Curr Epidemiol Rep (2017) 4:117–123
DOI 10.1007/s40471-017-0112-1

REPRODUCTIVE AND PERINATAL EPIDEMIOLOGY (R PLATT, SECTION EDITOR)

Causal Mediation Analysis in Pregnancy Studies: the Case of Environmental Epigenetics

Linda Valeri [1,2]

Published online: 29 April 2017

Abstract
Purpose of Review Studies on the mechanisms of action of environmental exposures in pregnancy are increasingly popular. In particular, it is of interest to investigate the role of genetic and epigenetic factors as mediators of the maternal environmental exposures' effects on perinatal outcomes. Causal mediation analysis lies at the center of environmental epigenetics research, and the methodological challenges that arise in this context have not yet been fully articulated.
Recent Findings Measurement error, unmeasured confounding, reverse causation, and multiple mediators are often disregarded issues in environmental epigenetic studies that can lead to important biases. Considering the study of maternal smoking effect on birth weight potentially mediated by DNA methylation as example, I discuss the impact of these phenomena on estimation and testing of causal pathways. Statistical methods have been recently introduced to account for these frequently encountered issues.
Summary Causal interpretation of pregnancy studies on the role epigenetic factors as mediators of environmental exposures effects can be improved by the adoption of recent methodological advancements in mediation analysis that correct for measurement error, use genetic instrumental variables, and account for the presence of multiple mediators.

This article is part of the Topical Collection on *Reproductive and Perinatal Epidemiology*

✉ Linda Valeri
lvaleri@mclean.harvard.edu

[1] Psychiatric Biostatistics Laboratory, McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA

[2] Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA

Keywords Environmental epigenetics · Measurement error · Mediation analysis · Mendelian randomization · Multiple mediators

Introduction

Pregnancy is a time of enhanced vulnerability in which environmental exposures exert effects on a variety of maternal and child perinatal outcomes that can have long-lasting consequences over the life-course. To investigate causal mechanisms in observational studies of environmental exposures and perinatal outcomes, studies screen human populations for biomarkers of exposures as well as mechanistic intermediates. Mediation analysis is a key tool to investigate causal pathways, biological mechanisms, and to design policy interventions [1]. Modern approaches to mediation have been inspired by the pioneering work of the geneticist Sewall Wright (1920), who developed the path analysis method. Path analysis is now viewed as a special case of structural equation modeling (SEM) and with the work of Baron and Kenny [2] became widely used in the context of linear models. Application of a counterfactual framework [3] has further provided a strong theoretical basis for causal inference in mediation analysis by precisely defining the causal contrasts along with necessary assumptions for their identifiability. Using the counterfactual framework has allowed for definitions of direct and indirect effects and for decomposition of a total effect into direct and indirect effects, even in models with interactions and nonlinearities [4, 5••]. The use of mediation analysis is now widespread in perinatal epidemiology and has resolved important paradoxes in the field, such as the birth weight paradox [6–8]. The application of mediation analysis in the context of environmental determinants of perinatal outcomes has been more recent but particularly fast growing [9–11, 12•,

13]. The purpose of this paper is to highlight methodological challenges in pregnancy studies that aim at investigating mediating mechanisms of the effect of environmental factors on perinatal outcomes and to provide some solutions based on the most recent methodological developments in mediation analysis. I introduce the study of the role of DNA methylation as mediator of the effect of maternal smoking on birth weight as motivating example. I then provide a review of important definitions and assumptions in causal mediation analysis. Finally, I delve into three issues that are often encountered: (1) measurement error, (2) unmeasured confounding and reverse causation, and (3) multiple mediators. For each of these challenges, I describe practical solutions for both estimation and testing of indirect effects.

The Role of Epigenetic Factors in Explaining the Effect of Environmental Exposures on Birth Outcomes

The role of epigenetic factors as mediators of the early life "programming" of health is becoming increasingly apparent [14]. Epigenetic mechanisms regulate all gene expression by determining the accessibility of DNA to drivers of gene activation, such as transcription factors [15, 16]. As they are not part of the genome, epigenetic marks are responsive to the environment and, in contrast to genetic mechanisms, are also cell type and developmental stage-specific [15, 16]. DNA methylation (DNAm) is the most well-studied epigenetic mechanism. It occurs when a methyl group has been added to a cytosine followed by a guanine (CpG site) base pair on the genome and is measured across the genome [15]. When a pattern of changes of DNAm is found to occur repeatedly at specific loci, discriminating the phenotypically affected cases from control individuals, this is regarded as an indication that epigenetic perturbation has taken place that is associated, possibly causally, with the phenotype. This approach is described as an epigenome-wide association study (EWAS) [17] and takes its cue from the association of genetic variability with phenotypes in genome-wide association studies (GWAS).

Evidence is accumulating from EWAS that environmental exposures modify the epigenome. In humans, the best-studied epigenetic modification is methylation and the best-studied exposure is smoking. Smoking has been reproducibly associated with alterations in methylation at specific loci in newborns whose mothers smoked during pregnancy [18]. These smoking methylation signals have been used to develop novel biomarkers of exposure [19]. Given strong evidence of differential methylation in newborns in relation to smoking by the mother, it has been of interest to consider whether these signals mediate the effects of maternal smoking on perinatal outcomes such as birth weight. It has recently been reported that differential DNAm of a single CpG site in placenta mediates up to 36% of the effect of smoking on lower birth weight [12•]. In another study, differential methylation in newborn blood at a single CpG site in a different gene was reported to mediate 19–46% of the relationship between smoking and birth weight [13]. Pregnancy studies focusing on other environmental exposures (e.g., air pollution, nutrition) and other perinatal outcomes (e.g., pre-term birth, preeclampsia) find similarly strong evidence of mediated effects through DNAm of single CpGs [9–11, 14, 20]. The striking results are obtained conducting mediation analyses that ignore in part or all the issues that we discuss here, namely exposure measurement error, confounding, reverse causation, and multiple mediators.

Mediation Analysis: Causal Contrasts and Assumptions

With reference to the example of mediation of the effect of maternal smoking during pregnancy on newborn birth weight by smoking-related differential methylation, let *A* denote the exposure, maternal smoking, and *M* denote the mediator, DNAm. Let *Y* denote the outcome, birth weight, and *C* denote a vector of covariates representing potential confounders. The directed acyclic graph in Fig. 1 describes the setting of mediation analysis. Mediation analysis can be employed to quantify how much of the total effect of maternal smoking on birth weight (Fig. 1a) is explained by the indirect effect of smoking on birth weight that is mediated by the DNA methylation level, relative to the direct effect of smoking on birth weight through pathways independent of DNA methylation (Fig. 1b). Under the counterfactual framework for causal inference, direct and indirect causal effects have been rigorously defined [20, 21].

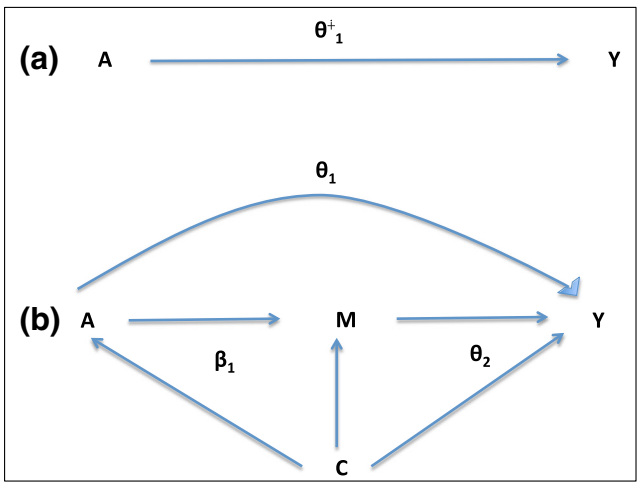

Fig. 1 **a** Directed acyclic graph for average causal effect of sustained smoking during pregnancy (*A*) on birth weight (*Y*) (TE = $\theta_1^{\dagger}$ from Eq. 1). **b** Directed acyclic graph for direct of sustained smoking during pregnancy (*A*) on birth weight (*Y*) and indirect effect of sustained smoking during pregnancy (*A*) on birth weight (*Y*) through DNA methylation (*M*) (NDE = θ_1 from Eq. 3, NIE = $\theta_1^{\dagger} - \theta_1 = \beta_1\theta_2$)

Let Y_a and M_a denote the value of the outcome and mediator that would have been observed if the exposure A had been set to level a. Let Y_{am} denote the birth weight that would have been observed if the smoking status and DNAm had been set to levels a and m, respectively. The average total effect, conditional on $C = c$, comparing exposure level 1 to 0, is defined by TE $= E(Y_1 - Y_0 \mid c)$, which compares the average outcome in subgroup $C = c$ if the mother had been a smoker with the average outcome in subgroup $C = c$ if the mother had been a non-smoker. The controlled direct effect is defined as CDE $= E(Y_{1m} - Y_{0m} \mid c)$, that is the effect of smoking on birth weight while fixing the mediator to an arbitrary value, m. The natural direct effect, conditional on $C = c$, is the effect of smoking status on birth weight when the mediator is fixed to the level, it would have naturally been had the mother not been a smoker, $A = 0$. It is defined in counterfactual notation by NDE $= E(Y_{1M_0} - Y_{0M_0} \mid c)$. The natural indirect effect, conditional on $C = c$, comparing the effect of DNAm at levels M_1 and M_0 while fixing the exposure at level 1, is defined by NIE $= E(Y_{1M_1} - Y_{1M_0} \mid c)$. It can be shown that on the risk difference scale, the total effect decomposes to the sum of natural direct and natural indirect effects (i.e., TE = NDE + NIE). In the example, this is interpretable as the indirect effect of smoking on birth weight that is mediated by the methylation level.

To validly estimate direct and indirect effects, the following four assumptions need to be satisfied. Conditioning on a vector of covariates $\boldsymbol{C}$, there is no unmeasured confounding of (i) the exposure–outcome relationship, (ii) the mediator–outcome relationship, (iii) the exposure–mediator relationship, and (iv) there are no mediator–outcome confounders affected by the exposure [22]. Furthermore, models for the outcome and mediator need to be correctly specified. For continuous outcome and mediator (as in the current setting of outcome birth weight and mediator methylation), under the assumption of no exposure–mediator interaction in the outcome model, typically made by published applications of mediation analysis in environmental epigenetics, if we specify three linear regression models

$$E(Y|A = a, \boldsymbol{C} = \boldsymbol{c}) = \theta_0^{\dagger} + \theta_1^{\dagger} a + \boldsymbol{\theta}^{\dagger'} \mathrm{c} \quad (1)$$

$$E(Y|A = a, M = m, \boldsymbol{C} = \boldsymbol{c}) = \theta_0 + \theta_1 a + \theta_2 m + \boldsymbol{\theta}' \mathrm{c} \quad (2)$$

$$E(M|A = a, \boldsymbol{C} = \boldsymbol{c}) = \beta_0 + \beta_1 a + \boldsymbol{\beta}' \mathrm{c}; \quad (3)$$

then, the estimators of total effect (TE), direct effect (NDE), and indirect effect (NIE) take the following form [2, 23]:

$$\mathrm{TE} = \theta_1^{\dagger} \quad (4)$$

$$\mathrm{NDE} = \theta_1 \quad (5)$$

$$\mathrm{NIE} = \beta_1 \theta_2 = \theta_1^{\dagger} - \theta_1. \quad (6)$$

Estimators for direct and indirect effects in the presence of exposure–mediator interactions and non-linear effects can be obtained under the counterfactual framework [4]. The most popular test for indirect effects is based on the product method, also known as the Sobel test [24]. This is a Wald test for the null hypothesis $H_0 : \beta_1 \theta_2 = 0$ based on the delta method standard error $\sigma_{\mathrm{NIE}} = \sqrt{\sigma_{\theta_2}^2 \beta_1^2 + \sigma_{\beta_1}^2 \theta_2^2}$, where $\sigma_{\theta_2}^2$ and $\sigma_{\beta_1}^2$ are the variances of the maximum likelihood estimates of θ_2 and β_1, respectively.

Is DNAm a Mediator or a Biomarker?

It is widely acknowledged that measurement of human environmental exposures, including smoking, is prone to error [25]. Random error exists for all exposures. Nonetheless, most studies that address whether methylation signatures from smoking mediate its perinatal outcomes have ignored the potential role of measurement error in assessment of smoking [12•, 13]. Given this measurement error, evaluation of mediation is complicated by the fact that the proposed mediators, DNA sites differentially methylated by smoking, are excellent biomarkers that may better capture the smoking exposure, which is almost always self-reported [19]. A recent study has shown that when exposure is measured with error, the exposure coefficient in the outcome model (Eqs. 1 and 2) is biased downward inducing, as expected, an underestimation of the total effect and of the direct effect. However, the bias of the naïve indirect effect estimator can be in either direction. In particular, when the mediator is a strong biomarker for the exposure (i.e., $\beta_1 \neq 0$; Eq. 3), as is the case for smoking methylation signals, the bias of the total effect estimator is larger than the bias of the natural direct effect estimator, leading to over-estimation of the indirect effect [26••]. In other words, when the mediator captures the variability of true latent smoking exposure better than the self-reported measure of smoking, some of the direct effect is incorrectly attributed to the mediator (the indirect effect). Exposure measurement error has implications on the validity of the Sobel test as well [26••]. Testing for an indirect effect implies evaluating a composite null hypothesis, as the indirect effect can be null under three scenarios: when (1) Neither the exposure is associated with the mediator nor the mediator is associated with the outcome, (2) The exposure is associated with the mediator but the mediator is not associated with the outcome, and (3) The exposure is not associated with the mediator but the mediator is associated with the outcome. The test for indirect effect is notoriously conservative [27]. However, when the mediator is a strong biomarker for the exposure, the exposure is not associated with the outcome (under the null case 3), and the exposure is measured with error, the type I error rate will not be preserved [26••]. Therefore, in reasonable scenarios of

mediation analysis in environmental epigenetic studies, the naïve mediation analysis is likely biased, and there is risk of reporting false positive findings of mediated effects through DNAm whenever the exposure is imperfectly measured and DNAm is a biomarker of the exposure.

Several steps can be taken to minimize this bias. First, the investigator should define carefully the exposure clarifying the timing and considering preferably a continuous measurement (e.g. number of cigarettes smoked per day during the first trimester, rather than (any) smoking during pregnancy). Second, resources permitting, the study design should include the collection of replicates or gold standard measurements of the exposure. Finally, at the analysis stage, statistical approaches for measurement error should be employed. To correct for measurement error or misclassification and obtain valid inferences on natural direct and indirect effects (defined in the previous section), a two-stage approach has been introduced [28–29]. In the first stage, assuming plausible values for the magnitude of measurement error, characterized by either the variance of the error for continuous exposure or misclassification probabilities for categorical exposures, mediator, and outcome regression coefficients can be estimated using either regression calibration, SIMEX (simulation and extrapolation), or the EM (expectation-maximization) algorithm approaches for measurement error correction [30–32]. In the second stage, the coefficient estimates are plugged into the formulas of NDE and NIE to obtain measurement error-corrected estimates of the causal contrasts of interest with standard errors obtained via the bootstrap. When the amount of error is not known from external validation data, a sensitivity analysis can be conducted.

Unmeasured Confounding and Reverse Causation

Most pregnancy studies on environmental and lifestyle factors measure the DNAm profile in the placenta; smoking has been shown to exert an effect of global and gene-specific placental methylation [33]. However, confounding, where a common exposure influences both epigenetic profile and phenotypic outcome in the absence of a causal link, must be considered. Even with careful measures of exposures to minimize residual confounding, there is always potential for unobserved effects on both epigenetic profile and outcome of interest (unmeasured confounding). Moreover, given the general requirement to measure epigenetic marks at birth, after any effects on early development are likely to have commenced [14], it is very difficult to ascribe a direct causal link to any observed epigenetic association. It is possible that epigenetic changes in the placenta observed at birth are a consequence of, and not a cause of, disrupted placental functioning and pregnancy development (i.e., reverse causation, where the intended outcome is observed to precede the effect). Sensitivity analyses for unmeasured confounding in mediation analyses have been developed [34–35] and are becoming routine practice in perinatal epidemiology. However, they have not yet been adopted to evaluate robustness of such environmental epigenetics studies to confounding bias.

Another approach that has the potential to address issues of measurement error, unmeasured confounding, and reverse causation is the Mendelian randomization (MR) approach [36, 37••]. In MR, if there are genetic variants robustly associated with the exposure of interest and other *independent* genetic variants robustly associated with the mediator of interest, these can be used to help infer causality. These genetic variants are correctly measured, are not associated with various confounders, and are not directly influenced by the outcome of interest. These variants must satisfy the assumptions of an instrumental variable (IV): are associated with the exposure of interest, are not associated with any confounder (including those that are unmeasured), and are not associated with the outcome given the exposure and all the confounders [38]. Such genetic variants divide the observed population into subgroups analogous to arms in a randomized controlled trial where the intervention is to change the level of the exposure. Let exposure and mediator each have corresponding genetic IVs, G_a and G_m, respectively. A causal DAG illustrating the relationships between these variables is given in Fig. 2. It has been shown that if all *effects are linear without interaction terms* [37••], the causal effects of A on Y, of A on M, and of M on Y can each be estimated by application of the ratio method [39] and then used to estimate the effects in Eqs. (4)–(6). The coefficient from the regression of the outcome on the exposure's IV, $\gamma_{Y \mid G_a}$, is divided by the coefficient from the regression of the exposure on the IV, $\gamma_{A \mid G_a}$ to obtain the total effect:

$$\mathrm{TE} = \gamma_{Y \mid G_a} / \gamma_{A \mid G_a}.$$

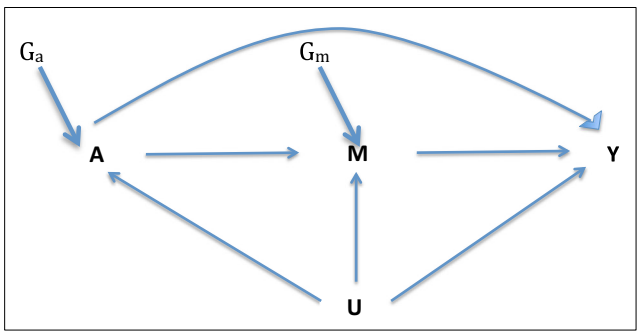

Fig. 2 Causal directed acyclic graph (DAG) leading to direct and indirect causal effects of variable A on Y with mediator M, associated instrumental variables G_a and G_m, and unmeasured confounders U

The natural indirect effect is estimated similarly upon obtaining the *A–M* and *M–Y* associations in a similar fashion:

$$\mathrm{NIE} = \left\{ \gamma_{M \mid G_a} / \gamma_{A \mid G_a} \right\} \times \left\{ \gamma_{Y \mid G_m} / \gamma_{M \mid G_m} \right\}.$$

Finally, the natural direct effect is given by the following:

$$\mathrm{NDE} = \mathrm{TE} - \mathrm{NIE}.$$

Although the MR approach appears simple, it relies on very strong assumptions. First, it is very hard to find genetic variants, which satisfy the IV assumptions for the exposure–mediator and mediator–outcome relationships. In the example considered here, variants for smoking exposure and certain DNA methylation CpGs have been uncovered [36]. Another important limitation is that it is possible that associations may reflect pleiotropy (multiple effects of a single gene) rather than mediation [37••]. If there are alternative pathways by which variants associated with the exposure may be associated with the mediator, then the assessment of mediation is more problematic. Discussion of these and other limitations of MR is given in [40•]. It is recommended to use MR only where the IV assumptions have a strong biological or scientific basis. Furthermore, results of MR should be compared with alternative approaches, such as sensitivity analyses for unmeasured confounding and measurement error. Finally, it is not clear yet how one would go about applying MR when more than one mediator is of interest, the final issue that I discuss in the next section.

Multiple Mediators

Another important yet understudied problem in mediation analysis in environmental epigenetics studies is how to estimate and test indirect effects in the presence of multiple or high-dimensional sets of mediators (Fig. 3). Currently, most mediation analyses in EWAS are conducted separately for each mediator (say DNAm of each CpG site separately) and post-hoc Bonferroni or permutation-based corrections for multiple comparisons are applied [12•, 13, 41, 42]. This approach is problematic for two main reasons. First, failure to adjust for other mediators could lead to inefficiency, if mediators are independent of each other, exacerbating the conservativeness of the Sobel test [43]. Second, this approach can introduce bias, and Bonferroni correction would be inappropriate, if mediators are correlated with each other [44••]. This latter issue may be more troublesome as the correlation among probes close to one another can be as high as 0.6 [45] in cell lines. It is therefore advisable to include multiple mediators in one model to determine to what extent the specific indirect effects are associated with mediators. As in GWAS, investigators can adopt a "candidate gene" approach and investigate multiple selected mediators. Alternatively, investigators can consider a high-dimensional set of mediators. Approaches for mediation analysis for multiple mediators are available and should be employed to estimate and test direct effects and joint indirect effects. Some of these approaches do not require modeling the mediators [44••, 46]. If mediators are correlated and temporally ordered (either by design or based on prior knowledge) certain path-specific effects can be estimated [47]. These approaches for multiple mediators do not accommodate high-dimensional mediators, and more research is needed to ensure valid causal inferences in this setting. Some initial proposals involve variable selection procedures before applying multiple mediators' approaches and provide joint testing procedures more appropriate to this setting [43, 48••].

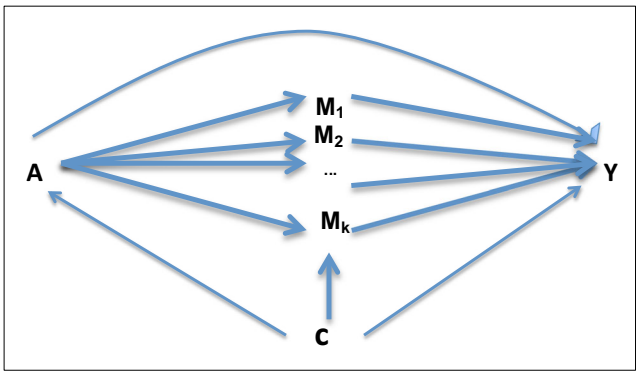

Fig. 3 Mediation analysis with multiple mediators

Concluding Remarks

Causal mediation analysis is a key approach to investigate the role of epigenetic factors in explaining the effect on environmental exposures on birth outcomes. Causal interpretation of such investigations can be strengthened by the adoption of recent methodological advancements in mediation analysis that correct for measurement error, use genetic instrumental variables, and account for the presence of multiple mediators.

Compliance with Ethical Standards

Conflict of Interest Linda Valeri declares no potential conflict of interest.

Human and Animal Rights and Informed Consent This article contains no studies with human or animal subjects performed by the author.

References

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

- • Of importance
- •• Of major importance

1. Pearl J. Interpretation and identification of causal mediation. Psychol Methods. 2014;19:459.
2. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51:1173–82. doi:10.1037/0022-3514.51.6.1173.
3. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66:688.
4. Valeri L, Vander Weele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013;18:137.
5. •• Vander Weele TJ. Explanation in causal inference: methods for mediation and interaction. London: Oxford University Press; 2015. ISBN-13: 978–0199325870; ISBN-10: 0199325871. **The book provides an accessible but comprehensive overview of methods for mediation. It includes an introduction to concepts and regression methods for mediation analysis. Furthermore, the book discusses sensitivity analysis techniques for unmeasured confounding and measurement error, methods for multiple mediators, and longitudinal data. Finally, the book devotes lengthy discussion on relations between mediation and other concepts involving intermediates such as instrumental variables and Mendelian randomization**.
6. Vander Weele TJ, Hernández-Diaz S. Is there a direct effect of pre-eclampsia on cerebral palsy not through preterm birth? Paediatr Perinat Epidemiol. 2011;25:111–5.
7. Ananth CV, Vander Weele TJ. Placental abruption and perinatal mortality with preterm delivery as a mediator: disentangling direct and indirect effects. Am J Epidemiol. 2011: kwr045.
8. Vander Weele TJ, Mumford SL, Schisterman EF. Conditioning on intermediates in perinatal epidemiology. Epidemiology (Cambridge, Mass). 2012;23(1):1–9. doi:10.1097/EDE.0b013e31823aca5d.
9. Ferguson KK, Chen YH, Vander Weele TJ, McElrath TF, Meeker JD, Mukherjee B. Mediation of the relationship between maternal phthalate exposure and preterm birth by oxidative stress with repeated measurements across pregnancy. Environ Health Perspect. 2017;125:488–94. doi:10.1289/EHP282.
10. Janssen BG, Byun HM, Gyselaers W, Lefebvre W, Baccarelli AA, Nawrot TS. Placental mitochondrial methylation and exposure to airborne particulate matter in the early life environment: an ENVIR ON AGE birth cohort study. Epigenetics. 2015;10(6):536–44.
11. Sharp GC, Lawlor DA, Richmond RC, Fraser A, Simpkin A, Suderman M, …, Gaunt TR. Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: findings from the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2015:dyv042.
12. • Küpers LK, Xiaojing X, Soesma AJ, et al. DNA methylation mediates the effect of maternal smoking during pregnancy on birth weight of the offspring. Int J Epidemiol. 2015;44(4):1224–37. **The article applies mediation analysis to assess the proportion of the effect of exposure to maternal smoking during pregnancy on birth weight explained by differential methylation in relation to this exposure at CpGs in the *GFI1* gene.**
13. Morales E, Vilahur N, Salas LA, et al. Genome-wide DNA methylation study in human placenta identifies novel loci associated with maternal smoking during pregnancy. Int J Epidemiol. 2016;45(5):1644–55.
14. Januar V, Desoye G, Novakovich B, Cvitic S, Saffery R. Epigenetic regulation of placental function and pregnancy outcome: considerations for causal inference. Am J Obstet Gynecol. 2015;213(4):S182–96.
15. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13:484–92.
16. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16:6–21.
17. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529–41. doi:10.1038/nrg3000.
18. Joubert BR, Felix JF, Yousefi P, et al. DNA methylation in newborns and maternal smoking in pregnancy: genomewide consortium meta-analysis. Am J Hum Genet. 2016;98(4):680–96.
19. Reese SE, Zhao S, Wu MC, et al. DNA methylation score as a biomarker in newborns for sustained maternal smoking during pregnancy. Environ Health Perspect. 2016; doi:10.1289/EHP333.
20. Lin V, Baccarelli AA, Burris HH. Epigenetics—a potential mediator between air pollution and pre-term birth. Environmental epigenetics. 2016;2(1):dvv008.
21. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992;3(2):143–55.
22. Pearl J. Direct and indirect effects. In: Proceedings of the seventeenth conference on uncertainty in artificial intelligence. CA: Morgan Kaufmann Publishers Inc; 2001. p. 411–420.
23. Vander Weele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Statistics and Its Interface. 2009;2:457–68.
24. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol Methodol. 1982;13:290–312.
25. Rhomberg LR, Chandalia JK, Long CM, Goodman JE. Measurement error in environmental epidemiology and the shape of exposure-response curves. Crit Rev Toxicol. 2011;41(8):651–71.
26. •• Valeri L, Reese SL, Zhao S, Page CM, Nystad W, Coull BA, London SJ. Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight? Epigenomics. 2017;9(3):253–65. **The paper studies bias due to misclassification of a binary exposure in mediation analysis and develops correction approaches. Furthermore, it investigates the impact of measurement error in environmental epigenetics studies that aim at testing and estimating the effect of environmental exposures on health outcomes potentially mediated by epigenetics factors in a large cohort. The study finds that when DNA methylation is a precisely measured, and strong biomarker for the exposure, then the indirect effect might be overestimated and type I error of the test for mediation might not be preserved.**
27. MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. Multivar Behav Res. 1995;30(1):41–62.
28. Valeri L, Vander Weele TJ. The estimation of direct and indirect causal effects in the presence of misclassified binary mediator. Biostatistics. 2014;15(3):498–512. doi:10.1093/biostatistics/kxu007.
29. Valeri L, Lin X, Vander Weele TJ. Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model. Stat Med. 2014;33(28):4875–90. doi:10.1002/sim.6295.
30. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. Am J Clin Nutr. 1997;65(4):1179S–86S.
31. Carroll RJ, Küchenhoff H, Lombard F, Stefanski LA. Asymptotics for the SIMEX estimator in nonlinear measurement error models. J Am Stat Assoc. 1996;91(433):242–50.

32. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Methodol. 1977:1–38.
33. Novakovic B, Saffery R. The ever-growing complexity of placental epigenetics role in adverse pregnancy outcomes and fetal programming. Placenta. 2012;33:959–70.
34. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. Stat Sci. 2010;25(1):51–71.
35. Vander Weele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology (Cambridge, Mass.). 2010;21 (4):540.
36. Relton CL, Smith GD. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012;41(1):161–76.
37.•• Burgess S, Daniel RM, Butterworth AS, Thompson SG, EPIC-InterAct Consortium. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol. 2015;44(2):484–95. **The paper extends the Mendelian randomization approach to investigate more complex networks of relationships between variables, in particular where some of the effect of an exposure on the outcome may operate through an intermediate variable (a mediator). The authors propose a regression-based method or structural equation models for instrumental variable analysis of direct and indirect effects of the exposure on the outcome and to assess the direction of effect between the exposure and a possible mediator. The approach is valid provided that instrumental variables for the exposure and mediator are available and the exposure and mediator do not interact in their effect on the outcome.**
38. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007;16(4):309–30.
39. Martens E, Pestman W, de Boer A, Belitser S, Klungel O. Instrumental variables: application and limitations. Epidemiology. 2006;17:260–7.
40.• Vander Weele TJ, Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. Epidemiology (Cambridge, Mass.). 2014;25(3):427. **The paper provides a clear presentation of the assumptions of Mendelian randomization approach and discusses their biological plausibility. In particular, the authors discuss issues of weak instruments, pleiotropy (multiple effects of a single gene), measurement error, the possibility of reverse causation, and the presence of linkage disequilibrium. Data analysis examples are given illustrating that the inappropriate use of instrumental variable techniques when the Mendelian randomization assumptions are violated can lead to biases of significant magnitude.**
41. van der Laan MJ, Dudoit S, Pollard KS. Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. Stat Appl Genet Mol Biol. 2004;3:Article 14.
42. Liu Y, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013;31:142–7.
43. Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P. Estimating and testing high-dimensional mediation effects in epigenetic studies. Bioinformatics. 2016:btw351.
44.•• Vander Weele T, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiologic methods. 2014;2(1):95–115. **The paper extends regression approaches within the generalized linear model framework for causal mediation analysis with multiple mediators. The article also discusses how certain pathways can be estimated in the presence of sequential mediators when the order is known.**
45. Moen EL, Zhang X, Mu W, Delaney SM, Wing C, McQuade J, Myers J, Godley LA, Dolan ME, Zhang W. Genome-wide variation of cytosine modifications between European and African populations and the implications for complex traits. Genetics. 2013;194 (4):987–96.
46. Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen EJ. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. Am J Epidemiol. 2015;181(5):349–56.
47. Vander Weele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology. 2014;2:300–6.
48.•• Huang YT, Pan WC. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. Biometrics. 2015. **The paper develops testing approaches for mediation analysis in high-dimensional mediator setting. The authors characterize the marginal mediation effect, the multivariate component-wise mediation effects, and the component-wise effect. Furthermore, a Monte-Carlo procedure for evaluating the statistical significance of the joint mediated effect and component-wise mediated effects is proposed**.