Air Pollution and Adverse Pregnancy and Birth Outcomes: Mediation Analysis Using Metabolomic Profiles

Kosuke Inoue[#1], **Qi Yan**[#1], **Onyebuchi A. Arah**[1,2,3], **Kimberly Paul**[1], **Douglas I. Walker**[4], **Dean P. Jones**[5,6], **Beate Ritz**[1,7,8]

[1]Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles (UCLA), 650 Charles E. Young Dr. South, Los Angeles, CA 90095, USA

[2]Department of Statistics, UCLA College of Letters and Science, Los Angeles, CA, USA

[3]Department of Public Health, Aarhus University, Aarhus, Denmark

[4]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[5]Clinical Biomarkers Laboratory, Division of Pulmonary, Allergy, and Critical Care Medicine, School of Medicine, Emory University, Atlanta, GA, USA

[6]Department of Medicine, Emory University, Atlanta, GA, USA

[7]Department of Environmental Health Sciences, UCLA Fielding School of Public Health, Los Angeles, CA, USA

[8]Department of Neurology, UCLA David Geffen School of Medicine, Los Angeles, CA, USA

[#] These authors contributed equally to this work.

Abstract

Purpose of Review—Review how to use metabolomic profiling in causal mediation analysis to assess epidemiological evidence for air pollution impacts on birth outcomes.

Recent Findings—Maternal exposures to air pollutants have been associated with pregnancy complications and adverse pregnancy and birth outcomes. Causal mediation analysis enables us to estimate direct and indirect effects on outcomes (i.e., effect decomposition), elucidating causal mechanisms or effect pathways. Maternal metabolites and metabolic pathways are perturbed by air pollution exposures may lead to adverse pregnancy and birth outcomes, thus they can be considered mediators in the causal pathways. Metabolomic markers have been used to explain the biological mechanisms linking air pollution and respiratory function, and of arsenic exposure and birth weight. However, mediation analysis of metabolomic markers has not been used to assess air pollution effects on adverse birth outcomes. In this article, we describe the assumptions and

applications of mediation analysis using metabolomic markers that elucidate the potential mechanisms of the effects of air pollution on adverse pregnancy and birth outcomes.

Summary—The hypothesis of mediation along specified pathways can be assessed within the structural causal modeling framework. For causal inferences, several assumptions that go beyond the data—including no uncontrolled confounding—need to be made to justify the effect decomposition. Nevertheless, studies that integrate metabolomic information in causal mediation analysis may greatly improve our understanding of the effects of ambient air pollution on adverse pregnancy and birth outcomes as they allow us to suggest and test hypotheses about underlying biological mechanisms in studies of pregnant women.

Keywords

Air pollution; Adverse birth outcomes; Metabolomics; Causal mediation analysis; 4-way decomposition

Introduction

Maternal exposures to air pollutants and their mixtures from various sources have been associated with pregnancy complications and adverse birth outcomes including preterm birth and low birth weight [1–4]. In addition, these complex exposures have also been shown to affect fetal development and neurodevelopment in offspring [4–7]. Since air pollution is widespread and increasing in certain countries, the substantial public health impacts of adverse birth outcomes and their long-term health sequelae will have to be addressed if these associations are deemed causal [8]. Determination of a causal relationship would have implications for the burden of disease measures and require strategies to mitigate these health effects of air pollution exposure. Experimental studies suggested that endocrine disruption, oxidative stress, inflammatory response, and DNA damage may be contributing to adverse pregnancy and birth outcomes associated with maternal air pollution exposures [9–11]. Yet, while little mechanistic insight has thus far been gained from studies of human pregnancy, if statistical tools combined with omics data to identify causal (biological) pathway from air pollution to reproductive health become feasible and available for epidemiologic studies, results based on these approaches may not only increase plausibility but also help establish causality.

In 2005, Christopher Wild proposed the concept of the exposome, which highlighted the need for comprehensive environmental exposure assessment tools to better understand causal links between complex exposures and disease [12]. The concept of the "exposome" has further evolved to not only encompass more sophisticated characterizations of the external environment through measurement and modeling but has been further defined to be the "cumulative measure of environmental influences and associated biological responses throughout the lifespan, including exposures from the environment, behavior, diet, and endogenous processes" [13]. The exposome definition that incorporates the biological responses to exposures was made possible through the rapid development of high-dimensional analytical platforms that provide researchers with opportunities to conduct omic-level biological response profiling of biofluids and tissues, facilitating an improved understanding of how the environment interacts with or disturbs human physiology, and

leaves biological signatures that ultimately predict human health imbalances and adverse outcomes [14••]. The metabolome includes all low molecular weight (< 2000 Da) chemical species present in biological matrices and represents a functional measure of interactions between the genome, diet, environment, and biochemical processes required for life [15]. Current high-resolution metabolomics (HRM) platforms have the capability to accurately measure more than 20,000 endogenous and exogenous chemical signals [14, 16]; of which the alterations in endogenous metabolites can be assembled into biological pathways and tested for their association with exposures or disease. Thus, metabolomics is beginning to be adopted more frequently in epidemiologic research to identify exposure signatures, disease risk traits, and biological pathways linking them.

From the epidemiological perspective, one of the major challenges of an "exposome" study is assigning causality to an exposure mixture-disease association that would eventually allow us to identify or design prevention approaches and modulate specific biological pathways found to be affected. For example, linking air pollution to adverse pregnancy and birth outcomes requires identifying intermediate metabolites that belong to a causal pathway between air pollution and adverse pregnancy and birth outcomes: "*Meet-in-the-middle*" (MITM) is one of the approaches to effectively identify such intermediate metabolites related to the exposure and the outcome [17]. Causal mediation analysis or effect decomposition, first popularized by Baron and Kenny [18], is a widely used epidemiological method to assess the role of intermediate variables or "mediators" that lie along the causal path(s) from the exposure (mixtures) to the outcome. Substantial advances in this method have been made by allowing for nonlinearity and interactions [19–22], as well as multiple mediators with path-specific effects (i.e., characterizing the mediation effect of a pathway via mediators) [23–25]. For ethical and feasibility reasons, in many cases, we cannot conduct large clinical trials in pregnant women in which we would experimentally vary exposures to potentially toxic air pollutants. The application of mediation analysis to inform this topic is a reasonable approach allowing for causal interpretations from observational studies. Furthermore, finding causal (biological) mediators might offer intervention targets (such as diets or medications) when it is difficult to intervene on the primary exposure in a timely manner. This is important when it is not easy to develop feasible interventions or in order to choose between alternative interventions as we develop strategies for mitigating risks for exposed mothers and infants. To date, mediation analysis has rarely been applied in environmental epidemiology studies that link air pollution and adverse pregnancy and birth outcomes, especially for mediators that represent metabolic changes.

In this article, we will first briefly discuss the current state of epidemiological findings for air pollution and adverse pregnancy and birth outcomes. Then, we will describe causal mediation analysis in general, followed by a discussion of how metabolomics can be used to generate an "omics-" type mediator in pregnancy outcome research. The integration of causal inference into an omics approach may facilitate the discovery of causal biomarker pathways and improve our knowledge of the pathophysiology involved in air pollution–related adverse pregnancy outcomes. Finally, we also discuss the potential limitations and challenges of implementing mediation analysis in general and also specifically with "omics" data.

Air Pollution and Adverse Pregnancy and Birth Outcomes: Implementing Mediation Analysis

Why Is This Topic Important?

Over the past two decades, numerous epidemiological studies have reported associations for maternal exposure to air pollutants including nitrogen oxides (NO_2, NOx), particulate matter (PM_{10}, $PM_{2.5}$), carbon monoxide (CO), and ozone (O_3), and pregnancy complications such as preeclampsia, hypertension, and gestational diabetes and adverse birth outcomes including preterm birth and low birth weight [1–3, 26–31]. The literature has been strengthened considerably by data-rich birth cohort studies such as studies that pooled data across 14 birth cohorts from 12 European countries showing that associations of $PM_{2.5}$ with term low birth weight did not change after adjustment for many potential confounding variables [5, 6]. Even though the overall evidence for adverse impacts of air pollution on pregnancy is now substantial, the heterogeneity of results still fuels debates and nurtures doubts about causality. This is the case even though inconsistencies may be attributable to differences in exposure levels, pollution sources (from traffic to coal burning), and their toxic components and mixtures, or differences due to vulnerable gestational windows for outcomes. For example, one birth cohort study located in a low exposure area—the Norwegian MOBA study [32]—did not observe positive associations overall between modeled traffic-related air pollution (NO_2) and preeclampsia (OR = 0.89; 95% CI 0.74–1.08); however, for women living in Oslo, the city with the highest exposures, an association was suggested (OR = 1.27; 95% CI 0.83–1.95). A three times larger study conducted in Southern Sweden, an area with similar air pollution levels as in Oslo, estimated relatively strong increases in preeclampsia (OR = 1.51; 95% CI 1.32–1.73) for the highest quartile of NO_x derived from a complex multiple source emissions model, and weaker effects for a simple measure of high traffic density within 200 m of a home (OR = 1.10; 95% CI 0.94–1.30) [30]. The same Swedish study also reported an increased risk for gestational diabetes both with the emissions model for NOx (OR = 1.69; 95% CI 1.41–2.03) and their simple traffic density measure (OR = 1.23; 95% CI 1.01–1.51). Meanwhile, an earlier study from the Netherlands did not find an association between residential proximity to traffic and preeclampsia (OR = 1.14; 95% CI 0.71–1.82) or gestational diabetes (OR = 0.79; 95% CI 0.35–1.81) but was only a tenth of the size of the Swedish study with less than half the gestational diabetes prevalence [33]. Thus, differences in air pollution levels and mixtures as well as differences in air pollution measurement approaches may contribute to inconsistencies in results. Also, weak exposure contrasts may require unrealistically large sample sizes or a very high disease prevalence to allow for the reliable estimation of air pollution effects in epidemiologic studies. Here, we suggest that some of these limitations preventing us from causally interpreting results from observational epidemiology studies can be overcome by using approaches that allow us to gain insights into biological pathways affected by air pollutants in human pregnancy.

Causal Mediation Analysis: a Causal Inference Tool in Environmental Epidemiology

Applying causal inference methods in environmental epidemiology is important for advancing the field as it is often challenging to generate experimental data in pregnant

women for environmental toxicants such as air pollution or many chemicals. Also, for pregnancy-related disorders, laboratory (animal or cell) model systems are rather limited in terms of generalizability to human conditions. Causal mediation analysis is one of the causal inference methods that allow us to decompose the total effect into indirect and direct effects. Estimating mediated effects through biomarkers for a system's response to exposure has received great interest in environmental health as these markers may reveal underlying biological mechanisms. These markers establish biological plausibility if they are indicative of alterations to endogenous pathways related to the health outcome under investigation and are associated with the studied exposure. In fact, they may represent biologically relevant mechanisms linking air pollution and adverse pregnancy and birth outcomes. However, the simple comparison of the models with and without adjusting for such intermediate biomarkers may not provide either an indirect or direct effect particularly due to the potential "collider-stratification" bias when we fail to fully adjust for confounders between the intermediate variables and outcomes [34, 35]. Thus, a growing literature in environmental epidemiology has used modern mediation analysis approaches that allow us to investigate causal effect mechanisms in terms of mediation and interaction [36, 37].

How Causal Mediation Analysis Works

Modern causal mediation analysis has been formalized under the structural causal model (SCM) and the potential outcomes framework to allow for identification and estimation of the indirect and direct effects of an exposure on an outcome [36]. The indirect effect refers to the effect of the exposure on the outcome that is mediated by the specified mediator. The direct effect refers to the effect of the exposure on the outcome through all other unspecified mechanisms. We can also estimate the proportion mediated—the proportion of the total effect that is due to the effect through a specific mediator on the causal path from the exposure to the outcome—based on the estimated indirect and total effect sizes. Here, we used directed acyclic graphs (DAGs)—a graphical tool to represent causal relationships among variables by linking them through arrows [38, 39]—to briefly introduce causal mediation analysis. In Fig. 1, we let X denote the exposure, Y denote the outcome, M denote the mediator, and C denote a set of confounders. Each arrow represents each causal effect of one variable on another. For example, the arrow X➔M➔Y represents the indirect effect of X on Y mediated through M, and the arrow X➔Y (not through M) represents the direct effect of X on Y not mediated through M.

Within the causal mediation analysis framework, there are several approaches to decompose the total effect including natural decomposition (where the mediator is set to the individual-specific value that it would attain under a specific exposure intervention), controlled decomposition (where the mediator is set to a fixed value for all individuals in the study population), and mixed decomposition (notably, the 4-way decomposition). Mediation analysis supports 2-, 3- or 4-way decomposition of the total effect into 2, 3, or 4 components, respectively. The 2-way decomposition is commonly used and yields the natural direct and indirect effects [36, 40]. A 4-way decomposition goes further and estimates (A) the effect of X on Y due to neither mediation by M nor interaction between X and M (controlled direct effect), (B) the effect of X on Y due to interaction between X and M only (reference interaction effect), (C) the effect of X on Y due to both mediation by M

and interaction between X and M (mediated interaction effect), and (D) the effect of X on Y due to mediation by M only (pure indirect effect). Figure 2 represents each effect using DAGs. In these DAGs, we let X × M denote the interaction between X and M. The 4-way decomposition approach is general in the sense that certain combinations of these four components correspond to other decompositions [41, 42]. For example, we can estimate the proportion attributable to interaction by summing the reference interaction effect and the mediated interaction effect. Further examples can be seen in Table 1 and elsewhere [41, 42]. Counterfactual expressions, empirical analogues, and each step of how to estimate these effects can be found in [41•]. Examples of SAS and SPSS macros [43], R packages [44, 45], and Stata command [46] that allow for exposure-mediator interactions are also available.

To causally interpret the estimated indirect and direct effects, causal mediation analysis requires rather strong assumptions [36••]. One of these assumptions is conditional exchangeability: (i) the causal path from the exposure (X) to the mediator (M) is not confounded given the covariates (C) controlled for; (ii) the causal path from the exposure (X) to the outcome (Y) is not confounded given the covariates controlled for; (iii) the causal path from the mediator (M) to the outcome (Y) is not confounded given the covariates (C); and (iv) there should not be a mediator-outcome (M-Y) confounder that is affected by the exposure (X). We need assumptions (iii), and (iv) even when using randomized controlled trials (RCT) because the RCT only assures no confounding between the exposure (X) and the outcome (Y), and between the exposure (X) and the mediator (M), i.e., the mediator (M) is not randomized. Bias analysis approaches have been proposed in the context of causal mediation analysis to assess the robustness of the findings to the assumption of unmeasured confounding [48, 49]. Other causal modeling assumptions needed are positivity, consistency, no model misspecification, and no other sources of bias.

Several approaches have been developed to conduct mediation analysis such as a regression-based approach [21, 42], a weighting-based approach [50], a g-computation approach [41, 51], a simulation-based approach [52], and a generalized mixed-effect model approach [53]. These models are flexible and allow us to conduct mediation analysis for complex situations including time-to-event outcomes, longitudinal data, and multiple mediators [23–25]. More details about each approach including counterfactual interpretations for each estimated effect can be found in cited papers and/or methodological reviews [36, 54].

Previous Examples Using Mediation Analysis for the Association Between Air Pollution and Adverse Pregnancy and Birth Outcomes

Very few studies to date have applied mediation analysis to answer research questions related to air pollution and reproductive health. One of these is a European study that examined whether placental mitochondrial DNA (mtDNA) changes implicate biological pathways leading from prenatal ambient nitrogen dioxide (NO_2) exposure to infant birth weight [55]. Using a regression-based mediation formula, the researchers found that changes in placental mtDNA content mediated 10% of the association between prenatal NO_2 exposure and birth weight and these observations suggested mitochondrial damage of placental tissue as one biological mechanism for adverse effects of early pregnancy exposure to air pollution on birth weight. Another example is a study that examined maternal thyroid

hormone levels (free thyroxine [fT4]) as mediators for inverse associations seen between maternal exposure to air pollution (particulate matter [$PM_{2.5}$], black carbon [BC], and ammonium [NH_4^+]) and birth weight [56]. Using a regression-based mediation approach, this study found that maternal fT4 levels explained 15–20% of the estimated effects of these air pollutants (i.e., $PM_{2.5}$, BC, NH_4^+) on birth weight Z-score. If this is confirmed in additional studies and since maternal thyroid function is clinically manageable [57], the findings from this mediation analysis not only revealed a potential underlying biological mechanism, but will have public health implication if it is possible to improve birth weight by intervening on maternal thyroid function among highly exposed mothers for whom a reduction of air pollution exposures—the intervention of choice—is not possible.

Air Pollution and Adverse Pregnancy and Birth Outcomes Using Mediation Analysis: Metabolome as the Mediator

Utilizing Metabolites as Mediators

HRM is an ideal tool to study environmental exposures and biological responses to these exposures in human tissues. With untargeted approaches using high-resolution mass spectrometry, it is possible to greatly expand coverage of the metabolome (from several hundred to several thousand chemical signals) and increase our ability to gain insights not only based on a few well-studied metabolites, but to integrate across many metabolites and biological processes. This presents us with the opportunity to more robustly identify general alterations of metabolic pathways that are implicated by changes in multiple metabolites.

The metabolome provides a central measure linking exposure to internal dose, biological response, and disease patho-biology. Specifically, if ambient air pollution perturbs exogenous and endogenous metabolites, the dysregulation identified for specific metabolic pathways may provide hints about disease etiology. The metabolome response patterns can also be considered intermediate factors, and it now becomes important to integrate causal inference method in the analyses in order to quantify the amount of variation in birth outcomes that are explained by ambient air pollution through pathways implicated by metabolomic disturbances. Specifically, using mediation analysis (i.e., by treating metabolites measured on metabolomic platforms as mediators), we can delineate quantitatively elements of the metabolome that appear to mediate how ambient air pollution affects adverse birth outcomes.

In the next section, we review current studies of ambient air pollution that used untargeted metabolomics approaches, with a focus on the pregnancy exposome. Then, we discuss potential metabolites/metabolic pathways that lie between air pollution exposure and adverse birth outcomes. Finally, we propose potential applications for mediation analysis on this topic.

Air Pollution Exposure During Pregnancy and Metabolomics Related to Adverse Pregnancy and Birth Outcomes

The pregnancy period is especially vulnerable to the impact of environmental exposures as the fetus not only develops and grows rapidly but organs change in terms of susceptibility to

stressors and their metabolic maturity [58]. Special efforts have recently been made to advance the role of the exposome in pregnancy and early life. For example, in Europe, the Human Early-Life Exposome (HELIX) project provides exposure and biological response data for more than 1000 mother-child pairs from 6 European birth cohorts [59]. In the USA, the Children's Health Exposure Analysis Resource now supports collaboration between laboratories with different analytic capabilities for assessing biological responses to exposure in pregnancy and early childhood samples available from different cohorts.

However, less than a handful of metabolomic studies focused on air pollution during pregnancy and adverse birth outcomes. One study targeted 37 oxylipins reflecting the cyclooxygenase (COX), lipoxygenase (5-LOX and 12/15-LOX), and cytochrome P450 (CYP) pathways in 197 cord blood samples, and found that exposure to particulate matter in-utero was associated with lipoxygenase pathways indicative of an inflammatory response to exposure in the newborn [60].

Our group recently conducted one of the first untargeted metabolomics studies of traffic-related air pollution exposure during pregnancy [61•]. We retrieved stored mid-pregnancy serum samples from 160 mothers who lived in the Central Valley of California known for its high particulate levels, and estimated traffic-related air pollution exposure (carbon monoxide, nitric oxides, and particulate matter less than2.5 μm) during the first trimester using the CALINE4 dispersion model. Using liquid chromatography high-resolution mass spectrometry to obtain untargeted metabolic profiles, we conducted feature selection and pathway analyses to identify biological pathways related to air pollution exposure. Maternal exposure to traffic-related air pollution affected oxidative stress and inflammation-related pathways, including linoleate, leukotriene, and prostaglandin pathways, and these results are further supported by in vitro and in vivo studies [62–66].

Our study also supports findings observed in other metabolomic studies of air pollution exposure. In a randomized, crossover trial that enrolled 55 healthy college students and used the same untargeted metabolomics approach, short-term (9 days) exposure to particulate matter less than 2.5 μm ($PM_{2.5}$) induced metabolic changes in serum associated with stress hormone levels, insulin resistance, and markers of oxidative stress and inflammation [67]. Two other experimental studies using a crossover design identified a variety of metabolic features associated with short-term traffic-related air pollution exposures including various acyl-carnitines, which are metabolites from fatty acid oxidation [68, 69]. Recently, a study measured traffic-related pollutants at multiple ambient and indoor sites at varying distances from a major highway artery for 12 weeks and collected plasma and saliva samples from 54 healthy college students living near (20 m) or far (1.4 km) from a highway. This research also identified metabolites and pathways associated with oxidative stress and inflammation, as well as with nucleic acid damage and repair [70].

Fewer studies focus on long-term air pollution exposure. A cross-sectional study conducted within the TwinsUK cohort found that oxidative stress and inflammation-related metabolites such as α-tocopherol, benzoate, and glycine were associated with both long-term air pollution (annual average $PM_{2.5}$ concentration) and lung function [71]. Another cross-sectional study with 59 healthy participants linked year-long ultrafine particle exposure that

was adjusted for time and activity with metabolic variations related to antioxidant pathways and endothelial function [72]. In addition, untargeted metabolomics approaches were used in two case-control studies nested within cohorts to apply the MITM concept; this study also implicated several oxidative stress and inflammation-related pathways mediating the effect of air pollution on two health outcomes, i.e., adult-onset asthma and cardio-cerebrovascular diseases [73].

It is important to note that oxidative stress in pregnancy may cause damage to all major cellular elements and affect the placenta and its function, therefore contributing to adverse birth outcomes such as spontaneous abortion, preeclampsia, intrauterine growth restriction, low birth weight, and preterm delivery [74–77]. Previous studies also have shown associations between elevated circulating levels of inflammatory markers and preterm birth and low birth weight [78–80]. Given the contribution of oxidative stress and inflammation-related pathways to adverse birth outcomes and their central role with air pollution exposure, these pathways are logical mediators (biomarkers) for the adverse effect of ambient air pollution on birth outcomes. Thus, it would be valuable to conduct a formal mediation analysis to quantify the contributions of these pathways for different outcomes.

Current Literature

A literature search using the electronic database MEDLINE (through November 30, 2019) did not identify any study that investigated air pollution and adverse birth outcomes as being mediated through individual metabolites or metabolic pathway-level alterations. This approach has however been introduced for other health outcomes. For example, a Dutch study focused on short-term ambient air pollution exposure changes (e.g., PM, NO_2, NO_x, O_3) by exposing 31 volunteers to ambient air pollution at five different locations for 5 h and collecting blood for untargeted metabolic profiling before and after exposure (493 blood samples and 3873 metabolic features) [69]. They employed causal mediation analyses for the effects of air pollutants on respiratory health outcomes (e.g., forced vital capacity (FVC), forced expiratory volume in 1s (FEV_1)) through a change in metabolic features that were associated with both the exposure and the outcome. This study serves as a proof of principle for the feasibility of conducting formal mediation analysis with complex air pollution exposures and untargeted metabolomics data even with a relatively small sample size. Despite the challenging issues of mediation analysis including the above-mentioned strong assumptions and statistical power issues, we believe that this causal inference tool may be useful in metabolomic studies to disentangle causal pathways from ambient air pollution to adverse pregnancy and birth outcomes. In the next section, we comment on potential challenges.

Implications for Future Studies and Challenges

Given the above-mentioned potential for assessing causal mechanisms between air pollution, metabolomics, and adverse birth outcomes, more research is needed that uses mediation analysis and metabolomics profiling. Figure 3 shows examples of causal diagrams for mediation analysis using metabolomic markers/profiles as a single or multiple mediator(s). Figure 3a is a simple setting that defines a single metabolomic pathway as a mediator on the causal pathway between air pollution and adverse birth outcome. However, this simple

situation may not be appropriate given the multi-factorial interaction between metabolites. Figure 3b represents a multiple mediator setting that requires more advanced methods (e.g., ordered or sequential mediation analysis) to identify mediation effects across multiple metabolic pathways. Figure 3c represents an alternate multiple mediator setting that does not require us to assume an order for mediating pathways. One of the approaches to estimate such a combined mediation effect is to consider multiple mediators (i.e., pathway 1, 2, … *k*) as a single set of mediators and estimate an indirect effect through at least one of the mediators without having to disentangling them. More details and discussions of how to use multiple mediators are described elsewhere [23].

One of the hurdles of using mediation analysis with omics data is the necessity of dealing with high-dimensional data. For example, some but not all metabolites included as potential mediators may be associated with the exposure and/or the outcome (after adjusting for the exposure). To overcome this problem, variable pre-selection may be needed to identify metabolites that do best represent a pathway using statistical approaches such as "sure independence screening" [81] and marginal association tests based on false discovery rate [82] in addition to prior knowledge about metabolites and pathways. Another solution is a transformation model using spectral decomposition [83], where mediators are transformed based on a principal component analysis and become uncor-related given exposure, and we can estimate mediation effects using a series of low-dimensional models. Further details on strategies for addressing these issues can be found elsewhere [83, 84].

In addition to the high-dimensionality problem, we need to consider the following issues when applying mediation analysis to address this research topic:

Confounding: As described above, we need fairly strong exchangeability assumptions for mediation analysis, i.e., the exposure-outcome, exposure-mediator, and mediator-outcome associations should not be confounded by unmeasured variables [36••]. To overcome this limitation and make causal claims, sensitivity analyses to assess the potential strengths of such a bias are recommended [22]. Moreover, as multiple pathways may contribute to the effects of air pollution on adverse birth outcomes, the assumption that no mediator-outcome confounder is being affected by exposure needs to be carefully considered.

Measurement error: Personal air pollution exposure measures are—on the surface—the most accurate exposure measures. However, collecting a personal measure is neither feasible in large populations nor over long periods of time and they are not necessarily the exposure of interest, in contrast to ambient air pollution and its sources that are subject to regulation. Ambient air pollution exposure is often estimated via complex spatial modeling based on emissions, land use, or monitoring data. This exposure assessment may be affected by measurement error due to the spatial misalignment of the monitoring data [85, 86]. On the other hand, ambient exposure measures may not suffer from confounding by personal risk factors and issues of reverse causation [87]. Measurement error may also affect the mediators (i.e., metabolomics), as their measured levels may depend on dynamic reactivity, structural diversity, and the timing of sample collection [88], such that it can be challenging to measure intensities accurately or annotate metabolites correctly. Careful sample preparation (quenching, extraction, etc.) is needed to reduce the inter-conversion among

metabolites. Metabolic profiles measured by LC-MS are also subject to limitations of the analytical platform in terms of limits of detection, dynamic ranges, and reproducibility [88]. However, these measurements are becoming more and more standardized and the use of repeated measurements and authentic standards can mitigate these measurement errors.

Temporality: In mediation analysis employed to infer causation, it is critical to establish temporality well such that the mediator (metabolomics measurement) occurs after exposure (air pollution), and before outcome (adverse birth outcomes). Timing can be established in longitudinal cohorts where the biosample extraction clearly precedes the diagnosis of a disease. However, for rare outcomes such as gestational diabetes or birth defects, only case-control data may be available, which makes it much more challenging to resolve the temporal ordering from exposure to metabolite to outcome. One possible solution is using stored biological samples that were collected early enough in pregnancy and prior to the development of the disease in a nested case-control study design [89].

Time window: Specific to the research question, there may be a cumulative effect of air pollution exposure and it might be challenging to identify the most relevant timing or period of exposure for the outcome of interest. Biological knowledge may help to decide on the critical window of exposure. Furthermore, the blood metabolome is dynamic and some of its pathways are influenced by factors such as medications, exercise, or diet [90]. Thus, if these pathways are of interest, it is critical to time and standardize the collection of biosamples for all study participants as much as possible.

Statistical power: To perform mediation analysis and obtain a valid effect estimate, we need a sufficiently large sample size. As with statistical power in general, the magnitude of the estimated effects in each pathway (i.e., the pathway from the exposure to the mediator(s) and the pathway from the mediator(s) to the outcome) is also important and may determine whether we are able to identify the mediation effect. Although previous studies have provided some first insights into issues of statistical power and sample size for mediation analysis [91, 92], power is hard to determine particularly when multiple mediators and/or interactions between exposure and mediator(s) need to be considered. Furthermore, determining the sample size for metabolomic studies is particularly challenging due to the high-dimensionality and multicollinearity of the variables. The number of measured metabolites depends on the analytical platform and is often unknown a priori without preliminary data. In addition, multiple comparison correction is required with tens of thousands of metabolic features being tested [93]. In summary, the power calculations for a mediation study that involves metabolic profiling depend on multiple factors including the analytical platform and choice of exposure and outcome, and there is still no standard method available.

Conclusion

Despite the substantial advances in formulating principles of mediation analysis and a growing number of applications of such methods in environmental epidemiology over the last decade, investigations into the causal relationship between ambient air pollution and adverse birth outcomes mediated through biomarkers are currently rare. Studies that

integrate metabolomic information in causal mediation analysis and carefully weigh assumptions may greatly improve our understanding of the effects of ambient air pollution on adverse pregnancy and birth outcomes. They may also allow us to suggest and test hypotheses about underlying biological mechanisms in studies of pregnant women.

Funding Information

References

Papers of particular interest, published recently, have been highlighted as:

• Of importance

•• Of major importance

1•. Stieb DM, Chen L, Eshoul M, Judek S. Ambient air pollution, birth weight and preterm birth: a systematic review and meta-analysis. Environ Res. 2012;117:100–11. 10.1016/j.envres.2012.05.007 [PubMed: 22726801] Comprehensive qualitative literature review and meta-analysis of ambient air pollution and pregnancy outcomes (birth weight and preterm birth).

2. Pedersen M, Giorgis-Allemand L, Bernard C, Aguilera I, Andersen AM, Ballester F, et al. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). Lancet Respir Med. 2013;1(9): 695–704. 10.1016/S2213-2600(13)70192-9. [PubMed: 24429273]

3. Lee PC, Talbott EO, Roberts JM, Catov JM, Sharma RK, Ritz B. Particulate air pollution exposure and C-reactive protein during early pregnancy. Epidemiology. 2011;22(4):524–31. 10.1097/EDE.0b013e31821c6c58. [PubMed: 21516040]

4. Glinianaia SV, Rankin J, Bell R, Pless-Mulloli T, Howel D. Particulate air pollution and fetal health: a systematic review of the epidemiologic evidence. Epidemiology. 2004;15(1):36–45. 10.1097/01.ede.0000101023.41844.ac. [PubMed: 14712145]

5. Guxens M, Garcia-Esteban R, Giorgis-Allemand L, Forns J, Badaloni C, Ballester F, et al. Air pollution during pregnancy and childhood cognitive and psychomotor development: six European birth cohorts. Epidemiology. 2014;25(5):636–47. 10.1097/EDE.0000000000000133. [PubMed: 25036432]

6. Becerra TA, Wilhelm M, Olsen J, Cockburn M, Ritz B. Ambient air pollution and autism in Los Angeles county, California. Environ Health Perspect. 2013;121(3):380–6. 10.1289/ehp.1205827. [PubMed: 23249813]

7. Suades-Gonzalez E, Gascon M, Guxens M, Sunyer J. Air pollution and neuropsychological development: a review of the latest evidence. Endocrinology. 2015;156(10):3473–82. 10.1210/en.2015-1403. [PubMed: 26241071]

8. Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A cohort study of traffic-related air pollution impacts on birth outcomes. Environ Health Perspect. 2008;116(5):680–6. 10.1289/ehp.10952. [PubMed: 18470315]

9. Hougaard KS, Jensen KA, Nordly P, Taxvig C, Vogel U, Saber AT, et al. Effects of prenatal exposure to diesel exhaust particles on postnatal development, behavior, genotoxicity and inflammation in mice. Part Fibre Toxicol. 2008;5:Artn 3 10.1186/1743-8977-5-3. [PubMed: 18394195]

10. Kelly FJ. Oxidative stress: its role in air pollution and adverse health effects. Occup Environ Med. 2003;60(8):612–6. [PubMed: 12883027]

11. Risom L, Moller P, Loft S. Oxidative stress-induced DNA damage by particulate air pollution. Mutat Res. 2005;592(1–2):119–37. 10.1016/j.mrfmmm.2005.06.012. [PubMed: 16085126]

12. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomark Prev. 2005;14(8):1847–50. 10.1158/1055-9965.EPI-05-0456.

13. Miller GW, Jones DP. The nature of nurture: refining the definition of the exposome. Toxicol Sci. 2014;137(1):1–2. 10.1093/toxsci/kft251. [PubMed: 24213143]

14••. Jones DP. Sequencing the exposome: a call to action. Toxicol Rep. 2016;3:29–45. 10.1016/j.toxrep.2015.11.009 [PubMed: 26722641] This review summarizes the current state of high-resolution metabolomics as a method for generating biological responses from environmental exposures.

15. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0–the human metabolome database in 2013. Nucleic Acids Res. 2013;41(Database issue):D801–7. 10.1093/nar/gks1065. [PubMed: 23161693]

16. Niedzwiecki MM, Walker DI, Vermeulen R, Chadeau-Hyam M, Jones DP, Miller GW. The exposome: molecules to populations. Annu Rev Pharmacol Toxicol. 2019;59:107–27. 10.1146/annurev-pharmtox-010818-021315. [PubMed: 30095351]

17. Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. Cancer Epidemiol Biomark Prev. 2007;16(10):1954–65. 10.1158/1055-9965.EPI-07-0457.

18. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51(6):1173–82. 10.1037/0022-3514.51.6.1173. [PubMed: 3806354]

19••. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992;3(2):143–55. 10.1097/00001648-199203000-00013 [PubMed: 1576220] One of the first papers to broadly introduce causal mediation analysis using the counterfactual approach in epidemiology.

20. Pearl J Direct and indirect effects. 2001.

21. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Am J Epidemiol. 2010;172(12):1339–48. 10.1093/aje/kwq332. [PubMed: 21036955]

22. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. Stat Sci. 2010;25:51–71.

23. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics. 2015;71(1):1–14. 10.1111/biom.12248. [PubMed: 25351114]

24. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiol Methods. 2014;2(1):95–115. [PubMed: 25580377]

25. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. Epidemiology. 2017;28(2):258–65. 10.1097/EDE.0000000000000596. [PubMed: 27922534]

26. Dadvand P, Figueras F, Basagana X, Beelen R, Martinez D, Cirach M, et al. Ambient air pollution and preeclampsia: a spatiotemporal analysis. Environ Health Perspect. 2013;121(11–12):1365–71. 10.1289/ehp.1206430. [PubMed: 24021707]

27. Freeze HH, Eklund EA, Ng BG, Patterson MC. Neurological aspects of human glycosylation disorders. Annu Rev Neurosci. 2015;38:105–25. 10.1146/annurev-neuro-071714-034019. [PubMed: 25840006]

28. Pedersen M, Stayner L, Slama R, Sorensen M, Figueras F, Nieuwenhuijsen MJ, et al. Ambient air pollution and pregnancy-induced hypertensive disorders: a systematic review and meta-analysis. Hypertension. 2014;64(3):494–500. 10.1161/HYPERTENSIONAHA.114.03545. [PubMed: 24935943]

29. Lee PC, Talbott EO, Roberts JM, Catov JM, Bilonick RA, Stone RA, et al. Ambient air pollution exposure and blood pressure changes during pregnancy. Environ Res. 2012;117:46–53. 10.1016/j.envres.2012.05.011. [PubMed: 22835955]

30. Malmqvist E, Jakobsson K, Tinnerberg H, Rignell-Hydbom A, Rylander L. Gestational diabetes and preeclampsia in association with air pollution at levels below current air quality guidelines. Environ Health Perspect. 2013;121(4):488–93. 10.1289/ehp.1205736. [PubMed: 23563048]

31. Jo H, Eckel SP, Chen JC, Cockburn M, Martinez MP, Chow T, et al. Associations of gestational diabetes mellitus with residential air pollution exposure in a large Southern California pregnancy cohort. Environ Int. 2019;130:104933 10.1016/j.envint.2019.104933. [PubMed: 31234004]

32. Madsen C, Haberg SE, Aamodt G, Stigum H, Magnus P, London SJ, et al. Preeclampsia and hypertension during pregnancy in areas with relatively low levels of traffic air pollution. Matern Child Health J. 2018;22(4):512–9. 10.1007/s10995-017-2417-6. [PubMed: 29285630]

33. van den Hooven EH, Jaddoe VW, de Kluizenaar Y, Hofman A, Mackenbach JP, Steegers EA, et al. Residential traffic exposure and pregnancy-related outcomes: a prospective birth cohort study. Environ Health. 2009;8:59 10.1186/1476-069X-8-59. [PubMed: 20028508]

34. Greenland S Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003;14(3): 300–6. [PubMed: 12859030]

35. Preston SH, Stokes A. Obesity paradox: conditioning on disease enhances biases in estimating the mortality risks of obesity. Epidemiology (Cambridge, Mass). 2014;25(3):–454.

36. Pearl J, editor. Direct and indirect effects Proceedings of the seventeenth conference on uncertainty in artificial intelligence; 2001: Morgan Kaufmann Publishers Inc.

37••. VanderWeele T Explanation in causal inference: methods for mediation and interaction. Oxford: Oxford University Press; 2015.Comprehensive text explaining the current state of knowledge about causal mediation analysis in epidemiology.

38. Pearl J Causal diagrams for empirical research. Biometrika. 1995;82(4):669–88. 10.1093/biomet/82.4.669.

39••. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37–48 [PubMed: 9888278] General introduction of directed acyclic graph (DAG) in epidemiology.

40. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013;42(5):1511–9. 10.1093/ije/dyt127. [PubMed: 24019424]

41•. Wang A, Arah OA. G-computation demonstration in causal mediation analysis. Eur J Epidemiol. 2015;30(10):1119–27 [PubMed: 26537707] Detailed step-by-step description of 4-way decomposition in causal mediation analysis using the g-computation algorithm.

42. VanderWeele TJ. A unification of mediation and interaction: a 4-way decomposition. Epidemiology (Cambridge, Mass). 2014;25(5):749–61. 10.1097/EDE.0000000000000121.

43. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013;18(2):137–50. 10.1037/a0031034. [PubMed: 23379553]

44. Steen J, Loeys T, Moerkerke B, Vansteelandt S. Medflex: an R package for flexible mediation analysis using natural effect models. J Stat Softw. 2017;76(11).

45. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. J Stat Softw. 2014;59(5).

46. Discacciati A, Bellavia A, Lee JJ, Mazumdar M, Valeri L. Med4way: a Stata command to investigate mediating and interactive mechanisms using the four-way effect decomposition. Oxford University Press; 2019.

47. Arah OA. Augmenting causal diagrams with effect modification, interaction and other parametric unformation. 48th Annual Meeting of the Society for Epidemiologic Research (SER) Denver, CO. 2015;2015: 283

48. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology. 2010;21(4):540–51. 10.1097/EDE.0b013e3181df191c. [PubMed: 20479643]

49. Bohnke JR. Explanation in causal inference: methods for mediation and interaction. Q J Exp Psychol (Hove). 2016;69(6):1243–4. 10.1080/17470218.2015.1115884. [PubMed: 28190383]

50. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol. 2012;176(3):190–5. [PubMed: 22781427]

51. Daniel RM, De Stavola BL, Cousens SN. gformula: estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. Stata J. 2011;11(4):479–517.

52. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15(4):309–34. 10.1037/a0020761. [PubMed: 20954780]
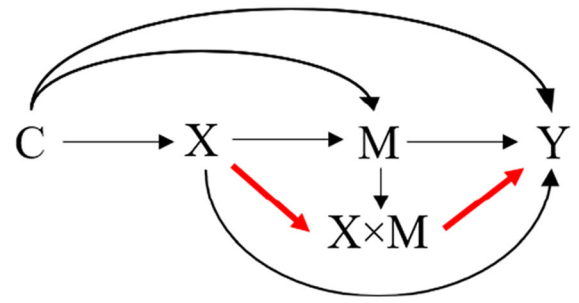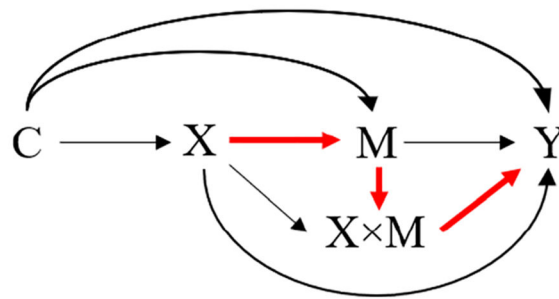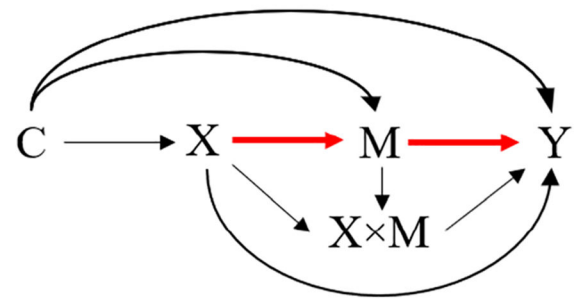
53. Bind MA, Vanderweele TJ, Coull BA, Schwartz JD. Causal mediation analysis for longitudinal data with exogenous exposure. Biostatistics. 2016;17(1):122–34. 10.1093/biostatistics/kxv029. [PubMed: 26272993]

54. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health. 2016;37:17–32. [PubMed: 26653405]

55. Clemente DB, Casas M, Vilahur N, Begiristain H, Bustamante M, Carsin AE, et al. Prenatal ambient air pollution, placental mitochondrial DNA content, and birth weight in the INMA (Spain) and ENVIRONAGE (Belgium) birth cohorts. Environ Health Perspect. 2016;124(5):659–65. 10.1289/ehp.1408981. [PubMed: 26317635]

56. Wang X, Liu C, Zhang M, Han Y, Aase H, Villanger GD, et al. Evaluation of maternal exposure to PM2.5 and its components on maternal and neonatal thyroid function and birth weight: a cohort study. Thyroid. 2019;29(8):1147–57. 10.1089/thy.2018.0780. [PubMed: 31298631]

57. Korevaar TIM, Medici M, Visser TJ, Peeters RP. Thyroid disease in pregnancy: new insights in diagnosis and clinical management. Nat Rev Endocrinol. 2017;13(10):610–22. 10.1038/nrendo.2017.93. [PubMed: 28776582]

58. Robinson O, Vrijheid M. The pregnancy exposome. Curr Environ Health Rep. 2015;2(2):204–13. 10.1007/s40572-015-0043-2. [PubMed: 26231368]

59. Vrijheid M, Slama R, Robinson O, Chatzi L, Coen M, van den Hazel P, et al. The human early-life exposome (HELIX): project rationale and design. Environ Health Perspect. 2014;122(6):535–44. 10.1289/ehp.1307204. [PubMed: 24610234]

60. Martens DS, Gouveia S, Madhloum N, Janssen BG, Plusquin M, Vanpoucke C, et al. Neonatal cord blood oxylipins and exposure to particulate matter in the early-life environment: an ENVIRONAGE birth cohort study. Environ Health Perspect. 2017;125(4):691–8. 10.1289/EHP291. [PubMed: 27814242]

61•. Yan Q, Liew Z, Uppal K, Cui X, Ling C, Heck JE, et al. Maternal serum metabolome and traffic-related air pollution exposure in pregnancy. Environ Int. 2019;130:104872 10.1016/j.envint.2019.05.066 [PubMed: 31228787] This is the first untargeted metabolomics study of maternal pregnancy exposure to air pollution illustrating how metabolomics can inform studies of air pollution in pregnancy.

62. Daher N, Saliba NA, Shihadeh AL, Jaafar M, Baalbaki R, Shafer MM, et al. Oxidative potential and chemical speciation of size-resolved particulate matter (PM) at near-freeway and urban background sites in the greater Beirut area. Sci Total Environ. 2014;470–471:417–26. 10.1016/j.scitotenv.2013.09.104.

63. Dick CAJ, Singh P, Daniels M, Evansky P, Becker S, Gilmour MI. Murine pulmonary inflammatory responses following instillation of size-fractionated ambient particulate matter. J Toxic Environ Health A. 2003;66(23):2193–207. 10.1080/716100636.

64. Ghio AJ, Carraway MS, Madden MC. Composition of air pollution particles and oxidative stress in cells, tissues, and living systems. J Toxicol Environ Health Part B. 2012;15(1):1–21. 10.1080/10937404.2012.632359.

65. Guerra R, Vera-Aguilar E, Uribe-Ramirez M, Gookin G, Camacho J, Osornio-Vargas AR, et al. Exposure to inhaled particulate matter activates early markers of oxidative stress, inflammation and unfolded protein response in rat striatum. Toxicol Lett. 2013;222(2):146–54. 10.1016/j.toxlet.2013.07.012. [PubMed: 23892126]

66. Happo MS, Uski O, Jalava PI, Kelz J, Brunner T, Hakulinen P, et al. Pulmonary inflammation and tissue damage in the mouse lung after exposure to PM samples from biomass heating appliances of old and modern technologies. Sci Total Environ. 2013;443:256–66. 10.1016/j.scitotenv.2012.11.004. [PubMed: 23201646]

67. Li H, Cai J, Chen R, Zhao Z, Ying Z, Wang L, et al. Particulate matter exposure and stress hormone levels: a randomized, double-blind, crossover trial of air purification. Circulation. 2017;136(7):618–27. 10.1161/CIRCULATIONAHA.116.026796. [PubMed: 28808144]

68. van Veldhoven K, Kiss A, Keski-Rahkonen P, Robinot N, Scalbert A, Cullinan P, et al. Impact of short-term traffic-related air pollution on the metabolome - results from two metabolome-wide experimental studies. Environ Int. 2018;123:124–31. 10.1016/j.envint.2018.11.034. [PubMed: 30522001]

69. Vlaanderen JJ, Janssen NA, Hoek G, Keski-Rahkonen P, Barupal DK, Cassee FR, et al. The impact of ambient air pollution on the human blood metabolome. Environ Res. 2017;156:341–8. 10.1016/j.envres.2017.03.042. [PubMed: 28391173]

70. Liang D, Moutinho JL, Golan R, Yu T, Ladva CN, Niedzwiecki M, et al. Use of high-resolution metabolomics for the identification of metabolic signals associated with traffic-related air pollution. Environ Int. 2018;120:145–54. 10.1016/j.envint.2018.07.044. [PubMed: 30092452]

71. Menni C, Metrustry SJ, Mohney RP, Beevers S, Barratt B, Spector TD, et al. Circulating levels of antioxidant vitamins correlate with better lung function and reduced exposure to ambient pollution. Am J Respir Crit Care Med. 2015;191(10):1203–7. 10.1164/rccm.201411-2059LE. [PubMed: 25978575]

72. Walker DI, Lane KJ, Liu K, Uppal K, Patton AP, Durant JL, et al. Metabolomic assessment of exposure to near-highway ultrafine particles. J Expo Sci Environ Epidemiol. 2018 10.1038/s41370-018-0102-5.

73. Jeong A, Fiorito G, Keski-Rahkonen P, Imboden M, Kiss A, Robinot N, et al. Perturbation of metabolic pathways mediates the association of air pollutants with asthma and cardiovascular diseases. Environ Int. 2018;119:334–45. 10.1016/j.envint.2018.06.025. [PubMed: 29990954]

74. Peter Stein T, Scholl TO, Schluter MD, Leskiw MJ, Chen X, Spur BW, et al. Oxidative stress early in pregnancy and pregnancy outcome. Free Radic Res. 2008;42(10):841–8. 10.1080/10715760802510069. [PubMed: 18985484]

75. Lavigne E, Burnett RT, Stieb DM, Evans GJ, Godri Pollitt KJ, Chen H, et al. Fine particulate air pollution and adverse birth outcomes: effect modification by regional nonvolatile oxidative potential. Environ Health Perspect. 2018;126(7):077012 10.1289/EHP2535. [PubMed: 30073952]

76. Duhig K, Chappell LC, Shennan AH. Oxidative stress in pregnancy and reproduction. Obstet Med. 2016;9(3):113–6. 10.1177/1753495X16648495. [PubMed: 27630746]

77. Al-Gubory KH, Fowler PA, Garrel C. The roles of cellular reactive oxygen species, oxidative stress and antioxidants in pregnancy outcomes. Int J Biochem Cell Biol. 2010;42(10):1634–50. [PubMed: 20601089]

78. Wilkinson AL, Pedersen SH, Urassa M, Michael D, Andreasen A, Todd J, et al. Maternal systemic or cord blood inflammation is associated with birth anthropometry in a Tanzanian prospective cohort. Tropical Med Int Health. 2017;22(1):52–62. 10.1111/tmi.12799.

79. Aye ILMH, Lager S, Ramirez VI, Gaccioli F, Dudley DJ, Jansson T, et al. Increasing maternal body mass index is associated with systemic inflammation in the mother and the activation of distinct placental inflammatory pathways. Biol Reprod. 2014;90(6):129 10.1095/biolreprod.113.116186. [PubMed: 24759787]

80. Kennelly MA, Ainscough K, Philips CM, Alberdi G, Lindsay KL, McAuliffe FM. 1012: maternal inflammation: potential mediators and effects on pregnancy outcomes. Am J Obstet Gynecol. 2019;220(1):S650–S1.

81. Fan JQ, Lv JC. Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc B. 2008;70:849–83. 10.1111/j.1467-9868.2008.00674.x.

82. Efron B Large-scale inference: empirical Bayes methods for estimation, testing, and prediction. Cambridge University Press; 2012.

83. Huang YT, Pan WC. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. Biometrics. 2016;72(2):402–13. 10.1111/biom.12421. [PubMed: 26414245]

84. Yang T, Niu J, Chen H, Wei P. Estimation of mediation effect for high-dimensional omics mediators with application to the Framingham Heart Study. bioRxiv. 2019:774877.

85. Paul KC, Haan M, Mayeda ER, Ritz BR. Ambient air pollution, noise, and late-life cognitive decline and dementia risk. Annu Rev Public Health. 2019;40:203–20. 10.1146/annurevpublhealth-040218-044058. [PubMed: 30935305]

86. Sheppard L, Burnett RT, Szpiro AA, Kim SY, Jerrett M, Pope CA 3rd, et al. Confounding and exposure measurement error in air pollution epidemiology. Air Qual Atmos Health. 2012;5(2):203–16. 10.1007/s11869-011-0140-9. [PubMed: 22662023]
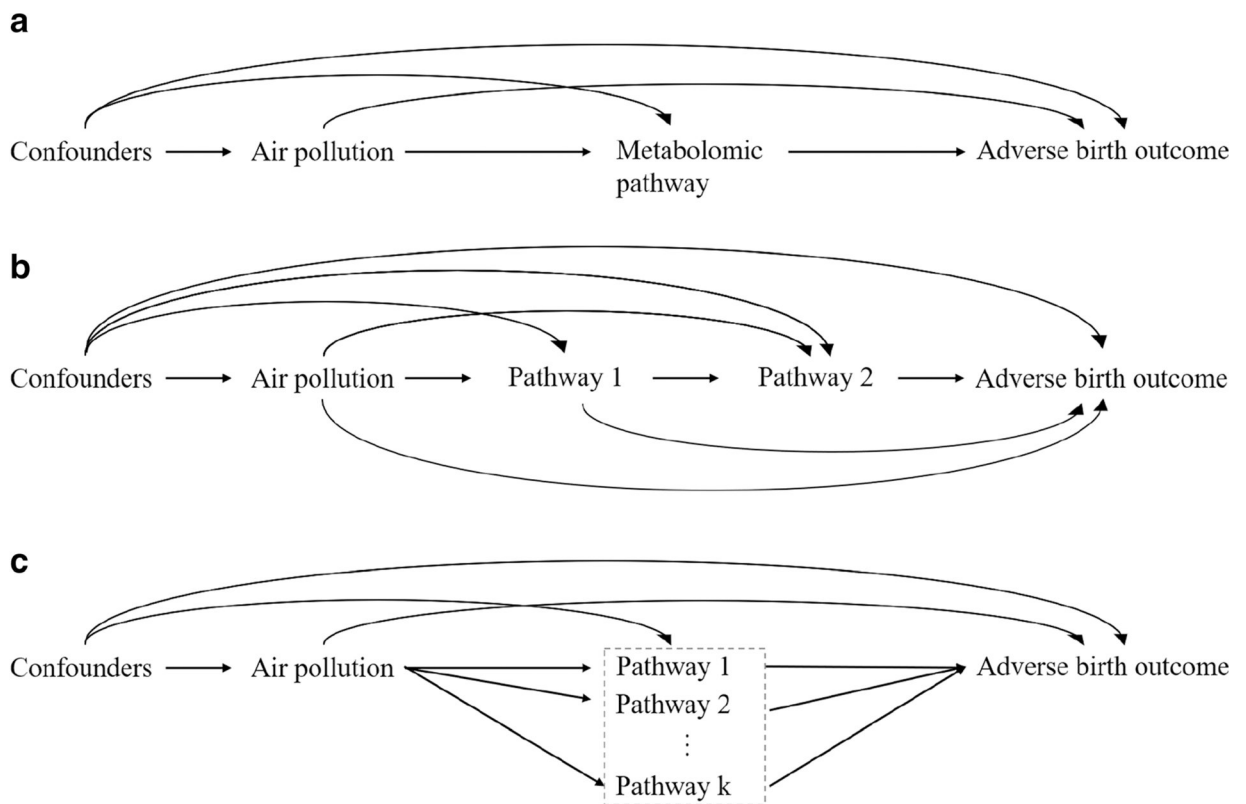
87. Weisskopf MG, Webster TF. Trade-offs of personal versus more proxy exposure measures in environmental epidemiology. Epidemiology. 2017;28(5):635–43. 10.1097/EDE.0000000000000686. [PubMed: 28520644]

88. Lu W, Su X, Klein MS, Lewis IA, Fiehn O, Rabinowitz JD. Metabolite measurement: pitfalls to avoid and practices to follow. Annu Rev Biochem. 2017;86:277–304. 10.1146/annurev-biochem-061516-044952. [PubMed: 28654323]

89. Renson A, Herd P, Dowd JB. Sick individuals and sick (microbial) populations: challenges in epidemiology and the microbiome. Annu Rev Public Health. 2020;41:63–80. 10.1146/annurev-publhealth-040119-094423. [PubMed: 31635533]

90. Rappaport SM. Redefining environmental exposure for disease etiology. NPJ Syst Biol Appl. 2018;4:30 10.1038/s41540-018-0065-0. [PubMed: 30181901]

91. Fritz MS, Mackinnon DP. Required sample size to detect the mediated effect. Psychol Sci. 2007;18(3):233–9. 10.1111/j.1467-9280.2007.01882.x. [PubMed: 17444920]

92. Kenny DA, Judd CM. Power anomalies in testing mediation. Psychol Sci. 2014;25(2):334–9. 10.1177/0956797613502676. [PubMed: 24311476]

93. Blaise BJ, Correia G, Tin A, Young JH, Vergnaud AC, Lewis M, et al. Power analysis and sample size determination in metabolic phenotyping. Anal Chem. 2016;88(10):5179–88. 10.1021/acs.analchem.6b00188. [PubMed: 27116637]
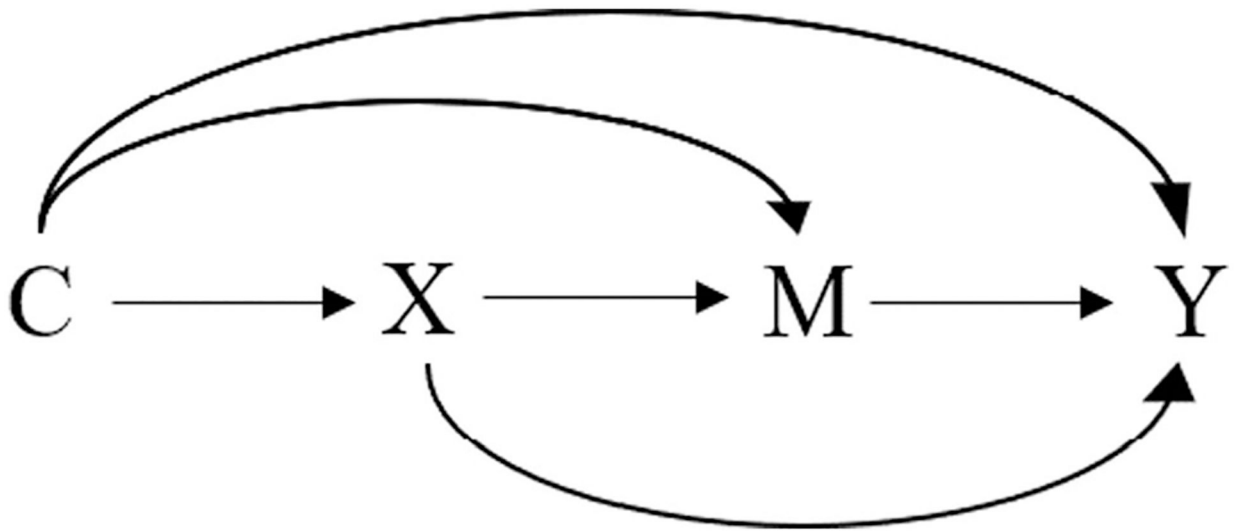
Fig. 1.
Causal diagram representing simple mediation. Legend: X: the exposure, M: the mediator, Y: the outcome, C: a set of confounders.
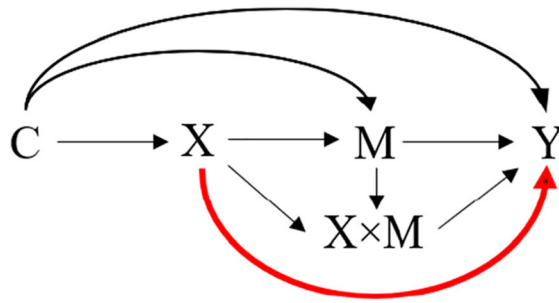
Fig. 2.
Causal diagram with the interaction representing a 4-way decomposition [47]. Legend: X: the exposure, M: the mediator, X × M: the interaction between the exposure and the mediator, Y: the outcome, C: a set of confounders. Red line shows each effect. **a** Controlled direct effect. **b** Reference interaction effect. **c** Mediated interaction effect. **d** Pure indirect effect

a

Confounders → Air pollution → Metabolomic pathway → Adverse birth outcome

b

Confounders → Air pollution → Pathway 1 → Pathway 2 → Adverse birth outcome

c

Confounders → Air pollution → Pathway 1 / Pathway 2 / ⋮ / Pathway k → Adverse birth outcome

Fig. 3.
Examples of causal diagrams of multi-omic research questions using mediation analysis. **a** represents a single metabolomic pathway as a mediator on the causal pathway between air pollution and adverse birth outcomes. **b** represents multiple ordered metabolomic pathways as mediators on the causal pathway between air pollution and adverse birth outcomes. **c** represents multiple metabolomic pathways treated as an unordered set of mediators on the causal pathway between air pollution and adverse birth outcomes

Table 1

Effect decomposition in causal mediation analysis

	4-way decomposition			
	Controlled direct effect	Reference interaction effect	Mediated interaction effect	Pure indirect effect
(i) 2-way decomposition: pattern A				
Pure direct effect	×	×		
Total indirect effect			×	×
(ii) 2-way decomposition: pattern B				
Total direct effect	×	×	×	
Pure indirect effect				×
(iii) 2-way decomposition: pattern C				
Controlled direct effect	×			
Portion eliminated		×	×	×
(iv) 3-way decomposition: pattern A				
Pure direct effect	×	×		
Mediated interaction effect			×	
Pure indirect effect				×
(v) 3-way decomposition: pattern B				
Controlled direct effect	×			
Portion attributable to interaction		×	×	
Pure indirect effect				×