

Developing the building blocks to elucidate the impact of the urban exposome on cardiometabolic-pulmonary disease

The EU EXPANSE project

Jelle Vlaanderen^a, Kees de Hoogh^{a,b,c}, Gerard Hoek^a, Annette Peters^d, Nicole Probst-Hensch^{b,c}, Augustin Scalbert^e, Erik Melén^f, Cathryn Tonne^g, G. Arline de Wit^{h,i}, Marc Chadeau-Hyam^{a,j}, Klea Katsouyanni^{j,k}, Tõnu Esko^l, Karin R. Jongsma^m, Roel Vermeulen^{a,h,j}; the EXPANSE Consortium

Abstract. By 2030, more than 80% of Europe's population will live in an urban environment. The urban exposome, consisting of factors such as where we live and work, where and what we eat, our social network, and what chemical and physical hazards we are exposed to, provides important targets to improve population health. The EXPANSE (EXposome Powered tools for healthy living in urbAN SETtings) project will study the impact of the urban exposome on the major contributors to Europe's burden of disease: Cardio-Metabolic and Pulmonary Disease. EXPANSE will address one of the most pertinent questions for urban planners, policy makers, and European citizens: "How to maximize one's health in a modern urban environment?" EXPANSE will take the next step in exposome research by (1) bringing together exposome and health data of more than 55 million adult Europeans and OMICS information for more than 2 million Europeans; (2) perform personalized exposome assessment for 5,000 individuals in five urban regions; (3) applying ultra-high-resolution mass-spectrometry to screen for chemicals in 10,000 blood samples; (4) evaluating the evolution of the exposome and health through the life course; and (5) evaluating the impact of changes in the urban exposome on the burden of cardiometabolic and pulmonary disease. EXPANSE will translate its insights and innovations into research and dissemination tools that will be openly accessible via the EXPANSE toolbox. By applying innovative ethics-by-design throughout the project, the social and ethical acceptability of these tools will be safeguarded. EXPANSE is part of the European Human Exposome Network.

Keywords: Exposome; Urban exposome; Cardiometabolic disease; Pulmonary disease; European Human Exposome Network; OMICS; Ultra-high-resolution massspectrometry; Life course epidemiology; Ethics parallel research

Background

Our environment has a significant impact on our health, explaining an estimated 70% of the (noncommunicable) chronic disease burden.¹ As most of the aspects of our environment are modifiable, this provides a huge potential for disease prevention. Yet, information on specific risk factors, their interactions, and the most effective interventions for disease prevention is far from complete. Reasons for the existence of these knowledge gaps include (1) limitations in the quantification of environmental exposures, both in coverage (the number of exposures that can be quantified) and resolution (the quality with which we

can quantify these exposures); (2) inability of current methods to agnostically explore risk factors; (3) misspecification of disease models, in particular when addressing exposure across the life course; and (4) inadequate approaches for tackling the complexity of real-life exposure scenarios, involving many concomitant exposures (i.e., mixtures) as well as complex and possibly genetically driven interactions within hosts.

To address these knowledge gaps scientists in Europe and the United States conceptualized the exposome.²⁻⁵ Derived from the term exposure, the exposome is the sum of all nongenetic drivers of health and diseases throughout the life course (Figure 1). Interacting with the genome and modifying its penetrance to

^aInstitute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands; ^bSwiss Tropical Health, Basel, Switzerland; ^cUniversity of Basel, Switzerland; ^dHelmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; ^eInternational Agency for Research on Cancer (IARC), Biomarkers Group, Lyon, France; ^fDepartment of Clinical Science and Education Södersjukhuset, Karolinska Institutet, Stockholm, Sweden; ^gBarcelona Institute for Global Health (ISGlobal), Universitat Pompeu Fabra, CIBER Epidemiología y Salud Pública, Barcelona, Spain; ^hDepartment of health care innovation and evaluation, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; ⁱCentre for Nutrition, Prevention and Healthcare. National Institute of Public Health and the Environment, Bilthoven, the Netherlands; ^jImperial College London, London, United Kingdom; ^kNational and Kapodistrian University of Athens, Athens, Greece; ^lUniversity of Tartu, Tartu, Estonia; and ^mDepartment of Medical Humanities, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.

EXPANSE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 874627 and is coordinated by Utrecht University.

Access to the tools and data generated within the EXPANSE project will be provided through the exposome Hub (see manuscript for details).

Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.enviroepidem.com).

The EXPANSE Consortium members can be viewed by accessing the following link (<http://links.lww.com/EE/A146>).

*Corresponding Author. Address: Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands. E-mail: r.c.h.vermeulen@uu.nl (R. Vermeulen).

Copyright © 2021 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The Environmental Epidemiology. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Environmental Epidemiology (2021) 00:e162

Received: 23 December 2020; Accepted 1 June 2021

Published online 1 July 2021

DOI: 10.1097/EE9.000000000000162

diseases, it defines individual health at different life stages. Different domains within the exposome have been studied, for example particular age groups (e.g., early or late-life),⁶ diseases,⁷ or exposure circumstances (chemical, social, and urban).⁸ The EXPANSE project (EXposome Powered tools for healthy living in urBAN Settings) focuses on the urban exposome.

Currently, 72% of Europe’s population lives in an urban environment. This percentage is expected to grow to more than 80% by 2030.⁹ It is therefore of utmost importance that European urban areas provide a healthy living environment for all and are designed to minimize health inequalities.¹⁰ Socioeconomically disadvantaged populations are more vulnerable to adverse

health effects due to their living environment and have limited opportunities to move to “healthier environments,” also within cities.¹¹ This is of great public health importance as significant differences in life expectancy are observed not only between different European urban populations but also within regions and cities.^{12,13} Traditionally, health research in urban environments has singled out exposures such as the social environment, indoor and outdoor air pollution, noise, heat, lack of green and blue space, or water and food contamination. Applying the exposome concept to the urban environment domain provides the opportunity to establish the interrelations between these features and upstream causes and investigate how these contribute

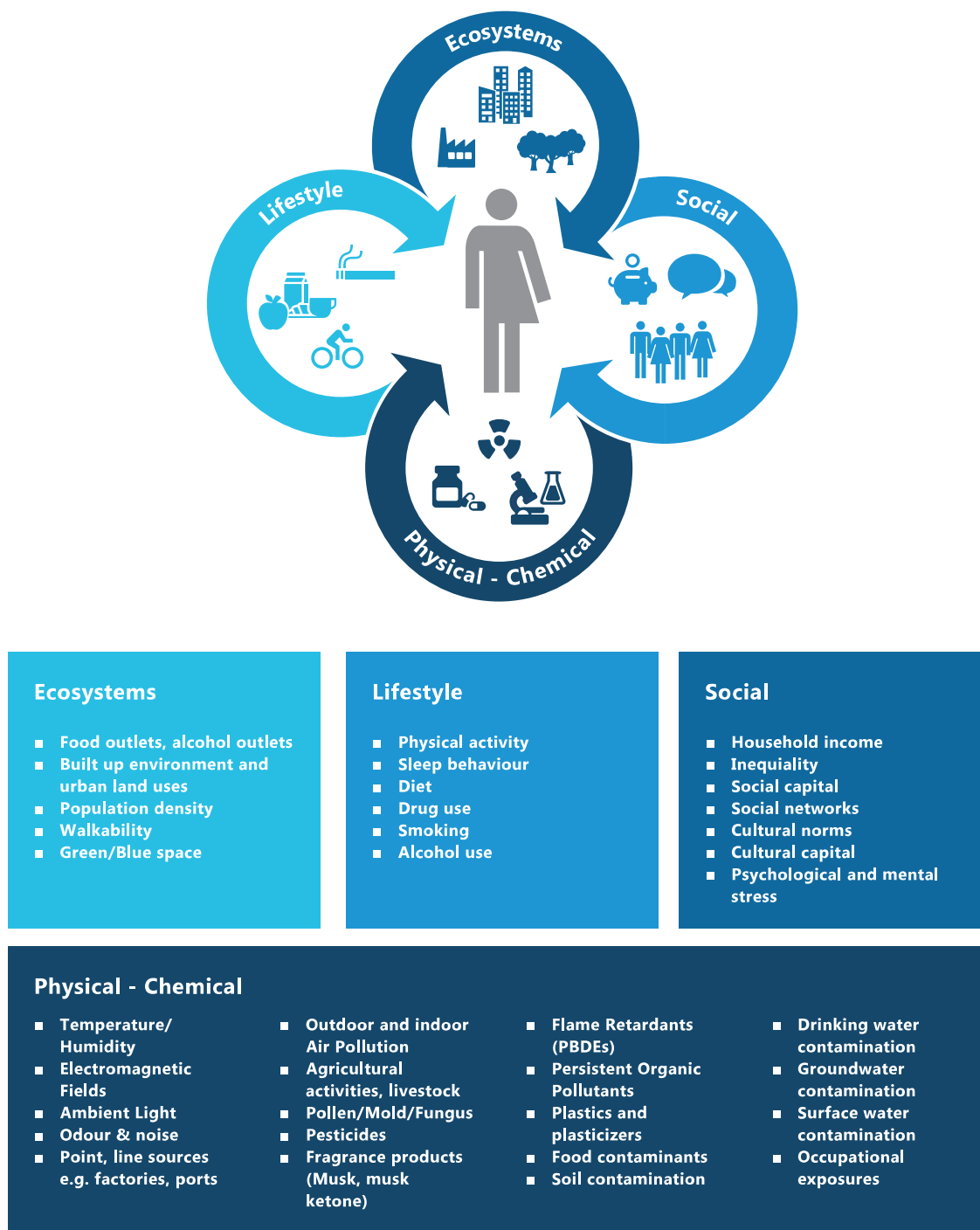


Figure 1. Four dimensions of the exposome and a nonexhaustive list of specific exposures within these dimensions.⁵ The exposome concept tries to capture the diversity and range of these four dimensions, as well as their corresponding biological responses. The urban exposome is an important and integral part of the exposome.

jointly to individual health and quality of aging. This will ultimately contribute to a system-level understanding of urban environments, the behavior of citizens in these environments, their consequence (both positive and negative) for health, and provides entry points for approaches to improve health such as structural interventions or agentostructural (structural changes facilitating healthier choices; i.e., nudging) interventions.¹⁴

Here, we present the design of the EXPANSE project. EXPANSE will apply the exposome concept to the complex interplay between the urban built, social, physicochemical, food, and lifestyle environment in relation to cardiometabolic and pulmonary disease across the life course.

Project description

EXPANSE will address one of the most pertinent issues for urban planners, policy makers, and European citizens: “How to improve one’s health and wellbeing in a modern urban environment?” EXPANSE thereby contributes to several United Nations Sustainable Development Goals (SDGs),¹⁵ in particular SDG 3 (“good health and well-being”), 9 (“industry, innovation, and infrastructure”), 11 (“sustainable cities and communities”), and 13 (“climate action”) and is aligned with the objectives in the European Green Deal and the EU zero pollution policies.

The main research questions addressed in EXPANSE are as follows:

- What aspects in the urban environment contribute most to cardiometabolic and pulmonary disease risk, and how do these factors interact?
- Can we use biological markers of (complex) exposures and of biological effect to get a better understanding of the etiological pathways that connect the urban environment to health, aging, and cardiometabolic and pulmonary disease risk toward improved causal understanding?
- Which intervention strategies reduce exposure to the most damaging aspects of the urban environment and which intervention strategies most effectively promote health?

Approach

EXPANSE brings together large datasets on urban populations, including cohorts based on administrative data (covering more than 55 million individuals) and individual adult and matured birth cohorts (covering more than 2 million individuals) with highly personalized data prospectively collected in EXPANSE’s Urban Laboratories. EXPANSE increases the coverage of the External exposome by extending the number of exposures that are included and by refining the resolution with which these exposures are assessed in space and time, collectively contributing to the number of data elements (e.g., environmental stressors, exposome factors, and molecular markers) that is collected per individual. As such all datasets used in EXPANSE can be placed on a two-dimensional spectrum formed by the number of individuals and number of data elements (Figure 2). Through the integration of datasets that have a different balance in number of individuals and number of data elements (e.g., administrative cohorts versus urban laboratories), we strive towards replication of results across study designs, studies, and molecular (OMICS) levels. This triangulation approach will allow interrogating mediating (molecular) pathways from the external exposome to preclinical and clinical morbidity, as well as mortality. By including both adult and matured birth cohorts (in which participants have reached adulthood), we will be able to apply accelerated longitudinal design and trajectory methods¹⁶ to study the exposome across the life course (i.e., from birth to adult health). The combination of large data sets and novel technologies (such as tracking) and OMICS approaches requires ethical reflection from the onset of the project. For example, to weigh the proportionality of the potentially large benefits and

harm, evaluate the goals, and investigate responsibilities. Ethics parallel research¹⁷ is therefore incorporated from the beginning of the project in order to enhance ethically sound innovation, which has taken into account the preferences and needs of relevant stakeholders.

EXPANSE’s objective is to translate its research findings into actionable tools for user-stakeholders such as the general public, urban planners, policy makers, and other researchers. In Figure 3 we illustrate the toolbox through which findings and methods of the research project will be disseminated. It contains research tools developed within EXPANSE, the Exposome Navigator (focused on dissemination toward the individual), the Reference Exposome (focused on dissemination toward other researchers), and the Exposome Maps (focused on dissemination toward policy makers and urban planners).

EXPANSE focuses on cardiometabolic and pulmonary disease

The EXPANSE infrastructure facilitates studying the impact of the urban exposome on a vast array of health outcomes, in particular cardiometabolic and pulmonary disease (CMPD). CMPD share important features: (1) strong environmental and lifestyle determinants with evidence for an impact across the life course¹⁸; (2) strong variation in prevalence related to socioeconomic status^{19,20}; (3) shared causative pathways involving genetic and epigenetic factors (including, among others, oxidative stress, inflammation, and mitochondrial dysfunction)^{18,21,22}; and (4) alarming increasing trends in incidence in urban societies in Europe and world-wide.²³ We will study the link between the urban exposome and the incidence of acute myocardial infarction (AMI) and stroke (which are among the leading causes of death in Europe), Type 2 diabetes, and chronic obstructive pulmonary disease (COPD). To fully understand the impact of the urban exposome on the burden of CMPD, we will move away from one exposure and one disease at a time toward the exploration of combinations of determinants and mediators over the full life span, accounting for variations in these factors over time. All used data sources (administrative, adult, and child cohorts) encompass multiple health endpoints allowing broadening of the disease scope of EXPANSE over time.

To answer the three main questions defined within EXPANSE, we developed five major focus areas: (1) measure the external urban exposome; (2) measure the internal urban exposome; (3) link the urban exposome to CMPD; (4) evaluate intervention strategies to prevent CMPD; and (5) develop and apply novel enabling analytical methods.

Focus areas and methods

Measure the external exposome

Our goal is to comprehensively characterize the external urban exposome, including the food, lifestyle, social, built, and physicochemical environments (Figure 1). A wealth of data on environmental stressors, such as regulated air pollutants and green space, has been developed in the past decade.^{24,25} For most of these stressors, the simple direct linking of existing geo-spatial maps to assess individual exposure is not optimal, as it generally lacks the relevant information on the spatiotemporal aspects of exposure. Additional work is therefore needed to improve exposure assessment, for example, by improving the spatial scale of existing data, such as fine-resolution characterization of “street-scapes,” or by better characterizing individual exposures using time survey data, and applying historical data to capture temporal trends in exposures.

One of the crucial aims for measuring the external exposome is to develop harmonized exposure datasets across the EU. For a number of stressors, these are already available, but often at a too coarse scale to capture intraurban variability, at a too low temporal

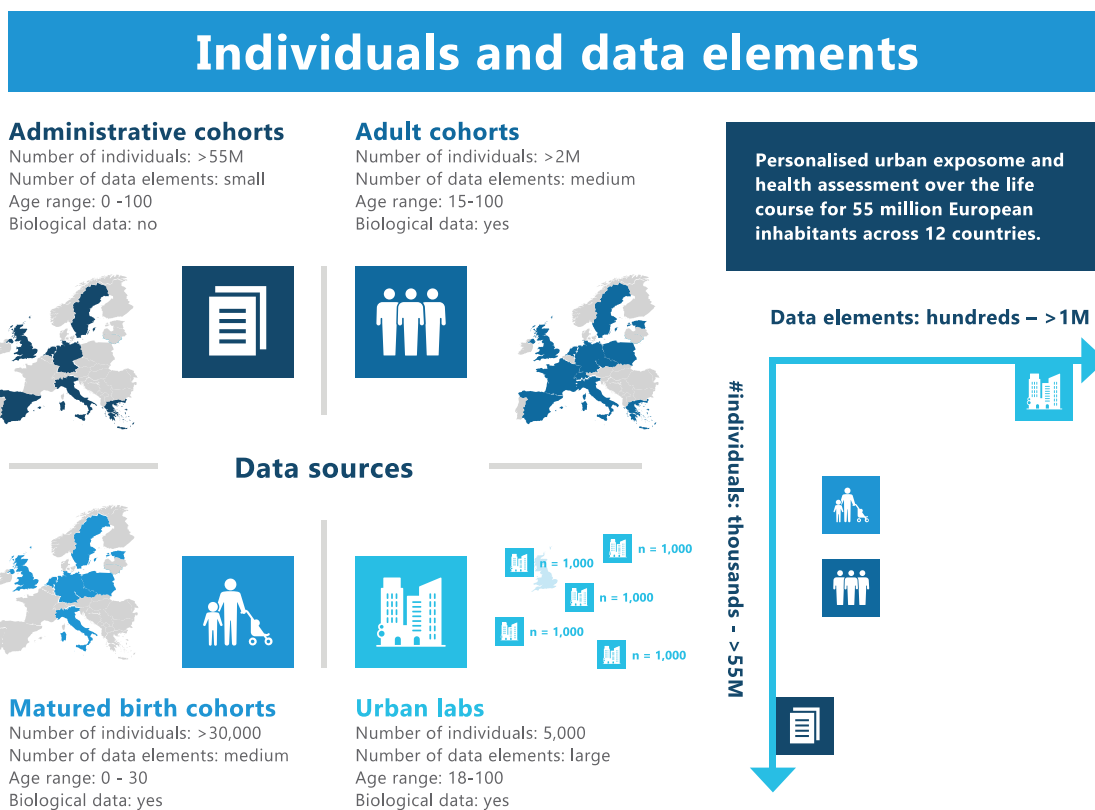


Figure 2. Geographical distribution and key characteristics of study types included in EXPANSE (left panel) and two-dimensional spectrum formed by the number of individuals and number of data elements (i.e., external and internal exposome factors) included in these study types (right panel).

resolution to capture temporal variation from the early 1990s, or covering only a fraction of the urban European population.²⁶ While increasing both the spatial and temporal resolution of the external exposome, we will aim to go beyond the State of the

Art in modeling a number of important stressors. For example, currently ultrafine particles (UFP) models exist for a number of cities/areas in Europe,^{27,28} but transferability of these models to other cities or areas in Europe is challenging.²⁹ To fill this gap,

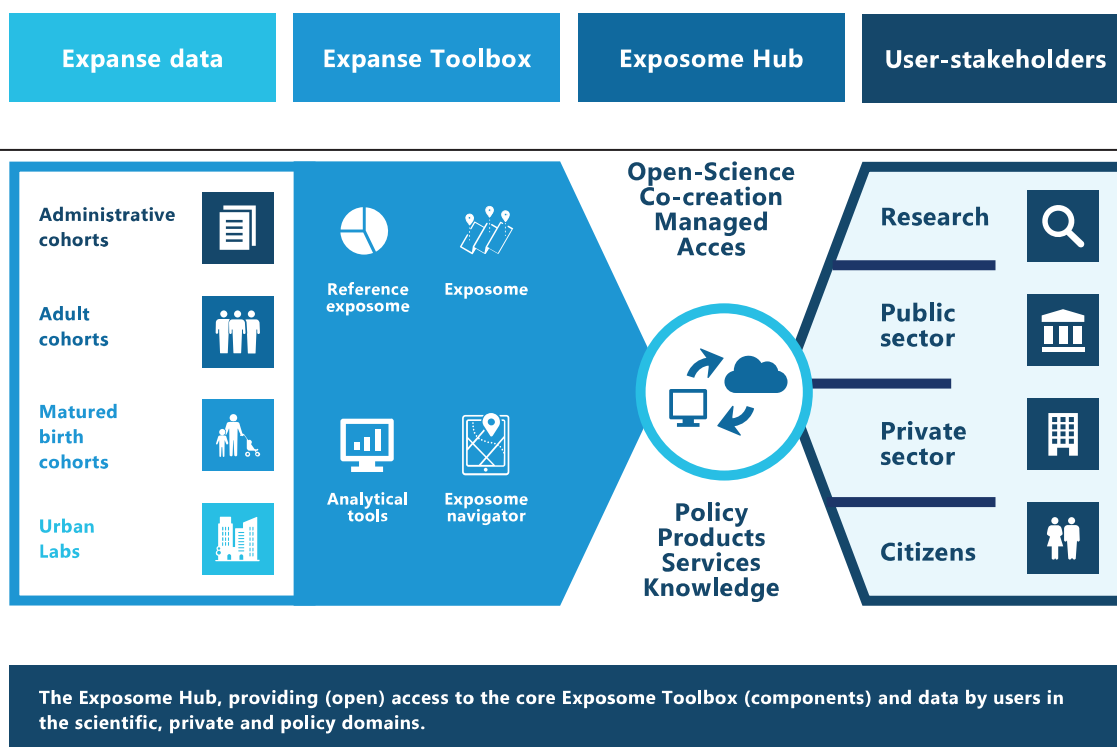


Figure 3. The exposome Hub, providing (open) access to the core exposome Toolbox (components) and data by users in the scientific, private and policy domains and citizens.

we will develop models to predict UFP concentrations across the EU combining measurement data (routine monitoring sites, past monitoring campaigns, and mobile monitoring data) with a number of predictor variables depicting UFP source activity. The development of UFP models will be based on approaches described by Apte,³⁰ Messier³¹ and van Nunen,²⁷ for the first time extending this to a European-wide scale. UFP was chosen as first priority for this exercise, due to the expertise available within the EXPANSE consortium, and the fact that such data is obtained by mobile sensors. Lessons learned from the use of data from multiple mobile sensors will be applied in the development of other exposome factors in a later stage of the project.

Individual exposures can be improved by combining classical field-based modeling (e.g., land-use regression and dispersion) with agent-based modeling, which incorporates behavior and time activity patterns. The latter addresses the geographical uncertainty problem as people are not static but move in space and time. To address the geographical uncertainty, we will incorporate data from large time activity surveys and population dynamics into our assessment of the external exposome.

An overview of the information that will be collected to assess the urban exposome is presented in Appendix I.

Measure the internal exposome

One of the biggest challenges of internal exposome research is its complexity and the plethora of environmental components, their metabolites, and their induced molecular signatures, which complicate the quest to associate features of the internal exposome with specific characteristics of the external exposome. In the last 10 years, a considerable progress in untargeted liquid- (LC-) and gas (GC-) chromatography with high-resolution mass spectrometry (HRMS) shows it is possible to routinely detect 1000s of metabolites and chemicals in a relatively small volume of blood (or other biological material). While traditional metabolomic approaches focus on using untargeted LC-HRMS to measure endogenous metabolites, systematic optimization of LC-HRMS and GC-HRMS platforms combined with innovative new data extraction approaches now enable detection of exposure-related chemicals present at very low concentrations. Application to human populations allows detection and characterization of a large range of exogenously derived small chemicals that include pharmaceuticals, pesticides, plasticizers, flame retardants, preservatives, dietary compounds, and microbial metabolites that provide a measure of chemical internal dose. Combined with endogenous metabolites also detected using these approaches, characterization of the internal exposome provides a central measure linking exposure to internal dose, biological response, and metabolic alterations consistent with disease pathobiology.

Integrating measurements of the internal exposome with exposures faces several major challenges. Despite the rapid development of metabolite databases such as the Human Metabolite Database (HMDB) over the last 10 years,³² a large fraction of chemical signals detected by LC-HRMS and GC-HRMS in human biospecimens do not match any of the known entries commonly present in chemical databases.³³ This is particularly true for exogenous compounds and pollutants, which have received less attention than endogenous metabolites. While lists of exogenous chemicals are available and can be used for annotation, many are specific to the form approved for commercial use and have limited information on transformation products arising from biotic and abiotic processes that in many cases represent the key exposure or detected biomarker in a biological sample. These can be further transformed by the host or gut microbiota into various metabolites, making their identification even more challenging.³³ Timing also impacts measurement of the human exposome. Many exogenous and endogenous compounds have short half-lives and show large intraindividual variability, and their use as biomarkers requires

optimal sampling strategies and appropriate sample collection timing relative to the occurrence of the exposure and health outcomes of interest.³⁴

The EXPANSE project will address these challenges and analyze the exposome in 10,000 blood samples from various adult and matured birth cohorts through Europe by both LC- and GC-HRMS. The resulting database will provide the largest characterization of the human internal exposome to-date, and will be a key and lasting resource for understanding the distribution and magnitude of environmental exposures and associated effects in the EU.

Annotation of the exposome will include in-house chemical libraries for over 1000 chemicals, including pollutants, contaminants or food compounds, and expert knowledge on the metabolism of xenobiotics, while databases (such as the recently developed CECscreen database³⁵) will be used to assign annotations of 1,000s of additional exposures. The evaluation of the value of these compounds as biomarkers of exposure will rely on literature data curated by IARC in the Exposome-Explorer database,³⁶ on analyses of repeated samples, and on measured correlations with the External exposome. This opens up a completely new avenue for population-based research, where, for the first time, comprehensive molecular snapshots of exposures and their downstream effects can be linked to health outcomes.

In addition, cohorts included in EXPANSE compile an extensive collection of other OMICS profiles describing the internal exposome including genome-wide methylomes, transcriptomes, targeted and untargeted metabolomics, proteomics including inflammatory markers, and the gut microbiome. Available OMICS data within EXPANSE ranges between 5000 and 45,000 individuals, depending on the OMICS marker (see Figure 4). We will use the Molgenis data platform³⁷ to create an inventory of this data, and align the inventory with existing relevant ontologies, such as developed by the ELIXIR consortium.³⁸ These existing OMICS data will be linked to the External exposome and the newly generated HRMS data to describe the biochemical effects of urban exposures and their links with CMPD.

Link the urban exposome to CMPD

Our goal is to define the role of the urban exposome in CMPD across the life course, to decrease the proportion of unexplained CMPD risk in urban populations, and to gain insights into mediating biological mechanisms and causality. To do so, we will develop data analytical tools to link the external and internal urban exposome to CMPD and related clinical endpoints such as lung function, BMI, and blood pressure. We will take a life-course perspective by coupling insights from matured birth cohorts with those of adult cohorts (e.g., accelerated longitudinal design¹⁶). We will explore how the urban exposome influences individual trajectories in the formation, activation, maintenance and reactivation of resources over the life course determining the functional and structural reserve of biological functions (see Figure 5), as has been observed for forced expiratory volume in one second and COPD.^{40,41}

Cohorts participating in EXPANSE

EXPANSE will bring together mega-sized administrative cohorts, comprising the health data of more than 55 million European citizens and large individual adult and matured birth cohorts (>2 million European citizens). EXPANSE will thereby feature an optimal balance between statistical power and deep exposure, biomarker, and health phenotyping (Figure 4). The inclusion of mega-sized administrative cohorts in which individual-based external exposome features will be made available enables the use of machine learning and big data analytics to identify exposome patterns with robust discriminative accuracy (i.e. robust area under the receiver-operating characteristic curve). These will be validated and

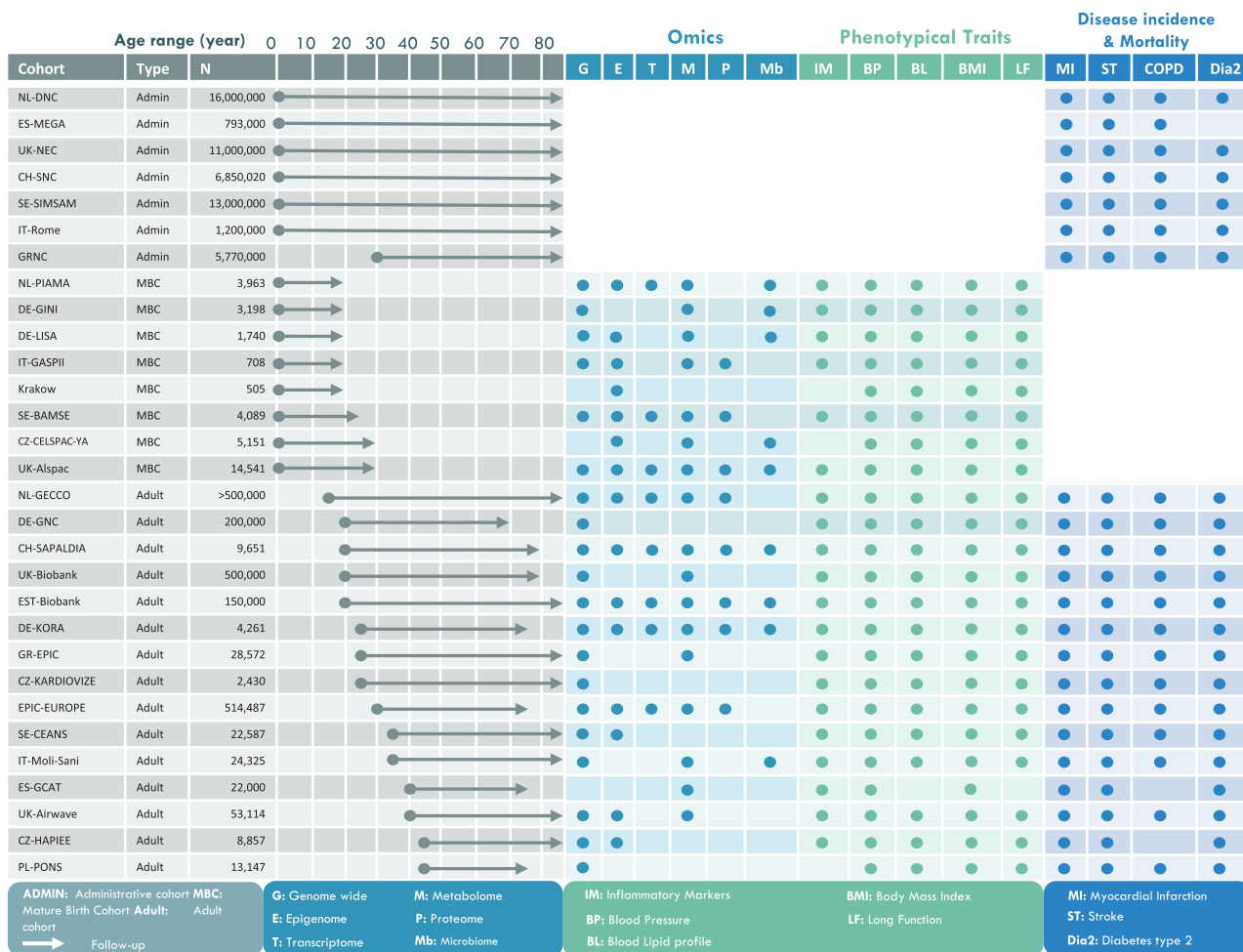


Figure 4. Overview of the three types of cohorts and their key characteristics, used in EXPANSE. NL-GECCO consists of several cohorts such as NTR, LifeLines, LifeWork, and EPIC-NL.³⁹ Blood samples are available in all adult and matured birth cohorts.

refined in matured birth and adult cohorts in which participants’ characteristics are available at a finer resolution.

A Pan-European panel study—Urban laboratories

The EXPANSE Urban laboratories are designed to allow deep phenotyping of the urban exposome as well as assessing its use and perception. It will involve 5000 individuals from five diverse urban areas spread across Europe (Amsterdam-Rotterdam area, Barcelona, Basel, Lodz, and Athens). The protocol applies novel and engaging data capturing features and includes tracking for spatial-temporal activity patterns, sensor-based environmental measurements, behavioral measurements, and minimally invasive biological sample collection for metabolomics analyses. Wellbeing and preclinical symptoms will be assessed as a potentially important mediator between the urban exposome and CMPD. The Urban Labs are expected to start in Fall 2021 (post COVID-19 or taking the impact of COVID-19 into consideration) and will run for ~2 years.

Epidemiological analyses

A unique feature of EXPANSE is the ability to cover the exposome and health outcomes over the full life-course by combining both matured birth cohorts and adult cohorts. As such we will be able to associate time-varying estimates of the external urban exposome with the development of preclinical CMPD phenotypes over time in adolescents and (young) adults (body weight, blood pressure, lipids, glycemia, and lung function), and

with the incidence of CMPD. In four case-control studies nested in the cohorts (for AMI, type 2 diabetes, stroke, and COPD) metabolomics and other OMICS markers that have been assessed in prospectively collected blood samples will be used to assess mediating or moderating biological pathways underlying the association of external urban exposome features and single or comorbid CMPD phenotypes (Figure 6). The same metabolomics measures will also be assessed in randomly selected samples in the matured birth cohorts, allowing for the assessment of the overlap between metabolomics patterns identified and combined trajectory analyses in adult and matured birth cohorts. Such a population-based trajectory will be compared to repeated samples available on a subset of the included individuals to study the intraindividual difference in trajectories as compared to the overall population trajectory.

In the case-control studies, underlying mechanisms involved in both disease inception and manifest clinical disease will be explored. We developed a detailed protocol for the selection of cohorts, individuals (cases, controls; N = 10,000 in total), and samples to participate in the metabolomics analyses. Prominent criteria were geographical spread, quality of the diagnosis, storage time, and availability of covariate and existing OMICS data.

EXPANSE will collect detailed information on relevant dimensions of time (e.g., the time between exposure(s) and the biological sample collection, age of the subject when the biological sample was collected, storage time of the biological sample at the time of analysis, the time between biological sample collection and the incidence date of the cases in the case-control studies, diurnal variation in OMICS markers and its relation to

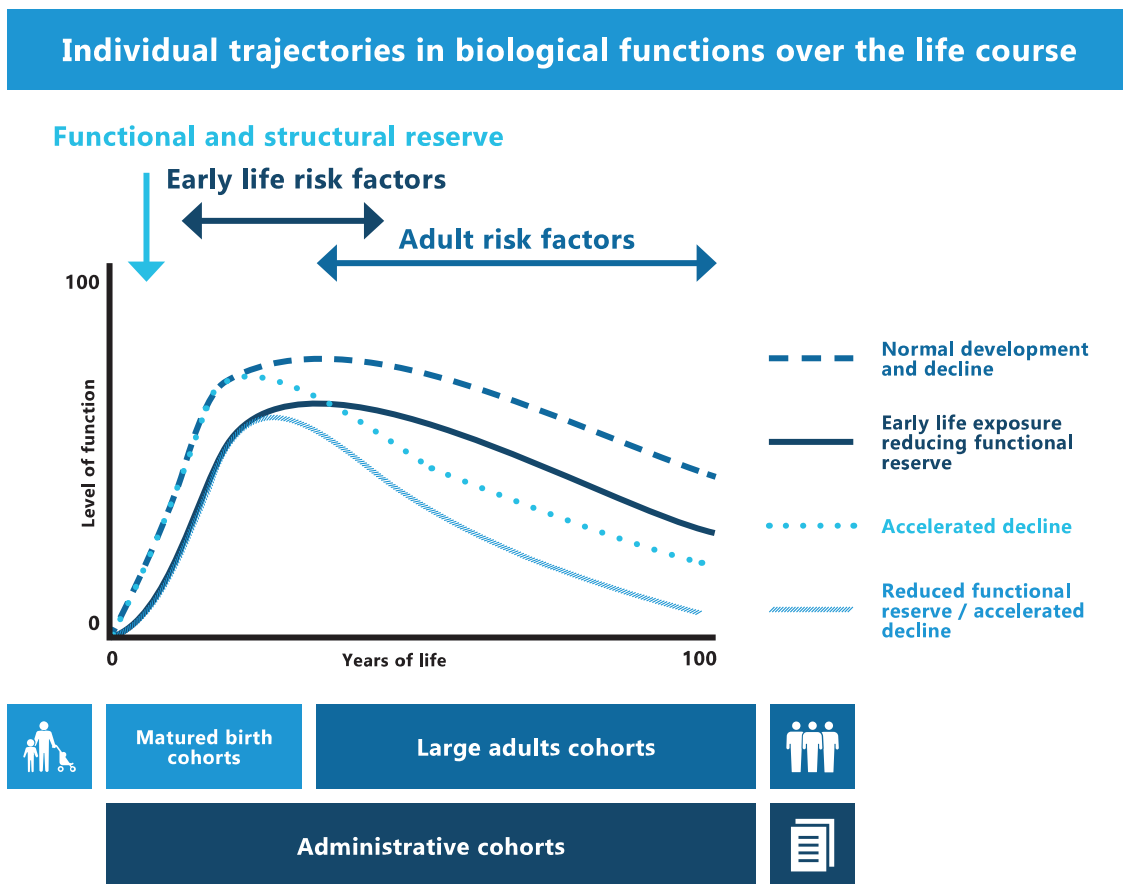


Figure 5. Top panel, Illustration of how individual trajectories in the formation, activation, maintenance and reactivation of resources over the life course determine the functional and structural reserve of biological functions (a disease example of such a biological function would be COPD). Bottom panel, Coverage of the life course by cohort types included in EXPANSE.

the time of day when the biological sample was collected, fasting state, and batch effects) with regards to the OMICS data that will be used in the epidemiological analyses. As EXPANSEs approach is based on existing collections of biological samples, there is limited freedom in harmonizing, and standardizing sample collection, yet collecting detailed information on all the relevant time dimensions will allow us to explore the impact of these dimensions on the OMICS markers and potentially adjust for these dimensions in our analyses when relevant.

Evaluate intervention strategies to prevent CMPD

Our goal is to develop and evaluate intervention strategies to prevent CMPD and promote health by modifying aspects of the urban exposome at the population and individual level.

Residential relocation to another neighborhood changes multiple dimensions of the urban exposome simultaneously, and likely leads to much larger changes in exposures compared to other urban interventions (e.g., local traffic management, urban redevelopment). EXPANSE will use residential relocation as a natural experiment affecting the external exposome and thereby the intermediate markers of CMPD risk. Using data on thousands of cohort participants who relocated, we will quantify the potential for residential relocation to shift the external exposome and how these changes in the external exposome are associated with time-varying intermediate CMPD risk factors (e.g., blood pressure) in longitudinal data.

EXPANSE will co-develop specific exposome-based interventions to prevent CMPD with diverse stakeholders (e.g., citizens, urban planners, and policy makers). EXPANSE will utilize detailed data from the Urban Labs on individuals' behaviors,

mobility and time-activity patterns, social network, and surrounding food and built environments to simulate the impact of these interventions using agent-based modeling, a powerful bottom-up approach to model individual behavior, accounting for nonlinearity and dynamic feedbacks within the model.

A sound economic evaluation of exposome-based prevention strategies does not only require proper identification and valuation of effects but also requires investments. These costs may fall upon a wide range of stakeholders, for example, individuals (healthier food choices, sports club membership, and wearables for direct feed-back on health behaviors), municipalities (investments to increase local green and blue spaces, bike lanes), the healthcare system (relocate facilities to neighborhoods with highest health need or easiest or most cost-effective health gains), industry (improved control of emissions), and government (legislation and enforcement, e.g., related to noise and pollutants). Stakeholders may have diverging interests, and for transparent decision-making, it is important to explicitly address which parties gain and which lose from voluntary or mandatory implementation of interventions. We will assess the welfare impact (net societal costs and benefits) of the urban exposome-based interventions for different stakeholders that either pay for these interventions or gain health benefits and other benefits, such as increases in productivity. This analysis aims to support decision makers at different levels (citizens, municipalities, government, and healthcare system) in their consideration which prevention strategies merit investments.

Statistical solutions for exposome challenges

The complex nature of the urban exposome (high number of variables that are collected, lack of knowledge on causal structures,

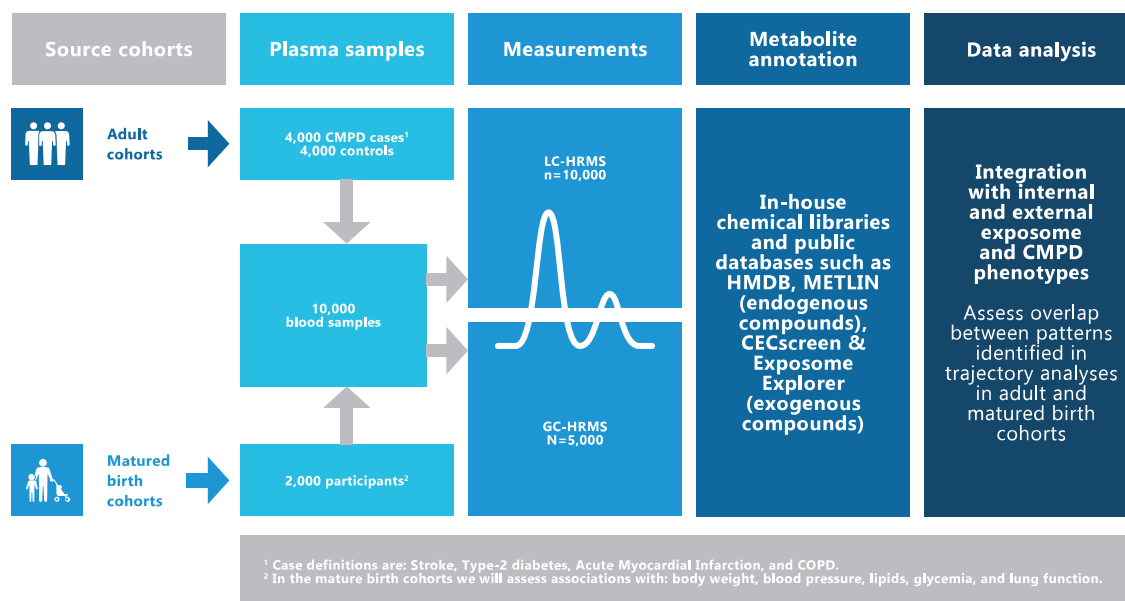


Figure 6. Schematic of the case-control and cross-sectional metabolomics studies that will be conducted in EXPANSE. 1. Case definitions are: stroke, type-2 diabetes, acute myocardial infarction, and chronic obstructive pulmonary disease. 2. In the matured birth cohorts, we will assess associations with: body weight, blood pressure, lipids, glycemia, and lung function. GC, gas chromatography; HRMS, high-resolution mass spectrometry; LC, liquid chromatography.

and complex measurement error structures) complicates the statistical analysis of its impact on CMPD. Previous exposome projects have highlighted the complexity of both the exposome data itself and of its molecular effects and downstream health consequences. This calls for the development of novel statistical approaches that can be applied in exposome research. EXPANSE will therefore move beyond existing profiling approaches and will adopt a complexity reduction approach that will (1) identify external exposome features jointly contributing to population stratification; (2) investigate if these clustering features are embodied and if molecular signatures of these features can be detected; and (3) assess if and to what extent both external and internal exposome features complementarily associate to CMPD outcomes.

Developing expotypes through cluster analyses

Borrowing from approaches developed in the population genetics sciences⁴² and leveraging the large population size in which external exposome features are available, we will apply clustering techniques to identify homogeneous groups of individuals sharing similar exposome experiences: “Expotypes.” These Expotypes will be subsequently characterized by exploring which features—(sets of) exposures—that primarily drive cluster membership. Clusters will be defined within studies, but also across studies and countries. To maximally use the longitudinal data that are available in the EXPANSE cohorts, we will expand cluster methods to incorporate repeated measures of the external exposome across the life course and will extend our clustering approach to trajectories (rather than cross-sectional data).

Investigating biological embodiment of expotypes

Expotypes could be viewed as the set of external exposome features maximizing the exposure contrast within the population. As such, the investigation of the molecular signature of the Expotypes will provide valuable information on the way these exposures are biologically translated. Further understanding of the complex correlation structures across the internal signatures of Expotypes can be provided by characterizing the topology of correlation networks combining external exposome features (in particular those contributing to population stratification) and (selected) internal exposome features.

Ethics-by-design

In exposome research, large amounts of data on the person and their environment are collected and combined. These data collections raise ethical questions that need to be identified and addressed to ensure ethically sound and responsible research and innovation. At the same time, there is a critical need to evaluate to which extent “society” will appreciate the dissemination of tools developed by projects such as EXPANSE. EXPANSE therefore will employ ethics parallel research, which is an approach for the early ethical guidance of novel (bio)medical research and innovation.¹⁷ We will identify and evaluate the ethical challenges raised by EXPANSE using an interdisciplinary approach that deploys empirical and normative methods as well as public and patient involvement (PPI) elements. The inclusion of a diverse group of citizens in the early phases of exposome research will enable us to understand and map how they anticipate exposome research will affect their lives and realities as well as their hopes and fears. In addition to the user stakeholder panel, we will set up citizen panels to empirically collect what citizens, researchers, and beyond perceive as the main ethical dimensions of this research. We will focus on what comprises, in their views, responsible research and data stewardship and what kind of ethical norms this would entail. This, in combination with conceptual clarification and normative analysis, will allow us to provide normative guidance for the design and goals of exposome research, the influence on our conceptions of health and illness, appropriate data governance as well as the interpretation of exposome data for health care allocation and public health policy. By conducting this research in parallel to the development of exposome research as an integral part of EXPANSE, ethics-by-design can be realized by anticipating and integrating ethical norms and societal values in exposome research.

Outreach and dissemination

EXPANSE will deliver three dissemination tools that together with the analytical methods developed within the project will be made open-access via the exposome Hub. The Reference exposome will be derived from urban dwellers whose external and internal exposome is quantified within cohort and panel studies. These region-specific reference exposomes can be used for future studies and for network analyses that allow the

understanding of differences in levels and connectivity of urban exposomes between regions, and between population demographics, including gender and age.

The Exposome Maps are Pan-European spatially resolved maps on multiple stressors that are intended to be used by policy makers and health studies. The main domains of the urban exposome (the food, physicochemical, social, and built environments) are represented in these maps. For each domain, multiple layers that reflect the geographical distribution of specific exposures or risk factors are included. By geolocating members of a cohort, for which CMPD status across the life course is known, and incorporating projected trajectories on the Exposome Map, the association between the urban exposome (or specific factors within the urban exposome) and CMPD risk can be estimated. The geographical distribution of the urban exposome factors and their association with CMPD risk (i.e., Exposome Risk Score⁵) can be used to calculate a weighted risk score. This urban exposome score can be used to identify hotspots of concern and can be used by policy makers to establish evidence-based public health interventions.

The Exposome Navigator is an envisioned application that allows individuals to quantify their own exposome score by integrating information on location, activities, lifestyle choices, and biological read-outs. Whereas the Exposome Maps are targeted to the population at large, the Exposome Navigator is targeted toward the individual. It will primarily serve to provide individuals exposome intervention options to improve their health. By adding individual level information to the application, individuals can also use the Exposome Navigator to refine, and possibly reduce their predicted personal risks of developing CMPD.

Main strengths of the project

EXPANSE will span the evidence space from mega-sized administrative cohorts, comprising the health data of more than 55 million European citizens, large individual adult and matured birth cohorts (>2 million European citizens), and Urban laboratories (5000 individuals). The large size of the cohorts, the use of different study designs, a life course approach from childhood to adulthood, and evidence across different biological layers allow for a pluralistic approach in terms of statistical analysis, and where possible replicating results in different populations allowing an improved causal interpretation of statistical associations. EXPANSE will expand on the number of factors currently measured as part of the urban exposome; make the urban exposome surfaces time-varying and more spatially refined; and incorporate time-activity patterns in estimating individualized urban exposome features. In addition to new methods to assess the external exposome, EXPANSE will generate new insights into the interplay between the external and internal exposome by investing in novel broad-scan HRMS. By analyzing 10,000 samples with both LC-HRMS and GC-HRMS, EXPANSE will capture both the exogenous (e.g., synthetic chemicals and dietary constituents) and endogenous (metabolites) parts of the internal exposome. By linking these profiles to the external exposome and existing OMICS data, we will make a significant leap toward an integrated view of the external–internal exposome continuum. The life course perspective in EXPANSE covers both the life course in exposome features and disease development. Furthermore, EXPANSE proactively considers the ethical and user perspectives of their research methods.

Main challenges of the project

The large number of cohorts, samples, and exposome information that will be gathered in EXPANSE introduces large challenges for the mapping of information and harmonization of variables. To facilitate these crucial steps EXPANSE uses a platform for the reliable, long-term storing and archiving of large amounts of research data (YODA).⁴³ YODA is a generic,

flexible, and scalable solution, which complies to EU's General Data Protection Regulation (GDPR) and Findable, Accessible, Interoperable and Reusable (FAIR) principles. EXPANSE implements several secure platforms to optimally analyze sensitive research data, consisting of a virtual research environment and federated analysis solutions.⁴⁴ Together with its partners in the European Human Exposome Network, EXPANSE will work toward cataloging and harmonization of the internal and external exposome data collected in the project. We will build our efforts upon an existing approach developed to catalog and harmonize heterogeneous types of data in epidemiological studies.^{38,45} Other challenges include organizational or administrative difficulties with the exchange of samples within the project, technical issues in the HRMS measurements, insufficient data (quality) to allow for methodological developments or to conduct analyses such as natural experiment analyses and agent based modeling.

Over the course of EXPANSE steps will be made toward a better understanding of the exposome on reducing the global burden of CMPD. While the contribution of EXPANSE to an overall reduction of the global burden of CMPD during the lifetime of the project will likely be small to moderate, methodological developments and the interaction with the European Human Exposome Network will facilitate the speed and level of detail with which the exposome can be studied and results can be translated into meaningful interventions.

Exchange with stakeholders and other exposome projects

EXPANSE is dedicated to an optimal uptake of its findings by user-stakeholders, consisting of public sector policy makers (e.g., municipalities, city planners, national ministries, the EU), researchers (e.g., universities and research institutes), the private sector (e.g., architects, urban planners, big-data analytics, sensor developers), Non-Governmental Organizations (NGOs, e.g., environmental advocacy groups), and citizens (including CMPD patients). By following an Exposome approach EXPANSE will contribute to a better understanding of society's understanding of tradeoffs between exposure and health (such as the balance between eating more fruit and receiving higher pesticide exposures) which will allow us to avoid the potentially conflicting messages that the 'one exposure at a time research' provides society. To promote the adoption and use of our results, a multisectoral user stakeholder panel in which these stakeholder groups are represented has been set up. When relevant, other stakeholder groups than those currently defined can be added to the stakeholder panel. This user stakeholder panel has an advisory role in EXPANSE and will be consulted throughout the project in the design and execution of our research. Engaging our targeted stakeholders throughout the project is expected to result in a sustainable uptake, and use of project outputs by these stakeholders after the project has ended and aligned project goals with stakeholder needs for new evidence and tools.

The EXPANSE infrastructure facilitates studying the impact of the urban exposome on a vast array of health outcomes, including cardiometabolic disease, pulmonary disease, and well-being. The EXPANSE Toolbox will be made accessible to stakeholder groups and organizations outside of EXPANSE via the Exposome Hub. The Exposome Hub is a web-enabled "open science, co-creation" portal where scientists, policymakers, and private sector users can upload, access, and share exposome data, methods, and tools while collaborating in virtual workspaces on joint research and development projects. The Hub will allow users to create their own workspaces, with specific (tiered) access rights and corresponding intellectual property and privacy protection measures.

Conclusion

EXPANSE has been designed to provide actionable data and insights to improve health and wellbeing in a modern urban

environment. With the explicit aim to generate policy relevant information, the project focuses on generating and collating research data at a scale that allows for meaningful inferences. The project will deliver innovations in the assessment of the urban exposome and the internal exposome, statistical, and bioinformatics approaches, strategies to evaluate the impact of potential interventions, and strategies to communicate results to stakeholders such as citizens and policy makers. These innovations will be made available in a set of tools that will be shared in the EXPANSE toolbox and the distributed toolbox of the European Human Exposome Network.⁴⁶

Conflict of interest statement

The authors declare that they have no conflicts of interest with regard to the content of this report.

References

- Forouzanfar MH, Alexander L, Anderson HR, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015;386:2287–2323.
- 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.
- Wild CP. The exposome: from concept to utility. *Int J Epidemiol*. 2012;41:24–32.
- Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. *Science*. 2010;330:460–461.
- Vermeulen R, Schymanski EL, Barabási AL, Miller GW. The exposome and health: where chemistry meets biology. *Science*. 2020;367:392–396.
- Vrijheid M, Slama R, Robinson O, et al. The human early-life exposome (HELIX): project rationale and design. *Environ Health Perspect*. 2014;122:535–544.
- Turner MC, Vineis P, Seleiro E, et al; EXPOsOMICS Consortium. EXPOsOMICS: final policy workshop and stakeholder consultation. *BMC Public Health*. 2018;18:260.
- Vineis P, Avendano-Pabon M, Barros H, et al. The biology of inequalities in health: the LIFEPATH project. *Longit Life Course Stud*. 2017;8:417–439.
- Cities in Europe - PBL Netherlands Environmental Assessment Agency. Available at: <https://themasites.pbl.nl/o/cities-in-europe/>. Accessed 21 September 2020.
- Tonne C, Adair L, Adlakha D, et al. Defining pathways to healthy sustainable urban development. *Environ Int*. 2021;146:106236.
- Organization WH. Environment and Health Risks: A Review of the Influence and Effects of Social Inequalities. 2010. Available at: https://digitalscholarship.tsu.edu/mlcejs_info/6/. Accessed 2 December 2020.
- Panczak R, Galobardes B, Voorpostel M, Spoerri A, Zwahlen M, Egger M; Swiss National Cohort and Swiss Household Panel. A Swiss neighbourhood index of socioeconomic position: development and association with mortality. *J Epidemiol Community Health*. 2012;66:1129–1136.
- Ribeiro AI, Fraga S, Kelly-Irving M, et al. Neighbourhood socioeconomic deprivation and allostatic load: a multi-cohort study. *Sci Rep*. 2019;9:8790.
- Backholer K, Beauchamp A, Ball K, et al. A framework for evaluating the impact of obesity prevention strategies on socioeconomic inequalities in weight. *Am J Public Health*. 2014;104:e43–e50.
- THE 17 GOALS | Sustainable Development. Available at: <https://sdgs.un.org/goals>. Accessed 2 December 2020.
- Galbraith S, Bowden J, Mander A. Accelerated longitudinal designs: an overview of modelling, power, costs and handling missing data. *Stat Methods Med Res*. 2017;26:374–398.
- Jongsma KR, Bredenoord AL. Ethics parallel research: an approach for (early) ethical guidance of biomedical innovation. *BMC Med Ethics*. 2020;21:81.
- Cosselman KE, Navas-Acien A, Kaufman JD. Environmental factors in cardiovascular disease. *Nat Rev Cardiol*. 2015;12:627–642.
- Schultz WM, Kelli HM, Lisko JC, et al. Socioeconomic status and cardiovascular outcomes: challenges and interventions. *Circulation*. 2018;137:2166–2178.
- Polak M, Szafraniec K, Kozela M, Wolfshaut-Wolak R, Bobak M, Pająk A. Socioeconomic status and pulmonary function, transition from childhood to adulthood: cross-sectional results from the polish part of the HAPIEE study. *BMJ Open*. 2019;9:e022638.
- Anupama N, Sindhu G, Raghu KG. Significance of mitochondria on cardiometabolic syndromes. *Fundam Clin Pharmacol*. 2018;32:346–356.
- Peters A, Nawrot TS, Baccarelli AA. Hallmarks of environmental insults. *Cell*. 2021;184:1455–1468.
- Kelli HM, Kassas I. Cardio metabolic syndrome: a global epidemic. *J Diabetes Metab*. 2016;6:513.
- Eeftens M, Beelen R, de Hoogh K, et al. Development of land use regression models for PM(2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. *Environ Sci Technol*. 2012;46:11195–11205.
- Vienneau D, de Hoogh K, Faeh D, Kaufmann M, Wunderli JM, Rösli M; SNC Study Group. More than clean air and tranquillity: residential green is independently associated with decreasing mortality. *Environ Int*. 2017;108:176–184.
- Chen J, de Hoogh K, Gulliver J, et al. Development of Europe-Wide models for particle elemental composition using supervised linear regression and random forest. *Environ Sci Technol*. 2020;54:15698–15709.
- van Nunen E, Vermeulen R, Tsai MY, et al. Land use regression models for ultrafine particles in six European areas. *Environ Sci Technol*. 2017;51:3336–3345.
- Kerckhoffs J, Hoek G, Gehring U, Vermeulen R. Modelling nationwide spatial variation of ultrafine particles based on mobile monitoring. *Environ Int*. 2021;154:106569.
- Yang Z, Freni-Sterrantino A, Fuller GW, Gulliver J. Development and transferability of ultrafine particle land use regression models in London. *Sci Total Environ*. 2020;740:140059.
- Apte JS, Messier KP, Gani S, et al. High-resolution air pollution mapping with Google Street view cars: exploiting big data. *Environ Sci Technol*. 2017;51:6999–7008.
- Messier KP, Chambliss SE, Gani S, et al. Mapping air pollution with Google Street view cars: efficient approaches with mobile monitoring and land use regression. *Environ Sci Technol*. 2018;52:12563–12572.
- Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*. 2018;46:D608–D617.
- Meijer J, Lamoree M, Hamers T, et al. An annotation database for chemicals of emerging concern in exposome research. *Environ Int*. 2021;152:106511.
- Pleil JD, Wallace MAG, Stiegel MA, Funk WE. Human biomarker interpretation: the importance of intra-class correlation coefficients (ICC) and their calculations based on mixed models, ANOVA, and variance estimates. *J Toxicol Environ Health B Crit Rev*. 2018;21:161–180.
- Meijer J, Lamoree M, Hamers T, et al. S71 | CECSCREEN | HBM4EU CECscreen: screening list for chemicals of emerging concern plus metadata and predicted phase 1 metabolites. 2020. doi: 10.5281/ZENODO.3957497
- Neveu V, Nicolas G, Salek RM, Wishart DS, Scalbert A. Exposome-Explorer 2.0: an update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Res*. 2020;48:D908–D912.
- Pang C, Sollie A, Sijtsma A, et al. SORTA: a system for ontology-based re-coding and technical annotation of biomedical phenotype data. *Database (Oxford)*. 2015;2015:bav089.
- ELIXIR Platform. Available at: <https://elixir-europe.org/platforms/interoperability>. Accessed 13 December 2020.
- Lakerveld J, Wagtendonk A, Vaartjes I, Karssen D; GECCO Consortium. Deep phenotyping meets big data: the Geoscience and Health Cohort Consortium (GECCO) data to enable exposome studies in The Netherlands. *Int J Health Geogr*. 2020;19:49.
- Lange P, Celli B, Agustí A, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med*. 2015;373:111–122.
- Cullati S, Kliegel M, Widmer E. Development of reserves over the life course and onset of vulnerability in later life. *Nat Hum Behav*. 2018;2:551–558.
- Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol*. 2003;18:189–197.
- Data Publication platform of Utrecht University. Available at: <https://public.yoda.uu.nl/>. Accessed 17 December 2020.
- Gaye A, Marcon Y, Isaeva J, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. *Int J Epidemiol*. 2014;43:1929–1944.
- Pang C, van Enckevort D, de Haan M, et al. MOLGENIS/connect: a system for semi-automatic integration of heterogeneous phenotype data with applications in biobanks. *Bioinformatics*. 2016;32:2176–2183.
- European Exposome Network – Human Exposome. Available at: <https://www.humanexposome.eu/>. Accessed 13 December 2020.

Appendix I:**Collected Variables Depicting the Urban Exposome**

Environment	Variable	Specification
Food and lifestyle	Food landscape	Fast food, supermarket, healthy food store density at different spatial resolution
Food and lifestyle	Food intake	Food intake by food diaries and questionnaires
Food and lifestyle	Contaminant intake	Linkage of food-intake information with contaminant databases
Social	Individual Socioeconomic Position (SEP)	Educational level, income, and occupational attainment
Social	Neighborhood SEP	Deprivation indices (income, unemployment rate education level, ethnicity, crime rate, and neighborhood characteristics)
Social	Psychosocial indicators	Family structure and composition, parenting style, family climate, and social support
Physicochemical	Noise	Average noise level throughout the day, evening, and night
Physicochemical	UV	Effective UV irradiance
Physicochemical	Pollen	10- to 15-year tri-monthly average
Physicochemical	Nearest distance to emission sources	Industrial site
Physicochemical		Line source
Physicochemical		Area source
Physicochemical		Landfill sites
Physicochemical		Gas/petrol station
Physicochemical		Air pollution
Physicochemical	Climate	Temperature and relative humidity
Physicochemical	Indoor sources	Data on cooking appliances, ventilation, smoking, etc.
Built	Greenspace	Surrounding greenness
Built		Access to green spaces
Built	Blue spaces	Access to blue spaces
Built	Gray spaces	Surrounding impervious land surface
Built	Walkability	Walkability index
Built	Light-at-night	Light intensity at night (LAN)
Built	Streetscape	Street level vegetation, water bodies, streetlights, topography, sidewalk, cycling lanes, and continuity of surrounding constructions