



Statistical Challenges in the Analysis of Biomarkers of Environmental Chemical Exposures for Perinatal Epidemiology

Janice M.Y. Hu¹ · Liheng Harry Zhuang¹ · Brendan A. Bernardo¹ · Lawrence C. McCandless¹

Published online: 4 July 2018

© Springer International Publishing AG, part of Springer Nature 2018

Abstract

Purpose of Review Biomarkers are widely used in perinatal epidemiology to examine the health effects of environmental chemical exposures during pregnancy. These measurements take the form of chemical concentrations measured in blood, urine, or other biospecimens. Biomarkers have the advantage of providing objective estimates of chemical exposures from multiple sources. However, they are difficult to handle at the data analysis stage. We review recent trends and developments in the statistical analysis of biomarkers with particular emphasis on exposure assessment and multivariable modeling.

Recent Findings Six statistical challenges are presented in the recent literature: (1) the analysis of biomarkers that fall below the limit of detection, (2) adjustment for dilution-dependent sample variation, (3) handling repeated biomarker measurements within a single pregnancy, (4) accounting for heterogeneity in biomarker levels between chemicals within the same chemical class, (5) variable selection and shrinkage for biomarkers in the same class, and finally, (6) dimension reduction strategies including the sum-of-chemical approach.

Summary The analysis of biomarkers of environmental chemical exposures remains immensely difficult, and the proper application of emerging statistical techniques requires input from experts in diverse disciplines. We highlight specific gaps in the literature where innovation in statistical methods is required.

Keywords Biomarkers · Environmental exposures · Epidemiology · Biostatistics · Children's health

Introduction

Biomarkers of environmental chemical exposures are widely used to assess etiologic relationships in perinatal epidemiology [1]. For example, there is great interest in understanding the effects of toxic chemicals, such as phthalates or polychlorinated biphenyls (PCBs), on pregnancy outcomes [2–5]. These effects can be studied using measurements from human biospecimens, such as blood or urine, which quantify the presence and the concentration of various toxic environmental agents. A key advantage of using biomarkers is that they unify diverse sources and routes of exposure to estimate the biological dose in the target tissue [6, 7]. They also allow

investigators to study multiple exposures simultaneously by using data taken from the same biospecimen [6, 7].

The analysis of biomarker data entails an extraordinary collection of statistical challenges [8–12]. A biomarker can describe a measurement in the causal sequence of events between an exposure to a hazardous factor and a health outcome [6]. However, a biomarker cannot be used to directly quantify the effects of interventions because the actual source and route of exposure are difficult to determine. Furthermore, attributing the cause of a health outcome to a single biomarker can be difficult because there are often dozens of biomarkers to take into consideration and they may be highly correlated. PCBs, for example, involve 209 congeners measured in blood plasma at very low levels [13]. Chemical concentrations measured in blood or urine samples from pregnant women may also be influenced by a variety of factors including dilution-dependent sample variation, censoring at low-levels due to detection limits, short chemical half-life, and variability due to physiology and metabolism.

This review aims to discuss recent developments in the statistical analysis of biomarkers with particular emphasis on

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

✉ Lawrence C. McCandless
lmccandl@sfu.ca

¹ Faculty of Health Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

chemical exposures during pregnancy. Given the breadth of the field, we limit our review to significant trends and developments related to two topics: (1) statistical methods for exposure assessment using biomarkers and (2) multivariable modeling of perinatal outcomes where multiple biomarkers are included as inputs in a linear regression model. Throughout, we focus on settings where the scientific objective is to estimate the confounder-adjusted association between individual biomarkers within a single chemical class and a continuous outcome.

Statistical Challenges in Exposure Assessment Using Biomarkers

The quality of exposure measurement plays a critical role in estimating the effects of environmental exposures. The use of biomarkers adds an additional layer of complexity [9, 10]. Errors in exposure measurement can introduce bias and uncertainty into estimates of health effects [14]. When working with biomarkers, the objective is to estimate the health effects of usual personal exposures over several months or during specific windows of vulnerability (e.g., trimesters of pregnancy) [15]. However, there are a myriad of data-analytic challenges that one may encounter. For example, non-persistent chemicals such as phthalates and bisphenol A tend to metabolize quickly and measured analyte concentrations from urine samples tend to reflect recent exposure [12].

In the following sections, we review three statistical challenges in exposure assessment that are specifically related to biomarkers. We note that there is a vast literature on biostatistical methods to correct for measurement error and misclassification [10, 14]. These approaches, including regression calibration and Bayesian approaches, model the relationship between the observed exposure and an unobserved true exposure, and they permit statistical adjustment for bias and uncertainty in the exposure-disease relation [10]. However, these methods are less commonly used in epidemiological studies with biomarkers of environmental chemical exposures because, in most settings, there is no unambiguous definition of the true exposure that can be measured objectively and then used to estimate the size of the measurement error [14].

Biomarkers That Fall Below the Limit of Detection

Biomarkers for low-level chemical exposure frequently fall below the minimum detectable capacity of the analytical instruments. These measurements, reported by many laboratories as less than the limits of detection (< LOD), are not usable in any statistical analysis. Substitution with a fixed value such as $LOD/2$ or $LOD/\sqrt{2}$ is the most common strategy for

handling the left-censored data. Deletion of all observations < LOD is also commonly seen. However, when the same value is substituted repeatedly or when data are not censored at random, then increased bias, decreased power and reduced variability will result [16–20]. Although deletion and substitution are widely used and easy to implement, the consensus in the literature is that they should be avoided when the frequency of non-detection exceeds 10% [18, 21].

In contrast, maximum likelihood estimation (MLE) is usually superior. It uses the uncensored data to maximize the likelihood function for the log-transformed censored data and iteratively produce estimates with low bias and high statistical power [17, 19, 20, 22–24]. MLE is the gold standard when the data follow a parametric distribution [17, 18]. However, MLE is not suitable for datasets with small sample size or non-parametric distributions. In these settings, multiple imputation (MI) is preferable [18, 22]. MI uses multiple datasets that are imputed individually to obtain a set of estimates that are combined to form a final estimate [18, 22, 25–27]. The number of recommended imputations is between five and ten [22, 28] and can be implemented easily with statistical software. However, the selection of a suitable imputation model may be complicated [22, 28, 29].

In general, there is no single method that is superior in all settings and researchers are encouraged to select methods that best fit their own scenarios [17]. As bias increases with the proportion of < LOD, it is sometimes recommended that a chemical be omitted completely from the analysis if the frequency of non-detection exceeds 20%. Furthermore, it is often incorrectly assumed that the measurement error in the reported values < LOD is greater than the errors in those above the LOD [26, 30]. It has been suggested that machine readings, which are the uncensored outputs from the analytical instruments, may be more accurate than the imputed values [18, 31]. Therefore, before choosing an imputation method, obtaining the machine readings is encouraged to better understand the measurement error process [32].

Adjustment for Lipid or Urine Dilution to Better Estimate Biologically Relevant Exposure

Variabilities in physiological processes can influence biomarker measurements in blood and urine samples. Lipophilic chemicals in blood samples require adjustment for lipid dilution [32, 33••], whereas urinary chemicals require creatinine or specific gravity adjustment to account for hydration level [34]. For example, when measuring PCBs, individuals with more body fat will have greater PCB concentrations compared to leaner individuals and, if uncorrected, this measurement error leads to an overestimation of PCB exposure levels in the target tissue [35]. Consequently, to estimate the

biologically relevant exposure, it is necessary to account for dilution-dependent sample variation.

Traditionally, lipid dilution is adjusted by dividing chemical concentration with serum lipid levels (e.g., $\mu\text{g/L}$ or $\mu\text{g/g}$ lipids) [34, 35], and urine diluteness is adjusted using either creatinine or specific gravity (SG) [36–38]. There is an ongoing debate of what the best adjustment methods are, but there has been no definitive answer that works in all settings [34]. O'Brien et al. [39••], using directed acyclic graphs and simulation studies, compared traditional adjustment methods with novel methods. They recommended adjusting for urinary samples by creating a “creatinine z-score,” which divides the chemical concentration by $C/C[\text{ratio}]$, which is the ratio of the observed creatinine level divided by the predicted value [39••]. An alternative option is to adjust for both the “creatinine z-score” and creatinine as covariates [39••]. As for serum samples, both the traditional adjustment method and the inclusion of lipids as covariate in the regression model are recommended [39••].

Further studies looking at the use of 2-stage models for creatinine have been proposed [40, 41], and other newer methods involve examining total lipid and creatinine levels across participants according to sociodemographic factors [42], biological media (e.g., blood or urine), and chemical characteristics (e.g., persistence, lipophilicity, or hydrophilicity) [43•]. Regardless of the adjustment method, it is important to keep in mind that it is logistically challenging to obtain tissue-specific measurements which could help better understand the toxicant distribution in the target tissues. Blood and urine samples, therefore, are proxies of how the chemicals affect inaccessible human tissues [33••, 39••]. Even after correcting for dilution differences, the exposure measurements may not reflect exposure in the target tissue.

Handling Repeated Measures of Biomarkers Within a Single Pregnancy

Biomarker exposures have traditionally been assessed using spot samples. Recent literature has revealed that this practice of assessing exposure at a single time point during pregnancy does not adequately represent the exposure throughout the whole pregnancy [44•, 45••, 46•]. Single exposure measurements typically ignore within-person variability and may lead to exposure misclassification [45••, 47] especially among non-persistent chemicals with low intraclass correlation coefficients (ICCs), such as BPAs and some phthalates where within-person variability is especially high [45••].

ICCs determine the reliability of the exposure measurements by calculating the ratio of between-person variability to total variability (e.g., between- plus within-person variabilities) over a period of time. For persistent chemicals such as polybrominated diphenyl ethers (PBDEs), where within-

person variability is small compared to between-person variability, the ICCs are close to 1 (e.g., 0.87 to 0.99), which indicates that single samples can reliably measure average exposures [48]. For chemicals with ICCs less than 0.6, such as BPAs and phthalates, it was found that as many as 35 samples per participant are needed to achieve adequate statistical power and limit bias [45••]. For these low ICCs, multiple measurements per participant collected over the course of the pregnancy are recommended [44•, 46•, 49]. However, the quality of exposure measurement also depends on the characteristics of the biological sample. For example, cotinine measured in hair reflects longer periods of exposure to secondhand smoke while serum cotinine reflects exposure in recent days [50].

Assaying multiple samples per participant results in higher assay costs. A simulation study by Perrier et al. [45••] showed that repeated measurements can be pooled before analysis with equal aliquots from each sample, and this reduces costs. Compared to using single measures, this cost-saving method saw improved power and bias [43•, 45••]. Alternatively, at the data analysis stage, repeated biomarkers measures can be summarized into a single measure of cumulative exposure [33••, 51]. Using the mean biomarker concentration across visits improves power to detect exposure-disease relationships even when exposure has poor stability over time [33••]. However, the mean concentration may not reflect exposure during biologically relevant periods, and any data that are missing not at random can bias the results. When an acute exposure is of interest, using the maximum rather than the mean exposure across visits is appropriate. It is important to note that the effects of the acute exposure represented by maximum exposure will also depend on its temporal relationship with an outcome. Recently, Chen et al. [33••] reviewed several statistical models for repeated measure of biomarkers in pregnant women including multivariable regression, regression-in-parallel and multiple informants modeling, two-stage methods, and clustering methods. There are also statistical approaches using measurement error models such as regression calibration and simulation extrapolation (SIMEX) to correct for exposure misclassification [10].

Besides providing a more accurate exposure assessment, repeated measurements can identify critical windows of vulnerability by analyzing changes in biomarker levels across pregnancy. The identification of critical windows is difficult statistically because the biomarkers are correlated within individuals, and this induces collinearity if the repeated measures are included in the outcome model [33••]. Furthermore, important time windows are mostly unknown and are highly dependent on the chemicals and outcomes studied. To determine differences in effects across time of exposure, while accounting for correlated exposures, several modeling approaches such as those examined by Chen et al. [33••] and Sanchez et al. [51] can be applied.

Statistical Challenges in Multivariable Modeling with Biomarkers of Environmental Chemical Exposure

When examining the relation between multiple biomarkers in the same chemical class and a health outcome, the critical issue is to determine the scientific objective of the analysis [12, 52, 53]. The investigator must carefully consider the causal questions, competing hypotheses, and the quantification and interpretation of evidence [52]. However, in the case of biomarkers of environmental chemical exposures the inferences about causality are primarily at the hypothesis-generating end of the research spectrum. For example, when examining the dose-response relationship over continuous levels of PCBs measured in blood, there are no clearly defined interventions that can directly modify biomarker levels in human tissue [7, 54]. Consequently, in the discussion that follows, we focus on statistical challenges where the goal is to estimate the confounder-adjusted association between individual biomarkers within a chemical class (e.g., PCBs) and a continuous outcome, where the dose-response is presumed to have a simple mathematical form (e.g., linear or monotonically increasing).

The most widely used analytic strategy is multiple linear regression where one or more of the biomarkers are included as predictor variables, along with confounders [55••]. Controlling for too many biomarkers can lead to data sparsity or multicollinearity, particularly when the number of predictors is large in relation to the sample size [56]. Consequently, researchers will seek to reduce the number of variables in the model. Biomarkers present a host of unique challenges in multivariable modeling because some analytes are chemically related. For example, phthalate metabolites in urine are often derived from the same parent compound and tend to be highly correlated. Three specific problems include the following: (1) accounting for heterogeneity in biomarker levels between chemicals within the same chemical grouping (e.g., PCBs), (2) variable selection and shrinkage techniques for biomarkers, and (3) the role of dimension reduction techniques such as summation of chemical concentrations.

Accounting for Heterogeneity in Biomarker Levels Between Chemicals Within the Same Class

When examining the association between several exposure biomarkers within a chemical class and a health outcome, it is desirable to report a measure of effect that conveys the magnitude of the association, per unit change in the biomarker. However, choosing a measure of effect that is easily interpreted is not simple. Consider, for example, the case of PCB153, which is more widely detected, and also more

variable between individuals, than PCB118 [1]. If the measured concentrations for both compounds are included directly in a multivariable model for the outcome, then the regression coefficient for PCB153 will be diminished relative to PCB118 merely because the exposure level for PCB153 is more heterogeneous in the study population. More generally, within any chemical class there are often dozens of analytes to take into consideration, many of which are detectable in fewer than 50% of the study subjects. Furthermore, different authors use different scales of measurement of exposure, such as $\mu\text{g/L}$ vs. ng/g lipid . Consequently, it is extremely difficult to assemble a collection of predictor variables for inclusion in a multivariable model.

It is common practice to log transform the biomarker concentrations, using base 2 or 10, before incorporating them into a multivariable model. In addition to limiting the influence of outliers, this approach also incorporates a non-linear dose-response between the biomarker concentration and the mean of the outcome variable, where the slope of the curve levels off at higher doses [5]. Nonetheless, the interpretation of the multivariable regression analysis results remains problematic. The regression coefficients correspond to changes in the mean of the outcome that are associated with a 2-fold (or 10-fold) increase in the level of the biomarker. This is difficult to interpret without referring to percentiles of the concentration on the original scale. For example, a 10-fold increase in concentration from the median may entail extrapolation of effects outside of the range of biologically plausible exposure levels.

A different strategy is to put the biomarkers onto a common scale by subtracting the mean and dividing by the standard deviation [57]. Subtracting the mean typically improves the interpretation of main effects in the presence of interactions [58]. The resulting regression coefficient is the expected difference in the outcome comparing participants that differ by one standard deviation in the input variable, conditional on the remaining predictors. More generally, scaling predictors is widely used in combination with shrinkage methods, such as lasso regression, because it can dramatically affect regression coefficient estimation [59]. However, scaling input variables using the standard deviation has been criticized because the resulting scale of measurement is not transportable, meaning that the regression coefficients cannot be compared between studies because they depend on the distribution of exposure, which is sensitive to arbitrary features of the study design [60, 61]. Furthermore, right-skewness of biomarker concentrations tends to distort the standard deviation to render it meaningless as a measure of variability.

We recommend two alternative approaches to account for heterogeneity in biomarker levels: either discretizing the predictor into quartiles or tertiles [62, 63], or rescaling the biomarker concentrations using a standard reference such as an interquartile range (IQR). The standard reference

should be expressed in natural units of the exposure, which should be established based on the distribution of the exposure in the target population. Ideally, it should correspond to levels of exposure where biologically meaningful changes in the outcome are anticipated. Discretization has been criticized in the statistical literature [64] because the presumed model for the relationship between the predictor and the outcome is typically unrealistic, which can introduce bias. However, discretization has the advantage that it balances the tradeoff between simplicity and interpretability versus proper fit [65].

Variable Selection and Shrinkage Techniques for Biomarkers Within the Same Chemical Class

Variable selection and model building using biomarkers of environmental chemical exposure is uniquely challenging [9]. Consider, for example, the case of PCBs, which can be measured by as many as 209 unique congeners in blood. In populations not subject to widespread pollution, most PCBs are detectable in fewer than 50% of the study participants. On the one hand, we could attempt to impute low concentrations. However, the low detection rates raise questions about which biomarkers should be incorporated into the model. Many investigators discard all but the most widely detected congeners and retain those that meet a particular detection threshold (e.g., 20%). Other authors limit their focus entirely to PCB153 based on the argument that it is correlated with other PCBs and with PCB summary metrics [13]. Discarding some PCBs is reasonable because, individually, we do not expect low-level exposures to predict measurable variation in the outcome. On the other hand, all PCBs reveal information about cumulative PCB exposure, and some PCBs may be toxic even at very low levels. The health effects of low-level chemical exposures during pregnancy are mostly unknown [66]. Therefore, new research is required to support multivariable modeling of high-dimensional low-level biomarkers [9, 67]. One important direction of research is to develop novel MI techniques for joint imputation of multiple correlated biomarkers that fall below the LOD and to characterize settings in which such methods will be most useful [67].

A related issue in multivariable modeling concerns the adjustment for co-pollutant confounding [55••, 68••, 69••]. Consider the case of PCB153 and PCB180, which are both widely detected in human populations [13]. The biomarkers are correlated because the congeners were manufactured together as commercial PCB mixtures. The investigators must decide whether both compounds should be included into the same multivariable model for a health outcome or not. If both PCBs are included, the regression coefficients become harder to interpret etiologically because they describe

changes in the mean response conditional on the respective PCBs. However, PCB153 and PCB180 do not vary independently and the standard error of the regression coefficients may increase dramatically due to collinearity. Conversely, if both PCBs are not included in the model, then the regression coefficients will be confounded by one another [69••]. It is likely that etiologically relevant parameters that describe the causal effects of individual chemicals can be recovered via prediction from a model that incorporates multiple biomarkers from the same class, for example, by using the parametric g-formula [58, 70, 71].

One widely used analytic strategy is to combine variable selection with shrinkage methods, such as Bayesian hierarchical modeling [72, 73], the lasso or elastic net [55••, 74, 75]. Bayesian methods use probability distributions to model uncertainty. These distributions, called prior distributions, are combined with the data to obtain the posterior distribution for model parameters. Bayesian hierarchical models generalize ordinary regression to distinguish multiple levels of information in a model. The hierarchical structure introduces shrinkage which reduces the regression coefficients towards zero or towards a group-specific mean parameter. Introducing bias is typically not desirable; however, shrinkage methods also reduce the standard error, and consequently, the overall mean squared error of estimation. Thus, shrinkage methods are particularly useful in the presence of collinearity because they stabilize parameters estimates and reduce prediction error of the outcome variable. As a result of these advantages and their usefulness in settings with multiple correlated exposures, shrinkage methods and Bayesian techniques have been used to analyze biomarkers [3, 76–80].

Dimension Reduction Methods for Biomarkers Within the Same Chemical Class, Including the Sum-Of-Chemical Approach

Data from highly correlated exposures can be combined using dimension reduction techniques to reduce the number of variables [74]. The idea is to create linear combinations of biomarkers that serve as lower dimensional summaries, and these can be included as predictors in the outcome model. Rather than looking at the effects of individual biomarkers, dimension reduction techniques address a different scientific objective, which is to examine the association between a broad class of chemical exposures and a health outcome [11]. One of the most common examples is to sum the individual biomarker concentrations [13], or alternatively, to sum the concentrations after dividing by the molar mass (see [81] and [5] for examples related to PCBs and phthalates). This approach is straightforward and intuitively appealing; however, it does not account for the potency of the individual biomarkers. If a

biomarker does not affect the outcome, then including it in the sum will introduce a measurement error in the estimated dose-response relationship between the summation and the outcome. Alternatively, sums can be created by weighting each biomarker by their toxicity, such as toxic equivalency factors (TEFs), which describe dioxin-related toxicity for PCBs [82]. It is also possible to use data-driven approaches to create lower dimensional summary variables based on relationships in the data. Examples include principal component analysis (PCA), partial least squares, cluster analysis, weighted quantile sum regression [83], environmental risk scores [84], and structural equation models [85] that use a latent variable for the true unobservable exposure.

However, a limitation of dimension reduction techniques is that the biological interpretation of the analysis results is challenging. A unit change in the low-dimensional summary (e.g., +1 IQR) need not correspond to changes in more distal exposures (see [15] for an example related to household dust). When analyzing biomarkers of environmental chemical exposure, there are two distinct scientific objectives: estimating the health effects of individual chemicals, and estimating the health effect of cumulative exposure to multiple chemicals. Dimension reduction techniques for biomarkers of environmental chemical exposures lie squarely within the second objective. Further research is needed to bridge the gap between dimension reduction techniques and the growing literature on causal inference.

Conclusions

We have highlighted several unique challenges in the analysis of biomarkers of environmental chemical exposures with particular focus on exposure assessment and multivariable

modeling. Please refer to Table 1 for key references listed according to the statistical challenges we have identified. Although our emphasis was on exposures during pregnancy, the challenges are also of relevant to epidemiologists working in other areas of environmental health. Given the broad scope of the review, it is inevitable that some topics are omitted. We did not discuss the literature on uncertainty quantification for high-dimensional data (e.g., multiple testing) or machine learning and non-linear modeling techniques, and we refer the reader to the textbooks [64, 74] for further details. Additionally, we fully acknowledge having sidestepped a detailed discussion of causal inference using biomarkers. We focused on settings where the scientific objective is to estimate the confounder-adjusted association between individual biomarkers within a chemical class and a health outcome. This is consistent with the usual custom in epidemiology of exercising caution in using language about causality [86]. When analyzing biomarkers, the topic of cause and effect relationships inevitably drifts to formulating difficult-to-answer questions such as “Had their measured biomarker levels been different, would the pregnancy outcome have been different?” [7]. Ignoring causality has the disadvantage that it does not shed light on the health effects of specific interventions [15]. However, it should be emphasized that association studies have the advantage that they can reveal specific biological mechanisms [15]. This contributes to the overall body of evidence linking chemical exposures and health outcomes. More generally, there is a vast literature on statistical methods for causal inference [87]. Much of this work has been in applied settings with a binary treatment comparison [88]. There is also an emerging body of work about the analysis of chemical mixtures [89, 90]. New research is needed to clearly articulate the definitions of causal contrasts that are relevant to the study of biomarkers of environmental chemical exposures.

Table 1 Key references according to statistical challenges

Topics	Statistical challenges	Key references
Exposure assessment	Biomarkers below the LOD	Whitcomb et al. 2008 [32] Helsel 2005 [18]
	Dilution-dependent sample variation	O’Brien et al. 2016 [39••]
	Repeated measures during pregnancy	Perrier et al. 2016 [45••] Chen et al. 2015 [33••]
Multivariable modeling	Heterogeneity in biomarker levels	Greenland et al. 1991 [60] James et al. 2013 [74]
	Variable selection and shrinkage techniques	Patel 2017 [68••] Agier et al. 2016 [55••] Weisskopf et al. 2018 [69••]
	Dimension reduction and sum-of-chemical approach	James et al. 2013 [74] Carrico et al. 2014 [83]

Acknowledgements The authors gratefully acknowledge input and thoughtful comments from Tim Takaro and Ryan Allen on an earlier version of the manuscript.

Author's Contributions JH and LM drafted the manuscript with input from HZ and BB. All authors approved the final version of the manuscript.

Funding Information This project was funded by a Catalyst Grant from the Canadian Institutes for Health Research (L-CIP-150736).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

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

- Of importance
- Of major importance

1. Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003–2004. *Environ Health Perspect*. 2011;119:878–85.
2. Govarts E, Nieuwenhuijsen M, Schoeters G, Ballester F, Bloemen K, De Boer M, et al. Birth weight and prenatal exposure to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE): a meta-analysis within 12 European Birth Cohorts. *Environ Health Perspect*. National Institute of Environmental Health Sciences. 2012;120:162.
3. Lenters V, Portengen L, Rignell-Hydbom A, Jönsson BAG, Lindh CH, Piersma AH, et al. Prenatal phthalate, perfluoroalkyl acid, and organochlorine exposures and term birth weight in three birth cohorts: multi-pollutant models based on elastic net regression. *Environ Health Perspect*. 2016;124:365–72.
4. Rauch SA, Braun JM, Barr DB, Calafat AM, Khoury J, Montesano AM, et al. Associations of prenatal exposure to organophosphate pesticide metabolites with gestational age and birth weight. *Environ Health Perspect*. 2012;120:1055–60.
5. Shoaff JR, Romano ME, Yolton K, Lanphear BP, Calafat AM, Braun JM. Prenatal phthalate exposure and infant size at birth and gestational duration. *Environ Res*. 2016;150:52–8.
6. Savitz DA, Wellenius GA. Exposure biomarkers indicate more than just exposure. *Am J Epidemiol* 2017; Available: <https://academic.oup.com/aje/advance-article-abstract/doi/10.1093/aje/kwx333/4636591>
7. Savitz DA. Invited commentary: interpreting associations between exposure biomarkers and pregnancy outcome. *Am J Epidemiol*. 2014;179:545–7.
8. Albert PS, Schisterman EF. Novel statistical methodology for analyzing longitudinal biomarker data. *Stat Med*. 2012;31:2457–60.
9. Schisterman EF, Albert PS. The biomarker revolution. *Stat Med*. 2012;31:2513–5.
10. Thomas DC. *Statistical methods in environmental epidemiology*: Oxford University Press; 2009.
11. Braun JM, Gennings C, Hauser R, Webster TF. What can epidemiological studies tell us about the impact of chemical mixtures on human health? *Environ Health Perspect*. 2016;124:A6–9.
12. Braun JM, Gray K. Challenges to studying the health effects of early life environmental chemical exposures on children's health. *PLoS Biol Public Library of Science*. 2017;15:e2002800.
13. Axelrad DA, Goodman S, Woodruff TJ. PCB body burdens in US women of childbearing age 2001–2002: an evaluation of alternate summary metrics of NHANES data. *Environ Res*. 2009;109:368–78.
14. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd. Philadelphia: Lippincott Williams & Wilkins. 2008;
15. Weisskopf MG, Webster TF. Trade-offs of personal versus more proxy exposure measures in environmental epidemiology. *Epidemiology*. 2017;28: 635–643.
16. Nie L, Chu H, Liu C, Cole SR, Vexler A, Schisterman EF. Linear regression with an independent variable subject to a detection limit. *Epidemiology*. 2010;21(Suppl 4):S17–24.
17. Hewett P, Ganser GH. A comparison of several methods for analyzing censored data. *Ann Occup Hyg* 2007;51: 611–632.
18. Helsel DR. More than obvious: better methods for interpreting nondetect data. *Environ Sci Technol*. 2005;39:419A–23A.
19. Domthong U, Parikh CR, Kimmel PL, Chinchilli VM. Assessment, Serial Evaluation, and Subsequent Sequelae of Acute Kidney Injury Consortium. Assessing the agreement of biomarker data in the presence of left-censoring. *BMC Nephrol*. 2014;15:144.
20. Lafleur B, Lee W, Billheimer D, Lockhart C, Liu J, Merchant N. Statistical methods for assays with limits of detection: serum bile acid as a differentiator between patients with normal colons, adenomas, and colorectal cancer. *J Carcinog*. 2011;10:12.
21. Lubin JH, Colt JS, Camann D, Davis S, Cerhan JR, Severson RK, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect*. 2004;112:1691–6.
22. Lynn HS. Maximum likelihood inference for left-censored HIV RNA data. *Stat Med*. 2001;20:33–45.
23. Jain RB, Caudill SP, Wang RY, Monsell E. Evaluation of maximum likelihood procedures to estimate left censored observations. *Anal Chem*. 2008;80:1124–32.
24. Huynh T, Ramachandran G, Banerjee S, Monteiro J, Stenzel M, Sandler DP, et al. Comparison of methods for analyzing left-censored occupational exposure data. *Ann Occup Hyg*. 2014;58: 1126–42.
25. Chen H, Quandt SA, Grzywacz JG, Arcury TA. A Bayesian multiple imputation method for handling longitudinal pesticide data with values below the limit of detection. *Environmetrics*. 2013;24:132–42.
26. Harel O, Perkins N, Schisterman EF. The use of multiple imputation for data subject to limits of detection. *Sri Lankan J Appl Stat*. 2014;5:227–46.
27. Rubin D. B.(1987) *Multiple imputation for nonresponse in survey*. New York: Wiley. New York: Wiley;
28. Schafer JL. *Multiple imputation: a primer*. *Stat Methods Med Res*. 1999;8:3–15.
29. Harel O, Zhou X-H. Multiple imputation: review of theory, implementation and software. *Stat Med*. 2007;26:3057–77.
30. Guo Y, Harel O, Little RJ. How well quantified is the limit of quantification? *Epidemiology*. 2010;21(Suppl 4):S10–6.
31. Succop PA, Clark S, Chen M, Galke W. Imputation of data values that are less than a detection limit. *J Occup Environ Hyg*. 2004;1: 436–41.
32. Whitcomb BW, Schisterman EF. Assays with lower detection limits: implications for epidemiological investigations. *Paediatr Perinat Epidemiol*. 2008;22:597–602.
- 33.•• Chen Y-H, Ferguson KK, Meeker JD, McElrath TF, Mukherjee B. Statistical methods for modeling repeated measures of maternal environmental exposure biomarkers during pregnancy in

- association with preterm birth. *Environ Health*. 2015;14:9. **This paper compared nine statistical models that are useful in assessing exposure and examining the dose-response relationship between repeated measures of maternal environmental exposure biomarkers and preterm birth.**
34. Schisterman EF, Whitcomb BW, Louis GMB, Louis TA. Lipid adjustment in the analysis of environmental contaminants and human health risks. *Environ Health Perspect*. 2005;113:853–7.
 35. Phillips DL, Pirkle JL, Burse VW, Bernert JT Jr, Henderson LO, Needham LL. Chlorinated hydrocarbon levels in human serum: effects of fasting and feeding. *Arch Environ Contam Toxicol*. 1989;18:495–500.
 36. Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, Pirkle JL. Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements. *Environ Health Perspect*. 2005;113:192–200.
 37. Meeker JD, Hu H, Cantonwine DE, Lamadrid-Figueroa H, Calafat AM, Ettinger AS, et al. Urinary phthalate metabolites in relation to preterm birth in Mexico city. *Environ Health Perspect* ncbnlmnhgov. 2009;117:1587–92.
 38. Holland N, Huen K, Tran V, Street K, Nguyen B, Bradman A, et al. Urinary phthalate metabolites and biomarkers of oxidative stress in a Mexican-American cohort: variability in early and late pregnancy. *Toxics*. 2016;4 <https://doi.org/10.3390/toxics4010007>.
 39. O'Brien KM, Upson K, Cook NR, Weinberg CR. Environmental chemicals in urine and blood: improving methods for creatinine and lipid adjustment. *Environ Health Perspect*. 2016;124:220–7. **Using directed acyclic graphs and simulations, this study compared adjustment methods for dilution-dependent creatinine and serum lipids and recommended the novel method for urine dilution and traditional method for serum lipids biomarkers.**
 40. Jain RB. Single-stage versus two-stage models to estimate creatinine corrected urinary analyte concentrations. *Toxicol Environ Chem Taylor & Francis*. 2017;99:710–29.
 41. Hunter DJ, Hankinson SE, Laden F, Colditz GA, Manson JE, Willett WC, et al. Plasma organochlorine levels and the risk of breast cancer. *N Engl J Med Mass Medical Soc*. 1997;337:1253–8.
 42. Costanza MC, Cayanis E, Ross BM, Flaherty MS, Alvin GB, Das K, Morabia A. Relative contributions of genes, environment, and interactions to blood lipid concentrations in a general adult population. *Am J Epidemiol academicoupcom*; 2005;161: 714–724.
 43. Calafat AM. Contemporary issues in exposure assessment using biomonitoring. *Curr Epidemiol Rep*. 2016;3:145–53. **This review described factors that affect biomarkers of exposure such as biomarker selection, variability in biomarker concentrations, biomarker collection, and storage issues.**
 44. Fisher M, Arbuckle TE, Mallick R, LeBlanc A, Hauser R, Feeley M, et al. Bisphenol A and phthalate metabolite urinary concentrations: daily and across pregnancy variability. *J Expo Sci Environ Epidemiol*. 2015;25:231–9. **This study assessed chemical variability in urine sample over time using intraclass correlation coefficients (ICCs) and showed that multiple biomarker of exposure measurements collected throughout the pregnancy better represent average exposures across pregnancy compared to single spot measure.**
 45. Perrier F, Giorgis-Allemand L, Slama R, Philippat C. Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. *Epidemiology*. 2016;27:378–88. **This study, using simulation, showed that as intraclass correlation coefficients (ICCs) of chemicals decrease, the number of biospecimens needed to limit bias increases and that within-subject pooling of biospecimens before assaying the chemical can reduce exposure misclassification without increasing assay costs.**
 46. Vernet C, Philippat C, Calafat AM, Ye X, Lyon-Caen S, Siroux V, et al. Within-day, between-day, and between-week variability of urinary concentrations of phenol biomarkers in pregnant women. *Environ Health Perspect*. 2018;126:037005. **This study showed that, during pregnancy, phenol biomarkers have a strong within-day variability and that one biospecimen does not represent the average exposure of phenol throughout the pregnancy.**
 47. Ferguson KK, O'Neill MS, Meeker JD. Environmental contaminant exposures and preterm birth: a comprehensive review. *J Toxicol Environ Health B Crit Rev*. 2013;16:69–113.
 48. Makey CM, McClean MD, Sjödin A, Weinberg J, Carignan CC, Webster TF. Temporal variability of polybrominated diphenyl ether (PBDE) serum concentrations over one year. *Environ Sci Technol*. 2014;48:14642–9.
 49. Frederiksen H, Kranich SK, Jørgensen N, Taboureaux O, Petersen JH, Andersson A-M. Temporal variability in urinary phthalate metabolite excretion based on spot, morning, and 24-h urine samples: considerations for epidemiological studies. *Environ Sci Technol*. 2013;47:958–67.
 50. Kalkbrenner AE, Hornung RW, Bernert JT, Hammond SK, Braun JM, Lanphear BP. Determinants of serum cotinine and hair cotinine as biomarkers of childhood secondhand smoke exposure. *J Expo Sci Environ Epidemiol*. 2010;20:615–24.
 51. Sánchez BN, Hu H, Litman HJ, Téllez-Rojo MM. Statistical methods to study timing of vulnerability with sparsely sampled data on environmental toxicants. *Environ Health Perspect*. 2011;119: 409–15.
 52. Zeger SL. Invited commentary: epidemiologic studies of the health associations of environmental exposures with preterm birth. *Am J Epidemiol* 2012;175: 108–110; discussion 111–3.
 53. Dominici F, Zigler C. Best practices for gauging evidence of causality in air pollution epidemiology. *Am J Epidemiol*. 2017;186: 1303–9.
 54. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes*. 2008;32(Suppl 3):S8–14.
 55. Agier L, Portengen L, Chadeau-Hyam M, Basagaña X, Giorgis-Allemand L, Siroux V, et al. A systematic comparison of linear regression-based statistical methods to assess exposome-health associations. *Environ Health Perspect*. 2016;124:1848–56. **This simulation study evaluated the performance of various linear regression-based statistical methods in exposome studies where it is difficult to discern between true predictors and correlated exposures.**
 56. Greenland S, Pearce N. Statistical foundations for model-based adjustments. *Annu Rev Public Health*. 2015;36:89–108.
 57. Gelman A. Scaling regression inputs by dividing by two standard deviations. *Stat Med*. 2008;27:2865–73.
 58. Gelman A, Pardoe I. 2. Average predictive comparisons for models with nonlinearity, interactions, and variance components. *Sociol Methodol*. 2007;37:23–51.
 59. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal Int Soc Bayesian Anal*. 2006;1:515–34.
 60. Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. Standardized regression coefficients: a further critique and review of some alternatives. *Epidemiology*. 1991;2:387–92.
 61. Greenland S, Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *Am J Epidemiol*. 1986;123:203–8.
 62. Gelman A, Park DK. Splitting a predictor at the upper quarter or third and the lower quarter or third. *Am Stat Taylor Francis*. 2009;63:1–8.
 63. Rothman KJ. Six persistent research misconceptions. *J Gen Intern Med*. 2014;29:1060–4.

64. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer Science & Business Media; 2013.
65. Szklo M, Nieto FJ, Miller D. Epidemiology: beyond the basics. Am J Epidemiol Oxford University Press. 2001;153:821–2.
66. Di Renzo GC, Conry JA, Blake J, DeFrancesco MS, DeNicola N, Martin JN Jr, et al. International Federation of Gynecology and Obstetrics opinion on reproductive health impacts of exposure to toxic environmental chemicals. *Int J Gynaecol Obstet.* 2015;131: 219–25.
67. Schisterman EF, Little RJ. Opening the black box of biomarker measurement error. *Epidemiology* 2010;21 Suppl 4: S1–S3.
- 68.●● Patel CJ. Analytic complexity and challenges in identifying mixtures of exposures associated with phenotypes in the exposome era. *Curr Epidemiol Rep.* 2017;4:22–30. **This review described the challenges in identifying co-occurring exposures and discussed machine-learning and data analytics methods useful in narrowing down correlated exposures.**
- 69.●● Weisskopf MG, Seals RM, Webster TF. Bias amplification in epidemiologic analysis of exposure to mixtures. *Environ Health Perspect.* 2018;126:047003. **This study found that bias increases when outcome model included highly correlated mixtures or more mixture components.**
70. Liu J, Gustafson P. On average predictive comparisons and interactions. *Int Stat Rev.* 2008;76:419–32.
71. Keil AP, Daza EJ, Engel SM, Buckley JP, Edwards JK. A Bayesian approach to the g-formula. *Stat Methods Med Res.* 2017;962280217694665
72. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. *Environ Health Perspect.* 1994;102(Suppl 8): 33–9.
73. Greenland S. When should epidemiologic regressions use random coefficients? *Biometrics.* 2000;56:915–21.
74. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. Springer New York; 2013.
75. Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. *Environ Health.* 2013;12:85.
76. MacLehose RF, Dunson DB, Herring AH, Hoppin JA. Bayesian methods for highly correlated exposure data. *Epidemiology.* 2007;18:199–207.
77. Herring AH. Nonparametric bayes shrinkage for assessing exposures to mixtures subject to limits of detection. *Epidemiology.* 2010;21(Suppl 4):S71–6.
78. Woods MM, Lanphear BP, Braun JM, McCandless LC. Gestational exposure to endocrine disrupting chemicals in relation to infant birth weight: a Bayesian analysis of the HOME Study. *Environ Health.* 2017;16:115.
79. Ashley-Martin J, Dodds L, Arbuckle TE, Bouchard MF, Fisher M, Morriset A-S, et al. Maternal concentrations of perfluoroalkyl substances and fetal markers of metabolic function and birth weight. *Am J Epidemiol.* 2017;185:185–93.
80. Buckley JP, Engel SM, Braun JM, Whyatt RM, Daniels JL, Mendez MA, et al. Prenatal phthalate exposures and body mass index among 4- to 7-year-old children: a pooled analysis. *Epidemiology.* 2016;27:449–58.
81. Sagiv SK, Tolbert PE, Altshul LM, Korrick SA. Organochlorine exposures during pregnancy and infant size at birth. *Epidemiology.* 2007;18:120–9.
82. Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M, et al. The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicol Sci.* 2006;93:223–41.
83. Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. *JABES Springer US.* 2014;20:100–20.
84. Park SK, Tao Y, Meeker JD, Harlow SD, Mukherjee B. Environmental risk score as a new tool to examine multi-pollutants in epidemiologic research: an example from the NHANES study using serum lipid levels. *PLoS One.* 2014;9: e98632.
85. Petit C, Blangiardo M, Richardson S, Coquet F, Chevrier C, Cordier S. Association of environmental insecticide exposure and fetal growth with a Bayesian model including multiple exposure sources: the PELAGIE mother-child cohort. *Am J Epidemiol.* 2012;175: 1182–90.
86. Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health.* 2018;108:616–9.
87. Hernán MA, Robins JM. Causal inference. CRC Boca Raton, FL.; 2010.
88. Hernán MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health BMJ Publishing Group Ltd.* 2004;58:265–71.
89. Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: lessons from an innovative workshop. *Environ Health Perspect.* 2016;124: A227–9.
90. Billionnet C, Sherrill D, Annesi-Maesano I, study GERIE. Estimating the health effects of exposure to multi-pollutant mixture. *Ann Epidemiol.* 2012;22:126–41.