Supplementary Material

Blood Metal and Volatile Organic Compound Mixture Patterns and Prevalent Cancer Among Adults with a Construction Occupational History: An Exploratory NHANES Study

Francis Boafo and Emmanuel Obeng-Gyasi

All descriptive summaries are unweighted and computed before imputation. The analytic sample contains 136 adults, including 16 with a reported lifetime cancer history and 120 without. No population prevalence is implied.

Figure S1. Participant flow for the analytic sample.

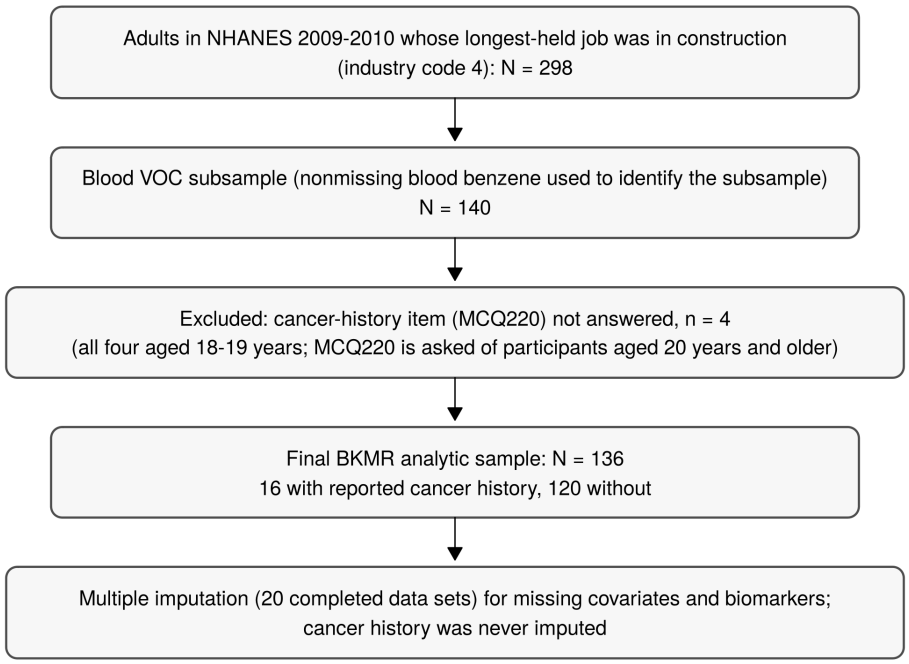


Table S1. Characteristics of the analytic sample by cancer-history status (unweighted). Values are *n* (%) or median [Q1, Q3]. Percentages for income and smoking use observed values as the denominator. Age is top-coded at 80 years by NHANES.

Characteristic	All (*n* = 136)	Cancer history (*n* = 16)	No cancer history (*n* = 120)
Participants, *n*	136	16	120
Age, years, median [Q1, Q3]	54.5 [39.0, 68.2]	74.5 [70.8, 80.0]	51.5 [36.8, 65.0]
Age 80 years (top-coded), *n* (%)	13 (9.6%)	5 (31.2%)	8 (6.7%)
Male, *n* (%)	125 (91.9%)	15 (93.8%)	110 (91.7%)
Non-Hispanic White, *n* (%)	76 (55.9%)	12 (75.0%)	64 (53.3%)
Hispanic (Mexican American or other Hispanic), *n* (%)	37 (27.2%)	1 (6.2%)	36 (30.0%)
Non-Hispanic Black, *n* (%)	20 (14.7%)	3 (18.8%)	17 (14.2%)

Characteristic	All (n = 136)	Cancer history (n = 16)	No cancer history (n = 120)
Other race/ethnicity, n (%)	3 (2.2%)	0 (0.0%)	3 (2.5%)
Less than high school, n (%)	53 (39.0%)	5 (31.2%)	48 (40.0%)
High school graduate or GED, n (%)	44 (32.4%)	6 (37.5%)	38 (31.7%)
More than high school, n (%)	39 (28.7%)	5 (31.2%)	34 (28.3%)
Annual household income under $20,000, n (% of observed)	49 (37.7%)	6 (37.5%)	43 (37.7%)
Income missing, n (%)	6 (4.4%)	0 (0.0%)	6 (5.0%)
Smoking indicator = 1, n (% of observed)	52 (53.6%)	3 (25.0%)	49 (57.6%)
Smoking indicator missing, n (%)	39 (28.7%)	4 (25.0%)	35 (29.2%)
Alcohol, drinks per drinking day, median [Q1, Q3] (observed)	3.0 [2.0, 6.0]	2.0 [1.0, 4.0]	3.5 [2.0, 6.0]
Alcohol missing, n (%)	47 (34.6%)	5 (31.2%)	42 (35.0%)
Duration of longest-held job, years, median [Q1, Q3]	18.0 [10.0, 30.0]	32.5 [20.2, 40.8]	17.5 [8.8, 30.0]
Longest-held job held 10 years or more, n (%)	104 (76.5%)	15 (93.8%)	89 (74.2%)

Note: the smoking rows describe the original binary current-smoking indicator (SMQ040) as it entered the fitted models, for which 39 participants had no entry. Reinspection of the linked SMQ_F file established that all 39 are questionnaire skips following fewer than 100 lifetime cigarettes (SMQ020 = "no") rather than smoking nonresponse. Table 2 of the main text reports the four-level smoking history derived from that linkage, under which no participant lacks a smoking category.

Table S2. Observed blood biomarker distributions (median [Q1, Q3]; observed n before imputation) and detection-limit summary. NHANES reports values below the limit of detection (LOD) as LOD/√2; the fill value and the number and percentage of observed values at the fill value are shown.

(a) Observed blood biomarker distributions

Characteristic	All (n = 136)	Cancer history (n = 16)	No cancer history (n = 120)
Blood cadmium (μg/L), median [Q1, Q3]; observed n	0.570 [0.280, 0.980]; n=135	0.555 [0.252, 0.657]; n=16	0.570 [0.280, 1.025]; n=119
Blood lead (μg/dL)	1.870 [1.250, 2.805]; n=135	2.085 [1.502, 2.775]; n=16	1.810 [1.230, 2.855]; n=119
Total blood mercury (μg/L)	0.730 [0.440, 1.340]; n=135	1.210 [0.447, 1.627]; n=16	0.720 [0.440, 1.230]; n=119
Blood benzene (ng/mL)	0.017 [0.017, 0.138]; n=136	0.017 [0.017, 0.047]; n=16	0.017 [0.017, 0.139]; n=120
Blood ethylbenzene (ng/mL)	0.032 [0.017, 0.077]; n=126	0.025 [0.017, 0.055]; n=13	0.032 [0.017, 0.077]; n=113
Blood m/p-xylene (ng/mL)	0.111 [0.050, 0.217]; n=128	0.075 [0.044, 0.169]; n=15	0.113 [0.051, 0.232]; n=113

(b) Detection-limit summary

Biomarker	Fill value (LOD/√2)	Observed n	n at fill value	% at fill value
Cadmium	0.14	135	18	13.3
Lead	0.47	135	1	0.7
Mercury	0.23	135	22	16.3
Benzene	0.0170	136	74	54.4
Ethylbenzene	0.0170	126	46	36.5

m/p-xylene	0.024	128	13	10.2

Table S3. Missing values before imputation in the analytic sample, *n* (% of column group). Cancer history is complete by construction; benzene is complete because a measured benzene value defined the blood VOC subsample.

Variable	All (*n* = 136)	Cancer history (*n* = 16)	No cancer history (*n* = 120)
Cancer history	0 (0.0%)	0 (0.0%)	0 (0.0%)
Age	0 (0.0%)	0 (0.0%)	0 (0.0%)
Gender	0 (0.0%)	0 (0.0%)	0 (0.0%)
Race/ethnicity	0 (0.0%)	0 (0.0%)	0 (0.0%)
Education	0 (0.0%)	0 (0.0%)	0 (0.0%)
Household income	6 (4.4%)	0 (0.0%)	6 (5.0%)
Smoking indicator	39 (28.7%)	4 (25.0%)	35 (29.2%)
Alcohol	47 (34.6%)	5 (31.2%)	42 (35.0%)
Duration of longest-held job	0 (0.0%)	0 (0.0%)	0 (0.0%)
Cadmium	1 (0.7%)	0 (0.0%)	1 (0.8%)
Lead	1 (0.7%)	0 (0.0%)	1 (0.8%)
Mercury	1 (0.7%)	0 (0.0%)	1 (0.8%)
Benzene	0 (0.0%)	0 (0.0%)	0 (0.0%)
Ethylbenzene	10 (7.4%)	3 (18.8%)	7 (5.8%)
m/p-xylene	8 (5.9%)	1 (6.2%)	7 (5.8%)

Note: missingness is tabulated for the variables as they entered the fitted models. The 39 absent smoking-indicator entries are questionnaire skips following fewer than 100 lifetime cigarettes, not smoking nonresponse; see Table 2 and Section 2.4 of the main text.

Table S4. Spearman correlation matrix for the six blood biomarkers (pairwise complete observations; pairwise *n* = 118 to 136).

Biomarker	Cadmium	Lead	Mercury	Benzene	Ethylbenzene	m/p-xylene
Cadmium	1.00	0.24	−0.13	0.67	0.53	0.43
Lead	0.24	1.00	0.23	0.04	−0.06	−0.11
Mercury	−0.13	0.23	1.00	−0.12	−0.09	−0.05
Benzene	0.67	0.04	−0.12	1.00	0.84	0.76
Ethylbenzene	0.53	−0.06	−0.09	0.84	1.00	0.95
m/p-xylene	0.43	−0.11	−0.05	0.76	0.95	1.00

Table S5. Pooled overall mixture-response estimates from the primary BKMR model: difference in h(z) (probit scale) when all six biomarkers are set to the given quantile versus their medians. Pooled posterior mean, pooled standard deviation (Rubin's rules), and 95% credible interval.

Quantile	Estimate	SD	95% CrI lower	95% CrI upper
0.25	0.080	0.078	−0.073	0.232
0.30	0.066	0.065	−0.061	0.194
0.35	0.056	0.055	−0.052	0.164
0.40	0.035	0.035	−0.033	0.104

Quantile	Estimate	SD	95% CrI lower	95% CrI upper
0.45	0.020	0.020	−0.019	0.060
0.50	0.000	0.000	0.000	0.000
0.55	−0.021	0.015	−0.050	0.009
0.60	−0.044	0.030	−0.103	0.014
0.65	−0.094	0.059	−0.209	0.021
0.70	−0.171	0.105	−0.377	0.035
0.75	−0.221	0.135	−0.485	0.043

Figure S2. Univariate exposure–response functions h(z) for each biomarker with the remaining biomarkers fixed at their medians, from the primary BKMR model. Thin blue lines are the 20 completed data sets; the black line is their mean. Exposures are standardized concentrations on the measured scale.

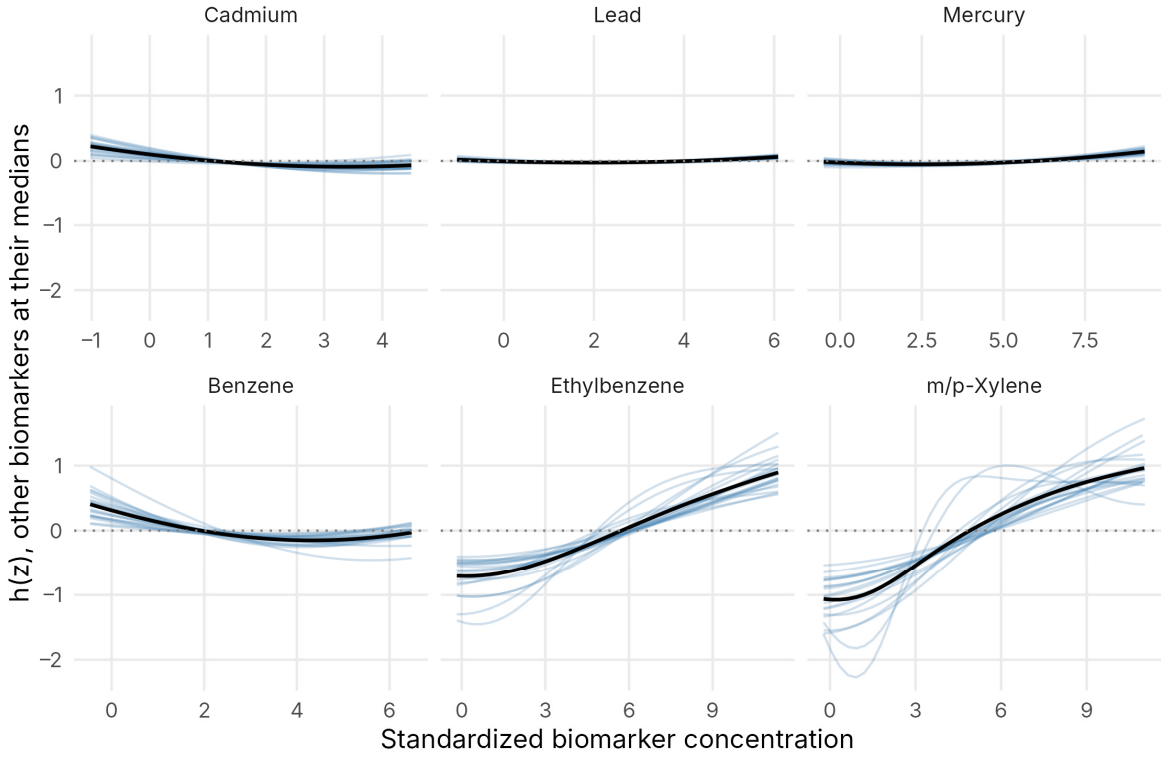


Figure S3. Bivariate exposure–response curves for cadmium and benzene (rows) against each other biomarker (columns) at the 25th, 50th, and 75th percentiles of the column biomarker, with the remaining biomarkers at their medians, for completed data sets 1, 10, and 20 (line types). Curves are posterior means from the primary BKMR model.

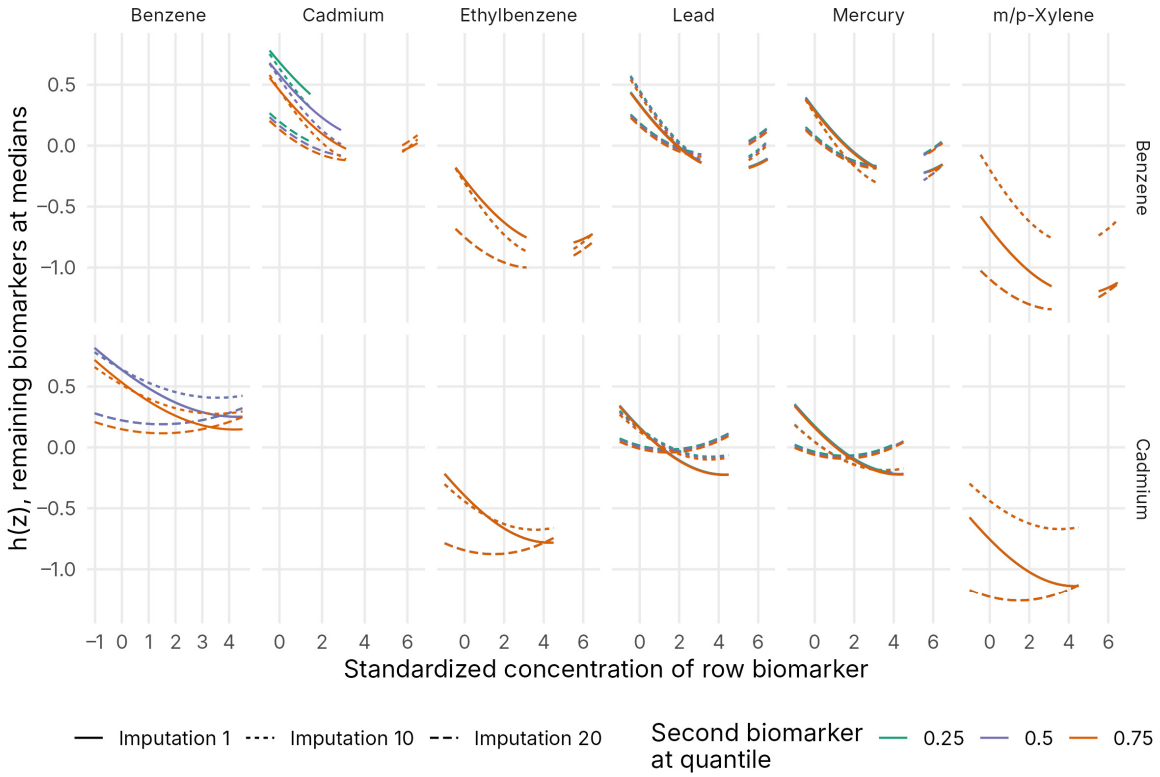


Figure S4. Conditional posterior inclusion probabilities for each biomarker across BKMR specifications (marginal PIPs for the non-hierarchical model). Values are means across the 20 completed data sets (single fit for the complete-case model).

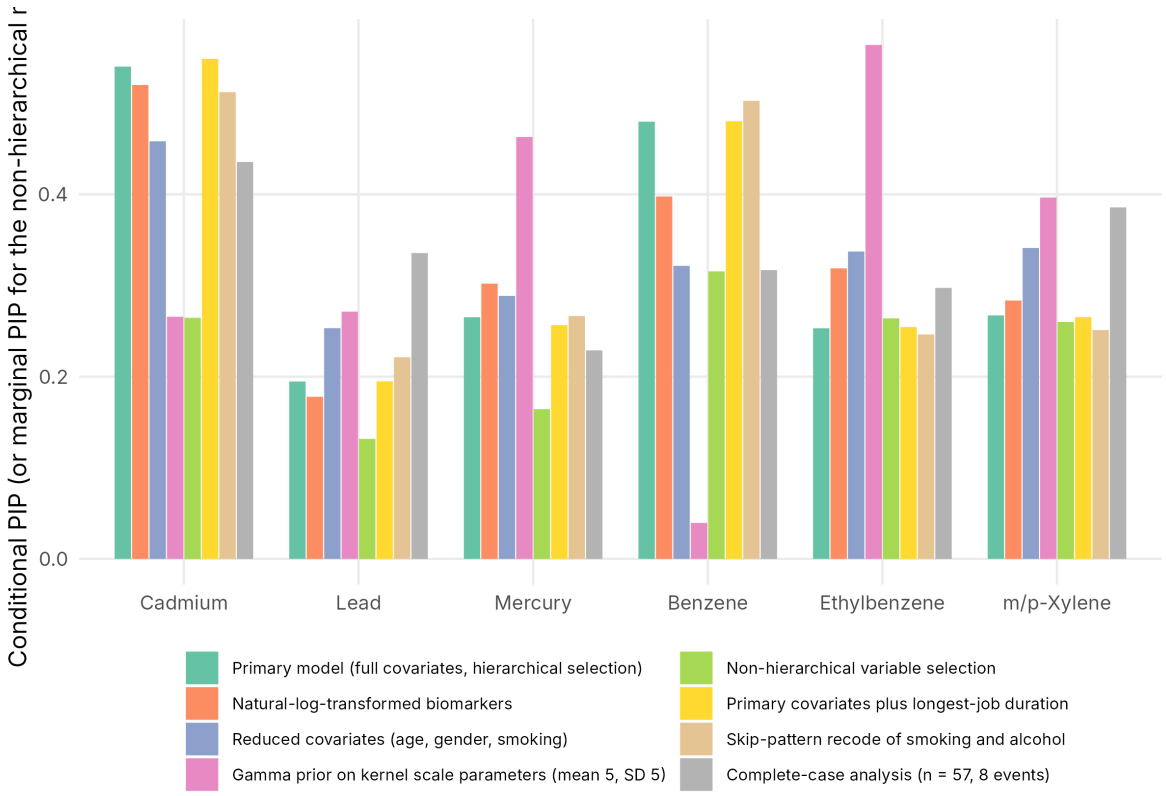


Figure S5. Single-biomarker contrasts (75th vs. 25th percentile, other biomarkers at their medians; probit scale) with 95% credible intervals across BKMR specifications.

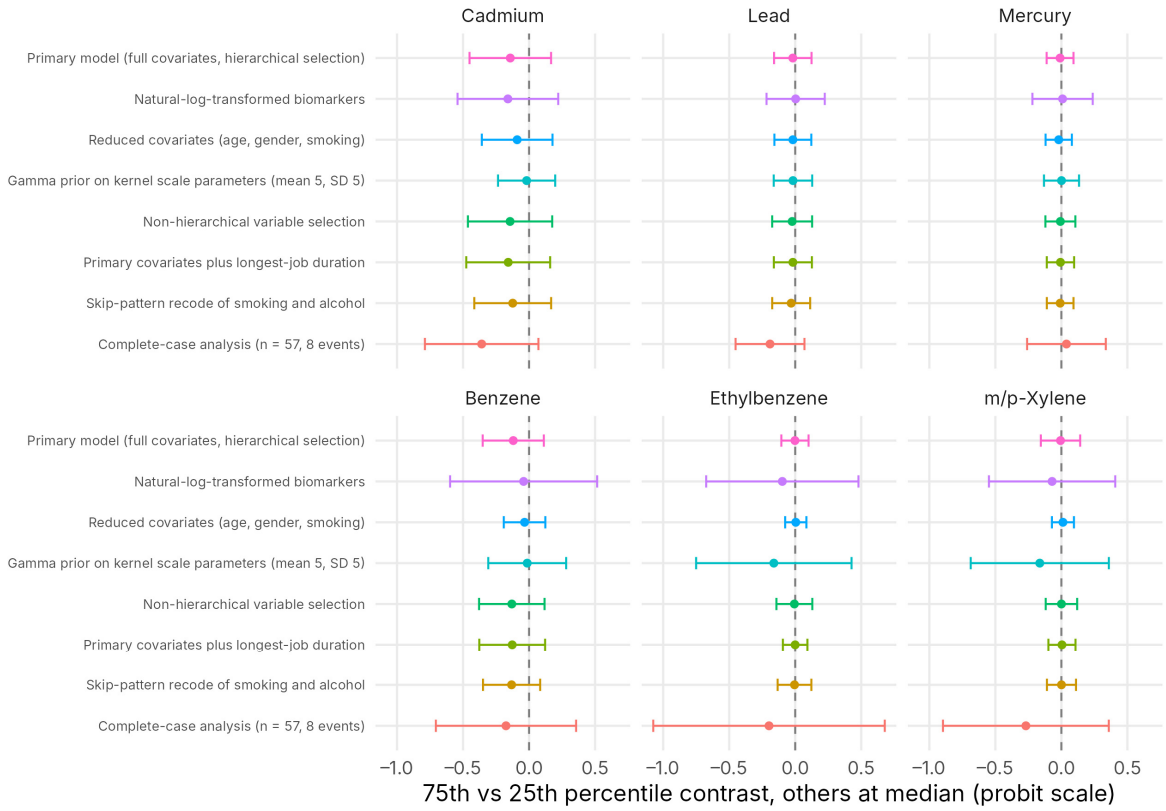


Table S6. Quantile g-computation: pooled positive and negative weights (mean across 20 completed data sets) for the primary covariate set. Weights sum to one within direction and indicate each biomarker's share of the positive or negative partial effect.

Biomarker	Positive weight	Negative weight
Cadmium	0.00	0.09
Lead	0.01	0.02
Mercury	0.14	0.00
Benzene	0.18	0.01
Ethylbenzene	0.67	0.00
m/p-xylene	0.00	0.88