

Methodological advances in designing a longitudinal urban exposome study: the UrbanX exposome study for the city of Paris within the CONSTANCES cohort

Nadine Haddad, MS¹, Emeline Lequy, PhD^{2,3}, Marie Zins, PhD^{4,5} , Marcel Goldberg, PhD² , Konstantinos C. Makris, PhD^{*1,6} 

¹Precision Exposomics and Health Laboratory, Cyprus International Institute for Environmental and Public Health, School of Health Sciences, Cyprus University of Technology, Limassol, Cyprus

²Unité "Cohortes en Population" UMS 011 Inserm/Université Paris-Cité/Université Paris Saclay/UVSQ, Villejuif, France

³Paris-Saclay University, UVSQ, Gustave Roussy, Inserm, U1018 CESP, Team "Exposome and Heredity", Villejuif, France

⁴Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Centre for Research in Epidemiology and Statistics (CRESS), Paris, France

⁵Centre d'Epidémiologie Clinique, Hôpital Hôtel Dieu, AP-HP, Centre d'Epidémiologie Clinique, Hôpital Hôtel Dieu, Paris, France

⁶Department of Rehabilitation Sciences, School of Health Sciences, Cyprus University of Technology, Limassol, Cyprus

*Corresponding author: Konstantinos C. Makris, PhD, Professor of Environmental Health, Precision Exposomics and Health Laboratory, Cyprus International Institute for Environmental and Public Health, School of Health Sciences, Cyprus University of Technology, Limassol, Cyprus (konstantinos.makris@cut.ac.cy)

Abstract

Background: The complexity of urban ecosystems and populations presents major challenges in studying how urban settings shape populations' wellbeing and health over space and time. The urban exposome methodological framework has been earlier defined as the continuous spatiotemporal monitoring of population indicators associated with urban external and internal exposome domains, using small area-level analysis. We present the design of a novel urban exposome methodological framework and outline the application of its exposomics tools as nested within a specific longitudinal setting, that is that of the CONSTANCES cohort in France.

Methods: CONSTANCES is a population-based cohort including >200 000 volunteers recruited between 2012 and 2020 and followed-up to date annually with a medical examination every 4 years. The urban Paris population and its setting is considered here, using IRIS small area level (IRIS division of the territory typically involves a population of 1800 to 5000 inhabitants). Using the human exposome domains' classification, we describe a novel methodological approach for tackling the spatiotemporal profile dynamics of the urban Paris exposome and investigate its prospective association with a health outcome, that is body mass index (BMI). Participants from the inclusion dataset are clustered into exposome-based spatial networks using hierarchical clustering on principal components, applying spatial constraints. Mixture effects of co-occurring and space- and time-varying exposome variables in their prospective association with BMI may be estimated using Bayesian kernel machine regression—lagged distribution models, random forest and/or penalized regression models.

Discussion: The application of an urban exposome methodological framework nested within a longitudinal cohort population presents with opportunities to advance our understanding of chronic disease process in complex urban settings. Advanced biostatistical algorithms that account for repeated measures of exposures and outcomes are warranted. This approach would be also applicable in multiple urban population studies nested within different cohorts.

Key words: exposome; urban exposome; cohort; longitudinal studies; BMI; obesity; prevention; repeated measures; exposomics; omics.

Introduction

In a rapidly urbanized world, more than half of the global population currently lives in urban settings. The Eleventh Sustainable Development Goal (SDG-11) places the cities, and their inclusive characteristics in the heart of sustainable development. Specific milestones set for 2030 aim to ensure the provision for all of safe, accessible and affordable basic health and education services, as well as improving the well-being of urban communities.¹ Increased attention is placed on urban populations and accompanying urban designs, types, uses, and forms that shape up relationships and complex mechanisms by which urban ecosystems may impact population health.^{2,3} Urban health represents

a growing research priority aiming to explore the relationship between population health and features of the urban ecosystem towards developing health, environment and climate policies that practically improve quality of life in urban settings.²⁻⁴

The impact of a city on its population's health and wellbeing has always been debated. Several challenges arise with the various approaches quantifying the relationship between urban features and the health status of a specific urban population. First, the "urban environment" concept often lacks a view wider than the urban typology,³ thus, putting forward a reductionist approach to a single physical characteristic(s) or the urban setting, while overlooking the potential synergies among demographic, social,

Received: July 30, 2025; Revised: December 8, 2025; Accepted: December 10, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

behavioral and other characteristics of the urban population.² Additionally, as urban features are not static, but rather dynamic over space and time, specific attention is warranted to explore the interactions and mechanisms of how the network of urban characteristics shapes the onset and progression of urban population health outcomes. As much of the published literature relies on a cross-sectional design, the spatio-temporal patterns between urban exposures and health outcomes are not well captured.⁵

The theoretical framework of the urban exposome

One of the novel concepts to address environment and health concerns is the human exposome concept originally coined by Dr Chris Wild in 2005.⁶ The human exposome concept encompasses the totality of environmental exposures and their endogenous response across the lifespan, and how they shape disease risk and disease development.⁶ However, when people live in urban settings with variable population densities, individuals often share their personal exposomes due to common characteristics, which in turn may be influenced by socioeconomic (such as ethnicity, social class) or physical and infrastructure parameters (such as built environment and neighborhood status).⁷ The urban exposome appears as the natural continuum of the human exposome towards the characterization of small(er) urban areas in the city and their interactions

with residents, city infrastructure, built environment and policies. The urban exposome was earlier defined as “the continuous temporal, and spatial surveillance/monitoring of quantitative and qualitative indicators associated with the urban external and internal parameters (belonging to the domains of the urban exposome) that would ultimately shape up the quality of life and the health of the urban population, using small areas of the city, such as neighborhoods, quarters, smaller administrative districts, as the point of reference.”⁷ This definition outlines the importance of the time dimension on one hand, underlining the need for continuous monitoring of the temporal variation in the urban exposome components, and the space dimension on the other, allowing for the geospatial detection and monitoring of exposomic perturbations and health trajectories within the urban setting. Within the framework of the urban exposome, individual-level measurements aggregated within small(er)-areas of the urban setting become the key level of analysis, allowing for an improved understanding of the variability in exposures and urban health disparities. The general external urban exposome domain includes global policies, decisions and factors relating to the urban setting; the specific external urban exposome domain includes climate, migration and demographic changes, etc, while the internal urban exposome includes parameters integral to the urban setting, and extending into the individual-level human exposome domains.⁷ The

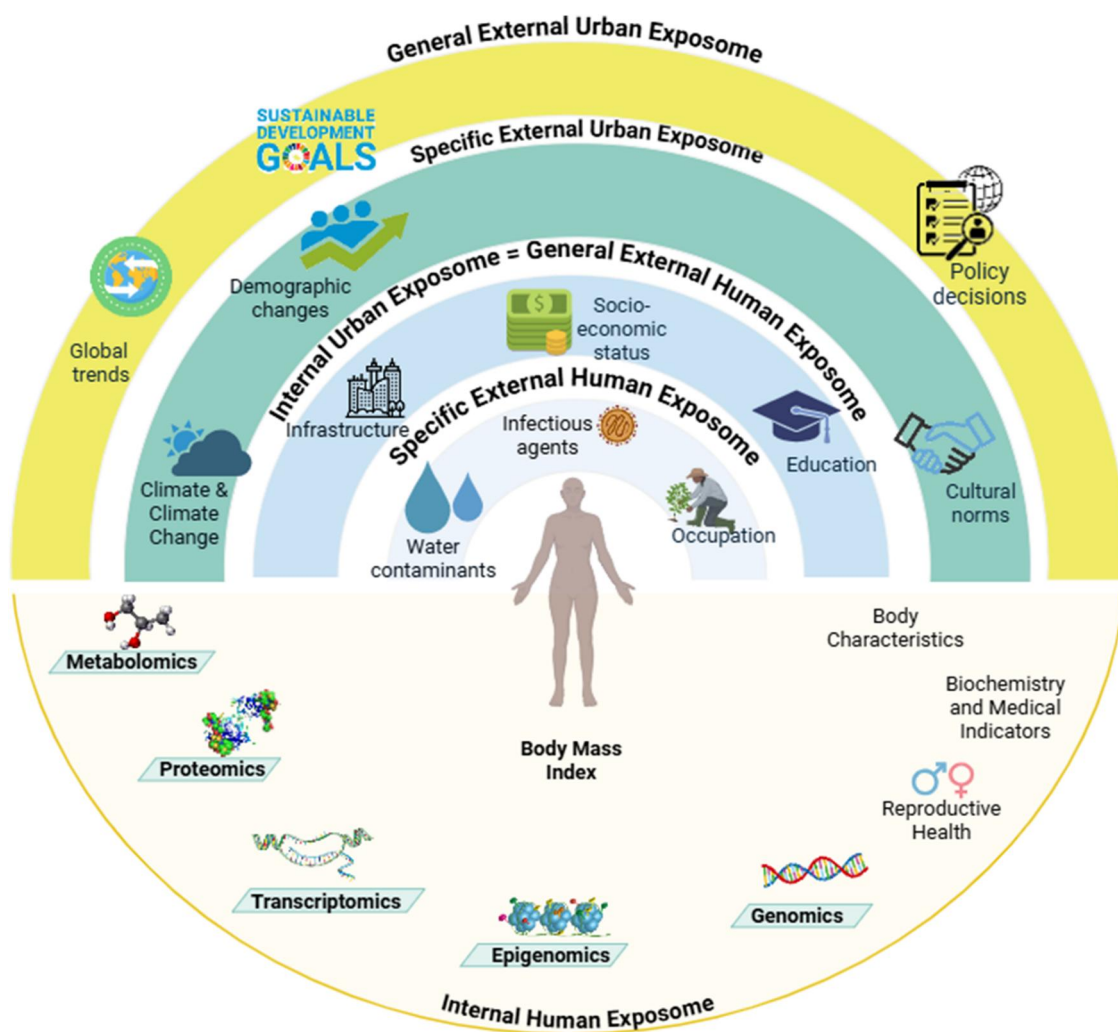


Figure 1. The natural continuum between the urban exposome and human exposome domains as influenced by the global dimensions of space and time.

integration of longitudinal multi-omics platforms presents with opportunities to better understand the complexity of urban exposures as depicted on population health indicator dynamics and associated health outcomes (Figure 1).^{8,9}

Urban exposome studies require the critical infrastructure and resources of pregnancy birth/population cohorts, registries/survey datasets, as well as the availability of routinely collected environmental (non-genetic) variables. One of the very first urban exposome studies within established cohorts, the Human Early Life Exposome (HELIX) project used exposomic data from six urban EU population-based birth cohorts linking them with -omics signatures and child health outcomes.¹⁰ The EXPANSE study studied the impact of the urban exposome on cardiometabolic and pulmonary diseases over the life course of 55 million European inhabitants across 12 countries.¹¹ The Child Cohort Network brought together several EU child cohorts into one data sharing platform, hence investigating the association of repeated measurements of the pregnancy and early life exposome with multi-omics and health outcomes followed up to 18 years of age.¹²

This methodological manuscript describes the incorporation of the urban exposome methodological framework taking the form of a study protocol describing the main elements (at a minimum) that such exposomic studies nested within existing prospective cohorts would possess, including their exposomics tools. The scientific approach presented hereafter aims to showcase a novel framework to articulate how longitudinal data from ongoing cohorts can be integrated into nested urban exposome study designs to capture and characterize the spatiotemporal dynamics of exposome profiles across urban populations—an area that has remained largely underexplored. The description of this methodological manuscript may operationalize either within an existing longitudinal study or applied to multiple cohort-nested exposome studies. By leveraging the existing CONSTANCES (“Cohorte des consultants des Centres d’examens de santé”) cohort data infrastructure and focusing on the urban exposome of the population of Paris (Paris UrbanX), this approach breaks new ground by laying out how the associations between geospatially dynamic, temporally resolved assessments of environmental exposures and their health impacts can be studied within a cohort infrastructure. As a theoretical illustration, we describe exposomic methods and tools for investigating the association between the urban exposome and body mass index (BMI), using small-area level units of measurement and analysis. This paper follows the “Strengthening The Reporting of Observational Studies in Epidemiology” (STROBE) guidelines (Table S1).¹³

Methods

General overview of the CONSTANCES cohort and the city of Paris

The French general population-based cohort CONSTANCES was designed according to a random sampling scheme stratified on age (aged 18–69 years at inception), sex, socioeconomic status (SES) from 21 selected health screening centers (HSCs) located in 20 “départements” of different regions of France, affiliated to the National Health Insurance Fund (Cnam: Caisse nationale d’assurance maladie) that covers salaried workers, professionally active or retired and their dependents (more than 85% of the French population), thus, excluding agricultural and self-employed workers which are affiliated to other health insurance funds.¹⁴ Those who agreed to participate to CONSTANCES received questionnaires to complete and attended one of the HSCs for a medical examination and completing specific questionnaires. After enrollment,

participants have been followed-up by annual self-administered paper-based or web-based questionnaires collecting various information, and they have been invited for a new medical examination every 4 years.¹⁴ The recruitment was initiated in late 2012 and included >200 000 volunteers recruited up to 2020 and followed-up to date.¹⁵ The CONSTANCES design and methodological framework for data and biospecimen collection have been extensively described elsewhere.^{14,16,17} CONSTANCES Cohort project has obtained the authorization related to confidentiality, safety, and security procedures from the relevant French legal authorities. Written informed consent was provided by participants to participate in this study.^{14,16} De-identified data are used throughout the process of data transfer, management, and analysis. A summary of the CONSTANCES data collection tools is provided in Table S2.

There are 20 municipal arrondissements in the city of Paris, each divided into several IRIS (Îlots Regroupés pour l’Information Statistique), the fundamental intra-municipal unit.^{18,19} As per the French National Institute of Statistics and Economic Studies (Institut National de la Statistique et des Etudes Economiques, INSEE), a residential IRIS is a geographical unit area with 1800–5000 inhabitants; there are 992 IRIS in Paris city¹⁸ during the study period (2012–2020).²⁰ The degree of urbanization of the provided Paris IRIS was verified using the DEGURBA database of France and cross-validated with the INSEE database of urban units.^{21,22} In the CONSTANCES cohort, each participant’s home address was recorded at inclusion and updated throughout the follow-up period and geocoded with a precision indicator ranging from the exact address to the centroid of the zip code, then linked to residential IRIS version of 2016, which corresponds to the middle of the inclusion period.

Study population within the CONSTANCES cohort

To allow temporal analysis of exposome variables, we suggest limiting the sample to CONSTANCES volunteers enrolled during 2012–2020 for whom follow-up information between 2013 and 2020 is available. To limit classification bias, we propose limiting the sample to volunteers residing in geocoded IRIS classified as “residential,” and those providing only one residential address throughout their study period. For the sake of this example, we also propose limiting the sample to those being alive during the complete study period of interest.

Urban exposome datasets and classification into human exposome domains

Variables from self-administered questionnaires and health screening examinations at inclusion and during follow-up²³ of the CONSTANCES population can be classified into the urban and human exposome domains, their components, and their sub-components (groups of variables)^{7,24} (Table S3). The temporal analysis of the urban exposome would require the inclusion of exposomic variables assessed at three or more time points of follow-ups (either at inclusion and twice or more during follow-up years; or collected at three or more follow-up years during the study period). Nevertheless, other variables (collecting socio-demographic information, personal medical history, physical limitations, family medical history) need to be selected even if available once at baseline (Table S4).

Urban exposome variables from auxiliary data sources

The availability of the participants’ geocoded residential (eg, IRIS) data in a cohort infrastructure enables their linkage to

contextual indicators (ie, deprivation index, localized accessibility indicators, etc.) and environmental databases,²⁵ such as annual mean concentrations of fine particulate matter (PM_{2.5}, with a diameter <2.5 µm), black carbon, and nitrogen dioxide (NO₂) at the residential addresses which can be estimated with (100 × 100 m) land-use regression models incorporating ground-based measurements, satellite-derived and chemical transport-modeled estimates, road density, land-use variables, and altitude.^{26,27} Satellite-based normalized difference vegetation index (NDVI) can be used to estimate residential greenness within 300 m of the participants' residential addresses.²⁸

As additional auxiliary data sources, further enrichment of CONSTANCES database with extra urban exposome variables can be achieved by screening national French governmental websites (e.g., www.info.gouv.fr, www.paris.fr, www.insee.fr) for articles, datasets and maps on urban exposome components covering any administrative level of Paris during the study period of interest (2012–2020).

Data management

All information obtained from the various auxiliary sources require harmonization and linkage to the CONSTANCES dataset as per the smallest geographic level for which each variable is available. To this, we propose statistical methods that can be conducted in separate packages using R version 4.3.2 and R studio 2024.04.1 and beyond.^{29,30} These statistical approaches can be further implemented using the hands-on ATHLETE tutorial, which describes their application across key exposome analysis steps: descriptive analysis, visualization, and association analysis.³¹ “ExposomeShiny” is another R-shiny web-based analytical tool that provides an interactive interface for exploring, visualizing and analyzing exposome datasets.³²

Table 1. Demographic characteristics of the urban Paris population at inclusion (N = 21805).

Variable	Mean (standard deviation)	N (%)
Age at inclusion	46.2 (13.2)	
Age categories at inclusion, years		
18–29		2382 (12)
30–39		4321 (22)
40–49		4823 (24)
50–59		4102 (21)
60–69		4283 (22)
Sex		
Women		11 264 (54)
Men		9625 (46)
Professional status		
I have a job		13 850 (70)
Unemployed or job seeker		1700 (8.6)
Retired or no longer in business		2595 (13)
In training (pupil, student, etc.)		592 (3.0)
Does not work for health reasons		130 (0.7)
No professional activity		386 (2.0)
Other		404 (2.1)
Recruitment center		
PARIS-CPAM		11 643 (56)
PARIS-IPC		9246 (44)
Inclusion year		
2012		675 (3.2)
2013		1795 (8.6)
2014		2702 (13)
2015		3319 (16)
2016		3594 (17)
2017		3654 (17)
2018		3119 (15)
2019		2031 (9.7)

Characterizing the exposomic profile of urban Paris population: data processing

A total of 214 exposome variables were selected from CONSTANCES at inclusion and classified into (n, %): the general external exposome domain (30, 14%), the specific external exposome domain (56, 26%) and the internal exposome domain (128, 60%). The UrbanX Paris population matching our selection criteria comprised 20 889 participants; most were women (54%) with a mean age of 46.6 years (Table 1). Descriptive statistics were applied to describe the human exposome profile by domain at inclusion. The distribution of selected variables by human exposome domain and component are provided in Tables S5–S8.

Participant residential locations were distributed over 849 (out of the 992) IRIS units of Paris (Figure 2). Around 1% of the IRIS (n = 10) included ≤5 participants, and 7% (n = 56) included <10 participants. Depending on the research question, it might be necessary to group participants into higher administrative units, such as per quarter or arrondissement (French administrative unit). The various selected exposome components and subcomponents of each human exposome domain at both inclusion and follow-ups of the CONSTANCES population can be depicted in Table 2.

Data pre-processing

Exposomic variables that are homogeneous (ie, those variables with >90% of observations with same values) and those with data missing for >70% are first excluded.^{33,34} For correlation and dimension reduction techniques, only the main exposures (eg, not including questions starting with “If yes, then”) are retained for better interpretation of the findings.³⁵ Also, only one exposure is retained from pairs with absolute Pairwise Spearman or Pearson correlation coefficients $r > 0.9$.^{34–36} The decision on which exposure to retain from the highly correlated pair is usually made using expert judgment and published studies. Circo plots and heat maps are produced for a better visualization of the correlations. Skewed exposures variables are transformed to approach normality using either Box-Cox transformation, optimal transformation, categorization, logarithmic, root square, inverse function or spline function as proposed in the exposome literature.^{33,36,37} To remove between-area variation, variables are centered by mean and scaled by unit variance.³⁴ Last, the imputation of missing data can be done using the method of multivariate imputation by chained equations using predictive mean matching in *mice* R package widely used in exposome studies.^{34,36–38} The *MissForest* R package is also gaining popularity in exposome studies as it can handle both categorical and continuous variables while capturing non-linear relationships, using a single iterative imputation algorithm.^{39–41} These pre-processing steps are applicable separately to the database at inclusion, and the annual follow-up database subsequently.

Dimension reduction

Dimension reduction techniques are useful for describing patterns of exposures within the exposome by concentrating on the variance of these exposures in a smaller set of factors or components, while minimizing loss of information. Depending on the urban exposome research question, the dimension reduction techniques can be either unsupervised (applied irrespectively of the outcome of interest) or supervised (considering the outcome of interest during feature extraction).^{35,42–44} Specialized R packages facilitate exposomics data processing, such as the *exposome* R package.⁴⁵ Unsupervised Principal Component Analysis (PCA) are usually proposed first to reduce data dimensionality from the

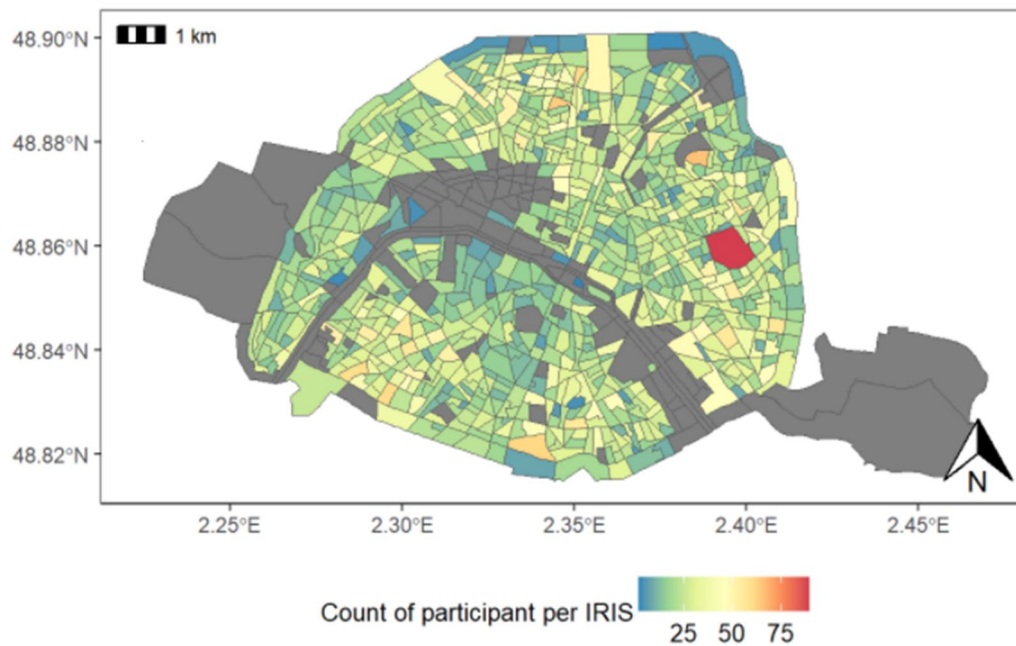


Figure 2. Geographic IRIS distribution of self-reported residential addresses in urban Paris, France.

Table 2. Availability of the human exposome variables selected from CONSTANCES dataset per exposome domain and component at inclusion and follow-up years.

Domain	Exposome component and Sub-components	Inclusion 2012– 2020 ^a	Follow-up ^a										
			2013	2014	2015	2016	2017	2018	2019	2020			
General External	Demographic Characteristics	■											
	Geographic origins	■											
	Socio-Economic status	■											
	Volunteer socio-professional category; Professional status; Education level	■											
	Volunteer Employment Situation	■	■	■	■	■	■	■	■	■	■	■	■
	Spouse Employment Situation; socio professional category	■	■	■	■	■	■	■	■	■	■	■	■
	Social Deprivation Score (EPICES)	■											
	Marital Status; Number of Children; Household composition	■								■			
	Living as a couple	■		■	■	■	■	■	■	■	■	■	■
	Income	■											
	Financial Difficulties/Foregone Medical Care	■	■	■	■	■							■
	Mental Health Center for Epidemiologic Studies Depression (CES-D) Scale CES-D Scale	■			■					■			
	Life events since the last 12 months Marriage; Arrival of child(ren); Death of a relative, etc.	■	■	■	■	■	■	■	■	■	■	■	■
	Specific External	Lifestyle	■										
Alcohol		■	■	■	■	■	■	■	■	■	■	■	■
Cannabis; Tobacco		■	■	■	■	■	■	■	■	■	■	■	■
Physical Activity		■						■					
Physical Limitations		■	■	■	■				■	■	■	■	■
Sleep		■	■		■				■				■
Diet		■											
Internal	Nutritional Habits	■							■				
	Diet	■	■	■	■				■				
	Intrinsic properties (age, sex)	■											
	Medical History (personal and family)	■											
	Women Health (eg, pregnancies; Menopausal treatments)	■	■	■	■	■			■	■	■	■	■
	Health problems (eg, diabetes, respiratory health, etc.)	■	■	■	■	■	■	■	■	■	■	■	■
Biological measurements (eg, BMI, blood biochemistry, blood pressure, visual capacity, etc.) ^c	■								c	c	c	c	

^aAccording to each participant.

^bSince 2017 during the health examination at follow-up.

^cAt inclusion and, during the follow-up, every 4 years if possible for the participant.

study subset of exposures, by producing a set of linearly uncorrelated variables called principal components (PCs) without reducing the number of exposures.⁴⁶

Clustering of observations sharing similar exposures

Clustering techniques, such as partitioning, hierarchical, or model-based approaches group observations into mutually exclusive clusters sharing a distinct exposomic profile that could potentially predict well the outcome.^{35,42,43} Hierarchical clustering has been applied in exposomics to determine exposure clusters, however, without considering the spatial dimensions of the dataset.^{37,47} In our example, we aim to group participants at inclusion by their shared exposomic profiles while keeping into consideration their geographic contiguity using the hierarchical clustering on Principal Components (HCPCA) approach that applies geographic constraints.⁴⁷ Briefly, HCPCA allows the combination of three methods, that is, principal component methods, hierarchical clustering, and partitioning clustering, accounting for dissimilarities from non-geographical features as well as geographical distances.⁴⁸

Assuming that collected variables do not vary between inclusion and throughout the follow-up period, the spatiotemporal profiling of the urban exposome would be extended to follow-up time points over the years, beyond its characterization at inclusion. By accurately pinpointing on the geocoding of residential addresses and the associated urban exposures being similarly collected across the follow-up period, one would assess the temporal evolution of the urban exposome, accounting for the repeated measures of the exposomic dataset across space.

Theoretical illustration of the present framework: the case of BMI as an outcome

Study outcome and covariates

Several exposomics studies have previously described human and urban exposome factors to explore the multifactorial biological plausibility of body mass index (BMI) disease process.^{37,43,49,50} In the context of Paris UrbanX, we propose limiting our sample to participants for whom BMI data was available with medical data at both inclusion (between 2012 and 2016) and follow-up (2017-2020). Furthermore, we propose limiting the sample to participants with no personal medical history of cardiovascular, respiratory, osteoarticular fractures and endocrine diseases. Age, sex, educational level, income, professional status, smoking status and physical activity will be considered as confounders based on other CONSTANCES similar studies.^{51,52}

Single exposure-based associations with BMI

The association between multiple exposures and the human trait or phenotype (BMI in this case) may be studied within an exposomic framework using: (i) exposome-wide association studies (ExWAS)^{50,53}; (ii) variable selection techniques, such as deletion-substitution-addition (DSA) algorithm, graphical unit evolutionary stochastic search (GUESS) algorithms or penalized regression methods such as elastic net (ENET) and (iii) dimension reduction techniques, such as sparse partial least squares (sPLS) regression.^{34,35,50} These statistical approaches have been also described in the context of repeated measures datasets, using both simulated and real datasets.⁵⁴ While ExWAS presents with high sensitivity in capturing true exposures associated with the outcome, it is accompanied by a high % of false positives and the repeated measures design is not explicitly considered.⁵⁴ On the other hand, while ENET and DSA are characterized by a low false discovery rate (FDR), yet they do not consider the repeated

measures design. The mixed performances of these methods showcase the lack of a gold method of choice and the need of a careful algorithm selection and interpretation of the findings in each context.⁵⁴

Repeated measures of multiple exposures and associations with BMI

Few methods are available to study the effect of a mixture of exposures on a health outcome.^{50,52} Briefly, these are grouped as: Bayesian Kernel Machine Regression (BKMR), Random Forest, Weighted quantile sum regression (WQS) and Penalized Distributed Lag (Non)Linear Models (DLNMPen).⁵⁵ Novel extensions of these approaches are developed to study the time-varying mixture of repeated exposures in association with health outcome. Next, we briefly describe these approaches and the suggested extensions (Figure 3).

The Bayesian Kernel Machine Regression (BKMR) allows for the estimation of individual and joint effects of multiple exposures, yet this may be troublesome for large population samples requiring large computational time; Bayesian Distributed Lagged Models (BDLM) are a useful extension of BKMR gaining popularity in dealing with repeated measurements of exposures. In BDLM, the outcome is regressed on exposures repeatedly measured during the preceding period.⁵⁶ This method allows identifying critical time windows of exposure mixtures, while accounting for non-linear and non-additive effects across multiple locations.^{57,58} The advantage of this method lies in its ability to dealing with highly correlated exposures in addition to providing relative importance of exposure variables within and among groups, due to the implementation of a hierarchical variable selection scheme.⁵⁹

The random forest algorithm is a decision-tree based method used for classification and regressions purposes as well as identifying the strongest predictors associated with the health outcome, exhibiting high performance in dealing with high-dimensional data.⁵⁵ Nevertheless, the interpretability of the findings is limited to the identification of significant predictors of the outcome, without further insights on the extent of effect, the interaction and dose-response shape.⁵⁵ On the other hand, WQS is highly effective for identifying cumulative mixture effects and is well suited to the complexity and high dimensionality of untargeted exposure profiles. However, it relies on the assumptions that exposures do not interact and that risk changes uniformly across quantile categories. These assumptions can be examined through sensitivity analyses using different quantile schemes (eg, quartiles instead of deciles) and by comparing results with alternative mixture modeling approaches designed for smaller exposure sets, such as BKMR.⁶⁰ Penalized Distributed Lag (Non) Linear Models (DLNMPen) considers the repeated measures design while presenting flexibility in modelling dose-reponse relationships, however, its performance depends on the dataset characteristics given the weakness of the method in handling categorical variables.⁵⁴

Studying exposome associations in a spatio-temporal context

When linking spatial and temporal exposomic data to the individual-level exposome data, two methods appear promising to address this data heterogeneity: calculating area- and time-weighted average to generate individual exposures for each spatio-temporal exposure window, or to apply multi-level models using the original spatial and temporal scales of the dataset.⁵⁶ Nevertheless, further efforts are warranted to evaluate the

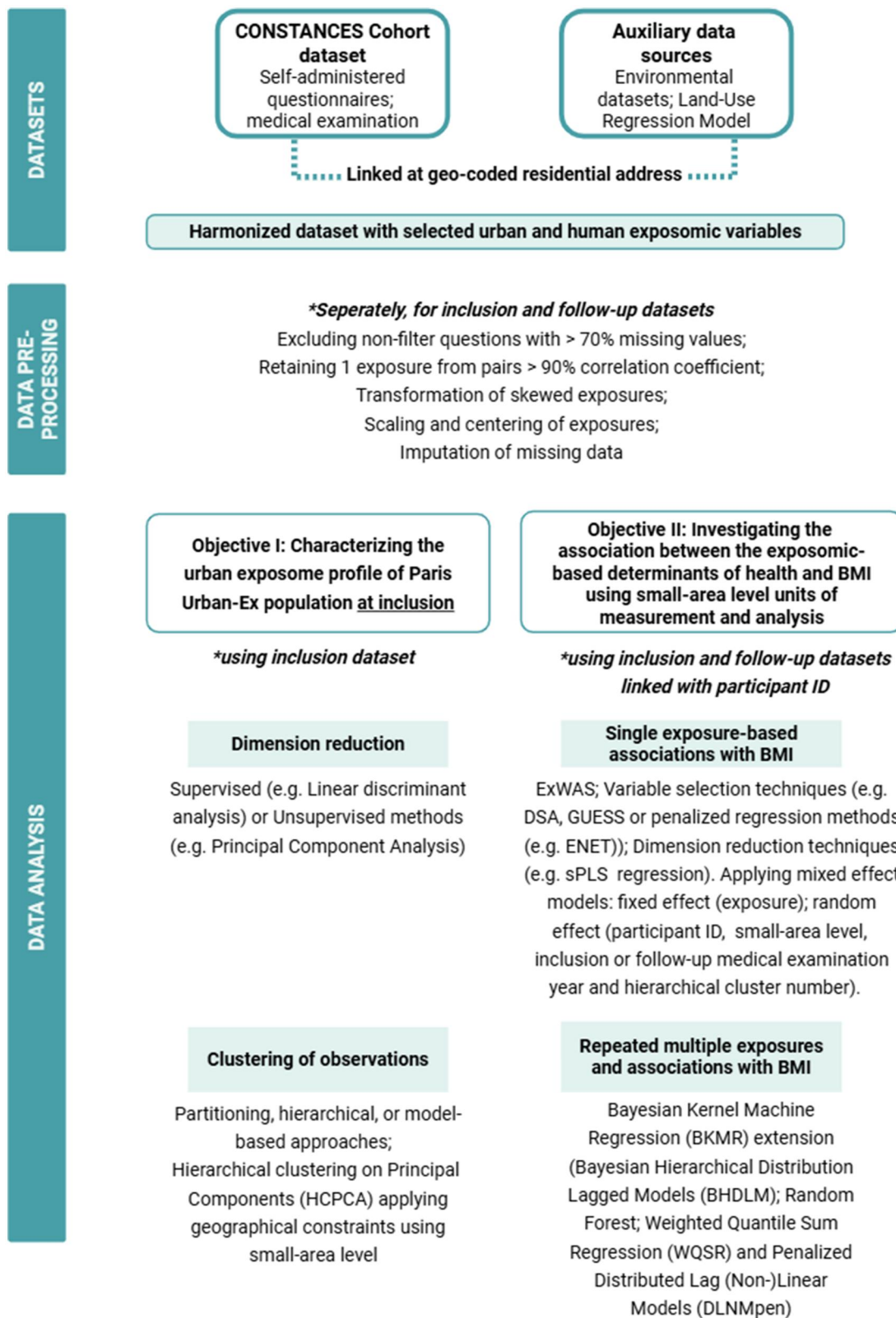


Figure 3. Diagram of the exposomics workflow of dataset harmonization, pre-processing and main data analysis using the urban exposome concept and its exposomics tools in urban settings.

performance of these strategies in their corresponding statistical methods.⁶¹ In our dataset, mixed effect models can be applied considering the exposure as a fixed effect, and the following random effect variables: the participant ID, the small-area level, and inclusion or follow-up medical examination year and

hierarchical cluster number (at inclusion and follow-up), respectively. ExWAS and Penalized Distributed Lag (Non-)Linear Models (DLNMpen) have the ability to consider repeated measures design using mixed effect models, yet with mixed performance.⁵⁴

The implementation of BDLM in a hierarchical framework (Bayesian Hierarchical Distribution Lagged Models (BHDL)) allows for more flexibility in the analysis of data distributed over space and time.⁵⁹ Nevertheless, its application in longitudinal datasets is limited in moderate sample sizes; in large cohort studies with a large number of exposure variables (>50), their performance may be challenged by model instability and high computing time.⁵⁵ A suggestion would be to apply this method to a smaller subsample with a subset of pre-selected variables. Extensions of random forest approaches and penalized regression methods were also recently suggested to accommodate for space- and time-dependent data using the R package *RandomForestsGLS* and *dlnm*, respectively.^{62,63} Yet, the application of these novel approaches is still scarce in urban exposome studies.

The situation may become more complex in multi-omics datasets to be used in urban exposomics studies by integrating a suite of transcriptomics, proteomics, metabolomics, microbiome datasets, etc, that may be available at multiple time points for the study population or not; there are novel algorithms that perform multi-omics integration by treating each omics platform as a separate block, yet there is limited work in this topic.^{9,64,65} The *mixOmics* R package offers a wide range of multivariate methods (eg, sparse PCA, or sparse partial least square analysis or canonical correlation analyses for the exploration and integration of omics datasets, allowing for feature selection.^{64,65}

Discussion

The availability of a multitude of time-varying and spatially-resolved data of a longitudinal cohort of any kind (pregnancy-birth, population-based, administrative cohort) promotes the utility of the urban exposome methodological framework and its exposomic tools in designing exposome studies nested within existing cohorts towards exploring linkages between multiple exposures and health outcomes at a small-area scale of analysis for an urban population. This paper presents a novel methodological framework for integrating longitudinal data into urban exposome research, using a comprehensive, real-world application to the urban population of Paris, France. This work presents an overview of the novel exposomic tools that aim to explore how the global dimensions of time (repeated measures of exposures) and space (small-area level) would influence associations between multiple exposures of the urban exposome profile and selected human traits, such as the BMI. This urban exposome framework within the CONSTANCES cohort is based on collected individual-level information to which layered geographical area-level information obtained from auxiliary sources or multi-omics datasets or sensor-based exposures may be integrated and merged into a single dataset. Our paper provides a comprehensive description of the statistical methods currently used in the exposomic literature, outlining their respective advantages and limitations, explaining the rationale for selecting these approaches, and offering a conceptual demonstration of their application in the CONSTANCES cohort using BMI as an example health outcome. We present a novel, structured synthesis to guide data-integration choices within a cohort setting, emphasizing that interpretation must remain grounded in the underlying data structure and research question.

To our knowledge, this type of an urban exposome study is the first of its kind within the CONSTANCES cohort that considers the simultaneous assessment of the intertwined urban exposures and their synergies across space and time, using IRIS-level data as a unit of analysis. In contrast to other CONSTANCES

population studies that often rely on the “one exposure-one outcome” associations at a time,^{51,52,66,67} the current methodological approach offers the opportunity to deploy agnostic statistical approaches to explore the totality of urban exposome determinants associated with a health outcome, taking BMI as example. Additionally, we provide examples of how the use of additional open-access sources of data regularly collected during our study period and for the small-scale areas of the city of Paris would enrich the span of urban exposome variables. Although some of the collected cohort data was not at IRIS level, the use of available higher-level data would still be beneficial, as it is collected in a standardized manner across administrative levels of France. On the other hand, the availability of small-area level of exposure measurements within an urban setting is usually limited in the globe, highlighting the unique strength of the CONSTANCES cohort in incorporating urban populations and enabling such fine-grained spatial analyses.^{36,68}

Limitations

Several challenges come along with the implementation of this complex urban in cohort studies, as illustrated in the current methodological study. First, the movement of the study population—whether within or outside Paris to the rest of France might not be reported timely or accurately by the participants. This becomes further challenging when attempting to study the duration of exposures at small-area residential level in association with the onset of health outcome. This issue of “migration analysis” is common to urban health studies that assume a long-term residence in the provided address. In our study, we have selected participants that provided only one residential address, assuming their address remained the same during the study period. A sensitivity analysis on the duration of residence including participants that provided two or more residential addresses can as well be suggested.⁶⁹ Also, as we limited our analysis to participants for which an IRIS identifier was available following linkage with geocoded address with an exact-address or street-match precision; an additional sensitivity analysis excluding street-match precision can be conducted.

Selecting variables available at three or more time points in CONSTANCES dataset for performing the repeated measures based longitudinal analysis resulted in the exclusion of certain exposome variables, and thus, of exposome components, particularly those relevant to occupational exposures (collected only once at inclusion), characteristics of housing and household products and exposure to pesticides (collected only once at follow-up). Additionally, the exposure to infectious contaminants is a missing exposure component, which is often understudied in human exposome research.^{24,70}

The lack of repeated measures for urban exposures is also another challenge for urban exposome studies; current strategy is to rely on two time points in cohort studies, or relying on mean values of survey or surveillance schemes in urban centers instead of time-varying exposomic variables, or relying on the sole available information.^{33,35,36,68,70} This underpins the importance of having routinely collected fine-scale information on urban environmental parameters such as air and water pollutants, noise, traffic, UV radiation and other indicators to feed into urban exposomics.

On one hand, in the proposed prospective associations of the Paris urban exposome and BMI, we suggested selecting participants with anthropometrics being assessed during the medical examination at inclusion and follow-up. On the other hand, we suggested excluding participants with a personal medical history

of five medical conditions (cardiovascular, respiratory, osteoarticular fractures and endocrine diseases). A comparative analysis can be suggested also, by excluding participants presenting with any personal medical history.

The novelty in applying the urban exposome framework to future longitudinal studies nested within cohorts

Future urban exposomics studies would benefit from the proposed methodological framework by providing a guided pipeline of steps to implement a longitudinal exposomics study in urban settings. Notably, this framework represents one of the first comprehensive approaches to operationalize the urban exposome concept within a longitudinal, spatially explicit design, integrating diverse exposome domains over time and space.⁷¹

Wealth of layered exposome data collected over time and space

The integration of urban exposome variables, obtained from auxiliary data sources, into the cohort dataset represents a novel approach that enables a more comprehensive opportunity for exploring the wider range of exposome components. In prospective studies, the incorporation of exposures collected at fine spatial and temporal scales using advanced personal or portable monitoring devices and sensors would provide added value to the data, providing to some extent a solution to the availability of standardized geo-spatial data for the study period of interest.

Novel exposomic tools

The wealth of urban exposome data obtained from longitudinal studies requires the use of methods dealing with high-dimensional and multi-level data, thus unraveling the exposures truly associated with the study outcome while accounting for the complex and highly-correlated exposomics structure.^{11,35} Additional considerations need to be paid to the spatio-temporal dynamics of urban external exposome features that are overlaid to multiple individual-level exposures. The exposomics tools should be capable of handling a large set of time-varying exposures spanning over the different urban and human exposome domains, while accounting for multicollinearity, possible nonlinearity, as well as the multi-level structure of the data.³⁵ Furthermore, the longitudinal aspect of the data implies the necessity for applying methods dealing with repeated assessment of exposures and outcomes, the time-varying cumulative sets of exposures and confounders, windows of susceptibility, among others. Despite the numerous statistical approaches proposed in the literature, many are still reserved to simulation studies with limited—and require further use in human exposome datasets.^{54,55,63} In the absence of a ‘gold standard’ data processing set of algorithms, the need for careful interpretation of findings in each context is warranted, particularly in relation to the study design characteristics and the dataset strengths and weaknesses keeping in mind the study’s specific objectives and research questions. Nevertheless, Artificial Intelligence (AI) is growing into a revolutionary asset in exposome research: with its advanced analytical tools, particularly machine learning (ML) and deep learning techniques, AI is unlocking new potential in enabling automated data processing, pattern recognition and predictive modelling using complex and exposome datasets layered over space and time.^{72,73}

Integration with multi-omics platforms

The collected biological samples as part of a cohort research infrastructure would allow for the implementation of multi-omics platforms, such as, transcriptomics, metabolomics, or microbiome, etc. Although none so far have been applied to CONSTANCES,¹⁶ the multi-omics opportunities in an urban exposomics setting represent a novelty that would be more frequently considered in the near future.

Expansion to multiple cohorts

Last, the application of this methodological framework could extend coverage from a single cohort to multiple cohorts spanning over different cities and urban policies. Such urban exposomics datasets provide policymakers with evidence-based knowledge to designing and implementing urban interventions that are tailored to the characteristics of the specific communities from the identified clusters within each city, accounting for the city’s exposome temporal trajectory changes.

Acknowledgments

The authors thank the team of the “Population-based cohorts unit” (Cohortes en population) that designed and manages the Constances cohort study. They also thank the French national health insurance fund (“Caisse nationale d’assurance maladie”, Cnam) and its Health screening centres (“Centres d’examens de sante”), which are collecting a large part of the data, as well as the French national old-age insurance fund (“Caisse nationale d’assurance vieillesse”, Cnav) for its contribution to the constitution of the cohort, and ClinSearch, Asqualab and Eurocell, which are conducting the data quality control.

Author contributions

Nadine Haddad (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Methodology [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Emeline Lequy (Investigation [equal], Project administration [equal], Validation [equal], Visualization [equal], Writing—review & editing [equal]), Marie Zins (Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Supervision [equal], Writing—review & editing [equal]), and Marcel Goldberg (Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Supervision [equal], Validation [equal], Writing—review & editing [equal]), Konstantinos C. Makris (Conceptualization [Lead], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Supervision [equal], Validation [equal], Writing—review & editing [equal]), Writing—original draft [equal], Writing—review & editing [equal])

Supplementary material

Supplementary material is available at *Exposome* online.

Funding

The CONSTANCES cohort study was supported and funded by the French national health insurance fund (“Caisse nationale de l’Assurance Maladie”, Cnam). CONSTANCES is a National infrastructure for biology and health (“Infrastructure nationale en

biologie et santé”) and benefits from a grant from the French national agency for research (ANR-11-INBS-0002). CONSTANCES is also partly funded to a small extent by industrial companies, notably in the healthcare sector, within the framework of Public-Private Partnerships (PPP). KCM would like to thank the funding by the Horizon Europe IHEN project on the International Human Exposome Network. None of these funding sources had any role in the design of the study, collection and analysis of data or decision to publish.⁷⁴

Conflicts of interest

None declared.

Data availability

In accordance with the Constances Charter, deidentified participant data from the Constances Cohort are available to researchers who meet the legal and ethical requirements set by the French National Commission governing data privacy laws. International researchers can access the dataset by following the procedure outlined here. Additionally, all study materials, including the study protocol and data dictionary of the Constances Cohort, are freely accessible.

Ethics approval and consent to participate

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures have been approved by the Institutional review board (IRB) of the French Institute of Health (Inserm) (Opinion n°01-011, then n°21-842), and authorized by the by the French Data Protection Authority (“Commission Nationale de l’Informatique et des Libertés”, CNIL) (Authorization #910486). All participants signed a written consent form for their participation in Constances. All information and regulatory authorizations relating to the publication are available on the French version page “Rights and data protection.”^{74,75}

References

1. United Nations, Department of Economic and Social Affairs. *Sustainable Development. Goal 11: Make Cities and Human Settlements Inclusive, Safe, Resilient and Sustainable*. United Nations Department of Economic and Social Affairs. Accessed December 19, 2025. https://sdgs.un.org/goals/goal11#targets_and_indicators.
2. Galea S, Vlahov D. Urban health: evidence, challenges, and directions. *Annu Rev Public Health*. 2005;26:341–365. <https://doi.org/10.1146/annurev.publhealth.26.021304.144708>
3. Schröder J, Moebus S, Skodra J. Selected research issues of urban public health. *Int J Environ Res Public Health*. 2022;19:5553. <https://doi.org/10.3390/ijerph19095553>
4. World Health Organization. *Urban Health*. Accessed October 10, 2025. <https://www.who.int/news-room/fact-sheets/detail/urban-health>.
5. Galea S, Schulz A. Methodological considerations in the study of urban health. How do we best assess how cities affect health? In *Cities and the Health of the Public*. Washington University in St Louis. Vanderbilt University Press; 2006. <https://doi.org/10.2307/j.ctv1622mv6>. <https://doi.org/10.2307/j.ctv1622mv6.17>
6. Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*. 2005;14:1847–1850. <https://doi.org/10.1158/5-9965.EPI-5-0456>
7. Andrianou X, Charisiadis P, Makris KC. The urban exposome framework and a proof-of-concept study. *Environ Epidemiol*. 2019;3:257. <https://doi.org/10.1097/01.EE9.0000608732.36531.e1>
8. Bodein A, Scott-Boyer MP, Perin O, Lê Cao KA, Droit A. timeOmics: an R package for longitudinal multi-omics data integration. *Bioinformatics*. 2022;38:577–579. <https://doi.org/10.1093/bioinformatics/btab664>
9. Bodein A, Chapleur O, Droit A, Lê Cao KA. A generic multivariate framework for the integration of microbiome longitudinal studies with other data types. *Front Genet*. 2019;10:963. <https://doi.org/10.3389/fgene.2019.00963>
10. Maitre L, de Bont J, Casas M, et al. Human early life exposome (HELIX) study: a European population-based exposome cohort. *BMJ Open*. 2018;8:e021311. <https://doi.org/10.1136/bmjopen-7-021311>
11. Vlaanderen J, de Hoogh K, Hoek G, et al. Developing the building blocks to elucidate the impact of the urban exposome on cardiometabolic-pulmonary disease: The EU EXPANSE project. *Environ Epidemiol*. 2021;5:e162. <https://doi.org/10.1097/EE9.000000000000162>
12. Vrijheid M, Basagaña X, Gonzalez JR, et al. Advancing tools for human early lifecourse exposome research and translation (ATHLETE): Project overview. *Environ Epidemiol*. 2021;5:e166. <https://doi.org/10.1097/EE9.000000000000166>
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet* 2007; 370:1453–1457. [https://doi.org/10.1016/S0140-6736\(07\)61602-X](https://doi.org/10.1016/S0140-6736(07)61602-X)
14. Zins M, Goldberg M; Constances Team. The French CONSTANCES population-based cohort: design, inclusion and follow-up. *Eur J Epidemiol*. 2015;30:1317–1328.
15. CONSTANCES Cohorte. Cohort of health examination center consultants. Accessed July 20, 2025. <https://www.constances.fr/en/home/>
16. Henny J, Nadif R, Got S, et al. The CONSTANCES Cohort Biobank: an open tool for research in epidemiology and prevention of diseases—PubMed. *Front Public Health*. 2020;8:605133.
17. Zins M, Berkman LF, Goldberg M. The CONSTANCES cohort, an epidemiological research infrastructure. Methods and results of the pilot phase. *Epidemiol Biostat Public Health* 2022;10. <https://doi.org/10.2427/8921>
18. Institut national de la statistique et des études économiques. *Définitions, méthodes et qualité. Géographie administrative et d'étude*. Département de Paris; 2024. Accessed April 27, 2024. <https://www.insee.fr/fr/metadonnees/geographie/departement/75-paris>
19. Wikipedia. *Administrative divisions of France*. In: 2024. Accessed April 27, 2024. https://en.wikipedia.org/w/index.php?title=Administrative_divisions_of_France&oldid=1214891182
20. Géoservices. *Contours IRIS*. Accessed April 27, 2024. <https://geoservices.ign.fr/contoursiris>
21. European Environment Agency. *Degree of Urbanisation (DEGURBA)*. 2024. Accessed April 16, 2024. <https://www.eea.europa.eu/data-and-maps/data/external/degree-of-urbanisation-degurba>.

22. Institut national de la statistique et des études économiques. *Table d'appartenance géographique des IRIS*. 2024. Accessed August 18, 2024. <https://www.insee.fr/fr/information/7708995>.
23. CONSTANCES Cohorte. *Scientific Area. Data*. 2024. Accessed August 18, 2024. <https://www.constances.fr/en/scientific-area/data/>
24. Haddad N, Andrianou XD, Makris KC. A scoping review on the characteristics of human exposome studies. *Curr Pollution Rep*. 2019;5:378–393. <https://doi.org/10.1007/s40726-9-00130-7>
25. CONSTANCES Cohorte. *Occupational and Residential Data*. 2024. Accessed August 18, 2024. <https://www.constances.fr/en/scientific-area/data/professional-and-residential-data>
26. de Hoogh K, Chen J, Gulliver J, et al. Spatial PM_{2.5}, NO₂, O₃ and BC models for Western Europe—Evaluation of spatiotemporal stability. *Environ Int*. 2018;120:81–92. <https://doi.org/10.1016/j.envint.2018.07.036>
27. Vienneau D, de Hoogh K, Bechle MJ, et al. Western european land use regression incorporating satellite- and ground-based measurements of NO₂ and PM₁₀. *Environ Sci Technol*. 2013;47:13555–13564. <https://doi.org/10.1021/es403089q>
28. Ramos A, Zare Sakhvidi MJ, Lafontaine A, et al. Association between greenspace exposure and different domains of cognitive function in the French CONSTANCES cohort. Presented at: *Annual Conference of the International Society of Environmental Epidemiology (ISEE)*; September 2022; Athens, Greece. The Comprehensive R Archive Network. Accessed July 2, 2024. <https://cran.r-project.org/>
29. The Comprehensive R Archive Network. Accessed July 2, 2024. <https://cran.r-project.org/>
30. Download RStudio—Posit. Accessed July 2, 2024. <https://www.rstudio.com/>
31. Dominguez A, Anguita-Ruiz A, Gonzalez JR, Basagaña X. Tutorial: Exposome data analysis. *Toolbox—Athlete*. 2024. Accessed April 16, 2024. <https://athleteproject.eu/toolbox>.
32. Escriba-Montagut X, Basagaña X, Vrijheid M, Gonzalez JR. Software application profile: exposomeShiny—a toolbox for exposome data analysis. *Int J Epidemiol*. 2022;51:18–26. <https://doi.org/10.1093/ije/dyab220>
33. Julvez J, López-Vicente M, Warembourg C, et al. Early life multiple exposures and child cognitive function: a multicentric birth cohort study in six European countries. *Environ Pollut*. 2021;284:117404. <https://doi.org/10.1016/j.envpol.2021.117404>
34. Nieuwenhuijsen MJ, Agier L, Basagaña X, et al. Influence of the urban exposome on birth weight. *Environ Health Perspect*. 2019;127:47007. <https://doi.org/10.1289/EHP3971>
35. Santos S, Maitre L, Warembourg C, et al. Applying the exposome concept in birth cohort research: a review of statistical approaches. *Eur J Epidemiol*. 2020;35:193–204. <https://doi.org/10.1007/s10654-0-00625-4>
36. Robinson O, Tamayo I, de Castro M, et al. The urban exposome during pregnancy and its socioeconomic determinants. *Environ Health Perspect*. 2018;126:077005. <https://doi.org/10.1289/EHP2862>
37. Ohanyan H, Portengen L, Huss A, et al. Machine learning approaches to characterize the obesogenic urban exposome. *Environ Int*. 2022;158:107015. <https://doi.org/10.1016/j.envint.2021.107015>
38. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Soft*. 2011;45:1–67. <https://doi.org/10.18637/jss.v045.i03>
39. Guimbaud J-B, Siskos AP, Sakhi AK, et al. Machine learning-based health environmental-clinical risk scores in European children. *Commun Med (Lond)*. 2024;4:98. <https://doi.org/10.1038/s43856-4-00513-y>
40. Shahbazi Z, Nowaczyk S. Towards personalized cardiometabolic risk prediction: A fusion of exposome and AI. *Heliyon* 2025; 11:e40859. <https://doi.org/10.1016/j.heliyon.2024.e40859>
41. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012; 28:112–118. <https://doi.org/10.1093/bioinformatics/btr597>
42. Stafoggia M, Breitner S, Hampel R, Basagaña X. Statistical approaches to address multi-pollutant mixtures and multiple exposures: the state of the science. *Curr Environ Health Rep*. 2017; 4:481–490. <https://doi.org/10.1007/s40572-7-0162-z>
43. Maitre L, Guimbaud J-B, Warembourg C, et al.; Exposome Data Challenge Participant Consortium. State-of-the-art methods for exposure-health studies: results from the exposome data challenge event. *Environ Int*. 2022;168:107422. <https://doi.org/10.1016/j.envint.2022.107422>
44. Kalia V, Walker DI, Krasnodemski KM, Jones DP, Miller GW, Kioumourtzoglou MA. Unsupervised dimensionality reduction for exposome research. *Curr Opin Environ Sci Health*. 2020;15: 32–38. <https://doi.org/10.1016/j.coesh.2020.05.001>
45. Bioinformatics Research Group in Epidemiology (BRGE) - Barcelona Global Health Institute. *rexposome project*. 2024. Accessed October 28, 2024. <https://isglobal-brge.github.io/rexposome>
46. Greenacre M, Groenen PJF, Hastie T, et al. Principal component analysis. *Nat Rev Methods Primers*. 2022;2:100. <https://doi.org/10.1038/s43586-2-00184-w>
47. >Song J, Du P, Yi W, et al. Using an exposome-wide approach to explore the impact of urban environments on blood pressure among adults in Beijing-Tianjin-Hebei and surrounding areas of China. *Environ Sci Technol*. 2022;56:8395–8405. <https://doi.org/10.1021/acs.est.1c08327>
48. Chavent M, Kuentz-Simonet V, Labenne A, Saracco J. ClustGeo: an R package for hierarchical clustering with spatial constraints. *Comput Stat*. 2018;33:1799–1822. <https://doi.org/10.1007/s00180-8-0791-1>
49. Haddad N, Andrianou X, Parrish C, Oikonomou S, Makris KC. An exposome-wide association study on body mass index in adolescents using the National Health and Nutrition Examination Survey (NHANES) 3–2004 and 3–2014 data. *Sci Rep*. 2022;12:8856. <https://doi.org/10.1038/s41598-2-12459-z>
50. Vrijheid M, Fossati S, Maitre L, et al. Early-life environmental exposures and childhood obesity: an exposome-wide approach. *Environ Health Perspect*. 2020;128:67009. <https://doi.org/10.1289/EHP5975>
51. Feral-Pierssens A-L, Carette C, Rives-Lange C, et al. Obesity and emergency care in the French CONSTANCES cohort. *PLoS One*. 2018;13:e0194831. <https://doi.org/10.1371/journal.pone.0194831>
52. Czernichow S, Renuy A, Rives-Lange C, et al. Evolution of the prevalence of obesity in the adult population in France, 3–2016: the Constances study. *Sci Rep*. 2021;11:14152. <https://doi.org/10.1038/s41598-1-93432-0>
53. Chung MK, House JS, Akhtari FS, et al.; Members of the Exposomics Consortium. Decoding the exposome: data science methodologies and implications in exposome-wide association studies (ExWASs). *Exposome*. 2024;4:osae001. <https://doi.org/10.1093/exposome/osae001>
54. Warembourg C, Anguita-Ruiz A, Siroux V, et al. Statistical approaches to study exposome-health associations in the context of repeated exposure data: a simulation study. *Environ Sci Technol*. 2023;57:16232–16243. <https://doi.org/10.1021/acs.est.3c04805>
55. Peters S, Wan W, Portengen L, Krone T, Ge C, Vermeulen R; WE-EXPOSE Consortium. *Report on Tutorial for the Application of a Suite of 'multiple exposure methods'*. WE-EXPOSE (Working Life Exposome for Policy, OSH and Science) Toolbox; 2022. Accessed December 18, 2025. https://www.we-expose.eu/multiple_exposure_0102022.pdf

56. Antonelli J, Wilson A, Coull BA. Multiple exposure distributed lag models with variable selection. *Biostatistics*. 2023;25:1–19. <https://doi.org/10.1093/biostatistics/kxac038>
57. Peng RD, Dominici F, Welty LJ. A Bayesian hierarchical distributed lag model for estimating the time course of risk of hospitalization associated with particulate matter air pollution. *J R Stat Soc Ser C*. 2009;58:3–24.
58. Liu SH, Bobb JF, Claus Henn B, et al. Modeling the health effects of time-varying complex environmental mixtures: mean field variational Bayes for lagged kernel machine regression. *Environmetrics*. 2018;29:e2504. <https://doi.org/10.1002/env.2504>
59. Wikle CK, Berliner LM, Cressie N. Hierarchical Bayesian space-time models. *Environ Ecol Stat*. 1998;5:117–154. <https://doi.org/10.1023/A:1009662704779>
60. Young AS, Gennings C, Eick SM, Liang D, Walker DI. A statistical workflow for analyzing the untargeted chemical exposome and metabolome in epidemiologic studies using high-dimensional mixture methods. *Exposome*. 2025;5:osaf010. <https://doi.org/10.1093/exposome/osaf010>
61. Hu H, Liu X, Zheng Y, et al. Methodological challenges in spatial and contextual exposome-health studies. *Crit Rev Environ Sci Technol*. 2023;53:827–846. <https://doi.org/10.1080/10643389.2022.2093595>
62. Saha A, Basu S, Datta A. RandomForestsGLS: An R package for Random Forests for dependent data. *J Open Source Softw*. 2022;7:3780. <https://doi.org/10.21105/joss.03780>
63. Gasparrini A, Scheipl F, Armstrong B, Kenward MG. A penalized framework for distributed lag non-linear models. *Biometrics*. 2017;73:938–948. <https://doi.org/10.1111/biom.12645>
64. Shen H, Huang JZ. Sparse principal component analysis via regularized low rank matrix approximation. *J Multivar Anal*. 2008;99:1015–1034. <https://doi.org/10.1016/j.jmva.2007.06.007>
65. Tenenhaus A, Philippe C, Guillemot V, Le Cao K-A, Grill J, Frouin V. Variable selection for generalized canonical correlation analysis. *Biostatistics*. 2014;15:569–583. <https://doi.org/10.1093/biostatistics/kxu001>
66. Franck JE, Ringa V, Rigal L, et al. Patterns of gynaecological check-up and their association with body mass index within the CONSTANCES cohort. *J Med Screen*. 2021;28:10–17. <https://doi.org/10.1177/0969141320914323>
67. Crampe-Casnabet C, Franck JE, Ringa V, Coeuret-Pellicer M, Chauvin P, Menvielle G. Role of obesity in differences in cervical cancer screening rates by migration history. The CONSTANCES survey. *Cancer Epidemiol*. 2019;58:98–103. <https://doi.org/10.1016/j.canep.2018.11.009>
68. Dzhambov AM, Markevych I, Lercher P. Associations of residential greenness, traffic noise, and air pollution with birth outcomes across Alpine areas. *Sci Total Environ*. 2019;678:399–408. <https://doi.org/10.1016/j.scitotenv.2019.05.019>
69. Yu J, Dwyer-Lindgren L, Bennett J, et al. A spatiotemporal analysis of inequalities in life expectancy and 20 causes of mortality in sub-neighbourhoods of Metro Vancouver, British Columbia, Canada, 0–2016. *Health Place*. 2021;72:102692. <https://doi.org/10.1016/j.healthplace.2021.102692>
70. Andrianou XD, Pronk A, Galea KS, et al. Exposome-based public health interventions for infectious diseases in urban settings. *Environ Int*. 2021;146:106246. <https://doi.org/10.1016/j.envint.2020.106246>
71. Makris KC, Baccarelli A, Silverman EK, Wright RO. How exposomic tools complement and enrich genomic research. *Cell Genom*. 2025;5:100952. <https://doi.org/10.1016/j.xgen.2025.100952>
72. Human exposome research: Potential, limitations and public policy implications. Think Tank. European Parliament; 2025. Accessed December 6, 2025. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2025\)765791](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2025)765791)
73. Guimbaud JB, Calabre E, de Cid R, et al. An informed machine learning-based environmental risk score for hypertension in European adults. *Artif Intell Med*. 2025;165:103139. <https://doi.org/10.1016/j.artmed.2025.103139>
74. CONSTANCES Cohorte. *You're about to Publish*. October 17, 2024. Accessed March 14, 2025. <https://www.constances.fr/en/scientific-area/youre-about-to-publish-2>.
75. CONSTANCES Cohorte. *Droits et protection des données—Constances*. October 18, 2023. Accessed March 14, 2025. <https://www.constances.fr/espace-volontaires/droits-et-protection-des-donnees>