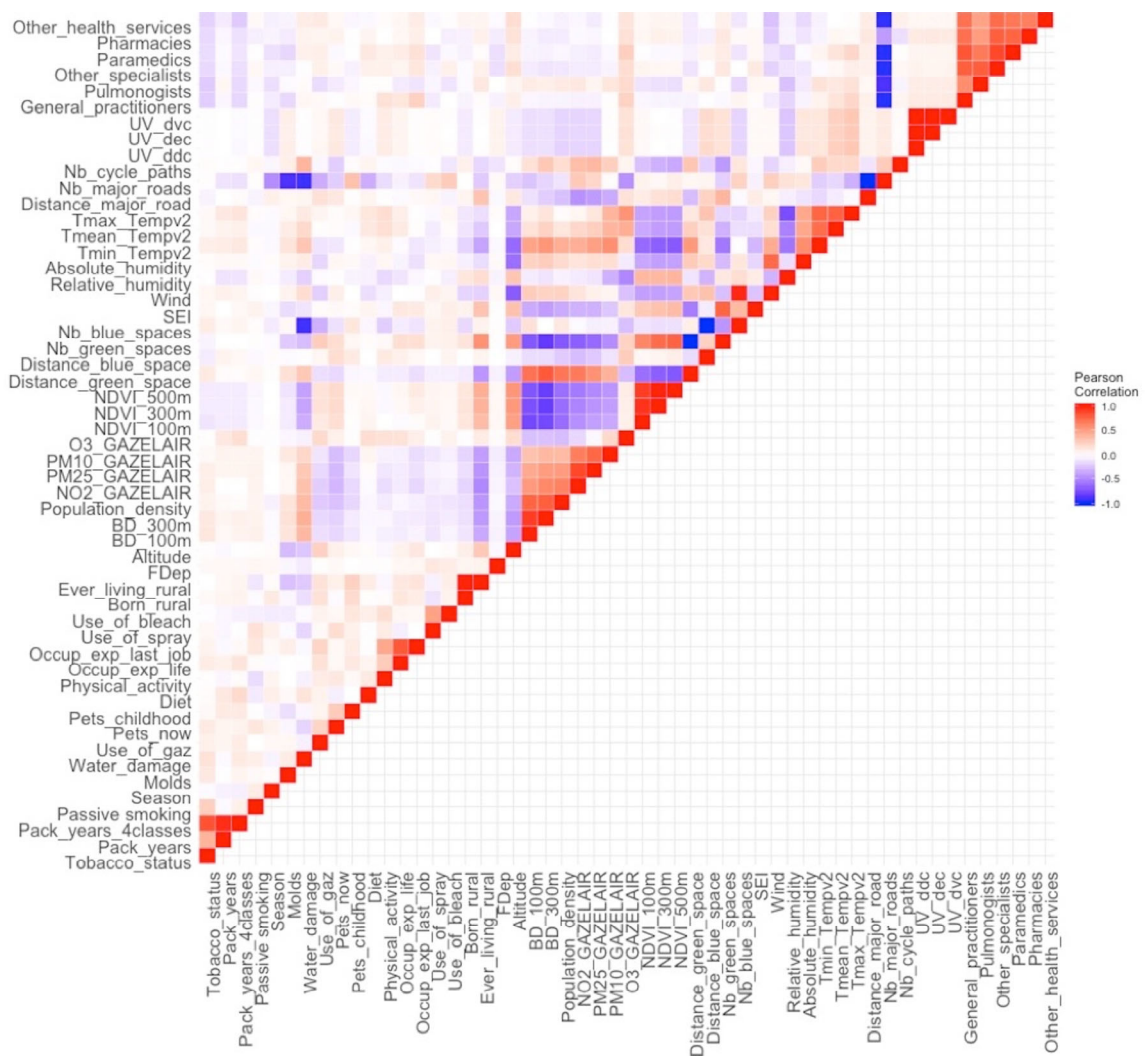
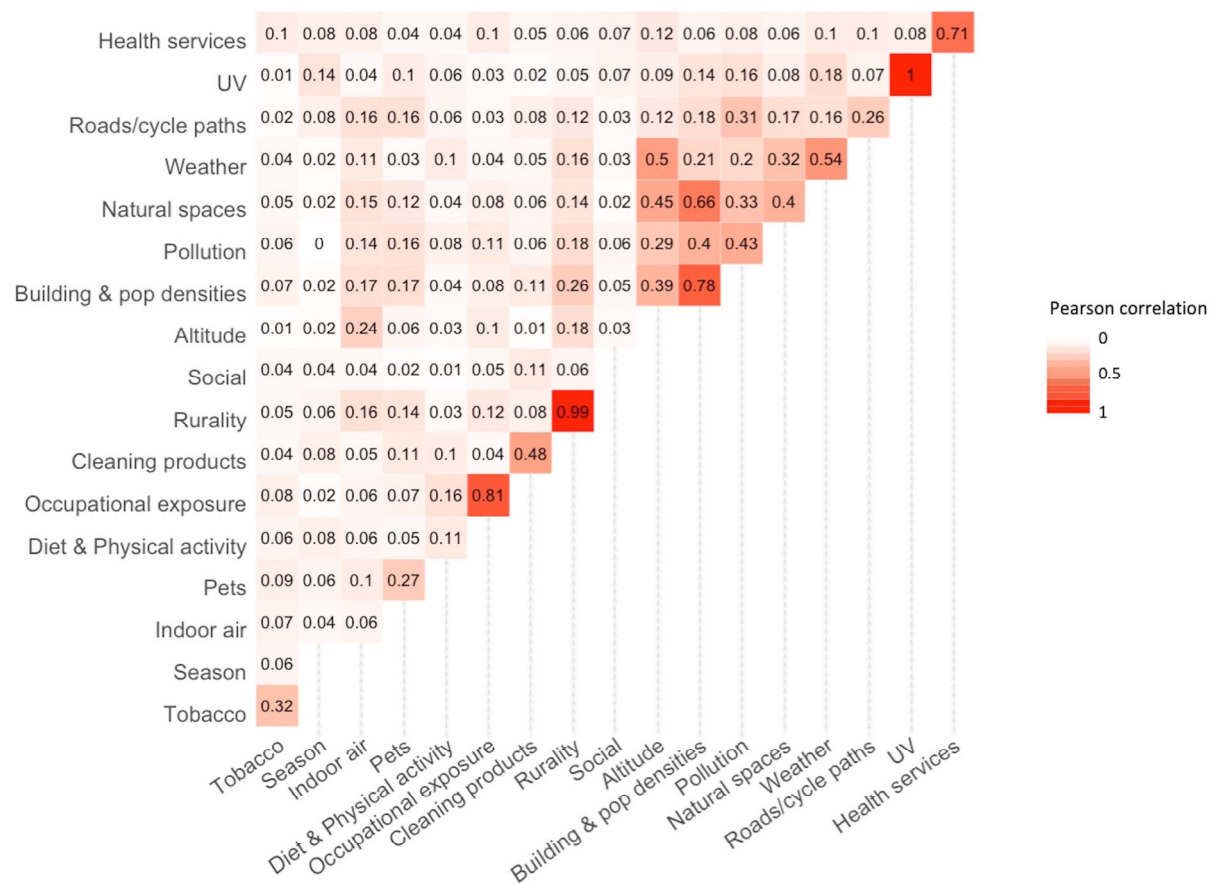
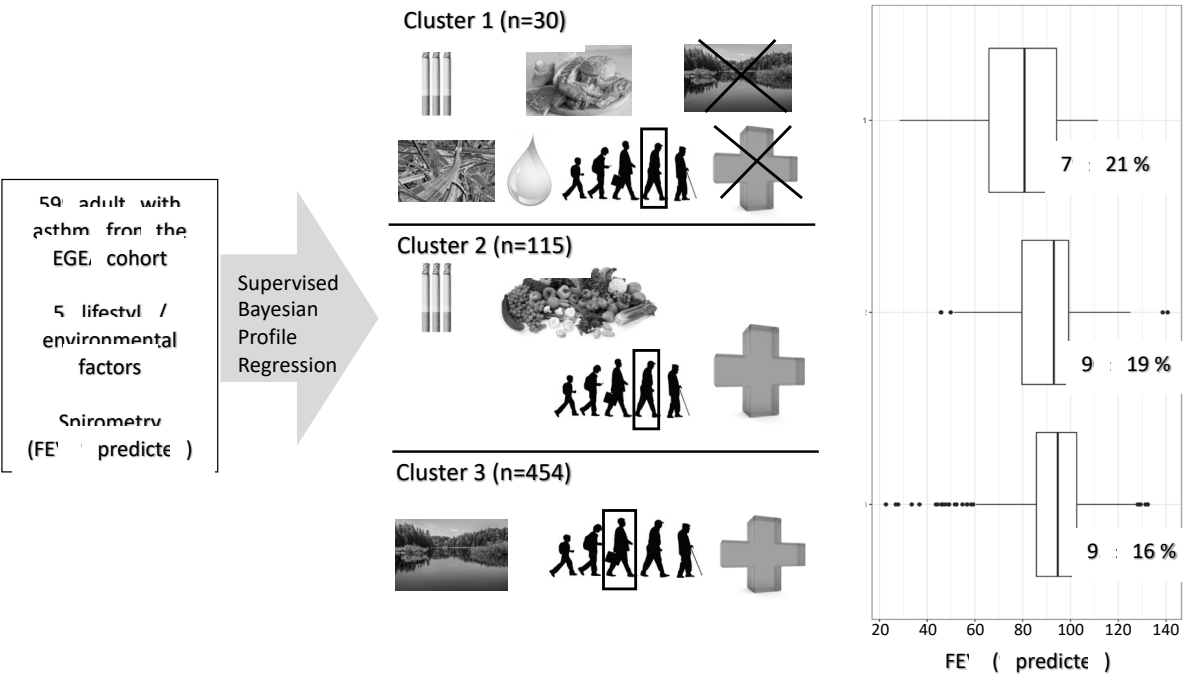


Figure 4.



Graphical abstract



Online Data Supplement

Profile of exposures and lung function in adults with asthma: an exposome approach in the EGEA study

Alicia Guillien¹, Johanna Lepeule¹, Emie Seyve¹, Nicole Le Moual^{2,3}, Isabelle Pin^{1,4}, Bruno Degano⁵, Judith Garcia-Aymerich^{6,7,8}, Jean-Louis Pépin⁵, Christophe Pison⁹, Orianne Dumas^{2,3}, Raphaëlle Varraso^{2,3}, Valérie Siroux¹

Correspondence:

Alicia Guillien

Institut pour l'Avancée des Biosciences

Equipe d'épidémiologie environnementale appliquée à la reproduction et à la santé
respiratoire

Centre de Recherche UGA / Inserm U 1209 / CNRS UMR 5309

Site Santé – Allée des Alpes

38700 La Tronche

Tel: +33 4 76 54 94 69

Email: alicia.guillien@univ-grenoble-alpes.fr

Online methods 41

Part I. Study population and spirometry	41
Study population.....	41
Spirometry testing	41
Part II. Exposome assessment method.....	42
Lifestyle and social factors	42
Indoor air.....	42
Tobacco	42
Season	42
Pets	43
Occupational exposure to asthmagens.....	43
Cleaning products	43
Rurality	43
Social	43
Physical activity and diet.....	43
Environmental exposures	45
Altitude	45
Building and population densities.....	45
Atmospheric pollutants.....	45
Natural spaces.....	45
Weather	46
Roads/cycle paths	47
Ambient ultra-violet (UV) radiation	47
Health services	47
Part III. Statistical analysis	48
Transformation of exposure variables.....	48
Imputation of missing data	48
Standardisation of exposure variables	48
Exposome description.....	48
Exposome dimension reduction from ExWAS results.....	48
Identification of profiles of joint exposures associated with FEV ₁ % using sBPR	49
References.....	50

Supplemental Tables 53

Supplemental Table E1. Characteristics of included and excluded individuals with asthma.	53
Supplemental Table E2. Comparisons of exposures by inclusion centre.....	54
Supplemental Table E3. Results of the ExWAS analyses assessing associations between each exposure variable and FEV ₁ (%predicted)	58
Supplemental Table E4. Characteristics of the studied population according to their participation to the EGEA3 follow-up about 7 years later.....	61

Online Figures 63

Part I. Study population and spirometry

Study population

EGEA is a French cohort based on three surveys over 20 years. Initially (EGEA1 study), 388 cases with asthma were recruited in the early 1990s in five cities (Paris, Lyon, Marseille, Montpellier and Grenoble), so as their first-degree relatives (n=1244) and 415 population-based controls. Two follow-ups were conducted in 2003-2007 (EGEA2) and 2011-2013 (EGEA3). At each survey, individuals responded to a detailed questionnaire on respiratory health and environment, based on international standardised tools, as previously described (Kauffmann et al., 1997; Kauffmann and Dizier, 1995). The EGEA collection and program management was certified ISO 9001 from 2006 to 2018.

Because part of geographic information systems (GIS)-based models did not fit for the EGEA1 study period and spirometry was not part of EGEA3 protocol, the present study focuses on adults with ever asthma at EGEA2. Ever-asthma was defined by a positive answer to either “Have you ever had attacks of breathlessness at rest with wheezing?” or “Have you ever had asthma attacks?” or being recruited as an asthma case in chest clinics.

Spirometry testing

Spirometry was performed using a flow-volume spirometer (SpiroDyn'R; SAS Dyn'R, Aix-en-Provence, France) according to the American Thoracic Society and European Respiratory Society guidelines (Miller et al., 2005). Standardized operating procedures were implemented and controlled, including calibration of all equipment before each measurement, and quality control visits by study coordinators to ensure correct protocols by field staff. Participants were asked to perform three to nine attempts to provide at least three technically acceptable spirometry manoeuvres. Pre-bronchodilator forced expiratory volume in 1 s (FEV₁) was expressed in percentage of the predicted value using the Global Lung Function Initiative (GLI-2012) reference equations, thereafter referred as FEV₁% (Quanjer et al., 2012).

Part II. Exposome assessment method

In total, 53 exposure variables (19 lifestyle or social factors and 34 environmental factors) were assessed.

Lifestyle and social factors

Data regarding behavioural factors were assessed mostly through self-completed questionnaires and provided information about 9 families of exposure: indoor air (3 variables), tobacco exposure (4 variables), season of examination (season at which individual performed the spirometry, 1 variable), pets (2 variables), occupational exposure (2 variables, assessed through a job-exposure matrix), household cleaning products (2 variables), rurality (2 variables), socio-economic status (SES) (assessed at the contextual level by the French Deprivation Index (FDep) from geocoded residential address, built to capture health inequalities at ecological level (Temam et al., 2017), 1 variable), and physical activity and diet (2 variables, assessed through the Metabolic Equivalent Tasks (METs) and the 2010 Alternative Healthy Eating Index (AHEI-2010) score, respectively).

Indoor air

Three variables were collected to characterise indoor air. **Water damage in the last 12 months** was defined by a positive answer to the question: “Has there been water damage in your building in the last 12 months”. Presence of **moulds in the bedroom/living-room/bathroom** was defined by positive answers to the questions: “Has there ever been any mould on any surface inside the house (food excluded)?” and “has this happened in the last 12 months?” and if the room was bedroom, living-room and/or bathroom. **Use of gaz for cooking** was defined by the answer “gas” to the question: “Currently, what is the fuel used for cooking?”.

Tobacco

Detailed information on smoking history was collected, included the number of cigarettes/pipe/cigars/cigarillos smoked per day for different time periods if the consumption of tobacco smoking has changed over time. From these data the number of **pack-years** (as a quantitative and divided in 3 classes (0; >0-10; >10-20; >20) was calculated and the **smoking status** was defined: never smokers were defined as those never smoking or having smoked on average less than one cigarette, cigar, cigarillos or one pipe a day for a year. Former smokers had stopped smoking since at least a month. No exposure to **passive smoking** was defined as exposure to tobacco smoke for less than one hour a day (all sources of exposure combined), intermediate exposure was defined as reporting exposure to tobacco smoke at home or work less than one hour a day and high exposure was defined as being exposed >1h a day (all sources of exposure combined).

Season

Season of examination was assessed from the date at which individuals performed spirometry and defined in 4 categories: spring (March 20th to June 20th), summer (June 21th to September 22th), autumn (September 23th to December 21th) and winter (December 22th and March 19th).

Pets

Having cat/dog currently was defined by at least one positive answer to the following questions: “Do you have a cat?” and “Do you have a dog?”. **Having cat/dog in childhood** was defined by a positive answer to the question: “In childhood, did you have a pet?” and have chosen the answer “cat” or “dog” to the question: “Was it a dog? A cat? A rodent? A bird? Other?”.

Occupational exposure to asthmagens

Occupational exposures to 18 asthmagens (high or low molecular weight agents, mixed environment) and to low risk agents (possible low level exposure to irritants or antigens) were evaluated at EGEA2 through an asthma-specific Job-Exposure Matrix (Kennedy, 2000) for both the last job and all occupational history. In addition, an expert re-evaluation step was performed to improve occupational exposure evaluation (Dumas et al., 2011).

Cleaning products

The variable **use of bleach** was defined according to the answer at the question: “how many days a week did you use bleach?” where the possible answers were: “never; <1 day/week; 1-3 days/week; 4-7 days/week”. For the present analysis, this variable was recorded in the three following categories: “<1 day/week; 1-3 days/week; 4-7 days/week” (Matulonga et al., 2016). **Use of sprays** was defined by the exposure to any of the eight types of sprays (furniture, glass-cleaning, carpet, mopping the floor, oven, ironing, air-refreshing, other use) and was composed of three categories: “none; 1/week; ≥ 2 /week” (Le Moual et al., 2012).

Rurality

Be **born in a rural municipality** was defined as those reporting having lived in a municipality of less than 5000 inhabitants at birth.

Using data from the General Agricultural Census (GAC), the commune at birth was categorised as rural when the number of inhabitants was lower than 5,000 (to have a sensitive definition of rural).

Ever-living in a rural municipality was defined as those reporting having lived in a municipality of less than 5000 inhabitants during their life. These variables were calculated from the zip code of each municipality the participants reported to have lived since birth.

Social

The **French Deprivation Index** (FDep), defined at the geographical IRIS level (division of the French territory of homogeneous size defined by 2000 inhabitants per unit), integrates four dimensions: the median income per unit of consumption in the household, the percentage of graduates in the population over 15, the percentage of workers in the labor force and the unemployment rate. Finally, this index was divided into 5 quintiles from a large french cohort (E3N: Etude épidémiologique auprès des femmes de la Mutuelle Générale de l'Education Nationale) (Temam et al., 2017).

Physical activity and diet

Dietary data were collected at EGEA2 using a 118-item self-administered food frequency questionnaire (FFQ), which is based on a validated French questionnaire (Bonifacj et al., 1997). Participants were asked to indicate their frequency of consumption of all food items by checking one of eight given categories, including “never or < 1 time/month”, “1–3

times/month”, “1 time/week”, “2–4 times/week”, “5–6 times/week”, “1 time/day” 2–3 times/day”, and “ ≥ 4 times/day”, for a standard portion size of each item. Standard portion sizes were listed for each food. The selected frequency category for each food item was converted to a daily intake. For example, a response of “1 serving/week” was converted to 0.14 servings/d. Through the use of the French food-composition data from the Vitamin and Antioxidant Element Supplementation (SU.VI.MAX) survey (Hercberg et al., 1998), the average daily intake of nutrients was calculated by multiplying the frequency of consumption of each item by its nutrient content per serving and totaling the nutrient intake for all food items. Out of the 1,601 participants from EGEA2, 35 adolescents (less than 16 years-old) were firstly excluded (no FFQ in adolescents). Among adults ($n = 1,566$), we excluded 109 participants who did not fill the FFQ, 60 who left more than 4 items blank on the diet questionnaire, and 161 participants with an implausibly high ($> 4,200$ kcal/d for men; $> 3,500$ kcal/d for women) or low total (< 800 energy intake kcal/d for men; < 500 kcal/d for women) (Varraso et al., 2012). In total, dietary data were available for 1,236 adults at EGEA2.

The scoring criteria of the **Alternate Healthy Eating Index 2010 (AHEI-2010)** was described in detail elsewhere (Chiuve et al., 2012). Briefly, the AHEI-2010 score includes 11 components. A high AHEI-2010 score denotes high intake of vegetables, fruits, whole grains, nuts and legumes, long-chain Ω -3 fatty acids and polyunsaturated fatty acids; moderate intake of alcohol; and low intake of sugar-sweetened drinks and fruit juice, red/processed meat, *trans* fat, and sodium. For each component, a minimal score of 0 and a maximal score of 10 was allocated based on its intake, and intermediate values scored proportionally. The total AHEI-2010 score sums scores of all the 11 components, ranging from 0 to 110; a higher score represents a healthier diet. In the EGEA study, we calculated a modified AHEI-2010 including all the items except for *trans* fat, which is not available in the French food composition table from the SU.VI.MAX study. Besides, cereal fiber intake was calculated instead of whole grain intake, because only intakes of whole wheat bread and of brown rice have been recorded as whole grains in the EGEA2 study. An intake of 15 g cereal fiber/day was considered as ideal based on the scoring criteria of the Alternate Healthy Eating Index (AHEI). Therefore, the highest value for the AHEI-2010 diet score in the EGEA study was 100 (rather than 110) (Li et al., 2017).

Assessment of habitual physical activity at EGEA2 was based on five questions on the usual distance walked daily (< 500 , $[500-2,000]$, and $\geq 2,000$ m), the average number of flights of stairs climbed daily (0, $[1-4]$, and ≥ 5), the average amount of time spent weekly doing vigorous recreational activity (such as jogging, aerobic, swimming, cycling, tennis, soccer, digging in the garden) (0, $[1-2]$, $[3-4]$ and ≥ 5 hours), moderate recreational activity (such as cleaning, walking, golf, slight gardening) (0, $[1-2]$, $[3-4]$ and ≥ 5 hours) and light recreational activity (such as reading, watching TV, office work, driving, eating) (0, $[1-2]$, $[3-4]$ and ≥ 5 hours), based on the questionnaire from Paffenbarger *et al.* (Paffenbarger et al., 1993). Physical habits at work (at work, you are seated most of the day; at work, you are standing, but your activity does not require any particular physical effort; at work, your activity requires physical efforts (lifting heavy loads, etc.) was also recorded based on the questionnaire from Baecke *et al.* (Baecke et al., 1982). Total physical activity was estimated by multiplying the metabolic equivalent cost (MET) of each activities according to the Compendium of Physical Activities

(Ainsworth et al., 2000) by their frequency and duration, and assessed in **metabolic equivalents per week (METs/w)**.

Environmental exposures

A broad range of environmental exposures was assessed through geographic information systems (GIS) that allow allocating a spatio-temporal estimate of exposure from geographical residential coordinates. Environmental exposures covered 8 families: altitude (1 variable), building and population densities (3 variables), atmospheric pollutants (4 variables), natural spaces (8 exposures), weather (6 variables), roads/cycle paths (3 variables), ambient ultra violet radiation (UV) (3 variables) and health services (6 variables). The GIS-based models are resting on National Geographic Institute (IGN) maps, the French National Institute for Statistics and Economic Research (INSEE) data, Météo-France data, geospatial models, land cover databases and satellite data. All these exposures were calculated from the geocoded residential address.

Altitude

Altitude (m) at the residential address of each participant was assessed from the ALTI database from IGN (“Institut de l’Information Géographique et Forestière”). Data were available for the year 2016 and had a spatial resolution of 25m.

Building and population densities

Building density was calculated within **100- and 300-meters buffers** by dividing the area of building cover (km²) by the area of each buffer (km²). The area of building cover was estimated from the 2016 data of the BDTOPO database (IGN).

Population density around residential address was calculated as the number of inhabitants in an area of **200*200m** surrounding the home address with the INSEE (“Institut Nationale de la Statistique et des Etudes Economiques”) database.

Atmospheric pollutants

Chronic exposure to four atmospheric pollutants (nitrogen dioxide (**NO₂**, g/m³), particulate matter with an aerodynamic diameter of less than 2.5 µm (**PM_{2.5}**, g/m³) and of less than 10 µm (**PM₁₀**, g/m³), and ozone (**O₃**, g/m³)) was assessed using the “GAZELAIR” model (Bentayeb et al., 2014). This model was defined through a geostatistical (Kriging) analysis, including data from the CHIMERE chemistry-transport model, a validated model in Europe (Alonso et al., 2011; Valari et al., 2011) and provides annual data from 1989 to 2008. The 12-month average levels of air pollutants at the participant’s residential address before the lung function test was estimated.

Natural spaces

Number of blue spaces in a 300m buffer, number of green spaces in a 300m buffer, distance to nearest blue space, and the distance to nearest green space were assessed using the 2006 Corine Land Cover (CLC) database (<https://www.geoportail.gouv.fr/donnees/corine-land-cover-2006>). This database provides information about land use in French territory, with a spatial resolution of 100m. Green spaces were defined as “green urban areas” (class 1.4.1) and blue spaces as “water bodies” (class 5). Distances from the nearest blue/green spaces were defined as the distance between residential address and nearest blue/green space identified in

the CLC database. Numbers of blue/green spaces in a 300m buffer were defined as the number of object of classes 5/1.4.1 in a 300m buffer surrounding residential address.

The **Shannon Evenness Index (SEI) in a buffer of 300m** was also calculated from the CLC database. The SEI is the degree of mixing of different types of land uses (such as residential, commercial, entertainment, and office development). A higher value indicates a more even distribution of land between the different types of land uses. This index was calculated as follow:

$$\frac{-\sum_{i=1}^m (p_i * \ln p_i)}{\ln m}$$

with:

p_i = proportion of the area occupied by land use type (class) i

m = number of land use types (classes) present in the study area

The SEI was calculated in a 300m buffer around residential address.

Normalized Difference Vegetation Index (NDVI) quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). NDVI values range from +1.0 to -1.0. Areas of barren rock, sand, or snow usually show very low NDVI values (for example, 0.1 or less). Sparse vegetation such as shrubs and grasslands or senescing crops may result in moderate NDVI values (approximately 0.2 to 0.5). High NDVI values (approximately 0.6 to 0.9) correspond to dense vegetation such as that found in temperate and tropical forests or crops at their peak growth stage. Negative values of NDVI (values approaching -1) correspond to water.

NDVI in a **100m, 300m and 500m buffers** were derived from the LANDSAT data. The original data set had a spatial resolution of 30m and provided daily values of NDVI from 1986 to 2017. We calculated the NDVI as the average of daily values from June to August in the year at which individuals performed spirometry.

Weather

Wind (m/s) and **relative humidity** (%) were calculated from the SAFRAN model (Météo-France). This model had a spatial resolution of 8km and provided daily values. The daily values of the 365 days preceding the day at which individuals performed spirometry were averaged to obtain the mean annual value.

Absolute humidity (g/m³) was calculated as follow:

$$wvp_mb = 6.11 * (10^{(((7.5 * dpt24h)/(237.7 + dpt24h)))))$$

$$ah_gm3 = (1000 * (100 * wvp_mb))/((273.15 + tempc24h) * 461.5)$$

where dpt24h is the dewpoint 24h and tempc24h was the average temperature in the last 24h (in °C)

Minimal, maximal and mean daily temperature were estimated by an updated version (Hough *et al.*, submitted) of a hybrid spatio-temporal model based on Moderate Resolution

Imaging Spectroradiometer (MODIS) satellite surface temperature data (Kloog et al., 2017). Briefly, daily satellite surface temperature data (in 1 km² grid cells) were obtained, and calibrated with air temperature data from Météo-France monitors, with adjustment for spatio-temporal predictors (NDVI, altitude, etc). These daily model predictions were used to generate temperature exposure estimates for every study participant, based on his home address.

Roads/cycle paths

Distance to nearest major road, number of major roads in a 100m buffer and number of cycle paths in a 100m buffer were derived from the BDTOPO database of IGN (<http://professionnels.ign.fr/bdtopo>). Major roads comprised “highways”, “nearly-highways”, “highway ramp” and “dual carriageway”. BDTopo data were derived in 2016.

Ambient ultra-violet (UV) radiation

Daily, weekly and monthly measurements of UV radiation (as **erythemal UV, Vitamin-D and DNA damaging UV**) at home at 0.5 x 0.5 degree resolution were derived from total ozone column measurements by the Global Ozone Monitoring Experiment-2 (GOME-2) instrument onboard the Meteorological Operational satellite program (MetOp) and SCIAMACHY instrument onboard ENVISAT satellite, assimilated into global ozone fields (Temis project, <http://www.temis.nl/uvradiation/UVindex.html>), and averaged over the year before the spirometry examination.

Health services

Health services information were derived from INSEE data: (<https://www.insee.fr/fr/statistiques/3568611?sommaire=3568656>). Data were collected in 2017 at the IRIS (“Ilots Regroupés pour l’Information Statistique”) scale, where an IRIS is composed of nearly 5,000 inhabitants (<https://www.insee.fr/fr/metadonnees/definition/c1523>). This database provided information regarding 42 types of health services. We individually considered presence of “**Medical practitioner**”, “**pulmonologist**” and “**pharmacy**” in the IRIS while we grouped other health services in 3 classes. “**Other specialists**” was defined as the sum of cardiologists, dermatologists/venerologists, gastroenterologists hepatologists, psychiatrists, ophthalmologists, otolaryngologists, paediatricians, radiologists, stomatologists, gynaecologists and dental surgeons in the IRIS. “**Paramedics**” were defined as the sum of midwives, nurses, physiotherapists, speech therapists, orthoptists, chiropodists, orthosist, occupational therapist, psychomotor therapist, dietician and psychologists in the IRIS. “**Other health services**” was defined as the sum of health care organization for short, medium and long stays, psychiatric hospitals, cancer centres, emergency rooms, maternities, health centres, psychiatric outpatient clinics, centre for preventive medicine, dialysis clinic, hospitalization at home, multidisciplinary health centre, medical analysis laboratories, ambulances, blood transfusion centre and thermal buildings.

Part III. Statistical analysis

Transformation of exposure variables

For each exposure variable X , eight histograms were plotted: X , $\log_{10}(X)$, $\log_2(X)$, $1/X^2$, $1/X$, $1/\sqrt{X}$, $\sqrt{X}$ and X^2 . The best transformation to achieve normality was chosen on the basis of these graphs and, if no transformation worked, the exposure variable was dichotomized (where the cut-off was the median value). Transformation was chosen to obtain reasonably symmetric distributions and, if possible, the same transformation was used for all exposure variables from the same exposure family.

Imputation of missing data

Missing values of exposures and cofounders (median (IQR) missing values = 0 (0-3)) were imputed using the method of chained equations (van Buuren and Groothuis-Oudshoorn, 2011; White et al., 2011) and 5 complete datasets were generated (Stuart et al., 2009). The imputation model included outcomes variables, centre of recruitment and all exposure variables with an absolute correlation greater than 0.2 with the exposure variable to be imputed. Predictive mean matching was used to impute continuous variables and logistic and multinomial regressions were used to impute binary and categorical exposures, respectively. $M=5$ complete datasets were created. Imputed missing values several times allows to account for uncertainty in the final standard errors, confidence intervals and p-value (Stuart et al., 2009).

Standardisation of exposure variables

Each quantitative transformed exposure variable was standardised by the interquartile range so as to allow between-covariate comparison of regression estimates. In that order, each transformed variable was divided by its 75th percentile (3rd quartile) minus 25th percentile (1st quartile).

Exposome description

All exposure variables were described as mean (standard deviation) or n (%) for quantitative and qualitative variables, respectively. A first heat map presented correlations between exposures. They were assessed through Pearson coefficient for two quantitative variables, polychoric coefficient for two qualitative variables and polyserial coefficient for a quantitative and a qualitative variable. A second heat map assessed absolute within (diagonal) and between (off-diagonal) families of exposure median correlations.

Exposome dimension reduction from ExWAS results

This step aimed to restrict the number of exposure variables to be included in the supervised Bayesian Profile Regression (sBPR) method to allow a good model stability. In that order, an Exposome-Wide Association Analysis (ExWAS) was performed to identify exposures individually associated with $FEV_1\%$. This method consists in fitting independent mixed linear regressions models for each exposure variable, with adjustment for the centre of inclusion, gender, age, height and socio-professional category as fixed effects and family as random effect. The Rubin's rule was used to combine the estimates from each of the 5 imputed datasets into one overall estimate together with an associated variance that incorporates both the within

and between imputation variability (Marshall et al., 2009; Rubin, 2004)[25,26]. Factors showing suggestive association with FEV₁% (p-value < 0.2) were selected in the reduced set of exposures. In addition, to address the significance of each factor independently, the number of independently effective tests was calculated with the Li formula (Li et al., 2012) applied to the correlation matrix of exposures variables and p-values of the association tests were compared to the significance threshold of 0.05 divided by this number (i.e. 1.71×10^{-3}) to correct for multiple hypothesis testing.

Identification of profiles of joint exposures associated with FEV₁% using sBPR

The second step consisted in identifying profiles of co-exposures associated with low lung function. In that order, a supervised cluster analysis, *i.e.* simultaneously considering in the model the outcome and exposures, was conducted using the sBPR clustering method (Molitor et al., 2010). This model is fitted using Markov Chain Monte Carlo sampling methods and outputs a different clustering of the data at each iteration of the sampler, thus coherently propagating uncertainty (Gilks et al., 1998). Here, the algorithm was run for 200,000 iterations. The sBPR was applied in the reduced set of exposures identified in step 1 by the ExWAS, FEV₁%, and included variables used as confounders in the ExWAS analysis (age, gender, height, centre of inclusion, number of smoked pack-years and socio-professional category) as covariates. When two exposures variables with a p-value < 0.20 in the ExWAS analysis had an absolute correlation coefficient > 0.70 (Ratner, 2009), only the variable with the lowest p-value in the ExWAS analysis was retained. This step allowed us to obtain well-characterized clusters with specific profiles of exposure. Socio-demographic characteristics and exposure factors were compared between clusters to identify the factors mainly driving the clustering. This analysis was also performed after adjustment for age, gender and centre to identify exposures associated with clusters independently of these 3 variables. Moreover, phenotypic characteristics at EGEA2, but also at EGEA3 about 7 years later, were compared between clusters.

As this method relies on a stochastic process, results may not be stable (final results depend from the initialization step). To test the stability of our final model, we ran it with 10 different seeds (all with 200,000 iterations). All these 10 seeds identified 3 clusters and the median [interquartile range] number of subjects remaining in the same cluster, independently from the seed, was 0.86 [0.73-0.94] (data not shown).

References

- Ainsworth, B.E., Haskell, W.L., Whitt, M.C., Irwin, M.L., Swartz, A.M., Strath, S.J., O'Brien, W.L., Bassett, D.R., Schmitz, K.H., Emplaineourt, P.O., Jacobs, D.R., Leon, A.S., 2000. Compendium of Physical Activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 32, S498–S516. <https://doi.org/10.1097/00005768-200009001-00009>
- Alonso, R., Vivanco, M.G., González-Fernández, I., Bermejo, V., Palomino, I., Garrido, J.L., Elvira, S., Salvador, P., Artíñano, B., 2011. Modelling the influence of peri-urban trees in the air quality of Madrid region (Spain). *Environ Pollut* 159, 2138–2147. <https://doi.org/10.1016/j.envpol.2010.12.005>
- Baecke, J.A., Burema, J., Frijters, J.E., 1982. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr* 36, 936–942. <https://doi.org/10.1093/ajcn/36.5.936>
- Bentayeb, M., Stempfelet, M., Wagner, V., Zins, M., Bonenfant, S., Songeur, C., Sanchez, O., Rosso, A., Brulfert, G., Rios, I., Chaxel, E., Virga, J., Armengaud, A., Rossello, P., Rivière, E., Bernard, M., Vasbien, F., Deprost, R., 2014. Retrospective modeling outdoor air pollution at a fine spatial scale in France, 1989–2008. *Atmos Environ* 92, 267–279. <https://doi.org/10.1016/j.atmosenv.2014.04.019>
- Bonifacj, C., Gerber, M., Scali, J., Daures, J.P., 1997. Comparison of dietary assessment methods in a southern French population: use of weighed records, estimated-diet records and a food-frequency questionnaire. *Eur J Clin Nutr* 51, 217–231.
- Chiuve, S.E., Fung, T.T., Rimm, E.B., Hu, F.B., McCullough, M.L., Wang, M., Stampfer, M.J., Willett, W.C., 2012. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. *J Nutr* 142, 1009–1018. <https://doi.org/10.3945/jn.111.157222>
- Dumas, O., Smit, L.A.M., Pin, I., Kromhout, H., Siroux, V., Nadif, R., Vermeulen, R., Heederik, D., Hery, M., Choudat, D., Kauffmann, F., Le Moual, N., on behalf of the Epidemiological Study on the Genetics and Environment of Asthma (EGEA), 2011. Do young adults with childhood asthma avoid occupational exposures at first hire? *Eur Respir J* 37, 1043–1049. <https://doi.org/10.1183/09031936.00057610>
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1998. Markov chain Monte Carlo in practice, Interdisciplinary statistics. Chapman & Hall, Boca Raton, Fla.
- Hercberg, S., Preziosi, P., Briançon, S., Galan, P., Triol, I., Malvy, D., Roussel, A.M., Favier, A., 1998. A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study--design, methods, and participant characteristics. *Supplementation en Vitamines et Minéraux AntioXydants. Control Clin Trials* 19, 336–351.
- Kauffmann, F., Dizier, M.H., 1995. EGEA (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy)--design issues. EGEA Co-operative Group. *Clin. Exp. Allergy* 25 Suppl 2, 19–22.

- Kauffmann, F., Dizier, M.-H., Pin, I., Paty, E., Gormand, F., Vervloet, D., Bousquet, J., Neukirch, F., Annesi, I., Oryszczyn, M.-P., Lathrop, M., Demenais, F., Lockhart, A., Feingold, J., 1997. Epidemiological Study of the Genetics and Environment of Asthma, Bronchial Hyperresponsiveness, and Atopy: Phenotype Issues. *Am J Respir Crit Care Med* 156, S123–S129. <https://doi.org/10.1164/ajrccm.156.4.12tac9>
- Kennedy, S.M., 2000. Development of an asthma specific job exposure matrix and its application in the epidemiological study of genetics and environment in asthma (EGEA). *Occup Environ Med* 57, 635–641. <https://doi.org/10.1136/oem.57.9.635>
- Kloog, I., Nordio, F., Lepeule, J., Padoan, A., Lee, M., Auffray, A., Schwartz, J., 2017. Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data. *Int J Climatol* 37, 296–304. <https://doi.org/10.1002/joc.4705>
- Le Moual, N., Varraso, R., Siroux, V., Dumas, O., Nadif, R., Pin, I., Zock, J.-P., Kauffmann, F., 2012. Domestic use of cleaning sprays and asthma activity in females. *Eur Respir J* 40, 1381–1389. <https://doi.org/10.1183/09031936.00197611>
- Li, M.-X., Yeung, J.M.Y., Cherny, S.S., Sham, P.C., 2012. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet* 131, 747–756. <https://doi.org/10.1007/s00439-011-1118-2>
- Li, Z., Kesse-Guyot, E., Dumas, O., Garcia-Aymerich, J., Leynaert, B., Pison, C., Le Moual, N., Romieu, I., Siroux, V., Camargo, C.A., Nadif, R., Varraso, R., 2017. Longitudinal study of diet quality and change in asthma symptoms in adults, according to smoking status. *Br J Nutr* 117, 562–571. <https://doi.org/10.1017/S0007114517000368>
- Marshall, A., Altman, D.G., Holder, R.L., Royston, P., 2009. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 9, 57. <https://doi.org/10.1186/1471-2288-9-57>
- Matulonga, B., Rava, M., Siroux, V., Bernard, A., Dumas, O., Pin, I., Zock, J.-P., Nadif, R., Leynaert, B., Le Moual, N., 2016. Women using bleach for home cleaning are at increased risk of non-allergic asthma. *Respir Med* 117, 264–271. <https://doi.org/10.1016/j.rmed.2016.06.019>
- Miller, M.R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Crapo, R., Enright, P., van der Grinten, C.P.M., Gustafsson, P., Jensen, R., Johnson, D.C., MacIntyre, N., McKay, R., Navajas, D., Pedersen, O.F., Pellegrino, R., Viegi, G., Wanger, J., ATS/ERS Task Force, 2005. Standardisation of spirometry. *Eur. Respir. J.* 26, 319–338. <https://doi.org/10.1183/09031936.05.00034805>
- Molitor, J., Papatthomas, 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>
- Paffenbarger, R.S., Blair, S.N., Lee, I.-M., Hyde, R.T., 1993. Measurement of physical activity to assess health effects in free-living populations: *Med Sci Sports Exerc* 25, 60–70. <https://doi.org/10.1249/00005768-199301000-00010>

Quanjer, P.H., Stanojevic, S., Cole, T.J., Baur, X., Hall, G.L., Culver, B.H., Enright, P.L., Hankinson, J.L., Ip, M.S.M., Zheng, J., Stocks, J., the ERS Global Lung Function Initiative, 2012. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 40, 1324–1343. <https://doi.org/10.1183/09031936.00080312>

Ratner, B., 2009. The correlation coefficient: Its values range between +1/–1, or do they? *J Targeting, Meas Anal Mark* 17, 139–142. <https://doi.org/10.1057/jt.2009.5>

Rubin, D.B., 2004. Multiple imputation for nonresponse in surveys, Wiley classics library ed. ed, Wiley classics library. Wiley-Interscience, Hoboken, NJ.

Stuart, E.A., Azur, M., Frangakis, C., Leaf, P., 2009. Multiple Imputation With Large Data Sets: A Case Study of the Children’s Mental Health Initiative. *Am J Epidemiol* 169, 1133–1139. <https://doi.org/10.1093/aje/kwp026>

Temam, S., Varraso, R., Pornet, C., Sanchez, M., Affret, A., Jacquemin, B., Clavel-Chapelon, F., Rey, G., Rican, S., Le Moual, N., 2017. Ability of ecological deprivation indices to measure social inequalities in a French cohort. *BMC Public Health* 17, 956. <https://doi.org/10.1186/s12889-017-4967-3>

Valari, M., Menut, L., Chatignoux, E., 2011. Using a Chemistry Transport Model to Account for the Spatial Variability of Exposure Concentrations in Epidemiologic Air Pollution Studies. *J Air Waste Manag Assoc* 61, 164–179. <https://doi.org/10.3155/1047-3289.61.2.164>

van Buuren, S., Groothuis-Oudshoorn, K., 2011. Mice : Multivariate Imputation by Chained Equations in R. *J Stat Softw* 45. <https://doi.org/10.18637/jss.v045.i03>

Varraso, R., Garcia-Aymerich, J., Monier, F., Le Moual, N., De Batlle, J., Miranda, G., Pison, C., Romieu, I., Kauffmann, F., Maccario, J., 2012. Assessment of dietary patterns in nutritional epidemiology: principal component analysis compared with confirmatory factor analysis. *Am J Clin Nutr* 96, 1079–1092. <https://doi.org/10.3945/ajcn.112.038109>

White, I.R., Royston, P., Wood, A.M., 2011. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 30, 377–399. <https://doi.org/10.1002/sim.4067>

Supplemental Tables

Supplemental Table E1. Characteristics of included and excluded individuals with asthma.

	Included individuals n = 599	Excluded individuals n = 99	p
Age, years	39 ± 16	34 ± 18	0.01
Gender			0.19
Female, n(%)	288 (48)	40 (40)	
Male, n(%)	311 (52)	59 (60)	
Smoking status			0.09
Never smoker, n(%)	301 (50)	52 (54)	
Former smoker, n(%)	151 (25)	15 (15)	
Active smoker, n(%)	147 (25)	30 (31)	
Socio-professional category			0.45
No profession, n(%)	84 (14)	12 (14)	
Executive, n(%)	189 (32)	22 (26)	
Technician, n(%)	254 (42)	35 (42)	
Farmer/Labourer, n(%)	72 (12)	15 (18)	
Asthma attack in the last year, n(%)	227 (38)	25 (30)	0.23
Use of inhaled corticosteroids in the last year, n(%)	239 (40)	35 (40)	1.00

Supplemental Table E2. Comparisons of exposures by inclusion centre

Family of exposures	Exposure	Paris n = 184	Lyon n = 94	Marseille n = 89	Montpellier n = 47	Grenoble n = 185	p
Indoor air	Mould in the bedroom/living-room/bathroom, n(%)	19 (10)	2 (2)	9 (10)	4 (9)	11 (6)	0.12
	Water damage in the last 12 months, n(%)	21 (12)	9 (10)	13 (15)	3 (7)	13 (7)	0.31
	Use of gaz for cooking, n(%)	99 (56)	68 (74)	60 (67)	38 (81)	132 (73)	0.001
Tobacco	Smoking status						0.03
	Never smoker, n(%)	101 (55)	37 (39)	40 (45)	27 (57)	96 (52)	
	Former smoker, n(%)	34 (19)	32 (34)	32 (36)	10 (21)	43 (23)	
	Active smoker, n(%)	49 (26)	25 (27)	17 (19)	10 (21)	46 (25)	
	Number of pack-years	4.2 ± 11.5	7.0 ± 10.8	8.8 ± 15.2	4.8 ± 8.5	3.4 ± 7.0	<0.001
	0, n(%)	101 (55)	37 (39)	40 (45)	27 (57)	96 (52)	
	< 10, n(%)	62 (34)	34 (36)	26 (29)	11 (23)	66 (36)	
	10 – 20, n(%)	11 (6)	13 (14)	7 (8)	6 (13)	16 (9)	
	> 20, n(%)	10 (5)	10 (11)	16 (18)	3 (6)	7 (4)	
	Passive smoking						0.001
	None, n(%)	83 (45)	37 (39)	31 (35)	28 (60)	78 (42)	
	Intermediate, n(%)	29 (16)	33 (35)	32 (36)	11 (23)	57 (31)	
	High, n(%)	72 (39)	24 (26)	25 (28)	8 (17)	50 (27)	
Season	Season of examination						0.13
	Summer, n(%)	49 (27)	19 (20)	19 (21)	9 (19)	58 (31)	
	Autumn, n(%)	40 (22)	21 (22)	30 (34)	10 (21)	37 (20)	
	Winter, n(%)	49 (27)	27 (29)	24 (27)	20 (43)	48 (26)	
	Spring, n(%)	46 (25)	27 (29)	16 (18)	8 (17)	42 (23)	
Pets	Having cat/dog currently, n(%)	65 (36)	43 (46)	41 (46)	24 (51)	71 (39)	0.16
	Having cat/dog in childhood, n(%)	105 (57)	66 (70)	62 (70)	35 (74)	114 (62)	0.05
	Occupational exposure ever						<0.001

Occupational exposures to asthmagens	None, n(%)	94 (57)	34 (36)	37 (44)	19 (41)	87 (53)	
	Low, n(%)	37 (22)	21 (22)	23 (27)	11 (24)	23 (15)	
	High, n(%)	35 (21)	37 (41)	25 (29)	16 (35)	52 (32)	
	Occupational exposure in the last job						<0.001
	None, n(%)	128 (78)	59 (63)	63 (74)	29 (63)	121 (74)	
	Low, n(%)	20 (12)	17 (18)	15 (18)	5 (11)	16 (10)	
	High, n(%)	17 (10)	18 (19)	7 (8)	12 (26)	26 (16)	
Cleaning products	Use of bleach						0.10
	<1 day/week, n(%)	148 (82)	65 (70)	71 (80)	33 (70)	141 (78)	
	1 to 3 days/week, n(%)	21 (12)	21 (23)	10 (11)	10 (21)	34 (19)	
	4 to 7 days/week, n(%)	12 (7)	7 (7)	8 (9)	4 (9)	5 (3)	
	Use of sprays						0.69
	None, n(%)	85 (54)	48 (63)	43 (62)	20 (56)	92 (64)	
	1 spray/week, n(%)	23 (17)	11 (14)	15 (22)	9 (25)	33 (23)	
Rurality	≥2 sprays/week, n(%)	28 (21)	17 (22)	11 (16)	7 (19)	18 (13)	
	Born in a rural municipality, n(%)	37 (21)	23 (25)	11 (13)	12 (26)	49 (26)	0.11
	Ever-living in a rural municipality, n(%)	79 (44)	62 (67)	35 (41)	31 (67)	120 (65)	<0.001
Social	Social Deprivation Index						0.09
	1st quintile, n(%)	92 (50)	52 (57)	47 (53)	11 (24)	95 (52)	
	2nd quintile, n(%)	33 (18)	16 (17)	20 (23)	12 (26)	41 (22)	
	3rd quintile, n(%)	29 (16)	14 (15)	11 (13)	9 (20)	15 (8)	
	4th quintile, n(%)	20 (11)	6 (7)	6 (7)	9 (20)	23 (13)	
	5th quintile, n(%)	10 (5)	4 (4)	4 (5)	5 (11)	9 (5)	
Physical activity and diet	METs	119 ± 39	132 ± 43	131 ± 37	148 ± 36	131 ± 41	<0.001
	Diet (AHEI-2010 score)	48 ± 10	51 ± 8	52 ± 9	53 ± 11	51 ± 10	0.001
Altitude	Altitude, m	85 ± 75	264 ± 147	99 ± 96	83 ± 118	261 ± 182	<0.001
Building and population densities	Building density in a 100m buffer	0.38 ± 0.23	0.24 ± 0.19	0.26 ± 0.17	0.30 ± 0.22	0.24 ± 0.17	<0.001
	Building density in a 300m buffer	0.27 ± 0.17	0.18 ± 0.14	0.21 ± 0.13	0.21 ± 0.15	0.18 ± 0.13	<0.001
	Population density (in a 200x200m area)	648 ± 670	284 ± 300	284 ± 343	233 ± 227	249 ± 257	<0.001

Pollution*	NO ₂ , µg/m ³	32.8 ± 9.7	29.6 ± 11.1	26.8 ± 8.5	22.3 ± 8.0	27.8 ± 8.4	<0.001
	PM _{2.5} , µg/m ³	19.2 ± 4.0	17.4 ± 3.8	18.6 ± 5.6	15.4 ± 5.0	17.9 ± 4.2	<0.001
	PM ₁₀ , µg/m ³	27.3 ± 4.4	26.2 ± 4.8	31.8 ± 7.5	25.7 ± 6.3	27.5 ± 5.0	<0.001
	O ₃ , µg/m ³	91.3 ± 8.7	105.3 ± 10.1	112.2 ± 9.7	110.5 ± 7.3	104.0 ± 8.3	<0.001
Natural spaces	NDVI in a 100m buffer	0.35 ± 0.15	0.38 ± 0.12	0.32 ± 0.09	0.32 ± 0.10	0.42 ± 0.12	<0.001
	NDVI in a 300m buffer	0.37 ± 0.15	0.40 ± 0.12	0.33 ± 0.09	0.34 ± 0.10	0.45 ± 0.12	<0.001
	NDVI in a 500m buffer	0.38 ± 0.15	0.41 ± 0.13	0.33 ± 0.09	0.35 ± 0.10	0.46 ± 0.12	<0.001
	Number of green spaces in a 300m buffer	2.5 ± 5.4	6.9 ± 9.5	4.1 ± 6.6	3.5 ± 7.6	5.1 ± 7.1	<0.001
	Number of blue spaces in a 300m buffer	0.4 ± 1.4	0.4 ± 1.7	0.0 ± 0.1	0.2 ± 1.0	0.1 ± 1.2	0.10
	Distance to nearest green space, m	1457 ± 15247	789 ± 966	676 ± 823	707 ± 728	481 ± 483	<0.001
	Distance to nearest blue space, m	3365 ± 2739	3508 ± 4240	9614 ± 7153	8359 ± 6248	2595 ± 3224	<0.001
	Shannon Evenness Index [†] in a 300m buffer	0.08 ± 0.09	0.14 ± 0.09	0.09 ± 0.10	0.08 ± 0.08	0.13 ± 0.09	<0.001
Weather*	Wind, m/s	3.0 ± 0.4	2.6 ± 0.5	3.4 ± 0.7	3.2 ± 0.6	1.5 ± 0.7	<0.001
	Relative humidity, %	76 ± 4	73 ± 3	67 ± 3	68 ± 2	75 ± 3	<0.001
	Absolute humidity, g/m ³	8.4 ± 0.3	8.1 ± 0.3	8.6 ± 0.5	8.8 ± 0.5	7.8 ± 0.8	<0.001
	Minimal temperature, °C	8.2 ± 1.3	7.9 ± 1.5	10.5 ± 1.7	9.8 ± 1.2	7.9 ± 1.5	<0.001
	Mean temperature, °C	11.6 ± 1.0	12.0 ± 1.6	13.8 ± 1.3	13.5 ± 1.2	12.0 ± 1.5	<0.001
	Maximal temperature, °C	16.8 ± 1.1	17.9 ± 1.2	20.4 ± 0.9	20.3 ± 0.9	17.5 ± 1.4	<0.001
Roads	Distance to nearest major road, m	1855 ± 1962	3670 ± 7078	3211 ± 6734	3973 ± 6353	1926 ± 4075	0.002
	Number of major roads in a 100m buffer	0.2 ± 0.8	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.5	0.08
	Number of cycle paths in a 100m buffer	0.1 ± 0.7	0.3 ± 1.0	0.2 ± 0.8	0.1 ± 0.3	0.3 ± 0.9	0.26
UV*	Erythematous UV dose	2.2 ± 0.2	2.2 ± 0.1	2.3 ± 0.1	2.4 ± 0.0	2.2 ± 0.2	<0.001
	Vitamin-D UV dose	3.7 ± 0.4	3.8 ± 0.1	4.0 ± 0.1	4.1 ± 0.1	3.8 ± 0.4	<0.001
	DNA-damage UV dose	0.9 ± 0.1	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	0.9 ± 0.1	<0.001
Health services[‡]	Number of general practitioners	1.3 ± 2.2	3.2 ± 4.5	4.0 ± 8.0	4.4 ± 6.3	1.9 ± 2.3	<0.001
	Number of pulmonologists	0.1 ± 0.5	0.2 ± 0.7	0.2 ± 0.5	0.2 ± 0.5	0.1 ± 0.5	0.41
	Number of other physicians	1.7 ± 3.5	5.1 ± 9.8	5.3 ± 10.7	8.5 ± 14.5	4.4 ± 6.6	<0.001
	Number of paramedics	6.0 ± 9.1	14.3 ± 15.2	13.5 ± 15.7	15.8 ± 11.9	8.9 ± 9.5	<0.001

	Number of pharmacies	0.6 ± 1.1	1.1 ± 1.3	1.0 ± 1.2	0.8 ± 1.0	1.0 ± 1.1	0.003
	Number of other health services	0.5 ± 0.9	0.9 ± 1.4	1.2 ± 2.4	0.7 ± 0.8	0.8 ± 1.4	0.002

AHEI: Alternative Healthy Eating Index; METs: Metabolic Equivalent of Tasks ; NO₂: nitrogen dioxide ; PM_{2.5}: particles matters with aerodynamical diameter less than 2.5 µm ; PM₁₀: particles matters with aerodynamical diameter less than 10 µm ; O₃: ozone; NDVI: normalized difference vegetation index; UV: ultraviolet

*: Daily values averaged on the 365 days before before performing the spirometry; †: index of area composition and richness ; ‡: number of health services in the IRIS (“Ilots Regroupés pour l’Information Statistique”) where an IRIS represents an area covering between 1,800 and 5,000 inhabitants.

Supplemental Table E3. Results of the ExWAS analyses assessing associations between each exposure variable and FEV₁(%predicted)

Family of exposures	Exposure	β	sd	p
Indoor air	Mould in the bedroom/living-room/bathroom	1.89	2.73	0.49
	Water damage in the last 12 months	1.95	2.31	0.40
	Use of gaz for cooking	-0.39	1.53	0.80
Tobacco	Smoking status			
	Never smoker	Ref		
	Former smoker	0.58	1.76	0.74
	Active smoker	-0.41	1.72	0.81
	Number of pack-years			
	0	Ref		
	< 10	0.34	1.54	0.83
	10 - 20	0.49	2.57	0.85
	> 20	-1.86	2.83	0.51
	Passive smoking			
Season	None	Ref		
	Intermediate	-0.05	1.70	0.98
	High	-2.26	1.72	0.19
	Season of examination			
	Summer	Ref		
Pets	Autumn	-3.98	2.03	0.05
	Winter	-3.04	1.95	0.12
	Spring	-2.77	2.03	0.17
	Having cat/dog currently	-0.06	1.48	0.97
Occupational exposures to asthmagens	Having cat/dog in childhood	-1.00	1.52	0.51
	Occupational exposure ever			
	None	Ref		
	Low	-0.22	1.89	0.91
	High	-0.64	1.76	0.72
	Occupational exposure in the last job			
	None	Ref		
	Low	-1.02	2.31	0.66
	High	-0.25	2.24	0.91
Cleaning products	Use of bleach			
	<1 day/week	Ref		
	1 to 3 days/week	-1.64	2.01	0.41
	4 to 7 days/week	-0.92	3.00	0.76
	Use of sprays			
	None	Ref		
	1 spray/week	0.28	1.79	0.88
	≥ 2 sprays/week	-1.52	2.15	0.48

Rurality	Born in a rural municipality	-3.01	1.73	0.08
	Ever-living in a rural municipality	-0.23	1.50	0.88
Social	Social Deprivation Index			
	1st quintile	Ref		
	2nd quintile	2.44	1.80	0.09
	3rd quintile	-2.28	2.14	0.29
	4th quintile	-0.37	2.35	0.88
	5th quintile	3.30	3.20	0.30
Physical activity and diet	METs ≥ 127	-0.53	1.64	0.75
	Diet (AHEI-2010 score)	0.13	0.09	0.13
Altitude	Altitude, m	0.16	0.61	0.79
Building and population densities	Building density in a 100m buffer	8.81	4.02	0.03
	Building density in a 300m buffer	8.42	4.71	0.07
	Population density (in a 200x200m area)	0.62	0.39	0.12
Pollution*	NO ₂ , $\mu\text{g}/\text{m}^3$	0.05	0.08	0.53
	PM _{2.5} , $\mu\text{g}/\text{m}^3$	0.01	0.17	0.95
	PM ₁₀ , $\mu\text{g}/\text{m}^3$	-0.05	0.14	0.72
	O ₃ , $\mu\text{g}/\text{m}^3$	-0.02	0.08	0.79
Natural spaces	NDVI in a 100m buffer	-5.69	5.94	0.34
	NDVI in a 300m buffer	-5.46	5.78	0.35
	NDVI in a 500m buffer	-6.33	5.82	0.28
	≥ 1 green spaces in a 300m buffer	-2.05	1.51	0.18
	≥ 1 blue spaces in a 300m buffer	5.08	3.77	0.18
	Distance to nearest green space, m	1.01	0.39	0.009
	Distance to nearest blue space, m	-0.77	0.49	0.11
	Shannon Evenness Index [†] in a 300m buffer	-0.18	7.89	0.98
Weather*	Wind ≥ 2 m/s	-4.08	2.61	0.12
	Relative humidity, %	0.37	0.22	0.09
	Absolute humidity, g/m^3	-0.60	1.29	0.65
	Minimal temperature, °C	-0.37	0.50	0.46
	Mean temperature, °C	-0.81	0.53	0.13
	Maximal temperature, °C	-1.09	0.62	0.08
Roads	Distance to nearest major road, m	0.74	0.47	0.12
	≥ 1 major roads in a 100m buffer	-17.18	6.34	0.007
	≥ 1 cycle paths in a 100m buffer	-3.84	2.60	0.14
UV*	Erythemat UV dose	4.15	5.37	0.44
	Vitamin-D UV dose	2.03	2.80	0.47
	DNA-damage UV dose	6.70	9.74	0.49
Health services³	≥ 2 general practitioners	1.68	1.48	0.26
	≥ 1 pulmonologists	-1.31	2.43	0.59
	≥ 2 other physicians	1.28	1.47	0.38
	≥ 7 paramedics	2.45	1.48	0.10
	≥ 1 pharmacies	2.00	1.44	0.16
	≥ 1 other health services	1.73	1.47	0.24

All linear regressions were adjusted for age, gender, height, socioprofessional category, inclusion centre as fixed effects and family as random effect.

ExWAS: Exposome-Wide Association Study ; FEV₁: forced expiratory volume in 1s ; sd: standard deviation ; AHEI: Alternative Healthy Eating Index; METs: Metabolic Equivalent of Tasks ; NO₂: nitrogen dioxide ; PM_{2.5}: particles matters with aerodynamical diameter less than 2.5 µm ; PM₁₀: particles matters with aerodynamical diameter less than 10 µm; O₃: ozone; NDVI: normalized difference vegetation index; UV: ultraviolet

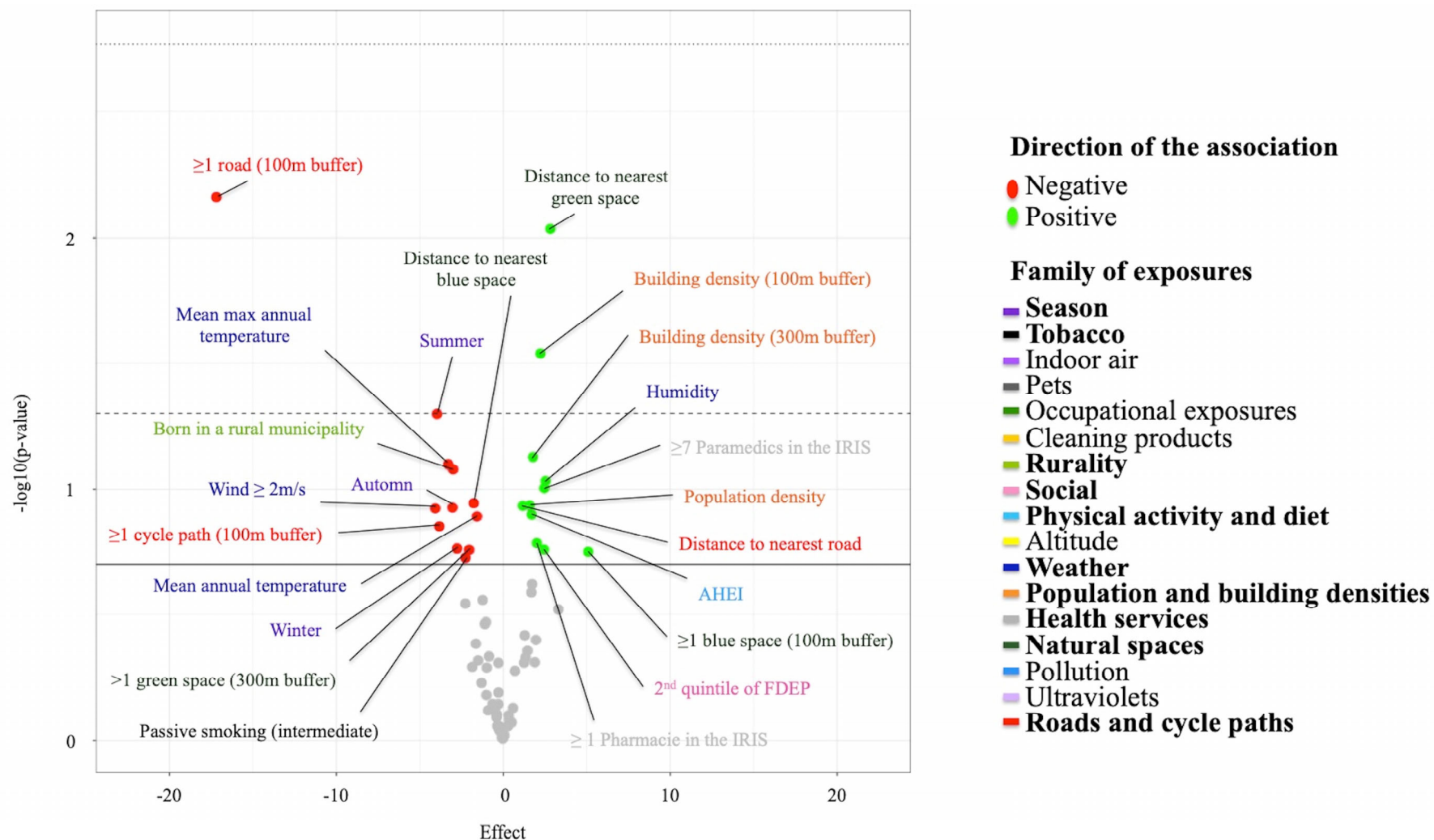
*: Daily values averaged on the 365 days before performing the spirometry; †: index of area composition and richness ; ‡: number of health services in the IRIS (“Ilots Regroupés pour l’Information Statistique”) where an IRIS represents an area covering between 1,800 and 5,000 inhabitants

Supplemental Table E4. Characteristics of the studied population according to their participation to the EGEA3 follow-up about 7 years later

	Individuals followed at EGEA3 study n = 502	Individuals lost to follow-up at EGEA3 study n = 97	p
Age, years	39.1 ± 16.3	38.6 ± 16.9	0.77
Gender			0.11
Female, n(%)	243 (50)	39 (40)	
Male, n(%)	253 (50)	58 (60)	
Smoking status			0.02
Never smoker, n(%)	265 (53)	36 (37)	
Former smoker, n(%)	121 (24)	30 (31)	
Active smoker, n(%)	116 (23)	31 (32)	
Socio-professional category			0.97
No profession, n(%)	70 (14)	14 (14)	
Executive, n(%)	160 (32)	29 (30)	
Technician, n(%)	211 (42)	43 (44)	
Farmer/Labourer, n(%)	61 (12)	11 (11)	
Asthma exacerbation* in the last year, n(%)	65 (14)	12 (14)	1.00
Use of corticosteroids in the last year, n(%)	201 (40)	38 (40)	0.98
FEV ₁ , %	92 ± 17	90 ± 18	0.21

*: exacerbation was defined by use of oral steroids or hospitalization for asthma or emergency visit for asthma in the last 12 months; FEV₁: forced expiratory volume in 1 second

Online Figures



Supplemental Figure E1. Volcano plot of the coefficient estimates for exposure variables versus p value in the ExWAS analysis of the exposures-FEV₁ (%predicted) association.

The solid line corresponds to a p-value of 0.20, the dashed line corresponds to a p-value of 0.05 (alpha) and the dotted line corresponds to a p-value of 1.71×10^{-3} (alpha corrected for multiple comparisons)