

Sex-specific associations between phenotypes and environmental exposures

Tomás González Zarzar (✉ tomas.gonzalez@umayor.cl)

Universidad Mayor <https://orcid.org/0000-0002-6410-0524>

Nicole Palmiero

<https://orcid.org/0000-0002-3321-4596>

Dokyo Kim

Perelman School of Medicine, University of Pennsylvania <https://orcid.org/0000-0002-4592-9564>

Li Shen

University of Pennsylvania <https://orcid.org/0000-0002-5443-0503>

Molly Hall

Pennsylvania State University

Article

Keywords: PheEWAS, environmental exposures, smoking exposure, albumin, homocysteine, red blood cell inflammation, CVD

Posted Date: October 12th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3246585/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Sex-specific associations between phenotypes and environmental exposures

Tomás González Zarzar^{1,2,3,*}, Nicole E. Palmiero⁴, Dokyoon Kim⁵, Li Shen⁵,
and Molly A. Hall^{4,6,7}

¹Society and Health Research Center, Universidad Mayor, Santiago, Chile

²Facultad de Ciencias Sociales y Artes, Universidad Mayor, Santiago, Chile

³Millennium Nucleus on Sociomedicine, Santiago, Chile

⁴Department of Veterinary and Biomedical Sciences, College of Agricultural Sciences, The Pennsylvania State University, University Park, PA 16802, USA

⁵Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

⁶The Huck Institutes of the Life Science, The Pennsylvania State University, University Park, PA 16802, USA

⁷Penn State Cancer Institute, The Pennsylvania State University, University Park, PA 16802, USA

*Corresponding author

Abstract

Sex and gender differences are ubiquitous in health and disease. We used data from the National Health and Nutrition Examination Survey to identify associations between environmental exposures and health-related phenotypes that differ between males and females. We found a total of 119 sex-specific exposure–phenotype associations, which included phenotypes such as albumin, homocysteine, bilirubin, and hemoglobin levels and exposures to cadmium, lead, benzene, toluene, ethylbenzene, and xylene. Smoking-related exposures were among the exposures with the greatest number of sex-specific phenotypic associations. The positive effects of smoking-related exposure on homocysteine and hemoglobin levels, hematocrit, and red blood cell counts were larger in females than in males. Conversely, the negative effects of smoking-related exposures on albumin and bilirubin levels were larger in males than in females. Our results show potential mechanisms by which women who smoke are at higher risk of cardiovascular diseases, and men have a heightened inflammatory response.

PheEWAS environmental exposures smoking exposure albumin homocysteine
red blood cell inflammation CVD

1 Introduction

2 Health and disease exhibit noticeable sex and gender differences [1]. Recent efforts to
3 better understand differences between males and females in disease risk and preva-
4 lence have demonstrated the importance of considering sex-specific health effects.
5 For example, females were found to have a greater risk than males of developing im-
6 mune diseases, whereas males have the greater risk of developing non-reproductive
7 cancers [2]. Furthermore, cardiovascular diseases were found to have different patho-
8 logic mechanisms in males and females [3], and females were shown to be more
9 susceptible than males to chronic obstructive pulmonary disease (COPD) [4]. Al-
10 though progress has been made in assessing sex differences in health and disease, as
11 shown by these brief examples, sex disparities in health still exist. Furthermore, few
12 biomedical studies report sex and gender differences [5, 6]. Without a doubt, lack
13 of research on sex differences in health and disease can negatively impact our un-
14 derstanding of the biological mechanisms underlying those disparities and maintain
15 or worsen sex disparities in health [6].

16 Lack of research is especially relevant when considering environmental influences
17 on health, which represents an underexplored area of study [7]. Nevertheless, one
18 notable area of research on sex differences in environmental influences on health has
19 been airway diseases [8, 9]. Occupational asthma is more prevalent in females than
20 males because of its high incidence among healthcare workers and teachers, who are
21 more likely to be female than male in the U.S. [10]. On the other hand, smoking
22 is more likely to cause mortality in females than in males [11], and female smokers
23 show a higher rate of decline in lung function than male smokers [12]. Some sex dif-
24 ferences, such as the difference in occupational asthma, are due to different degrees
25 of environmental exposure between males and females [8]. In contrast, other differ-
26 ences across sexes, such as the differences in mortality and lung function between

male and female smokers, are due to differential effects of the same environmental exposure [12]. Differential biological pathways might mediate the latter type of differences, as well as interactions with sex hormones or genetic and epigenetic mechanisms [13]. These studies have allowed a better understanding of the sex differences in airway diseases and improved diagnostic, therapeutic, and prognostic decisions [8]. Therefore, identifying environmental exposures with sex-specific effects on a wide range of phenotypes will enable a better assessment of environmental risks, help guide therapeutic recommendations, and lead to new biological and mechanistic hypotheses.

Genome-wide association studies (GWAS) have been successful in identifying sex-specific associations in multiple phenotypes such as coronary artery disease [14], Alzheimer’s disease [15, 16], glioma [17], and anthropometric measurements [18], among others [19, 20]. Such discoveries identified potential sex differences in the genetic architecture of multiple phenotypes and led to mechanistic hypotheses that aim to explain such differences across sexes. To achieve this purpose, considering the type of difference when evaluating sex differences is vital [21, 22]. Pure sex differences have a significant association in a single sex, quantitative sex differences show differences in the effect sizes between sexes, and qualitative sex differences show opposite effects. Each category can indicate different underlying biological mechanisms, providing greater insight into the nature of these sex differences but also warranting different methodological approaches.

Genome-wide studies have also benefited from using a phenome-wide approach, that is, from using multiple phenotypes. Specifically, phenome-wide association studies (PheWAS) examine the associations of genetic variants with multiple phenotypes, providing valuable information about the complex associations between genotype and phenotype [23]. On the other hand, environment-wide association

53 studies (EWAS) are an agnostic and hypotheses-generating research approach to
54 explore the associations of multiple environmental exposures, known collectively as
55 the exposome, with disease outcomes [24]. Multiple EWAS have identified associa-
56 tions between environmental exposures and type 2 diabetes [25, 26], cardiovascular
57 diseases [27], and high blood pressure [28], among other phenotypes. While eval-
58 uating associations between multiple exposures and a single phenotype in EWAS
59 has proven valuable, considering phenotypes in isolation can hinder the ability to
60 detect exposures with cross-phenotype associations or associations between a single
61 environmental exposure and multiple phenotypes; to our knowledge no studies have
62 interrogated the exposome simultaneously over a wide range of phenotypes.

63 Therefore, phenomic environment-wide association studies (PheEWAS) can re-
64 veal the exposome’s influence on various phenotypes. More specifically, sex-stratified
65 PheEWAS has the potential to reveal the influence of sex on the association between
66 the exposome and human health at the level of the phenome, identify sex-specific
67 environment-phenotype associations, and ultimately improve the potential for pre-
68 cision health in the clinical setting. Here, we present a sex-stratified PheEWAS
69 in the National Health and Nutrition Examination Survey (NHANES) to test the
70 association of the exposome and the phenome and to discover exposures with differ-
71 ential phenotypic associations in males and females. Using data from NHANES, we
72 evaluated 18,937 participants for 58 phenotypes and 272 environmental exposures.
73 We further classified sex-different exposure-phenotype associations into three cate-
74 gories [21]: pure sex-specific associations, in which exposures have an effect on one
75 sex but not the other; quantitative sex-specific associations, in which exposures af-
76 fect both sexes but with different effect sizes in each sex; and qualitative sex-specific
77 associations, in which exposures have inverse effects on males and females. Our
78 results suggest promising avenues to explore exposures that pose differential risks to

79 the sexes, elucidate biological pathways with significant sex-based differences, and
80 open the road to more personalized therapeutics.

81 **Materials and Methods**

82 **NHANES data sets**

83 The primary purpose of this study was to identify exposure–phenotype associations
84 that differ between the sexes using NHANES data sets. NHANES is a program
85 of studies whose purpose is to assess the health and nutritional status of adults
86 and children in the U.S. using data from physical examinations and interviews.
87 The physical examination consists of medical, dental, and physiological measure-
88 ments and laboratory tests. The interview includes demographic, socioeconomic,
89 dietary, and health-related questions. We used a curated set of 1,191 variables and
90 41,474 participants across four survey periods (1999–2000, 2001–2002, 2003–2004,
91 and 2005–2006), which are available on DataDryad [29] and through the NHANES
92 Dataset Explorer [30].

93 Similar to previous studies, we used a discovery (survey periods 1999–2000 and
94 2001–2002) and replication (survey periods 2003–2004 and 2005–2006) design [31,
95 32]. The discovery and replication data sets were divided into males and females,
96 resulting in four different cohorts: discovery females, discovery males, replication
97 females, and replication males. For clarity, we refer to these cohorts using the
98 subscripts df , dm , rf , rm , respectively. The procedures for quality control (QC) and
99 association discovery were performed in each cohort independently.

100 We generated a preliminary selection of phenotypes and exposures based on
101 the categories used in the NHANES. All variables under the NHANES categories
102 of biochemistry, blood, and hormones were selected as phenotypes, which included

103 blood and urine biomarkers, complete blood count measurements, and measure-
 104 ments of hormone levels in blood samples, respectively. Variables under the follow-
 105 ing NHANES categories were selected as exposures: alcohol use, bacterial infection,
 106 cotinine, diakyl, dioxins, food component recall, furans, heavy metals, housing,
 107 hydrocarbons, nutrients, occupation, pcbs, perchlorate, pesticides, phenols, phtha-
 108 lates, phytoestrogens, polybrominated ethers, polyflourochemicals, sexual behavior,
 109 smoking behavior, smoking family, social support, street drug, sun exposure, supple-
 110 ment use, viral infection, and volatile compounds. The preliminary selection yielded
 111 a total of 84 phenotypes and 739 exposures. Following the NHANES recommenda-
 112 tions, we used survey weights to account for the complex survey design. Specifically,
 113 4-year weights were used for the discovery cohorts, whereas 2-year weights were used
 114 for the replication cohorts after dividing by $\frac{1}{2}$ to generate 4-year weights. Three phe-
 115 notypes and 116 exposures did not have survey weight information and were removed
 116 from the data sets.

117 Most data cleaning, QC, association analyses, and visualizations were carried
 118 out using the CLARITE software [32]. All analytical steps and code are available
 119 in Zenodo under the DOI: 10.5281/zenodo.6483503. The repository contains in-
 120 structions and information about the packages and versions needed to replicate the
 121 analysis.

122 **Data quality control**

123 The QC process was carried out in each cohort independently. If a variable was
 124 removed in a single cohort, it was removed from all other cohorts. Only par-
 125 ticipants at least 18 years of age at the moment of data collection and with no
 126 missing data on any covariate were retained ($N = 18,937$). Additionally, low-
 127 density lipoprotein values (NHANES code name: LBDLDL) were adjusted by di-

128 viding the concentration value by 0.7 if the participants were on statins. Af-
 129 ter the complete data set was divided into four cohorts, the sample sizes were:
 130 $N_{df} = 4,724$; $N_{dm} = 4,339$; $N_{rf} = 5,123$; $N_{rm} = 4,751$.

131 Variables were categorized using CLARITE as constant, binary, categorical, and
 132 continuous. Constants, or variables with a single value that were neither covariates
 133 nor phenotypes, were removed. Furthermore, variables with only missing data and
 134 continuous variables with less than ten unique values were removed.

135 Several filters were applied in each cohort to ensure proper sample sizes. Vari-
 136 ables with less than 200 non-missing values and binary or categorical variables with
 137 less than 200 observations per category were removed. Continuous variables with
 138 more than 90% of their values equal to zero were also removed. Thus, from the initial
 139 set of phenotypes and exposures, 59 phenotypes and 281 exposures were retained.

140 Data transformation of continuous variables involved log2 transformation and
 141 z-score normalization. For the z-score normalization, we used the following formula:
 142 $X_{norm} = \frac{X - \mu}{\sigma}$. Outlier values, defined as being three standard deviations away from
 143 the mean, were considered missing values. After this addition of new missing values,
 144 all filters were applied again, and nine additional exposures were removed. Finally,
 145 a Pearson correlation was estimated for all possible pairs of variables, including
 146 exposures and phenotypes, to detect transformed identical variables. One phenotype
 147 was removed because it was found to be the inverse of another variable. The final
 148 data sets used in the PheEWAS included 10 covariates, 58 phenotypes, and 272
 149 exposures.

150 **Statistical Methods**

151 **PheEWAS**

152 A PheEWAS estimates the associations of multiple exposures and phenotypes. In
153 each cohort, an EWAS was performed using 58 phenotypes and 272 exposures and
154 controlling for 10 covariates, including sex, age, socioeconomic status (SES), body
155 mass index (BMI), survey year, and four race/ethnicity categories. Specifically, a
156 generalized linear model was adjusted for survey designs as implemented in CLAR-
157 ITE, using the 4-year weights provided by NHANES. A minimum of 200 complete
158 observations were required for a successful regression.

159 **Sex difference analysis**

160 Sex differences in associations were categorized according to differences in β values.
161 Pure sex differences were significant in one sex but not in the other. Quantitative
162 sex differences were significant in both sexes with β coefficients in the same direction
163 but significantly larger in one sex than in the other. Qualitative sex differences were
164 significant in both sexes but with β coefficients in opposite directions.

165 Two complementary approaches were used to detect sex differences. One ap-
166 proach was to evaluate sex differences across all phenome and environment-wide
167 associations. The other approach used filtering for significant overall associations
168 first. We refer to these approaches as the full evaluation approach and the filtering-
169 first approach, respectively. These two approaches were selected because they have
170 different powers to detect different types of sex differences and are recommended
171 when there is no previous hypothesis on what type of sex difference is likely to be
172 found. Specifically, the full evaluation approach is better suited to detect qualita-
173 tive sex differences, whereas the filtering-first approach is better suited to detect
174 quantitative and pure sex differences [21].

175 To estimate sex differences in exposure–phenotype associations, the following
 176 difference test was used:

$$Z_{diff} = \frac{\beta_f - \beta_m}{\sqrt{se_f^2 + se_m^2}}$$

177 The p –value corresponding to Z_{diff} is referred to as p_{diff} . Because there were
 178 four cohorts, a meta-analysis using an inverse variance method was used to generate
 179 sex-specific results in the female and male cohorts [21, 33].

180 In the full evaluation approach, Z_{diff} was estimated on all possible exposure–
 181 phenotype associations, and significant sex differences were identified as those whose
 182 $p_{diff} < \alpha_{diff}$, where α_{diff} is a Bonferroni-corrected α value based on M total
 183 number of tests, such that $\alpha_{diff} = \frac{0.05}{M}$.

184 In the filtering-first approach, a filter for significant overall association was used
 185 to reduce the number of exposure–phenotype associations. The filter is given by the
 186 following formula:

$$Z_{overall} = \frac{\frac{\beta_f}{se_f^2} + \frac{\beta_m}{se_m^2}}{\sqrt{\frac{1}{se_f^2} + \frac{1}{se_m^2}}}$$

187 The p –value corresponding to $Z_{overall}$ is referred to as $p_{overall}$. The exposure–
 188 phenotype associations whose $p_{overall} < 10^{-5}$ were kept. Next, a difference test was
 189 applied on the filtered subset of exposure–phenotype associations, and significant sex
 190 differences were identified as those whose $p_{diff} < \alpha_{diff}$, where α_{diff} is a Bonferroni-
 191 corrected α value based on the number of filtered tests N , such that $\alpha_{diff} = \frac{0.05}{N}$.

192 Finally, differences that were significant by either approach were considered sig-
 193 nificant sex differences. We classified these differences as qualitative, pure, or quan-
 194 titative sex differences according to the following rules:

- 195 • qualitative: $p_m < 0.05$, $p_f < 0.05$, and β_f and β_m point in different directions.
- 196 • quantitative: $p_m < 0.05$, $p_f < 0.05$, and β_f and β_m point in the same direction.
- 197 • pure: $p_m < 0.05$ or $p_f < 0.05$, but not both conditions together.

198 Results

199 We ran 15,999 association tests in each of the four cohorts from the NHANES, dis-
 200 covery females, discovery males, replication females, and replication males (see the
 201 Methods section for details and sample sizes). The number of converged tests and
 202 correct estimations of regression parameters fluctuated between 15,470 in the repli-
 203 cation males cohort and 15,994 in the discovery females and discovery males cohorts.
 204 These differences were due to some tests having less than 200 complete observations
 205 to run the regression. From the successful tests, 15,466 exposure–phenotype asso-
 206 ciations converged across all cohorts and were used to estimate sex differences. We
 207 identified 119 significant sex-specific associations, of which 49 were classified as pure,
 208 45 as quantitative, and 25 as qualitative.

209 Of the 58 evaluated phenotypes, 35 had sex-specific associations with at least
 210 one exposure. A single phenotype, serum albumin level, had sex-specific associa-
 211 tions with 30 different exposures, the phenotype with the highest number of cross-
 212 exposure associations. Nine of these associations were with smoking-related expo-
 213 sures, such as blood cotinine levels, the number of cigarettes smoked daily, and
 214 whether anyone smokes in the household. Eight associations were with nutrients,
 215 such as folate and vitamins A, D, and E. Six associations were with supplements,
 216 such as folic acid, calcium, and vitamin C. Four associations were with heavy met-
 217 als and volatile compounds. On the other hand, the phenotypes red cell count and
 218 c-reactive protein level had sex-specific associations with eight and six exposures, re-

spectively. Four associations of red cell counts were with heavy metals and volatile compounds, specifically cadmium, lead, ethylbenzene, and toluene. Two associations were with the smoking-related exposures cotinine blood levels, and whether anyone smokes in the household. Two associations of c-reactive protein were with the heavy metal exposures to cadmium and lead and two with food component recall of caffeine and retinol. Finally, the phenotypes hematocrit, high-density lipoprotein (HDL) cholesterol, homocysteine, and bilirubin levels had sex-specific associations with five different exposures each (Figure 1; Figure 2).

Of the 272 evaluated exposures, 53 had sex-specific associations with at least one phenotype. Ninety-five exposures with sex-specific associations belonged to the NHANES exposure categories of nutrients, heavy metals, smoking behavior, and food component recall. Within those categories, lead, cadmium, vitamin A, lutein and zeaxanthin, alcohol, and vitamin E were the exposures associated with the most phenotypes. Lead exposure was associated with 11 phenotypes, including albumin, bilirubin, bone alkaline phosphatase, c-reactive protein, HDL-Cholesterol, homocysteine, and red cell count. Cadmium was associated with eight phenotypes, including albumin, c-reactive protein, monocyte number, red cell count, and white blood cell count. Vitamin A was associated with seven phenotypes, including albumin, c-reactive protein, and lymphocyte number. Lutein and zeaxanthin were associated with six phenotypes, including albumin, HDL-Cholesterol, hematocrit, and red cell count. Alcohol was associated with five phenotypes, including cholesterol, gamma-glutamyl transferase, and iron. Finally, vitamin E was associated with five phenotypes, including albumin, HDL-Cholesterol, gamma-glutamyl transferase, and red cell count (Figure 1; Figure 2). Supplementary Table 1 shows the complete set of results in more detail.

244 Pure sex differences

245 Thirty-six exposures with pure sex-specific associations were categorized as nutri-
 246 ents, smoking behaviors, heavy metals, or food component recall. The phenotypes
 247 that had sex-specific associations with the greatest numbers of exposures were al-
 248 bumin level, red cell count, and HDL cholesterol level (Figure 2; Figure 3; Table 1).
 249 These pure sex-specific associations were partial towards males, as 18 were only
 250 significant in females, whereas 31 were only significant in males.

251 In females, the exposures with pure sex-specific associations were mainly cate-
 252 gorized as heavy metals, food component recall, nutrients, and volatile compounds.
 253 Red cell count, lymphocyte percent, albumin, HDL cholesterol, and homocysteine
 254 levels were among the phenotypes with the highest numbers of pure sex-specific
 255 associations in females. Blood levels of the volatile compounds ethylbenzene ($\beta_f =$
 256 $0.096, \beta_m = -0.020, P_{diff} = 1.53E-05$) and toluene ($\beta_f = 0.115, \beta_m = -0.015, P_{diff} =$
 257 $3.91E-06$) were both associated with red cell counts in females. In contrast, lead
 258 was associated with albumin ($\beta_f = 0.084, \beta_m = -0.028, P_{diff} = 4.31E-09$),
 259 creatinine ($\beta_f = 0.076, \beta_m = -0.005, P_{diff} = 1.96E-06$), and sodium levels
 260 ($\beta_f = 0.07, \beta_m = -0.024, P_{diff} = 4.49E-06$). The exposures with pure sex-specific
 261 associations in males were mainly categorized as nutrients, smoking behaviors, and
 262 supplement use. Albumin level had a pure sex-specific association with 17 different
 263 exposures in males, whereas c-reactive protein level and osmolality had two pure
 264 sex-specific associations in males each.

265 Of the 53 pure sex-specific associations, 19 involved smoking-related exposures.
 266 For example, cotinine levels were associated with albumin levels ($\beta_f = 0.005, \beta_m =$
 267 $-0.089, P_{diff} = 2.36E-13$) and osmolality ($\beta_f = -0.015, \beta_m = -0.086, P_{diff} =$
 268 $8.16E-06$) in males and with red cell count ($\beta_f = 0.074, \beta_m = -0.006, P_{diff} =$
 269 $3.9E-06$) in females. Similarly, the answer to the question “Does anyone smoke

in home?" was associated with albumin levels ($\beta_f = -0.020, \beta_m = -0.179; P_{diff} = 0.1E - 06$) in males and with HDL cholesterol ($\beta_f = -0.280, \beta_m = -0.051, P_{diff} = 3.38E - 06$) and red cell count ($\beta_f = 0.186, \beta_m = 0.002, P_{diff} = 4.52E - 06$) in females. Two of the ten survey questions related to smoking behavior were positively associated with homocysteine levels only in females. In contrast, seven were associated with albumin and one with c-reactive protein only in males (Figure 3).

Quantitative sex differences

Thirty-eight of the quantitative sex-specific associations involved exposures related to heavy metals, nutrients, or smoking behaviors. Hematocrit and albumin, gamma-glutamyl transferase (GGT), triglyceride, homocysteine, and bilirubin levels each had quantitative sex-specific associations with three different exposures (Figure 2; Figure 3; Table 2). Of the 45 quantitative sex-specific associations, 18 had greater effect sizes in females, whereas 27 had greater effect sizes in males. Among the heavy metal exposures, the effects of cadmium exposure on monocyte number ($\beta_f = 0.092, \beta_m = 0.166, P_{diff} = 7.48E - 06$), segmented neutrophils number ($\beta_f = 0.111, \beta_m = 0.206, P_{diff} = 2.2E - 08$), c-reactive protein level ($\beta_f = 0.045, \beta_m = 0.144, P_{diff} = 6.01E - 09$), mean cell hemoglobin level ($\beta_f = 0.047, \beta_m = 0.135, P_{diff} = 1.01E - 09$), mean cell volume ($\beta_f = 0.057, \beta_m = 0.145, P_{diff} = 1.27E - 10$), and white blood cell count ($\beta_f = 0.151, \beta_m = 0.252, P_{diff} = 9.7E - 10$) were greater in males than in females; however, the effect of cadmium exposure on red cell count was greater in females than in males ($\beta_f = 0.127, \beta_m = 0.027, P_{diff} = 6.97E - 11$). On the other hand, the effect of lead exposure on bone alkaline phosphatase level ($\beta_f = 0.233, \beta_m = 0.049, P_{diff} = 5.87E - 14$), homocysteine level ($\beta_f = 0.210, \beta_m = 0.130, P_{diff} = 5.17E - 06$), red cell count ($\beta_f = 0.190, \beta_m = 0.071, P_{diff} = 1.1E - 09$), and alkaline phosphatase level ($\beta_f = 0.193, \beta_m = 0.064, P_{diff} = 7.9E - 12$) was

greater in females than in males.

Several smoking-related exposures, including individual smoking, familial smoking, and cotinine levels, had quantitative sex-specific phenotypic associations. Cotinine levels and the answers to the survey questions “Does anyone smoke in home?” and “Current or Past Cigarette Smoker?” each had greater positive effects on hemocrit in females than in males (Figure 3). Similarly, quantitative associations between current or past cigarette smoking and hemoglobin level ($\beta_f = 0.429, \beta_m = 0.236, P_{diff} = 1.25E-06$) and between the average number of cigarettes smoked per day during the past 30 days and homocysteine level ($\beta_f = 0.154, \beta_m = 0.038, P_{diff} = 2.71E-06$) had greater effects in females than in males. Conversely, all other quantitative associations between smoking-related exposures and bilirubin level, creatinine level, segmented neutrophils number, mean cell hemoglobin level and white blood cell count had greater effects in males than in females (Figure 3).

Quantitative associations involving serum albumin, triglyceride, and GGT levels tended to have larger effects in males than in females. For example, the associations between g-tocopherol and GGT level ($\beta_f = 0.066, \beta_m = 0.129, P_{diff} = 2.08E-05$), g-tocopherol and triglyceride level ($\beta_f = 0.20, \beta_m = 0.319, P_{diff} = 1.35E-10$), retinyl palmitate and triglyceride level ($\beta_f = 0.273, \beta_m = 0.394, P_{diff} = 7.65E-11$), vitamin A and albumin level ($\beta_f = 0.145, \beta_m = 0.258, P_{diff} = 8.65E-11$), and vitamin E and GGT level ($\beta_f = 0.032, \beta_m = 0.145, P_{diff} = 4.3E-07$) had greater effects sizes in males than in females. Conversely, associations with homocysteine and urine albumin tended to have greater effects in females than in males. For example, the associations between iron supplement use and homocysteine level ($\beta_f = -0.127, \beta_m = -0.064, P_{diff} = 1.78E-07$) and between cesium level in urine and urine albumin level ($\beta_f = 0.504, \beta_m = 0.362, P_{diff} = 2.74E-09$) had greater effect sizes in females than in males.

Qualitative sex differences

Eighteen qualitative sex-specific associations involved exposure to nutrients, food component recall, and heavy metals. Exposures to lead, lutein and zeaxanthin, retinyl palmitate, and vitamin A each had qualitative sex-specific associations with more than one phenotype, whereas the phenotypes with more than one qualitative sex-specific association included albumin and c-reactive protein levels and lymphocyte percent (Figure 2; Figure 3; Table 3). Various folate and iron measurements were negatively associated with albumin levels in females but positively associated with albumin levels in males (Figure 3). There were qualitative sex-specific associations between albumin level and retinyl palmitate ($\beta_f = -0.053, \beta_m = 0.050, P_{diff} = 1.08E - 09$), blood o-xylene ($\beta_f = 0.072, \beta_m = -0.045, P_{diff} = 1.82E - 05$), and blood m-/p-xylene ($\beta_f = 0.076, \beta_m = -0.050, P_{diff} = 1.69E - 06$); between lead exposure and osmolality ($\beta_f = 0.044, \beta_m = -0.061, P_{diff} = 5.91E - 08$), bilirubin level ($\beta_f = 0.068, \beta_m = -0.029, P_{diff} = 6.11E - 07$), and c-reactive protein level ($\beta_f = -0.075, \beta_m = 0.049, P_{diff} = 8.48E - 10$); and between c-reactive protein level and various exposures, including caffeine ($\beta_f = -0.058, \beta_m = 0.034, P_{diff} = 4.45E - 07$), and retinol ($\beta_f = 0.034, \beta_m = -0.066, P_{diff} = 1.01E - 05$).

Discussion

We conducted a PheEWAS to identify sex-specific associations between 58 phenotypes and 272 environmental exposures among more than 18,000 NHANES participants. We identified environmental exposures with sex-specific effects on several blood phenotypes. The high-throughput and agnostic nature of our analysis provides avenues for further research into the biological, environmental, and mechanistic aspects of these sex differences. Furthermore, our study demonstrates the utility of

345 using an explicit PheEWAS framework to target potential differences between sexes
346 instead of evaluating sex differences as a complementary analysis.

347 We found a total of 119 exposure–phenotype associations that were significantly
348 different between the sexes. Many of the associations involved smoking-related ex-
349 posures, including smoking behavior, familial smoking, and cotinine levels. Further-
350 more, tobacco smoke can be a significant source of exposure to some heavy metals,
351 such as cadmium [34]. Exposure to other heavy metals, such as lead, does not oc-
352 cur predominantly through tobacco smoke; however, active and passive exposure
353 to tobacco smoking in adults and children has been associated with higher blood
354 concentrations of lead [35, 36]. Tobacco smoke is also an important source of expo-
355 sure to a variety of volatile compounds such as benzene, toluene, ethylbenzene, and
356 xylene [37]. Therefore, exposure to heavy metals and volatile compounds can be
357 correlated with active or passive exposure to tobacco smoke. In our study, results
358 from exposures to heavy metals and volatile compounds replicate those of more di-
359 rect assessments of smoking behavior, such as measurements of cotinine levels and
360 answers to smoking-related questionnaires. For example, cadmium exposure had a
361 greater effect on albumin levels in males, similar to the effects of cotinine exposure
362 and smoking behavior on albumin levels.

363 Previous studies identified sex differences in the prevalence and health effects of
364 smoking [38, 39]. We found sex differences in the associations of smoking behavior,
365 familial smoking, and cotinine levels with several diverse phenotypes. For example,
366 the number of cigarettes currently smoked per day and the frequency of cigarette
367 smoking during the past 30 days were positively associated with plasma homocys-
368 teine (Hcy) levels in females but not in males. Similarly, the average number of
369 cigarettes smoked per day in the last 30 days had a greater effect on Hcy levels
370 in females than in males. Hcy is an amino acid formed from the demethylation of

methionine [40]. Smoking has long been known to be associated with elevated Hcy levels [41], and both smoking and Hcy levels have been associated with increased risks of cardiovascular disease and COPD [42–45]. Although males on average have higher Hcy levels than females, smoking has a greater effect on Hcy levels in females than in males, and females might be at greater risk of the harmful effects of elevated Hcy levels than males [41, 46]. Furthermore, there is increasing evidence that females who smoke are at a greater risk than males who smoke of developing cardiovascular disease [3, 47] and COPD [4]. It has been hypothesized that the particular relation between smoking and Hcy levels in females might be mediated by hormonal mechanisms, particularly through estrogen metabolism and the influence of estrogen on methionine catabolism [46, 48, 49]. Although the mechanisms connecting smoking behavior and Hcy are still unknown, our results replicate previous findings that smoking has a greater effect on Hcy levels in females than in males, which might explain the heightened risk of cardiovascular disease in females who smoke.

Smoking behavior, familial smoking, cotinine levels, and blood levels of cadmium, lead, ethylbenzene, and toluene had female-specific positive associations with hematocrit, hemoglobin levels, and red cell counts. These findings are in line with previous studies showing that smoking increases hematocrit, hemoglobin levels, and red cell counts [50, 51]. One plausible hypothesis to explain these results is that carbon monoxide exposure from smoking induces hypoxia, which in turn induces changes in hematocrit, hemoglobin levels, and red cell counts [52].

Smoking behavior, familial smoking, cotinine levels, and cadmium levels had negative associations with serum albumin levels in males but not in females. Albumin is a water-soluble protein constituent of blood. Previous studies found that albumin levels are lower in smokers than in never-smokers [53, 54], and that smoking inten-

397 sity is inversely related to albumin levels [55, 56]. Another study found a negative
398 association between albumin levels and smoking behavior in males but not in fe-
399 males [57]. Lower albumin levels have also been associated with increased incidence
400 of type 2 diabetes [58], cardiovascular disease [53, 54], cancer-related mortality [59],
401 and all-cause mortality [60]. Because the associations between albumin levels and
402 diverse diseases are stronger in smokers than in never-smokers, cigarette smoking
403 has been hypothesized to directly affect albumin levels [54]. Low levels of albumin
404 are a characteristic of the acute phase of the inflammatory response, together with
405 increased c-reactive protein levels and white blood cell counts [55]. Therefore, it
406 has been hypothesized that the inflammatory response caused by smoking connects
407 smoking exposure, lower albumin levels, and increased disease risk [53]. Besides the
408 sex-specific association between smoking-related exposures and albumin levels, we
409 found that smoking behavior and blood cadmium levels were positively associated
410 with c-reactive protein level and white blood cell count, with greater effect sizes in
411 males. These results replicate previous findings on the role of smoking in inflam-
412 matory responses and suggest a heightened smoking-induced inflammatory response
413 that is specific to males.

414 We found that positive answers to several smoking behavior questions, including
415 daily cigarettes smoked, average cigarettes smoked daily in the past month, and
416 years smoked this amount, were negatively associated with bilirubin levels, with
417 greater effect sizes in males. Previous research found that smoking is associated
418 with low bilirubin levels [61–64]. Because bilirubin acts as an antioxidant, the
419 relatively low bilirubin levels in smokers might be due to oxidative stress produced
420 by the free radicals in tobacco smoke. Another hypothesis that has been proposed
421 is that tobacco consumption increases the activity of UDP-glucuronosyltransferase,
422 which is crucial for bilirubin conjugation [64]. Sex differences similar to the ones we

423 observed in bilirubin levels and smoking behaviors were found previously, although
424 a mechanistic explanation for these differences has not yet been outlined [61].

425 We found that serum albumin levels seem to be more susceptible to environmen-
426 tal exposures in males than in females. Of the 22 pure or quantitative sex-specific
427 associations involving serum albumin levels, 20 were either exclusive to males or had
428 a greater effect size in males. By contrast, associations involving serum Hcy levels
429 tended to be specific to females. In all five pure or quantitative sex-specific asso-
430 ciations involving serum Hcy, the effect was either exclusive to females or stronger
431 in females than in males. These results suggest that sex-specific mechanisms might
432 underlie the metabolism of Hcy and serum albumin, implicating sex differences in
433 methionine catabolism and inflammatory responses.

434 Of the 119 significant sex-specific associations identified in our PheEWAS, 94
435 were classified as pure or quantitative, and only 25 were classified as qualitative.
436 This favor toward quantitative and pure sex differences suggests that males and
437 females tend to differ in the degree of associations but not in the direction of ef-
438 fects. Furthermore, quantitative and pure sex differences suggest associations with
439 the same underlying mechanisms in males and females, whereas qualitative differ-
440 ences suggest associations with different mechanisms in males and females. It is
441 possible that some differences that are in fact quantitative might have been misclas-
442 sified as pure because of a lack of power to detect a significant association in one
443 of the sexes. On the other hand, qualitative differences suggest completely different
444 mechanisms or responses in each sex. The different underlying mechanisms of quali-
445 tative sex differences and pure or quantitative sex differences might also explain why
446 qualitative differences were rarer and more heterogeneous than pure or quantitative
447 differences. Pure and quantitative sex differences showed greater similarities than
448 qualitative sex differences in the exposures and phenotypes involved, for example,

449 smoking-related exposures. Nevertheless, the bias towards pure and quantitative
450 sex differences might be due to higher power to detect pure and quantitative sex
451 differences at the expense of qualitative differences.

452 Another caveat to our study is that sex differences in the range of variation
453 in the level of exposures might be observed as sex differences in the association
454 between exposures and phenotypes. For example, if males have a minimal range
455 of variation in a given exposure, there might not be enough variation to detect
456 an association with a given phenotype, whereas females have a greater range of
457 variation in the exposure, revealing a positive association between the exposure and
458 the phenotype. In such a case, a phenotype–exposure association would appear to
459 differ between males and females only because of the different degrees of exposure
460 and not necessarily because of different underlying mechanisms.

461 Although we identified several phenome–environment associations, the direction
462 of causality remains unknown, as in the case of EWAS more generally [65]. Several
463 sex-different associations might be due to differential behaviors that men and women
464 deploy in the face of similar events. For example, if levels of a particular pheno-
465 type are associated with a disease, sex-specific associations between disease-related
466 exposures and the phenotype might be due to different behaviors in the face of the
467 disease. Such behavioral differences are not trivial in analyses of sex differences,
468 because studies have shown gender differences in health-related behaviors such as
469 healthcare seeking and medication intake and adherence [66–68]. Therefore, our re-
470 sults cannot easily be ascribed simply to sex differences if we understand them solely
471 on the basis of biology. We hope that our study will encourage further research to
472 better understand the mechanisms of the sex-specific associations highlighted here.
473 In that way, intra-sex variation and hormone-based, genetic, and epigenetic mech-
474 anisms can be taken into account to explain sex-specific phenotype–environment

475 associations.

476 Author contributions.

477 TGZ, MAH, NEP, DK, and LS contributed to the conception and design of the
478 study. NEP and TGZ performed the cleaning of the databases and the analysis.
479 TGZ wrote the first draft of the manuscript. All authors contributed to the revision
480 of the manuscript, and read and approved the submitted version.

References

1. Mauvais-Jarvis, F. *et al.* Sex and Gender: Modifiers of Health, Disease, and Medicine. *The Lancet* **396**, 565–582. ISSN: 0140-6736. (2021) (Aug. 2020).
2. Klein, S. L. & Flanagan, K. L. Sex Differences in Immune Responses. *Nature Reviews Immunology* **16**, 626–638. ISSN: 1474-1741. (2020) (Oct. 2016).
3. Regitz-Zagrosek, V. & Kararigas, G. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. *Physiological Reviews* **97**, 1–37. ISSN: 1522-1210 (Jan. 2017).
4. DeMeo, D. L. *et al.* Women Manifest More Severe COPD Symptoms across the Life Course. *International Journal of Chronic Obstructive Pulmonary Disease* **13**, 3021–3029. ISSN: 1176-9106. (2021) (Oct. 2018).
5. Rogers, W. & Ballantyne, A. When Is Sex-Specific Research Appropriate? *IJ-FAB: International Journal of Feminist Approaches to Bioethics* **1**, 36–57. ISSN: 1937-4585. JSTOR: 40339177. (2018) (2008).
6. Lindemann, H. The Woman Question in Medicine: An Update. *Hastings Center Report* **42**, 38–45. ISSN: 0093-0334. (2018) (Apr. 2012).
7. Mallampalli, M. P. *et al.* Role of Environment and Sex Differences in the Development of Autoimmune Diseases: A Roundtable Meeting Report. *Journal of Women's Health* **22**, 578–586. ISSN: 1540-9996. (2023) (July 2013).
8. Raghavan, D. & Jain, R. Increasing Awareness of Sex Differences in Airway Diseases. *Respirology* **21**, 449–459. ISSN: 1440-1843. (2021) (2015).
9. DeBolt, C. & Harris, D. The Impact of Social Determinants of Health on Gender Disparities Within Respiratory Medicine. *Clinics in Chest Medicine. Gender and Respiratory Disease* **42**, 407–415. ISSN: 0272-5231. (2023) (Sept. 2021).

10. McHugh, M. K., Symanski, E., Pompeii, L. A. & Delclos, G. L. Prevalence of Asthma by Industry and Occupation in the U.S. Working Population. *American Journal of Industrial Medicine* **53**, 463–475. ISSN: 1097-0274. (2021) (2010).
11. Centers for Disease Control and Prevention (CDC). Annual Smoking-Attributable Mortality, Years of Potential Life Lost, and Economic Costs—United States, 1995-1999. *MMWR. Morbidity and mortality weekly report* **51**, 300–303. ISSN: 0149-2195 (Apr. 2002).
12. Gan, W. Q., Man, S. P., Postma, D. S., Camp, P. & Sin, D. D. Female Smokers beyond the Perimenopausal Period Are at Increased Risk of Chronic Obstructive Pulmonary Disease: A Systematic Review and Meta-Analysis. *Respiratory Research* **7**, 52. ISSN: 1465-993X. (2021) (Mar. 2006).
13. Pollard, K. M. Gender Differences in Autoimmunity Associated with Exposure to Environmental Factors. *Journal of Autoimmunity. Special Issue: Gender, Sex Hormones, Pregnancy and Autoimmunity* **38**, J177–J186. ISSN: 0896-8411. (2022) (May 2012).
14. Liu, L. Y., Schaub, M. A., Sirota, M. & Butte, A. J. Sex Differences in Disease Risk from Reported Genome-Wide Association Study Findings. *Human Genetics* **131**, 353–364. ISSN: 1432-1203. (2023) (Mar. 2012).
15. Dumitrescu, L. *et al.* Sex Differences in the Genetic Predictors of Alzheimer’s Pathology. *Brain* **142**, 2581–2589. ISSN: 0006-8950. (2023) (Sept. 2019).
16. Hu, Y. *et al.* Rs34331204 Regulates TSPAN13 Expression and Contributes to Alzheimer’s Disease with Sex Differences. *Brain* **143**, e95. ISSN: 0006-8950. (2023) (Nov. 2020).
17. Ostrom, Q. T. *et al.* Sex-Specific Glioma Genome-Wide Association Study Identifies New Risk Locus at 3p21.31 in Females, and Finds Sex-Differences in Risk at 8q24.21. *Scientific Reports* **8**, 7352. ISSN: 2045-2322. (2023) (May 2018).
18. Winkler, T. W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLOS Genetics* **11**, e1005378. ISSN: 1553-7404. (2020) (Oct. 2015).
19. Rask-Andersen, M., Karlsson, T., Ek, W. E. & Johansson, Å. Genome-Wide Association Study of Body Fat Distribution Identifies Adiposity Loci and Sex-Specific Genetic Effects. *Nature Communications* **10**, 339. ISSN: 2041-1723. (2023) (Jan. 2019).
20. Choquet, H. *et al.* A Large Multiethnic GWAS Meta-Analysis of Cataract Identifies New Risk Loci and Sex-Specific Effects. *Nature Communications* **12**, 3595. ISSN: 2041-1723. (2023) (June 2021).

21. Winkler, T. W. *et al.* Approaches to Detect Genetic Effects That Differ between Two Strata in Genome-Wide Meta-Analyses: Recommendations Based on a Systematic Evaluation. *PloS One* **12**, e0181038. ISSN: 1932-6203 (2017).
22. Khramtsova, E. A. *et al.* Quality Control and Analytic Best Practices for Testing Genetic Models of Sex Differences in Large Populations. *Cell* **186**, 2044–2061. ISSN: 0092-8674. (2023) (May 2023).
23. Bush, W. S., Oetjens, M. T. & Crawford, D. C. Unravelling the Human Genome–Phenome Relationship Using Phenome-Wide Association Studies. *Nature Reviews Genetics* **17**, 129–145. ISSN: 1471-0064. (2022) (Mar. 2016).
24. Zheng, Y. *et al.* Design and Methodology Challenges of Environment-Wide Association Studies: A Systematic Review. *Environmental Research* **183**, 109275. ISSN: 0013-9351. (2022) (Apr. 2020).
25. Patel, C. J., Bhattacharya, J. & Butte, A. J. An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *PLOS ONE* **5**, e10746. ISSN: 1932-6203. (2022) (May 2010).
26. Hall, M. A. *et al.* Environment-Wide Association Study (EWAS) for Type 2 Diabetes in the Marshfield Personalized Medicine Research Project Biobank. *Biocomputing 2014*, 200–211. (2023) (Nov. 2013).
27. Zhuang, X. *et al.* Toward a Panoramic Perspective of the Association between Environmental Factors and Cardiovascular Disease: An Environment-Wide Association Study from National Health and Nutrition Examination Survey 1999–2014. *Environment International* **118**, 146–153. ISSN: 0160-4120. (2022) (Sept. 2018).
28. McGinnis, D. P., Brownstein, J. S. & Patel, C. J. Environment-Wide Association Study of Blood Pressure in the National Health and Nutrition Examination Survey (1999–2012). *Scientific Reports* **6**, 30373. ISSN: 2045-2322. (2022) (July 2016).
29. Patel, C. J. *et al.* A Database of Human Exposomes and Phenomes from the US National Health and Nutrition Examination Survey. *Scientific Data* **3**, 160096. ISSN: 2052-4463. (2020) (Oct. 2016).
30. Patel, C. J. *et al.* Data from: A Database of Human Exposomes and Phenomes from the US National Health and Nutrition Examination Survey 2016. (2020).
31. Hall, M. A. *et al.* Detection of Pleiotropy through a Phenome-Wide Association Study (PheWAS) of Epidemiologic Data as Part of the Environmental Architecture for Genes Linked to Environment (EAGLE) Study. *PLOS Genetics* **10**, e1004678. ISSN: 1553-7404. (2020) (Dec. 2014).
32. Lucas, A. M. *et al.* CLARITE Facilitates the Quality Control and Analysis Process for EWAS of Metabolic-Related Traits. *Frontiers in Genetics* **10**. ISSN: 1664-8021. (2020) (2019).

33. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and Efficient Meta-Analysis of Genomewide Association Scans. *Bioinformatics* **26**, 2190–2191. ISSN: 1367-4803. (2020) (Sept. 2010).
34. Nordberg, G. F., Fowler, B. A. & Nordberg, M. *Handbook on the Toxicology of Metals* ISBN: 978-0-12-397339-9 (Academic Press, Aug. 2014).
35. Apostolou, A. *et al.* Secondhand Tobacco Smoke: A Source of Lead Exposure in US Children and Adolescents. *American Journal of Public Health* **102**, 714–722. ISSN: 0090-0036. (2022) (Apr. 2012).
36. Mannino, D. M., Homa, D. M., Matte, T. & Hernandez-Avila, M. Active and Passive Smoking and Blood Lead Levels in U.S. Adults: Data from the Third National Health and Nutrition Examination Survey. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco* **7**, 557–564. ISSN: 1462-2203 (Aug. 2005).
37. Chambers, D. M., Ocariz, J. M., McGuirk, M. F. & Blount, B. C. Impact of Cigarette Smoking on Volatile Organic Compound (VOC) Blood Levels in the U.S. Population: NHANES 2003-2004. *Environment International* **37**, 1321–1328. ISSN: 1873-6750 (Nov. 2011).
38. McHugh, R. K., Votaw, V. R., Sugarman, D. E. & Greenfield, S. F. Sex and Gender Differences in Substance Use Disorders. *Clinical Psychology Review. Gender and Mental Health* **66**, 12–23. ISSN: 0272-7358. (2022) (Dec. 2018).
39. Appelman, Y., van Rijn, B. B., ten Haaf, M. E., Boersma, E. & Peters, S. A. E. Sex Differences in Cardiovascular Risk Factors and Disease Prevention. *Atherosclerosis* **241**, 211–218. ISSN: 0021-9150. (2020) (July 2015).
40. Finkelstein, J. D. & Martin, J. J. Homocysteine. *The International Journal of Biochemistry & Cell Biology* **32**, 385–389. (2021) (2000).
41. Nygård, O., Refsum, H., Ueland, P. M. & Vollset, S. E. Major Lifestyle Determinants of Plasma Total Homocysteine Distribution: The Hordaland Homocysteine Study. *The American Journal of Clinical Nutrition* **67**, 263–270. ISSN: 0002-9165. (2021) (Feb. 1998).
42. O’Callaghan, P., Meleady, R., Fitzgerald, T., Graham, I. & European COMAC group. Smoking and Plasma Homocysteine. *European Heart Journal* **23**, 1580–1586. ISSN: 0195-668X (Oct. 2002).
43. Chaudhary, D., Sharma, N. & Senapati, S. Serum Homocysteine Could Be Used as a Predictive Marker for Chronic Obstructive Pulmonary Disease: A Meta-Analysis. *Frontiers in Public Health* **7**, 69. ISSN: 2296-2565 (2019).
44. Forey, B. A., Thornton, A. J. & Lee, P. N. Systematic Review with Meta-Analysis of the Epidemiological Evidence Relating Smoking to COPD, Chronic Bronchitis and Emphysema. *BMC pulmonary medicine* **11**, 36. ISSN: 1471-2466 (June 2011).

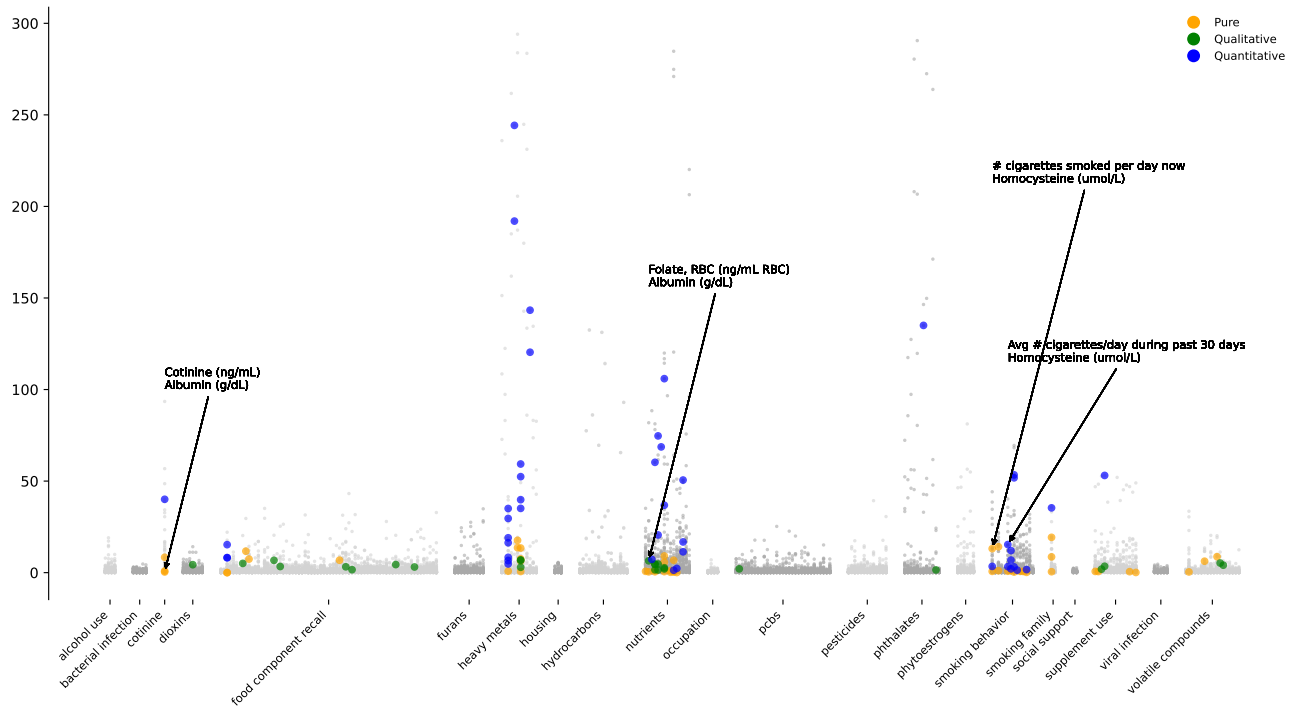
45. Chrysant, S. G. & Chrysant, G. S. The Current Status of Homocysteine as a Risk Factor for Cardiovascular Disease: A Mini Review. *Expert Review of Cardiovascular Therapy* **16**, 559–565. ISSN: 1477-9072. (2021) (Aug. 2018).
46. Stanisławska-Sachadyn, A. *et al.* Evidence for Sex Differences in the Determinants of Homocysteine Concentrations. *Molecular Genetics and Metabolism* **93**, 355–362. ISSN: 1096-7206 (Apr. 2008).
47. Huxley, R. R. & Woodward, M. Cigarette Smoking as a Risk Factor for Coronary Heart Disease in Women Compared with Men: A Systematic Review and Meta-Analysis of Prospective Cohort Studies. *Lancet (London, England)* **378**, 1297–1305. ISSN: 1474-547X (Oct. 2011).
48. Baron, J. A., La Vecchia, C. & Levi, F. The Antiestrogenic Effect of Cigarette Smoking in Women. *American Journal of Obstetrics and Gynecology* **162**, 502–514. ISSN: 0002-9378. (2021) (Feb. 1990).
49. Blom, H. J. *et al.* Differences between Premenopausal Women and Young Men in the Transamination Pathway of Methionine Catabolism, and the Protection against Vascular Disease. *European Journal of Clinical Investigation* **18**, 633–638. ISSN: 1365-2362. (2021) (1988).
50. Van Tiel, E. D. *et al.* Quitting Smoking May Restore Hematological Characteristics within Five Years. *Annals of Epidemiology* **12**, 378–388. ISSN: 1047-2797. (2021) (Aug. 2002).
51. Malenica, M. *et al.* Effect of Cigarette Smoking on Haematological Parameters in Healthy Population. *Medical Archives* **71**, 132–136. ISSN: 0350-199X. (2022) (Apr. 2017).
52. Helman, N. & Rubenstein, L. S. The Effects of Age, Sex, and Smoking on Erythrocytes and Leukocytes. *American Journal of Clinical Pathology* **63**, 35–44. ISSN: 0002-9173. (2021) (Jan. 1975).
53. Nelson, J. J. *et al.* Serum Albumin Level as a Predictor of Incident Coronary Heart Disease: The Atherosclerosis Risk in Communities (ARIC) Study. *American Journal of Epidemiology* **151**, 468–477. ISSN: 0002-9262. (2021) (Mar. 2000).
54. Shaper, A. G., Wannamethee, S. G. & Whincup, P. H. Serum Albumin and Risk of Stroke, Coronary Heart Disease, and Mortality: The Role of Cigarette Smoking. *Journal of Clinical Epidemiology* **57**, 195–202. ISSN: 0895-4356. (2021) (Feb. 2004).
55. Bakhru, A. & Erlinger, T. P. Smoking Cessation and Cardiovascular Disease Risk Factors: Results from the Third National Health and Nutrition Examination Survey. *PLOS Medicine* **2**, e160. ISSN: 1549-1676. (2021) (June 2005).

56. Jang, E. S. *et al.* Effects of Coffee, Smoking, and Alcohol on Liver Function Tests: A Comprehensive Cross-Sectional Study. *BMC Gastroenterology* **12**, 1–12. ISSN: 1471-230X. (2021) (Dec. 2012).
57. Fröhlich, M. *et al.* Independent Association of Various Smoking Characteristics with Markers of Systemic Inflammation in Men: Results from a Representative Sample of the General Population (MONICA Augsburg Survey 1994/95). *European Heart Journal* **24**, 1365–1372. ISSN: 0195-668X. (2021) (July 2003).
58. Stranges, S. *et al.* Additional Contribution of Emerging Risk Factors to the Prediction of the Risk of Type 2 Diabetes: Evidence From the Western New York Study. *Obesity* **16**, 1370–1376. ISSN: 1930-739X. (2021) (2008).
59. Phillips, A., Gerald Shaper, A. & Whincup, P. Association Between Serum Albumin and Mortality From Cardiovascular Disease, Cancer, and Other Causes. *The Lancet. Originally Published as Volume 2, Issue 8677* **334**, 1434–1436. ISSN: 0140-6736. (2021) (Dec. 1989).
60. Djoussé Luc, Rothman Kenneth J., Cupples L. Adrienne, Levy Daniel & Ellison R. Curtis. Serum Albumin and Risk of Myocardial Infarction and All-Cause Mortality in the Framingham Offspring Study. *Circulation* **106**, 2919–2924. (2021) (Dec. 2002).
61. Hoydonck, P. G. V., Temme, E. H. & Schouten, E. G. Serum Bilirubin Concentration in a Belgian Population: The Association with Smoking Status and Type of Cigarettes. *International Journal of Epidemiology* **30**, 1465–1472. ISSN: 0300-5771. (2021) (Dec. 2001).
62. Schwertner, H. A. Association of Smoking and Low Serum Bilirubin Antioxidant concentrations!The Views Expressed in This Article Are Those of the Author and Do Not Reflect the Official Policy of the Department of Defense or Other Departments of the US Government.1. *Atherosclerosis* **136**, 383–387. ISSN: 0021-9150. (2021) (Feb. 1998).
63. Jo, J., Kimm, H., Yun, J. E., Lee, K. J. & Jee, S. H. Cigarette Smoking and Serum Bilirubin Subtypes in Healthy Korean Men: The Korea Medical Institute Study. *Journal of Preventive Medicine and Public Health* **45**, 105–112. ISSN: 1975-8375. (2021) (Mar. 2012).
64. Kodal, J. B. *et al.* Smoking Reduces Plasma Bilirubin: Observational and Genetic Analyses in the Copenhagen General Population Study. *Nicotine & Tobacco Research* **22**, 104–110. ISSN: 1469-994X. (2021) (Jan. 2020).
65. Patel, C. J. & Ioannidis, J. P. A. Studying the Elusive Environment in Large Scale. *JAMA* **311**, 2173–2174. ISSN: 0098-7484. (2021) (June 2014).
66. Thompson, A. E. *et al.* The Influence of Gender and Other Patient Characteristics on Health Care-Seeking Behaviour: A QUALICOPC Study. *BMC Family Practice* **17**, 1–7. ISSN: 1471-2296. (2022) (Dec. 2016).

67. Chen, S.-L., Lee, W.-L., Liang, T. & Liao, I.-C. Factors Associated with Gender Differences in Medication Adherence: A Longitudinal Study. *Journal of Advanced Nursing* **70**, 2031–2040. ISSN: 1365-2648. (2022) (2014).
68. Manteuffel, M. *et al.* Influence of Patient Sex and Gender on Medication Use, Adherence, and Prescribing Alignment with Guidelines. *Journal of Women's Health* **23**, 112–119. ISSN: 1540-9996. (2022) (Feb. 2014).

Figures

Females



Males

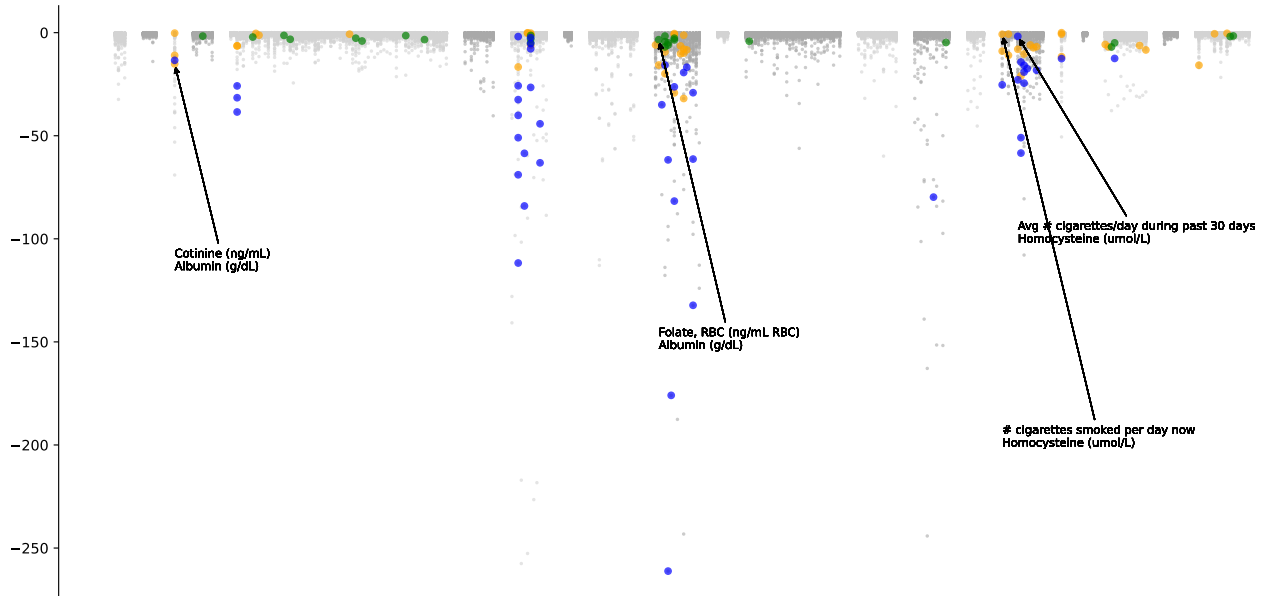


Figure 1: Miami Plot from PheEWS results stratified by sex. The Y axis shows the $-\log$ of the p-value, and the X axis is sorted by exposure categories. Significant sex differences are highlighted in color, and categorized as pure, quantitative, and qualitative sex differences. Relevant exposure-phenotype associations discussed in the manuscript are annotated.

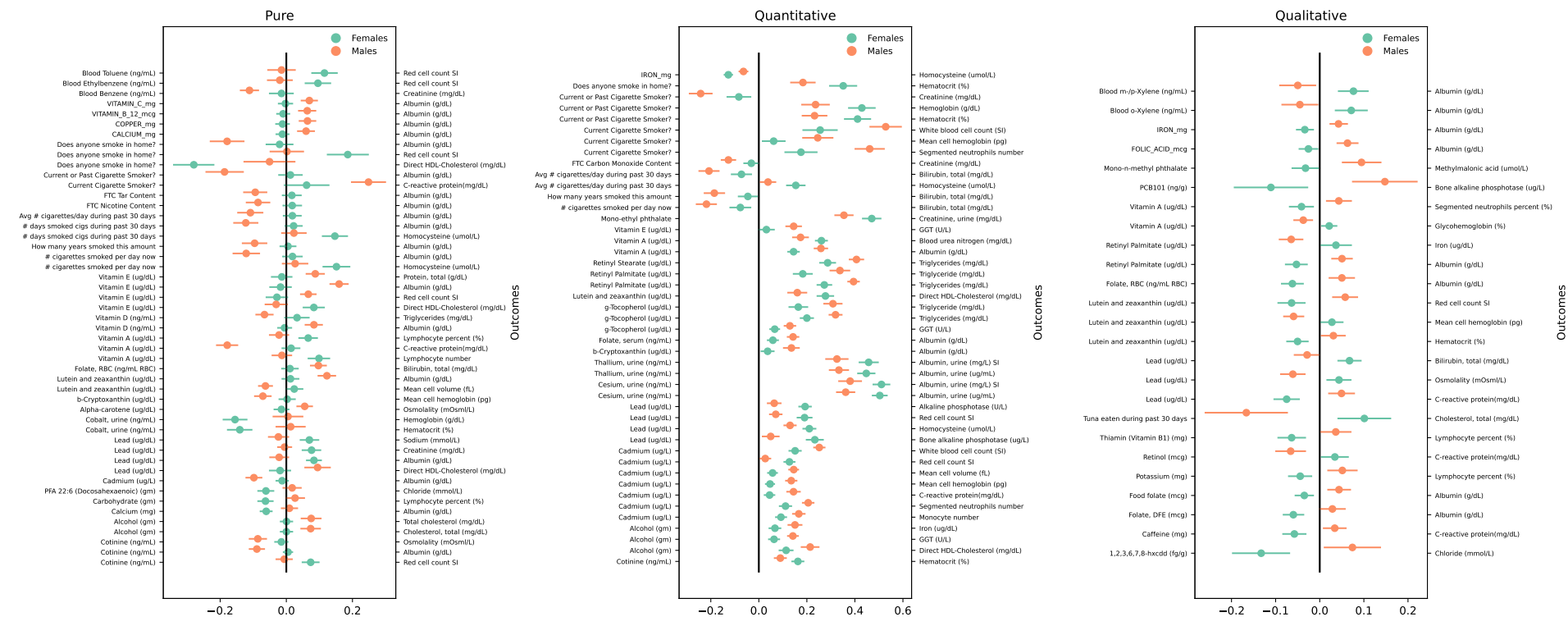


Figure 3: Forest plot showing the effect sizes and 95% confidence intervals in males and females. Exposures are defined on the left side and phenotypes on the right. Plots are divided into the three sex difference categories of pure, quantitative, and qualitative.

Tables

Table 1: Pure sex differences results.

Exposure Name	Outcome Name	β_f	P_f	β_m	P_m	P_{diff}
Cotinine	Albumin	0.005	4.41×10^{-1}	-0.089	6.80×10^{-16}	2.36×10^{-13}
Cotinine	Osmolality	-0.015	1.23×10^{-1}	-0.086	9.96×10^{-12}	8.16×10^{-6}
Cotinine	Red cell count	0.074	4.25×10^{-9}	-0.006	5.96×10^{-1}	3.90×10^{-6}
Alcohol	Cholesterol, total	0.0	9.79×10^{-1}	0.074	3.69×10^{-7}	1.67×10^{-5}
Alcohol	Total cholesterol	0.001	9.48×10^{-1}	0.075	4.29×10^{-7}	1.77×10^{-5}
Calcium	Albumin	-0.061	1.85×10^{-12}	0.009	4.40×10^{-1}	2.36×10^{-6}
Carbohydrate	Lymphocyte percent	-0.063	3.96×10^{-8}	0.026	5.76×10^{-2}	7.06×10^{-7}
PFA 22:6 (Docosaehaenoic)	Chloride	-0.062	1.57×10^{-7}	0.018	1.88×10^{-1}	8.34×10^{-6}
Cadmium	Albumin	-0.013	1.65×10^{-1}	-0.098	2.18×10^{-17}	6.12×10^{-9}
Cobalt, urine	Hematocrit	-0.141	2.29×10^{-14}	0.013	5.47×10^{-1}	7.75×10^{-8}
Cobalt, urine	Hemoglobin	-0.155	2.32×10^{-18}	0.005	8.24×10^{-1}	2.40×10^{-8}
Lead	1/Creatinine	-0.076	3.09×10^{-8}	0.005	6.00×10^{-1}	1.96×10^{-6}
Lead	Albumin	0.083	3.63×10^{-14}	-0.022	1.24×10^{-1}	4.31×10^{-9}
Lead	Direct HDL-Cholesterol	-0.019	2.33×10^{-1}	0.095	7.80×10^{-7}	4.95×10^{-6}
Lead	Sodium	0.07	4.30×10^{-7}	-0.024	1.12×10^{-1}	4.49×10^{-6}
Alpha-carotene	Osmolality	-0.015	1.85×10^{-1}	0.056	1.13×10^{-6}	1.01×10^{-5}
Folate, RBC	Bilirubin, total	0.011	3.69×10^{-1}	0.097	1.96×10^{-16}	4.46×10^{-7}
Lutein and zeaxanthin	Albumin	0.013	3.25×10^{-1}	0.123	1.22×10^{-20}	1.75×10^{-9}
Lutein and zeaxanthin	Mean cell volume	0.024	5.76×10^{-2}	-0.063	4.42×10^{-10}	7.05×10^{-8}
Vitamin A	C-reactive protein	0.014	2.80×10^{-1}	-0.179	6.20×10^{-30}	7.25×10^{-21}
Vitamin A	Lymphocyte number	0.099	8.29×10^{-10}	-0.013	3.53×10^{-1}	2.03×10^{-7}
Vitamin A	Lymphocyte percent	0.066	1.95×10^{-6}	-0.022	1.23×10^{-1}	9.64×10^{-6}
Vitamin D	Albumin	-0.005	5.98×10^{-1}	0.083	6.37×10^{-11}	5.63×10^{-8}
Vitamin D	Triglycerides	0.032	7.73×10^{-2}	-0.066	3.56×10^{-7}	1.16×10^{-5}
Vitamin E	Albumin	-0.018	2.81×10^{-1}	0.16	1.17×10^{-32}	4.02×10^{-17}
Vitamin E	Direct HDL-Cholesterol	0.083	1.77×10^{-7}	-0.031	6.29×10^{-2}	7.05×10^{-7}
Vitamin E	Protein, total	-0.014	3.71×10^{-1}	0.088	1.16×10^{-10}	7.56×10^{-7}
Vitamin E	Red cell count SI	-0.028	7.60×10^{-2}	0.066	1.05×10^{-8}	1.65×10^{-6}
b-Cryptoxanthin	Mean cell hemoglobin	0.003	8.35×10^{-1}	-0.071	4.61×10^{-9}	1.89×10^{-5}
# cigarettes smoked per day now	Albumin	0.019	1.98×10^{-1}	-0.121	1.14×10^{-9}	1.30×10^{-8}
# cigarettes smoked per day now	Homocysteine	0.152	7.87×10^{-14}	0.027	1.69×10^{-1}	7.91×10^{-6}
# days smoked cigs during past 30 days	Albumin	0.022	8.40×10^{-2}	-0.123	9.99×10^{-12}	5.51×10^{-11}
# days smoked cigs during past 30 days	Homocysteine	0.147	7.18×10^{-15}	0.023	2.08×10^{-1}	2.33×10^{-6}
Avg # cigarettes/day during past 30 days	Albumin	0.018	1.69×10^{-1}	-0.109	8.91×10^{-9}	3.98×10^{-8}
Current Cigarette Smoker?	C-reactive protein	0.061	7.15×10^{-2}	0.248	1.05×10^{-21}	1.21×10^{-5}
Current or Past Cigarette Smoker?	Albumin	0.012	4.82×10^{-1}	-0.187	4.29×10^{-11}	2.25×10^{-9}
FTC Nicotine Content	Albumin	0.018	1.99×10^{-1}	-0.086	1.06×10^{-6}	3.72×10^{-6}
FTC Tar Content	Albumin	0.017	2.30×10^{-1}	-0.094	4.60×10^{-8}	5.74×10^{-7}
How many years smoked this amount	Albumin	0.005	6.76×10^{-1}	-0.096	1.54×10^{-7}	3.36×10^{-6}
Does anyone smoke in home?	Albumin	-0.02	3.11×10^{-1}	-0.179	2.25×10^{-12}	1.06×10^{-6}
Does anyone smoke in home?	Direct HDL-Cholesterol	-0.28	5.27×10^{-20}	-0.051	1.89×10^{-1}	3.38×10^{-6}
Does anyone smoke in home?	Red cell count SI	0.186	2.49×10^{-9}	0.002	9.42×10^{-1}	4.52×10^{-6}
Calcium	Albumin	-0.012	2.08×10^{-1}	0.06	1.84×10^{-6}	4.67×10^{-6}
Copper	Albumin	-0.011	2.48×10^{-1}	0.064	1.24×10^{-7}	1.35×10^{-6}
Vitamin B12	Albumin	-0.01	2.99×10^{-1}	0.063	5.69×10^{-7}	3.51×10^{-6}
Vitamin C	Albumin	-0.003	7.83×10^{-1}	0.07	4.09×10^{-9}	4.43×10^{-6}
Blood Benzene	1/Creatinine	0.015	4.13×10^{-1}	0.111	1.60×10^{-16}	1.51×10^{-5}
Blood Ethylbenzene	Red cell count SI	0.096	5.62×10^{-7}	-0.02	2.91×10^{-1}	1.54×10^{-5}
Blood Toluene	Red cell count SI	0.115	1.90×10^{-9}	-0.015	4.80×10^{-1}	3.91×10^{-6}

Weighted generalized linear model results, indicating the environmental exposure (predictor), the phenotype (outcome), and the sex-stratified β and P values, and the P values of the sex difference test. For more detailed information on the results the Supplementary Table 1 shows the complete set of results in each cohort, including the standard errors, and units of measurement.

Table 2: Quantitative sex differences results.

Exposure Name	Outcome Name	β_f	P_f	β_m	P_m	P_{diff}
Cotinine	Hematocrit	0.163	8.81×10^{-41}	0.09	3.81×10^{-14}	1.65×10^{-5}
Alcohol	Direct HDL-Cholesterol	0.114	5.20×10^{-16}	0.214	2.75×10^{-32}	1.18×10^{-5}
Alcohol	GGT	0.063	7.31×10^{-9}	0.142	3.33×10^{-39}	3.46×10^{-7}
Alcohol	Iron	0.067	6.19×10^{-9}	0.151	1.47×10^{-26}	4.65×10^{-6}
Cadmium	C-reactive protein	0.045	1.77×10^{-5}	0.144	1.75×10^{-26}	6.01×10^{-9}
Cadmium	Mean cell hemoglobin	0.047	2.61×10^{-7}	0.135	3.23×10^{-33}	1.01×10^{-9}
Cadmium	Mean cell volume	0.057	4.26×10^{-9}	0.145	1.12×10^{-51}	1.26×10^{-10}
Cadmium	Monocyte number	0.092	4.60×10^{-17}	0.166	7.74×10^{-41}	7.48×10^{-6}
Cadmium	Red cell count SI	0.127	2.53×10^{-30}	0.027	1.22×10^{-2}	6.97×10^{-11}
Cadmium	Segmented neutrophils number	0.111	9.17×10^{-20}	0.205	1.16×10^{-69}	2.20×10^{-8}
Cadmium	White blood cell count (SI)	0.151	9.32×10^{-36}	0.252	1.82×10^{-112}	9.70×10^{-10}
Cesium, urine	Albumin, urineSI	0.511	1.02×10^{-192}	0.38	2.96×10^{-59}	7.22×10^{-6}
Cesium, urine	Albumin, urine	0.504	5.45×10^{-245}	0.362	8.58×10^{-85}	2.74×10^{-9}
Lead	Alkaline phosphatase	0.193	3.84×10^{-53}	0.064	4.85×10^{-6}	7.90×10^{-12}
Lead	Bone alkaline phosphatase	0.233	1.44×10^{-40}	0.049	3.97×10^{-3}	5.87×10^{-14}
Lead	Homocysteine	0.21	4.66×10^{-60}	0.13	2.82×10^{-27}	5.17×10^{-6}
Lead	Red cell count SI	0.19	8.14×10^{-36}	0.07	1.52×10^{-8}	1.10×10^{-9}
Thallium, urine	Albumin, urine SI	0.457	4.26×10^{-121}	0.325	6.08×10^{-45}	1.28×10^{-5}
Thallium, urine	Albumin, urine	0.447	4.64×10^{-144}	0.333	7.42×10^{-64}	1.57×10^{-5}
Folate, serum	Albumin	0.058	5.57×10^{-8}	0.142	1.07×10^{-35}	8.61×10^{-8}
Lutein and zeaxanthin	Direct HDL-Cholesterol (mg/dL)	0.277	5.63×10^{-61}	0.16	1.93×10^{-16}	5.32×10^{-6}
Retinyl Palmitate	Triglyceride	0.183	2.57×10^{-21}	0.338	1.97×10^{-62}	3.10×10^{-8}
Retinyl Palmitate	Triglycerides	0.273	2.11×10^{-75}	0.394	6.51×10^{-262}	7.65×10^{-11}
Retinyl Stearate	Triglycerides	0.287	2.13×10^{-69}	0.407	1.16×10^{-176}	3.17×10^{-8}
Vitamin A	Albumin	0.145	1.50×10^{-37}	0.258	1.98×10^{-82}	8.65×10^{-11}
Vitamin A	Blood urea nitrogen	0.261	1.15×10^{-106}	0.174	5.16×10^{-27}	1.45×10^{-5}
Vitamin E	GGT	0.032	4.80×10^{-2}	0.145	4.63×10^{-20}	4.30×10^{-7}
b-Cryptoxanthin	Albumin	0.037	5.03×10^{-3}	0.136	1.66×10^{-17}	1.52×10^{-6}
g-Tocopherol	GGT	0.066	4.09×10^{-12}	0.129	7.50×10^{-30}	2.08×10^{-5}
g-Tocopherol	Triglyceride	0.164	1.70×10^{-17}	0.309	4.81×10^{-62}	7.29×10^{-8}
g-Tocopherol	Triglycerides	0.2	2.72×10^{-51}	0.319	5.69×10^{-133}	1.35×10^{-10}
Mono-ethyl phthalate	Creatinine, urine	0.47	9.52×10^{-136}	0.354	1.68×10^{-80}	1.31×10^{-5}
# cigarettes smoked per day now	Bilirubin, total	-0.077	3.76×10^{-4}	-0.218	4.96×10^{-26}	2.42×10^{-6}
Avg # cigarettes/day during past 30 days	Bilirubin, total	-0.072	4.99×10^{-4}	-0.208	1.32×10^{-23}	4.13×10^{-6}
Avg # cigarettes/day during past 30 days	Homocysteine	0.154	4.58×10^{-16}	0.038	1.57×10^{-2}	2.71×10^{-6}
Current Cigarette Smoker?	Mean cell hemoglobin	0.062	8.80×10^{-3}	0.245	5.67×10^{-15}	2.93×10^{-6}
Current Cigarette Smoker?	Segmented neutrophils number	0.175	1.37×10^{-7}	0.462	1.17×10^{-51}	2.36×10^{-10}
Current Cigarette Smoker?	White blood cell count (SI)	0.255	9.13×10^{-13}	0.528	3.90×10^{-59}	1.68×10^{-8}
Current or Past Cigarette Smoker?	1/Creatinine	0.083	6.93×10^{-4}	0.242	3.26×10^{-25}	2.72×10^{-6}
Current or Past Cigarette Smoker?	Hematocrit	0.411	1.68×10^{-52}	0.232	7.51×10^{-20}	1.36×10^{-6}
Current or Past Cigarette Smoker?	Hemoglobin	0.429	3.86×10^{-54}	0.236	1.42×10^{-16}	1.25×10^{-6}
FTC Carbon Monoxide Content	1/Creatinine	0.031	4.14×10^{-2}	0.127	3.37×10^{-18}	6.00×10^{-6}
How many years smoked this amount	Bilirubin, total	-0.046	2.08×10^{-2}	-0.185	6.37×10^{-19}	1.18×10^{-6}
Does anyone smoke in home?	Hematocrit	0.351	4.12×10^{-36}	0.184	2.62×10^{-13}	8.24×10^{-6}
Iron	Homocysteine	-0.127	8.23×10^{-54}	-0.064	3.17×10^{-13}	1.78×10^{-7}

Weighted generalized linear model results, indicating the environmental exposure (predictor), the phenotype (outcome), and the sex-stratified β and P values, and the P values of the sex difference test. For more detailed information on the results the Supplementary Table 1 shows the complete set of results in each cohort, including the standard errors, and units of measurement.

Table 3: Qualitative sex differences results.

Exposure Name	Outcome Name	β_f	P_f	β_m	P_m	P_{diff}
1,2,3,6,7,8-hxcdd	Chloride	-0.133	4.55×10^{-5}	0.074	2.18×10^{-2}	6.46×10^{-6}
Caffeine	C-reactive protein	-0.058	9.03×10^{-6}	0.034	7.33×10^{-3}	4.45×10^{-7}
Folate, DFE	Albumin	-0.06	1.94×10^{-7}	0.029	4.70×10^{-2}	1.62×10^{-6}
Food folate	Albumin	-0.035	4.56×10^{-4}	0.044	6.20×10^{-4}	1.20×10^{-6}
Potassium	Lymphocyte per-cent	-0.044	6.40×10^{-4}	0.051	2.12×10^{-3}	6.13×10^{-6}
Retinol	C-reactive protein	0.034	2.42×10^{-2}	-0.066	9.09×10^{-5}	1.01×10^{-5}
Thiamin (Vitamin B1)	Lymphocyte per-cent	-0.064	3.94×10^{-5}	0.036	3.74×10^{-2}	1.82×10^{-5}
Tuna eaten during past 30 days	Cholesterol, total	0.101	8.21×10^{-4}	-0.166	4.31×10^{-4}	1.86×10^{-6}
Lead	Bilirubin, total	0.067	2.12×10^{-7}	-0.029	4.37×10^{-2}	6.11×10^{-7}
Lead	C-reactive protein	-0.075	1.03×10^{-7}	0.049	7.03×10^{-4}	8.48×10^{-10}
Lead	Osmolality	0.044	1.16×10^{-3}	-0.061	1.08×10^{-5}	5.90×10^{-8}
Folate, RBC	Albumin	-0.062	3.23×10^{-7}	0.05	3.84×10^{-4}	1.73×10^{-9}
Lutein and zeaxanthin	Hematocrit	-0.05	2.30×10^{-5}	0.031	2.20×10^{-2}	6.28×10^{-6}
Lutein and zeaxanthin	Mean cell hemoglobin	0.028	2.71×10^{-2}	-0.059	1.18×10^{-7}	2.21×10^{-7}
Lutein and zeaxanthin	Red cell count SI	-0.064	2.79×10^{-5}	0.058	4.04×10^{-5}	4.53×10^{-9}
Retinyl Palmitate	Albumin	-0.053	2.07×10^{-5}	0.05	1.22×10^{-5}	1.08×10^{-9}
Retinyl Palmitate	Iron	0.037	3.43×10^{-2}	-0.065	9.22×10^{-7}	3.24×10^{-6}
Vitamin A	Glycohemoglobin	0.021	9.55×10^{-3}	-0.037	3.06×10^{-4}	9.10×10^{-6}
Vitamin A	Segmented neutrophils percent	-0.041	2.19×10^{-3}	0.043	2.06×10^{-3}	1.40×10^{-5}
PCB101	Bone alkaline phosphatase	-0.111	8.63×10^{-3}	0.148	6.76×10^{-5}	4.16×10^{-6}
Mono-n-methyl phthalate	Methylmalonic acid	-0.032	3.23×10^{-2}	0.095	1.67×10^{-5}	1.93×10^{-6}
Folic Acid	Albumin	-0.026	1.28×10^{-2}	0.063	1.14×10^{-7}	1.75×10^{-8}
Iron	Albumin	-0.034	3.24×10^{-4}	0.043	1.13×10^{-5}	1.57×10^{-8}
Blood m-/p-Xylene	Albumin	0.076	5.52×10^{-6}	-0.05	1.41×10^{-2}	1.68×10^{-6}
Blood o-Xylene	Albumin	0.071	8.43×10^{-5}	-0.045	2.58×10^{-2}	1.81×10^{-5}

Weighted generalized linear model results, indicating the environmental exposure (predictor), the phenotype (outcome), and the sex-stratified β and P values, and the P values of the sex difference test. For more detailed information on the results the Supplementary Table 1 shows the complete set of results in each cohort, including the standard errors, and units of measurement.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DictionarySupplementaryTable1.txt](#)
- [SupplementaryTable1.csv](#)