Title

Using Birth Cohort Data to Estimate Prenatal Chemical Exposures for All Births around the New Bedford Harbor Superfund Site in Massachusetts.

Journal

Environmental Health Perspectives, 127(8)

Authors

Khalili, Roxana
Levy, Jonathan
Fabian, M
et al.

Publication Date

2019-08-01

DOI

10.1289/EHP4849

Peer reviewed

A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP4849.

Using Birth Cohort Data to Estimate Prenatal Chemical Exposures for All Births around the New Bedford Harbor Superfund Site in Massachusetts

Roxana Khalili,[1] Scott M. Bartell,[1,2,3,4] Jonathan I. Levy,[5,6] M. Patricia Fabian,[5,6] Susan Korrick,[6,7] and Verónica M. Vieira[1,2]

[1]Environmental Health Sciences Graduate Program, Susan and Henry Samueli College of Health Sciences, University of California, Irvine, Irvine, California, USA
[2]Program in Public Health, Susan and Henry Samueli College of Health Sciences, University of California, Irvine, Irvine, California, USA
[3]Department of Statistics, Donald Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, California, USA
[4]Department of Epidemiology, School of Medicine, Susan and Henry Samueli College of Health Sciences, University of California, Irvine, Irvine, California, USA
[5]Department of Environmental Health, Boston University School of Public Health, Boston, Massachusetts, USA
[6]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[7]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

Background: Children born near New Bedford, Massachusetts, have been prenatally exposed to multiple environmental chemicals, in part due to an older housing stock, maternal diet, and proximity to the New Bedford Harbor (NBH) Superfund site. Chemical exposure measures are not available for all births, limiting epidemiologic investigations and potential interventions.

Objective: We linked biomonitoring data from the New Bedford Cohort (NBC) and birth record data to predict prenatal exposures for all contemporaneous area births.

Methods: We used prenatal exposure biomarker data from the NBC, a population-based cohort of 788 mother–infant pairs born during 1993–1998 to mothers living near the NBH, linked to their corresponding Massachusetts birth record data, to build predictive models for cord serum polychlorinated biphenyls (expressed as a sum, ΣPCBs), *p*,*p*′-dichlorodiphenyl dichloroethylene (DDE), hexachlorobenzene (HCB), cord blood lead (Pb), and maternal hair mercury (Hg). We applied the best fit models (highest pseudo R^2), with multivariable smooths of continuous variables, to predict exposure biomarkers for all 10,270 births during 1993–1998 around the NBH. We used 10-fold cross validation to validate the exposure models and the bootstrap method to characterize sampling variability in the exposure predictions.

Results: The 10-fold cross-validated R^2 for the ΣPCBs, DDE, HCB, Pb, and Hg exposure models were 0.54, 0.40, 0.34, 0.46, and 0.40, respectively. For each exposure model, multivariable smooths of continuous variables improved the fit compared with linear models. Other variables with significant effects on exposure estimates were paternal education, maternal race/ethnicity, and maternal ancestry. The resulting exposure predictions for all births had variability consistent with the NBC measured exposures.

Conclusions: Predictive models using multivariable smoothing explained reasonable amounts of variance in prenatal exposure biomarkers. Our analyses suggest that prenatal chemical exposures can be predicted for all contemporaneous births in the same geographic area by modeling available biomarker data for a subset of that population. https://doi.org/10.1289/EHP4849

Introduction

Many potential epidemiologic analyses are constrained by the ability to obtain reliable individual-level chemical exposure estimates. It is cost prohibitive and logistically challenging to measure exposures for large populations, and for retrospective analyses or investigations using administrative data sets, it is often impossible. In multiple contexts, researchers have developed strategies to assign exposures to large populations by leveraging measurements for a subset of locations or participants. For example, in air pollution studies (Hoek et al. 2008), researchers collect measurements at a representative subset of geographic locations and construct land use regression (LUR) models to explain variability in those measurements. LUR models evaluate the relationship between observed air pollution concentrations and predictor variables such as land use and traffic conditions in a multivariable regression model to estimate concentrations at unmonitored locations. A defining feature of LUR models intended for epidemiologic studies is that they use as predictors covariates that are available for all geographic locations under study so that all study participants can have exposure assignment. Although less common, this regression approach can also be used to evaluate the relationship between measured biomarkers and predictor variables such as sociodemographic covariates in a subset of the population to predict exposure levels for the entire population.

Analogous to LUR models, we developed exposure regression models for biomarkers of prenatal exposure to organochlorines and metals using data obtained from pregnant women enrolled in the New Bedford Cohort (NBC) and predicted prenatal exposures for all contemporaneous births in the study area. The NBC is an ongoing birth cohort study of 788 mother–infant pairs born during 1993–1998 to mothers residing in one of the four towns adjacent to the polychlorinated biphenyl (PCB)–contaminated New Bedford Harbor (NBH) Superfund site (Figure 1) in Massachusetts. The NBH was designated a Superfund site by the U.S. Environmental Protection Agency (EPA) in 1982 because of extensive PCB contamination from local electronics manufacturing facilities (U.S. EPA 2015). Heavy metals, including lead (Pb), are also present in NBH sediments (Shine et al. 1995) and the area contains a number of other hazardous waste sites and industrial sources of pollution. The region's old housing stock (and associated Pb paint and Pb-containing water distribution systems) and relatively low socioeconomic status (SES) also contributes to Pb exposure risk (Friedrich 2000). Despite these local sources of chemical contamination, diet remains an important route of chemical exposure for the population; for example,

Address correspondence to Verónica M. Vieira, 653 E. Peltason Dr., AIRB 2084, Irvine, CA 92697 USA. Telephone: (949) 824-7017. Email: Vvieira@uci.edu

Supplemental Material is available online (https://doi.org/10.1289/EHP4849).
The authors declare they have no actual or potential competing financial interests.
Received 8 December 2018; Revised 30 July 2019; Accepted 5 August 2019; Published 26 August 2019.

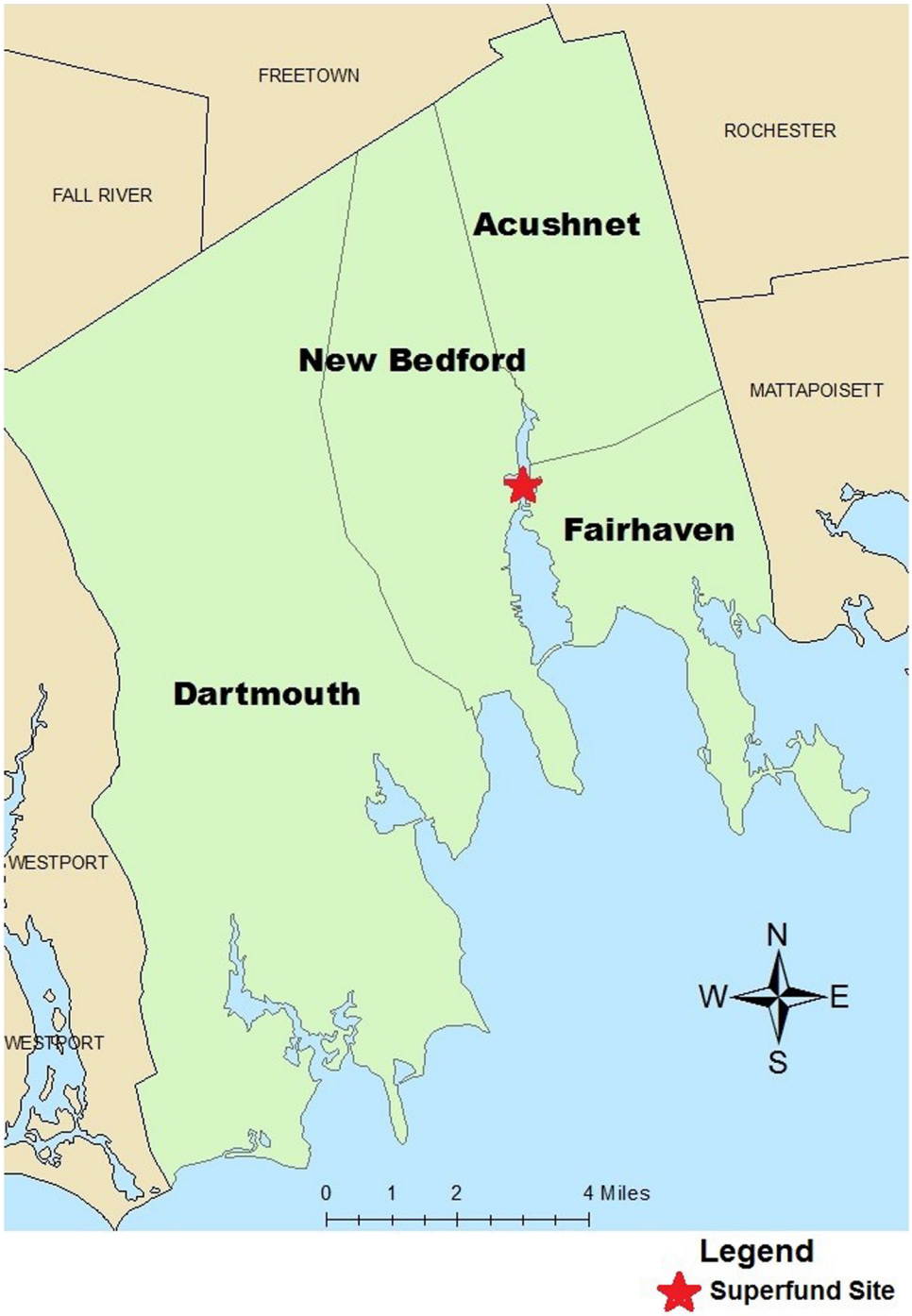

Figure 1. Four study area towns surrounding the New Bedford Harbor (NBH) Superfund site in southeast Massachusetts. The NBH was designated a Superfund site by the U.S. EPA in 1982 because of extensive polychlorinated biphenyl (PCB) contamination from local electronics manufacturing facilities.

fish intake correlates with PCB and methylmercury (MeHg) biomarker levels in the NBC (Choi et al. 2006; Sagiv et al. 2012b).

We modeled PCBs, *p*,*p*′-dichlorodiphenyl dichloroethylene (DDE), hexachlorobenzene (HCB), Pb, and mercury (Hg) because they have been associated with several childhood health outcomes of interest in this population. Prenatal exposure to organochlorines, including PCBs, has been associated with impaired neurodevelopment (Grandjean et al. 2001; Sagiv et al. 2008, 2010, 2012a), altered immune function (Dewailly et al. 2000), acute respiratory infections (Dallaire et al. 2004, 2006), and decreased birthweight (Govarts et al. 2012). Impaired neurological development has been also observed with prenatal exposure to Hg (Orenstein et al. 2014; Sagiv et al. 2012b) and Pb (Bellinger et al. 1987; Hu et al. 2006). Although the exposure routes to contaminants in this community include diet and old housing stock, proximity to the harbor has been associated with increased exposure to airborne PCBs (Martinez et al. 2017) as well as with important nonchemical stressors, including low SES and a poor home environment (Vieira et al. 2017). The significance of predicting prenatal exposures for all contemporaneous births in this community is that it allows for future epidemiologic analyses of health outcomes in a large study population for which it is impractical or infeasible to obtain exposure biomarkers.

Methods

Study Population

New Bedford Cohort. During 1993–1998, 788 New Bedford area newborns and their mothers were recruited as participants in the NBC birth cohort study. During pregnancy, the NBC mothers lived in one of the four towns surrounding the NBH Superfund site, including New Bedford, Acushnet, Fairhaven, and Dartmouth. The NBC was designed to assess relationships of prenatal and early life chemical exposures with subsequent child neurodevelopment (Orenstein et al. 2014; Sagiv et al. 2010, 2012a, 2012b). Study infants' prenatal exposures to organochlorines (PCBs, DDE, and HCB) were measured in cord serum and exposure to Pb was measured in cord whole blood; maternal peripartum hair was analyzed for total Hg as a proxy biomarker for MeHg exposure. The majority of Hg in hair is methylated and it correlates with MeHg levels in blood and integrates exposure over a longer time period than in blood, thereby decreasing measurement error in individual exposure assignments (Bartell et al. 2004; IPCS 1990). For PCBs, as previously, we focused on the sum of four prevalent PCB congeners: 118, 138, 153, and 180 (ΣPCB_4).

NBC cord blood samples were collected at birth with one aliquot centrifuged for removal of the serum fraction and a second aliquot collected for whole blood analyses. Organochlorines were measured in the cord serum at the Harvard T.H. Chan School of Public Health Organic Chemistry Laboratory (Boston, MA) using liquid–liquid extraction with extracts analyzed by gas chromatography with electron capture detection. Details of the method and quality assurance/quality control performance are reported elsewhere and include excellent reproducibility and sensitivity; for example, for most PCB congeners, detection limits were $<0.01\ ng/g$ (Korrick et al. 2000). Because of insufficient volume, lipid content could not be measured in NBC cord serum, so organochlorine concentrations are expressed on a wet weight basis as nanograms per gram serum.

Metal measurements were done at the Harvard T.H. Chan School of Public Health Trace Metals Laboratory (Boston, MA). Cord whole blood Pb levels were measured using isotope dilution inductively coupled plasma–mass spectrometry with excellent sensitivity [limit of detection (LOD) of 0.02 μg/dL] and precision (<5%). As a proxy for MeHg, total Hg was measured in maternal peripartum hair samples using atomic absorption spectroscopy with an average LOD of $50\ ng/g$ and excellent quality control performance (Sagiv et al. 2012b).

For both organochlorine and metal analyses, we used quantifiable values below the detection limit to optimize statistical power and avoid bias associated with censoring at the method detection limit (Kim et al. 1995). Questionnaire and medical record data on sociodemographic and birth outcome characteristics were available for NBC participants and are shown in Table 1. After excluding three sets of twins, the analysis included data for 782 NBC mother–infant pairs.

Massachusetts Birth Record Cohort. The Massachusetts Birth Record Cohort (MBRC) includes data from birth certificate data for all births in Massachusetts during 1993–1998, including the mothers recruited into the NBC. Birth records in Massachusetts are linked to several subsequent health outcomes data sets including maternal and child hospital discharge records, emergency department data, the birth defects registry, newborn hearing screening data and the Massachusetts Healthy Start, Early Intervention, and Women, Infants and Children (WIC) Programs (Barfield et al. 2008; Girguis et al. 2016, 2018; Khalili et al. 2018; Kotelchuck 2010; Manning et al. 2011). All Massachusetts children born in the four NBC towns were identified ($n=$ 10,270). Many of the same sociodemographic variables available in the NBC are also available for the MBRC (Table 1).

NBC Chemical Exposure Models

Exposure models were built to explain variability in measured NBC exposure biomarkers as a function of covariates available in the MBRC. Births in the NBC cohort were linked to Massachusetts birth record data. Previously developed prenatal exposure models for the NBC (Fabian et al. 2013, 2016) were used as a reference although these models did not exclusively focus on covariates available in the MBRC and did not include all exposures in the present analysis. Table 1 shows that the continuous variables available in the birth records are highly correlated with the same characteristics from the NBC study questionnaires and medical record reviews (Pearson correlation) and that there was high percentage agreement among categorical variables. Previous lactation is not directly available in the MBRC, but a proxy measure was constructed using breastfeeding initiation at the hospital and parity. Mothers with previous live births who initiated breastfeeding at the hospital were coded as having previous lactation. There was 85% agreement between this categorical proxy variable and the previous lactation variable available in the NBC data (Table 1). The other birth record variable considered in the exposure models that was not directly available in the NBC study was adequacy of prenatal care. For the purposes of comparing data collected in the NBC to data in the birth record, Table 1 includes NBC data on prenatal vitamin use as a proxy for adequate prenatal care.

In addition to the available birth record data, we calculated proximity measures for all geocoded birth addresses. Residential distance in meters to the nearest major roads (Class = 1–4) was estimated using geographic information systems (GIS) and road segments obtained from the Massachusetts Department of Transportation. Major roads include limited-access highways, multilane highways, numbered routes, and major roads. Residential distance to roadway was included in the analysis for the cord blood Pb model, as was age of home, which was obtained from the Massachusetts tax assessor database for 2016/2017. Distance to the harbor was also calculated using GIS and was included in the ΣPCB_4, DDE, HCB, and Hg models based on previous spatial associations with important sociodemographic predictors (Vieira et al. 2017). Geocoded addresses were also used to determine the median block group–level household income for the NBC and the MBRC using 2000 decennial U.S. Census data (U.S. Census Bureau 2000).

Exposure models were built for ΣPCB_4, DDE, HCB, Pb, and Hg using generalized additive models (GAMs) of complete data for each chemical exposure biomarker. The following GAM framework was used for each exposure model:

$$y(\text{exposure}) = S(x1, x2, \ldots, xN) + \boldsymbol{b}'\boldsymbol{z}$$

where y(exposure) is the predicted log exposures for a given chemical, $S(x1,\ x2,\ \ldots,\ xN)$ is the multivariable loess (locally weighted scatter plot smoothing) term(s) for the N continuous predictors, $\boldsymbol{b}'$ is the vector of parameters, and $\boldsymbol{z}$ is the vector of parametric covariates. The GAM was selected as the primary modeling approach because of its flexibility and ease of interpretation (Hastie and Tibshirani 1990); specifically, a loess smooth adapts to changes in data density and the model reduces to a linear regression without the smooth.

The exposure models were built using log-transformed chemical biomarkers due to their skewed distribution. We considered variables in Table 1 as predictors, starting with the full model and using backward elimination to select the final variables for inclusion. First, all continuous variables were modeled using univariable and multivariable smooth terms as well as linear terms, and the model structure that maximized the pseudo R^2 was

Table 1. Characteristics of mother–infant pairs born during 1993–1998 to mothers residing near the New Bedford Harbor from three sources: the New Bedford Cohort (NBC) from study data ($n = 782$), the NBC Massachusetts birth record data ($n = 779$), and the Massachusetts Birth Record Cohort (MBRC) birth record data ($n = 10,270$).

Characteristics	NBC: study data ($n = 782$)	NBC: birth record data ($n = 779$)	MBRC: birth record data ($n = 10,270$)	PCC or percentage agreement (%)[a]
Child's birthweight (g) (mean ± SD)	3,394 ± 455	3,396 ± 455	3,382 ± 659	0.99
Missing [*n* (%)]	0 (0)	3 (0.4)	36 (0.4)	
Child's year of birth [*n* (%)]				1.00
1993–1994	257 (32.9)	256 (32.9)	3,617 (35.2)	
1995–1996	300 (38.4)	298 (38.3)	3,262 (31.8)	
1997–1998	225 (28.8)	225 (28.9)	3,391 (33.0)	
Maternal age at birth (y) (mean ± SD)	26.3 ± 5.4	25.9 ± 5.4	26.0 ± 5.9	1.00
Maternal pregnancy weight gain (kg) (mean ± SD)	15.1 ± 6.4	13.6 ± 5.7	13.6 ± 5.8	1.00
Missing [*n* (%)]	153 (19.6)	10 (1.3)	136 (1.3)	
Maternal race/ethnicity [*n* (%)]				95%
Non-Hispanic white	533 (68.2)	607 (77.9)	8,050 (78.4)	
Non-Hispanic African American	33 (4.2)	47 (6.0)	474 (4.6)	
Hispanic/Latino	60 (7.7)	74 (9.5)	1,055 (10.3)	
Non-Hispanic, other race	62 (7.9)	49 (6.3)	665 (6.5)	
Missing	94 (12.0)	2 (0.3)	26 (0.2)	
Maternal ancestry [*n* (%)]				77%
Azores/Portugal	209 (26.7)	186 (23.4)	2,573 (25.1)	
Cape Verde	74 (9.5)	77 (10.0)	821 (8.0)	
Other	351 (44.9)	516 (66.2)	6,841 (66.6)	
Missing	148 (18.9)	3 (0.4)	35 (0.3)	
Maternal smoking during pregnancy [*n* (%)]				90%
Yes	204 (26.1)	196 (25.2)	2,306 (22.5)	
No	476 (60.9)	580 (74.5)	7,931 (77.2)	
Missing	102 (13.0)	3 (0.4)	33 (0.3)	
Maternal alcohol use in pregnancy [*n* (%)]				79%
Yes	137 (17.5)	16 (2.1)	179 (1.7)	
No	494 (63.2)	760 (97.6)	10,058 (98.0)	
Missing	151 (19.3)	3 (0.4)	33 (0.3)	
Previous lactation [*n* (%)][b]				85%
Yes	193 (24.7)	197 (25.3)	4,171 (40.6)	
No	441 (56.4)	578 (74.2)	6,030 (58.7)	
Missing	148 (18.9)	4 (0.5)	69 (0.7)	
Annual block group household income at birth [*n* (%)]				94%
<$20,000/y	175 (22.4)	167 (21.4)	2,235 (21.8)	
≥$20,000/y	606 (77.5)	608 (78.0)	8,029 (78.2)	
Missing	1 (0.1)	4 (0.5)	6 (0.05)	
Married at birth [*n* (%)]				93%
Yes	402 (51.4)	431 (55.3)	5,877 (57.2)	
No	325 (41.6)	348 (44.7)	4,393 (42.8)	
Missing	55 (7.0)	0 (0)	0 (0)	
Maternal education at birth [*n* (%)]				93%
<High school	131 (16.8)	180 (23.1)	2,940 (28.6)	
≥High school	594 (76.0)	595 (76.4)	7,285 (70.9)	
Missing	57 (7.3)	4 (0.5)	45 (0.4)	
Parity [*n* (%)]				95%
0	301 (38.5)	312 (40.1)	4,551 (44.3)	
1	315 (40.3)	307 (39.4)	3,388 (33.0)	
2	119 (15.2)	115 (14.8)	1,467 (14.3)	
≥3	47 (6.0)	45 (5.8)	812 (7.9)	
Missing	0 (0)	0 (0)	52 (0.5)	
Adequate prenatal care during pregnancy [*n* (%)][c]				93%
Yes	589 (75.3)	757 (97.2)	9,669 (94.1)	
No	41 (5.2)	14 (1.8)	463 (4.5)	
Missing	152 (19.4)	8 (1.0)	138 (1.3)	
Paternal education at birth [*n* (%)]				89%
<High school	188 (24.0)	257 (33.0)	3,388 (33.1)	
≥High school	526 (67.3)	443 (56.9)	5,730 (55.8)	
Missing	68 (8.7)	79 (10.1)	1,152 (11.2)	
Residential distance to harbor (m) (mean ± SD)[d]	1,572 ± 1,934	1,557 ± 1,911	1,617 ± 1,912	0.95
Residential distance to major roadway (m) (mean ± SD)[d]	191 ± 359	189 ± 366	200 ± 403	0.95
Build year for home at birth (mean ± SD)[e]	1925 ± 37.7	1928 ± 38.6	1929 ± 39.5	0.60
Missing	92 (11.8)	245 (31.5)	3,404 (33.1)	

Note: Pb, lead; PCC, Pearson Correlation Coefficient; SD, standard deviation.
[a]PCC and percentage agreement were calculated for the nonmissing data values.
[b]Previous lactation was not considered in the Hg exposure model.
[c]Adequate prenatal care during pregnancy was not directly measured in the NBC. Prenatal vitamin use was used as a proxy measure.
[d]Residential distance to the harbor was not considered in the Pb exposure model.
[e]Residential distance to the nearest major roadway and build year for home at birth were considered only in the Pb-exposure model.

Table 2. Estimates of the percentage difference in cord serum ΣPCB_4, DDE, and HCB in the NBC for parametric categorical terms in the generalized additive exposure models built using complete case data.

	ΣPCB_4 ($n = 659$) CV $R^2 = 0.54$		DDE ($n = 658$) CV $R^2 = 0.40$		HCB ($n = 590$) CV $R^2 = 0.34$	
Characteristics	Est.[a]	95% CI	Est.[a]	95% CI	Est.[a]	95% CI
Maternal race/ethnicity						
Non-Hispanic African American	−19.3	−43.4, 14.9	23.8	−18.2, 87.5	NS	NS
Hispanic/Latino	17.9	−2.5, 42.6	39.1	11.1, 74.2	NS	NS
Non-Hispanic, other race	13.1	−19.0, 58.4	114.0	44.4, 217.1	NS	NS
Maternal ancestry						
Azores/Portugal	36.0	20.6, 53.4	33.3	15.7, 53.5	13.6	1.2, 27.4
Cape Verde	19.8	−13.1, 65.2	−8.5	−37.3, 33.5	19.2	−0.5, 42.7
Maternal smoking during pregnancy	NS	NS	NS	NS	7.8	−4.7, 22.0
Maternal alcohol use in pregnancy	−1.7	−32.7, 43.5	2.1	−34.5, 59.2	−13.6	−40.2, 25.0
Previous lactation	−4.7	−16.6, 8.9	−8.2	−21.5, 7.4	−4.6	−16.2, 8.6
Block group household income at birth <$20,000/y	NS	NS	−13.0	−25.8, 2.0	2.7	−9.9, 17.0
Married at birth	NS	NS	NS	NS	NS	NS
Maternal education (<high school at birth)	11.4	−2.7, 27.5	−0.5	−15.2, 16.6	4.0	−8.7, 18.6
Parity						
1	−4.9	−16.6, 8.4	−3.8	−17.5, 12.2	1.8	−10.3, 15.5
2	−18.4	−31.7, −2.5	−13.4	−29.8, 6.6	−6.4	−21.1, 11.1
≥3	−16.8	−35.2, 6.7	−21.5	−41.5, 5.2	−5.0	−25.6, 21.3
Adequate prenatal care	NS	NS	NS	NS	NS	NS
Paternal education <high school at birth	14.2	2.0, 27.9	−2.5	−14.6, 11.4	−4.0	−14.0, 7.2

Note: NS indicates variable that were not significant ($p > 0.2$) and did not improve the R^2 and so were dropped from the final exposure model. CI, confidence interval; CV, cross validated; DDE, dichlorodiphenyl dichloroethylene; est., estimate; HCB, hexachlorobenzene; NBC, New Bedford Cohort; NS, not significant; ΣPCB_4, sum of four prevalent PCB congeners (118, 138, 153, 180).

[a]Percentage difference using the following reference groups: non-Hispanic white, ancestry other than Azores, Portugal, and Cape Verde, no maternal smoking during pregnancy, no maternal alcohol consumption during pregnancy, no previous lactation, block group–level household income >$20,000/y, unmarried at birth, maternal education >high school at birth, nulliparous, prenatal care not adequate, and paternal education >high school at birth.

selected. We calculated the pseudo R^2 as 1 – (residual deviance/null deviance) for variable selection. Based on the higher pseudo R^2, all continuous variables in the exposure models were modeled with smooth terms. We then eliminated categorical variables with p-values generated from the F-statistic that were greater than 0.2 in order of highest p-value if they did not increase the pseudo R^2 value. Variables that were tested but not retained in the final exposure models are indicated as not significant (NS) in Tables 2 and 3.

Continuous predictors in the exposure models included space and time represented by latitude and longitude of residential address, infant year of birth, and maternal age at birth. Results of a previous NBC study showed infant year of birth was strongly associated with PCB exposures (Choi et al. 2006). Furthermore, exposure to lipophilic organochlorines often increases with maternal age due to bioaccumulation (Axelrad et al. 2009; Birch et al. 2014). Birthweight, maternal weight gain (in kg), distance measures, and year the home was built were also tested as continuous predictors. Because the age of the home is often associated with the potential presence of Pb in plumbing or paint, we considered only build year in the Pb exposure model; build year was not

Table 3. Estimates of the percentage difference in cord blood Pb and maternal hair total Hg in the NBC for parametric terms in the generalized additive exposure models built using complete case data.

	Pb ($n = 442$) CV $R^2 = 0.46$		Hg ($n = 459$) CV $R^2 = 0.40$	
Characteristics	Est.[a]	95% CI	Est.[a]	95% CI
Maternal race/ethnicity				
Non-Hispanic African American	−5.0	−37.0, 43.3	−20.5	−54.5, 38.9
Hispanic/Latino	−2.2	−21.0, 21.1	−6.1	−29.0, 24.2
Non-Hispanic, other race	28.3	−13.8, 90.9	−28.2	−55.2, 15.1
Maternal ancestry				
Azores/Portugal	3.6	−9.9, 19.0	32.2	14.1, 53.3
Cape Verde	−6.3	−35.4, 36.0	19.2	−26.1, 92.3
Maternal smoking during pregnancy	31.5	13.2, 52.8	NS	NS
Maternal alcohol use in pregnancy	−26.5	−53.4, 16.2	23.7	−23.4, 99.8
Previous lactation	−18.8	−29.9, −5.9	NT	NT
Block group household income at birth (<$20,000/y)	NS	NS	8.4	−8.8, 28.9
Married at Birth	5.2	−9.9, 19.0	NS	NS
Maternal education (<high school at birth)	20.6	4.1, 39.8	NS	NS
Parity				
1	0.5	−13.4, 16.5	−19.3	−30.6, −6.3
2	−4.3	−21.1, 16.0	−16.6	−32.5, 3.2
≥3	−2.6	26.5, 29.0	−29.4	−47.4, −5.1
Adequate prenatal care	−1.0	−41.9, 68.9	NS	NS
Father's education (<high school at birth)	3.7	−8.5, 17.4	7.9	−6.3, 24.2

Note: NS indicates variable that were not significant ($p > 0.2$) and did not improve the R^2 and so were dropped from the final exposure model. NT indicates the variable was not tested for inclusion in the exposure model. CI, confidence interval; CV, cross validated; est., estimate; Hg, mercury; NBC, New Bedford Cohort; NS, not significant; NT, not tested; Pb, lead.

[a]Percentage difference using following reference groups: non-Hispanic white, other ancestry, no maternal smoking during pregnancy, no maternal alcohol consumption during pregnancy, no previous lactation, household income >$20,000/y, unmarried at birth, maternal education >high school at birth, nulliparous, prenatal care not adequate, and paternal education >high school at birth.

tested (NT; Table 3) as a predictor in the Hg or organochlorine models. Categorical variables considered in all exposure models included maternal race/ethnicity and ancestry, parental education, mother's marital status, block group income, parity, adequate prenatal care, maternal smoking and alcohol consumption during pregnancy, and previous lactation. Breastfeeding was tested in all models except the Hg model because it is not believed to be relevant for excretion of Hg (Table 3). The distributions for all variables are presented in Table 1.

We calculated a predictive 10-fold cross-validation R^2 value by randomly dividing the NBC data into 10 groups. For each fold, 1 group was used as the test data and the remaining groups combined as the training data set. The models were fit on the training set and evaluated against the test set. Each group served as a test set once. The correlation between the observed exposures and predicted exposures for the 10 test sets were then averaged to calculate the 10-fold cross-validation R^2 value. The optimal degree of smoothing in the GAMs was chosen by maximizing the 10-fold cross-validation R^2 for each model.

Exposure Predictions for the MBRC

Each NBC exposure model was used to predict the exposures for all births with complete data in the MBRC from 1993 through 1998 for the study towns of New Bedford, Acushnet, Fairhaven, and Dartmouth; we calculated the mean, median, and the 25th and 75th percentiles of those estimates. To account for the effects of sampling variability on our NBC exposure prediction model, we also applied the bootstrap method to each NBC exposure model to generate a distribution of plausible exposure biomarker values. We drew repeated samples 1,000 times, with replacement, from the NBC. The exposure model was refit each time, arriving at a different set of coefficients for the model terms, and applied to data from the MBRC to obtain 1,000 predicted log-transformed exposure values. After back-transforming those values, we calculated the median, the 25th and 75th percentiles, and the corresponding interquartile range (IQR) for the population and individual bootstrapped exposure estimates. For the individual estimates, the median and the 25th and 75th percentiles were calculated across the 1,000 bootstrapped values predicted for each birth and then averaged. For the population estimates, the median and the 25th and 75th percentiles were calculated across the entire cohort population of 10,270 births for each bootstrap, resulting in a distribution of 1,000 values for each statistic (the median and the 25th and 75th percentiles). We then took the mean and 95% probability intervals (PIs) of each statistic's distribution to assess the population exposures. Because the predictive characteristics of the population varied, we expected that the IQR of population exposures would be wider than the IQR for the exposures predicted for each individual birth using the bootstrap method. All calculations were performed using R (version 3.4.4; R Development Core Team), and the "gam" package was used for the models.

Exposure Model Performance

We also compared the performance of the GAMs to individual model and ensemble results from the SuperLearner package in R. SuperLearner evaluates the performance of multiple machine learning models, creates an optimal weighted average of the best performing models referred to as an ensemble, and provides comparisons of the ensemble to the other models using cross validation (Kennedy 2017). Using the SuperLearner function, we first fit the following models simultaneously: GAM, lasso, randomForest, XGBoost, SVM, and the mean of Y. The output provides a risk estimate for each as a measure of model performance, with lower risk estimates indicating better performance. The function also creates an ensemble with a weighted average of the individual models. We then calculated a predictive 10-fold cross-validation R^2, using the ensemble to compare with the GAM models we developed for each exposure. To evaluate the performance of the individual models in the SuperLearner to the ensemble, we used the CV.SuperLearner function, which calculates the standard errors of each via a nested 10-fold cross validation. The risk estimates and standard errors were then plotted for comparison.

Results

NBC and MBRC Characteristics

We successfully linked 779 of the 782 NBC births to their Massachusetts birth record data. When the birth record characteristics for the NBC subset were compared with the entire MBRC for the study area, we saw very similar proportions. In general, sociodemographic and behavioral characteristics collected in the NBC were similar to those available for the MBRC from the same time period (Table 1). In both data sets, mothers were on average 26 y of age, and the majority of mothers were non-Hispanic white and married at the time of the child's birth. The MBRC had a higher proportion of maternal ancestry from other than the Azores/Portugal or Cape Verde compared with the NBC variable (66% vs. 45%). Agreement between the two data sources for maternal ancestry was generally good (84%), and the higher proportion of missing data in the NBC may account for differences in the proportion of other ancestry.

Reported smoking during pregnancy was high in both the NBC (26%) and the MBRC (23%), and there was 90% agreement between the two data sources, but there were significant differences in reported alcohol consumption during pregnancy (18% in the NBC vs. 2% in the MBRC; 79% agreement). In addition, differences in maternal education levels were observed, with 16.8% of the NBC reporting less than a high school education versus 28.6% of the MBRC. This difference may be due to the amount of missing data for this variable in the NBC (7.3% vs. 0.5% in the linked Massachusetts birth data) given that the percentage of mothers with at least a high school education was similar between the NBC data and the linked Massachusetts birth data (76.0% and 76.4% respectively). With paternal education, 24.0% of the NBC compared with 33.1% of the MBRC reported less than a high school education. However, when using the linked Massachusetts birth data for paternal education among the NBC subset, the proportion with a high school education was similar to the larger MBRC.

NBC Exposure Models

The 10-fold cross-validated R^2 values for the ΣPCB_4, DDE, HCB, Pb, and Hg exposure models were 0.54, 0.40, 0.34, 0.46, and 0.40, respectively (Tables 2 and 3). Tables 2 and 3 present the estimates of the percentage difference in biomarker levels of PCB, DDE, HCB, Pb, and Hg associated with categorical variables modeled parametrically. For all the exposure models except Pb, maternal ancestry was a significant predictor, with higher biomarker levels seen for mothers born in the Azores/Portugal. Compared with non-Hispanic whites, Hispanics and non-Hispanic women in other nonwhite racial groups had higher DDE levels. Conversely, non-Hispanic whites had higher Hg levels compared with other groups. For mothers who smoked during pregnancy, there was a significant increase in cord blood Pb compared with mothers who did not smoke, consistent with smoking's association with blood Pb in other populations (Deutch and Hansen 1999; Rodosthenous et al.

2017; Saoudi et al. 2018). Paternal education was a significant predictor of PCB levels, with a 14% increase [95% confidence interval (CI): 2%, 28%] for infants whose fathers had less than a high school level education. Cord blood Pb levels were 21% higher (95% CI: 4%, 40%) among infants whose mothers had not completed high school compared with those whose mothers had at least a high school education.

To assess the effect of continuous covariates modeled using smooths, exposures were predicted by varying one covariate at a time while holding the other variables constant at the median value (Table 4). Older mothers tended to have higher exposure estimates, whereas nonlinear effects in maternal age were observed for HCB and Hg, with women at the 75th percentile value of maternal age (30 y) having higher exposures than the oldest mothers (41 y of age). Although a change from the 25th to 75th percentile value for maternal age nearly doubled the prediction for DDE (0.232 to 0.422 ng/g serum), an interquartile change in year of infant birth had little effect on the predictions (0.302 to 0.304 ng/g serum). Nonlinear effects were also observed with birth weight. Cord blood Pb level predictions were higher for older homes compared with newer homes, and higher at the 25th than 75th percentile of distance to nearest road (471 m and 1.4 μg/dL compared with 1,620 m and 1.2 μg/dL). Varying location, represented by longitude and latitude, also resulted in different exposure predictions, with the highest exposures generally at lower latitudes (closer to the ocean, Figure 1).

Predicted Exposures for the MBRC

Biomarker levels predicted for the MBRC using the entire NBC (prior to bootstrapping) showed similar distributions as the NBC biomarkers (Figure 2). Table 5 presents the mean and 95% PIs of the median and the 25th and 75th percentiles of the predicted biomarker concentrations and corresponding IQR across the entire

Table 4. Exposure predictions in the NBC calculated for a range of values of nonparametric continuous covariates in the generalized additive exposure models built using complete case data.

Variable		Value	ΣPCB_4 (ng/g)	DDE (ng/g)	HCB (ng/g)	Hg (μg/g)	Pb (μg/dL)
Longitude	Min	−71.043	0.120	0.243	0.024	0.465	1.48
	25th	−70.939	0.158	0.297	0.017	0.537	1.34
	Median	−70.931	0.162	0.303	0.017	0.544	1.32
	75th	−70.922	0.167	0.309	0.016	0.552	1.31
	Max	−70.833	0.206	0.360	0.012	0.616	1.21
Latitude	Min	41.538	0.193	0.410	0.012	0.793	1.47
	25th	41.624	0.167	0.320	0.016	0.583	1.35
	Median	41.646	0.162	0.303	0.017	0.544	1.32
	75th	41.666	0.160	0.294	0.017	0.526	1.31
	Max	41.750	0.138	0.230	0.021	0.386	1.20
Infant year of birth	Min	1993	0.196	0.302	0.024	0.625	1.75
	25th	1994	0.178	0.302	0.023	0.583	1.52
	Median	1995	0.162	0.303	0.017	0.544	1.32
	75th	1997	0.134	0.304	0.016	0.474	1.01
	Max	1998	0.122	0.304	0.022	0.442	0.88
Maternal age (y)	Min	18	0.083	0.178	0.012	0.562	1.11
	25th	21	0.113	0.232	0.015	0.314	1.21
	Median	25	0.162	0.303	0.017	0.544	1.32
	75th	30	0.256	0.422	0.023	0.610	1.48
	Max	41	0.540	0.874	0.016	0.171	1.88
Pregnancy weight gain (kg)	Min	0	0.162	0.415	0.019	0.487	1.83
	25th	9.5	0.156	0.288	0.017	0.531	1.45
	Median	13.6	0.162	0.303	0.017	0.544	1.32
	75th	16.8	0.156	0.303	0.016	0.548	1.31
	Max	41.7	0.090	0.154	0.010	0.429	1.40
Birth weight (g)	Min	1,871	0.167	0.247	0.011	0.558	1.38
	25th	3,090	0.167	0.295	0.017	0.565	1.30
	Median	3,402	0.162	0.303	0.017	0.544	1.32
	75th	3,685	0.155	0.296	0.017	0.528	1.35
	Max	5,216	0.134	0.322	0.015	0.431	1.41
Distance to harbor (m)[a]	Min	0	0.155	0.259	0.017	0.426	—
	25th	471	0.159	0.291	0.017	0.528	—
	Median	929	0.162	0.303	0.017	0.544	—
	75th	1,620	0.167	0.281	0.016	0.576	—
	Max	8,526	0.170	0.351	0.012	0.601	—
Distance to road (m)[a]	Min	0	—	—	—	—	1.11
	25th	43	—	—	—	—	1.41
	Median	103	—	—	—	—	1.32
	75th	193	—	—	—	—	1.23
	Max	4,057	—	—	—	—	1.18
Build year for home at birth[a]	Min	1728	—	—	—	—	1.49
	25th	1903	—	—	—	—	1.41
	Median	1920	—	—	—	—	1.32
	75th	1960	—	—	—	—	1.19
	Max	1998	—	—	—	—	1.07

Note: Predictions were made varying only the continuous variable of interest, holding the other continuous variables constant at the median. Categorical variables were held constant at the referent groups: non-Hispanic white, other ancestry, no maternal smoking during pregnancy, no maternal alcohol consumption during pregnancy, no previous lactation, household income >$20,000/y, unmarried at birth, maternal education >high school at birth, nulliparous, prenatal care not adequate, and paternal education >high school at birth. —, not applicable; DDE, dichlorodiphenyl dichloroethylene; HCB, hexachlorobenzene; Hg, mercury; max, maximum; min, minimum; NBC, New Bedford Cohort; ΣPCB_4, sum of four prevalent PCB congeners (118, 138, 153, 180); Pb, lead.

[a]Distance to harbor was not included in the Pb model. Distance to nearest road and build year for home at birth were included only in the Pb model.

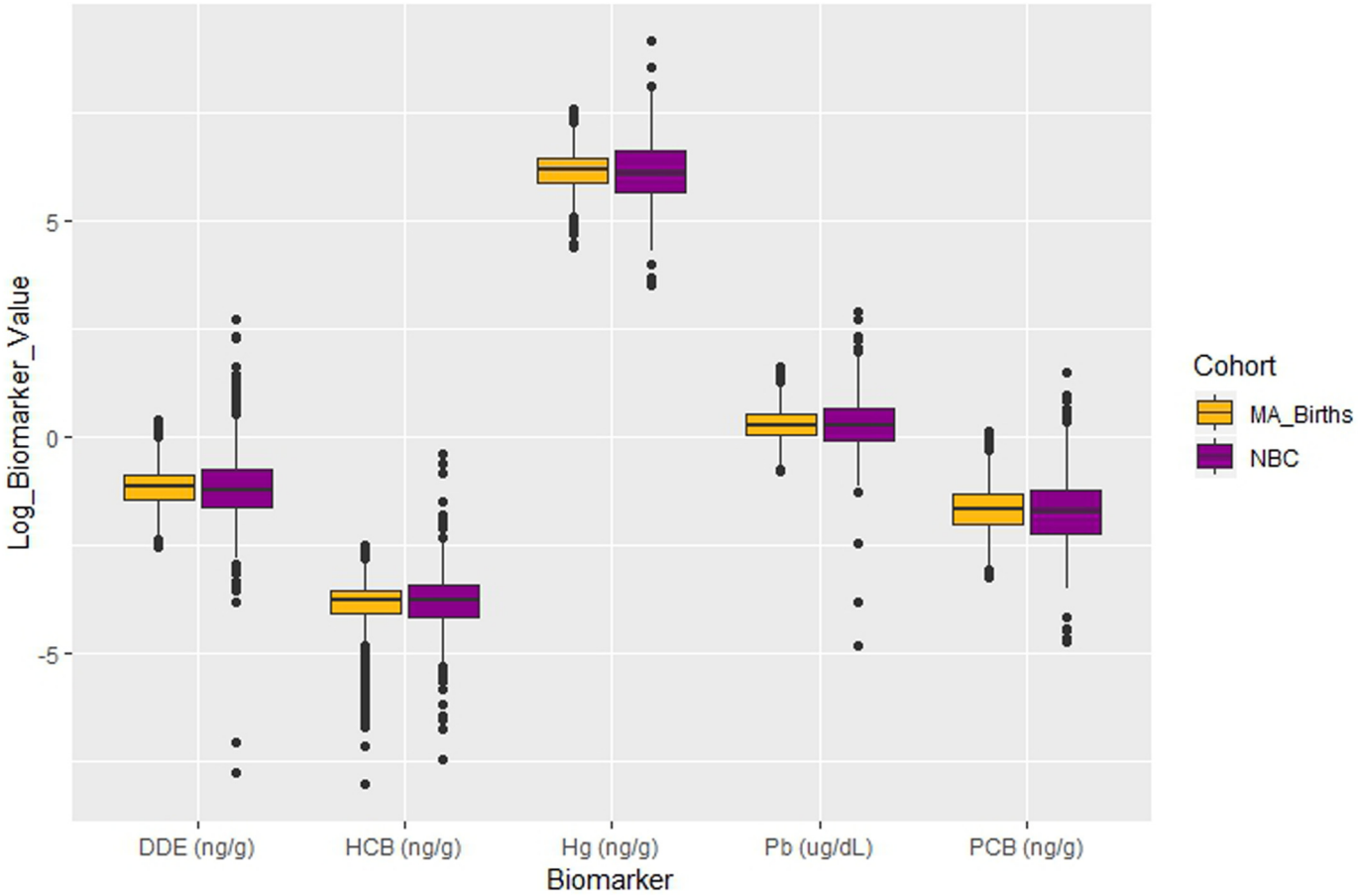

Figure 2. Box plot of exposure biomarker predictions for the Massachusetts Birth Records Cohort (MA_Births; in gold) compared with measured exposure biomarker concentrations for the New Bedford Cohort (NBC; in red-violet). The lower and upper boundaries of the boxes represent the exposure values for the 25th to 75th percentiles of the exposure distribution. The line in each box represents the median of the distribution, and the whiskers represent 1.5 times the interquartile range. The points indicate outlier values beyond this range.

population of the Massachusetts birth cohort using the bootstrap method. Based on these predicted biomarkers, it was possible to characterize highly exposed subpopulations. For example, women with ancestry from the Azores/Portugal had predicted median PCB exposure of 0.22 ng/g, whereas women with other ancestry had predicted median PCB exposure of 0.16 ng/g. In addition, women above the 90th percentile in PCBs were disproportionately older mothers with ancestry from the Azores/Portugal. In addition, for each birth, we found the median, the 25th and 75th percentiles, and IQR difference across the 1,000 bootstrapped exposure estimates predicted for that birth and then averaged over all the births (Table 6). The average individual IQRs (Table 6) are much smaller than the population IQRs (Table 5) as expected given that the characteristics for predicting exposures vary across the population but not at the individual level.

Exposure Model Performance Comparison

The 10-fold cross-validated R^2 values for the ΣPCB_4, DDE, HCB, Pb, and Hg exposure models using the GAMs we developed (0.54, 0.40, 0.34, 0.46, and 0.40, respectively) were similar to the 10-fold cross-validated R^2 values using the ensemble output from the SuperLearner package in R (cross-validated R^2 values of 0.54, 0.45, 0.33, 0.44, 0.32, respectively). The results of the nested SuperLearner cross validation showed that the GAM model performance was statistically similar to other modern modeling approaches, including lasso, randomForest, and the ensemble (see Figures S1–S5); the risk estimates and standard errors overlap. In building the GAM exposure models, modeling continuous variables using nonlinear smooths increased the pseudo R^2 compared with linear models (see Tables S1–S5).

Table 5. Mean and 95% probability interval (PI) of the median and the 25th and 75th percentiles of the biomarker exposure concentrations and corresponding interquartile range (IQR) across the entire population of the MBRC ($n = 10,270$) predicted using the NBC exposure models with bootstrapping.

Biomarkers	Median (95% PI)	25th percentile (95% PI)	75th percentile (95% PI)	IQR
ΣPCB_4 (ng/g)	0.18 (0.17–0.20)	0.13 (0.12–0.14)	0.26 (0.25–0.28)	0.13
DDE (ng/g)	0.31 (0.29–0.33)	0.23 (0.21–0.25)	0.42 (0.39–0.46)	0.19
HCB (ng/g)	0.022 (0.021–0.024)	0.017 (0.016–0.018)	0.029 (0.028–0.031)	0.012
Hg (μg/g)	0.47 (0.44–0.50)	0.35 (0.32–0.38)	0.63 (0.58–0.69)	0.28
Pb (μg/dL)	1.31 (1.25–1.42)	1.01 (1.00–1.07)	1.73 (1.62–1.86)	0.72

Note: DDE, dichlorodiphenyl dichloroethylene; HCB, hexachlorobenzene; Hg, mercury; MBRC, Massachusetts Birth Record Cohort; NBC, New Bedford Cohort; Pb, lead; ΣPCB_4, sum of four prevalent PCB congeners (118, 138, 153, 180).

Table 6. Mean of the median, the 25th and 75th percentiles of the predicted biomarker concentrations, and interquartile range (IQR) across the 1,000 bootstrapped results for each individual birth in the MBRC ($n = 10{,}270$).

Biomarker	Median	25th percentile	75th percentile	IQR
ΣPCB_4 (ng/g)	0.19	0.17	0.21	0.04
DDE (ng/g)	0.32	0.28	0.35	0.07
HCB (ng/g)	0.022	0.019	0.026	0.007
Hg (μg/g)	0.47	0.41	0.54	0.13
Pb (μg/dL)	1.34	1.21	1.49	0.28

Note: DDE, dichlorodiphenyl dichloroethylene; HCB, hexachlorobenzene; Hg, mercury; MBRC, Massachusetts Birth Record Cohort; Pb, lead; ΣPCB_4, sum of four prevalent PCB congeners (118, 138, 153, 180).

Discussion

The exposure models built in this study explained variability across several different chemicals using predictors available from the MBRC linked to biomarkers measured in the NBC, allowing for both exposure predictions for epidemiological studies and identification of susceptible subgroups. Our previous research with the NBC indicated that the range of exposures modeled in the MBRC, as well as the differential exposures associated with key sociodemographic covariates, are sufficient to be associated with adverse child health outcomes in epidemiologic analyses (Sagiv et al. 2010, 2012a, 2012b). In considering susceptible subgroups, a significant predictor across all exposure models, with the exception of Pb, was maternal ancestry from the Azores/Portugal. This could be due to distinctive dietary patterns, such as fish consumption (FAO 2011), or to differential early life exposures to bioaccumulative chemicals among mothers either born in these countries or exposed transgenerationally via the previous generation's exposures. The fact that maternal ancestry was a more consistent predictor than race/ethnicity emphasizes the potential importance of more granular demographic predictors when available. Location should also be considered in exposure models when available. We observed nonlinear associations with longitude and latitude, with births closer to the coastline, where industrial activities are located in this community, having higher exposures. Individuals living closer to the coastline may also be more likely to have higher seafood consumption, which is a known route of exposure for some of these contaminants. Inverse associations were generally observed in parous compared with nulliparous mothers, consistent with chemical elimination from breastfeeding or pregnancy (ATSDR 2004; Nickerson 2006). Although some of these associations do not reflect modifiable risk factors, they do allow for differential exposure assignment for epidemiology, and potential exploration of underlying exposure pathways for which intervention may be possible.

The strength of paternal (as well as maternal) educational attainment in predicting many biomarker levels suggests that this is also an indicator of household SES within this population, especially in the absence of individual-level or household-level measures of income. Inclusion of residential distance to the nearest major road and build year for home at birth as predictors of Pb reinforces previous findings that residential exposure pathways persist, especially in lower-income urban settings with an older housing stock. One limitation of the analysis was that paternal education and home build year were missing for a large proportion of the birth records. Although our exposure models used complete data only, future epidemiologic analyses that use the exposure predictions may consider multiple imputation methods to address the issue of missing data.

Another limitation of the analysis is that variables in the NBC such as diet that could have increased the variance explained by the exposure models were not available in the MBRC. Diet, including fish and organ meat consumption, are important predictors for many of the exposures we targeted for study (Choi et al. 2006). Although maternal ancestry may have served as an indicator of dietary habits, having additional food intake information may have improved our models and increased their ability to ascertain routes of exposure. That said, administrative data sets rarely have dietary and other behavioral information, so our ability to explain significant variance in the absence of these data is notable.

Last, our exposure regression models may not be generalizable to other geographic areas. Although residential distance to the harbor (a term specific to the New Bedford area) was not associated with umbilical cord serum PCB levels in a previous analysis (Choi et al. 2006), we included it in the exposure models (except for Pb) to capture any residual sociodemographic effects based on results from a spatial study that showed associations with proximity to the harbor (Vieira et al. 2017). Other sociodemographic variables might have different meanings in different settings. That said, the approach is generalizable to any setting where biomarker measurements and sociodemographic information are available for a contemporaneous subset of study area participants.

We used GAMs to build our exposure models but acknowledge that other models with comparable flexible modeling features may have performed similarly. We were able to increase the pseudo R^2 value of each model, compared with linear models, by smoothing continuous variables. Using the SuperLearner package in R, we showed that the GAM performed as well as other modeling techniques, including lasso and randomForest, and that the 10-fold cross-validated R^2 values for the GAM exposure models we developed were similar to the ensemble results from the SuperLearner. Moreover, GAMs have been applied previously to NBC data (Vieira et al. 2017) and can yield valuable insight regarding functional forms that facilitate model interpretability. Our cross-validated R^2 values showed comparable performance to exposure regression models used in other successful epidemiologic studies, including simulation studies that illustrated the robustness of exposure–outcome associations (Avanasi et al. 2016; Baxter et al. 2010; Shin et al. 2011).

Although there are several approaches for building exposure models, we used the model that resulted in the highest R^2 value to decrease the bias in our estimates. Using this method would suggest putting all of the variables into the model, given that increasing variables increases the R^2, but we allowed our sample size to change due to missingness from each variable to determine which variables in the model would give the largest pseudo R^2 value. Although using a liberal variable selection method decreased the bias in our estimates and is generally recommended for model selection when a subset of the data is available (Rubin 1996), it contributed to uncertainty in our exposure predictions. The bootstrap method accounted for sampling variability without assuming a parametric distribution, as reflected by the 95% PIs shown in Table 5. Those intervals are reasonably narrow, indicating that the central tendency and IQR of predicted exposure are fairly stable. Our approach does not account for other sources of uncertainty. For example, our model was not able to predict all the variation in observed biomarker concentrations; unexplained residual variation contributes to the overall uncertainty in biomarker predictions but is not assessed by ordinary bootstrap estimation methods. Future work might include generating parametric or nonparametric bootstrap prediction distributions for more comprehensive evaluation of the effects of individual exposure uncertainty in epidemiological analyses.

In summary, our study showed how multiple prenatal chemical exposures can be estimated in a general population administrative data set by modeling available measured biomarker data

for a subset of the population using GAMs and key sociodemographic and behavioral information. Our models can be used to predict exposure biomarker concentrations in future epidemiological investigations of health outcomes. In addition, they yielded valuable insights about residents in a community near a Superfund site who may have been at increased risk of exposure to a range of chemicals.

Acknowledgments

This work was supported by grants P42 ES007381, P42 ES005947, and R01 ES014864 from the National Institute of Environmental Health Sciences, National Institutes of Health.

References

ATSDR (Agency for Toxic Substances and Disease Registry). 2004. *Interaction Profile for Persistent Chemicals Found in Breast Milk*. Atlanta, GA: ATSDR, U.S. Department of Health and Human Services.

Avanasi R, Shin HM, Vieira VM, Savitz DA, Bartell SM. 2016. Impact of exposure uncertainty on the association between perfluorooctanoate and preeclampsia in the C8 Health Project population. Environ Health Perspect 124(1):126–132, PMID: 26090912, https://doi.org/10.1289/ehp.1409044.

Axelrad DA, Goodman S, Woodruff TJ. 2009. PCB body burdens in US women of childbearing age 2001–2002: an evaluation of alternate summary metrics of NHANES data. Environ Res 109(4):368–378, PMID: 19251256, https://doi.org/10.1016/j.envres.2009.01.003.

Barfield WD, Clements KM, Lee KG, Kotelchuck M, Wilber N, Wise PH. 2008. Using linked data to assess patterns of early intervention (EI) referral among very low birth weight infants. Matern Child Health J 12(1):24–33, PMID: 17562149, https://doi.org/10.1007/s10995-007-0227-y.

Bartell SM, Griffith WC, Faustman EM. 2004. Temporal error in biomarker-based mean exposure estimates for individuals. J Expo Anal Environ Epidemiol 14(2):173–179, PMID: 15014548, https://doi.org/10.1038/sj.jea.7500311.

Baxter LK, Wright RJ, Paciorek CJ, Laden F, Suh HH, Levy JI. 2010. Effects of exposure measurement error in the analysis of health effects from traffic-related air pollution. J Expo Sci Environ Epidemiol 20(1):101–111, PMID: 19223939, https://doi.org/10.1038/jes.2009.5.

Bellinger D, Leviton A, Waternaux C, Needleman H, Rabinowitz M. 1987. Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. N Engl J Med 316(17):1037–1043, PMID: 3561456, https://doi.org/10.1056/NEJM198704233161701.

Birch RJ, Bigler J, Rogers JW, Zhuang Y, Clickner RP. 2014. Trends in blood mercury concentrations and fish consumption among U.S. women of reproductive age, NHANES, 1999–2010. Environ Res 133:431–438, PMID: 24602558, https://doi.org/10.1016/j.envres.2014.02.001.

Choi AL, Levy JI, Dockery DW, Ryan LM, Tolbert PE, Altshul LM, et al. 2006. Does living near a Superfund site contribute to higher polychlorinated biphenyl (PCB) exposure? Environ Health Perspect 114(7):1092–1098, PMID: 16835064, https://doi.org/10.1289/ehp.8827.

Dallaire F, Dewailly É, Muckle G, Vézina C, Jacobson SW, Jacobson JL, et al. 2004. Acute infections and environmental exposure to organochlorines in Inuit infants from Nunavik. Environ Health Perspect 112(14):1359–1364, PMID: 15471725, https://doi.org/10.1289/ehp.7255.

Dallaire F, Dewailly É, Vézina C, Muckle G, Weber J-P, Bruneau S, et al. 2006. Effect of prenatal exposure to polychlorinated biphenyls on incidence of acute respiratory infections in preschool Inuit children. Environ Health Perspect 114(8):1301–1305, PMID: 16882544, https://doi.org/10.1289/ehp.8683.

Deutch B, Hansen JC. 1999. High blood levels of persistent organic pollutants are statistically correlated with smoking. Int J Circumpolar Health 58(3):214–219, PMID: 10528472.

Dewailly E, Ayotte P, Bruneau S, Gingras S, Belles-Isles M, Roy R. 2000. Susceptibility to infections and immune status in Inuit infants exposed to organochlorines. Environ Health Perspect 108(3):205–211, PMID: 10706525, https://doi.org/10.1289/ehp.00108205.

Fabian MP, Korrick SA, Peters J, Levy J. 2013. Modeling population exposure to multiple chemical and non-chemical stressors in a low income community near a hazardous waste site. Joint Conference of the ISEE/ISES/International Society of Indoor Air Quality (ISIAQ), Basel, Switzerland, https://ehp.niehs.nih.gov/doi/10.1289/isee.2013.O-2-26-02.

Fabian MP, Levy JI, Vieira V, Peters JL, Korrick S. 2016. *Behavioral and Sociodemographic Predictors of Exposure to Multiple Chemicals Associated with ADHD-Related Behavior in a Low Income Community*. Durham, NC: National Institute of Environmental Health Sciences Science Fest.

FAO, *Profil alimentaire et nutritionnel du Cap-Vert*. 2011. Food and Agriculture Organization of the United Nations: Rome, Italy.

Friedrich MJ. 2000. Poor children subject to "environmental injustice." JAMA 283(23):3057–3058, PMID: 10865284, https://doi.org/10.1001/jama.283.23.3057-JMN0621-3-1.

Girguis MS, Strickland MJ, Hu X, Liu Y, Bartell SM, Vieira VM. 2016. Maternal exposure to traffic-related air pollution and birth defects in Massachusetts. Environ Res 146:1–9, PMID: 26705853, https://doi.org/10.1016/j.envres.2015.12.010.

Girguis MS, Strickland MJ, Hu X, Liu Y, Chang HH, Kloog I, et al. 2018. Exposure to acute air pollution and risk of bronchiolitis and otitis media for preterm and term infants. J Expo Sci Environ Epidemiol 28(4):348–357, PMID: 29269754, https://doi.org/10.1038/s41370-017-0006-9.

Govarts E, Nieuwenhuijsen M, Schoeters G, Ballester F, Bloemen K, de Boer M, et al. 2012. Birth weight and prenatal exposure to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE): a meta-analysis within 12 European birth cohorts. Environ Health Perspect 120(2):162–170, PMID: 21997443, https://doi.org/10.1289/ehp.1103767.

Grandjean P, Weihe P, Burse VW, Needham LL, Storr-Hansen E, Heinzow B, et al. 2001. Neurobehavioral deficits associated with PCB in 7-year-old children prenatally exposed to seafood neurotoxicants. Neurotoxicol Teratol 23(4):305–317, PMID: 11485834, https://doi.org/10.1016/S0892-0362(01)00155-6.

Hastie T, Tibshirani R. 1990. *Generalized Additive Models*. London, UK: Chapman and Hall.

Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, et al. 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 42(33):7561–7578, https://doi.org/10.1016/j.atmosenv.2008.05.057.

Hu H, Téllez-Rojo MM, Bellinger D, Smith D, Ettinger AS, Lamadrid-Figueroa H, et al. 2006. Fetal lead exposure at each stage of pregnancy as a predictor of infant mental development. Environ Health Perspect 114(11):1730–1735, PMID: 17107860, https://doi.org/10.1289/ehp.9067.

IPCS (International Programme on Chemical Safety). 1990. IPCS Environmental Health Criteria 101, Methylmercury. http://www.inchem.org/documents/ehc/ehc/ehc101.htm [accessed 17 August 2019].

Kennedy C. 2017. Guide to SuperLearner. https://cran.r-project.org/web/packages/SuperLearner/vignettes/Guide-to-SuperLearner.html [accessed 7 June 2019].

Khalili R, Bartell SM, Hu X, Liu Y, Chang HH, Belanoff C, et al. 2018. Early-life exposure to PM2.5 and risk of acute asthma clinical encounters among children in Massachusetts: a case-crossover analysis. Environ Health 17(1):20, PMID: 29466982, https://doi.org/10.1186/s12940-018-0361-6.

Kim R, Aro A, Rotnitzky A, Amarasiriwardena C, Hu H. 1995. K x-ray fluorescence measurements of bone lead concentration: the analysis of low-level data. Phys Med Biol 40(9):1475–1485, PMID: 8532760, https://doi.org/10.1088/0031-9155/40/9/007.

Korrick SA, Altshul LM, Tolbert PE, Burse VW, Needham LL, Monson RR. 2000. Measurement of PCBs, DDE, and hexachlorobenzene in cord blood from infants born in towns adjacent to a PCB-contaminated waste site. J Expo Anal Environ Epidemiol 10(6 Pt 2):743–754, PMID: 11138666, https://doi.org/10.1038/sj.jea.7500120.

Kotelchuck M. 2010. Evaluating the Healthy Start program: a life course perspective. Matern Child Health J 14(5):649–653, PMID: 20582457, https://doi.org/10.1007/s10995-010-0629-0.

Manning SE, Davin CA, Barfield WD, Kotelchuck M, Clements K, Diop H, et al. 2011. Early diagnoses of autism spectrum disorders in Massachusetts birth cohorts, 2001–2005. Pediatrics 127(6):1043–1051, PMID: 21576313, https://doi.org/10.1542/peds.2010-2943.

Martinez A, Hadnott BN, Awad AM, Herkert NJ, Tomsho K, Basra K, et al. 2017. Release of airborne polychlorinated biphenyls from New Bedford Harbor results in elevated concentrations in the surrounding air. Environ Sci Technol Lett 4(4):127–131, PMID: 28413805, https://doi.org/10.1021/acs.estlett.7b00047.

Nickerson K. 2006. Environmental contaminants in breast milk. J Midwifery Womens Health 51(1):26–34, PMID: 16399607, https://doi.org/10.1016/j.jmwh.2005.09.006.

Orenstein ST, Thurston SW, Bellinger DC, Schwartz JD, Amarasiriwardena CJ, Altshul LM, et al. 2014. Prenatal organochlorine and methylmercury exposure and memory and learning in school-age children in communities near the New Bedford Harbor Superfund site, Massachusetts. Environ Health Perspect 122(11):1253–1259, PMID: 25062363, https://doi.org/10.1289/ehp.1307804.

Rodosthenous RS, Burris HH, Svensson K, Amarasiriwardena CJ, Cantoral A, Schnaas L, et al. 2017. Prenatal lead exposure and fetal growth: smaller infants have heightened susceptibility. Environ Int 99:228–233, PMID: 27923585, https://doi.org/10.1016/j.envint.2016.11.023.

Rubin DB. 1996. Multiple imputation after 18+ years. J Am Stat Assoc 91(434):473–489, https://doi.org/10.2307/2291635.

Childhood (Elfe). Int J Hyg Environ Health 221(3):441–450, PMID: 29352707, https://doi.org/10.1016/j.ijheh.2018.01.007.

Sagiv SK, Nugent JK, Brazelton TB, Choi AL, Tolbert PE, Altshul LM, et al. 2008. Prenatal organochlorine exposure and measures of behavior in infancy using the Neonatal Behavioral Assessment Scale (NBAS). Environ Health Perspect 116(5):666–673, PMID: 18470320, https://doi.org/10.1289/ehp.10553.

Sagiv SK, Thurston SW, Bellinger DC, Altshul LM, Korrick SA. 2012a. Neuropsychological measures of attention and impulse control among 8-year-old children exposed prenatally to organochlorines. Environ Health Perspect 120(6):904–909, PMID: 22357172, https://doi.org/10.1289/ehp.1104372.

Sagiv SK, Thurston SW, Bellinger DC, Amarasiriwardena C, Korrick SA. 2012b. Prenatal exposure to mercury and fish consumption during pregnancy and attention-deficit/hyperactivity disorder-related behavior in children. Arch Pediatr Adolesc Med 166(12):1123–1131, PMID: 23044994, https://doi.org/10.1001/archpediatrics.2012.1286.

Sagiv SK, Thurston SW, Bellinger DC, Tolbert PE, Altshul LM, Korrick SA. 2010. Prenatal organochlorine exposure and behaviors associated with attention deficit hyperactivity disorder in school-aged children. Am J Epidemiol 171(5):593–601, PMID: 20106937, https://doi.org/10.1093/aje/kwp427.

Saoudi A, Dereumeaux C, Goria S, Berat B, Brunel S, Pecheux M, et al. 2018. Prenatal exposure to lead in France: cord-blood levels and associated factors: results from the perinatal component of the French Longitudinal Study since

Shin HM, Vieira VM, Ryan PB, Steenland K, Bartell SM. 2011. Retrospective exposure estimation and predicted versus observed serum perfluorooctanoic acid concentrations for participants in the C8 Health Project. Environ Health Perspect 119(12):1760–1765, PMID: 21813367, https://doi.org/10.1289/ehp.1103729.

Shine JP, Ika RV, Ford TE. 1995. Multivariate statistical examination of spatial and temporal patterns of heavy metal contamination in New Bedford Harbor marine sediments. Environ Sci Technol 29(7):1781–1788, PMID: 22176450, https://doi.org/10.1021/es00007a014.

U.S. Census Bureau. 2000. *Census 2000 Summary File 3*. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF3_P053&prodType=table [accessed 17 August 2019].

U.S. EPA (U.S. Environmental Protection Agency). 2015. EPA Cleanups: Communities Around New Bedford Harbor. https://www.epa.gov/new-bedford-harbor [accessed 17 August 2019].

Vieira VM, Fabian MP, Webster TF, Levy JI, Korrick SA. 2017. Spatial variability in ADHD-related behaviors among children born to mothers residing near the New Bedford Harbor Superfund site. Am J Epidemiol 185(10):924–932, PMID: 28444119, https://doi.org/10.1093/aje/kww208.