A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP14942.

Partial Effects in Environmental Mixtures: Evidence and Guidance on Methods and Implications

Maria E. Kamenetsky,[1] Barrett M. Welch,[2] Paige A. Bommarito,[3] Jessie P. Buckley,[4] Katie M. O'Brien,[3] Alexandra J. White,[3] Thomas F. McElrath,[5] David E. Cantonwine,[5] Kelly K. Ferguson,[3] and Alexander P. Keil[1]

[1]Occupational and Environmental Epidemiology, Division of Cancer Epidemiology & Genetics, National Cancer Institute, Rockville, Maryland, USA
[2]School of Public Health, University of Nevada, Reno, Reno, Nevada, USA
[3]Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
[4]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA
[5]Division of Maternal-Fetal Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Background: The effects of a mixture of exposures on health outcomes are of interest to public health but pose methodological hurdles. These exposures may impact the outcome in opposing ways, which we call the positive and negative partial effects of a mixture. There has been growing interest in estimating these partial effects and their ability to inform public health interventions.

Objectives: Methods like quantile g-computation (QGC) and weighted quantile sums regression (WQSr) were originally developed for estimating an overall mixture effect. These approaches, however, have not been comprehensively evaluated in their ability to estimate partial effects. We study the bias–variance tradeoffs of these approaches in estimating partial effects.

Methods: We compare QGC with sample-splitting (QGCSS) and WQSr with sample-splitting (WQSSS) and new methods including *a*) QGC *a priori* (QGCAP) and WQS *a priori* (WQSAP), which use prior knowledge to determine the positive and negative exposures prior to partial effects estimation; *b*) model-averaging (QGC-MA); and *c*) elastic net to determine the split (QGC-Enet). We also considered WQSr with no sample-splitting (WQSNS), repeated holdout sets (RH-WQS), and two-index model with penalized weights (WQS2i). We compared performance under *a*) exposure correlations, *b*) varying sample sizes, *c*) spread in the negative effect across exposures, and *d*) imbalance in the partial effects.

Results: Our simulation results showed that the estimation of negative and positive partial effects grows in root mean squared error and average bias as correlation among exposures increases, sample sizes shrink, the negative effect is spread over more exposures, or the imbalance between the negative and positive effects increases. Our results are demonstrated in two examples of mixtures in relation to oxidative stress biomarkers and telomere length.

Discussion: Outside of having *a priori* knowledge, no method is optimally reliable for estimating partial effects across common exposure scenarios. We provide guidance for practitioners of when partial effects might be most accurately estimated under particular settings. https://doi.org/10.1289/EHP14942

Introduction

In public health, we are often interested in understanding how exposures affect human health for the purpose of identifying possible interventions. Specifically for environmental exposures, these interventions may come in a form which can affect multiple co-occuring exposures, either due to common sources standards that target multiple pollutants simultaneously. We call these co-occuring exposures a "mixture" and often we estimate the overall effect of the mixture on human health. Environmental mixtures are complex: The constituent exposures are often highly correlated, may be high-dimensional, may have complex interactions and dose–response patterns, and can vary in the direction of effects on study outcomes. To estimate the effect of mixtures in public health, many approaches have been developed,[1–5] which have focused on estimating the overall effect of the mixture.

Quantile g-computation[6] (QGC) and weighted quantile sums regression (WQSr)[7] are two approaches commonly used for the estimation of the overall effect of a mixture. This overall measure represents the effect of a mixture as a whole, and these two approaches converge to the same answer in some circumstances.[6] As defined here, an overall effect can be partitioned into partial effects. We focus on one particular partitioning: the negative and positive partial effects. A positive (or negative) partial effect for a health outcome is defined as the additive and linear effect of all exposures with independent effects in the direction of increased (or decreased) risk. Positive and negative partial effect estimation necessarily assumes linearity and additivity and, thus, should be estimated with models that can leverage those assumptions when appropriate. Many studies of mixtures have used QGC[8–13] and WQSr[14–17] to estimate some type of partial effect. While some studies, such as Keil et al.[6] have examined how well these methods perform at estimating overall effects,[6,18] there have been no rigorous studies dedicated to exploring how well these methods estimate partial effects. It is unknown how these approaches perform under the range of scenarios in which environmental mixtures occur, where such studies vary with respect to exposure correlation, sample sizes, and the relative balance of exposures with negative vs. partial effects. This research gap means it is unclear how well methods for estimating partial effects can be used to accurately guide interventions based on mixtures results.

In this study, we aim to characterize the bias and accuracy of QGC and WQSr and their extensions under various settings using simulation studies, highlighting key challenges. We demonstrate partial effect estimation using two epidemiologic examples for which there is reasonable *a priori* knowledge regarding specific exposures and associations with a health outcome. First, we explore the association of urinary paraben, phthalate, and phenol concentrations on inflammation-related biomarkers[19–23] using data from an ongoing pregnancy cohort, LIFECODES. Exposure to these comes from many routes including use of personal care

Address correspondence to Maria E. Kamenetsky. Email: maria.kamenetsky@nih.gov

Supplemental Material is available online (https://doi.org/10.1289/EHP14942).

The authors declare they have no conflicts of interest related to this work to disclose.

Received 7 March 2024; Revised 14 March 2025; Accepted 20 March 2025; Published 9 May 2025.

products[22,24,25] and through food and drinking water.[11] Urinary levels of the lipid 8-isoprostane during pregnancy has been associated with adverse birth outcomes, thought to be driven by inflammation and oxidative stress.[26] Our approach aims to evaluate partial effects where we expect an effect only in the positive direction. Second, we used data from the 2001–2002 cycle of the National Health and Nutrition Examination Survey (NHANES)[27] to investigate the association of persistent organic pollutant concentrations in blood on average leukocytic telomere length. Telomeres (noncoding segments of DNA) located at the end of chromosomes are shortened with cell division.[28] Polychlorinated biphenyls (PCBs), dioxins, and furans have previously been investigated in their role in telomere length.[28–30] Previous work has hypothesized negative relationships between telomere length and nondioxin-like PCBs[31] and positive relationships with dioxin-like PCBs and furans.[28,29] We summarize our results and provide guidance to practitioners on the use of partial effects estimation in environmental epidemiology.

Methods

Simulation Methods

Overall and partial mixture effects. Overall effects allow us to understand how to intervene to improve public health.[6] Consider a mixture consisting of four exposures, x_1, x_2, x_3, x_4. Suppose that increases in x_1 and x_3 cause the health outcome to decrease, and increases in x_2 and x_4 cause the health outcome to increase. An overall effect of the mixture can be given by contrasts of the health outcome that would be expected comparing two different scenarios in which all exposures differ (e.g., "exposed to all exposures vs. unexposed" is a comparison of $x_1, x_2, x_3, x_4 = 1$ vs. $x_1, x_2, x_3, x_4 = 0$). Environmental settings typically consider an "overall effect" to refer to a joint exposure–response, given by the parameter ψ, which quantifies the expected difference in the outcome that would be observed if all exposures selected into the mixture were increased by one unit. It has previously been shown that under linearity and additivity, the overall effect of a mixture can be estimated as the sum of independent effects using QGC or in large samples assuming directional homogeneity using WQSr.[6] Both methods can also be used, to various extents, to estimate overall effects under nonlinearity and nonadditivity.

A partial effect within the mixture can be considered as an effect of changing only a subset of exposures, while holding other exposures constant. A positive (or negative) partial effect is an effect of increasing only the exposures that increase (or decrease) the expected value of the outcome (e.g., the positive partial effect is the partial effect of x_1 and x_3, given that both exposures increase risk). Notably, positive and negative partial effects must assume linearity and additivity, which makes QGC and WQSr suitable for assessing partial effects. WQSr was originally considered as an approach to estimate effects "in the direction of increased risk,"[7] while use of QGC for estimating positive and negative partial effects is more recent.[13]

QGC. QGC is a regression method used for estimating the effects of a mixture, based on the g-computation estimator from the causal effect estimation literature.[32,33] QGC estimates the effect of simultaneously increasing all exposures by one quantile. While QGC allows nonlinear and nonadditive effects of exposure, under linearity and additivity, QGC simplifies to a generalized linear model on the quantized exposures (with P confounders) given by

$$g(\widehat{\mu}_i) = \widehat{\beta}_0 + \sum_{j=1}^{J} \widehat{\beta}_j x_{ij}^q + \sum_{p=1}^{P} \widehat{\theta}_p u_{ip}, \tag{1}$$

where $g(\cdot)$ is a generalized linear model link function, $\widehat{\mu}_i$ is the mean outcome for participant i, $\widehat{\beta}_0$ is the estimated intercept, x_{ij}^q is the quantized exposure j for participant i, $\widehat{\theta}_p$ are the estimated regression parameters for those covariates, and u_{ip} are covariates measured for each participant i. Under linearity and additivity, $\widehat{\psi} = \sum_{j=1}^{J} \widehat{\beta}_j$ is the estimate of the overall effect of the mixture. QGC also reports weights w_j for each exposure j in the mixture, which are given by $\widehat{w}_j = \widehat{\beta}_j / \sum_{j=1}^{J} \widehat{\beta}_j$ (when all β_j have the same sign, otherwise there are sets of positive and negative weights).

WQSr. WQSr develops an index which is a weighted average of all exposures in the mixture after each exposure has been transformed into a categorical variable defined by its quantiles. The weights for each exposure, $\widehat{w}_j$, are estimated either using bootstrap samples[7] or using random subsets.[34] The overall mixture effect, under directional homogeneity, is estimated by $\widehat{\psi}$, the coefficient for the index exposure in an adjusted model for the outcome. While WQSr can be approached numerous ways, the default approach in the R (version 4.2.2; R Development Core Team) package "gWQS"[35] uses the following steps: First, define the effect direction of interest. Second, assign each exposure an ordinal "score" according to quantiles of that exposure. Next, an optimization algorithm[36] is used to repeatedly fit a (possibly constrained) generalized linear model to bootstrap samples from the data, given all exposures and covariates. This step yields weights that are used to construct a univariate exposure index in a single direction. The study outcome is then regressed on the exposure index and covariates to yield effect estimates.

Partial effect of a mixture. Under linearity and additivity, the overall effect of a mixture can be subdivided into "partial effects." In this study, we compare QGC *a priori* (QGCAP) and WQSr *a priori* (WQSAP) methods as well as data-driven methods using sample-splitting (QGCSS and WQSSS) and approaches using the elastic net (QGCEN), model-averaging (QGCMA), repeated hold-out sets (RH-WQS), and two-index approaches (WQS2i).

In order to better understand possible interventions or causal relationships, it may be important to characterize the partial effects within the overall mixture. WQSr was originally envisioned as a way to estimate "in the direction of increased risk, thereby averting the focus from the environmental toxins that may be found to have a protective effect," while QGC was developed for estimating the overall effect of a mixture. Comparison of these methods has been restricted to settings in which the overall effect equaled a partial effect (e.g., all exposures with effects in one direction) or assessing bias for the overall effect. The properties of how well these methods estimate these partial effects has not been studied.

Partial effects can represent effects of any subset of the mixture, all of which sum to ψ. Then,

$$\psi^- = \sum_{j=1}^{J} \mathbb{I}(\beta_j < 0)\beta_j \text{ and } \psi^+ = \sum_{j=1}^{J} \mathbb{I}(\beta_j > 0)\beta_j, \tag{2}$$

where $\mathbb{I}(\cdot)$ is the indicator function which takes 1 if the jth β coefficient is positive or negative, respectively, and is 0 otherwise; β_j are the coefficients over J exposures. In this case, ψ^- and ψ^+ may represent the expected value of the outcome per unit increase in all exposures with negative or positive coefficients, respectively. We define ψ^- as the negative partial effect as the effect of all exposures which would be expected to decrease the average value of the outcome ($x_1, x_3 = 1$ vs. $x_1, x_3 = 0$, holding other exposures constant) and the positive partial effect (ψ^+) as the effect of all exposures which would be expected to increase average value of the outcome ($x_2, x_4 = 1$ vs. $x_2, x_4 = 0$), holding all other exposures constant. Under this definition, each exposure in a mixture contributes to (at most) only one partial effect. Thus, the grouping of exposures into positive or negative partial effects can be referred to as a "partitioning" of the mixture into distinct sets of exposures. For an outcome

in which higher values imply worse health, the positive partition is the set of exposures that increase the outcome and, thus, are the "harmful" exposures, while the negative partition would include salutary exposures. If this partitioning is known *a priori* (e.g., via published literature), then methods like QGC can be used to estimate partial effects by simply estimating the effect of each exposure group separately. For example, Niehoff et al.[12] separately estimated effects of toxic and essential metals on body mass index (BMI), presuming that toxic metals might be associated with higher BMI and essential metals might be associated with lower BMI.

Partial effects—QGC. The current implementation of QGC (via the qgcomp[37] version 2.15.2 package in R) yields calculated partial effects based on Equation 2. However, it does not report confidence intervals or variance, which require sample-splitting (QGCSS). Implementation in the qgcomp R package without sample-splitting uses the same data for determining which exposures are positive or negative as well as estimating those effects and standard errors, resulting in artificially narrow confidence intervals. Sample-splitting avoids this "double-dipping" problem and provides valid inference. Software has been developed using the following three steps as implemented in the function qgcomp.partials() in the qgcomp[37] R package (developed by the authors): *a*) define the partial effect direction of interest and split the data into a training and validation set, *b*) transform the exposure variables into quantiles and use training data to partition exposures into positive or negative partial effects based on Equation 2, *c*) in the validation set, estimate separate partial effects (ψ^- and ψ^+) for each respective set of negative exposures and positive exposures while adjusting for the remaining exposures as covariates. Under QGCSS, weights are calculated as the proportional contribution of each exposure in the training set, and the partial negative or positive effect is estimated separately. Exposures are not excluded from the model so that a quantized exposure with a negative coefficient still acts as a confounder for the positive partial effect. For the current manuscript, the data are partitioned into 60% training and 40% validation sets following Carrico et al.[7]

Partial effects—WQSr. For a given partial effect direction, "weights" for each exposure are estimated according to its relative contribution to the partial effect in a chosen direction, averaged over bootstrap samples. WQSr then creates a summary index of the exposure mixture as a whole based on the weighted average of all transformed exposures. Exposures with zero weights are effectively excluded from the summary index. In large samples and under "directional homogeneity" (where $\psi = \psi^-$ or $\psi = \psi^+$), QGC and WQSSS converge to the same estimate.[6]

***A priori* approaches.** Ideally, *a priori* groupings of exposure within a mixture should be decided based on mechanistic knowledge or estimated negative or positive health associations from replicated epidemiologic and toxicologic studies. Both QGC and WQSr can incorporate prior knowledge about the exposures into an *a priori* partition of exposures according to whether they are hypothesized to be positively or negatively associated with the health outcome. Separate effects for both sets of exposures are estimated, adjusting for the remaining exposures as covariates. The use of this subject-matter knowledge-based approach is referred to as QGC *a priori* (QGCAP) and WQS *a priori* (WQSAP). We note that in addition to our novel extensions to QGC, the use of WQSAP represents another novel approach that, to our knowledge, has not been applied in the literature. Generally, authors of WQSr methods have advocated fitting the methods in both directions with all exposures included. QGC has been shown to provide unbiased estimates for the overall effect of a mixture under correct model specification.[6] In this case, the partial effect is a joint effect of a subset of the mixture so that under QGCAP with correct model specification, the partial effect will be unbiased. However, it will not correspond to the special case of the positive (or negative) partial effect (a joint effect of a subset of the mixture that is associated with the outcome in a single direction). However, very often it is difficult to clearly specify *a priori* groups of exposures in the mixture, particularly as studies continue to increase the number and complexity of exposures.[38] If this partitioning cannot be determined *a priori*, then the partitioning of the partial effects must be performed using data-driven methods such as with sample-splitting or with other extensions. The partitioning step of QGCSS relies on fitting a model with all exposures in a subsample of the data, so it is prone to variance inflation. We, therefore, used two common variance reduction procedures (elastic net and model averaging) to assess whether this step of QGCSS could be improved.

***Data-driven extensions.* QGC-elastic net.** The elastic net approach combines the variable selection properties of the Lasso with the coefficient shrinkage of ridge regression.[39] QGCSS is modified by using the elastic net in the data-partitioning step to identify the sign of the coefficient using training data. We use the elastic net to partition exposures into positive and negative partial effects and then use quantile g-computation without any selection or shrinkage in the second step to estimate the partial effects. We use the cv.glmnet() function from the glmnet[40] 4.1.8 R package.

QGC-model averaging. With QGCEN, a combination of variable selection and shrinkage is used to partition the exposures in the mixture. However, this does not capture the uncertainty in the model. Model-averaging is an approach that combines several plausible models into a meta-model using weights that incorporate the model uncertainty.[41] In a frequentist framework, model weights may be derived using likelihood-based weights, such as using Akaike's information criterion (AIC).[41] Let $m^{(1)}, \ldots, m^{(K)}$ represent K models based on different exposure combinations for the proposed mixture. Define $\mathrm{AIC}^{(k)} = -2\log \mathcal{L}^{(k)} + 2p^{(k)}$ where $\log \mathcal{L}^{(k)}$ is the log-likelihood of the kth model, and $p^{(k)}$ is the number of parameters in the kth model. AIC weights are given by $\omega^{(k)} = \frac{\mathrm{AIC}^{(k)}}{\sum_{k=1}^{K} \mathrm{AIC}^{(k)}}$ and models $m^{(k)}, \ldots, m^{(K)}$ are combined using a weighted average of weights $\omega^{(k)}$ and coefficients $\beta_j^{(k)}$ across k models. QGC with model-averaging (QGCMA) partitions the data into positive and negative exposures using model-averaging and then estimates the respective partial effects using quantile g-computation. First, split the dataset into a training and validation set using a 60/40 split. On the training set, permute all possible exposure combinations for the mixture into K models and estimate the AIC weights for each model. We note that all permuted combinations of exposures may be quite large as the number of exposures grows, increasing computation time. Using the likelihood-based weights, create a weighted-average of the K models and use the meta-model to determine exposures that were negative and positive. Use those negative and positive exposures, determined via the model-averaging step, to then estimate the negative and positive partial effects using separate QGC models for each respective partial effect. We use the MuMIn[42] version 1.47.5 R package.

Weighted quantile sums regression-no split. With WQSr methods, sample-splitting into a training (estimating the weights) and validation (estimating the partial effect) set is recommended to avoid overfit by using the data twice. Sample-splitting, however, limits statistical power because it reduces the sample size available for estimation.[18] WQSr that omits sample-splitting ("no split" or WQSNS)[18] has been used with a permutation-based test to circumvent this problem. In WQSNS, all data are used to estimate both the weights and the respective partial effects.

Weighted quantile sums regression with repeated holdout sets. To overcome the reduction in power and instability in estimates due to sample-splitting, Tanner et al. proposed a repeated holdout approach to WQSr.[36] This approach randomly partitions the dataset with replacement ($R = 100$ times), and WQS regression is used on each set to simulate a distribution of validated results. A bootstrap step ($B = 100$) is used within each set to ensure the stability of weights, giving a total of $B \times R$ estimations of the positive and negative weights.

Weighted quantile sums regression with penalized weights and two indices. The weighted quantile sums regression two-index model[43] allows for the estimation of a mixture effect in both positive and negative directions simultaneously. The model has the form,

$$g(\widehat{\mu}_i) = \widehat{\beta}_0 + \widehat{\psi}^+ \left(\sum_{j=1}^{J} \widehat{w}_j^+ x_{ij}^q \right) + \widehat{\psi}^- \left(\sum_{j=1}^{J} \widehat{w}_j^- x_{ij}^q \right) + \sum_{p=1}^{P} \widehat{\theta}_p u_{ip}. \quad (3)$$

Partition and estimation steps across methods are summarized in Table 1.

Simulation Study

Simulation set-up. In our simulation studies, we explored five simulation scenarios: *a*) correlated vs. uncorrelated exposures, *b*) sample size, *c*) spread in the partial effect among exposures, and *d*) imbalance in the magnitude of partial effects. For each setting, we simulated 200 random datasets on the quantile basis under the linear model as follows:

$$y = \beta_0 + \sum_{j=1}^{10} \beta_j x_j^q + \epsilon, \quad (4)$$

where β_0 is the model intercept; x_j^q is the jth "quantized" exposure that takes on values 0, 1, 2, and 3; and $\varepsilon \sim N(0, 1)$. This model corresponds to the limiting case of QGC and WQSr under directional homogeneity. Exposure data were randomly generated as 10 independent, quantized mixture exposures; $\beta_1 - \beta_4$ were nonzero ($x_1^q - x_4^q$ were causal exposures), and $\beta_5 - \beta_{10}$ were zero ($x_5^q - x_{10}^q$ were "noise" exposures). The intercept β_0 was set to 0. The negative and positive partial effects were set to -0.2 and 0.2, which were spread over the four nonzero exposures each with a β coefficient of 0.1, unless otherwise noted (i.e., $\beta_1 = \beta_3 = 0.1$, $\beta_2 = \beta_4 = -0.1$).

Aside from the base case scenario, we also explored the following:

1. Increasing exposure correlation: the correlations between the first exposure (with a positive effect) and remaining three non-zero exposures were set to 0.5, 0.75, and 0.8 (base case = 0). Correlation represents a range observed over environmental studies.
2. Increasing sample sizes: sample sizes of 300, 500, 1,000, and 10,000 in scenario (base case = 200). Selected sample sizes were chosen to reflect a range commonly found in environmental epidemiologic studies.
3. Increasing spread of the partial effect across exposures: we kept both the negative and positive partial effects at ± 0.2 but spread the negative partial effect (-0.2) over one, four, and eight exposures (base case was spread across two exposures). While spread is unknown in any observational study, low spread occurs when one exposure drives the overall effect and high spread occurs when mixtures have many harmful constituents.
4. Increasing imbalance between the negative and positive partial effects: We keep the negative partial effect of -0.2 from the base case but then change the positive partial effect by setting it to 1 (large imbalance), 0.6 (medium imbalance), or 0.3 (small imbalance) (base case = 0.2, no imbalance). While imbalance is unknown in any study, we note that a mixture of air pollutants would often be expected to have a different balance (large imbalance) from a dietary mixture of both healthful and harmful constituents (possibly no imbalance).

A summary of simulation scenarios can be found in Table S1. We calculated the root mean squared error (RMSE) and average

Table 1. Comparison of partition and estimation steps for mixtures methods.

Method	Partition step	Estimation step
QGCAP (quantile g-computation with *a priori* knowledge)	Partition exposures into negative or positive based on *a priori* knowledge	Estimate separate effects for negative and positive sets using QGC, keeping all exposures in the model
QGCSS (quantile g-computation with sample-splitting)	Use training data to identify negative and positive coefficients using QGC	Use validation data to fit each of the negative and positive partial effects using QGC, keeping all exposures in the model
QGCMA (quantile g-computation with model-averaging)	Use training data to permute exposures into K models and partition into negative and positive coefficients based on the meta-model (weighted-average of the K models)	Use validation data to fit each of the negative and positive partial effects using QGC, keeping all exposures in the model
QGCEN (quantile g-computation with elastic net)	Use elastic net with training data to select negative and positive exposures	Use validation data to fit each of the negative and positive partial effects using QGC, keeping all exposures in the model
WQSSS (weighted quantile sums regression with sample-splitting)	Define positive or negative target and estimate weights for that target in a subsample of the data (training set)	Estimate overall association for all exposures with non-zero weight by regressing the outcome on a weighted index exposure in the validation set, dropping exposures with null weights
WQSNS (weighted quantile sums regression with no sample-splitting)	Define positive or negative target	Use all data to *a*) estimate weights and *b*) estimate respective partial effects, given those weights, dropping exposures with null weights
RH-WQS (repeated holdouts weighted quantile sums regression)	Randomly partition dataset with replacement	Same estimation as WQSSS but include a bootstrap step to stabilize weights within each partitioned dataset.
WQSAP (weighted quantile sums regression with *a priori* knowledge)	Partition exposures into negative or positive based on *a priori* knowledge	Same estimation as WQSSS, keeping all exposures in the model
WQS2i (weighted quantile sums regression two-index model with penalization)	Define positive and negative target in same model and estimate weights for targets in repeated holdouts of dataset (randomly partitioned data with replacement)	Take the mean of exposure weights and both positive and negative partial effects across holdout sets and bootstrap repetitions

bias (Equations 5 and 6) across these methods for each positive and negative partial effect, where M is the number of simulated datasets. We adjust for confounding by including the opposing sign exposures as covariates. We assess bias in terms of direction and relative magnitude of bias among the methods, rather than absolute magnitude. We note that in the base case, the magnitude of the partial effects is 0.2, so an absolute bias of 0.2 for all scenarios other than the "imbalance" scenario would be a 100% absolute bias. All analyses were performed using the R statistical computing software version 4.2.2. The high-performance computing cluster at the National Institutes of Health was used for performing the simulation studies.[44]

$$\text{RMSE}\left(\widehat{\psi}^{+}\right)=\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\widehat{\psi}^{+}-\psi^{+}\right)^{2}}\ \text{and}\ \text{RMSE}\left(\widehat{\psi}^{-}\right)$$
$$=\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\widehat{\psi}^{-}-\psi^{-}\right)^{2}}, \quad (5)$$

$$\text{Average Bias}\left(\widehat{\psi}^{+}\right)=\frac{1}{M}\sum_{m=1}^{M}\widehat{\psi}^{+}-\psi^{+}\ \text{and}\ \text{Average Bias}\left(\widehat{\psi}^{-}\right)$$
$$=\frac{1}{M}\sum_{m=1}^{M}\widehat{\psi}^{-}-\psi^{-}. \quad (6)$$

Data Example Methods

Our data examples focus on two scenarios: in example 1, we have strong *a priori* knowledge and expect only a positive partial effect. In example 2, we expect partial effects in both directions. Since QGCSS and QGCMA resulted in very similar RMSE and bias in both examples, we restrict our QGC-based analyses to QGCSS and QGCAP. We omit analysis by QGCEN, as our simulation studies showed poor mean squared error and bias overall.

Example 1: phthalates, phenols, parabens, and oxidative stress in the LIFECODES study. We used the LIFECODES Fetal Growth Study, a case–cohort sample nested within the prospective LIFECODES pregnancy cohort, designed to examine consumer product exposures and their effects on fetal growth.[45] Our data were composed of 899 pregnant participants enrolled between 2007 and 2018; the protocol for the LIFECODES birth cohort was approved by the institutional review board (IRB) at Brigham and Women's Hospital.[46] Participants provided urine samples at multiple study visits, and we used urinary measures of 24 phthalate and phenol metabolites from samples provided at the first study visit (median: 11 wk) and urinary concentrations of 8-isoprostane measured at the same study visit. Urinary concentrations of metabolites and 8-isoprostane were adjusted for specific gravity using Boeniger's method.[47] Following Bommarito et al.,[48] we created a weighted sum of di-2-ethylhexyl phthalate (DEHP) metabolites [mono-2ethylhexyl phthalate (MEHP), mono (2-ethyl-5-hydroxyhexyl) phthalate (MEHHP), mono (2-ethyl-5-oxohexyl) phthalate (MEOHP), and mono-(2-ethyl-5-carboxypentyl) phthalate (MECPP)] using their respective molar weights, [$\sum$DEHP = [(MEHP/278 + MEHHP/294 + MEOHP/292 + MECPP/308) × 308], yielding a final set of 21 metabolites. Our covariates included age (continuous), prepregnancy BMI (continuous), maternal education level [no high school degree, high school degree (or equivalent), technical school, junior college/some college, college degree (referent), graduate school], conception year [categorical, 2007 (referent) to 2018], alcohol use [yes, no (referent)], smoking status during pregnancy [yes, no (referent)], and gestational age (continuous). Eight percent of participants had missing values in any of the 21 metabolites or covariates, which we considered to be negligible. After excluding those participants, our final analytic sample included 824 participants (Table 2).

Previous studies have found that many of the phthalates, parabens, and phenols of interest have been associated with increased levels of 8-isoprostane.[21,25,26,49] Our *a priori* hypothesis was that all 21 metabolites would be cross-sectionally associated with higher concentrations of 8-isoprostane. We expected the positive partial effect to be equal to the overall. Relevant to our simulations, we expect there to be a large imbalance in the partial effect and low to moderate correlation among exposures, but it is unknown whether this effect is spread across many or only a few of the exposures.

Example 2: persistent organic pollutant biomarkers and leukocytic telomere length. We used participant data from the 2001–2002 cycle of the National Health and Nutrition Examination Survey (NHANES). NHANES study protocols are reviewed and approved by the National Center for Health Statistics (NCHS) Research Ethics Review Board; details on the study design have been described elsewhere.[28,29] Following Mitro et al.[28] and Gibson et al.,[29] of the 11,039 participants originally interviewed, we excluded participants if they were under 20 years old, did not consent to DNA analysis or did not have sufficient samples to estimate telomere length ($n = 6{,}779$), did not have the environmental chemical analysis ($n = 2{,}850$), were missing key covariates (BMI, education, serum cotinine) ($n = 80$), or had any missing values on any one of the 24 PCBs, dioxins, or furans that were used to form toxic equivalence factors ($n = 327$). After excluding an additional 24 participants who were missing data on any of the additional PCBs, dioxins, or furans considered, the final analytic sample comprised of 979 participants. Missing data were considered negligible at 3.9% of the original sample due to small numbers. We excluded congeners with <60% of sample concentrations above the limit of detection,[29] which resulted in a final sample (Table 3) including 18 persistent organic pollutant exposure biomarkers, corrected for lipid serum lipid concentrations via standardization.[50] We used mean telomere length relative to a standard DNA referent (T:S ratio), taken across six assays per sample.[51,52] In line with prior analyses, we adjusted for age (continuous) and age^2, body mass index (continuous), NHANES race/ethnicity group [non-Hispanic white (referent), non-Hispanic black, Mexican-American, other (including multi-race, non-Hispanic respondents whose primary race was unknown and were not members of the other groups)], education category [less than high school, high school graduate, some college, and college graduate or more (referent)], lipid cotinine levels (continuous), and white blood cell count (continuous). We use race/ethnicity (self-reported) following Mitro et al.[28] as a way to adjust for confounding from other social determinants of health. Percent cell type (lymphocytes, monocytes, neutrophils, eosinophils) were linear combinations of other values and were omitted, and sampling weights were not used in the interest of generalizability of the approach. We also log-transformed the congeners, cotinine, and mean leukocytic telomere length.

Results

Simulation Results

As sample size increased, we observed all methods converged to the same value of mean squared error across both partial effects, except for WQS2i (Figure 1). As the negative partial effect was spread over more exposures, RMSE for all methods except for WQSNS and WQS2i appeared to increase for that partial effect (but not the opposing partial effect). With no imbalance in the partial effects, the mean squared error was lowest for QGCAP

Table 2. Study sample characteristics of participants and congeners from the LIFECODES fetal growth study, 2007–2018.

Variable	Study sample ($n = 824$)	Missing (%)
Age [y (mean ± SD)]	32.20 (5.65)	0
Prepregnancy BMI [kg/m^2 (mean ± SD)]	26.50 (6.68)	0
Gestational age at time 1 [wk (mean ± SD)]	11.39 (3.44)	0
Conception year [*n* (%)]		0
2007	115 (14.0)	—
2008	37 (4.5)	—
2009	59 (7.2)	—
2010	78 (9.5)	—
2011	126 (15.3)	—
2012	138 (16.7)	—
2013	92 (11.2)	—
2014	37 (4.5)	—
2015	52 (6.3)	—
2016	38 (4.6)	—
2017	44 (5.3)	—
2018	8 (1.0)	—
Smoking during pregnancy [*n* (%)]	54 (6.6)	0
Yes	54 (6.6)	—
No	770 (93.4)	—
Maternal education [*n* (%)]	—	1.1
Did not graduate high school	24 (2.9)	—
Graduated from high school	93 (11.3)	—
Junior college/some college	155 (18.8)	—
Technical school	15 (1.8)	—
College graduate	250 (30.3)	—
Graduate school	278 (33.7)	—
Alcohol use during pregnancy [*n* (%)]	—	0.4
No	771 (93.6)	—
Yes	50 (6.1)	—
8-isoprostane at time 1 [ng/mL (mean ± SD)]	1.30 (1.15)	0
Phthalate metabolites at time 1 [ng/mL (mean ± SD)]	—	—
$\sum$ *DEHP*	96.90 (393.28)	0
MCNP	7.21 (21.95)	0
MCOP	58.00 (110.52)	0
MBP	24.91 (85.05)	0
MBZP	16.58 (79.00)	0
MCPP	11.34 (92.35)	0
MEP	294.44 (1,736.65)	0
MIBP	16.48 (44.73)	0
MNP	0.41 (0.92)	0
Phenol metabolites at time 1 [ng/mL (mean ± SD)]	—	—
2,4-DCP	1.36 (7.12)	0
2,5-DCP	11.36 (68.72)	0
BP3	386.17 (1,255.99)	0
BPA	1.91 (2.80)	0
BPS	0.63 (1.29)	0
TCB	1.57 (5.09)	0
TCS	138.08 (359.79)	0
BPF	1.97 (20.85)	0
Paraben metabolites at time 1 [ng/mL (mean ± SD)]	—	—
BPB	2.16 (8.23)	0
EPB	23.47 (127.64)	0
MPB	324.17 (983.33)	0
PPB	85.35 (227.68)	0

Note: Phthalates ($\sum$DEHP–MNP), phenols (2,4-DCP–BPF), and parabens (BPB–PPB). —, no data; 2,4-DCP, 2,4-DCP concentration; 2,5-DCP, 2,5 DCP concentration; BMI, body mass index; BP3, benzophenone 3 concentration; BPA, bisphenol A concentration; BPB, butyl paraben concentration; BPF, bisphenol F concentration; BPS, bisphenol S concentration; $\sum$*DEHP*, weighted molar sum of MEHP, MEHHP, MEOHP, MECPP (phthalates); EPB, ethyl paraben; MBP, mono-*n*-butyl phthalate; MBZP, mono-benzyl phthalate; MCNP, mono-carboxy isononyl phthalate; MCOP, mono-carboxy octyl phthalate; MCPP, mono (3-carboxypropyl); MECPP, mono-(2-ethyl-5-carboxypentyl) phthalate; MEHP, mono-2ethylhexyl; MEHHP, mono (2-ethyl-5-hydroxyhexyl) phthalate; MEOHP, mono (2-ethyl-5-oxohexyl) phthalate; MEP, monoethyl phthalate; MIBP, mono-isobutyl phthalate; MNP, mono-isononyl phthalate; PPB, propyl paraben; SD, standard deviation; TCB, triclocarban concentration; TCS, triclosan concentration.

Table 3. Study sample characteristics of participant baseline characteristics and congeners from the 2001–2002 cycle of the national health and nutrition examination survey (NHANES).

Variable	Study sample ($n = 979$)
Age [y (mean ± SD)]	48.34 (18.35)
BMI [kg/m^2 (mean ± SD)]	28.26 (6.17)
Cotinine [ng/mL (mean ± SD)]	52.36 (115.25)
Telomere length (mean ± SD)	1.06 (0.25)
White blood cell count (mean ± SD)	7.21 (2.14)
Race/ethnicity [*n* (%)]	
Non-Hispanic white	501 (51.2%)
Mexican-American	225 (23.0%)
Non-Hispanic black	176 (18.0%)
Other Hispanic	42 (4.3%)
Other, including mixed race	35 (3.6%)
Education [*n* (%)]	
<9th grade	140 (14.3%)
9th–11th grade	174 (17.8%)
HS/GED	222 (22.7%)
Some college/AA	260 (26.6%)
College or above	183 (18.7%)
Sex at birth [*n* (%)]	
Female	549 (56.1%)
Male	430 (43.9%)
Furans [ng/g (mean ± SD)]	
1,2,3,4,6,7,8-HpCDF	11.57 (10.72)
1,2,3,6,7,8-HxCDF	5.42 (4.19)
1,2,3,4,7,8-HxCDF	6.44 (4.94)
2,3,4,7,8-PeCDF	6.66 (5.67)
Dioxins [ng/g (mean ± SD)]	
1,2,3,4,6,7,8,9-OCDD	494.32 (517.50)
1,2,3,4,6,7,8-HpCDD	57.63 (55.78)
1,2,3,6,7,8-HxCDD	48.09 (40.57)
PCBs [ng/g (mean ± SD)]	
PCB169	24.47 (20.62)
PCB118	16.70 (23.31)
PCB126	38.32 (51.82)
PCB194	10.45 (10.83)
PCB187	12.10 (12.27)
PCB180	37.80 (39.56)
PCB170	15.52 (14.42)
PCB153	52.22 (53.54)
PCB138	38.10 (40.33)
PCB99	9.98 (10.28)
PCB74	12.62 (13.59)

Note: Furans, dioxins, and PCBs are reported as lipid-adjusted serum levels and reported as ng/g lipid. AA, associate of arts (two-year undergraduate degree); BMI, body mass index; GED, general educational development; HS, high school; SD, standard deviation.

and WQSAP. Though we examined impacts of increasing imbalance in the positive partial effect, we also observed changes in the RMSE in the negative partial effect across all methods except QGCAP and WQSAP.

As correlation among exposures increased (and approached collinearity), bias for both partial effects went through the null for all methods except QGCSS and QGCEN, and the two *a priori* methods (Figure 2). For all methods, RMSE and bias decreased with sample size (Figures 1 and 2). At higher levels of imbalance, QGCAP and WQSAP consistently had the lowest RMSE. QGCMA and QGCSS had near identical negative biases in the negative and positive effect across all settings; WQSSS and RH-WQS had near identical average bias in both partial effect directions.

Increasing exposure correlation. In the base case, the patterns across the negative and positive partial effects were identical, so we report only results for the positive partial effect. WQSSS, WQSNS, and RH-WQS showed a bias away from the null while the QGC methods showed a negative bias (except for QGCAP and WQSAP). QGCAP and WQSAP had the lowest RMSE, followed by RH-WQS, QGCSS, QGCMA, WQSSS, QGCEN, WQS2i, and WQSNS. As correlation with x_1 increased among the exposures, RMSE increased for all estimators except WQSNS.

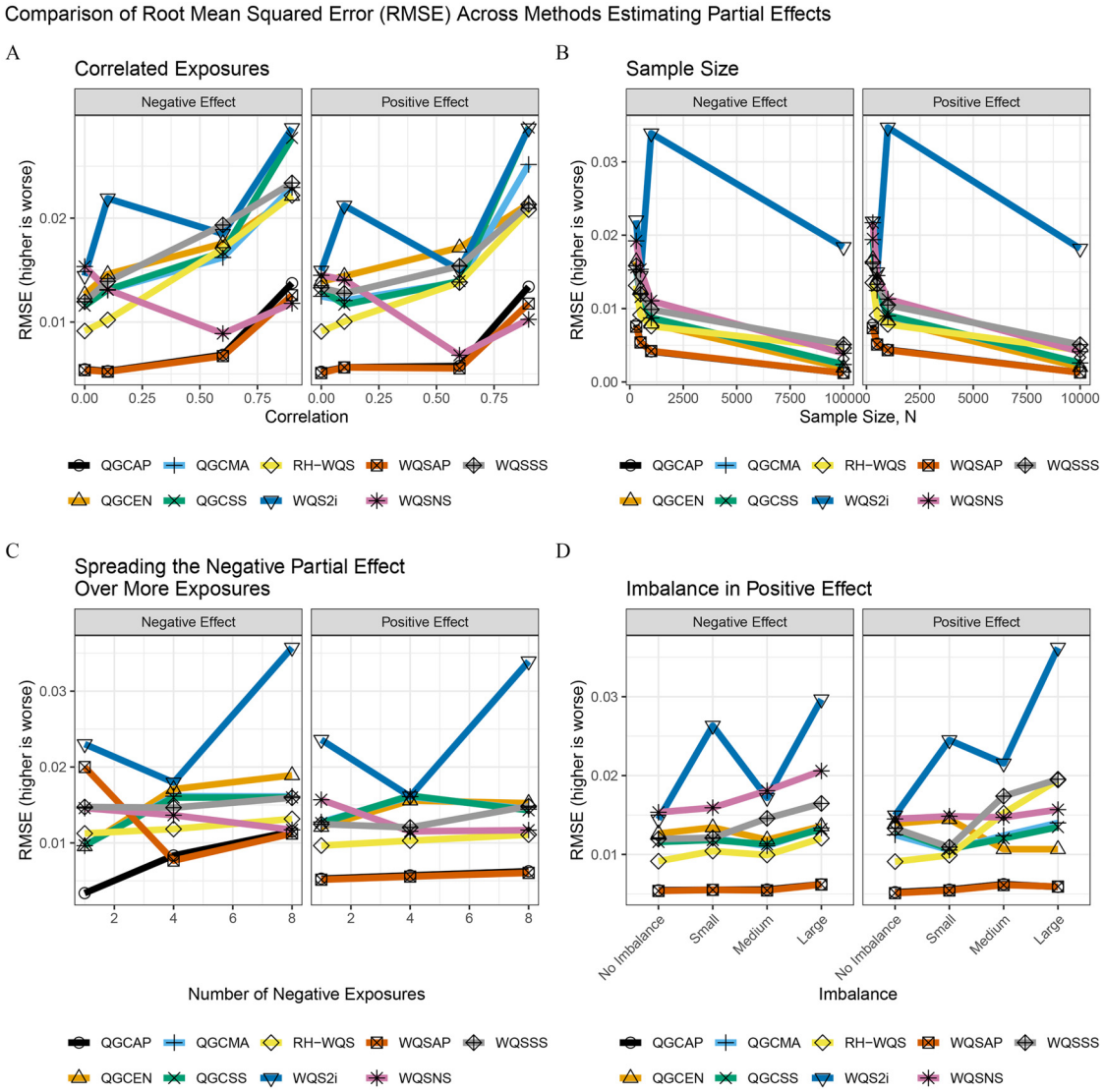

Figure 1. Root mean squared error (RMSE) over 200 realizations across four simulation settings [(A) correlated exposures, (B) sample size, (C) spread in effect, and (D) imbalance in positive effect]. Except in panels where each feature is varied, exposures are uncorrelated, sample size is 500, each partial effect is spread over two exposures, and there is no imbalance in the magnitude of the partial effects. See Tables S2–S5. Note: QGCAP, quantile g-computation with *a priori* knowledge; QGCEN, quantile g-computation with elastic net; QGCMA, quantile g-computation with model-averaging; QGCSS, quantile g-computation with sample-splitting; RH-WQS, repeated holdouts weighted quantile sums regression; WQSAP, weighted quantile sums regression *a priori*; WQSNS, weighted quantile sums regression with no sample-splitting; WQSSS, weighted quantile sums regression with sample-splitting; WQS2i, weighted quantile sums regression two-index model with penalization.

Increasing sample size. At the largest sample size of 10,000, QGCAP and WQSAP had the lowest bias and RMSE in both the negative and positive effects, while WQS2i had the highest RMSE in both. All QGC-based methods maintained a constant near zero bias once sample size reached 1,000, except for QGCEN (Figure 2). WQSSS, WQSNS, RH-WQS, and WQS2i demonstrated larger biases that persisted to the largest sample sizes studied. Results for WQSSS, WQSNS, WQS2i, and RH-WQS suggest that it does not appear to be converging toward zero bias. This result is surprising in light of the large sample results of Keil et al.[6] where both methods were unbiased in large samples under directional homogeneity. The current analysis was not done under directional homogeneity but, in the base case exposures, was uncorrelated so the large sample bias of WQSr methods (aside from WQSAP) did not result from uncontrolled confounding by exposures with opposing effects.

Spreading the partial effect over more exposures. In both the negative and positive partial effects, QGCAP consistently had the lowest RMSE and bias; WQSAP performed similarly only for the positive partial effect—for the negative partial effect, bias was high when the effect was concentrated over fewer exposures. For QGCEN, bias and RMSE increased most dramatically with spread, likely indicating that spread contributes to "selecting out" exposures from the model and bias toward the null. RMSE was highest for WQS2i and grew as spread increased. Increasing spread had a unique effect on WQSSS, RH-WQS, and WQS2i where bias went across and away from the null as the spread increased.

Increasing imbalance toward the positive effect. RMSE increased in WQSSS, WQSNS, and WQS2i most dramatically from a small imbalance to a large imbalance in both effects. On the positive side (where the imbalance was placed), QGCAP, WQSAP, QGCSS, and QGCMA maintained a relatively constant RMSE across the small, medium, and large imbalances. The largest increase in RMSE in the positive effect was observed in

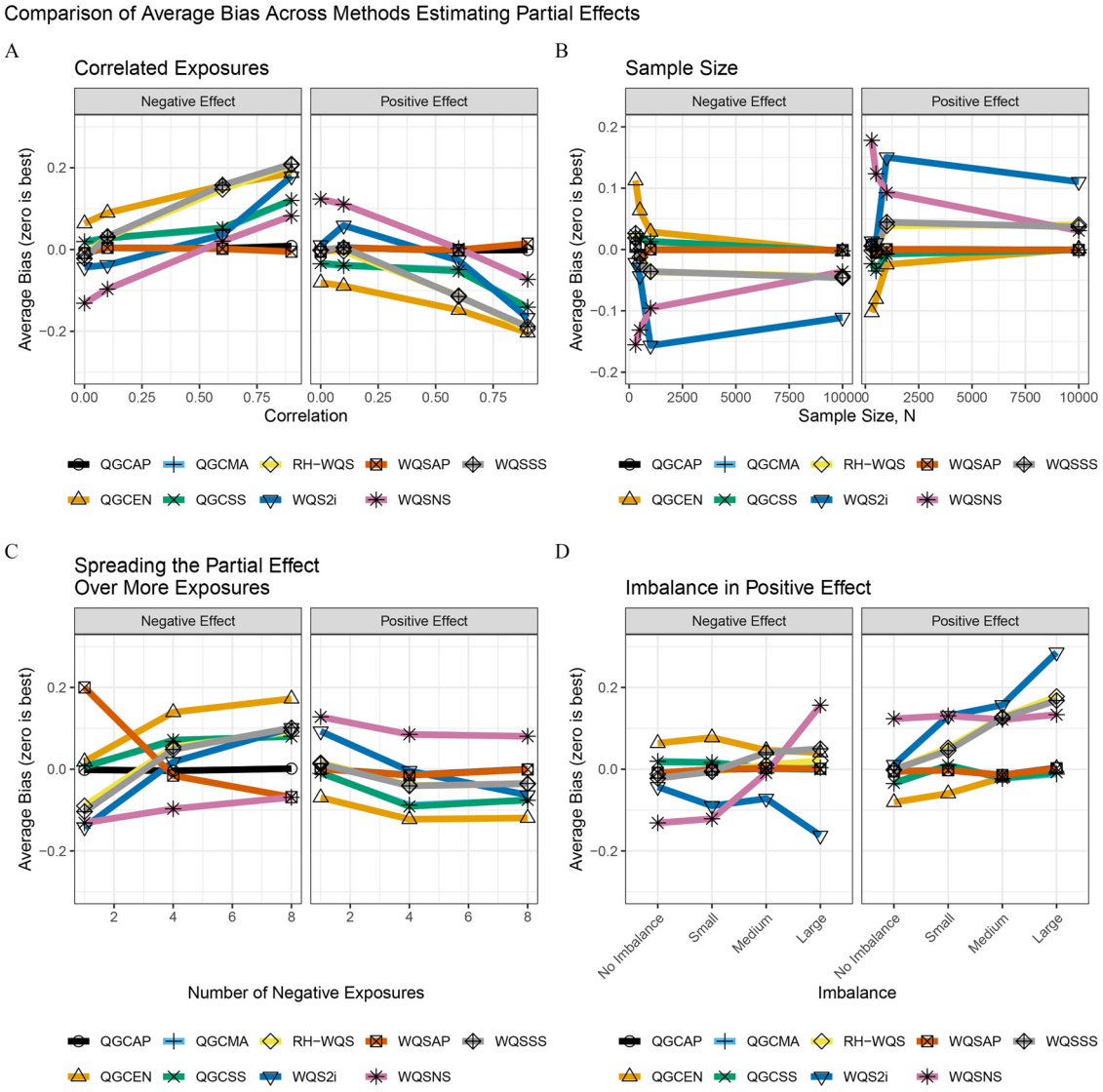

Figure 2. Average bias over 200 realizations across four simulation settings [(A) correlated exposures, (B) sample size, (C) spread in effect, (D) imbalance in positive effect]. Except in panels where each feature is varied, exposures are uncorrelated, sample size is 500, each partial effect is spread over two exposures, and there is no imbalance in the magnitude of the partial effects. See Tables S2–S5. Note: QGCAP, quantile g-computation with *a priori* knowledge; QGCEN, quantile g-computation with elastic net; QGCMA, quantile g-computation with model-averaging; QGCSS, quantile g-computation with sample-splitting; RH-WQS, repeated holdouts weighted quantile sums regression; WQSAP, weighted quantile sums regression *a priori*; WQSNS, weighted quantile sums regression with no sample-splitting; WQSSS, weighted quantile sums regression with sample-splitting; WQS2i, weighted quantile sums regression two-index model with penalization.

WQS2i; WQSNS also increased in RMSE in both effects, though not as dramatically, from a small imbalance to a large imbalance. In both the negative and positive partial effects, the average bias was maintained near 0 for QGCAP and WQSAP even as the imbalance in the positive effect increased. In the negative effect, WQSNS had a distinct pattern where bias decreased from small to medium imbalance and across the null in the large imbalance scenario. In the positive effect, where the larger effects were placed to create the imbalance, the most drastic change in bias was observed in WQS2i.

Data Example Results

Example 1: phthalates, phenols, parabens, and oxidative stress in the LIFECODES study. Since our *a priori* knowledge was that all metabolites would increase 8-isoprostane concentrations, there was no QGCAP nor WQSAP negative effect estimated. Figure 3 shows the results for the partial effect estimates across methods (Figure S1 and Table S8 show the correlations among exposures). In accordance with our hypothesis of no negative partial effects, WQSNS and WQSSS did not estimate any negative partial effect. Although QGCSS estimates a negative partial effect, it did not differ significantly from the null [QGCSS: $\hat{\psi} - 0.110$ (95% CI: -0.288, 0.062)]. In contrast, all methods estimated a positive partial effect that was different from zero. QGCSS had the highest point estimate (0.390) followed by WQSNS, WQSAP, WQS2i, WQSSS, and QGCAP (Table 4).

As defined here, the true overall effect of all exposures is equivalent to the sum of the true partial effects. We also explored how the sum of the estimated partial effects differed from the estimated overall effect across different seed values (Figure S2; Table S6) for QGCSS and WQSS, in which both had a sample-splitting component in the supplemental material.

Example 2: persistent organic pollutant biomarkers and leukocytic telomere length. We expected that most PCBs (PCB74, PCB99, PCB138, PCB153, PCB170, PCB180, PCB187, and

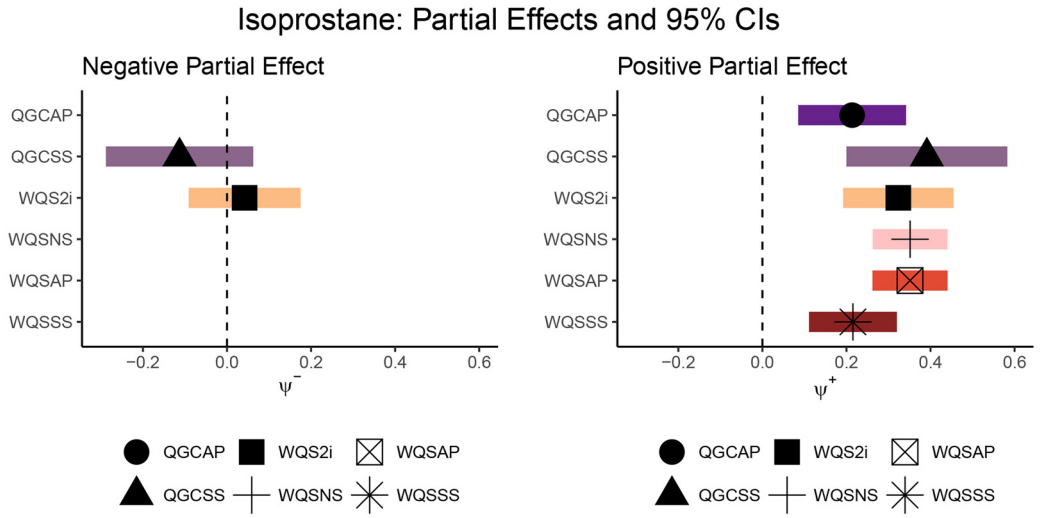

Figure 3. LIFECODES Fetal Growth Study, 2007–2018 example. Estimated negative and positive partial effects with 95% confidence intervals (CIs). Analyses were adjusted for age, prepregnancy BMI, maternal education, conception year, alcohol use, smoking during pregnancy, and gestational age. Estimates are missing for a method when a partial effect was ruled out *a priori* (QGCAP) or when the method failed to yield an estimate (WQSNS), and confidence intervals are not reported for WQSNS because they are known to be invalid. See Table 4. Note: BMI, body mass index; QGCAP, quantile g-computation with *a priori* knowledge; QGCSS, quantile g-computation with sample-splitting; RH-WQS, repeated holdouts weighted quantile sums regression; WQSAP, weighted quantile sums regression *a priori*; WQSNS, weighted quantile sums regression with no sample-splitting; WQSSS, weighted quantile sums with sample-splitting; WQS2i, weighted quantile sums regression two-index model with penalization.

PCB194) would have a negative effect on leukocytic telomere length while dioxins and furans (1,2,3,6,7,8-HxCDD, 1,2,3,4,6,7,8-HpCDD, 1,2,3,4,6,7,8,9-OCDD, 2,3,4,7,8-PeCDF, 1,2,3,4,7,8-HxCDF, 1,2,3,6,7,8-HxCDF, and 1,2,3,4,6,7,8-HpCDF) as well as PCB126, PCB169, and PCB118 would have a positive effect on length.

Figure 4 shows the estimated negative and positive partial effects across methods (Figure S3 and Table S9 show the correlations among exposures). No method showed sufficient evidence of a negative partial effect, as 95% confidence intervals mostly spanned across the null (Figure 4; Table 5), except for WQSSS (in the negative partial effect) and WQSNS and WQSAP (in the positive partial effect). Interestingly, WQSSS [0.039 (0.007, 0.073)] estimated a positive value for the negative partial effect. Though it is possible to estimate a partial effect of the opposite sign, it indicates a replication failure driven by either incorrect *a priori* assumptions (*a priori* methods) or sampling error (sample-splitting methods), implying insufficient evidence for a partial effect. WQSNS did not estimate a negative partial effect. In the positive partial effect, WQSAP and WQSNS estimated a significant positive effect [0.046 (0.023, 0.070) and 0.048 (0.027, 0.069)], whereas QGCAP, QGCSS, WQSSS, and WQS2i did not.

To follow up on the unexpected finding in the WQSr methods, we explored congeners whose estimated effect was opposite that of the partial effect. Congeners that were identified as having conflicting effect directions in non-*a priori* methods included PCB74, PCB126, 1,2,3,4,6,7,8-HpCDF, 1,2,3,4,6,7,8,9-OCDD, and 1,2,3,4,6,7,8-HpCDD. We found weak to no evidence of negative or positive partial effects in the persistent organic pollutants mixture on leukocyte telomere length. Though this is in contrast

Table 4. Positive and negative partial effect estimates and membership across top six methods in the data example analysis of the LIFECODES fetal growth study, 2007–2018.

Method	Negative partial effect, ψ^- (95% confidence interval)	Negative partial effect membership	Positive partial effect ψ^+ (95% confidence interval)	Positive partial effect membership
QGCAP	(not estimated)	—	0.21 (0.085, 0.342)	MCNP, MCOP, MBP, MBZP, MCPP, MEP, MIBP, MNP, 2,4-DCP, 2,5-DCP, BP3, BPA, BPF, BPS, BPB, EPB, MPB, PPB, TCB, TCS, $\sum DEHP$
QGCSS	−0.11 (−0.288, 0.062)	BP3, TCS, MPB, MNP, BPF, TCB	0.39 (0.2, 0.583)	2,4-DCP, MCOP, PPB, $\sum DEHP$, MBZP, MCPP, MIBP, BPA, MBP, BPB
WQSSS	(not estimated)	—	0.22 (0.111, 0.32)	MCPP, PPB, MIBP, MBZP, MCOP, $\sum DEHP$
WQSNS	(not estimated)	—	0.35 (0.262, 0.441)	MBZP, MCPP, PPB, $\sum DEHP$, MCOP, 2,4-DCP, MIBP, MBP
WQSAP	(not estimated)	—	0.35 (0.262, 0.441)	MCNP, MCOP, MBP, MBZP, MCPP, MEP, MIBP, MNP, 2,4-DCP, 2,5-DCP, BP3, BPA, BPF, BPS, BPB, EPB, MPB, PPB, TCB, TCS, $\sum DEHP$
WQS2i	0.04 (−0.091, 0.175)	MBP, TCS, MNP, BPF, TCB, BP3, MEP	0.32 (0.192, 0.455)	PPB, $\sum DEHP$, 2,4-DCP, MCOP, BPB, MCPP, MIBP, MBZP

Note: LIFECODES Fetal Growth Study, 2007–2018 ($n = 824$) example. Estimated negative and positive partial effects with 95% confidence intervals (CIs). Analyses were adjusted for age, prepregnancy BMI, maternal education, conception year, alcohol use, smoking during pregnancy, and gestational age. Estimates are missing for a method when a partial effect was ruled out *a priori* (QGCAP) or when the method failed to yield an estimate (WQSNS), and confidence intervals are not reported for WQSNS because they are known to be invalid. —, no data; BMI, body mass index; QGCAP, quantile g-computation with *a priori* knowledge; QGCSS, quantile g-computation with sample-splitting; WQSAP, weighted quantile sums regression *a priori*; WQSNS, weighted quantile sums regression with no sample-splitting; WQSSS, weighted quantile sums with sample-splitting; WQS2i, weighted quantile sums regression two-index model with penalization; 2,4-DCP, 2,4-DCP concentration; 2,5-DCP, 2,5 DCP concentration; BP3, benzophenone 3 concentration; BPA, bisphenol A concentration; BPB, butyl paraben concentration; BPF, bisphenol F concentration; BPS, bisphenol S concentration; weighted molar sum of MEHP, MEHHP, MEOHP, MECPP (phthalates); EPB, ethyl paraben; MBP, mono-n-butyl phthalate; MBZP, mono-benzyl phthalate; MCNP, mono-carboxy isononyl phthalate; MCOP, mono-carboxy octyl phthalate; MCPP, mono (3-carboxypropyl); MECPP, mono-(2-ethyl-5-carboxypentyl) phthalate; MEHP, mono-2ethylhexyl; MEHHP, mono (2-ethyl-5-hydroxyhexyl) phthalate; MEOHP, mono (2-ethyl-5-oxohexyl) phthalate; MEP, monoethyl phthalate; MIBP, mono-isobutyl phthalate; MNP, mono-isononyl phthalate; PPB, propyl paraben; SD, standard deviation; TCB, triclocarban concentration; TCS, triclosan concentration.

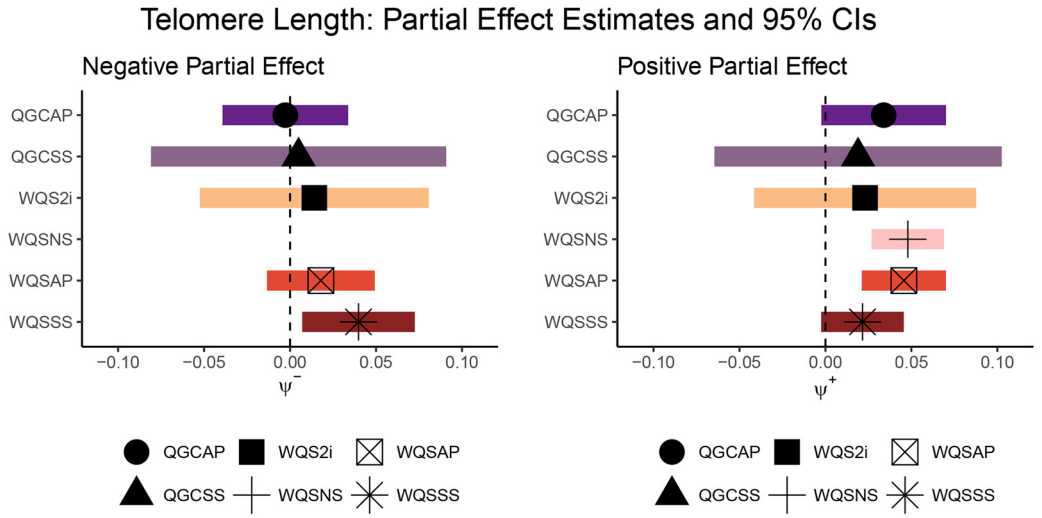

Figure 4. National Health and Nutrition Examination Survey (NHANES), 2001–2002 cycle example. Estimated negative and positive partial effects with 95% confidence intervals (CIs). Confidence intervals are not included for WQSNS. Analyses were adjusted for age, age^2, BMI, race/ethnicity, sex, cotinine, and white blood cell count. See Table 5. Note: BMI, body mass index; QGCAP, quantile g-computation with *a priori* knowledge; QGCSS, quantile g-computation with sample-splitting; RH-WQS, repeated holdouts weighted quantile sums regression; WQSAP, weighted quantile sums regression *a priori*; WQSNS, weighted quantile sums regression with no sample-splitting; WQSSS, weighted quantile sums with sample-splitting; WQS2i, weighted quantile sums regression two-index model with penalization.

to previous findings, null results could be driven by noisy exposure measurements or imperfect *a priori* knowledge. The more variability in exposure measurements, the more challenging it is to identify an effect, particularly as sample size is reduced with sample-splitting. Pulling from the simulation results, high correlations could lead all of the methods to behave somewhat unpredictably in terms of bias (except for QGCAP and WQSAP), which can manifest in highly variable results.

Estimates, their confidence intervals, and negative/positive partial effect exposure memberships are summarized in Tables 4 and 5. We also explored how the sum of the partial effects differed from the estimated overall effect across different seed values in the supplemental material (Figure S4; Table S7).

Discussion

Environmental mixtures necessitate innovative methods to answer the variety of research questions posed in the field. While QGC and WQSr have become frequently used in environmental epidemiology to analyze mixture data and estimate an overall

Table 5. Positive and negative partial effect estimates and membership across top six methods in the data example analysis of the 2001–2002 cycle of the National Health and Nutrition Examination Survey (NHANES).

Method	Negative partial effect, ψ^- (95% confidence interval)	Negative partial effect membership	Positive partial effect ψ^+ (95% confidence interval)	Positive partial effect membership
QGCAP	−0.003 (−0.039, 0.034)	PCB74, PCB99, PCB138, PCB153, PCB170, PCB180, PCB187, PCB194	0.034 (−0.002, 0.07)	PCB126, PCB169, PCB118, 1,2,3,6,7,8-HxCDD, 1,2,3,4,6,7,8-HpCDD, 1,2,3,4,6,7,8,9-OCDD, 2,3,4,7,8-PeCDF, 1,2,3,4,7,8-HxCDF, 1,2,3,6,7,8-HxCDF, 1,2,3,4,6,7,8-HpCDF
QGCSS	0.005 (−0.081, 0.091)	PCB180, PCB74, 1,2,3,4,6,7,8-HpCDD	0.019 (−0.065, 0.103)	PCB153, 2,3,4,7,8-PeCDF, PCB126, 1,2,3,4,6,7,8-HpCDF, PCB169
WQSSS	0.04 (0.007, 0.073)	1,2,3,4,6,7,8,9-OCDD; 1,2,3,4,6,7,8-HpCDD; 1,2,3,6,7,8-HxCDD; 1,2,3,4,7,8-HxCDF; 1,2,3,6,7,8-HxCDF; PCB126; PCB74; 1,2,3,4,6,7,8-HpCDF; PCB118	0.022 (−0.002, 0.046)	2,3,4,7,8-PeCDF, 1,2,3,4,6,7,8-HpCDF, PCB126, 1,2,3,4,6,7,8,9-OCDD, 1,2,3,4,6,7,8-HpCDD
WQSNS	(not estimated)	—	0.048 (0.027, 0.069)	2,3,4,7,8-PeCDF; PCB126; PCB153; 1,2,3,4,6,7,8-HpCDF
WQSAP	0.018 (−0.013, 0.049)	PCB74, PCB99, PCB138, PCB153, PCB170, PCB180, PCB187, PCB194	0.046 (0.021, 0.07)	PCB126; PCB169; PCB118; 1,2,3,6,7,8-HxCDD; 1,2,3,4,6,7,8-HpCDD; 1,2,3,4,6,7,8,9-OCDD; 2,3,4,7,8-PeCDF; 1,2,3,4,7,8-HxCDF; 1,2,3,6,7,8-HxCDF; 1,2,3,4,6,7,8-HpCDF
WQS2i	0.014 (−0.052, 0.081)	1,2,3,4,7,8-HxCDF; PCB118; PCB138; 1,2,3,6,7,8-HxCDD; PCB180	0.023 (−0.042, 0.088)	2,3,4,7,8-PeCDF; PCB99; PCB126; 1,2,3,4,6,7,8-HpCDF; PCB153; PCB170; 1,2,3,4,6,7,8,9-OCDD; PCB74; PCB194

Note: National Health and Nutrition Examination Survey (NHANES) ($n = 979$), 2001–2002 cycle example. Estimated negative and positive partial effects with 95% confidence intervals (CIs). Confidence intervals are not included for WQSNS. Analyses were adjusted for age, age^2, BMI, race/ethnicity, sex, cotinine, and white blood cell count. BMI, body mass index; QGCAP, quantile g-computation with *a priori* knowledge; QGCSS, quantile g-computation with sample-splitting; RH-WQS, repeated holdouts weighted quantile sums regression; WQSAP, weighted quantile sums regression *a priori*; WQSNS, weighted quantile sums regression with no sample-splitting; WQSSS, weighted quantile sums with sample-splitting; WQS2i, weighted quantile sums regression two-index model with penalization.

effect, no study has examined the ability of these methods to estimate their respective partial effects. These partial effects are of interest in disentangling different types of mixture effects on health outcomes. For QGC, all models included the opposing sign exposures as covariates to adjust for confounding. For WQSr approaches, variables are effectively selected out if they receive very small weights for a partial effect, which will reduce or remove control for potential confounding by these variables. In this study, we explored the estimation of negative and positive partial effects using these two commonly used methods in mixtures analysis and novel variants under the assumptions of linearity and additivity.

Due to the nature of mixture data, the estimation of partial effects is challenging, and we found no one method that is universally best in terms of bias or RMSE. Once we added the complexity of correlated exposures, small sample sizes, imbalanced partial effects, and spread in partial effects over more exposures, we observed differentiation across the methods that can drive choices between methods to address specific settings. When sample sizes were large (>500), all approaches apart from WQSNS and WQS2i had low RMSE and average bias. Our simulation studies showed that QGCAP and WQSAP had the smallest RMSE and lowest average bias. However, this approach assumes that we have perfect knowledge, which outside of simulation studies is not realistic. Nonetheless, the simulations can be used to help guide some decisions, which we describe here.

In the LIFECODES example, we observed consistent results where the methods all estimated a positive partial effect, as expected *a priori*. Following Figure 5, most exposures in the mixture were not highly correlated with each other, so we expected low bias from QGCAP, WQSAP, and QGCSS (though medium RMSE from QGCSS). The sample size is large and so QGCAP, WQSAP, QGCSS, and WQSSS all have low bias and low RMSE. Finally, we expected the partial effects to be unbalanced since we only expect a positive partial effect. In this case, QGCAP and WQSAP had the lowest RMSE and bias, followed by QGCSS (medium RMSE and bias). WQSSS and WQSNS had high RMSE and bias. Based on this and our strong *a priori* knowledge that phthalates and phenols are positively associated with increases in oxidative stress markers,[22,53] we recommend using results by either *a priori* methods. The results of this approach suggest that the mixture contributes to increase 8-isoprostane, and there is not a set of metabolites that seem to decrease 8-isoprostane.

In the telomere length example, however, we observed that these methods gave conflicting results in estimation of both the negative and positive partial effects. Observing high correlations among exposures and following guidance from Figure 5, we expect that QGCAP, WQSAP, and WQSNS to have the lowest RMSE and bias. There was little prior information to suggest whether the effect would be concentrated or spread across multiple exposures. From simulation results, when the partial effect was concentrated among few exposures, then QGCAP had the lowest bias followed by WQSSS; WQSAP had conflicting RMSE and bias across the negative and positive partial effects. In the data, we expected an imbalance in partial effects, indicating either QGCAP, WQSAP, or

Correlation	Sample Size	Spread	Imbalance	Recommendation
Low QGCAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** QGCSS **Medium RMSE**, ψ^+ and ψ^- have **slight bias** WQSSS **Medium RMSE**, ψ^+ and ψ^- have **moderate bias** away from the null WQSNS **Medium RMSE**, ψ^+ and ψ^- have **moderate bias** toward null (and **invalid CIs**) WQSAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** WQS2i **High RMSE,** ψ^+ and ψ^- have **moderate bias** away from the null	**Small ($n < 500$)** QGCAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** QGCSS **Medium RMSE**, ψ^+ and ψ^- have **Medium bias** WQSSS **Medium RMSE**, ψ^+ and ψ^- have **Medium bias** WQSNS **High RMSE**, ψ^+ and ψ^- have **high bias** WQSAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** WQS2i **High RMSE**, ψ^+ and ψ^- have **high bias**	**Concentrated effect** QGCAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** QGCSS **Medium RMSE**, ψ^+ and ψ^- have **low bias** (except Enet) WQSSS **Low RMSE**, ψ^+ and ψ^- have **medium bias** WQSNS **Low RMSE**, ψ^+ and ψ^- have **high bias** WQSAP **High RMSE** and **high bias** in ψ^-, **low RMSE** and **low bias** in ψ^+ WQS2i **High RMSE, high bias**	**Partial Effects are balanced ($\lvert\psi^+\rvert \approx \lvert\psi^-\rvert$)** QGCAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** QGCSS **Medium RMSE**, ψ^+ and ψ^- have **medium bias** (Enet high bias) WQSSS **Medium RMSE**, ψ^+ and ψ^- have **medium bias** WQSNS **Medium-high RMSE**, ψ^+ and ψ^- have **high bias** WQSAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** WQS2i **Medium-high RMSE**, ψ^+ and ψ^- have **medium bias**	QGCAP and WQSAP, when perfect *a priori* knowledge is available, are most robust to high correlations, low sample sizes, and imbalance across the partial effects. QGCAP also had low RMSE with a concentrated effect across few exposures whereas WQSAP had high RMSE in this scenario QGCSS and WQSSS had overall similar performance across the measures, though in some scenarios QGCSS had superior performance QGCEN had overall poorer performance across measures in comparison to all methods. QGCSS and QGCMA had near identical performance RH-WQS and WQSSS had near identical performance in average bias, but RH-WQS had lower RMSE Highest RMSE and bias observed in WQS2i and QGCEN **Summary:** If you have good *a priori* knowledge, it is recommended you use QGCAP or WQSAP. If *a priori* knowledge is poor, triangulate with multiple approaches using patterns of results: e.g. if correlations are low and WQSSS and QGC disagree, then there is likely an imbalance of effects and the lower bias or lower RMSE method should be slightly preferred.
Medium/High QGCAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** QGCSS **Medium RMSE**, ψ^+ and ψ^- have **medium bias** (Enet has larger bias) WQSSS **Medium RMSE**, ψ^+ and ψ^- have **large bias** away from null WQSNS **Good RMSE**, ψ^+ and ψ^- have **slight bias** away from null (and **invalid CIs**) WQSAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** WQS2i **High RMSE,** ψ^+ and ψ^- have **moderate bias** away from the null	**Large ($n \geq 500$)** QGCAP **Low RMSE,** ψ^+ and ψ^- have **low bias** QGCSS **Low RMSE,** ψ^+ and ψ^- have **negligible bias** WQSSS **Low RMSE,** ψ^+ and ψ^- have **slight bias** WQSNS **Medium-low RMSE,** ψ^+ and ψ^- have **slight bias** WQSAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** WQSAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** WQS2i **High RMSE**, ψ^+ and ψ^- have **high bias**	**Spread out effect** QGCAP **Low RMSE**, ψ^+ and ψ^- have **low bias** QGCSS **Medium RMSE**, ψ^+ and ψ^- have **medium bias** (except Enet has high RMSE/bias) WQSSS **Medium RMSE**, ψ^+ and ψ^- have **medium-low bias** WQSNS **Medium RMSE**, ψ^+ and ψ^- have **high bias** WQSAP **Low RMSE**, ψ^+ and ψ^- have **low bias** WQS2i **High RMSE, moderate bias** in ψ^+ and ψ^-	**Partial Effects are unbalanced ($\lvert\psi^+\rvert \approx \lvert\psi^-\rvert$)** QGCAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** QGCSS **Medium RMSE**, ψ^+ and ψ^- have **medium bias** WQSSS **High RMSE**, ψ^+ and ψ^- have **medium-high bias** WQSNS **High RMSE**, ψ^+ and ψ^- have **medium-high bias** WQSAP **Low RMSE**, ψ^+ and ψ^- have **negligible bias** WQS2i **High RMSE**, ψ^+ and ψ^- have **medium-high bias**	

Figure 5. Partial effects summary and recommendations across methods; recommendations are based on the assumptions of linearity and additivity. Concentrated effect indicates that the largest weights in either positive or negative direction are concentrated on few exposures and spread out effect indicates that the largest weights in either positive or negative direction are spread across many exposures. Note: CI, confidence interval; QGCAP, quantile g-computation with *a priori* knowledge; QGCEN, quantile g-computation with elastic net; QGCMA, quantile g-computation with model-averaging; QGCSS, quantile g-computation with sample-splitting; RH-WQS, repeated holdouts weighted quantile sums regression; RMSE, root mean squared error; WQSAP, weighted quantile sums regression *a priori*; WQSNS, weighted quantile sums regression with no sample-splitting; WQSSS, weighted quantile sums regression with sample-splitting; WQS2i, weighted quantile sums regression two-index model with penalization.

QGCSS as the method with most optimal properties based on simulation results. While the *a priori* methods had the lowest RMSE and bias according to our simulations, the extent of imbalance that we would expect is unclear from prior literature. In this case, we recommend performing either QGCSS or WQSSS across multiple seeds. The simulations with high correlation showed that nearly all estimators will have higher RMSE/higher variance as correlation grows. This can manifest in the sensitivity to seed value for the sample-splitting step and can be conceptualized as a challenge to statistical power, as also observed by Taylor et al.[54]: As exposure correlation grows and the estimation demands grow, then statistical power will decrease and RMSE will tend to increase. Though the NHANES study should have been adequately powered for some analyses, our results demonstrate how this power may not translate to estimating partial effects when procedures like sample-splitting and adjustment for many covariates is needed.

Given the complexities of environmental mixtures, it is possible that at least some exposures in a mixture will have nonlinear relationships and interactions. In this case, it is recommended to instead use methods for estimating the overall mixture effect. Partial effects and the underlying methods used to estimate (as defined and implemented in our study following common usage in the literature) rely on linearity and additivity of effects of exposures. Other commonly used methods that allow nonlinearity and interactions, such as Bayesian kernel machine regression (BKMR),[5] Bayesian multiple index model[55] (BMIM), and generalized linear models[56] (as well as nonlinear implementations of QGC and WQSr), do not result in a single positive and negative partial effect estimate. It is our view that if interactions and nonlinearity are present at meaningful levels, then estimating a single partial effects can be misleadingly simple and methods for the overall effect (e.g., BKMR or QGC without partial effects) should be preferred. Therefore, assessment of interaction and nonlinearity, through either exploratory analysis or formal methods,[57] is warranted prior to focusing on partial effects. These other approaches, as well as exploring nonlinearity and possible interactions in the data, are therefore outside the scope of the study, and we leave further comparisons in light of interactions across methods to future work. We did not assess the performance of Bayesian approaches to either WQSr[58] or QGC. We are unaware of implementations of Bayesian WQSr that estimate partial effects because the default implementation yields a posterior distribution over a parameter that simultaneously represents the positive and negative partial effects, rather than two separate parameters as would be necessary here.

Our study compares QGC and WQSr along with several variant methods in their ability to estimate partial effects in mixtures analysis. In accordance with other studies in mixtures that found improved estimation with the incorporation of prior knowledge,[59] our broad recommendations to practitioners, summarized in Figure 5, are to *a*) use *a priori* knowledge about mixtures wherever possible rather than relying on data to infer harmful from helpful exposures and use either QGCAP or WQSAP to estimate partial effects and *b*) when *a priori* knowledge is poor, report partial effects estimates using several methods to compare and contrast estimates. WQSr methods were designed to fit models with all exposures included,[18,36] though our results suggest that leveraging *a priori* knowledge can reduce bias and error, when available. From our study, we found that in some cases, results are consistent across multiple methods, while in others, different methods pointed to partial effects that differed greatly. We provide illustrative code for our study at https://github.com/mkamenet3/PartialEffects.

Toward addressing inconsistency across results, we have shown that approaches like QGC and WQSr may be unpredictable in their estimation of partial effects in settings where poor prior information exists for partitioning exposures, even when the overall effect of a mixture can be well-estimated. More research and methods development are needed in partial effects estimation of mixtures, particularly approaches that can take into account the complexity of mixtures, such as correlations among exposures, nonlinearity, and interactions. While there has been work in the Bayesian literature on directional homogeneity constraints for the identification and estimation of specific exposures using posterior inclusion probabilities,[55,59] continued development of methods specific to the estimation of partial effects must be performed. The partial effects of interest in a particular study should be justified in terms of the types of insights they can tease out of epidemiologic data and whether partial effects are warranted given the potential for nonlinearity and nonadditivity that is incompatible with their causal interpretation.

References

1. Billionnet C, Sherrill D, Annesi-Maesano I, GERIE study. 2012. Estimating the health effects of exposure to multi-pollutant mixture. Ann Epidemiol 22(2):126–141, PMID: 22226033, https://doi.org/10.1016/j.annepidem.2011.11.004.
2. Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, et al. 2013. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. Environ Health 12(1):85, PMID: 24093917, https://doi.org/10.1186/1476-069X-12-85.
3. Hamra GB, Buckley JP. 2018. Environmental exposure mixtures: questions and methods to address them. Curr Epidemiol Rep 5(2):160–165, PMID: 30643709, https://doi.org/10.1007/s40471-018-0145-0.
4. Joubert BR, Kioumourtzoglou M-A, Chamberlain T, Chen HY, Gennings C, Turyk ME, et al. 2022. Powering research through innovative methods for mixtures in epidemiology (PRIME) program: novel and expanded statistical methods. Int J Environ Res Public Health 19(3):1378, PMID: 35162394, https://doi.org/10.3390/ijerph19031378.
5. Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, et al. 2015. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics 16(3):493–508, PMID: 25532525, https://doi.org/10.1093/biostatistics/kxu058.
6. Keil AP, Buckley JP, O'Brien KM, Ferguson KK, Zhao S, White AJ. 2020. A quantile-based g-computation approach to addressing the effects of exposure mixtures. Environ Health Perspect 128(4):047004, PMID: 32255670, https://doi.org/10.1289/EHP5838.
7. Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. 2015. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. J Agric Biol Environ Stat 20(1):100–120, PMID: 30505142, https://doi.org/10.1007/s13253-014-0180-3.
8. Liu X, Zhang D, Wu X, Tu J, Gong C, Li Y, et al. 2022. Urinary metals as influencing factors of coronary heart disease among a population in Guangzhou, China. Ecotoxicol Environ Saf 241:113746, PMID: 35689886, https://doi.org/10.1016/j.ecoenv.2022.113746.
9. Lee K-S, Kim K-N, Ahn YD, Choi Y-J, Cho J, Jang Y, et al. 2021. Prenatal and postnatal exposures to four metals mixture and IQ in 6-year-old children: a prospective cohort study in South Korea. Environ Int 157:106798, PMID: 34339957, https://doi.org/10.1016/j.envint.2021.106798.
10. Rager JE, Clark J, Eaves LA, Avula V, Niehoff NM, Kim YH, et al. 2021. Mixtures modeling identifies chemical inducers versus repressors of toxicity associated with wildfire smoke. Sci Total Environ 775:145759, PMID: 33611182, https://doi.org/10.1016/j.scitotenv.2021.145759.
11. Bommarito PA, Welch BM, Keil AP, Baker GP, Cantonwine DE, McElrath TF, et al. 2021. Prenatal exposure to consumer product chemical mixtures and size for gestational age at delivery. Environ Health 20(1):68, PMID: 34112176, https://doi.org/10.1186/s12940-021-00724-z.
12. Niehoff NM, Keil AP, O'Brien KM, Jackson BP, Karagas MR, Weinberg CR, et al. 2020. Metals and trace elements in relation to body mass index in a prospective study of US women. Environ Res 184:109396, PMID: 32209500, https://doi.org/10.1016/j.envres.2020.109396.
13. Eaves LA, Keil AP, Jukic AM, Dhingra R, Brooks JL, Manuck TA, et al. 2023. Toxic metal mixtures in private well water and increased risk for preterm birth in North Carolina. Environ Health 22(1):69, PMID: 37845729, https://doi.org/10.1186/s12940-023-01021-7.
14. Czarnota J, Gennings C, Colt JS, De Roos AJ, Cerhan JR, Severson RK, et al. 2015. Analysis of environmental chemical mixtures and non-Hodgkin lymphoma risk in the NCI-SEER NHL study. Environ Health Perspect 123(10):965–970, PMID: 25748701, https://doi.org/10.1289/ehp.1408630.

15. Zhang Y, Dong T, Hu W, Wang X, Xu B, Lin Z, et al. 2019. Association between exposure to a mixture of phenols, pesticides, and phthalates and obesity: comparison of three statistical models. Environ Int 123:325–336, PMID: 30557812, https://doi.org/10.1016/j.envint.2018.11.076.
16. Pacyga DC, Patti MA, Papandonatos GD, Haggerty DK, Calafat AM, Gardiner JC, et al. 2023. Associations of individual and cumulative urinary phthalate and replacement biomarkers with gestational weight gain through late pregnancy. Sci Total Environ 855:158788, PMID: 36116648, https://doi.org/10.1016/j.scitotenv.2022.158788.
17. Wallace ER, Ni Y, Loftus CT, Sullivan A, Masterson E, Szpiro AA, et al. 2022. Prenatal urinary metabolites of polycyclic aromatic hydrocarbons and toddler cognition, language, and behavior. Environ Int 159:107039, PMID: 34902794, https://doi.org/10.1016/j.envint.2021.107039.
18. Day DB, Sathyanarayana S, LeWinn KZ, Karr CJ, Mason WA, Szpiro AA. 2022. A permutation test-based approach to strengthening inference on the effects of environmental mixtures: comparison between single-index analytic methods. Environ Health Perspect 130(8):087010, PMID: 36040702, https://doi.org/10.1289/EHP10570.
19. Ashrap P, Watkins DJ, Calafat AM, Ye X, Rosario Z, Brown P, et al. 2018. Elevated concentrations of urinary triclocarban, phenol and paraben among pregnant women in Northern Puerto Rico: predictors and trends. Environ Int 121(pt 1):990–1002, PMID: 30316544, https://doi.org/10.1016/j.envint.2018.08.020.
20. Watkins DJ, Ferguson KK, Toro LVAD, Alshawabkeh AN, Cordero JF, Meeker JD. 2015. Associations between urinary phenol and paraben concentrations and markers of oxidative stress and inflammation among pregnant women in Puerto Rico. Int J Hyg Environ Health 218(2):212–219, PMID: 25435060, https://doi.org/10.1016/j.ijheh.2014.11.001.
21. Holland N, Huen K, Tran V, Street K, Nguyen B, Bradman A, et al. 2016. Urinary phthalate metabolites and biomarkers of oxidative stress in a Mexican-American cohort: variability in early and late pregnancy. Toxics 4(1):7, PMID: 28008399, https://doi.org/10.3390/toxics4010007.
22. Ferguson KK, Cantonwine DE, Rivera-González LO, Loch-Caruso R, Mukherjee B, Anzalota Del Toro LV, et al. 2014. Urinary phthalate metabolite associations with biomarkers of inflammation and oxidative stress across pregnancy in Puerto Rico. Environ Sci Technol 48(12):7018–7025, PMID: 24845688, https://doi.org/10.1021/es502076j.
23. Cantonwine DE, Cordero JF, Rivera-González LO, Anzalota Del Toro LV, Ferguson KK, Mukherjee B, et al. 2014. Urinary phthalate metabolite concentrations among pregnant women in Northern Puerto Rico: distribution, temporal variability, and predictors. Environ Int 62:1–11, PMID: 24161445, https://doi.org/10.1016/j.envint.2013.09.014.
24. Eick SM, Ferguson KK, Milne GL, Rios-McConnell R, Vélez-Vega C, Rosario Z, et al. 2020. Repeated measures of urinary oxidative stress biomarkers and preterm birth in Puerto Rico. Free Radic Biol Med 146:299–305, PMID: 31704372, https://doi.org/10.1016/j.freeradbiomed.2019.11.003.
25. Ferguson KK, Cantonwine DE, McElrath TF, Mukherjee B, Meeker JD. 2016. Repeated measures analysis of associations between urinary bisphenol-A concentrations and biomarkers of inflammation and oxidative stress in pregnancy. Reprod Toxicol 66:93–98, PMID: 27751756, https://doi.org/10.1016/j.reprotox.2016.10.002.
26. Welch BM, McNell EE, Edin ML, Ferguson KK. 2022. Inflammation and oxidative stress as mediators of the impacts of environmental exposures on human pregnancy: evidence from oxylipins. Pharmacol Ther 239:108181, PMID: 35367517, https://doi.org/10.1016/j.pharmthera.2022.108181.
27. NCHS (National Center for Health Statistics). 2013. *The National Health and Nutrition Examination Survey (NHANES) Analytics and Reporting Guidelines*. Hyattsville, MD: NCHS.
28. Mitro SD, Birnbaum LS, Needham BL, Zota AR. 2016. Cross-sectional associations between exposure to persistent organic pollutants and leukocyte telomere length among U.S. Adults in NHANES, 2001–2002. Environ Health Perspect 124(5):651–658, PMID: 26452299, https://doi.org/10.1289/ehp.1510187.
29. Gibson EA, Nunez Y, Abuawad A, Zota AR, Renzetti S, Devick KL, et al. 2019. An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. Environ Health 18(1):76, PMID: 31462251, https://doi.org/10.1186/s12940-019-0515-1.
30. Keil AP, O'Brien KM. 2024. Considerations and targeted approaches to identifying bad actors in exposure mixtures. Stat Biosci 16:459–481.
31. Ziegler S, Schettgen T, Beier F, Wilop S, Quinete N, Esser A, et al. 2017. Accelerated telomere shortening in peripheral blood lymphocytes after occupational polychlorinated biphenyls exposure. Arch Toxicol 91(1):289–300, PMID: 27146145, https://doi.org/10.1007/s00204-016-1725-8.
32. Robins J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Math Model 7(9–12):1393–1512, https://doi.org/10.1016/0270-0255(86)90088-6.
33. Snowden JM, Rose S, Mortimer KM. 2011. Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. Am J Epidemiol 173(7):731–738, PMID: 21415029, https://doi.org/10.1093/aje/kwq472.
34. Curtin P, Kellogg J, Cech N, Gennings C. 2021. A random subset implementation of weighted quantile sum (WQS_{RS}) regression for analysis of high-dimensional mixtures. Commun Stat Simul Comput 50(4):1119–1134, https://doi.org/10.1080/03610918.2019.1577971.
35. Renzetti S, Curtin P, Just AC, Bello G, Gennings C. 2021. gWQS: Generalized Weighted Quantile Sum Regression. https://CRAN.R-project.org/package=gWQS [accessed 12 November 2024].
36. Tanner EM, Bornehag CG, Gennings C. 2019. Repeated holdout validation for weighted quantile sum regression. MethodsX 6:2855–2860, PMID: 31871919, https://doi.org/10.1016/j.mex.2019.11.008.
37. Keil A. 2023. qgcomp: Quantile G-Computation. https://CRAN.R-project.org/package=qgcomp [accessed 12 November 2024].
38. Gao P. 2021. The exposome in the era of one health. Environ Sci Technol 55(5):2790–2799, PMID: 33417434, https://doi.org/10.1021/acs.est.0c07033.
39. Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 67(2):301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x.
40. Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, et al. 2023. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. https://cran.r-project.org/web/packages/glmnet/index.html [accessed 7 October 2024].
41. Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference*. 2nd ed. New York, NY: Springer.
42. Bartoń K. 2024. MuMIn: multi-Model Inference. https://cran.r-project.org/web/packages/MuMIn/index.html [accessed 7 October 2024].
43. Renzetti S, Gennings C, Calza S. 2023. A weighted quantile sum regression with penalized weights and two indices. Front Public Health 11:1151821, PMID: 37533534, https://doi.org/10.3389/fpubh.2023.1151821.
44. NIH HPC Group. 2025. Biowulf: High Performance Computing at the NIH. https://hpc.nih.gov/ [accessed 8 November 2024].
45. Bommarito PA, Cantonwine DE, Stevens DR, Welch BM, Davalos AD, Zhao S, et al. 2023. An application of group-based trajectory modeling to define fetal growth phenotypes among small-for-gestational-age births in the LIFECODES fetal growth study. Am J Obstet Gynecol 228(3):334.e1–334.e21, PMID: 36027952, https://doi.org/10.1016/j.ajog.2022.08.041.
46. Kim SS, Meeker JD, Aung MT, Yu Y, Mukherjee B, Cantonwine DE, et al. 2020. Urinary trace metals in association with fetal ultrasound measures during pregnancy. Environ Epidemiol 4(2):e075, PMID: 32201854, https://doi.org/10.1097/ee9.0000000000000075.
47. Kuiper JR, O'Brien KM, Ferguson KK, Buckley JP. 2021. Urinary specific gravity measures in the U.S. population: implications for the adjustment of non-persistent chemical urinary biomarker data. Environ Int 156:106656, PMID: 34062395, https://doi.org/10.1016/j.envint.2021.106656.
48. Bommarito PA, Stevens DR, Welch BM, Weller D, Meeker JD, Cantonwine DE, et al. 2023. Temporal trends and predictors of phthalate, phthalate replacement, and phenol biomarkers in the LIFECODES fetal growth study. Environ Int 174:107898, PMID: 37001215, https://doi.org/10.1016/j.envint.2023.107898.
49. Ferguson KK, Rosen EM, Rosario Z, Feric Z, Calafat AM, McElrath TF, et al. 2019. Environmental phthalate exposure and preterm birth in the PROTECT birth cohort. Environ Int 132:105099, PMID: 31430608, https://doi.org/10.1016/j.envint.2019.105099.
50. O'Brien KM, Upson K, Cook NR, Weinberg CR. 2016. Environmental chemicals in urine and blood: improving methods for creatinine and lipid adjustment. Environ Health Perspect 124(2):220–227, PMID: 26219104, https://doi.org/10.1289/ehp.1509693.
51. Cawthon RM. 2002. Telomere measurement by quantitative PCR. Nucleic Acids Res 30(10):e47, PMID: 12000852, https://doi.org/10.1093/nar/30.10.e47.
52. Lan Q, Cawthon R, Shen M, Weinstein SJ, Virtamo J, Lim U, et al. 2009. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of non-Hodgkin lymphoma. Clin Cancer Res Off Res 15(23):7429–7433, PMID: 19934287, https://doi.org/10.1158/1078-0432.CCR-09-0845.
53. Ferguson KK, Lan Z, Yu Y, Mukherjee B, McElrath TF, Meeker JD. 2019. Urinary concentrations of phenols in association with biomarkers of oxidative stress in pregnancy: assessment of effects independent of phthalates. Environ Int 131:104903, https://doi.org/10.1016/j.envint.2019.104903.
54. Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. 2016. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: lessons from an innovative workshop. Environ Health Perspect 124(12):A227–A229, PMID: 27905274, https://doi.org/10.1289/EHP547.
55. McGee G, Wilson A, Webster TF, Coull BA. 2023. Bayesian multiple index models for environmental mixtures. Biometrics 79(1):462–474, PMID: 34562016, https://doi.org/10.1111/biom.13569.

56. McCullagh P, Nelder JA. 1991. *Generalized Linear Models*. 2nd ed. Boca Raton, FL: Chapman and Hall.
57. Midya V, Alcala CS, Rechtman E, Gregory JK, Kannan K, Hertz-Picciotto I, et al. 2023. Machine learning assisted discovery of interactions between pesticides, phthalates, phenols, and trace elements in child neurodevelopment. Environ Sci Technol 57(46):18139–18150, PMID: 37595051, https://doi.org/10.1021/acs.est.3c00848.
58. Colicino E, Pedretti NF, Busgang SA, Gennings C. 2020. Per- and poly-fluoroalkyl substances and bone mineral density. Environ Epidemiol 4(3):e092, PMID: 32613152, https://doi.org/10.1097/EE9.0000000000000092.
59. McGee G, Wilson A, Coull BA, Webster TF. 2023. Incorporating biological knowledge in analyses of environmental mixtures and health. Stat Med 42(17):3016–3031, PMID: 37161723, https://doi.org/10.1002/sim.9765.