

RESEARCH

Open Access



# Associations between mixtures of urinary phthalate metabolites with gestational age at delivery: a time to event analysis using summative phthalate risk scores

Jonathan Boss<sup>1†</sup>, Jingyi Zhai<sup>1†</sup>, Max T. Aung<sup>2</sup>, Kelly K. Ferguson<sup>3</sup>, Lauren E. Johns<sup>2</sup>, Thomas F. McElrath<sup>4</sup>, John D. Meeker<sup>2</sup> and Bhramar Mukherjee<sup>1,5\*</sup>

## Abstract

**Background:** Preterm birth is a significant public health concern and exposure to phthalates has been shown to be associated with an increased odds of preterm birth. Even modest reductions in gestational age at delivery could entail morbid consequences for the neonate and analyzing data with this additional information may be useful. In the present analysis, we consider gestational age at delivery as our outcome of interest and examine associations with multiple phthalates.

**Methods:** Women were recruited early in pregnancy as part of a prospective, longitudinal birth cohort at the Brigham and Women's Hospital in Boston, Massachusetts. Urine samples were collected at up to four time points during gestation for urinary phthalate metabolite measurement, and birth outcomes were recorded at delivery. From this population, we selected all 130 cases of preterm birth (< 37 weeks of gestation) as well as 352 random controls. We conducted analysis with both geometric average of the exposure concentrations across the first three visits as well as using repeated measures of the exposure. Two different time to event models were used to examine associations between nine urinary phthalate metabolite concentrations and time to delivery. Two different approaches to constructing a summative phthalate risk score were also considered.

**Results:** The single-pollutant analysis using a Cox proportional hazards model showed the strongest association with a hazard ratio (HR) of 1.21 (95% confidence interval (CI): 1.09, 1.33) per interquartile range (IQR) change in average log-transformed mono-2-ethyl-5-carboxypentyl phthalate (MECPP) concentration. Using the accelerated failure time model, we observed a 1.19% (95% CI: 0.26, 2.11%) decrease in gestational age in association with an IQR change in average log-transformed MECPP. We next examined associations with an environmental risk score (ERS). The fourth quartile of ERS was significantly associated with a HR of 1.44 (95% CI: 1.19, 1.75) and a reduction of 2.55% (95% CI: 0.76, 4.30%) in time to delivery (in days) compared to the first quartile.

**Conclusions:** On average, pregnant women with higher urinary metabolite concentrations of individual phthalates have shorter time to delivery. The strength of the observed associations are amplified with the risk scores when compared to individual pollutants.

**Keywords:** Hazard ratio, Phthalate, Pregnancy, Environmental risk score, Time to delivery, Mixtures

\* Correspondence: [bhramar@umich.edu](mailto:bhramar@umich.edu)

<sup>†</sup>Jonathan Boss and Jingyi Zhai contributed equally to this work.

<sup>1</sup>Department of Biostatistics, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, USA

<sup>5</sup>Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA

Full list of author information is available at the end of the article



## Background

Phthalate diesters are produced in large quantities yearly in the US for use in everyday products such as polyvinyl flooring, shower curtains, food packaging plastics, and personal care products. Exposure occurs through contact with these products as well as the consumption of contaminated food and drinking water [1–3]. Phthalate exposure has been related to various health outcomes in humans, including altered thyroid and reproductive hormone levels [4, 5], decreased semen quality in males [6], and asthma and allergic symptoms [7]. Exposure to phthalates in utero has been linked to adverse birth outcomes as well, including altered reproductive tract development in male infants [8], neurodevelopment in both sexes [9, 10], and both prematurity and small size at birth [11–14]. Preterm birth, defined as delivery before 37 weeks completed gestation, is a particularly important endpoint of interest due to: 1) its contribution to neonatal mortality and morbidity and consequent cost to society; 2) the apparent increase in rates over the last three decades; and 3) poorly understood causes and lack of effective interventions [15]. Not only is preterm birth the leading cause of perinatal and infant mortality, but it is also associated with adverse developmental outcomes in children, including chronic conditions such as cardiovascular disease and endocrine disorders [15–17]. The societal costs of preterm birth comprise high medical expenditures and utilization and consequently places economic burdens on healthcare systems [15]. Research to uncover contributing causes, particularly those in connection with environmental contaminant exposures, is a public health priority [18].

We recently demonstrated clear associations between maternal urinary phthalate metabolite concentrations averaged from multiple time points during pregnancy and increased odds of preterm birth in a nested case-control study ( $N=130$  cases,  $N=352$  controls) of women who delivered at the Brigham and Women's Hospital in Boston [14]. In follow-up analyses we examined variability in phthalate levels across pregnancy, attempted to identify any patterns in levels by gestational age, and assessed associations between phthalate exposure biomarkers at individual time points during pregnancy and preterm birth in order to identify windows of vulnerability [19]. Upon human exposure, phthalates are metabolized quickly, therefore single spot urine samples are less predictive of long-term exposure, and measurement of multiple urine samples are more reliable [20]. While these results suggested that the third trimester of pregnancy may be particularly sensitive for the relationship between phthalate exposure and early delivery, the strongest observed effect estimates were with the more stable metric of average phthalate exposure over gestation.

Studying preterm birth as a binary outcome is reasonable because variation in gestational age at delivery around 40 weeks can be due to misclassification. Thus, examining a cutoff such as 37 weeks focuses on pregnancies that are most likely to be truly early, and that are clinically significant. However, some studies indicate that “preterm birth” is not a homogeneous condition; early delivery—regardless of gestational age—is associated with poorer health outcomes in childhood. For example, late preterm birth (delivery at 34–36 weeks gestation) is associated with increased likelihood of cerebral palsy and other developmental disorders such as reduced mental index scores compared to births that occur at term (after 37 weeks gestation) [21, 22]. Thus, examining the relationship between environmental exposures and the rest of the gestational age distribution could be biologically meaningful.

In this present analysis, we consider time to delivery as our primary outcome of interest instead of the binary outcome of term vs. preterm birth. We consider both average exposure during pregnancy as well as repeated measures of exposure across pregnancy as potential correlates of time to delivery in this framework. In addition to analyzing each compound separately, we constructed two aggregate summaries of total phthalate exposure, an environmental risk score [22] and a weighted quantile sum [23], in relation to time to delivery. These aggregate summary analyses attempt to capture the effect of mixtures on time to delivery.

## Methods

### Study population

Participants were part of an ongoing prospective cohort study of pregnant women with initial prenatal visits at clinics in the Boston area. All women who wished to participate were included if they planned to deliver at the Brigham and Women's Hospital and if their initial visit was prior to 17 weeks gestation. Subjects were followed throughout the course of pregnancy and provided information (e.g., health status, weight) and urine samples at up to four visits. Urine samples were refrigerated (4 C) for a maximum of two hours before being processed and frozen (–80 C) for long-term storage. At delivery, birth outcome characteristics such as mode of delivery and fetal measurements were recorded. From 2006 to 2008 approximately 1600 women were recruited, and 1181 were followed until delivery and had live singleton infants. From this population, the present nested case-control study includes all 130 mothers who delivered preterm, as well as 352 controls selected randomly from subjects who had a urine sample from visit 1 and from at least one additional visit.

Gestational ages at individual visits and at delivery were calculated based on last menstrual period (LMP)

and confirmed by first trimester ultrasound. Study participants provided written informed consent and institutional review board approval was obtained from Brigham and Women's Hospital and the University of Michigan. Within this study, visit 1 urine samples were taken at median 9.71 weeks gestation (range 4.71 to 16.1 weeks), visit 2 at median 17.9 weeks (range 14.9 to 21.9 weeks), visit 3 at median 26.0 weeks (range 22.9 to 29.3 weeks), and visit 4 at median 35.1 weeks (range 33.1 to 38.3 weeks). The number of subjects with samples available decreased slightly with increasing visit, with the fourth visit having the smallest number of samples. Visit 4 also had a smaller proportion of cases with urine samples, since some had delivered by this time point.

### Phthalate exposure

Nine phthalate metabolites were measured in each available urine sample ( $N = 1693$ ) by NSF International in Ann Arbor, MI, following methods developed by the Centers for Disease Control (CDC), described in detail elsewhere [24, 25]. The final number of samples analyzed for all phthalate metabolites were as follows by visit (cases, controls): Visit 1 (129, 350); Visit 2 (118, 304); Visit 3 (111, 301); and Visit 4 (66, 314). Phthalate measurements below the limit of detection (LOD) were replaced with the LOD divided by  $\sqrt{2}$  [26].

To adjust for urinary dilution, specific gravity (SG) levels were also measured in each urine sample using a digital handheld refractometer (ATAGO Company Ltd., Tokyo, Japan). For univariate analyses phthalate levels were corrected for urinary SG using the following formula:  $P_C = P[(M_{SG} - 1)/(SG - 1)]$ , where  $P_C$  represents the SG-corrected phthalate concentration (micrograms per liter),  $P$  represents the measured concentration in urine,  $M_{SG} = 1.015$  is the median SG of all samples measured, and  $SG$  represents the SG of the individual sample [12].

For regression models, unadjusted phthalate levels were used and urinary SG was included as a covariate, since modeling adjusted phthalate levels may incur bias [27]. In the analysis of individual phthalate metabolites, we additionally examined a summed measure of di(2-ethylhexyl) phthalate (DEHP) metabolites ( $\Sigma$ DEHP; nanomoles/liter) that is typically calculated as an index of total exposure to the parent compound. All individual metabolites and  $\Sigma$ DEHP were log-normally distributed and ln-transformed for analysis.

### Descriptive analysis

The nested case-control sample was appropriately weighted via inverse probability weighting in order to compute representative descriptive statistics and to make

inference on time to birth for the overall cohort. All analyses utilized these weights. Population-level summary statistics were computed for demographic characteristics, including race, maternal age, education, and health insurance provider (public vs. private). Race, maternal age, and education were included as covariates in all single-pollutant and multi-pollutant analysis models. The distribution of each log-transformed phthalate metabolite and final gestational age were assessed via histogram. To evaluate the concordance between different phthalates, a Pearson correlation matrix between log-transformed and specific gravity corrected average phthalate metabolite concentrations was tabulated.

### Analysis with average exposure

#### Single-pollutant models

One common strategy to model repeated exposure measurements is to compute the average contaminant concentration for each individual and subsequently use the average exposure concentration in the model of interest. In this context, average exposure for a particular phthalate metabolite refers to the average of the log-transformed concentrations obtained at the first three visits. We excluded the fourth visit from our average exposure analysis because women with preterm deliveries were less likely to have a fourth visit. We will consider three such mean exposure analysis models: (1) Logistic Regression, (2) Cox Proportional Hazards Model, and (3) Accelerated Failure Time Model. All coefficients are reported in the unit of one interquartile range (IQR) change of the exposure under consideration. Standardizing by IQR allows us to compare two subjects with identical covariates, one of which is at the 75th percentile of exposure and the other is at the 25th percentile of exposure.

*Logistic Regression with Preterm Birth as Outcome:* Let  $T_i$  denote the gestational age at delivery for subject  $i$  (in days) and let  $E_i = I(T_i < 37 \text{ wks})$  indicate whether subject  $i$  had a preterm birth, where  $i = 1, \dots, 479$ . Then the single-pollutant logistic regression model for the  $k$ th phthalate can be expressed as:

$$\text{logit}(\pi_{ik}) = \beta_{0k} + \beta_{1k} \bar{X}_{ik} + \mathbf{Z}_i^T \boldsymbol{\beta}_{2k}, \quad (1)$$

where  $\pi_{ik} = P(E_i = 1 | \bar{X}_{ik}, \mathbf{Z}_i)$ ,  $\bar{X}_{ik}$  is the mean of the  $k$ th phthalate (log-transformed) divided by the IQR of the log-transformed  $k$ th phthalate for the  $i$ th individual,  $k = 1 \dots, 9$ , and  $\mathbf{Z}_i^T$  is a vector of baseline covariates for the  $i$ th individual (race, education, maternal age, average specific gravity across the first three visits, and health insurance status). The fitted coefficients  $\exp(\hat{\beta}_{1k})$  provide an estimated odds ratio of preterm birth for one IQR increase in average log-transformed phthalate levels. Similar results were presented in Ferguson and colleagues (2014), but are

included to compare with the repeated measures single-pollutant exposure results [14].

While logistic regression is easy to implement and interpret, it has the drawback of discretizing gestational age at delivery. Thus, we consider gestational age at delivery as a continuous outcome and fit two commonly used time to event models. The distinction from a standard survival context is that everyone in the study experiences the event and the distribution of time to delivery is left skewed (instead of a survival time distribution, that is typically right skewed).

**Cox Proportional Hazards Model:** The first obvious and natural way to correlate the outcome of gestational age at delivery to phthalate levels is to model the hazard of having birth at time  $t$ ,  $\lambda_k(t)$ , as a function of the covariates and the  $k$ -th phthalate metabolite,

$$\lambda_k(t) = \lambda_{0k}(t) \exp(\alpha_{1k}\bar{X}_{ik} + \mathbf{Z}_i^T \alpha_{2k}), \tag{2}$$

for  $k = 1, \dots, 9$ , where  $\bar{X}_{ik}$  and  $\mathbf{Z}_i^T$  are defined above. The fitted coefficients  $\exp(\hat{\alpha}_{1k})$  provide an estimated hazard ratio of giving birth at time  $t$  for a one IQR change in average phthalate exposure (log-transformed). Numbers larger than one indicate an increased hazard ratio of delivery and, therefore, a shorter time to delivery.

**Accelerated Failure Time Model:** Although Cox proportional hazards model is the most commonly used model for time to event data, our direct objective of associating gestational age at delivery (rather than modeling the instantaneous hazard at time  $t$ ) to phthalate levels is better addressed by the accelerated failure time model which, in this case, simply reduces to modeling the log of gestational age as a normal linear regression model:

$$\log(T_i) = \gamma_{0k} + \gamma_{1k}\bar{X}_{ik} + \mathbf{Z}_i^T \gamma_{2k} + \sigma \epsilon_i, \tag{3}$$

where  $\bar{X}_{ik}$  and  $\mathbf{Z}_i^T$  are defined above and the errors  $\epsilon_i$  are independent standard normal variates. Both models provide natural interpretations of their respective parameter estimates. For the accelerated failure time model,  $100 \times (\exp(\hat{\gamma}_{1k}) - 1)$ , yields the percent change in gestational age at delivery per IQR increase in the  $k$ th log-transformed phthalate exposure. Negative numbers indicate shorter time to delivery.

**Multi-pollutant models**

While the standard practice has been to consider single-pollutant models, in reality, we are exposed to mixtures of multiple agents. A direct approach to modeling multiple pollutants is to construct a joint multivariate model with all 9 phthalate metabolites. This approach is often not feasible due to lack of sample size and potential multicollinearity among exposures. Two recently proposed strategies consider weighted sum type metrics that represent the composite effect of mixtures.

The first such method is to construct an environmental risk score (ERS) [22]. ERS is calculated as a linear combination of the individual contaminant exposures, weighted by their associated regression coefficients obtained from a given model. Another method of quantifying aggregate exposure to multiple environmental contaminants is through a weighted quantile sum (WQS), where adaptive weights corresponding to chosen quantiles of phthalates are estimated by bootstrapping the data [23]. In both summative risk scores, the weights are derived from a model of the association between chemical mixtures and the health outcome of interest.

To avoid multicollinearity issues incurred by the inclusion of all 9 phthalates, we adopted two approaches to select subsets of phthalates for ERS and WQS construction. The first approach examined the correlation structure of the phthalates (Additional file 1: Table S1). Five phthalates in the upper-left block show strong correlation (greater than 0.5) and, among them, we chose the one that showed the strongest association with gestational age at delivery (MECPP). The ERS and WQS determined by this empirical examination of correlations were subsequently based on six phthalate metabolites: MECPP, MBzP, MBP, MiBP, MEP, and MCPP. We call the two risk scores based on such ad hoc screening of the correlation structure “ERS-Corr” and “WQS-Corr.” The second approach constituted of running stepwise logistic regression with all 9 candidate phthalates and selecting the phthalates that were retained at the end of the stepwise procedure. The risk scores based on the phthalates selected by stepwise variable selection are called “ERS-Stepwise” and “WQS-Stepwise.” Details regarding data adaptive weight construction for these risk scores are relegated to Additional file 1: Appendix A1.

Analogous to single-pollutant models (1), (2), and (3), similar models with the continuous summative risk scores (or categorized values of the risk scores), generically denoted as  $RS$  in the following expressions, were fit. All risk scores were standardized by their respective IQR in order to facilitate comparisons between risk scores.

**Logistic Regression:**  $\text{logit}(\pi_i^{RS}) = \beta_0^{RS} + \beta_1^{RS} RS_i + \mathbf{Z}_i^T \beta_2^{RS}$ , where  $\pi_i^{RS} = P(E_i = 1 | RS_i, \mathbf{Z}_i)$ .

**Cox proportional hazards model:**  $\lambda^{RS}(t) = \lambda_0^{RS}(t) \exp(\alpha_1^{RS} RS_i + \mathbf{Z}_i^T \alpha_2^{RS})$ .

**Accelerated failure time model:**  $\log(T_i) = \gamma_0^{RS} + \gamma_1^{RS} RS_i + \mathbf{Z}_i^T \gamma_2^{RS} + \sigma \epsilon_i$ .

**Analysis with repeated measures of exposure**

**Single-pollutant models**

In a traditional repeated measures situation, repeated measures are taken on the outcome of interest, potentially adjusted for time independent covariates. However, in our scenario, we have up to four repeated measures per exposure in each subject. To capture the variation in phthalate

levels across pregnancy, we consider a two-step method as described in Chen and colleagues (2015) [28]. The two-step method consists of: (a) fitting a linear mixed effects model with random intercepts to the repeated measures of the phthalate levels (b) extracting the estimated subject-specific intercepts to be used as a predictor in the second-step outcome model, akin to the mean analysis presented through models (1), (2), and (3).

Note that the differences between average exposure across visits and subject-specific intercepts are small. However, using subject-specific intercepts is a more general approach as random slopes or other features may be incorporated into the stage 1 linear mixed model (LMM). Moreover, the LMM framework better addresses subjects with differing numbers of visits, because BLUPs are shrinkage estimates of subject-specific averages relative to the population average. In that respect, we feel that the LMM framework is a statistically principled analog to the more ad hoc approach of taking a simple exposure average. The details of this fitting process are provided in Additional file 1: Appendix A2.

**Multi-pollutant models**

For the repeated measures analysis, we will only focus on ERS-Corr and ERS-Stepwise (we cannot use WQS, because the outcome in the linear mixed effects model needs to be continuous and WQS is inherently discrete by construction). We repeat the process of constructing the ERS at each time point (Visits 1–4) and fit a random intercept linear mixed model to the repeated measures of ERS at each time point. That is, we first fit:

$$ERS_{ij} = b_{0i} + \phi_0 + \phi_1 T_{ij} + \phi_2 SG_{ij} + \epsilon_{ij},$$

where  $SG_{ij}$  is the specific gravity for the  $i$ th subject at the  $j$ th visit and  $b_{0i} \sim N(0, \sigma_b^2)$  and  $\epsilon_{ij} \sim N(0, \sigma^2)$  are independent. Let  $\hat{b}_{0i}$  be the best linear unbiased predictors (BLUP) of the subject-specific random intercepts, extracted from a standard linear mixed effects model output, and let  $\hat{b}_{0i}^*$  denote the IQR standardized BLUP. Then our final analysis models are of the form:

*Logistic regression model:*

$$\text{logit}(\pi_i^{RS}) = \beta_0^{RS} + \beta_1^{RS} \hat{b}_{0i}^* + \mathbf{Z}_i^T \boldsymbol{\beta}_2^{RS}, \text{ where } \pi_i^{RS} = P(E_i = 1 | \hat{b}_{0i}^*, \mathbf{Z}_i).$$

*Cox proportional hazards model:*

$$\lambda^{RS}(t) = \lambda_0^{RS}(t) \exp\left(\alpha_1^{RS} \hat{b}_{0i}^* + \mathbf{Z}_i^T \boldsymbol{\alpha}_2^{RS}\right).$$

*Accelerated failure time model:*

$$\log(T_i) = \gamma_0^{RS} + \gamma_1^{RS} \hat{b}_{0i}^* + \mathbf{Z}_i^T \boldsymbol{\gamma}_2^{RS} + \sigma^* \epsilon_i.$$

All analyses were performed using R statistical software, version 3.4.4 ([www.r-project.org](http://www.r-project.org)). WQS was implemented using the gWQS package in R [29].

**Results**

Summary statistics for the study population with respect to demographic characteristics such as race, education, maternal age, and health insurance status can be found in Additional file 1: Table S2. Overall, the cohort primarily had private health insurance and was highly educated, with 79.9% of the study participants having privatized health insurance and 83.6% of women completing some postsecondary education at a college or technical school. There were minimal differences between cases and controls with respect to race, education, maternal age, and health insurance status. Descriptive summary characteristics for the distribution of phthalate metabolite concentrations are provided in supplementary Table S3. All contaminant distributions are right-skewed and each phthalate metabolite has a very low percentage of non-detects, with the largest being 4.70% of values below LOD for MEHP. Additional file 1: Figure S1 shows that the distribution of gestational length is heavily left-skewed.

Table 1 summarizes the odds ratios, hazard ratios, and percent change in the single-pollutant mean exposure analysis models (see Additional file 1: Table S4 for single-pollutant IQR values used in interpreting model coefficients). MEHP (OR: 1.50, 95% CI: 1.10, 2.07), MECPP (OR: 1.66, 95% CI: 1.20, 2.30), and summed DEHP metabolites (OR: 1.47, 95% CI: 1.06, 2.03) all showed elevated odds of preterm birth per IQR change in their respective mean log-transformed concentrations adjusted for average specific gravity, race, education, maternal age, and health insurance status. Considering gestational age as a continuous outcome, the Cox proportional hazards model identifies MECPP (HR: 1.21, 95% CI: 1.09, 1.33), summed DEHP metabolites (HR: 1.14, 95% CI: 1.04, 1.26), MBzP (HR: 1.15, 95% CI: 1.03, 1.27), MBP (HR: 1.17, 95% CI: 1.05, 1.29), and MCPP (HR: 1.10, 95% CI: 1.01, 1.20), as having a significant HR of delivery per IQR change in their respective mean log-transformed concentrations. The single-pollutant accelerated failure time models identify MECPP as having a 1.19% (95% CI: 0.26, 2.11%) decrease in final gestational age in days and summed DEHP as having a 1.03% (95% CI, 0.01, 1.95%) decrease in final gestational age in days for one IQR higher in average log-transformed MECPP and summed DEHP, respectively.

Table 2 summarizes the odds ratios, hazard ratios, and percent change in the single-pollutant repeated measures analysis models using random intercepts (see Additional file 1: Table S4 for BLUP IQR values used in interpreting model coefficients). First-step models adjusted for time-varying specific gravity and all second-step models adjusted for race, education, maternal age, and health insurance status. MEHP (OR: 1.40, 95% CI: 1.06, 1.85), MECPP (OR: 1.43, 95% CI: 1.12,

**Table 1** Single-pollutant associations between average phthalate exposures and gestational age

Phthalate Metabolite	Logistic*		Cox		AFT	
	OR	95% CI	HR	95% CI	% Change	95% CI
MEHP	<b>1.50</b>	<b>(1.10, 2.07)</b>	1.07	(0.98, 1.17)	-0.79%	(-1.67, 0.09%)
MEHHP	1.05	(0.75, 1.47)	1.04	(0.94, 1.15)	-0.51%	(-1.01, 0.32%)
MEOHP	1.23	(0.87, 1.75)	1.09	(0.98, 1.21)	-0.76%	(-1.74, 0.23%)
MECPP	<b>1.66</b>	<b>(1.20, 2.30)</b>	<b>1.21</b>	<b>(1.09, 1.33)</b>	<b>-1.19%</b>	<b>(-2.11, -0.26%)</b>
Σ DEHP	<b>1.47</b>	<b>(1.06, 2.03)</b>	<b>1.14</b>	<b>(1.04, 1.26)</b>	<b>-1.03%</b>	<b>(-1.95, -0.01%)</b>
MBzP	1.27	(0.79, 1.65)	<b>1.15</b>	<b>(1.03, 1.27)</b>	-0.60%	(-1.63, 0.43%)
MBP	1.35	(0.99, 1.87)	<b>1.17</b>	<b>(1.05, 1.29)</b>	-0.74%	(-1.75, 0.28%)
MiBP	0.98	(0.66, 1.45)	1.00	(0.90, 1.12)	0.30%	(-0.82, 1.44%)
MEP	1.22	(0.87, 1.71)	1.01	(0.91, 1.12)	-0.11%	(-1.06, 0.86%)
MCPP	1.27	(0.93, 1.72)	<b>1.10</b>	<b>(1.01, 1.20)</b>	-0.28%	(-1.18, 0.62%)

Phthalates were averaged across the first three visits and log-transformed. All models were adjusted for average specific gravity, maternal age at first visit, race, and education. The single-pollutant models for MBzP, MBP, MiBP, MEP, and MCPP are also adjusted for health insurance provider. Bolded cells indicate significant ( $p < 0.05$ ) odds ratios (OR), hazard ratios (HR), and percent changes (% Change). Odds ratios, hazard ratios and percent changes are all calculated on IQR (Interquartile Range) scale. Abbreviations: CI, Confidence Interval; Logistic, Logistic Regression; Cox, Cox Proportional Hazards Model; AFT, Accelerated Failure Time Model; MEHP, Mono-(2-ethyl)-hexyl phthalate; MEHHP, Mono-(2-ethyl-5-hydroxyhexyl) phthalate; MEOHP, Mono-(2-ethyl-5-oxohexyl) phthalate; MECPP, Mono-(2-ethyl-5-carboxypentyl) phthalate; Σ DEHP, Summed DEHP metabolites; MBzP, Mono-benzyl phthalate; MBP, Mono-n-butyl phthalate; MiBP, Mono-isobutyl phthalate; MEP, Mono-ethyl phthalate; MCPP, Mono-(3-carboxypropyl) phthalate. \*Results are similar to Ferguson and colleagues (2014) [14]

1.83), and summed DEHP metabolites (OR: 1.32, 95% CI: 1.01, 1.74) showed higher odds of a preterm birth per IQR change in the subject-specific random intercept. MECPP (HR: 1.11, 95% CI: 1.03, 1.19), MBzP (HR: 1.13, 95% CI: 1.05, 1.22), MBP (HR: 1.11, 95% CI: 1.04, 1.19), and MCPP (HR: 1.06, 95% CI: 1.00, 1.12) showed an elevated risk of shortened gestational length per IQR change in their respective mean log-transformed concentrations. Moreover, summed DEHP metabolites had a nearly significant hazard ratio after accounting for the repeated measures of DEHP metabolites (HR: 1.07, 95% CI: 0.99, 1.16). In the accelerated failure time model,

MECPP (% Change: -0.74, 95% CI: -1.14, -0.03%) was the only metabolite that was significantly associated with a percent decrease in final gestational age per IQR change in the subject-specific MECPP random intercept.

Overall, results for the repeated measures analysis are consistent with the mean exposure analysis. One notable difference is that the results in Table 2 are generally attenuated relative to the results in Table 1. Intuitively, this is because extracting summaries of phthalates by random effects and associating them with gestational length are “noisier” than directly using averaged measurements.

**Table 2** Single-pollutant associations between repeated measures of phthalate exposure and gestational age

Phthalate Metabolite	Logistic		Cox		AFT	
	OR	95% CI	HR	95% CI	% Change	95% CI
MEHP	<b>1.40</b>	<b>(1.06, 1.85)</b>	1.03	(0.96, 1.12)	-0.52%	(-1.28, 0.26%)
MEHHP	0.97	(0.74, 1.26)	0.98	(0.91, 1.07)	-0.14%	(-0.90, 0.63%)
MEOHP	1.10	(0.85, 1.42)	1.02	(0.95, 1.10)	-0.33%	(-1.07, 0.40%)
MECPP	<b>1.43</b>	<b>(1.12, 1.83)</b>	<b>1.11</b>	<b>(1.03, 1.19)</b>	<b>-0.74%</b>	<b>(-1.14, -0.03%)</b>
Σ DEHP	<b>1.32</b>	<b>(1.01, 1.74)</b>	1.07	(0.99, 1.16)	-0.64%	(-1.14, 0.13%)
MBzP	1.08	(0.81, 1.42)	<b>1.13</b>	<b>(1.05, 1.22)</b>	-0.50%	(-1.28, 0.28%)
MBP	1.18	(0.97, 1.46)	<b>1.11</b>	<b>(1.04, 1.19)</b>	-0.44%	(-1.11, 0.24%)
MiBP	0.93	(0.72, 1.19)	0.98	(0.92, 1.05)	0.33%	(-0.38, 1.04%)
MEP	1.10	(0.82, 1.46)	0.98	(0.90, 1.06)	0.04%	(-0.78, 0.87%)
MCPP	1.19	(0.95, 1.48)	<b>1.06</b>	<b>(1.00, 1.12)</b>	-0.19%	(-0.84, 0.46%)

Phthalate measurements at each visit were log-transformed. All models were adjusted for specific gravity at the respective visit, time of sample collection, maternal age at first visit, race, and education. The single-pollutant models for MBzP, MBP, MiBP, MEP, and MCPP are also adjusted for health insurance provider. Bolded cells indicate significant ( $p < 0.05$ ) odds ratios (OR), hazard ratios (HR), and percent changes (% Change). Odds ratios, hazard ratios and percent changes are all calculated on IQR (Interquartile Range) scale. Abbreviations: CI, Confidence Interval; Logistic, Logistic Regression; Cox, Cox Proportional Hazards Model; AFT, Accelerated Failure Time Model; MEHP, Mono-(2-ethyl)-hexyl phthalate; MEHHP, Mono-(2-ethyl-5-hydroxyhexyl) phthalate; MEOHP, Mono-(2-ethyl-5-oxohexyl) phthalate; MECPP, Mono-(2-ethyl-5-carboxypentyl) phthalate; Σ DEHP, Summed DEHP metabolites; MBzP, Mono-benzyl phthalate; MBP, Mono-n-butyl phthalate; MiBP, Mono-isobutyl phthalate; MEP, Mono-ethyl phthalate; MCPP, Mono-(3-carboxypropyl) phthalate

Descriptive analyses for the phthalate risk scores showed that ERS-Corr and ERS-Stepwise (Additional file 1: Figure S2) are approximately normally distributed, whereas WQS-Corr and WQS-Stepwise are decidedly non-normal. Additional file 1: Figure S3 contains a Pearson correlation matrix between the four risk scores. There are moderately strong, pairwise correlations (average correlation around  $r = 0.6$ ) between ERS-Corr, WQS-Corr, and WQS-Stepwise, but ERS-Stepwise is weakly correlated with WQS-Corr and WQS-Stepwise. Across average exposure and repeated measures models, MEOHP and MECPP had the largest contribution to the construction of ERS-Stepwise and WQS-Stepwise, MECPP, MBP, and MiBP had the largest contribution to the construction of ERS-Corr, and MECPP, MBzP, and MEP had the largest contribution to the construction of WQS-Corr (see Additional file 1: Table S5 for a list of exact weights used in ERS and WQS construction).

Table 3 summarizes the odds ratios, hazard ratios, and percent change in the ERS and WQS average exposure analysis models, where ERS and WQS were determined using mean log-transformed phthalate concentrations. All models were adjusted for specific gravity, race, education, maternal age, and health insurance status. One IQR change in ERS-Corr (OR: 1.81, 95% CI: 1.32, 2.52), ERS-Stepwise (OR: 2.14, 95% CI: 1.62, 2.87), WQS-Corr (OR: 1.66, 95% CI: 1.06, 2.64), and WQS-Stepwise (OR: 1.64, 95% CI: 1.01, 2.72) were all associated with a higher odds of preterm birth. For the Cox proportional hazards model, ERS-Stepwise (HR: 1.30, 95% CI: 1.16, 1.46) and WQS-Corr (HR: 1.21, 95% CI: 1.06, 1.38) showed a significantly higher risk for shortened gestational length per IQR change, while ERS-Corr (HR: 1.06, 95% CI: 0.98, 1.14) and WQS-Stepwise (HR: 1.06, 95% CI: 0.92, 1.23) did not show a significantly higher risk for shortened gestational length per IQR change. In the accelerated failure time model, ERS-Corr (% Change: -1.86, 95% CI: -2.98, -0.73%),

ERS-Stepwise (% Change: -1.84, 95% CI: -2.78, -0.88%), and WQS-Corr (% Change: -1.12, 95% CI: -2.25, -0.08%) showed a significant percent decrease in gestational length per IQR change in the respective risk score. See Additional file 1: Table S4 for the IQR ranges of each multi-pollutant risk score.

Figure 1 depicts the odds ratios, hazard ratios, and percent change in gestational age for ERS and WQS quartiles (see Additional file 1: Table S6 for numerical summaries). For WQS-Corr, ERS-Corr, and ERS-Stepwise, there is generally an increasing trend in the odds ratios and hazard ratios and a decreasing trend in the % change as the respective risk score quartile increases. Namely, WQS-Corr shows significantly higher odds of preterm birth (OR: 3.33, 95% CI: 1.44, 7.69), significantly higher risk for shortened gestational length (HR: 1.48, 95% CI: 1.16, 1.89), and a significant percent decrease in gestational length (% Change: -2.89, 95% CI: -5.01, -0.71%) for quartile 4 compared to quartile 1.

Two single-pollutant mean exposure models for MECPP and MBP are also included in Fig. 1 for comparison. Notably, the odds ratios, hazard ratios, and % change corresponding to the single-pollutant models are attenuated relative to the ERS-Corr, ERS-Stepwise, and WQS-Corr models. Specifically, if we compare quartile 4 to quartile 1, then we see that ERS-Corr has an odds ratio of 3.77 (95% CI: 1.96, 7.25), a hazard ratio of 1.44 (95% CI: 1.19, 1.75), and a % change of -2.55% (95% CI: -4.30, -0.76%), while the single-pollutant mean exposure models for MBP have an odds ratio of 2.25 (95% CI: 1.16, 4.36), a hazard ratio of 1.30 (95% CI: 1.00, 1.69), and % change of -1.63% (95% CI: -4.03, 0.82%). Thus, measures of aggregate phthalate exposure, notably ERS-Corr, ERS-Stepwise, and WQS-Stepwise, quantify a stronger association between phthalate exposure and gestational length, compared to single-pollutant models.

**Table 3** Association of gestational age with summative phthalate risk scores

Risk Score	Logistic		Cox		AFT	
	OR	95% CI	HR	95% CI	% Change	95% CI
Average Exposure Analysis						
ERS-Corr	<b>1.81</b>	<b>(1.32, 2.52)</b>	1.06	(0.98, 1.14)	<b>-1.86%</b>	<b>(-2.98, -0.73%)</b>
ERS-Stepwise	<b>2.14</b>	<b>(1.62, 2.87)</b>	<b>1.30</b>	<b>(1.16, 1.46)</b>	<b>-1.84%</b>	<b>(-2.78, -0.88%)</b>
WQS-Corr	<b>1.66</b>	<b>(1.06, 2.64)</b>	<b>1.21</b>	<b>(1.06, 1.38)</b>	<b>-1.12%</b>	<b>(-2.25, -0.08%)</b>
WQS-Stepwise	<b>1.64</b>	<b>(1.01, 2.72)</b>	1.06	(0.92, 1.23)	-0.57%	(-1.83, 0.70%)
Repeated Measures Analysis						
ERS-Corr	<b>1.89</b>	<b>(1.45, 2.51)</b>	<b>1.19</b>	<b>(1.10, 1.27)</b>	<b>-1.33%</b>	<b>(-2.03, -0.63%)</b>
ERS-Stepwise	<b>1.77</b>	<b>(1.37, 2.31)</b>	<b>1.23</b>	<b>(1.14, 1.34)</b>	<b>-0.85%</b>	<b>(-1.60, -0.10%)</b>

Phthalate measurements at each visit were log-transformed. All average exposure models were adjusted for average specific gravity, maternal age at first visit, race, education, and health insurance provider. All repeated measures models were adjusted for specific gravity at each visit, time of sample collection, maternal age at first visit, race, education, and health insurance provider. Bolded cells indicate significant ( $p < 0.05$ ) odds ratios (OR), hazard ratios (HR), and percent changes (% Change). Odds ratios, hazard ratios and percent changes are all calculated on IQR (Interquartile Range) scale. Abbreviations: CI, Confidence Interval; Logistic, Logistic Regression; Cox, Cox Proportional Hazards Model; AFT, Accelerated Failure Time Model

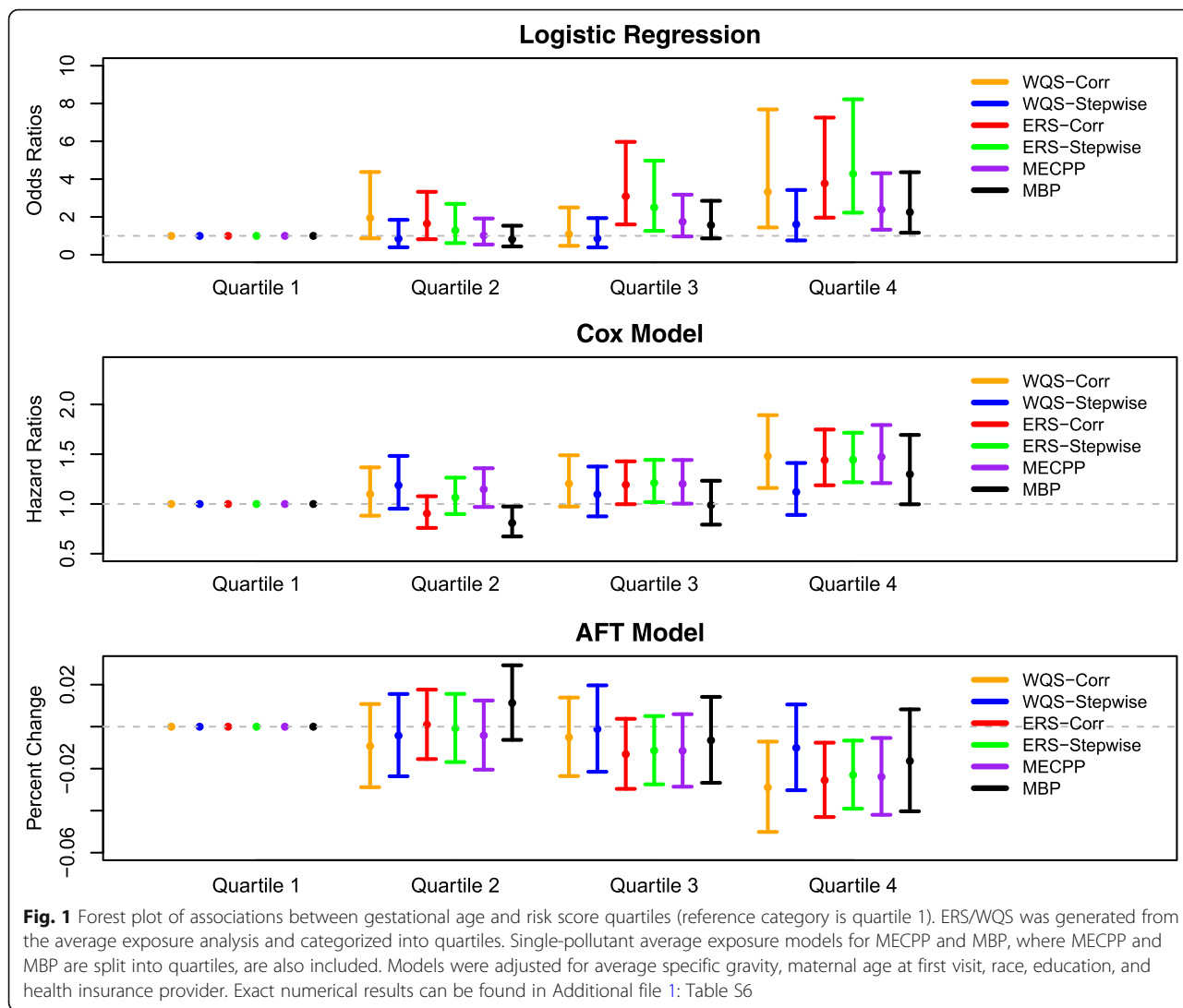


Table 3 also summarizes the odds ratios, hazard ratios, and percent change in the ERS exposure analysis models, where ERS is determined using repeated measures of phthalate concentrations. For ERS-Corr (OR: 1.89, 95% CI: 1.45, 2.51; HR: 1.19, 95% CI: 1.10, 1.27; % Change: -1.33, 95% CI: -2.03, -0.63%) and ERS-Stepwise (OR: 1.77, 95% CI: 1.37, 2.31; HR: 1.23, 95% CI: 1.14, 1.34; % Change: -0.85, 95% CI: -1.60, -0.10%) we observe an elevated odds, higher risk of lower gestational length, and a percent decrease in gestational length per IQR change in their respective ERS (see Random Intercept column in Additional file 1: Table S4 for repeated measures ERS IQR ranges).

**Discussion**

In this paper, we make two primary contributions. The first is to analyze gestational age at delivery in a time to event framework. Modeling time to delivery as a continuous variable addresses the limitations of using

dichotomous outcomes such as term versus preterm birth, which may oversimplify the pathological consequences of shorter gestational periods. The second is to use phthalate risk scores, such as ERS and WQS, as summary measures to estimate the cumulative effect of phthalate mixtures. We examined four different risk scores, and found that three of the four, ERS-Corr, ERS-Stepwise, and WQS-Corr, were significantly associated with time to delivery. Simulation studies need to be conducted to better understand the analytical benefits and drawbacks of using ERS compared to WQS.

Overall, this study provides further evidence that select phthalates are associated with risk factors for adverse reproductive and birth outcomes. Namely, several individual phthalates, such as MECPP and summed DEHP metabolites, were associated with an increased odds of preterm delivery, an increased risk of delivering, and a reduced gestational duration. Associations from the repeated measures analysis for individual pollutants appear



to be attenuated compared to the associations from the average exposure analysis. This is primarily due to the limited number of repeated measures for each subject and the additional variability that is introduced as a result of computing the BLUP estimates in the first-stage model. When comparing the models with multi-pollutant risk scores to single pollutant models, we found that ERS-Corr and WQS-Corr were associated with an even greater reduction in gestational duration and higher odds of preterm birth in comparison to individual phthalate metabolites. These findings align with the hypothesis that mixtures of multiple pollutants may have greater adverse effects in comparison to single-pollutants evaluated in isolation.

Analytically, logistic regression, Cox proportional hazards model, and accelerated failure time model (AFT) estimate different quantities, and thus the significance and interpretation of single phthalate metabolites and summative phthalate risk scores are likely to vary across the three models. Logistic regression estimates odds ratios, intrinsically related to a dichotomized outcome and most commonly used measure in this context. The hazard ratio estimate obtained from the Cox proportional hazards model can be understood as a ratio of the hazard rates for women with high and low phthalate exposure, respectively. A hazard ratio greater than one indicates that a woman with greater phthalate exposure has a higher probability of instantaneously giving birth compared to a woman with lower phthalate exposure, given that both women have not delivered up to that time point. Although the Cox proportional hazards model is very popular for time to event data, the interpretation in the context of gestational duration is somewhat unusual as everybody experiences the event and the time to delivery data is left skewed instead of being right skewed (typically noted for survival outcomes). The AFT model has a much simpler interpretation; namely, a negative coefficient implies that, on average, women with higher phthalate exposure would expect a reduction in gestational duration compared to women with lower phthalate exposure. Given the ease of interpretation on the direct gestational age scale, we believe that AFT is better suited for studies of time to gestation.

Preterm delivery, defined as gestational duration less than 37 weeks, can be an informative and clinically relevant outcome to predict maternal and child health outcomes. However, there are also other dichotomous cutoffs, such as late preterm (34–36 weeks gestation), moderate preterm (32–33 weeks gestation), very preterm (< 32 weeks gestation) and extremely premature birth (< 28 weeks gestation) [30, 31]. Even among term pregnancies that result in delivery after 37 weeks gestation, there is considerable variation in days of gestation prior to delivery [32]. Furthermore, post-term births that take place after 42 weeks

gestation also contribute to perinatal morbidity [30]. There may be distinct and overlapping pathological consequences associated with each of these gestational age ranges, therefore, analyzing continuous gestational age as an outcome variable may characterize a more accurate understanding of the relationship between maternal phthalate levels and overall duration of pregnancy. Our study reported a reduction in gestational duration that ranged between 0 and 3% for most predictor variables, which we recognize may not be clinically significant on the individual level. However, given the ubiquity of phthalate exposure, we emphasize that the reduction in gestational duration associated with phthalate exposure, averaged among all births that take place in the U.S. population, could have widespread societal level effects.

Several investigators have previously characterized associations between concentrations of phthalate metabolites during pregnancy and either gestational age or preterm birth, however, their methodological approach differed from our present study – previous studies have mostly estimated associations with gestational age through linear regression [12, 13, 31, 33–39]. Although these previous studies do not address phthalate mixtures, some of these studies have found parallel findings to our single-pollutant analyses. Weinberger and colleagues (2013) assessed the relationship between phthalate metabolites and gestational age in 72 women from New Jersey, and reported that maternal urinary MEHHP was associated with a decrease in gestational age [31]. In another study of 68 women from Michigan, Watkins and colleagues (2016) observed an inverse relationship between the sum of DBP metabolites (MBP, MHBP, MCPP) at delivery and gestational age [38]. Polanska and colleagues (2016) observed significant inverse associations between maternal MEP in the third trimester and gestational age from a prospective birth cohort in Poland ( $N=165$ ) [36]. In a cohort of strictly African American and Dominican women in New York ( $N=331$ ), Whyatt and colleagues (2009) found that shorter gestational duration was significantly associated with higher maternal urinary concentrations of MEHP, MEHHP, MEOHP, and MECPP in the third trimester [13]. In addition to these studies, a small case-control study of women from Mexico ( $N=60$ ) reported that maternal urinary MECPP, MBP, and MCPP during the third trimester was significantly associated with an increase in the odds of preterm birth [12].

Aside from maternal urinary levels of phthalate metabolites, one of the studies focused on metabolite concentrations in cord blood [34]. In this study of 207 women from China, associations between cord blood levels of phthalates and gestational age were estimated [34]. Concentrations of several phthalates in cord blood (DMP, DEP, DEEP, DPP, BMPP, DNHP, BBP, DNOP, DMEP,

DBP, DIBP, DBEP, and DNP) were significantly associated with shorter gestational age [34]. Another study in Italy ( $N = 84$ ) found lower gestational age among infants with detectable cord blood concentrations of MEHP in comparison to infants without detectable MEHP [35].

Among the existing studies reviewed here, some have also found contrary or null results in comparison to our study. Adibi and colleagues (2009) drew from a multi-center U.S. pregnancy cohort ( $N = 283$ ), and found maternal urinary concentrations of the metabolites MEHP, MEOHP, and MEHPP to be significantly associated with lower odds of preterm birth [33]. Meanwhile, these investigators also reported significant increase in odds for delivery past 41 weeks gestation in relation to higher urinary concentrations of MEHP, MEOHP, and MEHPP [33]. Similar to these findings, a study of 404 women in New York also reported a positive association between maternal MEHP concentrations in the third trimester in relation to longer gestational age [39]. Another study of pregnant women in Japan ( $N = 149$ ) by Suzuki and colleagues (2010) resulted in non-significant associations between 9 different phthalate metabolites (MMP, MEP, MnBP, MBzP, MEHP, MEHHP, and MEOHP) and gestational age [37]. Contrasting findings from previous studies may be due to differences in exposure assessment – most of these studies measured phthalates using single spot urine samples. Another reason could be due to geographical dissimilarities in the location of study participants, given that phthalate exposure may vary by country and region.

Phthalates are metabolized quickly in the body, and as a mixture, they may be interacting with several target tissues to elicit changes in various endogenous signaling molecules, such as hormones, and markers of inflammation and oxidative stress [20, 40]. There are several potential mechanisms by which phthalate mixtures can disrupt the production and circulation of endogenous biomarkers, due to their ability to interact with nuclear receptors and transcription factors, such as estrogen and progesterone receptors, aryl hydrocarbon receptors, peroxisome proliferator-activated receptors, and thyroid receptors [40–42]. Human and animal studies indicate that select phthalate metabolites have been associated with disruption of several circulating hormones, inflammation and oxidative stress markers [42–45]. With concern for reproductive health, phthalate exposure may alter gestational duration by acting through these mechanistic pathways. Through shortened gestational duration, phthalate exposure may contribute to adverse neonatal outcomes and child morbidity later in life.

Though we present a comprehensive analytical framework to capture time and multiple pollutants in an omnibus analysis, there are several methodological limitations. First, the accelerated failure time model in our

analysis, which is typically used for right-skewed outcomes, is modeling a left-skewed outcome, namely gestational duration. Therefore, as a sensitivity check, we fit accelerated failure time models on a transformed version of gestational age, such that the log of the transformed gestational age was normally distributed (results not presented). In terms of significance and direction, transformed and untransformed time to delivery produced consistent findings. Regression parameters for the models with untransformed time to delivery are more straightforward to interpret, however, in modeling a log-transformed left-skewed outcome, estimates of the percent reduction in gestational age may be heavily influenced by a small number of subjects with very short gestational duration. Second, ERS is calculated and used on the same data and thus has the potential for overfitting. Before using the phthalate ERS as a prognostic tool, one needs to validate it in an independent cohort. Third, we did not collect data on dietary patterns prior to or during pregnancy, which may confound the relationship between phthalates and gestational duration.

One major challenge in multipollutant modeling is the selection of etiologically relevant contaminant mixtures in the presence of potentially highly collinear exposures. WQS is specifically designed to handle moderately correlated predictors, however variable selection properties of WQS under a high degree of multicollinearity are not well-studied. Czarnota and colleagues (2015) argued that variable selection in the presence of moderate multicollinearity using an ad hoc threshold for WQS weights outperforms regularized regression methods such as elastic net, but there is no theoretical justification for their claim or the choice of the threshold [46]. Moreover, Czarnota and colleagues (2015) note that they expect WQS to have worse performance when subject to highly correlated contaminants [46]. Given that logistic regression is known to have poor performance under strong multicollinearity, we would also expect our ERS construction method to have difficulty with highly collinear phthalate metabolites [47]. In such situations, ridge regression type methods may have more desirable properties for constructing risk scores though they do not lead to unbiased estimates of each of the separate coefficients.

In our multipollutant models, ERS is based on a model with linear phthalate main effects and does not capture potential interactions or non-linearity in the response surface. However, there are multiple strategies for constructing exposure risk scores that do simultaneously account for nonlinearity and high order interactions in the response surface [48]. Bayesian Additive Regression Trees (BART) sum individual regression trees together to estimate a flexible multivariable function of exposures that is associated with the health outcome of interest

[49]. Similar to BART, Bayesian Kernel Machine Regression (BKMR) also aims to estimate a multivariable function of exposures that is associated with the health outcome of interest, but instead uses kernels to approximate a wide array of possible functional forms [50]. Both methods should be considered when it is contextually important to incorporate interactions between exposures into the ERS generative model.

Additionally, we want to point out that ERS and WQS are both typically calculated using the same structure of the analysis model. Ideally, the ERS should be constructed based on fitting Cox regression on training data and validating on the test data. However, in our present analysis, ERS and WQS are each generated from a logistic regression model and are subsequently used as explanatory variables in Cox proportional hazards models and AFT models. The main issue with using continuous gestational age in ERS/WQS construction is that ERS/WQS needs to be generated separately for Cox regression and AFT, because model parameters for Cox regression and AFT correspond to different interpretable quantities (hazard ratio and percent reduction in gestational age, respectively) and are on different scales. Another reason for using ERS from models with preterm birth as an outcome is that most of the published data are available on this outcome rather than considering gestational age as a time to event outcome. Thus if other investigators wanted to construct ERS based on coefficients reported in other published studies, the ERS we proposed would be comparable.

Lastly, our study is also limited by the reality that phthalates are highly variable, and measurements reflect recent exposures [20]. We previously reported interclass correlation coefficients (ICC) of phthalates from this study population [19], which represents the ratio of intra-individual variability to the sum of intra and inter-individual variability and range from zero to one, where values equal to one indicate no intra-individual variability [51]. ICC for phthalates in this study population ranged from 0.19 to 0.61, indicating low to moderate intra-individual variability. As such, our exposure assessment of phthalates may suffer from some degree of non-differential measurement error.

Despite these limitations, our study has several strengths. First, we obtained up to four urine samples from a large cohort of pregnant women. Compared to single spot urine measurements, having multiple repeated measurements affords a robust exposure assessment, and reduces non-differential measurement error due to intra-individual phthalate variability. Our study also used highly sensitive analytical methods to measure urinary phthalate metabolites, which resulted in high detection rates. Finally, our study was strengthened by our assessment of gestational dates, which were validated both clinically and with first-trimester ultrasonography.

Finally, we would like to emphasize that for non-detects below the LOD, substitution by  $\text{LOD}/\sqrt{2}$  may not always lead to optimal statistical properties. This is less of a concern in our analysis as a very small proportion of phthalate metabolite concentrations are below their respective LODs (Table S3). For studies with higher proportions of non-detects one may want to use more rigorous approaches for handling non-detects, such as multiple imputation or censored likelihood maximization [52, 53]. In such situations,  $\text{LOD}/\sqrt{2}$  substitution could heavily bias regression parameter estimates for constructing ERSs, even if the proportion of non-detects is relatively small, i.e., around 15–20% [54–56].

## Conclusions

Environmental exposure to phthalates remain a persistent public health concern, especially within the context of pregnancy. The present study determined that several phthalates and phthalate risk scores, which quantify the isolated effect of a single phthalates and the aggregate effect of multiple phthalates, respectively, were associated with shortened gestational duration in the Cox proportional hazards models, the accelerated failure time models, and logistic regression models. Furthermore, this study provides a novel statistical framework for investigators to analyze the simultaneous effect of multiple pollutants. Future studies should aim to characterize potential biological mediators that relate phthalate exposure and gestational duration.

## Additional file

**Additional file 1:** Outlining descriptive characteristics of the nested case-control study population, ERS and WQS Construction, and additional modeling details. **Table S1.** Pearson correlation between mean sg-adjusted phthalate exposures. **Table S2.** Descriptive statistics for pregnant women in the nested case-control sample. **Table S3.** Descriptive Statistics for Phthalate Metabolites. **Table S4.** Interquartile Range (IQR) Standardization Values. **Table S5.** Weights for ERS and WQS Construction. **Table S6.** Association between preterm birth / gestational age and categorized phthalate risk score. **Figure S1.** Distribution of final gestational age for subjects included in the nested case-control sample. **Figure S2.** Distribution of ERS and WQS corresponding to the average exposure analysis. **Figure S3.** Scatterplot matrix and Pearson correlation between phthalate risk scores. Appendix A1: Construction of ERS and WQS in Average Exposure Analysis. Appendix A2: Single-pollutant models with repeated measures of exposure. (DOCX 469 kb)

## Abbreviations

BLUP: Best linear unbiased predictors; DEHP: Di(2-ethylhexyl) phthalate; ERS: Environmental risk score; IQR: Interquartile range; LMP: Last menstrual period; LOD: Limit of detection; MBP: Mono-n-butyl phthalate; MBzP: Mono-benzyl phthalate; MCPP: Mono-(3-carboxypropyl) phthalate; MECPP: Mono-(2-ethyl-5-carboxypentyl) phthalate; MEHHP: Mono-(2-ethyl-5-hydroxyhexyl) phthalate; MEHP: Mono-(2-ethyl)-hexyl phthalate; MEOHP: Mono-(2-ethyl-5-oxohexyl) phthalate; MEP: Mono-ethyl phthalate; MiBP: Mono-isobutyl phthalate; SG: Specific gravity; WQS: Weighted quantile sum

## Acknowledgements

Subject recruitment and sample collection was originally funded by Abbott Diagnostics.

### Funding

This work was supported by National Institute of Environmental Health Sciences, National Institutes of Health (grants R01ES018872, P42ES017198, P01ES022844, P30ES017885, P50ES026049, U2CES026553). Support for Max Aung was provided in part by a grant from the Robert Wood Johnson Foundation Health Policy Research Scholars program. Support for Kelly Ferguson was provided by the Intramural Research Program of the National Institute of Environmental Health Sciences.

### Availability of data and materials

The dataset analyzed during the current study is not publicly available due to the sensitive nature of biological samples and demographic variables of the human subjects dataset, but is available from the corresponding author on reasonable request. A simulated dataset and R code supporting the conclusions of this article are available at: <https://github.com/kktzjy/Phthalate-and-gestational-age-analysis>. The simulated dataset was created to closely mimic study population characteristics, phthalate metabolite concentrations, and the gestational duration distribution.

### Authors' contributions

JZ conducted statistical analyses for the study. JB constructed tables and figures for the results, and wrote part of the manuscript. MA and LJ contributed in literature review and writing part of the manuscript. KF contributed in cleaning and preparing the data for analysis and wrote part of the manuscript. TM, JM, and BM contributed in study design. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

The present study received institution review board approval from the Brigham and Women's Hospital.

### Consent for publication

All authors have participated in preparing and reviewing this manuscript, and have subsequently approved the attached, submitted manuscript.

### Competing interests

The authors declared that they have no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Department of Biostatistics, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, USA. <sup>2</sup>Department of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, MI, USA. <sup>3</sup>Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, Durham, NC, USA. <sup>4</sup>Division of Maternal and Fetal Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>5</sup>Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA.

Received: 23 February 2018 Accepted: 8 June 2018

Published online: 20 June 2018

### References

- Agency for Toxic Substances and Disease Registry (ATSDR). 2002. Toxicological profile for Di (2-ethylhexyl) phthalate (DEHP). Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service.
- Agency for Toxic Substances and Disease Registry (ATSDR). 2001. Toxicological profile for Di-n-butyl Phthalate. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service.
- Agency for Toxic Substances and Disease Registry (ATSDR). 1995. Toxicological profile for diethyl phthalate. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service.
- Mendiola J, Meeker JD, Jorgensen N, Andersson A-M, Liu F, Calafat AM, et al. Urinary concentrations of di(2-ethylhexyl) phthalate metabolites and serum reproductive hormones: pooled analysis of fertile and infertile men. *J Androl*. 2012;33:488–98.
- Meeker JD, Ferguson KK. Relationship between urinary phthalate and bisphenol A concentrations and serum thyroid measures in U.S. adults and adolescents from the National Health and nutrition examination survey (NHANES) 2007–2008. *Environ Health Perspect*. 2011;119:1396–402.
- Hauser R, Meeker JD, Duty S, Silva MJ, Calafat AM. Altered semen quality in relation to urinary concentrations of phthalate monoester and oxidative metabolites. *Epidemiology*. 2006;17:682–91.
- Bornehag CG, Nanberg E. Phthalate exposure and asthma in children. *Int J Androl*. 2010;33:333–45.
- Swan SH, Main KM, Liu F, Stewart SL, Kruse RL, Calafat AM, et al. Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environ Health Perspect*. 2005;113:1056–61.
- Engel SM, Zhu C, Berkowitz GS, Calafat AM, Silva MJ, Miodovnik A, et al. Prenatal phthalate exposure and performance on the neonatal behavioral assessment scale in a multiethnic birth cohort. *Neurotoxicology*. 2009;30:522–8.
- Engel SM, Miodovnik A, Canfield RL, Zhu C, Silva MJ, Calafat AM, et al. Prenatal phthalate exposure is associated with childhood behavior and executive functioning. *Environ Health Perspect*. 2010;118:565–71.
- Meeker JD. Exposure to environmental endocrine disruptors and child development. *Arch Pediatr Adolesc Med*. 2012;166:E1–7.
- 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*. 2009;117:1587–92.
- Whyatt RM, Adibi JJ, Calafat AM, Camann DE, Rauh VA, Bhat HK, et al. Prenatal di(2-ethylhexyl) phthalate exposure and length of gestation among an inner-city cohort. *Pediatrics*. 2009;124:e1213–20.
- Ferguson KK, McElrath TF, Meeker JD. Environmental phthalate exposure and preterm birth. *JAMA Pediatr*. 2014;168:61–7.
- Institute of Medicine (US) Committee on Understanding Premature Birth and Assuring Healthy Outcomes. Preterm birth: Causes, Consequences, and Prevention. Washington D.C.: National Academies Press; 2007.
- Moster D, Lie RT, Markestad T. Long-term medical and social consequences of preterm birth. *N Engl J Med*. 2008;359:262–73.
- Callaghan WM, MacDorman MF, Rasmussen SA, Qin C, Lackritz EM. The contribution of preterm birth to infant mortality rates in the United States. *Pediatrics*. 2006;118:1566–73.
- Ashton DM, Lawrence HC III, Adams NL III, Fleischman AR. Surgeon general's conference on the prevention of preterm birth. *Obstet Gynecol*. 2009;113:925–30.
- Ferguson KK, McElrath TF, Ko Y-A, Mukherjee B, Meeker JD. Variability in urinary phthalate metabolite levels across pregnancy and sensitive windows of exposure for the risk of preterm birth. *Environ Int*. 2014;70:118–24.
- Barr DB, Wang RY, Needham LL. Biologic monitoring of exposure to environmental chemicals throughout the life stages: requirements and issues for consideration for the National Children's study. *Environ Health Perspect*. 2005;113:1083–91.
- Petrini JR, Dias T, McCormick MC, Massolo ML, Green NS, Escobar GJ. Increased risk of adverse neurological development for late preterm infants. *J Pediatr*. 2009;154:169–76.
- Woythaler MA, McCormick MC, Smith VC. Late preterm infants have worse 24-month neurodevelopmental outcomes than term infants. *Pediatrics*. 2011;127:e622–9.
- Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. *J Agric Biol Environ Stat*. 2014;20:100–20.
- Silva MJ, Samandar E, Preau JL Jr, Reidy JA, Needham LL, Calafat AM. Quantification of 22 phthalate metabolites in human urine. *J Chromatogr B*. 2007;860:106–12.
- Lewis RC, Meeker JD, Peterson KE, Lee JM, Pace GG, Cantoral A, et al. Predictors of urinary bisphenol A and phthalate metabolite concentrations in Mexican children. *Chemosphere*. 2013;93:2390–8.
- Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hygiene*. 1990;5:46–51.
- 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.
- 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 <https://doi.org/10.1186/1476-069X-14-9>.
- Renzetti S, Curtin P, Just AC, Gennings C. gWQS: generalized weighted quantile sum regression. R; 2016. Available from: <https://cran.r-project.org/package=gWQS>

30. Slama R, Ballester F, Casas M, Cordier S, Merete E, Iniguez C, et al. Epidemiologic tools to study the influence of environmental factors on fecundity and pregnancy-related outcomes. *Epidemiol Rev.* 2014;36:148–64.
31. Weinberger B, Ventrano AM, Archer FE, Marcella SW, Buckley B, Wartenberg D, et al. Effects of maternal exposure to phthalates and bisphenol a during pregnancy on gestational age. *J Matern Neonatal Med.* 2014;27:323–7.
32. Jukic AM, Baird DD, Weinberg CR, McConaughy DR, Wilcox AJ. Length of human pregnancy and contributors to its natural variation. *Hum Reprod.* 2013;28:2848–55.
33. Adibi JJ, Hauser R, Williams PL, Whyatt RM, Calafat AM, Nelson H, et al. Maternal urinary metabolites of Di-(2-Ethylhexyl) phthalate in relation to the timing of labor in a US multicenter pregnancy cohort study. *Am J Epidemiol.* 2009;169:1015–24.
34. Huang Y, Li J, Garcia JM, Lin H, Wang Y, Yan P, et al. Phthalate levels in cord blood are associated with preterm delivery and fetal growth parameters in Chinese women. *PLoS One.* 2014;9:e87430.
35. Latini G, De Felice C, Presta G, Del Vecchio A, Paris I, Ruggieri F, et al. In utero exposure to Di-(2-ethylhexyl)phthalate and duration of human pregnancy. *Environ Health Perspect.* 2003;111:1783–5.
36. Polanska K, Ligocka D, Sobala W, Hanke W. Effect of environmental phthalate exposure on pregnancy duration and birth outcomes. *Int J Occup Med Environ Health.* 2016;29:683–97.
37. Suzuki Y, Niwa M, Yoshinaga J, Mizumoto Y, Serizawa S, Shiraishi H. Prenatal exposure to phthalate esters and PAHs and birth outcomes. *Environ Int.* 2010;36:699–704.
38. Watkins DJ, Milewski S, Domino SE, Meeker JD, Padmanabhan V. Maternal phthalate exposure during early pregnancy and at delivery in relation to gestational age and size at birth: a preliminary analysis. *Reprod Toxicol.* 2016;65:59–66.
39. Wolff MS, Engel SM, Berkowitz GS, Ye X, Silva MJ, Zhu C, et al. Prenatal phenol and phthalate exposures and birth outcomes. *Environ Health Perspect.* 2008;116:1092–7.
40. Benjamin S, Masai E, Kamimura N, Takahashi K, Anderson RC, Faisal PA. Phthalates impact human health: epidemiological evidences and plausible mechanism of action. *J Hazard Mater.* 2017;340:360–83.
41. Kiyama R, Wada-Kiyama Y. Estrogenic endocrine disruptors: molecular mechanisms of action. *Environ Int.* 2015;83:11–40.
42. Thompson PA, Khatami M, Baglole CJ, Sun J, Harris SA, Moon E-Y, et al. Environmental immune disruptors, inflammation and cancer risk. *Carcinogenesis.* 2015;36:S232–53.
43. Andra SS, Makris KC. Thyroid disrupting Chemicals in Plastic Additives and Thyroid Health. *J Environ Sci Heal Part C Environ Carcinog Ecotoxicol Rev.* 2012;30:107–51.
44. Lovekamp-Swan T, Davis BJ. Mechanisms of phthalate Ester toxicity in the female reproductive system. *Environ Health Perspect.* 2003;111:139–45.
45. Sharpe RM. Pathways of endocrine disruption during male sexual differentiation and masculinisation. *Best Pract Res Clin Endocrinol Metab.* 2006;20:91–110.
46. Czarnota J, Gennings C, Wheeler DC. Assessment of weighted quantile sum regression for modeling chemical mixtures and cancer risk. *Cancer Inform.* 2015;14:159–71.
47. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in Epidemiologic studies. *Epidemiology.* 2016;6:227.
48. Park SK, Zhao Z, Mukherjee B. Construction of environmental risk score beyond standard linear models using machine learning methods: application to metal mixtures, oxidative stress and cardiovascular disease in NHANES. *Environ Health.* 2017;16:102.
49. Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. *Ann Appl Stat.* 2010;4:266–98.
50. Bobb JF, Valeri L, Henn BC, Christiani DC, Wright RO, Mazumdar M, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics.* 2015;16:493–508.
51. Rosner BA. *Fundamentals of biostatistics.* 7th ed. Boston: Brooks Cole, Cengage Learning; 2011.
52. Bernhardt PW, Wang HJ, Zhang D. Statistical methods for generalized linear models with covariates subject to detection limits. *Stat Biosci.* 2015;7:68–89.
53. May RC, Ibrahim JG, Chu H. Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits. *Stat Med.* 2011;30:2551–61.
54. Wu H, Chen Q, Ware LB, Koyama T. A Bayesian approach for generalized linear models with explanatory biomarker measurement variables subject to detection limit - an application to acute lung injury. *J Appl Stat.* 2012;39:1733–47.
55. Nie L, Chu H, Liu C, Cole SR. Linear regression with an independent variable subject to a detection limit. *Epidemiology.* 2010;21:S17–24.
56. Lynn HS. Maximum likelihood inference for left-censored HIV RNA data. *Stat Med.* 2001;20:33–45.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

