Environmental Science & Technology

pubs.acs.org/est

Perspective

Vision for Exposomics-Scale Investigation: A Heuristic for Exposure-Wide Association Studies (ExWAS)

Published as part of Environmental Science & Technology special issue "60th Anniversary of Environmental Science and Technology".

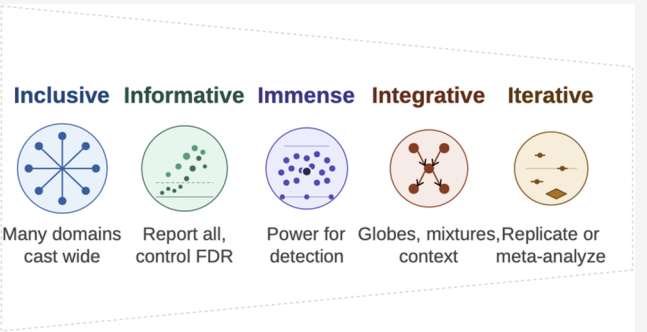
Chirag J. Patel,* Gary W. Miller, and Alison A. Motsinger-Reif

Cite This: *Environ. Sci. Technol.* 2026, 60, 15511−15515

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Exposomics has the potential to revolutionize biomedical research. However, the transition from hypothesis-driven to discovery-driven research has created tension between expansive screening and the risk of false discoveries. To meet the rigorous expectations of funders, editors, and the broader scientific community, the field requires clear benchmarks for a specific discovery-oriented design within exposomics: the Exposure-Wide Association Study (ExWAS). Here, we propose a heuristic framework for ExWAS based on the 5 I's: Inclusive, Informative, Immense, Integrative, and Iterative. We distinguish ExWAS from the broader enterprise of exposomics: high-resolution personal sensor studies, high-throughput screening in model systems, targeted biomonitoring, and single-exposure causal studies are essential and complementary. We operationalize each 'I' as a set of graded thresholds rather than absolute gates, so that smaller groups and under-resourced cohorts can participate through federated and meta-analytic designs. Mirroring the systematic rigor of genomics, these criteria provide a roadmap for minimizing spurious associations while preserving discovery power and supporting the eventual creation of an "Exposome-Phenome Catalog" analogous to the GWAS Catalog.



KEYWORDS: *exposome, exposome-wide association study, phenome, architecture, heuristic*

FROM EXPOSOMICS TO EXWAS: SCOPE OF THIS PIECE

Traditional environmental health studies identify determinants of health through hypothesis-driven investigations of specific exposures. The *exposome*—defined as "the integrated compilation of all physical, chemical, biological, and (psycho)social influences that impact biology"[1]—is a conceptual umbrella that spans many designs: targeted biomonitoring, high-resolution mass spectrometry, personal sensors, model-system screens, and population-scale association studies. With databases documenting over 500,000 exposure−phenotype hypotheses[2] and densely correlated exposures spanning multiple domains[3] that change across the life course,[4] the time is ripe for discovery-oriented approaches in exposomics.

The heuristic in this Viewpoint is scoped to one such approach, the *exposure-wide association study* (ExWAS),[5] which we consider to be the exposomics analogue of the *genome-wide association study* (GWAS). We make this explicit to avoid the implication that every exposomics study must satisfy these criteria: personal sensor time series, model-system high-throughput screens, and carefully designed single-exposure causal studies will contribute valuable exposomic evidence, such as individual-level phenotypic response, without meeting population-scale ExWAS thresholds. Just as genomics is not exclusively GWAS, the exposomics is not exclusively ExWAS.

MOVING FROM HYPOTHESIS-DRIVEN TO DISCOVERY-DRIVEN STUDIES

Analogous to genome-wide association studies (GWASs), exposure-wide association studies (ExWASs) systematically examine the relationships between hundreds to thousands of environmental factors and health outcomes. Early studies used the term *environment-wide association study* (EWAS);[6] because that acronym is now widely used for *epigenome*-wide

Received: February 13, 2026
Revised: May 1, 2026
Accepted: May 4, 2026
Published: May 27, 2026

Table 1. Graded Heuristic to Standardize ExWAS Investigation

heuristic	description	role in discovery	proposed operational threshold	rationale
inclusive	breadth of exposures measured across multiple domains (physical, chemical, biological, psychosocial, and contextual, e.g., built, occupational, digital, and policy environments)	reduces false negatives by ensuring that relevant factors are not left unmeasured. Scale is defined by the number and diversity of exposures	≥3 exposure domains and ≥100 individual exposure variables (e.g., chemical biomarkers, geospatial layers, and lifestyle/behavioral measures). Where domains are unavailable, authors should declare them as scope limitations	single domains cannot capture the complexity of the exposomes; transparency about included/excluded domains supports comparability across studies of different sizes and settings
informative	agnostic, preregistered, and transparent reporting of all tested associations under an explicit multiplicity framework	controls false positives by eliminating selective reporting and by making the multiple testing burden explicit, given correlated exposures	report all tested associations with effect sizes, CIs, and *p*-values (not only "significant" ones). Prespecify an error rate: FDR ≤ 5% (Benjamini–Hochberg) for discovery, or a Bonferroni-style threshold that accounts for the effective number of independent tests given exposure correlation. Provide preregistration or an analysis plan where feasible	operationalizes multiplicity control and transparent testing as an explicit, cross-cutting principle rather than a background issue; enables meta-analysis and prevents cherry-picking
immense	adequately powered samples, ideally with longitudinal or repeated exposure assessment, sized to the error rate and effect range of interest	maximizes power to detect true associations of modest magnitude, typical of environmental exposures	minimum sample size justified by a formal power calculation at the prespecified multiplicity-corrected threshold, typically adequate to detect $r = \sim 0.05–0.10$ (continuous) or HR = ∼1.1–1.5 (time-to-event). When a single cohort cannot reach this threshold, federated analysis or meta-analysis across smaller cohorts is an acceptable substitute	exposome associations exhibit heterogeneous, often small effects. Federated and meta-analytic designs allow smaller groups to participate meaningfully rather than being excluded
integrative	explicit linkage of exposures to demographic, social, biological, genetic, or other 'omic context; and quantification of coexposure structure	controls confounding and bias by placing exposures within biological, demographic, geographic, and social context; reveals mixtures and coexposure patterns	adjust models for ≥1 minimally sufficient confounder set report pairwise exposure–exposure and exposure–confounder correlation structure ("globes") Where biospecimens and external data permit, link to ≥1 additional data modality (e.g., genomics, metabolomics, geospatial exposures)	places exposures in biological and social context; identifies potential confounders; reveals coexposure patterns critical for mixture interpretation. Linkage requirements are graded so that studies without 'omics can still meet the criterion
iterative	independent replication and/or meta-analysis across cohorts and study designs; clear separation of discovery and replication samples to avoid double dipping	the final filter: only associations robust across populations and designs are promoted to "candidate" status for mechanistic or interventional follow-up	replication in ≥1 independent cohort, OR inclusion in a preregistered meta-analysis discovery and replication samples must be nonoverlapping; replication threshold prespecified if no replication cohort is available, findings should be explicitly labeled as "discovery-only" pending replication	independent replication is the most durable safeguard against false positives and avoids "double dipping" (reusing discovery data for inference). Labeling discovery-only findings keeps smaller single-cohort studies in the literature without overstating claims

https://doi.org/10.1021/acs.est.6c02553
Environ. Sci. Technol. 2026, 60, 15511−15515

association studies, we use ExWAS throughout for terminological clarity.

A priori evidence supports discovery-driven designs: recent investigations suggest that environmental factors may explain variation not captured by germline genetics,[7] although stochastic chance remains a competing explanation.[8] Unlike genetics, exposures are time varying, correlated, and socially patterned, requiring different assumptions about causality, measurement errors, and replication.

Traditional environmental studies isolate single exposures. ExWAS inverts this by measuring hundreds to thousands of exposures and asking without preconceived hypotheses, which are associated with phenotypes. This mirrors GWAS successes, which include systematic measurement, rigorous multiple-testing control, cross-cohort meta-analysis, and independent replication.[9] Without standardization, however, the literature risks accumulating false discoveries and the proliferation of "exposomics" as a marketing term.[10]

Traditional and exposomic studies are interdependent; traditional cohort studies establish causality, quantify magnitudes, and inform policy. ExWAS investigates associations of multiple exposures with phenotypes without prespecified hypotheses, reduces file-drawer bias in hypothesis generation, and establishes candidates for targeted studies.[11] In short, ExWAS generates hypotheses, and traditional designs, including randomized, quasi-experimental, and Mendelian randomization studies, test them. Both are essential for informing public health and precision medicine. ExWAS may integrate heterogeneous data sources that capture environmental influences across domains, including biochemical measurements from biospecimens, self-reported and sensor-based behavioral and psychosocial data, and externally linked contextual factors derived from geospatial information systems (e.g., air pollution, built environment).

HEURISTIC FOR EXWAS: THE 5 I'S

Discovery-based sciences operate at scale to achieve efficiency. They must also balance false positives and false negatives.[11,12] As exposome measurements expand, the potential for both true discoveries and spurious findings increases. False positives are spurious findings due to bias or inadequate statistical thresholds, and false negatives are true findings missed when preselecting exposures. Biases such as confounding inflate both. We propose a heuristic (Table 1) to ease this tension and maximize discovery at scale. Each 'I' addresses a distinct failure mode in large-scale environmental discovery, and multiplicity control is treated as a cross-cutting principle that binds the Informative, Integrative, and Iterative criteria together: significance thresholds under Informative define the error rates; Integrative analyses condition on the correlation structure that determines the effective number of independent tests; and Iterative replication separates discovery from inference so that the nominal error rate remains interpretable.

These thresholds are intended as graded standards. Studies that do not meet every threshold remain publishable and useful; authors and reviewers should state which criteria are met, which are partially met, and which are infeasible in their setting (e.g., absence of genomic linkage, lack of repeated measures, single-cohort analysis). This grading is meant to widen rather than restrict participation in exposomics; federated analysis, meta-analysis, consortium-level sharing, and explicit labeling of "discovery-only" findings allow all groups to contribute without requiring every individual study to carry the full infrastructural burden of a mega-cohort, but make clear the heuristics achieved, so that results may be interpreted in context.

We also recognize that current ExWAS efforts undersample several important exposure dimensions and that cohorts in low- and middle-income settings (and also in higher income settings) often lack the 'omic linkages and repeated measurements recommended by the heuristic. The Inclusive and Integrative criteria are therefore written to reward transparency about scope as much as breadth itself, so that gaps in domain coverage and data linkage are visible and comparable across studies.

HEURISTIC IN PRACTICE: EXAMPLES

Recent ExWAS illustrate the heuristic in practice. None fully satisfies all five I's; each meets the criteria unevenly, and we state this explicitly.

The Phenome-Exposome Atlas,[13] which tested 619 chemical biomarkers across 278 clinical phenotypes (~1,19,000 associations) in >50,000 participants, incorporates FDR-controlled discovery,[14] coexposure network mapping,[15] and replication across independent data. This strongly illustrates Informative (explicit FDR, all associations reported), parts of Immense (large n), Integrative (coexposure structure), and Iterative (independent replication). This partially illustrates Inclusive, as it focuses on a single-exposure domain (chemical biomarkers); however, it does not satisfy the longitudinal element of Immense. Single-exposure variance was modest (median ~0.5%), while the top 20 coexposures explained more, comparable to genetic risk scores. Reverse causality is a residual concern.

The China Serum Exposome Atlas,[16] which biomonitored 267 chemicals in 5696 participants across 12 chronic diseases, illustrates chemical and disease *Inclusive* coverage and parts of *Informative* and *Immense*. It is limited by *Iterative* (independent replication outside China) and multidomain *Inclusive* (no geospatial or behavioral linkage).

A UK Biobank lifestyle/mortality ExWAS[17] that analyzed 164 lifestyle factors in ~5,00,000 participants showed that the exposome explained ~17% of mortality variance versus <2% for polygenic risk scores. This illustrates *Immense* (large sample size and longitudinal follow-up to some degree), *Integrative* ('omics and genetic linkage), and *Inclusive* within the lifestyle/geospatial domains, but is limited with respect to chemical biomonitoring and generalizability beyond a predominantly European ancestry sample.

Taken together, these exemplars demonstrate that the 5 I's are most useful as an auditable and gradable checklist and a way for authors, reviewers, and editors to see which criteria a given study meets, which it partially meets, and which it cannot meet in its current design.

CHALLENGES AND THE PATH FORWARD

Current ExWAS reveal modest single-exposure effects but substantial cumulative exposome contributions.[7,17] This mirrors the "missing heritability" story in genetics: individual exposures may explain little, but collectively they may matter. How they cumulatively explain variation and how much of that variation is causal versus confounded requires resolution. Key design challenges include reverse causality, time-varying and correlated exposures, and differential measurement errors across exposure domains. Addressing these issues requires

https://doi.org/10.1021/acs.est.6c02553
Environ. Sci. Technol. 2026, 60, 15511−15515

strategies that explicitly model coexposures temporally rather than treating them independently to ultimately address heuristics.

Priorities that our groups and others[18] and that animate the vision of the Network of Exposomics in the United States (NEXUS) include the following.

High-Throughput Measurement Approach. The most direct way to operationalize the heuristic is to use technology for high-throughput, low-cost measurement of exposure factors.[18] This will dictate both the number of exposures measured (*Inclusive*) and the sample sizes needed to detect them (*Immense*). Cheap genotyping arrays have launched GWASs, and exposome arrays will ensure robust discoveries.

Build a Repository of Associations. All tested exposure−phenotype associations are to be documented with effect sizes, confidence intervals, p-values, and replication status, and were centralized analogously to the GWAS Catalog.[19] Common biomonitoring panels, harmonized survey instruments, and shared geospatial exposure layers are critical for enabling cross-study replication and meta-analysis. Such a resource requires reporting of null results, standardizing of exposure identifiers, and minimal metadata describing the measurement context, timing, and population (*Iterative*).

Comprehensive but High-Powered Testing. Exposomics scientists are often asked: "Isn't this just *p*-hacking?" Taking a cue from genetics, GWAS enforces stringent *p*-value thresholds (5×10^{-8}), mandatory prespecified replication, and public data sharing. ExWAS should adopt analogous standards by making the *effective* number of independent tests, the chosen error rate, and the replication plan explicit for each study (*Informative* and *Immense*).

Lower the Floor, Not the Ceiling. Replication and large samples are ideal, not gatekeepers. Federated analyses, cross-cohort meta-analyses, and explicit "discovery-only" labeling allow single-cohort and modestly resourced studies to participate without overstating their claims. Globally, the heuristic should adapt to cohorts that lack 'omic linkage, long follow-up, or independent replication samples by rewarding transparency about scope rather than penalizing its absence.

Exposomics, operationalized through rigorous ExWAS, represents a paradigm shift in environmental health: from studying one exposure at a time to systematically discovering the totality of environmental influences on health. Our heuristic provides a minimum standard and a stake in the ground and is open for discussion. Some elements will guide editors, funders, and researchers, while others will evolve. Over time, a heuristic can morph into a standard that journals require (a reporting checklist), funders demand (statements about power, domains, and replication), and researchers embrace as shared infrastructure.

Early findings have demonstrated that while individual environmental factors explain modest variance, cumulative exposome effects rival genetic contributions. As methods mature and infrastructure develops, we envision a future in which exposome risk scores[20] are evaluated for precision prevention, exposome catalogs complement the GWAS Catalog[19] in completeness, and environmental health interventions include exposomic evidence in the evidentiary chain that is alongside, not in place of, the broader family of exposomics designs that this heuristic intentionally does not cover.

AUTHOR INFORMATION

Corresponding Author

Chirag J. Patel − *Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, United States;* orcid.org/0000-0002-8756-8525; Email: chirag_patel@hms.harvard.edu

Authors

Gary W. Miller − *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States;* orcid.org/0000-0001-8984-1284

Alison A. Motsinger-Reif − *National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health, Durham, North Carolina 27709, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.6c02553

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.J.P. and G.W.M. are supported by NIEHS U24ES036819. C.J.P. is supported by NIEHS R01ES032470, NIDDK R01DK137993, and NIA RF1AG074372. The funders and employers had no role in the writing or preparation of this paper.

REFERENCES

(1) Miller, G. W.; Banbury Exposomics Consortium; et al. Integrating exposomics into biomedicine. *Science* **2025**, *388*, 356−358.

(2) Davis, A. P.; Wiegers, T. C.; Johnson, R. J.; Sciaky, D.; Wiegers, J.; Mattingly, C. J. Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Res.* **2023**, *51*, D1257−D1262.

(3) Vermeulen, R.; Schymanski, E. L.; Barabási, A.-L.; Miller, G. W. The exposome and health: Where chemistry meets biology. *Science* **2020**, *367*, 392−396.

(4) Athersuch, T. J. The Role of Metabolomics in characterizing the human exposome. *Bioanalysis* **2012**, *4*, 2207−2212.

(5) Chung, M. K.; House, J. S.; Akhtari, F. S.; Makris, K. C.; Langston, M. A.; Islam, K. T.; et al. Decoding the exposome: data science methodologies and implications in exposome-wide association studies (ExWASs). *Exposome* **2024**, *4*, No. osae001.

(6) Patel, C. J.; Bhattacharya, J.; Butte, A. J. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* **2010**, *5*, No. e10746.

(7) Lakhani, C. M.; Tierney, B. T.; Manrai, A. K.; Yang, J.; Visscher, P. M.; Patel, C. J. Repurposing large health insurance claims data to estimate genetic and environmental contributions in 560 phenotypes. *Nat. Genet.* **2019**, *51*, 327−334.

(8) Smith, G. D. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *Int. J. Epidemiol.* **2011**, *40*, 537−562.

(9) Ioannidis, J. P. A.; Loy, E. Y.; Poulton, R.; Chia, K. S. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci. Transl. Med.* **2009**, *1*, No. 7ps8.

(10) Eisen, J. A. Badomics words and the power and peril of the ome-meme. *Gigascience* **2012**, *1*, No. 6.

(11) Ioannidis, J. P. A.; Tarone, R.; McLaughlin, J. K. The False-positive to False-negative Ratio in Epidemiologic Studies. *Epidemiology* **2011**, *22*, 450−456.

(12) Lloyd, D.; House, J. S.; Akhtari, F. S.; Schmitt, C. P.; Fargo, D. C.; Scholl, E. H.; et al. Questionnaire-based exposome-wide

association studies for common diseases in the Personalized Environment and Genes Study. *Exposome* **2024**, *4*, No. osae002.
(13) Patel, C. J.; Ioannidis, J. P. A.; Manrai, A. K. An atlas of exposome-phenome associations in health and disease risk. *Nat. Med.* **2026**, *32*, 1501−1510.
(14) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **1995**, *57*, 289−300.
(15) Patel, C. J.; Manrai, A. K. Development of exposome correlation globes to map out environment-wide associations. *Pac. Symp. Biocomput.* **2015**, 231−242.
(16) You, L.; Kou, J.; Wang, M.; Ji, G.; Li, X.; Su, C.; et al. An exposome atlas of serum reveals the risk of chronic diseases in the Chinese population. *Nat. Commun.* **2024**, *15*, No. 2268.
(17) Argentieri, M. A.; Amin, N.; Nevado-Holgado, A. J.; Sproviero, W.; Collister, J. A.; Keestra, S. M.; et al. Integrating the environmental and genetic architectures of aging and mortality. *Nat. Med.* **2025**, *31*, 1−10.
(18) Manrai, A. K.; Cui, Y.; Bushel, P. R.; Hall, M.; Karakitsios, S.; Mattingly, C. J.; et al. Informatics and Data Analytics to Support Exposome-Based Discovery for Public Health. *Annu. Rev. Public Health* **2017**, *38*, 279−294.
(19) Welter, D.; MacArthur, J.; Morales, J.; Burdett, T.; Hall, P.; Junkins, H.; et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **2014**, *42*, D1001−D1006.
(20) He, Y.; Lakhani, C. M.; Rasooly, D.; Manrai, A. K.; Tzoulaki, I.; Patel, C. J. Comparisons of Polyexposure, Polygenic, and Clinical Risk Scores in Risk Prediction of Type 2 Diabetes. *Diabetes Care* **2021**, *44*, 935−943.

https://doi.org/10.1021/acs.est.6c02553
Environ. Sci. Technol. 2026, 60, 15511−15515