Environment International 208 (2026) 110117

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Review article

Computational standards and tools for exposome-wide association studies linking the human exposome with health outcomes

Dimosthenis Sarigiannis [a,b,c,d,*], Ourania Anesti [b,e], Nafsika Papaioannou [b,c], Achilleas Karakoltzidis [b,c], Spyros Karakitsios [a]

[a] National Hellenic Research Foundation, 48 Vasileos Constantinou Ave., 11635 Athens, Greece
[b] HERACLES Research Center on Health and the Exposome, Center for Interdisciplinary Research and Innovation, Aristotle University of Thessaloniki, Greece
[c] Environmental Engineering Laboratory (EnvE Lab), School of Chemical Engineering, Aristotle University of Thessaloniki, Greece
[d] Research Center for Complex Risk and Data Analysis, University Institute for Advanced Studies IUSS, 27100 Pavia, Italy
[e] Medical School, University of Crete, Heraklion, Greece

ARTICLE INFO

Keywords:
Exposome
ExWAS
Environmental health
Multi-omics integration
Computational tools
Data harmonization
Precision prevention
Genome–environment interaction

ABSTRACT

The human exposome (the cumulative measure of environmental exposures across the life course) offers a critical complement to genomics in deciphering the multifactorial origins of complex diseases. Exposome-wide association studies (ExWAS) represent an emerging class of high-dimensional epidemiological analyses designed to systematically assess associations between diverse environmental exposures and health outcomes. However, ExWAS requires advanced computational standards and tools capable of handling exposure complexity, temporal variability, co-exposure correlation, and multi-omics data integration. This review synthesizes current computational methodologies and platforms for ExWAS, highlighting recent advances in statistical modeling, exposure quantification, and bioinformatics tools. We conducted a PRISMA-ScR-guided scoping review across PubMed, Scopus, and Web of Science (2010–2025), with dual-reviewer screening in Rayyan, standardized data charting, and SWiM-aligned narrative synthesis. We explore multivariable and mixture modeling approaches (e.g., weighted quantile sum regression, Bayesian kernel machine regression), integration of external and internal exposome domains, and the application of longitudinal designs and environmental risk scoring. Key platforms such as the rexposome suite, exposomeShiny, and the integrative INTEGRA framework are examined for their role in operationalizing exposomic analyses at population scale. We also discuss the importance of data standardization, including exposure ontologies, harmonization protocols, and federated data infrastructure supporting cross-cohort analyses. Moreover, we discuss how computational exposomics can elucidate mechanistic pathways linking environmental exposures to disease, particularly when integrated with transcriptomic and metabolomic data. Finally, we outline future directions for the field, including genome–exposome integration, AI-driven causal inference, and translational pipelines for regulatory and clinical implementation. Beyond listing methods, we assess computational maturity and reproducibility (open licensing, containerization, federation readiness) and connect standards + tools to ExWAS workflows and translation. Computationally mature and mechanistically anchored, ExWAS are poised to become central tools in precision environmental health, enhancing the interpretability of genome–environment interactions and the predictive power of integrated omics frameworks.

1. Introduction

Exposome captures the totality of environmental exposures and related biological responses across the life course (Wild, 2005; Wild, 2012). As a complement to genomics, it motivates exposome-wide association studies (ExWAS), which systematically screen many exposures for association with health outcomes in human populations, analogous to GWAS but focused on environmental variables (Patel et al., 2010; Patel & Ioannidis, 2014).

Over the past decade, exposome research has expanded through major international initiatives and consortia (e.g., HELIX, EXPOsOMICS, HEALS, HERCULES, CHEAR/HHEAR) and newer European programs

* Corresponding author at: National Hellenic Research Foundation, 48 Vasileos Constantinou Ave., 11635 Athens, Greece.
E-mail address: sarigiannis@eie.gr (D. Sarigiannis).

https://doi.org/10.1016/j.envint.2026.110117
Received 21 June 2025; Received in revised form 7 January 2026; Accepted 29 January 2026
Available online 3 February 2026

(ATHLETE, LifeCycle, EPHOR, EHEN, NEUROSOME, URBANOME, ENVESOME). In 2025, the ExpoHealthNet (EHN) cluster was launched, pulling together efforts of 6 European projects focusing on emerging environmental pollutant and hazardous waste impact on human health. In parallel, the Network for Exposomics in the United States (NEXUS) has been established as a national center for exposome research coordination, together with the IHEN (International Health and Exposome Network) in the EU and the NEXUS Global Network in the USA that connect exposomics partners worldwide. As of May 2025, the Global Exposome Forum was launched aiming at creating a truly global exposome community effectively addressing the link between the human exposome, public health and precision prevention and medicine. These efforts have proven the feasibility of measuring hundreds–to–thousands of exposure variables in epidemiological cohorts and integrating them with high-dimensional molecular data. Early ExWAS analyses on existing datasets (e.g. using NHANES, the US National Health and Nutrition Examination Survey) successfully identified unexpected environmental correlates of disease – for example, blood cadmium and lead levels associated with cardiovascular risk, and certain persistent organic pollutants associated with type 2 diabetes – findings that might have been missed in hypothesis-driven (Patel et al., 2010; Tzoulaki et al., 2012). Such discoveries illustrate the promise of the exposome approach in uncovering novel environmental risk factors.

However, exploiting the full potential of ExWAS requires addressing significant computational and data challenges. By design, ExWAS involves high-dimensional data: large numbers of exposures (from chemical contaminants and dietary nutrients to social factors) are examined in relation to health outcomes, often with complex correlation structures and time-varying behavior. This complexity calls for advanced statistical methods, robust data processing pipelines, and standardized frameworks to ensure results are reliable and comparable across studies (Manrai et al., 2017). Equally important are data harmonization standards – common protocols for how exposure data are measured, curated, and shared – so that different studies can be integrated or compared, much as genomic consortia rely on standardized genotyping and data formats. Without such standards, the heterogeneity of exposure data (differences in units, collection methods, timing, and metadata) can impede *meta*-analyses and the pooling of evidence (Zhang et al., 2021).

In this review, we survey current computational practices in exposome-wide association studies (ExWAS) in human epidemiology, with an emphasis on linking exposome data to health outcomes. We synthesize three connected areas: (1) methodological advances in study design and statistical analysis, (2) software platforms and tools that support exposome data processing and ExWAS workflows, and (3) data harmonization and standardization efforts that improve integration, reproducibility, and cross-cohort analysis. We include illustrative case studies, identify key bottlenecks, and outline priorities to make exposome research more actionable and clinically translatable. Although prior reviews have addressed parts of exposomics, no methods-oriented synthesis has systematically connected computational standards (e.g., ontologies, common data models, FAIR practices, federated analytics) with the practical toolchain needed to run ExWAS end to end. Our contribution is to (i) map the computational "substrate" of ExWAS across methods, software, and standards; (ii) assess tool and platform maturity through a reproducibility-focused lens; and (iii) link these components to emerging needs in multi-omics integration, exposure modeling (including PBPK), and privacy-preserving, federated infrastructures for scalable cross-cohort studies.

2. Methods

2.1. Review planning, search, and synthesis approach

Aim and review questions. We undertook a scoping, methods-focused review to map the computational standards and tools that operationalize exposome-wide association studies (ExWAS), and to appraise their maturity, reproducibility, and adoption. Our pre-specified questions were:

RQ1: Which computational standards (ontologies, common data models, FAIR practices, and federated architectures) are used in human ExWAS?

RQ2: Which statistical and computational methods constitute current practice for mixtures, longitudinal exposures, and multi-omics integration?

RQ3: Which open-source tools/platforms implement these methods in practice?

RQ4: What evidence exists for performance, reproducibility, and adoption (e.g., code availability, containerization, multi-cohort use)?

RQ5: Where are the gaps that limit translation to precision public health?

Methodological framework. Given the heterogeneity of evidence (methods papers, software notes, standards documents, exemplars), we followed PRISMA-ScR for scoping reviews and used Synthesis without Meta-analysis (SWiM) guidance for transparent narrative/quantitative synthesis when *meta*-analysis is not applicable. Reporting of search,

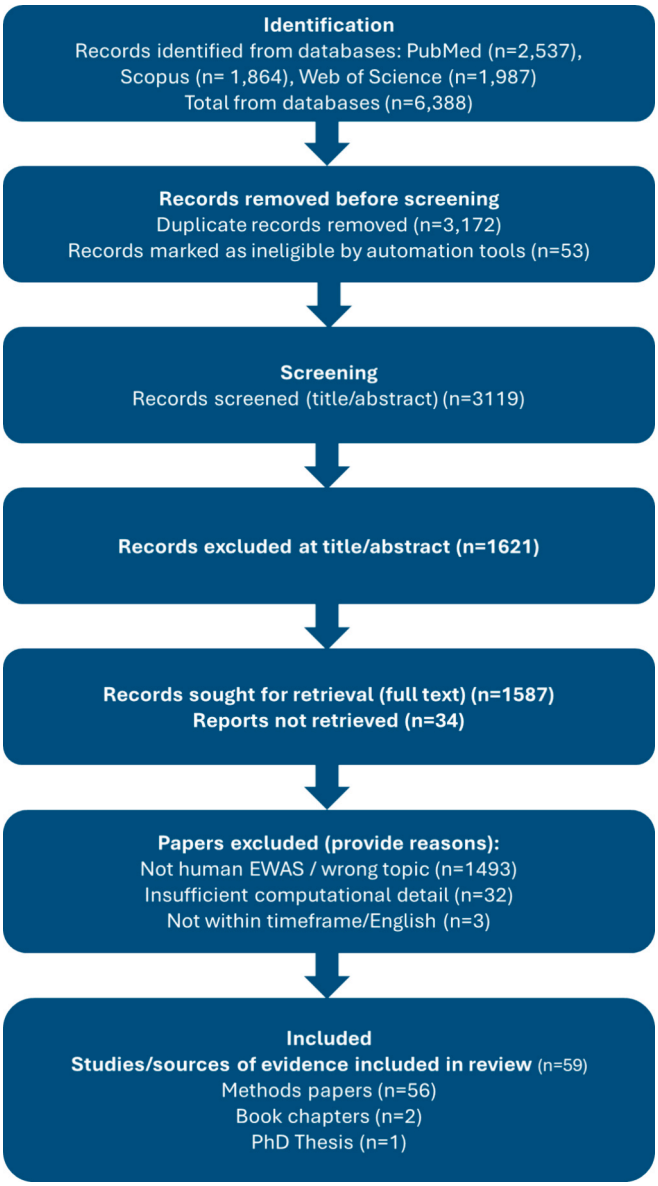

Fig. 1. Applied PRISMA methodology.

selection, and synthesis aligns with PRISMA 2020 good practice (Tricco et al., 2018), as outlined in Fig. 1. We also drew on the JBI scoping review methodology for protocol structure and data charting.

Eligibility criteria.

- **Population/Domain**: Human exposome/ExWAS or directly enabling computational methods (mixture models, life-course modeling, exposome–omics integration), data standards (ontologies, Common Data Models (CDMs), FAIR), and platforms for analysis or federation.
- **Evidence types**: Peer-reviewed methods papers, software articles, standards/guidance from recognized bodies, and major project deliverables where they define or implement reusable standards/tools.
- **Timeframe**: 2010–2025 (period of rapid maturation of computational exposomics).
- **Exclusions**: Single-chemical toxicology without computational generalizability to ExWAS; animal-only studies unless they define cross-cutting computational standards; purely conceptual commentaries without operational detail.

Information sources and search development.

We searched PubMed/MEDLINE, Web of Science Core Collection, and Scopus, complemented by backward/forward citation search from benchmark papers that define the computational landscape (e.g., rexposome/ExposomeSet, exposomeShiny, DataSHIELD, BKMR, WQS, HHEAR/Exposome-Explorer entries). This "seed-and-snowball" step secured conceptual coverage across standards, methods, and tools. Representative anchors include Hernandez-Ferrer et al., 2019 (rexposome), Vrijheid et al. 2021 (exposomeShiny), Marcon et al. 2021 (DataSHIELD), Bobb et al. 2018 (BKMR), Gennings et al. 2020 (WQS), Patel et al. 2016 (NHANES exposome/phenome), and Neveu et al. (2016) (Exposome-Explorer).

Search strings. We combined controlled vocabulary and free-text terms; an example PubMed string is:

(exposome OR "exposome-wide" OR EWAS OR ExWAS) AND

(method* OR algorithm* OR pipeline* OR software OR tool* OR platform OR "common data model" OR ontology OR federat* OR FAIR OR "data harmonization" OR "mixture model*" OR BKMR OR "weighted quantile sum" OR WQS OR "DataSHIELD" OR rexposome OR exposomeShiny OR "OMOP" OR FHIR)

Filters: 2010/01/01–present; English

Screening and selection. Citations were de-duplicated and screened in Rayyan by two reviewers working independently at title/abstract and full-text stages, with disagreements resolved by a third senior reviewer.

Data extraction. We piloted and then finalized a standardized extraction form capturing:

- bibliometrics; study type; cohort(s) if any;
- method class (pre-processing/QC, imputation, mixture modeling, longitudinal/life-course, mediation/causal, omics integration, PBPK/exposure modeling);
- tooling (language, repository/DOI, license, containerization, test data, documentation, update activity);
- standards (ontology/CDM used; FAIR elements; federation capability);
- validation (simulation benchmarks, external replication, cross-cohort reuse);
- adoption signal (citations, cross-project use).

Evaluation/quality considerations. For methods/software, conventional "risk-of-bias" tools are not fit-for-purpose; instead we assessed reproducibility and software quality using a pragmatic approach aligned with open science and software quality guidance (e.g., versioned releases, OSI-approved license, container image, CI tests, documentation, example workflows), and high-level software quality attributes from ISO/IEC 25,010 (e.g., reliability, maintainability, portability) adapted to research tools. We do *not* score studies of health effects; where such studies were used to evidence *tool performance*, we reported their own design/replication status narratively (ISO, 2011).

Synthesis and analysis. Following SWiM and narrative synthesis best practice, we: (i) tabulated evidence by thematic domains (methods, tools, standards/federation); (ii) mapped each item to the ExWAS workflow (intake → QC → modeling → omics integration → interpretation → governance), (iii) aggregated simple counts (e.g., number of tools supporting containerization or federated analysis), and (iv) summarized performance/validation evidence qualitatively (no vote-counting of effect significance). Where feasible we provide descriptive statistics (e.g., proportion of tools with open licenses, number of independent cohorts using a tool) and evidence-gap heatmaps.

2.2. Methodological developments in Exposome-Wide association studies

An ExWAS scans numerous exposures for associations with a health outcome, agnostic to specific hypotheses (Chung et al., 2024; Patel et al., 2010). Early ExWAS adopted GWAS methods, using exposure-by-exposure regression models with multiple testing corrections like false discovery rate control (Benjamini & Hochberg, 1995). This approach identifies candidates for further study but faces challenges from multiple testing (increasing false positives) and correlated exposures (e.g., traffic-related pollutants), which violate independence assumptions (Patel & Manrai, 2015). Traditional ExWAS uses strict significance thresholds or permutation tests and reports exposure correlation clusters to avoid overinterpreting single factors (Patel et al., 2016). Tools like the exposome correlation globe visualize exposure correlation (Chung et al., 2018; Patel & Manrai, 2015). Recent ExWAS methods emphasize multivariable and multivariate approaches to address these limitations.

Before specific methods are detailed, a distinction between the term "mixture models" and the term "exposome wide computational strategies" is made in this review. "Mixture models" are used to refer to statistical approaches in which the joint health effects of a relatively well defined subset of exposures are estimated, typically for a class of correlated environmental agents such as chemical pollutants, and in which interaction and nonlinear dose response are often accommodated (for example, WQS regression and BKMR) (Bobb et al., 2018; Gennings et al., 2020). By contrast, "exposome wide computational strategies" are used to denote workflows and methods by which the full multidimensional exposome is analyzed (Manrai et al., 2017; Miller & Jones, 2014; Rappaport & Smith, 2010). Within these strategies, heterogeneous external and internal exposures (chemical, physical, lifestyle, and multi-omics) are integrated and data structures, harmonization steps, and analytic pipelines are provided to scan and interpret this broader exposure space (Hernandez-Ferrer et al., 2019; Vrijheid et al., 2021). In this sense, mixture models are regarded as one important class of methods within the wider exposomics computational toolbox, but they are not regarded as synonymous with exposome wide analysis.

In exposome-wide analyses, the extremely high dimensionality of exposure datasets often leads to the well-known "curse of dimensionality," where the number of predictors greatly exceeds the number of observations (Altman & Krzywinski, 2018; Anguita-Ruiz et al., 2023). This phenomenon increases model instability, dilutes information, and can reduce statistical power. For this reason, many exposomics workflows incorporate an initial dimension-reduction step to extract latent structures or reduce redundancy before applying mixtures models or Exposome-Wide Association Studies frameworks (Gibson et al., 2019; Kalia et al., 2020; Liu et al., 2024; Reich et al., 2020). Most frequently used methods are described in the next paragraph. Overall, this dimension-reduction strategy helps stabilize estimation, control false positives linked to correlation and confounding, and improve interpretability in downstream analyses.

Principal component analysis (PCA) condenses correlated exposures into fewer composite variables (e.g., principal components) that reflect

exposure clusters, such as vehicle exhaust pollutants or lifestyle patterns, for association with health outcomes. However, PCA assumes linear, orthogonal components and can be sensitive to outliers and left-censored data, which may limit its ability to recover interpretable source patterns in complex exposure mixtures. Alternative unsupervised approaches more tailored to mixtures have therefore been proposed, such as principal component pursuit (PCP), which decomposes the exposure matrix into a low-rank structure capturing consistent patterns and a sparse matrix isolating extreme or unique exposure events, and has been adapted for environmental mixtures with missing and LOD-censored values (Gibson et al., 2022). Penalization and variable selection methods, like lasso or elastic net regression, identify key exposures predictive of outcomes while minimizing less relevant ones (Billionnet & Sherrill, 2012). Mixture-focused approaches, such as weighted quantile sum (WQS) regression and quantile g-computation, estimate joint effects of multi-component exposure mixtures (Gennings et al., 2020; Keil et al., 2020). Quantile g-computation generalizes WQS by relaxing directional homogeneity and estimating a marginal joint effect of simultaneous quantile increases in mixture components (Keil et al., 2020). BKMR flexibly models non-linear and non-additive mixture effects and can identify key exposures and interactions (Bobb et al., 2018). Recent pipelines and decision frameworks (e.g., Hao et al. (2024); NIEHS PRIME) provide practical guidance on method choice given the scientific question, mixture structure, and sample size.

Longitudinal and life-course exposure data integration is a key methodological advance in ExWAS. Unlike traditional cross-sectional ExWAS, life-course approaches account for time-varying exposures and susceptibilities (Smith et al., 2016). Methods like distributed lag models and trajectory analysis use repeated exposure measures (e.g., pollutants or dietary factors) to pinpoint critical exposure windows (e.g., prenatal vs. early childhood) impacting outcomes. Combining longitudinal exposure and outcome trajectories, as in birth cohort studies, enhances causal inference and timing insights. The exposome's temporal dimension highlights the relevance of exposure history for chronic diseases. Computational ExWAS frameworks now support repeated-measures or cumulative exposure indices, such as polyexposure scores (PXS), sometimes also termed environmental risk scores (ERS). Adapted from polygenic risk scores, PXS/ERS approaches weight and sum selected exposures to generate an individual-level risk index that can be evaluated in independent populations (He et al., 2021; Pries et al., 2022; Tarraf et al., 2017). PXS-type indices may enable risk stratification across populations if proven transportable.

Exposome research integrates omics data to capture internal biochemical changes from exposures, extending beyond external measurements. Modern ExWAS leverage high-throughput molecular profiling, including untargeted metabolomics, epigenomics (DNA methylation, histone modifications), transcriptomics, and proteomics. These data support multiple approaches: a two-stage workflow first identifies disease-associated external exposures via ExWAS, then examines downstream molecular changes; alternatively, parallel omics-wide association studies (e.g., epigenome-wide studies) link exposures and molecular changes to the same health outcome (Relton & Davey Smith, 2012). Integrative methods like mediation analysis test if molecular changes mediate exposure-outcome relationships, while network/pathway analyses combine exposure and omics data. Agnostic metabolome profiling reveals exposure-perturbed metabolic pathways, generating mechanistic hypotheses (Sarigiannis, 2017). Advanced methods, such as integrative pathway enrichment analysis (Paczkowska et al., 2020) and multi-omics factor analysis (Alemu et al., 2025), identify co-varying exposure-molecular clusters influencing disease risk. Linking external and internal exposome domains strengthens causal inference, as exposure-disease associations corroborated by related gene expression or metabolite changes increase confidence in causality (Furman et al., 2019). Sarigiannis et al. (2018) advocate integrating transcriptomic and metabolomic profiling to identify exposure-related molecular signatures and connect them to adverse outcome pathways.

Assessing exposures in ExWAS is complex due to their dynamic nature, unlike fixed genetic data. Exposome studies combine multiple methods: direct measurements (e.g., pollutant levels in blood/urine, wearable sensors for air quality or activity, GPS trackers), questionnaires (capturing lifestyle, occupation, residential history), and environmental modelling with geospatial data linkage. Spatial models, like satellite-based remote sensing or land-use regression, estimate exposures (e.g., PM2.5 or greenness) at fine geographic scales, assigned to individuals via addresses (Di et al., 2017). Integrating these with epidemiological data requires advanced computational tools, often using GIS and large datasets. Reconstructing long-term exposure histories, such as lifetime chemical exposure, involves combining historical emissions data, dispersion modelling, and residential timelines (Russ et al., 2021). Exposure modelling frameworks, like the MENTOR system, integrate environmental concentrations with human activity patterns to estimate personal exposure metrics (Georgopoulos & Lioy, 2006).

Physiologically based pharmacokinetic (PBPK) models enhance exposome studies by quantifying biologically effective doses of pollutants, moving beyond simple external exposure measurements (Sarigiannis et al., 2020). For example, in an acute dioxin emission incident, PBPK modeling calculated the area-under-the-curve (AUC) of blood dioxin concentration over time, providing a robust cumulative internal dose metric. These models allow classification of participants by internal dose or use as covariates in association analyses. Integrating PBPK models with epidemiology, termed computational exposure biology, connects emission sources to health outcomes via external exposure, internal dose, and biological effects (Sarigiannis et al., 2018). This approach demands significant computational infrastructure for complex simulations but improves exposure assessment accuracy in ExWAS.

In practical ExWAS applications, PBPK modelling is often integrated through a sequence of linked steps. Compound specific PBPK models are specified and calibrated on the basis of external exposure information such as monitoring data or reconstructed intake, together with physiological characteristics and substance specific parameters. Internal concentration profiles over time are then simulated for each participant. These simulated profiles are summarized into biologically meaningful internal dose metrics, for example peak concentration, area under the curve for a defined time window such as pregnancy or early childhood, or time weighted averages for periods of particular susceptibility. The internal dose metrics are merged with the rest of the exposome dataset and are treated as additional exposure variables in the epidemiological analysis. They are included in standard regression based ExWAS, in mixture models such as WQS or BKMR together with other exposures, or in the construction of Environmental Risk Scores. Through this integration the PBPK model is used as an intermediate layer between external measurements and biologically effective dose, so that the health associations estimated by ExWAS are interpreted as associations with internal dose indicators rather than only with external concentrations or short-lived biomarkers. In this sense, PBPK modelling can be seen as one component of a broader family of New Approach Methodologies (NAMs) that combine in vitro, in silico, and mechanistic information to derive human-relevant internal dose metrics and to interface naturally with exposome-scale datasets (Schmeisser et al., 2023; Wambaugh et al., 2019).

High-dimensional exposome data in ExWAS requires robust data cleaning and quality control (QC). Missing data, common in chemical assays or questionnaires, are often driven by left censoring below the analytical limit of detection (LOD) and are addressed using multiple imputations or survival analysis for censored data below detection limits such as e.g., replacing non-detects with $\mathrm{LOD}/\sqrt{2}$. However, simple substitution rules such as $\mathrm{LOD}/\sqrt{2}$ can bias effect estimates, particularly in mixture analyses, and recent work has shown that different imputation strategies for censored exposures can materially influence ExWAS mixture inferences and variable selection (Lee et al., 2024). More principled alternatives include parametric and semi-parametric multiple

imputation drawing from truncated normal or log-normal distributions conditional on covariates, quantile regression imputation of left-censored data (e.g., QRILC), likelihood-based censored regression models (such as Tobit or survival-type approaches), and fully Bayesian joint models that treat values below the LOD as latent variables and propagate this uncertainty into mixture effect estimates. Exposome software frameworks support more flexible imputation methods for censored and missing exposure data. Data normalization, such as log-transformations, handles skewed distributions, while outlier detection ensures data integrity. Combining disparate data sources (e.g., questionnaires, sensors, labs) necessitates consistent units, scaling, and encoding. Reproducible workflows, using scripted R/Python pipelines or containerization (e.g., Docker for exposomeShiny), enhance consistency, mirroring genomics trends. Standardized data objects like ExposomeSet (rexposome R package) integrate exposure and phenotype data. These methodological advances in statistical models, integrative analyses, exposure assessment, and data quality improve the sensitivity, specificity, and credibility of ExWAS linking exposomes to health outcomes.

2.3. Software platforms and computational tools for exposome data analysis

The complexity of exposome data has driven the development of a diverse ecosystem of software tools for ExWAS and related analyses (Table 1). R/Bioconductor packages such as the rexposome suite (including ExposomeSet and omicRexposome) and workflow-oriented tools like CLARITE and Elja support core tasks including data cleaning, harmonization, regression-based ExWAS, and omics integration. User-friendly interfaces such as exposomeShiny lower the barrier for non-programmers, while DataSHIELD enables federated ExWAS across cohorts without sharing individual-level data. Integrated platforms like

Table 1
Representative software tools, platforms, and resources for exposome-wide analyses.

Tool/platform/resource	Type/environment	Main role in EcWAS/exposomics	Notable features/notes	Literature
rexposome suite (R exposome package, ExposomeSet, rexposome, omicRexposome, CTDquerier)	R/Bioconductor packages	Core exposome data model and analysis toolkit for ExWAS and exposome–omics integration	ExposomeSet object for structured exposome data; functions for PCA, clustering, regression-based ExWAS; omicRexposome for omics integration; CTDquerier links to CTD for exposure–gene–disease pathways	Hernandez-Ferrer et al. (2019)
CLARITE/"CLAIRE"	R package(s)	Cleaning, harmonization, and ExWAS workflows for large epidemiologic datasets	Designed for datasets like NHANES; variable harmonization, missing-data handling, automated regression modeling across many exposures/phenotypes	Lucas et al. (2019); Passero et al. (2020)
Elja	R package	Automated environment-wide association studies	Loops over exposures, fits regression models with confounder adjustment, applies multiple-testing correction, and produces standardized summary reports	El Homsi and Annesi-Maesano (2024)
exposomeShiny	Shiny web application (R-based)	Point-and-click exposome data analysis interface	Supports preprocessing (imputation, normalization), descriptives, visualization, PCA/clustering, single-exposure and lasso-based ExWAS, and omics integration; CTD queries; open-source and Docker-ready for local deployments	Vrijheid et al. (2021)
DataSHIELD	Federated analysis framework (R server/client)	Privacy-preserving multi-cohort EcWAS and pooled analyses	Sends code to remote nodes and combines only non-disclosive summaries; widely used in consortia (e.g. LongITools) to enable cross-cohort exposome analyses without sharing individual-level data	Marcon et al. (2021)
Human Exposome Database	Data resource/database	Repository of NHANES-derived exposure–health associations	Curated exposure–outcome associations to support hypothesis generation, replication, and *meta*-analysis	Patel et al. (2016)
Exposome-Explorer	Data resource/database	Catalogue of biomarkers of exposure (dietary, chemical, etc.)	Standardized biomarker definitions, matrices, and sources; useful for study design, biomarker selection, and harmonization across cohorts	Neveu et al. (2016)
Comparative Toxicogenomics Database (CTD)	Data resource/knowledge base	Mechanistic interpretation of ExWAS findings	Curated chemical–gene–disease relationships; used via CTDquerier and exposomeShiny to link exposures to pathways and disease mechanisms	Davis et al. (2020)
INTEGRA	Integrated computational platform	End-to-end lifecourse exposure and risk assessment	Integrates external monitoring, dispersion modeling, PBPK simulation, biomarker data, and omics; used in complex scenarios such as the Aspropyrgos dioxin incident	Sarigiannis et al. (2014)
HEALS framework/platform	Project-level integrated exposome platform	Large-scale spatiotemporal exposome assessment with omics	Combines smart sensing, satellite and GIS data, big-data analytics, and multi-omics to link exposures with health outcomes; supports personalized exposure forecasting	Sarigiannis (2019)
scikit-learn, TensorFlow	General-purpose machine-learning libraries (Python)	Predictive modeling of health outcomes and exposure profiles	Implement algorithms such as random forests, gradient boosting, and deep learning; used to build prediction models from high-dimensional exposome and omics data	Atehortúa et al. (2023)
Gephi, Cytoscape, igraph	Network analysis and visualization tools	Exploration of exposure and omics networks	Visualize correlation/association networks between exposures, biomarkers, and outcomes; help identify clusters and key nodes in exposome structures	Patel et al. (2016) (for network-type exposome analyses)
Google Colab, RStudio Server, HPC clusters	Cloud/remote computing and HPC environments	Scalable execution of computationally intensive ExWAS workflows	Support parallelized modeling, resampling, and omics integration; enable shared analytic environments (e.g., browser-based R/Python sessions) for teams	
Nextflow	Workflow management system	Orchestration of complex exposome/omics pipelines	Enables modular, version-controlled pipelines that run on local, HPC, or cloud backends; improves reproducibility and portability of ExWAS workflows	
Snakemake	Workflow management system	Rule-based automation of analysis steps	Declarative workflow language with built-in support for provenance tracking, environments, and scalable execution; widely used for omics and can be adapted for exposome pipelines	
HERCULES Exposome Research Center (Emory University)	Methods/resource hub and training center	Centralized resources for exposome tools, training, and best practices	Curates exposomics tools and educational materials; supports capacity building and dissemination of exposome methods, including many tools listed above	

INTEGRA and the HEALS framework demonstrate how external exposure modeling, PBPK simulation, biomonitoring, and multi-omics can be combined in end-to-end pipelines. These specialized tools are complemented by general-purpose machine-learning libraries, network analysis software, high-performance computing environments, and workflow managers (e.g. Nextflow, Snakemake), which together support scalable and reproducible exposome analysis.

2.4. Data harmonization standards in exposome research

Data harmonization is essential in exposome research because datasets span heterogeneous sources, chemical measurements, sensors, GIS metrics, clinical data, surveys, biomarkers, and multi-omics, and must be comparable and reusable across studies. Harmonization typically relies on standardized metadata, controlled vocabularies/ontologies, and common data elements and models to support interoperable pipelines and pooled or federated analyses (Schmitt et al., 2023; Wilkinson et al., 2016). Ontologies address inconsistent exposure terminology (e.g., "PM2.5" vs. "fine airborne particles") by providing defined terms and relationships; key resources include the Exposure Ontology (ExO) (Mattingly et al., 2012), EPA's Environmental Ontology, and the NCI Thesaurus. Recent reviews note gaps in coverage of the external exposome, specially built environment and socio-economic factors, prompting calls to expand frameworks to capture neighborhood quality, climate, and social capital (Heacock et al., 2022; Zhang et al., 2021). In practice, the NIH PhenX Toolkit offers standardized exposure measurement protocols (CDEs) that improve cross-study consistency (Hamilton et al., 2011). Common data models (e.g., OMOP extensions for environmental exposures or cohort-specific models such as HELIX) enable integration of environmental and omics data (Vrijheid et al., 2014). Large harmonization efforts include ATHLETE, which retrospectively aligned variables like maternal smoking and urban greenness across birth cohorts (Tamayo-Uria et al., 2019), and Hu et al. (2020), which standardized over 5,500 external exposome variables linked to a birth registry by aligning spatiotemporal scales and metadata.

The growing volume of exposome data has increased interest in federated data ecosystems, as outlined in a 2023 NIEHS workshop roadmap (Schmitt et al., 2023). This model allows institutions to maintain their own databases (e.g., specialized chemical data) while adhering to common standards for virtual querying and integration. Key components include standardized data access protocols, semantic ontologies, and data governance addressing privacy and proprietary concerns. The NIH's HHEAR Data Center mandates standardized data formats with detailed metadata, and Exposome-Explorer provides templates for biomarker documentation. For legacy datasets, "crosswalks" map variables to standard formats, ensuring interoperability. Standardized ExWAS reporting, similar to MIAME for microarrays, is also emerging, requiring clear documentation of tested exposures, measurement methods, multiple testing corrections, confounders, and significant associations, with shared analytic code for reproducibility (Buckley et al., 2022). These practices are increasingly expected by environmental health journals.

Challenges and ongoing work: Challenges remain because new data streams (e.g., wearables) and cross-disciplinary standards evolve faster than uptake. Community coordination (e.g., EXPOSOME-IT; Human Exposome Forum) and standard-setting efforts (e.g., ISO discussions on reporting biomonitoring/exposure data) are important levers for accelerating adoption.

In sum, ontologies, common data models, and FAIR-aligned reporting are increasingly enabling multi-cohort ExWAS and more portable tool reuse. Continued incentives from funders and journals will accelerate adoption (Chung et al., 2024).

2.5. Positioning against existing reviews and method surveys

We systematically identified related reviews to establish novelty. In brief:

- Manrai et al., 2017 focus on informatics opportunities and challenges at a conceptual level; they do not map standards/CDMs or evaluate software maturity.
- Zhang et al., 2021 cover semantic standards for the external exposome but do not connect them to ExWAS analytics or the tooling ecosystem.
- Heacock et al., 2022 emphasize data integration and reuse; the piece is forward-looking but not a methods/toolchain map for ExWAS.
- Lai et al., 2024b treat high-resolution mass-spectrometry for the chemical exposome; analytics standards and ExWAS platforms are out of scope.
- Hernandez-Ferrer et al., 2019 present rexposome packages; this is a tool release paper rather than a landscape review.
- Marcon et al., 2021 address federated analysis (DataSHIELD) but not the broader ExWAS computational stack.

Our contribution is thus complementary: it unifies standards with methods and tooling and appraises reproducibility, openness, and adoption to support scalable cross-cohort ExWAS.

3. Results and discussion

3.1. Challenges in conducting exposome-wide association studies

Despite the exciting advances in methods, tools, and data infrastructure described above, exposome-wide association studies face numerous challenges that researchers must navigate. These challenges stem from both the inherent nature of exposome data and practical limitations in current research practices. Some of the most salient challenges are outlined below, along with insights from recent studies on how to address them.

ExWAS often involve testing hundreds of exposures, but many epidemiological studies have relatively modest sample sizes (hundreds to a few thousands of subjects). This imbalance can lead to low statistical power for detecting modest associations after correcting for multiple comparisons. The multiple testing burden in ExWAS is high – for example, testing 1000 exposures at $\alpha = 0.05$ significance level requires a p-value threshold on the order of 5E-05 (using Bonferroni) or controlling false discovery rate at say 5%. If the study has only a few hundred individuals, only very strong exposure–outcome associations can reach such stringent thresholds, potentially missing weaker (but still important) signals (Chung et al., 2024; Patel et al., 2010). Furthermore, the correlation among exposures means that the effective number of independent tests is somewhat lower, but it also means collinearity can inflate variance and make it difficult to attribute effects to a specific exposure. This challenge is being met by designing larger studies and consortia (to increase sample size) and by using dimension reduction or targeted testing as described earlier to mitigate the severe multiple testing penalty. Still, the reality is that many published ExWAS have suffered from underpower, and some null results could be due to insufficient sample rather than absence of effect. As exposome datasets grow (e.g., through pooling data or leveraging biobank-scale studies that now collect environmental data), this challenge will gradually ease, but careful power calculations and possibly Bayesian approaches (which can incorporate prior information to stabilize estimates) are recommended.

Unlike genotypes, which are static and measured with near-perfect accuracy nowadays, exposure measures are prone to measurement error and temporal variability. A single measurement of a chemical in blood or urine might not reflect the long-term exposure level of an individual due to short half-life or day-to-day variation; questionnaires about diet or stress are subject to recall bias; satellite-based estimates of pollution are proxies that have their own error margins. Measurement error typically biases associations towards the null (attenuation), making ExWAS findings conservative. Additionally, if error differs by

population subgroup or correlates with the outcome (differential error), it can even create spurious associations. Dealing with measurement error requires either improved exposure assessment methods (e.g., multiple measurements, use of biomarkers with longer integration time like hair or toenail samples for some metals) or statistical correction techniques (e.g., regression calibration if calibration data are available). Some researchers have incorporated external validation studies – for instance, in a subset of participants, use personal exposure monitors or gold-standard methods to quantify the error in a cheaper exposure metric, then adjust for that error in the full dataset. While this is still uncommon in ExWAS due to complexity, it is increasingly recognized. The HEALS project explicitly noted the importance of quantifying exposure uncertainty and propagating it through the analysis (Sarigiannis, 2019) and developed some computational tools to do so (e.g., probabilistic exposure estimates in addition to point estimates).

Environmental exposures are not randomly assigned; they cluster with social and behavioral factors. This raises the issue of confounding in ExWAS. For example, if lower socioeconomic status (SES) leads to both higher exposure to pollutants and worse health outcomes, an exposome study might find pollutant exposure associated with the outcome, but SES could be the real underlying cause. Traditional epidemiology addresses this by adjusting for confounders (such as SES indicators) in regression models. In exposome studies, since we measure many variables including social factors, some of these confounders are themselves part of the exposome. A key challenge is that adjusting for certain variables might obscure part of the exposome's effect (if, say, low income is considered part of the exposome construct rather than a nuisance variable). Recent thinking encourages treating factors like socioeconomic position, psychosocial stress, etc., not merely as confounders to "control away", but as integral parts of the exposome that interact with chemical exposures. For instance, effect modification by social factors is of great interest: does environmental exposure have a stronger effect on disadvantaged groups? Evidence suggests this for some outcomes (e.g., air pollution impacting those of lower SES more severely, possibly due to co-morbid stress or limited healthcare access). ExWAS must navigate controlling for confounders versus over-adjusting. Advanced causal inference methods (like directed acyclic graphs, DAGs) are used to decide what to adjust for. Nonetheless, residual confounding remains a concern – there could be unmeasured factors correlated with many exposures (e.g., a general "unhealthy lifestyle" phenotype) that drive findings. Researchers are increasingly using negative control outcomes or exposures to detect spurious associations (e.g., see if exposures correlate with outcomes they biologically should not, to gauge bias). The complexity of causal pathways (with many mediators and moderators in play) means that ExWAS results need cautious interpretation and, ideally, external validation.

One striking challenge observed in the literature is that different ExWAS on the same outcome sometimes identify different sets of significant exposures. For example, one study may find certain dietary factors and pollutants associated with obesity, while another similar study finds a somewhat different list. This between-study heterogeneity can arise from differences in populations (geography, genetics, culture), differences in exposure prevalence, or simply statistical randomness when many tests are near the significance threshold. It has led to some skepticism about ExWAS findings, echoing past concerns about reproducibility in GWAS and other high-dimensional searches (Ioannidis et al., 2009). To bolster confidence, replication studies can be used: if an ExWAS discovers a novel exposure-outcome link, a separate cohort should attempt the same analysis to see if it holds. Initiatives like the Exposome Collaboration (somewhat akin to the GIANT consortium in genetics) are starting to facilitate multi-cohort analyses that improve reproducibility. Moreover, *meta*-analytic techniques can combine results from several ExWAS, increasing power and examining consistency. Multi-center studies (like those combining data from different European regions) provide a methodological answer by replicating analyses in parallel and pooling results (Sarigiannis et al., 2021b). Reproducibility is also enhanced by pre-registering exposome study protocols, when possible, to reduce data-driven selection of results.

Handling exposome data can be computationally demanding. A dataset with, say, 1000 participants and 5000 exposure variables, plus genomic data plus metabolomic profiles, quickly reaches millions of data points, and biobank-scale studies can be orders of magnitude larger. This is particularly challenging for flexible mixture methods based on Gaussian processes: for example, BKMR relies on Markov chain Monte Carlo sampling and repeated inversion of large covariance matrices and can exhibit slow convergence and extremely long run times as sample size and dimensionality increase (Bobb et al., 2018). Several methodological developments have begun to address this scalability problem. Within the BKMR framework itself, reduced-rank or predictive-process approximations to the Gaussian process have been proposed to accelerate inference while retaining much of the model's flexibility (Bobb et al., 2018). More recently, divide-and-conquer strategies such as fast BKMR via median posterior aggregation (FBKMR) split very large cohorts into manageable subsets, fit Gaussian process mixture models in parallel, and then combine posterior summaries to recover full-sample estimates for ultra-massive datasets (Sonabend et al., 2024). Complementary approaches use kernel approximations based on supervised random Fourier features, recasting BKMR as a high-dimensional linear mixed model and achieving substantial reductions in computation time in large simulations and air-pollution applications (Zhang et al., 2025). Related kernel-machine extensions, such as generalized BKMR and collapsible kernel machine regression for exposomic analyses, also aim to improve flexibility, power, and interpretability as the number of exposures grows (McGee et al., 2025; Mou et al., 2025). Ensuring efficient storage (e.g., using database systems or sparse matrices) and designing workflows that can exploit parallel computing remains a genuine computational challenge even with these advances. The learning curve for using specialized tools can also be steep, raising the need for better training and more streamlined pipelines. Cloud-based solutions and collaborative platforms can alleviate local computational limitations but require comfort with those technologies. There is also the challenge of data security – exposome data often include personal identifiers (location, etc.), so when using cloud or sharing data, privacy must be maintained (hence the interest in federated analysis as mentioned earlier). In summary, while tools and scalable methods exist, not all researchers have equal access or expertise to use them optimally, which can be a barrier. The community is trying to address this via workshops (e.g., Exposome Boot Camps at Columbia University and elsewhere to train researchers in exposomic methods) and online tutorials.

After applying an ExWAS, interpretation of the discovered associations in a biologically meaningful way remains challenging. This obstacle has been extensively reported in exposomics methodology reviews (Patel, 2017; Ponzano et al., 2025; Price et al., 2022; Sarigiannis et al., 2025). As ExWAS analyses may include hundreds to thousands of correlated exposures, it is frequently necessary to combine information from multiple data sources to achieve mechanistic interpretation. Resources such as the Comparative Toxicogenomics Database (CTD), pathway enrichment tools, and high-resolution metabolomics can help reveal converging biological pathways, although these analyses are typically hypothesis-generating and prone to post-hoc narrative if not supported by independent evidence (Warth et al., 2017). To address these challenges, exposomics increasingly integrates statistical associations with experimental and in silico toxicology approaches (Lai et al., 2024a; Sarigiannis et al., 2025).

Laboratory assays, new approach methodologies, and QSAR/AOP-based predictions are employed to assess whether implicated chemicals disrupt relevant molecular events (Escher et al., 2022; Sheng et al., 2025). These approaches, encompassing in vitro, in silico, and alternative non-mammalian models, are increasingly used in chemical hazard evaluation (Bajard et al., 2023). Integrated workflows that combine transcriptomics, metabolomics, and pathway analysis enable the joint

identification of candidate adverse outcome pathways, which can subsequently be evaluated using targeted multi-omics and functional assays (González-Ruiz et al., 2019; Martins et al., 2019). Such multi-omics integration helps to capture changes across the biological cascade, moving beyond single-omics limitations to provide a systemic understanding of toxicity pathways (Canzler et al., 2020). This functional exposomics strategy enhances biological plausibility and advances the analysis from correlation toward causal inference by providing mechanistic insights and linking internal exposure to potential effects (Chung et al., 2021; Foreman et al., 2024; Orešič et al., 2020; Rappaport, 2018).

These approaches align with the rapid evolution of mechanistic toxicology, where NAMs—spanning high-throughput in vitro testing, alternative model systems, and computational models (QSAR, PBPK/IVIVE, AOP networks)—provide human-relevant mechanistic evidence while reducing reliance on traditional animal studies (Lovegrove et al., 2024; Ponzano et al., 2025). Analogous to Mendelian randomization in genomics, exposome research is also exploring causal inference tools leveraging genetic proxies, longitudinal trajectories, and repeated measures, although suitable instruments remain scarce for many exposures.

As mentioned earlier, the development of standards is ongoing. A practical challenge is ensuring adoption of these standards by the research community. It often lags behind their development. Many current studies still use idiosyncratic methods and do not share data in a FAIR way. Harmonizing efforts and convincing researchers to put in the extra work for standardization (which might not yield immediate publications) is a socio-technical challenge. Funding agencies and journals are starting to require data sharing and use of common protocols, which will push things in the right direction. The NIEHS exposome workshop emphasized forming an international consortium to drive these practices, akin to what the Human Genome Project did for genomics. Overcoming this challenge will likely require top-down encouragement (policy) as well as bottom-up demonstration of the benefits (showing that those who adopt standards can collaborate more widely and get more citations, etc.).

In summary, challenges in ExWAS are multifaceted – statistical, technical, and conceptual. Recognizing these challenges is the first step to overcoming them. The field is actively tackling them through innovations in study design (e.g., bigger and more diverse cohorts), advanced analytical methods (to tackle confounding and mixtures), better computational tools (for scaling and managing data), and community-wide initiatives for harmonization and collaboration. It is an evolving process, but each challenge met brings exposomics closer to fulfilling its promise of illuminating how the environment shapes health.

3.2. Key-points in future exposome research

As exposome research moves forward, future ExWAS development will be shaped by advances in environmental sensing, data science, systems biology, and implementation in public health (Miller, 2021; Vermeulen et al., 2020). One of the most promising future directions is the tighter integration of exposome data with genomic and other omic data to enable gene-environment interaction studies and improved risk prediction (Makris et al., 2025; Sarigiannis et al., 2025). To date, most ExWAS have analyzed environment and health in parallel to genetic factors, often adjusting for genetic principal components or including polygenic risk scores as covariates. However, this "parallel" treatment of genetics and environment risks ad hoc covariate selection and bias amplification if heritable covariates or colliders are inappropriately adjusted for (Weisskopf et al., 2018). Recent mixtures roadmaps therefore recommend beginning with an explicit conceptual model, often formalized as a directed acyclic graph (DAG), to distinguish exposures, outcomes, mediators, confounders and effect modifiers, and to derive minimal sufficient adjustment sets and pre-specified interactions before fitting high-dimensional models (Joubert et al., 2024). Within such a framework, age, sex, ancestry, and socioeconomic position may be treated as effect modifiers of interest, whereas other variables such as e.g. downstream biomarkers are deliberately not adjusted for to avoid blocking causal pathways.

Against this backdrop, the real power of exposome-genome integration may come from explicitly modeling interactions between specific exposures and genetic variants, rather than treating genetic information only as a nuisance covariate. Future large-scale studies (and biobanks) that contain both genomic and exposomics information could perform genome-wide environment interaction scans (GEWIS) to identify genetic polymorphisms that modulate the effect of exposures (and vice versa). For example, such studies might find that a certain antioxidant gene variant protects against air pollution effects on lung function, or that a metabolic gene variant makes some individuals more susceptible to pesticide exposure. This will require even larger sample sizes to have power for interaction tests, but with the advent of national biobank projects incorporating environmental data (like the UK Biobank collecting air pollution and neighborhood data, or the All of Us program in the US incorporating geospatial data), this becomes feasible. On the analysis front, new methods are being developed for exposome–genome integration, such as exposome-wide polygenic risk scoring (combining environmental risk scores with polygenic risk scores) and network-based methods that overlay gene and exposure networks (Chung et al., 2024; Colwell et al., 2023; Price et al., 2022). Bridging the gap between the genome and the exposome can enhance our understanding of disease etiology beyond what either alone can offer (Manrai et al., 2017; Sarigiannis et al., 2025). We anticipate that future "Exposome-Genome-Wide" studies will become more common, especially as computational platforms are now being proposed to handle the dual big-data streams (Zhou & Lee, 2021). The ultimate goal is a precision-medicine predictive model that integrates both genetic and environmental risk factors, consistent with the emerging exposome–genome paradigm in which the exposome complements the genome in explaining disease risk (Barouki et al., 2018; Sarigiannis et al., 2025). We refer to this integrated perspective as an "Exposome-Enabled Genome" approach. In cancer epidemiology, for example, researchers are beginning to incorporate detailed exposure histories into risk models that traditionally only used genetics and family history, hopefully improving individual risk stratification for screening through a combined risk score that integrates exposome-based risk with polygenic risk scores.

On the exposure measurement side, wearable and mobile sensors enabling more individual-specific, high-resolution exposure and physiologic data (e.g., personal air quality, UV, noise, heart-rate variability, glucose) (Helbig et al., 2021; Novak et al., 2023). Smartphone-based sensing (e.g., GPS plus sensor attachments) already supports minute-level "personal exposome" profiles (Turner et al., 2017). Integrating these high-frequency time-series into ExWAS will require scalable pipelines and time-series methods, alongside standardized device outputs and data formats to enable pooling across platforms and linkage to clinical and multi-omics endpoints.

While association studies are valuable, the field of exposomics recognizes a need to strengthen causal inference. Future work will employ methods such as Mendelian randomization (MR), where genetic variants (that relate to exposures) serve as instrumental variables to infer causality (Grau-Perez et al., 2019; Li et al., 2024). For example, variants affecting the metabolism of vitamin D have been used as IVs to test causal effects of vitamin D levels (an exposure) on health outcomes. Similarly, metabolism of toxicants like cadmium or lead may have genetic determinants that can be used in MR. Another approach is the use of negative exposure controls – variables that are known not to causally affect the outcome can test if an ExWAS analysis is picking up spurious associations due to confounding or bias. Additionally, the rise of causal discovery algorithms (from machine learning, like Bayesian networks) may help to hypothesize causal relationships from complex data (Huang et al., 2021). As datasets become richer (with multiple time points and multi-omics), algorithms can be deployed to suggest likely causal chains (e.g., exposure → epigenetic change → gene expression change →

metabolome change → disease) which can then be validated experimentally. Interventional studies, though not usually feasible for unethical exposures (you cannot randomize people to pollution), can be done for some exposome factors – e.g., dietary interventions or stress reduction programs – to see if modifying part of the exposome leads to health changes, thereby confirming causality.

A future direction aligned with toxicology is functional exposomics, which integrates the exposome concept with the Adverse Outcome Pathway (AOP) framework by linking measured exposures and bioactivity readouts to AOP-defined key events and networks (Ankley & Edwards, 2018; Wittwehr et al., 2016). AOPs are structured representations of biological pathways from a molecular initiating event (like a chemical binding a receptor) through intermediate key events to an adverse health outcome. In practical terms, integration occurs along two complementary axes. On the toxicology side, mixtures that reflect the human exposome such as e.g., extracts of water, dust, serum, or synthetic mixtures assembled from biomonitoring data are tested in batteries of mechanistic in vitro assays or −omics platforms. The resulting "bioactivity fingerprints" are then mapped onto AOP key events and AOP networks to identify which pathways and adverse outcomes are likely to be engaged and to rank risk-driving mixture components. On the population side, exposome studies in human cohorts generate panels of molecular biomarkers such as e.g., transcriptomic or metabolomic signatures) that can be annotated as key events within specific AOPs, providing evidence that those pathways are perturbed under real-world exposure conditions. For example, the metabolomic changes observed in the Aspropyrgos case study (altered lipid profiles due to AhR activation) correspond to key events in the dioxin AOP for cancer (Sarigiannis & Karakitsios, 2018), effectively using an AOP as a scaffold for interpreting human exposome data. Future ExWAS might routinely include panels of biomarkers for key events (transcriptomic or metabolomic signatures) to monitor activation of known toxicity pathways in people. The research community is already discussing creating exposome-AOP knowledgebases where human data can be overlaid on mechanistic pathway maps (Smith et al., 2011). The convergence of human data and experimental toxicology in a computational framework could enhance risk assessment: instead of relying on animal data alone, we might use human exposome data to directly inform which pathways are actually perturbed in exposed populations. Sarigiannis et al. (2021b) have argued that the exposome paradigm can be a new foundation for risk assessment, moving it toward a more evidence-based, mechanistic footing. This suggests that in the future, regulatory science will incorporate exposomics evidence. For instance, if an industrial site is under evaluation, not only environmental measurements but also exposomics biomarker data from nearby residents mapped to relevant AOP networks might be considered to truly gauge impact.

Establishing a global exposome data infrastructure is a key future direction (Fayet et al., 2024). We anticipate the formation of a Global Exposome Forum, where multiple countries and agencies collaborate to share data and resources. This could manifest as a federated network of exposome data hubs on different continents, all adhering to certain standards so that a researcher could query "exposures associated with asthma" and gather results from studies worldwide. The technical underpinnings (federated databases, common APIs for data access, privacy-preserving analytics) are being actively worked on. If successful, this will vastly increase the scale of exposome research – effectively a "Human Exposome Project" analogous to the Human Genome Project. In fact, China has launched a national exposome initiative, and the European Union has its Human Exposome Network (EHEN) and the more recent Exposome to Health Network (EXPOHEALTHNET), indicating that large-scale coordination is starting. In the next 5–10 years, we may see exposome reference datasets established – e.g., reference exposomes for different regions or populations, which could serve as baselines or comparators for local studies. We might also see the development of exposome chips or other multiplex assays that can cost-effectively measure a broad suite of biomarkers (much as genotyping arrays do for SNPs) – efforts in the exposome field like the "chemical exposome" via non-targeted mass spectrometry are aiming for that, but standardization and coverage need improvement (Lai et al., 2024b).

Future directions are not only technical; they also involve how exposome research engages with society. As personal exposure monitoring grows, issues of privacy will loom large. GPS traces and detailed logs of one's environment could inadvertently reveal sensitive personal information. Ensuring data is used ethically and with individuals' consent will be paramount. At the same time, the concept of citizen science in exposomics is emerging, where individuals actively participate in measuring and understanding their own exposome (e.g., community members carrying sensors or using smartphone apps to track their exposures and sharing data). This democratization of data could raise public awareness of environmental health issues and empower communities to lobby for changes (e.g., if a community crowd-sources their exposure data and finds high pollution levels, they can use that evidence). Researchers need to provide tools and platforms that are accessible to non-scientists for this to scale up. One can imagine a future where people have personal "exposome reports" analogous to personal genetic testing, highlighting key exposures and suggestions for mitigation (perhaps managed through healthcare systems). Of course, interpreting and acting on such information safely will be a challenge – this is where guidelines and likely some regulation will be needed to avoid misinterpretation or anxiety over one's exposome profile.

Finally, a crucial future direction is strengthening the link from exposome research to public health action and policy. To date, many exposome studies are proof-of-concept or discovery. The next phase will require converting findings into interventions. For example, if an exposome study robustly shows that a combination of housing conditions and air pollution drives asthma, interventions might combine environmental cleanup with housing improvements. Policymakers could use exposome data to prioritize which environmental factors, among many, will yield the biggest health benefit if addressed. This will likely involve modeling studies that simulate various intervention scenarios on a population's exposome and estimate health outcomes (some initial work using such simulations exists). Exposome data could also refine risk assessment practices: rather than assessing one chemical at a time, regulatory agencies may begin to assess cumulative risk using exposomics evidence about co-exposures (Sarigiannis & Hansen, 2012). In the EU, concepts like the "mixture assessment factor" are being informed by human data that show certain mixtures are common and have combined effects. Over the long term, success for exposomics will be measured in part by whether it leads to concrete improvements: cleaner environments, targeted preventive measures for high-risk groups, or personalized recommendations to reduce individual risks. Given that environmental factors are modifiable (in contrast to genes), the ultimate promise of exposome research is preventive medicine. We foresee "precision prevention" strategies where an individual's exposomic and genomic profile guides advice such as diet changes, supplementation, or relocation from high-risk areas, much as precision medicine might guide therapy based on genetics. Achieving this will require robust evidence and careful evaluation through interventions or trials, so this is likely a longer-term horizon, but it is a key aspirational direction.

In summary, the future of exposome-wide association studies is poised to be even more data-rich, integrative, and impactful. By incorporating cutting-edge technology (sensors, AI analytics), expanding scale (global data federation), and deepening integration with biology and medicine (genomics, AOPs, precision health), ExWAS will evolve from finding correlations to establishing causes and solutions. The trajectory suggests that in a decade, we may talk about exposomics in the same breath as genomics as a cornerstone of understanding health – enabled by the computational standards and tools that are being developed today.

3.3. Hands-on case studies from the HEALS project

Within the HEALS project, several European birth cohorts were leveraged to optimize ExWAS workflows, each contributing complementary paradigms before their application in the EXHES case study. Across all HEALS cohorts, untargeted LC–MS and NMR metabolomics were applied to maternal samples and processed using harmonized R-based pipelines for data preprocessing, annotation, and pathway enrichment, followed by the ExWAS paradigm to reveal the links between exposures and perturbed molecular pathways, and subsequently connecting those pathways to adverse health outcomes.

In particular, the PHIME cohort (Slovenia, n = 179) evaluated prenatal methylmercury and mixed metal exposures, identifying associations with early-childhood neurodevelopment, birth size, and language outcomes, supported by mechanistic indications of disrupted synaptic and mitochondrial pathways (Anesti et al., 2022). The REPRO_PL cohort (Poland, n = 146) examined combined prenatal exposure to phthalates and metals and found consistent associations with reduced fetal growth and early neurodevelopment, including effects of DEHP, DnBP, and DINP on cognitive, language and motor outcomes during infancy (Sarigiannis et al., 2021a). Building on these earlier paradigms, the Exposure and Health Examination Survey mother–child(ren) cohort (EXHES) (n = 2352 from 9 EU countries) was established as a proof-of-concept for a fully integrated exposome–health analysis pipeline under HEALS (Errahmani et al., 2025). The EXHES Tarragona (Spain, n = 135), for example, linked prenatal exposure to metals, phthalates/DINCH, and pesticides with adverse birth outcomes. ExWAS analyses showed associations between organophosphate and organochlorine biomarkers and reduced fetal growth, as well as links between multiple phthalates/DINCH and decreased birth height (Papaioannou, 2023).

The initial HEALS case studies established a pioneering integrated exposome–health analytical pipeline in Europe, meticulously combining external exposure modeling, physiologically based biokinetic-derived internal dosimetry, high-resolution metabolomics, and Exposome-Wide Association Studies. Following this, numerous large-scale exposome cohorts and initiatives have emerged, significantly bolstering research capacity and expanding the breadth of exposome research. These include consortia such as ATHLETE (Vrijheid et al., 2021) and U.S.-based initiatives including HHEAR (Viet et al., 2021) and HERCULES (Niedzwiecki & Miller, 2019), which are designed to leverage extensive multi-omics data and Exposome-Wide Association Studies to unravel complex gene-environment interactions and understand disease etiology (Avery et al., 2022; Cadiou et al., 2020; Gao et al., 2022; Maitre et al., 2022). Furthermore, these initiatives heavily rely on advanced, high-resolution non-targeted screening techniques for comprehensive exposome characterization (Manz et al., 2023). The successful transfer and expansion of the HEALS analytical backbone into subsequent initiatives like URBANOME and ENVESOME highlight the scalability, robustness, and continued relevance of such integrated methodologies for addressing broader exposure dimensions and understanding the complex interplay between the environment and human health in contemporary exposome research.

4. Conclusions

The emergence of exposome-wide association studies (ExWAS) marks a transformative step in environmental health research, analogous to how GWAS revolutionized genetic epidemiology. In this review, we have reviewed the state-of-the-art computational standards and tools that underpin ExWAS, focusing on methodological developments, software platforms, and data harmonization efforts that collectively enable linking the human exposome to health outcomes. Key takeaways from our review can be summarized as follows:

Methodological innovations: ExWAS methodology has rapidly advanced from simple single-exposure scans to sophisticated multi-dimensional analyses. Techniques such as high-dimensional penalized regression, exposure grouping by PCA/clustering, and multi-omics integration are routinely applied to tackle the challenges of correlated exposures and to extract mechanistic insights. Approaches like PBPK modeling and longitudinal data analysis further enrich ExWAS by accounting for internal dose dynamics and life-course exposure effects. These methodological tools enhance our ability to detect true associations amidst the noise of complex exposure data and to build causal narratives linking exposures to biological changes and disease.

Software and computational tools: A robust ecosystem of software platforms has been established to support exposome data analysis. R packages (e.g. rexposome, omicRexposome, CLARITE) and interactive applications like exposomeShiny encapsulate best practices, making ExWAS more accessible and reproducible for researchers. At the same time, specialized frameworks (like the INTEGRA platform) demonstrate how tailored integration of environmental modeling, biomonitoring, and −omics is feasible in a single workflow. These tools not only increase efficiency and consistency in analysis but also serve as vehicles to disseminate novel methods to the community. The trend toward open-source development and containerized deployment (Docker images, GitHub repositories, etc.) aligns exposomics with modern computational science, emphasizing transparency and collaboration.

Data harmonization and standards: The power of ExWAS can only be fully realized when data are comparable and combinable across studies. Significant progress has been made on this front: the adoption of ontologies for exposure terminology, creation of common data models and elements for epidemiological studies, and adherence to FAIR principles are increasingly seen in exposomics projects. Initiatives like the HHEAR Data Center and Exposome-Explorer database are promoting standardized data formats and deposition practices. Moreover, international efforts are coalescing around building federated data infrastructures and consensus protocols for exposome research. Although challenges remain, especially given the heterogeneity of exposure data sources, a foundation is laid such that a researcher in one study can leverage data and findings from others with confidence in their compatibility. This is crucial for increasing statistical power via pooled analyses and for ensuring that discoveries are reproducible.

Contributions of exposome pioneers: Throughout the paper, we highlighted contributions from research teams which have significantly shaped the computational landscape of exposomics. Their work on integrated methodologies (e.g., within the HEALS project), case studies in complex real-world scenarios (industrial contamination and acute chemical disasters), and pushing the envelope of exposure biology integration (transcriptomics, metabolomics within exposome studies) exemplify how multi-disciplinary expertise can be fused into practical research frameworks. These contributions underscore the importance of crossing traditional silos – combining environmental engineering, computational modeling, and epidemiology – to unravel the complexity of the exposome. As exposome research moves forward, such integrative and systems-oriented approaches championed by leading groups will be instrumental in training the next generation of researchers and setting benchmarks for what comprehensive exposure-health investigations look like.

Challenges and future outlook: We discussed ongoing challenges such as high-dimensional statistical issues, exposure measurement errors, confounding, and data privacy concerns. Recognizing these challenges is in itself leading to improvements: for instance, larger consortium studies and replication ExWAS are becoming more common to ensure robustness of findings; measurement technologies are improving to reduce error; and causal inference techniques are being more widely applied. The future directions we outlined, from incorporating genomics to leveraging wearables and AI, paint an optimistic picture of an exposomics field that is both deepening (in mechanistic understanding) and broadening (in scope and scale). Notably, there is a strong drive to translate ExWAS findings to actionable strategies – whether policy changes to reduce harmful exposures, or personalized advice for at-risk individuals. This translational ethos will likely define the next decade of

exposome research.

In conclusion, computational practices in ExWAS have matured significantly in a short time, offering researchers a rich toolkit to unravel the environmental drivers of health and disease. The convergence of robust data standards, powerful analytical tools, and comprehensive study designs allows us to move beyond studying one exposure-disease pair at a time, toward analyzing the human environment as a complex system of interacting factors – and doing so in human populations at scale. This systems approach, enabled by computation, illuminates health determinants that were previously opaque or underestimated. Insights gained from exposome studies demonstrate tangible benefits: identifying high-risk exposure mixtures, vulnerable population subgroups, early biomarkers of harmful exposure, and modifiable factors that can mitigate risk. As these insights accumulate, we anticipate they will increasingly inform public health interventions and policies, making the exposome paradigm a cornerstone of modern epidemiology and preventive medicine.

The journey is ongoing. Continued advancements in computational methods, along with international collaboration and data sharing, will be key. If the genome project taught us "what could be," the exposome project teaches us "what is" – and with that knowledge, how we might improve it. By fully harnessing the computational tools and standards at our disposal, exposome-wide association studies are poised to deliver a more profound understanding of how our environment in its entirety shapes our wellbeing, and how we can engineer healthier environments for future generations.

Funding sources

This work was supported in part by the European Commission research grants URBANOME (Grant No. 945391), and ENVESOME (Grant No. 101157269).

CRediT authorship contribution statement

Dimosthenis Sarigiannis: Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ourania Anesti:** Writing – original draft, Investigation, Formal analysis. **Nafsika Papaioannou:** Writing – original draft, Methodology, Investigation, Formal analysis. **Achilleas Karakoltzidis:** Writing – original draft, Methodology, Investigation. **Spyros Karakitsios:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Alemu, R., Sharew, N.T., Arsano, Y.Y., Ahmed, M., Tekola-Ayele, F., Mersha, T.B., Amare, A.T., 2025. Multi-omics approaches for understanding gene-environment interactions in noncommunicable diseases: techniques, translation, and equity issues. Hum. Genomics 19 (1), 8. https://doi.org/10.1186/s40246-025-00718-9.

Altman, N., Krzywinski, M., 2018. The curse(s) of dimensionality. Nat. Methods 15 (6), 399–400. https://doi.org/10.1038/s41592-018-0019-x.

Anesti, O., Papaioannou, N., Gabriel, C., Karakoltzidis, A., Dzhedzheia, V., Petridis, I., Stratidakis, A., Dickinson, M., Horvat, M., Snoj Tratnik, J., Tsatsakis, A., Karakitsios, S., Sarigiannis, D.A., 2022. An exposome connectivity paradigm for the mechanistic assessment of the effects of prenatal and early life exposure to metals on neurodevelopment. Front. Public Health 10, 871218. https://doi.org/10.3389/fpubh.2022.871218.

Anguita-Ruiz, A., Amine, I., Stratakis, N., Maitre, L., Julvez, J., Urquiza, J., Luo, C., Nieuwenhuijsen, M., Thomsen, C., Grazuleviciene, R., Heude, B., McEachan, R., Vafeiadi, M., Chatzi, L., Wright, J., Yang, T.C., Slama, R., Siroux, V., Vrijheid, M., Basagaña, X., 2023. Beyond the single-outcome approach: a comparison of outcome-wide analysis methods for exposome research. Environ. Int. 182, 108344. https://doi.org/10.1016/j.envint.2023.108344.

Ankley, G.T., Edwards, S.W., 2018. The adverse Outcome Pathway: a Multifaceted Framework supporting 21(st) Century Toxicology. Curr. Opin Toxicol. 9, 1–7. https://doi.org/10.1016/j.cotox.2018.03.004.

Atehortúa, A., Gkontra, P., Camacho, M., Diaz, O., Bulgheroni, M., Simonetti, V., Chadeau-Hyam, M., Felix, J.F., Sebert, S., Lekadir, K., 2023. Cardiometabolic risk estimation using exposome data and machine learning. Int. J. Med. Inf. 179, 105209. https://doi.org/10.1016/j.ijmedinf.2023.105209.

Avery, C.L., Howard, A.G., Ballou, A.F., Buchanan, V.L., Collins, J.M., Downie, C.G., Engel, S.M., Graff, M., Highland, H.M., Lee, M.P., Lilly, A.G., Lu, K., Rager, J.E., Staley, B.S., North, K.E., Gordon-Larsen, P., 2022. Strengthening causal inference in exposomics research: application of genetic data and methods. Environ. Health Perspect. 130 (5), 55001. https://doi.org/10.1289/ehp9098.

Bajard, L., Adamovsky, O., Audouze, K., Baken, K., Barouki, R., Beltman, J.B., Beronius, A., Bonefeld-Jørgensen, E.C., Cano-Sancho, G., de Baat, M.L., Di Tillio, F., Fernández, M.F., FitzGerald, R.E., Gundacker, C., Hernández, A.F., Hilscherova, K., Karakitsios, S., Kuchovska, E., Long, M., Blaha, L., 2023. Application of AOPs to assist regulatory assessment of chemical risks – Case studies, needs and recommendations. Environ. Res. 217, 114650. https://doi.org/10.1016/j.envres.2022.114650.

Barouki, R., Audouze, K., Coumoul, X., Demenais, F., Gauguier, D., 2018. Integration of the human exposome with the human genome to advance medicine. Biochimie 152, 155–158. https://doi.org/10.1016/j.biochi.2018.06.023.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc.: Ser. B (Methodol.) 57 (1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Billionnet, C., Sherrill, D., 2012. Estimating the health effects of exposure to multi-pollutant mixture. Ann. Epidemiol. 22, 126–141. https://doi.org/10.1016/j.annepidem.2011.11.004.

Bobb, J.F., Claus Henn, B., Valeri, L., Coull, B.A., 2018. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. Environ. Health 17 (1), 67. https://doi.org/10.1186/s12940-018-0413-y.

Buckley, J.P., Kuiper, J.R., Bennett, D.H., Barrett, E.S., Bastain, T., Breton, C.V., Chinthakindi, S., Dunlop, A.L., Farzan, S.F., Herbstman, J.B., Karagas, M.R., Marsit, C.J., Meeker, J.D., Morello-Frosch, R., O'Connor, T.G., Romano, M.E., Schantz, S., Schmidt, R.J., Watkins, D.J., Woodruff, T.J., 2022. Exposure to contemporary and emerging chemicals in commerce among pregnant women in the United States: the environmental influences on child health outcome (ECHO) program. Environ. Sci. Technol. 56 (10), 6560–6573. https://doi.org/10.1021/acs.est.1c08942.

Cadiou, S., Bustamante, M., Agier, L., Andrusaityte, S., Basagaña, X., Carracedo, A., Chatzi, L., Grazuleviciene, R., Gonzalez, J.R., Gutzkow, K.B., Maitre, L., Mason, D., Millot, F., Nieuwenhuijsen, M., Papadopoulou, E., Santorelli, G., Saulnier, P.J., Vives, M., Wright, J., Slama, R., 2020. Using methylome data to inform exposome-health association studies: an application to the identification of environmental drivers of child body mass index. Environ. Int. 138, 105622. https://doi.org/10.1016/j.envint.2020.105622.

Canzler, S., Schor, J., Busch, W., Schubert, K., Rolle-Kampczyk, U.E., Seitz, H., Kamp, H., von Bergen, M., Buesen, R., Hackermüller, J., 2020. Prospects and challenges of multi-omics data integration in toxicology. Arch. Toxicol. 94 (2), 371–388. https://doi.org/10.1007/s00204-020-02656-y.

Chung, M.K., House, J.S., Akhtari, F.S., Makris, K.C., Langston, M.A., Islam, K.T., Holmes, P., Chadeau-Hyam, M., Smirnov, A.I., Du, X., Thessen, A.E., Cui, Y., Zhang, K., Manrai, A.K., Motsinger-Reif, A., Patel, C.J., Consortium, M. o. t. E, 2024. Decoding the exposome: data science methodologies and implications in exposome-wide association studies (ExWASs). Exposome 4 (1). https://doi.org/10.1093/exposome/osae001.

Chung, M.K., Kannan, K., Louis, G.M., Patel, C.J., 2018. Toward capturing the exposome: exposure biomarker variability and coexposure patterns in the shared environment. Environ. Sci. Technol. 52 (15), 8801–8810. https://doi.org/10.1021/acs.est.8b01467.

Chung, M.K., Rappaport, S.M., Wheelock, C.E., Nguyen, V.K., Meer, T.P.v.d., Miller, G. W., Vermeulen, R., Patel, C.J., 2021. Utilizing a biology-driven approach to map the exposome in health and disease: an essential investment to drive the next generation of environmental. Discovery 129 (8), 085001. https://doi.org/10.1289/EHP8327.

Colwell, M.L., Townsel, C., Petroff, R.L., Goodrich, J.M., Dolinoy, D.C., 2023. Epigenetics and the exposome: DNA methylation as a proxy for health impacts of prenatal environmental exposures. Exposome 3 (1), osad001. https://doi.org/10.1093/exposome/osad001.

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., Wiegers, J., Wiegers, T.C., Mattingly, C.J., 2020. Comparative toxicogenomics database (CTD): update 2021. Nucleic Acids Res. 49 (D1), D1138–D1143. https://doi.org/10.1093/nar/gkaa891.

Di, Q., Rowland, S., Koutrakis, P., Schwartz, J., 2017. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. J. Air Waste Manag. Assoc. 67 (1), 39–52. https://doi.org/10.1080/10962247.2016.1200159.

El Homsi, M., Annesi-Maesano, I., 2024. An R package to perform environment-wide association studies (Ewas/Envwas) analysis. SM J. Environ. Toxicol. 7 (4).

Errahmani, M.-Y., Maio, S., Baldacci, S., Tagliaferro, S., Stanisci, I., Sarno, G., Viegi, G., La Grutta, S., Malizia, V., Rovira, J., Ramos, E., Perchard, R., Burton, O., Johnstone, E., Kabesch, M., Palir, N., Mazej, D., Tratnik, J.S., Falnoga, I., Annesi-Maesano, I., 2025. The exposure and health examination survey mother-child(ren) cohort profile: applying the exposome to the comprehension of child's health and development joining singletons and twins data. Exposome 5 (1). https://doi.org/10.1093/exposome/osaf009.

Escher, S. E., Partosch, F., Konzok, S., Jennings, P., Luijten, M., Kienhuis, A., de Leeuw, V., Reuss, R., Lindemann, K.-M., & Bennekou, S. H. (2022). Development of a roadmap for action on new approach methodologies in risk assessment. 19(6), 7341E. DOI: 10.2903/sp.efsa.2022.EN-7341.

Fayet, Y., Bonnin, T., Canali, S., Giroux, E., 2024. Putting the exposome into practice: an analysis of the promises, methods and outcomes of the European human exposome network. Soc. Sci. Med. 354, 117056. https://doi.org/10.1016/j.socscimed.2024.117056.

Foreman, A.L., Warth, B., Hessel, E.V.S., Price, E.J., Schymanski, E.L., Cantelli, G., Parkinson, H., Hecht, H., Klánová, J., Vlaanderen, J., Hilscherova, K., Vrijheid, M., Vineis, P., Araujo, R., Barouki, R., Vermeulen, R., Lanone, S., Brunak, S., Sebert, S., Karjalainen, T., 2024. Adopting mechanistic molecular biology approaches in exposome research for causal understanding. Environ. Sci. Technol. 58 (17), 7256–7269. https://doi.org/10.1021/acs.est.3c07961.

Furman, D., Campisi, J., Verdin, E., Carrera-Bastos, P., Targ, S., Franceschi, C., Ferrucci, L., Gilroy, D.W., Fasano, A., Miller, G.W., Miller, A.H., Mantovani, A., Weyand, C.M., Barzilai, N., Goronzy, J.J., Rando, T.A., Effros, R.B., Lucia, A., Kleinstreuer, N., Slavich, G.M., 2019. Chronic inflammation in the etiology of disease across the life span. Nat. Med. 25 (12), 1822–1832. https://doi.org/10.1038/s41591-019-0675-0.

Gao, P., Shen, X., Zhang, X., Jiang, C., Zhang, S., Zhou, X., Schüssler-Fiorenza Rose, S.M., Snyder, M., 2022. Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. Genome Res. 32 (6), 1199–1214. https://doi.org/10.1101/gr.276521.121.

Gennings, C., Curtin, P., Bello, G., Wright, R., Arora, M., Austin, C., 2020. Lagged WQS regression for mixtures with many components. Environ. Res. 186, 109529. https://doi.org/10.1016/j.envres.2020.109529.

Georgopoulos, P., Lioy, P., 2006. From a theoretical framework of human exposure and dose assessment to computational system implementation: the modeling environment for total risk studies (MENTOR). J. Toxicol. Environ. Health B Crit. Rev. 9, 457–483. https://doi.org/10.1080/10937400600755929.

Gibson, E.A., Nunez, Y., Abuawad, A., Zota, A.R., Renzetti, S., Devick, K.L., Gennings, C., Goldsmith, J., Coull, B.A., Kioumourtzoglou, M.-A., 2019. An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. Environ. Health 18 (1), 76. https://doi.org/10.1186/s12940-019-0515-1.

Gibson, E.A., Zhang, J., Yan, J., Chillrud, L., Benavides, J., Nunez, Y., Herbstman, J.B., Goldsmith, J., Wright, J., Kioumourtzoglou, M.A., 2022. Principal Component Pursuit for Pattern Identification in Environmental Mixtures. Environ. Health Perspect. 130 (11), 117008. https://doi.org/10.1289/ehp10479.

González-Ruiz, V., Schvartz, D., Sandström, J., Pezzatti, J., Jeanneret, F., Tonoli, D., Boccard, J., Monnet-Tschudi, F., Sanchez, J.-C., & Rudaz, S. (2019). An integrative multi-omics workflow to address multifactorial toxicology experiments. *9*(4), 79. https://www.mdpi.com/2218-1989/9/4/79.

Grau-Perez, M., Agha, G., Pang, Y., Bermudez, J.D., Tellez-Plaza, M., 2019. Mendelian randomization and the environmental epigenetics of health: a systematic review. Curr. Environ. Health Rep. 6 (1), 38–51. https://doi.org/10.1007/s40572-019-0226-3.

Hamilton, C.M., Strader, L.C., Pratt, J.G., Maiese, D., Hendershot, T., Kwok, R.K., Hammond, J.A., Huggins, W., Jackman, D., Pan, H., Nettles, D.S., Beaty, T.H., Farrer, L.A., Kraft, P., Marazita, M.L., Ordovas, J.M., Pato, C.N., Spitz, M.R., Wagener, D., Haines, J., 2011. The PhenX Toolkit: get the most from your measures. Am. J. Epidemiol. 174 (3), 253–260. https://doi.org/10.1093/aje/kwr193.

Hao, W., Cathey, A. L., Aung, M. M., Boss, J., Meeker, J. D., & Mukherjee, B. (2024). Statistical methods for chemical mixtures: a roadmap for practitioners. *medRxiv*, 2024.2003. 2003.24303677. DOI: 10.1101/2024.03.03.24303677.

He, Y., Lakhani, C.M., Rasooly, D., Manrai, A.K., Tzoulaki, I., Patel, C.J., 2021. Comparisons of polyexposure, polygenic, and clinical risk scores in risk prediction of type 2 diabetes. Diabetes Care 44 (4), 935–943. https://doi.org/10.2337/dc20-2049.

Heacock, M.L., Lopez, A.R., Amolegbe, S.M., Carlin, D.J., Henry, H.F., Trottier, B.A., Velasco, M.L., Suk, W.A., 2022. Enhancing data integration, interoperability, and reuse to address complex and emerging environmental health problems. Environ. Sci. Technol. 56 (12), 7544–7552. https://doi.org/10.1021/acs.est.1c08383.

Helbig, C., Ueberham, M., Becker, A.M., Marquart, H., Schlink, U., 2021. Wearable sensors for human environmental exposure in urban settings. Curr. Pollut. Rep. 7 (3), 417–433. https://doi.org/10.1007/s40726-021-00186-4.

Hernandez-Ferrer, C., Wellenius, G.A., Tamayo, I., Basagaña, X., Sunyer, J., Vrijheid, M., Gonzalez, J.R., 2019. Comprehensive study of the exposome and omic data using rexposome Bioconductor packages. Bioinformatics 35 (24), 5344–5345. https://doi.org/10.1093/bioinformatics/btz526.

Hu, H., Zhao, J., Savitz, D.A., Prosperi, M., Zheng, Y., Pearson, T.A., 2020. An external exposome-wide association study of hypertensive disorders of pregnancy. Environ. Int. 141, 105797. https://doi.org/10.1016/j.envint.2020.105797.

Huang, S.-Y., Yang, Y.-X., Chen, S.-D., Li, H.-Q., Zhang, X.-Q., Kuo, K., Tan, L., Feng, L., Dong, Q., Zhang, C., Yu, J.-T., 2021. Investigating causal relationships between exposome and human longevity: a Mendelian randomization analysis. BMC Med. 19 (1), 150. https://doi.org/10.1186/s12916-021-02030-4.

Ioannidis, J.P., Loy, E.Y., Poulton, R., Chia, K.S., 2009. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. Sci. Transl. Med. 1 (7), 7ps8. https://doi.org/10.1126/scitranslmed.3000247.

ISO. (2011). International Organization for Standardization ISO/IEC, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models, ISO/IEC Standard 25010, 2011. In.

Joubert, B. R., Palmer, G., Dunson, D., Kioumourtzoglou, M.-A., & Coull, B. A. (2024). Environmental Mixtures Analysis (E-MIX) Workflow and Methods Repository. *medRxiv*, 2024.2012.2020.24318087. DOI: 10.1101/2024.12.20.24318087.

Kalia, V., Walker, D.I., Krasnodemski, K.M., Jones, D.P., Miller, G.W., Kioumourtzoglou, M.-A., 2020. Unsupervised dimensionality reduction for exposome research. Curr. Opin. Environ. Sci. Health 15, 32–38. https://doi.org/10.1016/j.coesh.2020.05.001.

Keil, A.P., Buckley, J.P., O'Brien, K.M., Ferguson, K.K., Zhao, S., White, A.J., 2020. A quantile-based g-computation approach to addressing the effects of exposure mixtures. Environ. Health Perspect. 128 (4), 047004. https://doi.org/10.1289/EHP5838.

Lai, Y., Ay, M., Hospital, C.D., Miller, G.W., Sarkar, S., 2024a. Seminar: functional exposomics and mechanisms of toxicity-insights from model systems and NAMs. Environ. Health Perspect. 132 (9), 94201. https://doi.org/10.1289/ehp13120.

Lai, Y., Koelmel, J.P., Walker, D.I., Price, E.J., Papazian, S., Manz, K.E., Castilla-Fernández, D., Bowden, J.A., Nikiforov, V., David, A., Bessonneau, V., Amer, B., Seethapathy, S., Hu, X., Lin, E.Z., Jbebli, A., McNeil, B.R., Barupal, D., Cerasa, M., Miller, G.W., 2024b. High-resolution mass spectrometry for human exposomics: expanding chemical space coverage. Environ. Sci. Technol. 58 (29), 12784–12822. https://doi.org/10.1021/acs.est.4c01156.

Lee, M., Saha, A., Sundaram, R., Albert, P.S., Zhao, S., 2024. Accommodating detection limits of multiple exposures in environmental mixture analyses: an overview of statistical approaches. Environ. Health 23 (1), 48. https://doi.org/10.1186/s12940-024-01088-w.

Li, D., Zhou, L., Cao, Z., Wang, J., Yang, H., Lyu, M., Zhang, Y., Yang, R., Wang, J., Bian, Y., Xu, W., Wang, Y., 2024. Associations of environmental factors with neurodegeneration: an exposome-wide Mendelian randomization investigation. Ageing Res. Rev. 95, 102254. https://doi.org/10.1016/j.arr.2024.102254.

Liu, S.H., Chen, Y., Kuiper, J.R., Ho, E., Buckley, J.P., Feuerstahler, L., 2024. Applying Latent Variable Models to Estimate Cumulative Exposure Burden to Chemical Mixtures and Identify Latent Exposure Subgroups: a critical Review and Future Directions. Stat. Biosci. 16 (2), 482–502. https://doi.org/10.1007/s12561-023-09410-9.

Lovegrove, C.E., Howles, S.A., Furniss, D., Holmes, M.V., 2024. Causal inference in health and disease: a review of the principles and applications of Mendelian randomization. J. Bone Min. Res. 39 (11), 1539–1552. https://doi.org/10.1093/jbmr/zjae136.

Lucas, A.M., Palmiero, N.E., McGuigan, J., Passero, K., Zhou, J., Orie, D., Ritchie, M.D., Hall, M.A., 2019. CLARITE facilitates the quality control and analysis process for EWAS of metabolic-related traits [Original Research]. Front. Genet. 10–2019. https://doi.org/10.3389/fgene.2019.01240.

Maitre, L., Bustamante, M., Hernández-Ferrer, C., Thiel, D., Lau, C.-H.E., Siskos, A.P., Vives-Usano, M., Ruiz-Arenas, C., Pelegrí-Sisó, D., Robinson, O., Mason, D., Wright, J., Cadiou, S., Slama, R., Heude, B., Casas, M., Sunyer, J., Papadopoulou, E. Z., Gutzkow, K.B., Vrijheid, M., 2022. Multi-omics signatures of the human early life exposome. Nat. Commun. 13 (1), 7024. https://doi.org/10.1038/s41467-022-34422-2.

Makris, K.C., Baccarelli, A., Silverman, E.K., Wright, R.O., 2025. How exposomic tools complement and enrich genomic research. Cell Genomics 5 (8), 100952. https://doi.org/10.1016/j.xgen.2025.100952.

Manrai, A.K., Cui, Y., Bushel, P.R., Hall, M., Karakitsios, S., Mattingly, C.J., Ritchie, M., Schmitt, C., Sarigiannis, D.A., Thomas, D.C., Wishart, D., Balshaw, D.M., Patel, C.J., 2017. Informatics and data analytics to support exposome-based discovery for public health. Annu. Rev. Public Health 38, 279–294. https://doi.org/10.1146/annurev-publhealth-082516-012737.

Manz, K.E., Feerick, A., Braun, J.M., Feng, Y.-L., Hall, A., Koelmel, J., Manzano, C., Newton, S.R., Pennell, K.D., Place, B.J., Godri Pollitt, K.J., Prasse, C., Young, J.A., 2023. Non-targeted analysis (NTA) and suspect screening analysis (SSA): a review of examining the chemical exposome. J. Eposure Sci. Environ. Epidemiol. 33 (4), 524–536. https://doi.org/10.1038/s41370-023-00574-6.

Marcon, Y., Bishop, T., Avraam, D., Escriba-Montagut, X., Ryser-Welch, P., Wheater, S., Burton, P., González, J.R., 2021. Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD. PLoS Comput. Biol. 17 (3), e1008880. https://doi.org/10.1371/journal.pcbi.1008880.

Martins, C., Dreij, K., Costa, P.M., 2019. The state-of-the art of environmental toxicogenomics: challenges and perspectives of "Omics". App. Directed Tox. Mixt. 16 (23), 4718. https://www.mdpi.com/1660-4601/16/23/4718.

Mattingly, C.J., McKone, T.E., Callahan, M.A., Blake, J.A., Hubal, E.A.C., 2012. Providing the missing link: the exposure science ontology ExO. Environ. Sci. Technol. 46 (6), 3046–3053. https://doi.org/10.1021/es2033857.

McGee, G., Coull, B.A., Wilson, A., 2025. Collapsible kernel machine regression for exposomic analyses. Stat. Med. 44 (20–22), e70258. https://doi.org/10.1002/sim.70258.

Miller, G.W., 2021. Integrating the exposome into a multi-omic research framework. Exposome 1 (1). https://doi.org/10.1093/exposome/osab002.

Miller, G.W., Jones, D.P., 2014. The nature of nurture: refining the definition of the exposome. Toxicol. Sci. 137 (1), 1–2. https://doi.org/10.1093/toxsci/kft251.

Mou, X., Zhang, H., Arshad, S.H., 2025. Generalized Bayesian kernel machine regression. Stat. Methods Med. Res. 34 (2), 243–257. https://doi.org/10.1177/09622802241280784.

Neveu, V., Moussy, A., Rouaix, H., Wedekind, R., Pon, A., Knox, C., Wishart, D.S., Scalbert, A., 2016. Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. Nucleic Acids Res. gkw980. https://doi.org/10.1093/nar/gkw980.

NEXUS. *Network for Exposomics in the United States (NEXUS). NEXUS Global Network.* Retrieved 11/2025 from https://www.nexus-exposomics.org/.

Niedzwiecki, M.M., Miller, G.W., 2019. HERCULES: an academic center to support exposome research. In: Dagnino, S., Macherone, A. (Eds.), Unraveling the Exposome: A Practical View. Springer International Publishing, pp. 339–348. https://doi.org/10.1007/978-3-319-89321-1_13.

Novak, R., Robinson, J.A., Frantzidis, C., Sejdullahu, I., Persico, M.G., Kontić, D., Sarigiannis, D., Kocman, D., 2023. Integrated assessment of personal monitor applications for evaluating exposure to urban stressors: a scoping review. Environ. Res. 226, 115685. https://doi.org/10.1016/j.envres.2023.115685.

Orešič, M., McGlinchey, A., Wheelock, C.E., Hyötyläinen, T., 2020. Metabolic signatures of the exposome—quantifying the impact of exposure to environmental chemicals on human. Health 10 (11), 454. https://www.mdpi.com/2218-1989/10/11/454.

Paczkowska, M., Barenboim, J., Sintupisut, N., Fox, N.S., Zhu, H., Abd-Rabbo, D., Mee, M.W., Boutros, P.C., Reimand, J., 2020. Integrative pathway enrichment analysis of multivariate omics data. Nat. Commun. 11 (1), 735. https://doi.org/10.1038/s41467-019-13983-9.

Papaioannou, N., 2023. High Dimensional Biology To Assess The Effects Of The Exposure To Heavy Metals, Plasticizers And Pesticides To Child Neurodevelopment. Aristotle University of Thessaloniki. National Archive of PhD Theses (EADD/Didaktorika).

Passero, K., He, X., Zhou, J., Mueller-Myhsok, B., Kleber, M.E., Maerz, W., Hall, M.A., 2020. Phenome-wide association studies on cardiovascular health and fatty acids considering phenotype quality control practices for epidemiological data. In: Biocomputing 2020, pp. 659–670. DOI: 10.1142/9789811215636_0058.

Patel, C.J., 2017. Analytic complexity and challenges in identifying mixtures of exposures associated with phenotypes in the exposome era. Curr. Epidemiol. Rep. 4 (1), 22–30. https://doi.org/10.1007/s40471-017-0100-5.

Patel, C.J., Bhattacharya, J., Butte, A.J., 2010. An environment-wide association study (EWAS) on type 2 diabetes mellitus. PLoS One 5 (5), e10746. https://doi.org/10.1371/journal.pone.0010746.

Patel, C.J., Ioannidis, J.P., 2014. Studying the elusive environment in large scale. J. Am. Med. Assoc. 311 (21), 2173–2174. https://doi.org/10.1001/jama.2014.4129.

Patel, C.J., Manrai, A.K., 2015. Development of exposome correlation globes to map out environment-wide associations. Pac. Symp. Biocomput. 20, 231–242.

Patel, C.J., Pho, N., McDuffie, M., Easton-Marks, J., Kothari, C., Kohane, I.S., Avillach, P., 2016. A database of human exposomes and phenomes from the US National Health and Nutrition Examination Survey. Sci. Data 3, 160096. https://doi.org/10.1038/sdata.2016.96.

Ponzano, M., Rotem, R.S., Bellavia, A., 2025. Complex methods for complex data: key considerations for interpretable and actionable results in exposome research. Eur. J. Epidemiol. https://doi.org/10.1007/s10654-025-01281-2.

Price, E.J., Vitale, C.M., Miller, G.W., David, A., Barouki, R., Audouze, K., Walker, D.I., Antignac, J.P., Coumoul, X., Bessonneau, V., Klánová, J., 2022. Merging the exposome into an integrated framework for "omics" sciences. iScience 25 (3), 103976. https://doi.org/10.1016/j.isci.2022.103976.

Pries, L.-K., Moore, T.M., Visoki, E., Sotelo, I., Barzilay, R., Guloksuz, S., 2022. Estimating the association between exposome and psychosis as well as general psychopathology: results from the ABCD study. Biol. Psych. Global Open Sci. 2 (3), 283–291. https://doi.org/10.1016/j.bpsgos.2022.05.005.

Rappaport, S.M., 2018. Redefining environmental exposure for disease etiology. npj Syst. Biol. Appl. 4 (1), 30. https://doi.org/10.1038/s41540-018-0065-0.

Rappaport, S.M., Smith, M.T., 2010. Epidemiology. Environment and disease risks. Science 330 (6003), 460–461. https://doi.org/10.1126/science.1192603.

Reich, B.J., Guan, Y., Fourches, D., Warren, J.L., Sarnat, S.E., Chang, H.H., 2020. Integrative statistical methods for exposure mixtures and health. Ann. Appl. Stat. 14 (4), 1945–1963. https://doi.org/10.1214/20-aoas1364.

Relton, C.L., Davey Smith, G., 2012. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int. J. Epidemiol. 41 (1), 161–176. https://doi.org/10.1093/ije/dyr233.

Russ, T.C., Cherrie, M.P.C., Dibben, C., Tomlinson, S., Reis, S., Dragosits, U., Vieno, M., Beck, R., Carnell, E., Shortt, N.K., Muniz-Terrera, G., Redmond, P., Taylor, A.M., Clemens, T., van Tongeren, M., Agius, R.M., Starr, J.M., Deary, I.J., Pearce, J.R., 2021. Life course air pollution exposure and cognitive decline: modelled historical air pollution data and the lothian birth cohort 1936. J. Alzheimers Dis. 79 (3), 1063–1074. https://doi.org/10.3233/jad-200910.

Sarigiannis, D., Karakitsios, S., 2018. Addressing complexity of health impact assessment in industrially contaminated sites via the exposome paradigm. Epidemiol. Prev. 42 (5–6), 37–48.

Sarigiannis, D., Karakitsios, S., Anesti, O., Stem, A., Valvi, D., Sumner, S.C.J., Chatzi, L., Snyder, M.P., Thompson, D.C., Vasiliou, V., 2025. Advancing translational exposomics: bridging genome, exposome and personalized medicine. Hum. Genomics 19 (1), 48. https://doi.org/10.1186/s40246-025-00761-6.

Sarigiannis, D., Karakitsios, S., Gotti, A., Loizou, G., Cherrie, J., Smolders, R., De Brouwere, K., Galea, K., Jones, K., Handakas, E., 2014. Integra: from global scale contamination to tissue dose. 7th International Congress on Environmental Modelling and Software International Environmental Modelling and Software Society.

Sarigiannis, D.A., 2017. Assessing the impact of hazardous waste on children's health: the exposome paradigm. Environ. Res. 158, 531–541. https://doi.org/10.1016/j.envres.2017.06.031.

Sarigiannis, D.A., 2019. The HEALS project. In: Dagnino, S., Macherone, A. (Eds.), Unraveling the Exposome: A Practical View. Springer, pp. 405–422. https://doi.org/10.1007/978-3-319-89321-1_16.

Sarigiannis, D.A., Gotti, A., Handakas, E., Karakitsios, S.P., 2018. Informatics and data analytics to support exposome-based discovery: part 2-computational exposure biology. In: Applying Big Data Analytics in Bioinformatics and Medicine. IGI Global, pp. 145–187. DOI: 10.4018/978-1-5225-2607-0.ch007.

Sarigiannis, D.A., Hansen, U., 2012. Considering the cumulative risk of mixtures of chemicals - a challenge for policy makers. Environ. Health 11 (SUPPL.1). http://www.scopus.com/inward/record.url?eid=2-s2.0-84863309869&partnerID=40&md5=6bc972da6ae1e3a068c01251a892ae0f.

Sarigiannis, D.A., Papaioannou, N., Handakas, E., Anesti, O., Polanska, K., Hanke, W., Salifoglou, A., Gabriel, C., Karakitsios, S., 2021. Neurodevelopmental exposome: the effect of in utero co-exposure to heavy metals and phthalates on child neurodevelopment. Environ. Res. 197, 110949. https://doi.org/10.1016/j.envres.2021.110949.

Sarigiannis, D.A., Karakitsios, S.P., Handakas, E., Gotti, A., 2020. Development of a generic lifelong physiologically based biokinetic model for exposome studies. Environ. Res. 185, 109307. https://doi.org/10.1016/j.envres.2020.109307.

Schmeisser, S., Miccoli, A., von Bergen, M., Berggren, E., Braeuning, A., Busch, W., Desaintes, C., Gourmelon, A., Grafström, R., Harrill, J., Hartung, T., Herzler, M., Kass, G.E.N., Kleinstreuer, N., Leist, M., Luijten, M., Marx-Stoelting, P., Poetz, O., van Ravenzwaay, B., Tralau, T., 2023. New approach methodologies in human regulatory toxicology – not if, but how and when! Environ. Int. 178, 108082. https://doi.org/10.1016/j.envint.2023.108082.

Schmitt, C.P., Stingone, J.A., Rajasekar, A., Cui, Y., Du, X., Duncan, C., Heacock, M., Hu, H., Gonzalez, J.R., Juarez, P.D., 2023. A roadmap to advance exposomics through federation of data. Exposome 3 (1), osad010. https://doi.org/10.1093/exposome/osad010.

Sheng, Q.-S., Liu, B., Wang, X., Hua, L., Zhao, S.-C., Sun, X.-Z., Li, M.-Y., Zhang, X.-Y., Wang, J.-X., Hu, P.-L., 2025. Revolutionizing toxicological risk assessment: integrative advances in new approach methodologies (NAMs) and precision toxicology. Arch. Toxicol. 99 (12), 4697–4707. https://doi.org/10.1007/s00204-025-04169-y.

Smith, A.D., Hardy, R., Heron, J., Joinson, C.J., Lawlor, D.A., Macdonald-Wallis, C., Tilling, K., 2016. A structured approach to hypotheses involving continuous exposures over the life course. Int. J. Epidemiol. 45 (4), 1271–1279. https://doi.org/10.1093/ije/dyw164.

Smith, M., Zhang, L., McHale, C., Skibola, C., Rappaport, S., 2011. Benzene, the Exposome and Future Investigations of Leukemia Etiology. Chem. Biol. Interact. 192, 155–159. https://doi.org/10.1016/j.cbi.2011.02.010.

Sonabend, A., Zhang, J., Schwartz, J., Coull, B. A., & Lu, J. (2024). Scalable Gaussian process regression via median posterior inference for estimating multi-pollutant mixture health effects. *arXiv preprint arXiv:2411.10858*. DOI: 10.48550/arXiv.2411.10858.

Tamayo-Uria, I., Maitre, L., Thomsen, C., Nieuwenhuijsen, M.J., Chatzi, L., Siroux, V., Aasvang, G.M., Agier, L., Andrusaityte, S., Casas, M., de Castro, M., Dedele, A., Haug, L.S., Heude, B., Grazuleviciene, R., Gutzkow, K.B., Krog, N.H., Mason, D., McEachan, R.R.C., Basagaña, X., 2019. The early-life exposome: Description and patterns in six European countries. Environ. Int. 123, 189–200. https://doi.org/10.1016/j.envint.2018.11.067.

Tarraf, W., Rodríguez, C.J., Daviglus, M.L., Lamar, M., Schneiderman, N., Gallo, L., Talavera, G.A., Kaplan, R.C., Fornage, M., Conceicao, A., González, H.M., 2017. Blood pressure and hispanic/latino cognitive function: hispanic community health study/study of latinos results. J. Alzheimers Dis. 59 (1), 31–42. https://doi.org/10.3233/jad-170017.

Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D.J., Horsley, T., Weeks, L., Hempel, S., Akl, E.A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M.G., Garritty, C., Straus, S.E., 2018. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann. Intern. Med. 169 (7), 467–473. https://doi.org/10.7326/M18-0850.

Turner, M.C., Nieuwenhuijsen, M., Anderson, K., Balshaw, D., Cui, Y., Dunton, G., Hoppin, J.A., Koutrakis, P., Jerrett, M., 2017. Assessing the exposome with external measures: commentary on the state of the science and research recommendations. Annu. Rev. Public Health 38, 215–239. https://doi.org/10.1146/annurev-publhealth-082516-012802.

Tzoulaki, I., Patel, C., Okamura, T., Chan, Q., Brown, I., Miura, K., Ueshima, H., Zhao, L., Horn, L., Daviglus, M., Stamler, J., Butte, A., Ioannidis, J., Elliott, P., 2012. A nutrient-wide association study on blood pressure. Circulation 126. https://doi.org/10.1161/CIRCULATIONAHA.112.114058.

Vermeulen, R., Schymanski, E.L., Barabási, A.L., Miller, G.W., 2020. The exposome and health: where chemistry meets biology. Science 367 (6476), 392–396. https://doi.org/10.1126/science.aay3164.

Viet, S.M., Falman, J.C., Merrill, L.S., Faustman, E.M., Savitz, D.A., Mervish, N., Barr, D. B., Peterson, L.A., Wright, R., Balshaw, D., O'Brien, B., 2021. Human health exposure analysis resource (HHEAR): a model for incorporating the exposome into health studies. Int. J. Hyg. Environ. Health 235, 113768. https://doi.org/10.1016/j.ijheh.2021.113768.

Vrijheid, M., Basagaña, X., Gonzalez, J.R., Jaddoe, V.W.V., Jensen, G., Keun, H.C., McEachan, R.R.C., Porcel, J., Siroux, V., Swertz, M.A., Thomsen, C., Aasvang, G.M., Andrušaitytė, S., Angeli, K., Avraam, D., Ballester, F., Burton, P., Bustamante, M., Casas, M., Slama, R., 2021. Advancing tools for human early lifecourse exposome

research and translation (ATHLETE): Project overview. Environ. Epidemiol. 5 (5), e166.
Vrijheid, M., Slama, R., Robinson, O., Chatzi, L., Coen, M., van den Hazel, P., Thomsen, C., Wright, J., Athersuch, T.J., Avellana, N., Basagaña, X., Brochot, C., Bucchini, L., Bustamante, M., Carracedo, A., Casas, M., Estivill, X., Fairley, L., van Gent, D., Nieuwenhuijsen, M.J., 2014. The human early-life exposome (HELIX): project rationale and design. Environ. Health Perspect. 122 (6), 535–544. https://doi.org/10.1289/ehp.1307204.
Wambaugh, J.F., Bare, J.C., Carignan, C.C., Dionisio, K.L., Dodson, R.E., Jolliet, O., Liu, X., Meyer, D.E., Newton, S.R., Phillips, K.A., Price, P.S., Ring, C.L., Shin, H.-M., Sobus, J.R., Tal, T., Ulrich, E.M., Vallero, D.A., Wetmore, B.A., Isaacs, K.K., 2019. New approach methodologies for exposure science. Curr. Opin Toxicol. 15, 76–92. https://doi.org/10.1016/j.cotox.2019.07.001.
Warth, B., Spangler, S., Fang, M., Johnson, C.H., Forsberg, E.M., Granados, A., Martin, R. L., Domingo-Almenara, X., Huan, T., Rinehart, D., Montenegro-Burke, J.R., Hilmers, B., Aisporna, A., Hoang, L.T., Uritboonthai, W., Benton, H.P., Richardson, S. D., Williams, A.J., Siuzdak, G., 2017. Exposome-scale investigations guided by global metabolomics, pathway analysis, and cognitive computing. Anal. Chem. 89 (21), 11505–11513. https://doi.org/10.1021/acs.analchem.7b02759.
Weisskopf, M.G., Seals, R.M., Webster, T.F., 2018. Bias Amplification in Epidemiologic Analysis of Exposure to Mixtures. Environ. Health Perspect. 126 (4), 047003. https://doi.org/10.1289/EHP2450.
Wild, C.P., 2005. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidem Biomar 14 (8), 1847–1850. http://www.scopus.com/inward/record.url?eid=2-s2.0-23844523425&partnerID=40&md5=0354765569e43db3e95fa25bcac97de6.
Wild, C.P., 2012. The exposome: from concept to utility. Int. J. Epidemiol. 41 (1), 24–32. https://doi.org/10.1093/ije/dyr236.
Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3 (1), 160018. https://doi.org/10.1038/sdata.2016.18.
Wittwehr, C., Aladjov, H., Ankley, G., Byrne, H.J., de Knecht, J., Heinzle, E., Klambauer, G., Landesmann, B., Luijten, M., MacKay, C., Maxwell, G., Meek, M.E., Paini, A., Perkins, E., Sobanski, T., Villeneuve, D., Waters, K.M., Whelan, M., 2016. How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. Toxicol. Sci. 155 (2), 326–336. https://doi.org/10.1093/toxsci/kfw207.
Zhang, D., Eick, S. M., & Chang, H. H. (2025). Approximate Bayesian Kernel Machine Regression via Random Fourier Features for Estimating Joint Health Effects of Multiple Exposures. *arXiv preprint arXiv:2502.13157*. DOI: 10.48550/arXiv.2502.13157.
Zhang, H., Hu, H., Diller, M., Hogan, W.R., Prosperi, M., Guo, Y., Bian, J., 2021. Semantic standards of external exposome data. Environ. Res. 197, 111185. https://doi.org/10.1016/j.envres.2021.111185.
Zhou, X., Lee, S.H., 2021. An integrative analysis of genomic and exposomic data for complex traits and phenotypic prediction. Sci. Rep. 11 (1), 21495. https://doi.org/10.1038/s41598-021-00427-y.

Further reading

Joubert, B. R., Palmer, G., Dunson, D., Kioumourtzoglou, M.-A., & Coull, B. A. Workflow for Statistical Analysis of Environmental Mixtures. *Environmental Health Perspectives*, *0*(0). DOI: 10.1289/EHP16791.