HHS Public Access

Author manuscript

Environ Int. Author manuscript; available in PMC 2020 September 01.

Published in final edited form as:

Environ Int. 2019 September ; 130: 104877. doi:10.1016/j.envint.2019.05.071.

Using Phenome-Wide Association Studies to Examine the Effect of Environmental Exposures on Human Health

Joseph M. Braun[1], **Geetika Kalloo**[1], **Samantha L. Kingsley**[1], **Nan Li**[1]

[1]Department of Epidemiology, Brown University, Providence, RI

Introduction

In the last two decades, the field of environmental epidemiology has used several "-omics" platforms in an untargeted fashion to gain new insights into the complex relations between environmental pollutant exposures and human health.[1] For instance, exposomics has been used to understand how exposure to complex mixtures of air pollution, chemicals, and metals affect health.[2–4] Metabolomics and methylomics have been used to identify putative biological pathways affected by environmental exposures, as well as biological responses to exposures.[5,6] Despite progress in understanding the health effects of and biological responses to complex exposures, relatively little has been done to understand how environmental pollutants affect complex disease phenotypes. Many environmental exposures – air pollution, lead, and secondhand tobacco smoke – have been associated with multiple diseases, but potentially informative patterns of multimorbidity have often been ignored.[7–9]

We propose to use the phenome as a novel approach to study the health effects of environmental exposures. We define the phenome as the patterns and profiles of human disease that individuals experience from birth to death; this includes disease diagnoses, continuous traits related to disease, and biological pathways underlying disease states. Thus, the phenome represents a continuum that spans from the biological pathways underpinning disease to the clinical manifestations of disease. Quantifying the patterns of multimorbidity associated with an environmental pollutant exposure may provide new information about the health effects of that exposure, as well as potential biological pathways related to an exposure. Here we describe how the Phenome-Wide Association Study (PheWAS) can be used as a tool to better understand how environmental exposures impact the multitude of health states that humans experience across the life course.

Contact Information Joseph M. Braun, Department of Epidemiology, Box G-S121-2, Brown University, Providence, RI 02912, joseph_braun_1@brown.edu.

Conflicts of Interest: JMB was financially compensated for serving as an expert witness for plaintiffs in litigation related to tobacco smoke exposures.

PheWAS Background

In the early 2000's, as the Human Genome Project was nearing completion, scientists were contemplating if phenotypes could be "sequenced" by the Human Phenome Project in an effort to understand how individual genes were associated with multiple phenotypes, thus gaining insight into pleiotropic effects of genes.[10,11] In the context of studying individual single nucleotide polymorphisms (SNPs) or pollutant exposures, the PheWAS can be thought of as a reverse genome-wide association study (GWAS) (Figure 1). The GWAS estimates associations between thousands of SNPs and one or a few phenotypes or diseases, much like an environment-wide association study (EWAS) that examines the associations between multiple exposures and a phenotype.[12] In contrast to these approaches, the PheWAS estimates associations of one or a few SNPs (or exposures) with hundreds or thousands of phenotypes or health states in order to identify patterns of multimorbidity related to a given gene or exposure.

Note, the terms comorbidity and multimorbidity have been used interchangeably in the literature with little consensus on the best definition.[13] Here, we use the term multimorbidity and cumulative hierarchy proposed by van den Akker.

1. Simple multimorbidity, which includes causal, correlated, and coincidental disease co-occurrence. For example, the co-occurrence of cardiovascular disease and osteoarthritis is likely coincidental.
2. Associative multimorbidity implies a statistical relation between two or more diseases, and thus, includes causally-related and correlated diseases. For example, the symptoms of some micronutrient deficiencies are correlated, but non-causal since they are all related to the same common cause (e.g., Vitamin C deficiency).
3. Causal multimorbidity implies a causal relation between two or more diseases. For example, the co-occurrence of type 2 diabetes and diabetic retinopathy is causal since the former causes the latter.

The first PheWAS we are aware of examined the association between five SNPs and 776 diseases or phenotypes in over 6,000 adults.[14] They used the International Classification of Disease-9 (ICD-9) codes to define binary disease phenotypes (i.e., cases or controls). Many subsequent PheWAS adopted this model and have examined the association of SNPs with a range of clinical outcomes, usually derived from electronic medical records (EMRs).[15] Some studies have examined clinical biomarkers such as white blood cell counts or autoantibodies instead of SNPs as their primary exposure.[16,17] For instance, Liao et al. reported associations between autoantibodies and clinical diagnoses defined by ICD-9 codes, finding that antinuclear antibodies were associated with Sjögren's/sicca syndrome.[17] Table 1 summarizes the design and results of several PheWAS studies.

We are unaware of any PheWAS examining an environmental pollutant as the exposure. However, Chen and VanderWeele used a PheWAS approach (referred to as an Outcome-Wide Association Study by the authors) to examine the relations of religious service

attendance and prayer/meditation with 26 character strengths and psychological, mental, behavioral, and physical health outcomes.[18,19]

Phenome and PheWAS Framework

Regardless of the exposure of interest, a PheWAS begins by carefully defining the phenome and selecting appropriate data resources and study designs. Most PheWAS have used the ICD codes to identify specific clinical diseases. Other approaches to define phenotypes include using clinical text data from EMRs or data from large population-based studies.[18,20,21] While prospective designs are the most robust in terms of causal inference, cross-sectional studies could be used to generate new hypotheses.

Phenotypes can be classified using available ontologies like Phecodes or the Human Phenotype Ontology.[22–24] Phecodes was developed to aggregate the ICD9 or ICD10 codes into hierarchical trait or disease-relevant groupings that can be used for biomedical research.[23] The Human Phenotype Ontology also uses hierarchical groupings, but they are based on phenotypic abnormalities encountered in human disease and not billing codes.[22,24] It is important to note that these ontologies use binary classifications of the phenotype (e.g., disease vs. no disease). However, in many epidemiologic studies, outcomes are measured as continuous traits (e.g., blood pressure), that can also be characterized as clinically vs. non-clinically significant (e.g., hypertension). Thus, continuous traits are important to study because they may detect earlier manifestations of disease that clinical diagnosis would otherwise miss and provide a relative ranking of the outcomes while enhancing statistical power.

Here we expand the scope of the phenome beyond clinical or disease diagnoses as has been done in previous studies. We propose that the phenome includes clinical diagnoses, continuous traits underlying these diagnoses, and biological pathways related to these traits or diagnoses. For example, individuals can be diagnosed as obese or normal weight based on their body mass index (BMI), which in turn is a continuous trait that is used to assess an individual's adiposity. Biological pathways related to the development or maintenance of obesity include hormones produced by adipose tissue (e.g., adipocytokines).

Clinical Data

Many PheWAS take advantage of EMRs that include the ICD codes. These types of data could be used to examine associations between environmental exposures and clinical disease diagnoses. The ICD-9 coding system contains a wide spectrum of phenotypes, including over 17,000 disease codes grouped in a multi-level hierarchy.[25,26] Because the ICD coding system was designed primarily for billing and administrative functions, customized groupings of ICD codes are needed to approximate clinical disease phenotypes for a PheWAS. For example, similar ICD codes like primary tuberculosis and late effects of tuberculosis should be combined, but similar codes representing distinctly different diseases, like Type 2 and Type 2 diabetes, should be separated.[14] Finally, another approach that may capture more detailed information, is to use clinical text, examination, or laboratory data from EMRs instead of ICD-9 codes to define the phenome. For instance, Hebbring and

colleagues used EMR data to develop a text-based phenome by documenting clinical text and reducing it to clinically relevant phenotypes.[20]

Two limitations of using administrative databases containing ICD codes or EMRs for a PheWAS are worth noting. First, these databases have a limited number of environmental pollutant exposures available. Ambient air pollution, temperature, or other built environment factors could be assessed by linking participant addresses to publicly available datasets. In addition, for some sub-populations, individual-level environmental exposure data might be available (e.g., childhood blood lead concentrations). Second, outcome misclassification may arise when relying on the ICD codes or EMRs, which would result in reduced statistical power, assuming non-differential misclassification. For instance, the ICD codes are specific, but not sensitive, at classifying cardiovascular and chronic kidney disease.[27,28]

Large Datasets with Detailed Phenotyping

Large datasets, such as the National Health and Nutrition Examination Surveys (NHANES), conduct biomonitoring for a wide range of ubiquitous environmental chemicals and assess a large and diverse set of phenotypes.[29] A number of disease diagnoses, continuous phenotypes, and underlying biological pathway data have been systematically assessed using questionnaires, direct assessments, and biomarkers in the NHANES (Table 1). These include anthropometry, oral health, metabolic and endocrine biomarkers, neurodevelopment, respiratory health, allergies, and questionnaire data related to numerous disease diagnoses.[30] Other data resources that would have detailed phenotype information include ongoing prospective cohort studies of adults or children, including the National Institutes of Health Environmental Determinants of Child Health Outcomes (ECHO) Study.[31–33]

Several limitations to using these types of datasets for PheWAS are worth noting. First, some databases, like the NHANES, are cross-sectional, thus, creating temporal ambiguity between exposure and phenotypes. Second, cross-sectional data could only be used to study prevalent conditions. Third, some of these databases would have low statistical power for rare conditions (e.g., specific forms of cancer). Finally, some self-reported diagnoses may not be completely accurate, but in some cases they could be augmented by clinical examination data (e.g., measured blood pressure instead of self-reported diagnosis of hypertension).

Analyzing and Interpreting Phenome Data

To facilitate analysis and interpretation, phenotypic information could be hierarchically classified based on available ontologies or characterized in coarser groupings based on organ systems (e.g., cardiovascular vs. metabolism).[22–24] These classes could serve as a “backbone” that can be used to organize the array of assessed phenotypes and facilitate interpretation of associations between a given pollutant and multiple phenotypes. This is akin to chromosomes in a GWAS or groups of exposure (e.g., metals, phthalates, pesticides, etc.) in a EWAS. For example, using the Human Phenotype Ontology, Type 2 diabetes and hypothyroidism are both classified as endocrine system abnormalities, but further distinctions can be made based on the specific endocrine organs affected.

Some additional considerations should be made when curating phenotype data. First, the same phenotype is often measured with multiple measures and some of these measures are

highly correlated (e.g., different measures of adiposity).[34] Thus, it may only be necessary to include one of these measures depending on the degree of correlation and goal of the specific PheWAS. Additionally, some diseases or phenotypes with similar etiologies may need to be distinguished based on lifestage (e.g., Type 2 diabetes vs. gestational diabetes).

When analyzing phenome data, the entire phenome could be examined or a specific class of the phenome. This latter approach could be used when there is limited phenotype data available for some classes or there is an *a priori* hypothesis about the potential effects of an exposure. For instance, one could conduct a PheWAS to examine the association between a potentially immunotoxic compound and immune-related outcomes.

Interpreting PheWAS results can be facilitated by examining patterns of associations between the exposure and outcomes within and across phenotype classes. Using ontologies and the observed pattern of exposure-associated multimorbidity, decisions about the causal or non-causal nature of the relations between an exposure and outcome(s) can be made. For instance, observing associations between an exposure and multiple cardiovascular endpoints might suggest a common biological mechanism of action for that agent. However, an exposure associated with two biologically unrelated diseases could suggest a non-causal multimorbidity.

Like GWAS and EWAS, replication is necessary for PheWAS to ensure that significant associations are not spurious. Replication studies could be conducted on a portion of the original data (e.g. 20%) or another dataset with similar features of the original data set.

Advantages of PheWAS

The PheWAS has several potential applications to the field of environmental epidemiology that could help enhance our knowledge about specific pollutants and biological pathways related to these pollutants.

First, the PheWAS can be used as a tool to generate new hypotheses about specific exposures and human health. By examining a multitude of phenotypes, the PheWAS can efficiently provide information for exposures with little or no data about their potential health effects. Thus, the PheWAS can be used to guide the development of more targeted studies in cases where human health data are lacking. For instance, this can be quite important as some chemicals are phased out of commerce and industry, and replaced with compounds that have little or no toxicity data available (e.g., phthalate and perfluoroalkyl substance replacements).

Second, the PheWAS can improve our understanding of environmental exposures and related biological pathways by examining patterns of phenotypes associated with a single exposure. Because many disease processes are related to common biological pathways, exposure-induced effects on a given pathway or set of pathways could produce ‘environmentally pleiotropic’ effects. Thus, the PheWAS can provide evidence that an exposure alters specific biological pathways if that exposure is associated with multiple diseases or phenotypes related to that pathway. For example, active and secondhand tobacco smoke exposures are associated with the metabolic syndrome, a constellation of symptoms that includes excess

central adiposity, hypertension, dyslipidemia, impaired glucose tolerance, and insulin resistance.[35–37] Thus, tobacco smoke exposure may cause these effects by altering inflammatory, epithelial, and vascular pathways that are related to features of the metabolic syndrome.

Finally, the PheWAS can be used to generate new hypotheses about established toxicants (e.g., lead or tobacco smoke exposure) in an effort to more comprehensively assess their potential health effects. Novel exposure-phenotype associations would be difficult to identify when studying individual outcomes one-at-a-time. Exposure-associated patterns of multimorbidity may occur when an exposure affects a biological pathway related to multiple diseases or phenotypes. For instance, children often have multiple neurodevelopmental disorders (e.g., both attention-deficit/hyperactivity disorder and conduct disorder), and this pattern of multimorbidity may be related to perturbations of the same biological pathway(s).[38] Indeed, some environmental neurotoxicants, like lead, have been associated with both attention-deficit/hyperactivity disorder and conduct disorder.[39,40] Moreover, the PheWAS approach avoids selective reporting and publication bias by describing an exposure's association with all outcomes, even those that are null.

Challenges to Conducting PheWAS

Despite the advantages of PheWAS, there are several challenges in implementing them related to multiple testing, data availability, phenotyping quality, sample size, analyzing and characterizing phenotypes, the dynamic nature of exposure and outcomes, and controlling for confounding.

As is the case with all high dimensional data, there is a risk of false positives when examining associations between a single exposure and hundreds or thousands of phenotypes. For instance, there are over 17,000 potential ICD-9 codes and >155,000 ICD-10 codes.[41] Traditionally, null hypothesis testing with correction for multiple comparisons is used to "filter" out potentially false positive results using Bonferroni correction, family-wise error rates correction, or false discovery rate control.[42] Alternatively, statistical techniques can be used to reduce the dimensionality of the phenotype data (e.g., principal components). However, these statistical techniques could produce components or clusters that are difficult to interpret, not related to the exposure, or be method-dependent.[43]

Another potential concern when conducting a PheWAS is the availability and quality of the phenotyping data. One cannot acquire phenotype data in a similar to that used for other "-omics" technologies. In genomics, epigenomics, and metabolomics, thousands of features can be interrogated simultaneously on a single platform (e.g., sequencing, microarrays, or mass-spectrometry based approaches). Despite their high cost on a per-assay level, these platforms are quite efficient on a per-feature level. However, these platforms do not exist for phenotyping, thus making it more challenging to conduct a PheWAS.

PheWAS require studies that collectively have a large sample size and common protocol for phenotype assessment. Examples include EMR databases, the NHANES, or extraordinary cohorts such as the Nurses' Health Study or ECHO.[44,45] While smaller cohort studies with

detailed and research-quality phenotype measures could be used to conduct a PheWAS, they may assess a small number of diseases or phenotypes, have limited statistical power in the face of multiple testing correction, and be unable to examine rare diseases. Larger studies using EMRs will have access to a fuller spectrum of clinical disorders and sufficient statistical power to analyze most rare diseases, but there may be misclassification of some outcomes and inability to examine biological pathways. A hybrid approach could be employed where larger studies are used for discovery and smaller studies for replication and interrogation of specific biological pathways.

The dynamic nature of exposure must be considered in PheWAS. In genetic studies employing PheWAS, the exposure (i.e., single nucleotide polymorphisms) is static across the lifespan. However, in the environmental PheWAS, exposures change across the lifespan and there may be discrete periods of vulnerability for some exposures that differ with respect to phenotype.[46] There are at least three strategies to deal with this. First, PheWAS studies could examine exposures exhibiting less within-person variation (e.g., persistent pollutants) since exposure misclassification would be reduced relative to exposures with more within-person variation (e.g., bisphenol A).[47] Second, cumulative measures of exposure representing specific periods of life could be used (e.g., deciduous tooth biomarkers).[48] Third, exposure during discrete periods of life could be examined (e.g., early childhood or concurrent), acknowledging that they may not be relevant for some health outcomes

In addition, phenotypes change over time. For example, some phenotypes might not manifest until a specific age (e.g., pubertal development) or some diseases might resolve (e.g., eczema). Moreover, the development of one disease may increase the risk of another disease or phenotype. For instance, adults with type 2 diabetes have impairments in executive functions, which might arise because of diabetes-induced damage to brain microvasculature.[49] The possibility of a chain of risk makes it necessary to consider the longitudinal nature of health trajectories and comorbidities when conducting a PheWAS study.

Finally, in any observational study, proper confounding control requires adjustment for predictors of both exposure and outcome, while not adjusting for causal intermediates or colliders.[50] As Vanderweele points out, this can be relatively easily accomplished in a PheWAS since only one exposure is being considered and it is sufficient to adjust only for predictors of exposure, even if not all these variables predict the outcome.[19] This is in contrast to exposomic studies, which consider a multitude of exposures, some of which are correlated to one another due to shared exposure sources (e.g., phthalates),[51] or may be related to each other through various causal pathways (e.g., built environment factors affecting activity patterns). Additionally, measuring confounder data prospectively with respect to exposure mitigates the risk of adjusting for colliders or intermediates. Investigators could minimize bias from confounding and adjustment for colliders or intermediates by carefully select confounding variables using directed acyclic graphs or single world intervention graphs.[52,53]

Future Directions

The time has come for the field of environmental epidemiology to embrace the phenome given the increasing emphasis on studying complex exposures and biological pathways using different "-omics" methods. As a first step, we propose that PheWAS studies of relatively well-characterized exposures be conducted to demonstrate the utility of these studies and identify hurdles to implementing them. This will require the identification and curation of data resources that have assessed at least one environmental exposure and the phenome. Potential resources include the NHANES, large cohorts studies (e.g., Nurse's Health Study),[32] or EMR databases. More specialized resources focused on distinct life stages (i.e., children's health) could pool data from existing National Institute of Environmental Health Sciences funded Children's Environmental Health Centers or the National Institute of Health funded ECHO Study.[31,54]

Several issues bear additional reflection as we incorporate the PheWAS into our "-omics" toolbox. First, it will be important to consider how to incorporate life course approaches into studies of the phenome given that many diseases and phenotypes have early life origins and are dynamic in nature.[55,56] Second, there will be the need to consider how to combine and analyze different forms of highly dimensional data (e.g., the exposome and phenome). Already, there have been calls for more integration of different types of molecular data, but this call could be extended to include the exposome and phenome as well.[57] Finally, and related to this, it will be necessary to examine the relations between environmental pollutant mixtures and the phenome.[58] Ultimately, studying the relations between complex exposures, biological pathways, and the phenome across the lifespan may lead to new insights about the contribution of environmental exposures to human health and wellbeing.

Acknowledgement

We thank Dr. David Savitz for his feedback on an earlier draft of this manuscript.

Grant Funding: This work was supported by grants from the National Institute of Environmental Health Sciences grants (R01 ES025214, R01 ES024381, and R01 ES027408) and National Institutes of Health (UG3 OD023313).

References

1. Coughlin SS. Toward a road map for global -omics: a primer on -omic technologies. American journal of epidemiology 2014;180:1188–95. [PubMed: 25409464]
2. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 2005;14:1847–50. [PubMed: 16103423]
3. Vrijheid M The exposome: a new paradigm to study the impact of environment on health. Thorax 2014;69:876–8. [PubMed: 24906490]
4. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PLoS One 2010;5:e10746. [PubMed: 20505766]
5. Walker DI, Uppal K, Zhang L, et al. High-resolution metabolomics of occupational exposure to trichloroethylene. Int J Epidemiol 2016;45:1517–27. [PubMed: 27707868]
6. Joubert BR, Haberg SE, Nilsen RM, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect 2012;120:1425–31. [PubMed: 22851337]

7. Brook RD, Rajagopalan S, Pope CA, 3rd, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation 2010;121:2331–78. [PubMed: 20458016]
8. NTP monograph on health effects of low-level lead. NTP Monogr 2012:xiii, xv–148.
9. The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General. Atlanta (GA)2006.
10. Freimer N, Sabatti C. The Human Phenome Project. Nature Genetics 2003;34:15–21.
11. International Human Genome Sequencing C. Initial sequencing and analysis of the human genome. Nature 2001;409:860. [PubMed: 11237011]
12. Patel CJ, Rehkopf DH, Leppert JT, et al. Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States national health and nutrition examination survey. Int J Epidemiol 2013;42:1795–810. [PubMed: 24345851]
13. van den Akker M, Buntinx F, Knottnerus J. Comorbidity or multimorbidity: what's in a name? A review of literature. Eur J Gen Pract 1996;2:65–70.
14. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 2010;26:1205–10. [PubMed: 20335276]
15. Ye Z, Mayer J, Ivacic L, et al. Phenome-wide association studies (PheWASs) for functional variants. European Journal of Human Genetics 2015;23:523–9. [PubMed: 25074467]
16. Warner JL, Alterovitz G. Phenome Based Analysis as a Means for Discovering Context Dependent Clinical Reference Ranges. AMIA Annual Symposium Proceedings 2012;2012:1441–9. [PubMed: 23304424]
17. Liao KP, Kurreeman F, Li G, et al. Associations of Autoantibodies, Autoimmune Risk Alleles, and Clinical Diagnoses From the Electronic Medical Records in Rheumatoid Arthritis Cases and Non–Rheumatoid Arthritis Controls. Arthritis & Rheumatism 2013;65:571–81. [PubMed: 23233247]
18. Chen Y, Kim ES, Koh HK, Frazier AL, VanderWeele TJ. Sense of Mission and Subsequent Health and Well-being among Young Adults:An Outcome-Wide Analysis. American journal of epidemiology 2019.
19. VanderWeele TJ. Outcome-wide Epidemiology. Epidemiology 2017;28:399–402. [PubMed: 28166102]
20. Hebbring SJ, Rastegar-Mojarad M, Ye Z, Mayer J, Jacobson C, Lin S. Application of clinical text data for phenome-wide association studies (PheWASs). Bioinformatics 2015;31:1981–7. [PubMed: 25657332]
21. Polimanti R, Kranzler HR, Gelernter J. Phenome-Wide Association Study for Alcohol and Nicotine Risk Alleles in 26394 Women. Neuropsychopharmacology 2016;41:2688–96. [PubMed: 27187070]
22. Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res 2019;47:D1018–D27. [PubMed: 30476213]
23. Wu P, Gifford A, Meng X, et al. Developing and Evaluating Mappings of ICD-10 and ICD-10-CM codes to Phecodes. bioRxiv (2018), 10.110/62077
24. Human Phenotype Ontology. Introduction to Human Phenotype Ontology. https://hpo.jax.org/app/help/introduction Accessed on May 1st, 2019.
25. Israel RA. The International Classification of Disease. Two hundred years of development. Public Health Rep 1978;93:150–2. [PubMed: 635089]
26. Israel RA. The history of the International Classification of Diseases. Health bulletin 1991;49:62–6. [PubMed: 1938377]
27. Kern EF, Maney M, Miller DR, et al. Failure of ICD-9-CM codes to identify patients with comorbid chronic kidney disease in diabetes. Health services research 2006;41:564–80. [PubMed: 16584465]
28. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. Med Care 2005;43:480–5. [PubMed: 15838413]
29. Nature, 409 (6822) (2001), pp. 860–921 [PubMed: 11237011]

30. Curtin LR, Mohadjer LK, Dohrmann SM, et al. The National Health and Nutrition Examination Survey: Sample Design, 1999–2006. Vital Health Stat 2 2012:1–39.
31. Gillman MW, Blaisdell CJ. Environmental influences on Child Health Outcomes, a Research Program of the National Institutes of Health. Curr Opin Pediatr 2018;30:260–2. [PubMed: 29356702]
32. Bao Y, Bertoia ML, Lenart EB, et al. Origin, Methods, and Evolution of the Three Nurses' Health Studies. American journal of public health 2016;106:1573–81. [PubMed: 27459450]
33. Braun JM, Kalloo G, Chen A, et al. Cohort Profile: The Health Outcomes and Measures of the Environment (HOME) study. International journal of epidemiology 2016;46:24.
34. Boeke CE, Oken E, Kleinman KP, Rifas-Shiman SL, Taveras EM, Gillman MW. Correlations among adiposity measures in school-aged children. BMC Pediatr 2013;13:99. [PubMed: 23799991]
35. Sun K, Liu J, Ning G. Active smoking and risk of metabolic syndrome: a meta-analysis of prospective studies. PloS one 2012;7:e47791. [PubMed: 23082217]
36. Weitzman M, Cook S, Auinger P, et al. Tobacco smoke exposure is associated with the metabolic syndrome in adolescents. Circulation 2005;112:862–9. [PubMed: 16061737]
37. Xie B, Palmer PH, Pang Z, Sun P, Duan H, Johnson CA. Environmental tobacco use and indicators of metabolic syndrome in Chinese adults. Nicotine Tob Res 2010;12:198–206. [PubMed: 20056689]
38. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry 2003;60:837–44. [PubMed: 12912767]
39. Braun JM, Kahn RS, Froehlich T, Auinger P, Lanphear BP. Exposures to environmental toxicants and attention deficit hyperactivity disorder in U.S. children. Environ Health Perspect 2006;114:1904–9. [PubMed: 17185283]
40. Braun JM, Froehlich TE, Daniels JL, et al. Association of environmental toxicants and conduct disorder in U.S. children: NHANES 2001–2004. Environ Health Perspect 2008;116:956–62. [PubMed: 18629321]
41. ICD-10. (Accessed July 17, 2018, at http://www.cms.gov/Medicare/Coding/ICD10/index.html?redirect=/icd10.)
42. Chadeau-Hyam M, Campanella G, Jombart T, et al. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. Environ Mol Mutagen 2013;54:542–57. [PubMed: 23918146]
43. Ng SK, Tawiah R, Sawyer M, Scuffham P. Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. International journal of epidemiology 2018.
44. Gillman MW, Blaisdell CJ. Environmental influences on child health outcomes, a research program of the National Institutes of Health. Current opinion in pediatrics 2018;30:260–2. [PubMed: 29356702]
45. Colditz GA, MANSON JE, HANKINSON SE. JE, HANKINSON SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. Journal of Women's Health 1997;6:49–62.
46. Buckley JP, Hamra GB, Braun JM. Statistical Approaches for Investigating Periods of Susceptibility in Children's Environmental Health Research. Curr Environ Health Rep 2019.
47. Braun JM, Gray K. Challenges to studying the health effects of early life environmental chemical exposures on children's health. PLoS biology 2017;15:e2002800. [PubMed: 29257831]
48. Andra SS, Austin C, Wright RO, Arora M. Reconstructing pre-natal and early childhood exposure to multi-class organic chemicals using teeth: Towards a retrospective temporal exposome. Environment international 2015;83:137–45. [PubMed: 26134987]
49. Vincent C, Hall PA. Executive Function in Adults With Type 2 Diabetes: A Meta-Analytic Review. Psychosomatic medicine 2015;77:631–42. [PubMed: 25469685]
50. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 2009;20:488–95. [PubMed: 19525685]

51. Robinson O, Basagana X, Agier L, et al. The Pregnancy Exposome: Multiple Environmental Exposures in the INMA-Sabadell Birth Cohort. Environmental science & technology 2015;49:10632–41. [PubMed: 26168307]

52. Breskin A, Cole SR, Hudgens MG. A practical example demonstrating the utility of single-world intervention graphs. Epidemiology (Cambridge, Mass) 2018;29:e20–1.

53. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology (Cambridge, Mass) 1999;10:37–48.

54. Buckley JP, Engel SM, Braun JM, et al. Prenatal Phthalate Exposures and Body Mass Index Among 4- to 7-Year-old Children: A Pooled Analysis. Epidemiology (Cambridge, Mass 2016;27:449–58.

55. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. International journal of epidemiology 2002;31:285–93. [PubMed: 11980781]

56. Barker DJ. Sir Richard Doll Lecture. Developmental origins of chronic disease. Public Health 2012;126:185–9. [PubMed: 22325676]

57. Everson TM, Marsit CJ. Integrating -Omics Approaches into Human Population-Based Studies of Prenatal and Early-Life Exposures. Curr Environ Health Rep 2018.

58. Braun JM, Gennings C, Hauser R, Webster TF. What Can Epidemiological Studies Tell Us about the Impact of Chemical Mixtures on Human Health? Environmental health perspectives 2016;124:A6–9. [PubMed: 26720830]

Highlights

- Epidemiology has not considered whether pollutants have pleiotropic effects.
- The phenome is the patterns/profiles of disease experienced from birth to death.
- Phenome Wide Association Studies (PheWAS) examine a pollutant and all phenotypes.
- Using PheWAS could improve our understanding of the health effects of pollutants.

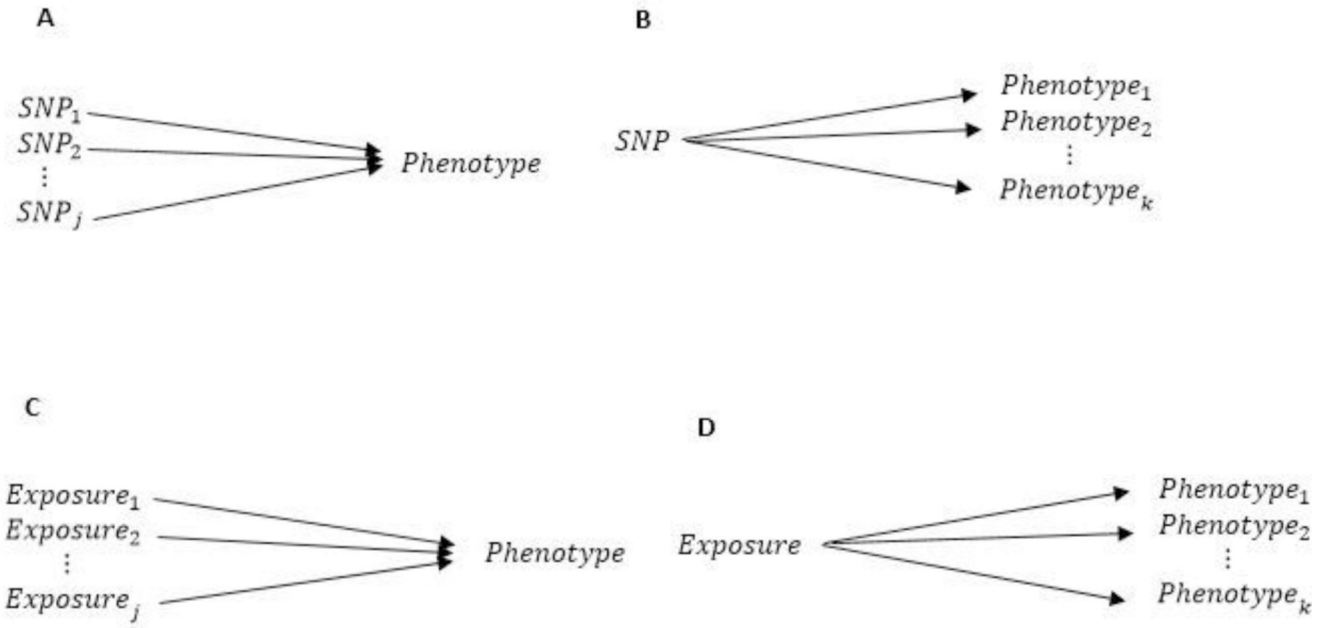

Figure 1:

Graphical depiction of a GWAS (A), genotype PheWAS (B), EWAS (C), and exposure PheWAS (D)

*-GWAS: Genome Wide Association Study, PheWAS: Phenome Wide Association Study, EWAS: Exposome Wide Association Study.

Table 1:

Selected examples of prior phenome wide association studies

Paper	N	Exposure	Phenome	Number of phenotypes	Results
Denny et al. 2010	6,005	SNPs (5) previously linked to atrial fibrillation, Crohn's disease, carotid artery stenosis, coronary artery disease, multiple sclerosis, systemic lupus erythematosus and rheumatoid arthritis	ICD9 codes	776	Four of the known SNP-disease associations were replicated and 19 new associations were identified
Polimanti et al. 2016	26,394	SNPs (8) in CHRNA3–CHRNA5 locus, ADH1B, and ALDH2	Large cohort database	360	Replicated findings that these SNPs are associated with drinking and smoking behaviors as well as novel findings that these SNPs were associated with psychological traits
Hebbring et al. 2015	4,235	SNPs (5) previously linked to multiple sclerosis, ankylosing spondylitis, triglyceride levels, atrial fibrillation and age-related macular degeneration	Clinical text data from EMRs	23,384	Replicated findings and demonstrated that raw text data can be used to define a phenome
Warner and Alterovitz 2012	36,095	White blood cell counts	ICD9 codes	5,675	Peak WBC counts between 15–45 K/μl were associated with *Clostridium difficile* and bacterial sepsis
Liao et al. 2013	1,290 cases 1,236 controls	Autoantibodies (anti-citrullinated protein antibodies, antinuclear antibodies, antitissue transglutaminase antibodies, antithyroid peroxidase)	ICD9 codes	512 in cases 698 in controls	In cases, the presence of antinuclear antibodies (ANA) was associated with Sjögren's/sicca syndrome In controls, higher ANA was associated with chronic nonalcoholic liver disease In both cases and controls, anti-thyroid peroxidase antibodies was associated with hypothyroidism

Table 2:

Health dimensions assessed in NHANES, corresponding measurements, and measurement method(s)[a]

Health Dimension	NHANES Measurements	Method Of Measurement
General Health		
	Current Health Status	Questionnaire
	Physician Exam	Physical Examination
	Medical Conditions	Questionnaire
Physical Health		
	Physical Activity and Physical Fitness	Questionnaire
	Physical Activity Monitor	Physical Examination
	Physical Functioning	Questionnaire
	Physical Functioning-Timed Walks	Physical Examination
	Isokinetic Knee Extensors Strength	Physical Examination
Mental Health		
	Attention Deficit Hyperactivity Disorder	Questionnaire
	Anxiety	Questionnaire
	Conduct Disorders	Questionnaire
	Depression	Questionnaire
	Eating Disorders	Questionnaire
	Elimination Disorders	Questionnaire
	Panic Disorder	Questionnaire
	Depression	Questionnaire
Cognitive Functioning		
	Cognitive Functioning	Examination
Body Composition/Bone Health		
	Anthropometry Measurements	Physical Examination
	Bioelectric Impedance Analysis	Physical Examination
	Dual Energy X-Ray Absorptiometry	Physical Examination
	Body Composition	Physical Examination
	Bone Density-Hip and Spine	Physical Examination
	Vertebral Fracture Assessment	Physical Examination
	Weight History	Questionnaire
	Osteoporosis	Questionnaire
	Bone Alkaline Phosphatase	Laboratory Measurement
	N-telopeptide (NTX)	Laboratory Measurement
Muscular Health		
	Grip Strength Test	Physical Examination
	Muscle Pain	Questionnaire
	Creatinine Kinase	Laboratory Measurement
	Creatinine Phosphokinase	Laboratory Measurement
	Creatinine	Laboratory Measurement
Dermatologic Health		

Health Dimension	NHANES Measurements	Method Of Measurement
	Dermatology	Questionnaire, Physical Examination
Ocular Health		
	Vision	Questionnaire, Physical Examination
	Retinal Photography	Physical Examination
	Visual Fields	Physical Examination
Oral Health		
	Oral Health	Questionnaire
	Dental Fluorosis Imaging	Physical Examination
Auditory Health		
	Audiometry	Physical Examination
	Hearing/Audiometry	Questionnaire
Respiratory Health		
	Respiratory Health and Disease	Questionnaire
	Exhaled Nitric Oxide	Laboratory Measurement
	Spirometry	Laboratory Measurement
Cardiovascular Health		
	Cardiovascular Disease	Questionnaire
	Cardiovascular Fitness	Physical Examination
	Blood Pressure	Questionnaire, Physical Examination
	Peripheral Vascular Disease	Physical Examination
	Fibrinogen	Laboratory Measurement
Thyroid Function		
	Thyroid Hormones	Laboratory Measurement
	Parathyroid Hormone	Laboratory Measurement
Gastrointestinal Health		
	Bowel Health	Questionnaire
	Celiac Disease	Laboratory Measurement
Renal and Urinary Health		
	Urology	Questionnaire
	Kidney Conditions	Questionnaire
	Urine Flow Rate Calculations	Laboratory Measurement
	Urine Osmolality	Laboratory Measurement
	Chemistry Panel	Laboratory Measurement
	Prostrate Conditions	Questionnaire
	Prostrate Health Specific Antigens	Laboratory Measurement
Hepatobiliary System		
	Chemistry Panel	Laboratory Measurement
	Albumin	Laboratory Measurement
	C-Reactive Protein	Laboratory Measurement
Reproductive Health and Gonadal Hormone Function		
	Reproductive Health	Questionnaire
	Pubertal Maturation	Questionnaire

Health Dimension	NHANES Measurements	Method Of Measurement
	Testosterone	Laboratory Measurement
	Sex Hormone Binding Globulin	Laboratory Measurement
	Follicle Stimulating Hormone	Laboratory Measurement
	Luteinizing Hormone	Laboratory Measurement
Immune System		
	Complete Blood Count	Laboratory Measurement
	White Blood Count	Laboratory Measurement
	Deoxyribonucleic Acid	Laboratory Measurement
Sleep		
	Sleep Disorders	Questionnaire
Arthritis		
	Arthritis Body Measures	Physical Examination
	Inflammatory Arthritis Pain	Questionnaire
	Arthritis Biomarkers	Laboratory Measurement
Glucose Metabolism/Diabetes		
	Diabetes	Questionnaire
	Oral Glucose Tolerance	Laboratory Measurement
	Glucose	Laboratory Measurement
	Insulin/C-peptide	Laboratory Measurement
	Glycohemoglobin	Laboratory Measurement
	Peripheral Neuropathy	Physical Examination
Dyslipidemia		
		Questionnaire, Laboratory
	Cholesterol	Measurement
	High Density Lipoprotein	Laboratory Measurement
	Low Density Lipoprotein	Laboratory Measurement
	Triglycerides	Laboratory Measurement
	Lipoprotein(a)	Laboratory Measurement
	Apolipoprotein	Laboratory Measurement
Allergy		
	Allergy	Questionnaire
	Immunoglobulin E-allergens	Laboratory Measurement
Nutritional Biomarkers		
	Erythrocyte Protoporphyrin	Laboratory Measurement
	Ferritin	Laboratory Measurement
	Total Iron Binding Capacity/Transferrin Saturation	Laboratory Measurement
	Transferrin Receptor	Laboratory Measurement
	Methylmalonic Acid	Laboratory Measurement
	Vitamin A	Laboratory Measurement
	Vitamin E	Laboratory Measurement
	Carotenoids	Laboratory Measurement

Health Dimension	NHANES Measurements	Method Of Measurement
	Vitamin B6	Laboratory Measurement
	Vitamin B12	Laboratory Measurement
	Vitamin C	Laboratory Measurement
	Vitamin D	Laboratory Measurement
Vestibular Function		
	Balance	Questionnaire, Physical Examination

[a]Some measurements are assessed using multiple modalities. For instance, diabetes can be assessed by questionnaire and laboratory-based measurements.