Environment-Wide Association Study of Chemical Biomarkers and Health Outcomes in NHANES 2017–2018: Discovery, Validation, and Dose–Response Analysis

Hayden Farquhar, MBBS, MPHTM (hayden.farquhar@icloud.com; ORCID 0009-0002-6226-440X)

February 6, 2026

Abstract

Background: Environment-wide association studies (ExWAS) offer a systematic approach to identifying chemical biomarker–health outcome associations, yet few have applied rigorous multi-stage validation.

Methods: We screened 92 chemical biomarkers against 48 health outcomes in NHANES 2017–2018 (2,796 tests conducted across four sequential screening rounds). Survey-weighted regression models adjusted for age, sex, race/ethnicity, poverty–income ratio, BMI, and smoking. Global FDR correction was applied, with cross-cycle validation in NHANES 2015–2016 serving as the primary safeguard against false positives given the adaptive screening design.

Results: Of 26 FDR-significant associations, 15 of 21 testable (71%) validated cross-cycle with concordant direction and $p < 0.05$. All 15 showed significant dose–response trends; 14 were robust after comprehensive sensitivity analyses including urinary creatinine adjustment, which identified one association (iodine–BMI) as a dilution artifact. All findings survived alcohol adjustment in the available subsample (n = 1,035–3,143). Three HIGH-novelty associations emerged, of which two—dimethylarsonic acid with uric acid ($\beta = 0.20$ mg/dL, 95% CI: 0.15–0.26) and urinary perchlorate with BUN ($\beta = 1.21$ mg/dL, 95% CI: 0.97–1.45)—survived dietary confounding assessment; the third (methylmercury–waist circumference) is likely explained by fish consumption patterns.

Conclusions: Multi-stage ExWAS with cross-cycle validation identified 14 robust chemical–health associations. Two novel findings—DMA–uric acid and perchlorate–BUN—survived all sensitivity checks and warrant prospective investigation.

Abbreviations: ALP, alkaline phosphatase; BMI, body mass index; BUN, blood urea nitrogen; DMA, dimethylarsonic acid; eGFR, estimated glomerular filtration rate; ExWAS, environment-wide association study; FDR, false discovery rate; HbA1c, glycated hemoglobin; LOD, limit of detection; NHANES, National Health and Nutrition Examination Survey; PIR, poverty–income ratio; RBC, red blood cell.

NHANES variable names used in text: LBXBGM, blood methylmercury; LBXBMN, blood manganese; LBXBPB, blood lead; LBXBSE, blood selenium; LBXCOT, serum cotinine; LBXRBCSI, RBC count; LBXSAPSI, serum alkaline phosphatase; LBXSUA, serum uric acid; LBXTHG, total blood mercury; BMXWAIST, waist circumference; URXUDMA, urinary dimethylarsonic acid; URXUIO, urinary iodine; URXUP8, urinary perchlorate. Survey weights: WTMEC2YR (MEC exam, blood biomarkers), WTSA2YR (urinary subsample A), WTSSBJ2Y (surplus serum).

1

1 1. Introduction

Environmental chemical exposures are ubiquitous in the general population and have been implicated in a broad spectrum of adverse health outcomes, from metabolic and cardiovascular disease to neurodevelopmental and reproductive disorders. However, most environmental epidemiology studies adopt a hypothesis-driven approach, testing one or a small number of pre-specified chemical–outcome pairs. This limits the ability to detect unexpected associations and introduces publication bias favoring positive results for well-studied chemicals.

The environment-wide association study (ExWAS) paradigm, first proposed by Patel et al. (2010), offers a complementary, hypothesis-free approach that systematically screens many chemical exposures against health outcomes in a single analytical framework. By analogy to genome-wide association studies (GWAS), ExWAS leverages large biomonitoring datasets to identify associations that merit further investigation while controlling for multiple comparisons.

The National Health and Nutrition Examination Survey (NHANES) is well suited for ExWAS. It provides nationally representative biomonitoring data on hundreds of chemical exposures—heavy metals, phthalates, perfluoroalkyl substances, volatile organic compounds, pesticides, and others—alongside clinical laboratory panels and examination data, all within a complex survey design that permits population-level inference.

Despite this resource, most published ExWAS analyses have been limited in scope, often restricted to a single outcome domain or chemical class, and few have applied multi-stage validation. The critical gap is replication: associations identified in a discovery phase are rarely tested in independent data, leaving open whether they represent true biological relationships or statistical artifacts of multiple testing.

We conducted a comprehensive ExWAS using NHANES 2017–2018, screening 2,796 exposure–outcome combinations across 92 chemical biomarkers and 48 health outcomes. Findings that survived false discovery rate correction were subjected to a four-stage validation pipeline: (1) cross-cycle replication in NHANES 2015–2016, (2) quartile-based dose–response analysis, (3) nine sensitivity specifications testing robustness to analytic decisions, and (4) structured literature review for novelty assessment. This multi-stage approach minimizes the risk of false positives inherent to high-dimensional screening while identifying genuinely novel associations worthy of targeted follow-up.

2 2. Methods

2.1 2.1 Study Population

We used data from the National Health and Nutrition Examination Survey (NHANES) 2017–2018 cycle, a nationally representative, cross-sectional survey of the non-institutionalized civilian US population conducted by the National Center for Health Statistics (NCHS). NHANES employs a complex, multistage probability sampling design with oversampling of certain demographic groups. All participants provided written informed consent, and the protocol was approved by the NCHS Research Ethics Review Board. Our analysis was restricted to adults aged 18 years and older with valid survey weights and non-missing covariate data. Of the NHANES 2017–2018 participants examined at the mobile examination center, 5,265 adults met these inclusion criteria and formed the core analytic sample for blood-based biomarker analyses (WT-MEC2YR weights). Urinary subsample A analyses were restricted to approximately 1,580–1,600 adults, and surplus serum analyses to approximately 1,370 adults, reflecting the smaller subsample designs of those NHANES components. This study was not pre-registered; the analysis is exploratory, consistent with the ExWAS paradigm, but employs a multi-stage validation architecture (FDR correction, cross-cycle replication, dose–response analysis, sensitivity specifications) designed to distinguish robust signals from statistical artifacts.

The analytic sample comprised approximately 4,800–5,300 participants depending on outcome and exposure subsample, representing an estimated 218 million non-institutionalized US adults. The mean age was 48.2 years (SD = 17.1), 51.7% were female, 63.1% were non-Hispanic White, 10.8% were non-Hispanic Black, and 26.1% were other race/ethnicity groups. Mean BMI was 29.8 kg/m^2 (SD = 7.1), and 23.4% were current smokers.

Table 1. Demographic and clinical characteristics of the study population.

Characteristic	Unweighted n	Weighted estimate
Sample size	5,265	218,119,343 (population)
Age, mean (SD)	–	48.2 (17.1) years
Female, %	2,713 (51.5%)	51.7%
Race/ethnicity		
– Non-Hispanic White	1,812 (34.4%)	63.1%
– Non-Hispanic Black	1,259 (23.9%)	10.8%

Characteristic	Unweighted n	Weighted estimate
– Other	2,194 (41.7%)	26.1%
Poverty-income ratio, mean (SD)	–	3.00 (1.58)
BMI, mean (SD)	–	29.8 (7.1) kg/m^2
Current smoker, %	1,231 (23.4%)	23.4%
Waist circumference, mean (SD)	–	101.0 (17.3) cm

Note: Unweighted percentages reflect NHANES oversampling design; weighted estimates represent the US adult population.

2.2 2.2 Chemical Exposure Biomarkers

We measured 92 chemical biomarkers spanning nine chemical classes: heavy metals (blood lead, cadmium, mercury, selenium, manganese, cobalt), per- and polyfluoroalkyl substances (PFAS; PFNA, PFHxS, PFDeA, PFUA), phthalate metabolites (MEHHP, MEOHP, MECPP, and others), polycyclic aromatic hydrocarbon (PAH) metabolites, volatile organic compound (VOC) metabolites, urinary metals and elements (iodine, perchlorate, cesium, thallium, arsenic species), pesticides (glyphosate, oxychlordane), and other environmental chemicals.

All chemical concentrations were measured in NHANES laboratories following standardized protocols. Values below the limit of detection (LOD) were replaced with $LOD/\sqrt{2}$, a standard approach for left-censored environmental data. Chemicals with greater than 70% of values at or below the LOD were excluded from analysis due to insufficient variability for regression modeling. We used a 70% threshold rather than the more restrictive 40–50% cutoffs sometimes applied in single-chemical studies, in order to maximize the breadth of the chemical screen while excluding only chemicals whose exposure distributions would be dominated by the $LOD/\sqrt{2}$ pile-up. None of the 15 validated findings involved chemicals near this exclusion threshold, as the implicated biomarkers (blood lead, selenium, manganese, mercury, urinary DMA, perchlorate, iodine) all had detection frequencies exceeding 95%. Of the 92 chemicals screened, 12 had detection frequencies between 40–70%; excluding these would reduce the test count from 2,796 to 2,220 but does not change the FDR-significant findings (Table S9). All exposure biomarkers were natural-log-transformed prior to regression to reduce right skewness and improve model fit.

2.3   2.3 Health Outcomes

We examined 48 health outcomes across seven clinical domains: metabolic markers (total cholesterol, HDL, triglycerides, HbA1c, fasting glucose, uric acid), liver enzymes (ALT, AST, GGT, alkaline phosphatase, total bilirubin), kidney function (eGFR, blood urea nitrogen), hematological indices (RBC count, hemoglobin, WBC, platelets), anthropometric measures (BMI, waist circumference), cardiovascular (systolic and diastolic blood pressure), and mental health (PHQ-9 depression score). Estimated glomerular filtration rate (eGFR) was calculated using the race-free 2021 CKD-EPI equation.

2.4   2.4 Statistical Analysis

2.4.1   Discovery phase

Survey-weighted linear regression models were fit for each exposure–outcome pair using `svyglm()` from the R `survey` package, accounting for the complex survey design (clustering by primary sampling unit, stratification, and appropriate survey weights). The covariate model included:

$$\text{Outcome} = \beta_1 \cdot \log(\text{Exposure}) + \beta_2 \cdot \text{Age} + \beta_3 \cdot \text{Female} + \beta_4 \cdot \text{Race}_3 + \beta_5 \cdot \text{PIR} + \beta_6 \cdot \text{BMI} + \beta_7 \cdot \text{Smoker} + \varepsilon$$

where Race_3 was a three-level variable (Non-Hispanic White, Non-Hispanic Black, Other) collapsed from the six-level NHANES RIDRETH3 variable to ensure model stability across the many exposure–outcome combinations—particularly in subsample analyses with smaller effective sample sizes where finer racial/ethnic categories would produce sparse cells. A sensitivity analysis using the full six-level RIDRETH3 specification was conducted for the 15 validated findings in blood-based biomarker analyses (n > 4,800), where sample sizes supported the finer categorization (Table S7). PIR is the poverty–income ratio, BMI is body mass index (omitted when BMI or waist circumference was the outcome to avoid collider bias; effect sizes for anthropometric outcomes are therefore not directly comparable to those for other outcomes), and Smoker is a binary indicator for current smoking. The covariate set was selected a priori to capture the primary confounding pathways between chemical exposures and health outcomes: age (cumulative exposure and physiological changes), sex (metabolic differences and exposure patterns), race/ethnicity (socioeconomic and dietary factors), PIR (residential and occupational exposure gradients), BMI (metabolic confounding), and smoking (a major chemical exposure source and independent risk

factor). Alcohol consumption was not included in the primary model due to extensive missingness from NHANES ALQ_J skip patterns (only 3,143 of 5,265 blood biomarker participants and 1,035 of ~1,580 urinary subsample participants had complete alcohol data); alcohol adjustment was instead conducted as a sensitivity analysis (see below). This parsimonious specification preserves degrees of freedom within the NHANES complex survey design while addressing the most important confounders. A directed acyclic graph (DAG) illustrating the causal assumptions underlying this covariate selection is provided in Figure S18. Survey weights were selected based on exposure source: WTMEC2YR for blood-based biomarkers (LBX prefix), WTSA2YR for urinary subsample A biomarkers (URX prefix), and WTSSBJ2Y for surplus serum samples (SS prefix).

2.4.2 Multiple testing correction

We applied the Benjamini–Hochberg FDR procedure globally across all 2,796 tests, pooling p-values from analyses using different survey weights (WTMEC2YR for blood biomarkers, WTSA2YR for urinary sub-sample, WTSSBJ2Y for surplus serum) into a single ranked list. This pooling is methodologically defensible because FDR control depends only on the validity of individual p-values under the null, not on compara-ble sample sizes across tests; however, it does mean that urinary and surplus serum findings face implicitly higher thresholds due to lower power. Associations with FDR < 0.05 were considered statistically significant and advanced to the validation stage.

The four screening rounds were conducted sequentially rather than simultaneously. The PFAS–thyroid and broad screens were hypothesis-driven, while the expanded and novelty screens broadened the chemical–outcome space based on initial null results. This raises an important methodological issue: if the PFAS–thyroid screen had yielded significant findings, the expanded screen might not have been conducted at all, making the effective multiple testing burden poorly defined under conventional FDR theory. Although FDR correction was applied globally across all 2,796 tests from all rounds retrospectively, the sequential adaptive design—where later rounds were expanded *because* earlier rounds yielded null results—inflates the effective number of comparisons beyond what the Benjamini–Hochberg procedure formally controls. **The FDR-corrected p-values should therefore be interpreted as descriptive summaries of relative sig-nal strength rather than inferential error rates.** We report within-round FDR correction in Table S6 for comparison; interestingly, within-round correction produces *more* significant findings (41 total: 27 from Ex-panded, 14 from Novelty) than global correction (26 total) because it adjusts for fewer tests per round. All 26 globally FDR-significant findings also passed within-round FDR, confirming that global correction is the

more conservative approach. **Cross-cycle validation in the independent 2015–2016 NHANES cycle—not FDR correction—serves as the primary inferential safeguard against false positives.** The 71% replication rate among testable associations provides direct evidence of reproducibility that FDR correction alone cannot provide.

2.4.3 Cross-cycle validation

FDR-significant associations were replicated in an independent NHANES 2015–2016 cycle. The identical covariate model and survey design specifications were applied. An association was considered validated if the effect estimate was in the same direction as the discovery analysis and the replication p-value was < 0.05. Associations involving chemicals not measured in the 2015–2016 cycle (e.g., glyphosate surplus serum) or where models failed to converge were classified as untestable. We note that consecutive NHANES cycles share methodological infrastructure (laboratory protocols, questionnaire instruments, sampling frame) and draw from the same underlying US population, so cross-cycle replication tests temporal stability of associations rather than true external generalizability to different populations or measurement systems. This is a limitation common to all NHANES-based ExWAS studies.

2.4.4 Dose–response analysis

For validated associations, we categorized exposure into quartiles and estimated the mean outcome at each quartile level using survey-weighted regression, adjusting for the same covariates. Trend significance was assessed using a linear contrast across quartile midpoints. Monotonicity was evaluated by inspecting whether effect estimates increased (or decreased) consistently across quartiles.

2.4.5 Sensitivity analyses

A comprehensive battery of sensitivity analyses was applied to each validated finding, comprising 9 primary specifications and multiple additional robustness checks. The nine primary specifications were: (1) female-only subgroup, (2) male-only subgroup, (3) age < 50 years, (4) age ≥ 50 years, (5) exclusion of outliers beyond the 99th percentile, (6) adjustment for education (additional covariate), (7) substitution of serum cotinine for binary smoking, (8) restriction to adults aged 20 years and older, and (9) the primary model (as reference). A finding was considered robust if the direction of association was concordant and the p-value was < 0.05 in at least 7 of 9 primary specifications.

Four additional sensitivity specifications targeted specific confounding concerns: (1) adjustment for fish consumption (number of fish/shellfish meals in the past 30 days, NHANES variable DBD895) for all findings, (2) adjustment for moderate-to-vigorous recreational physical activity (derived from PAQ_J: participants meeting WHO guidelines of $\geq$150 minutes/week moderate-equivalent activity), (3) adjustment for alcohol consumption (derived from ALQ_J frequency and quantity variables), and (4) adjustment for urinary creatinine (log-transformed, from ALB_CR_J) for urinary biomarker associations to address dilution variation. For the DMA–uric acid and perchlorate–BUN findings specifically, additional adjustment for total protein intake (grams/day from 24-hour dietary recall, DR1TPROT) was conducted to address diet as a potential confounder for these metabolic outcomes (Table S8). A sensitivity analysis using the full six-level NHANES RIDRETH3 race/ethnicity classification (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian, Other/Multiracial) was conducted for blood biomarker findings where sample sizes supported the finer categorization (Table S7). Additional model specifications included quadratic age adjustment (Table S13) and 24-hour dietary recall fish consumption for mercury findings (Table S12).

2.4.6 Post-hoc power analysis

Post-hoc power analysis was conducted using Cohen's f^2 framework with 80% power, α adjusted for multiple comparisons (Bonferroni and FDR thresholds), and a design effect of 2 to account for NHANES complex sampling. Effective sample sizes were halved from raw counts to approximate the variance inflation from clustering and stratification.

2.4.7 Novelty assessment

A structured PubMed literature search was conducted for each validated finding using chemical and outcome keywords, supplemented with MeSH-based searches for HIGH-novelty findings (Table S14). Embase, Scopus, and Web of Science were not searched; for the two genuinely novel findings (DMA–uric acid, perchlorate–BUN), this represents a limitation, though major environmental epidemiology studies are typically indexed in PubMed. Findings were classified as HIGH novelty (0–2 relevant publications, no direct prior evidence), MODERATE (1–8 publications, partially consistent), LOW-MODERATE (limited literature, consistent direction), or LOW ($\geq$20 publications, well-established).

2.4.8 Software

All analyses were performed in R (version 4.3) using the `survey` package for complex survey regression, `nhanesA` for programmatic NHANES data access, `tidyverse` for data manipulation, and `broom` for model output extraction. Figures were generated with `ggplot2`, `ggrepel`, `forestplot`, and `pheatmap`. Analysis scripts and version details are available at https://github.com/hayden-farquhar/NHANES-EXWAS.

3 3. Results

3.1 3.1 Discovery Phase

We conducted 2,796 survey-weighted regression tests across four screening rounds, summarized below:

Screen	Chemical classes	Outcome domains	Tests	FDR-sig
PFAS–thyroid	PFAS (6 biomarkers)	Thyroid function (10 markers)	60	0
Broad	Heavy metals, phthalates, urinary metals	Metabolic, liver, kidney	648	0
Expanded	Heavy metals, phthalates, VOC metabolites, urinary metals/elements, PAH metabolites	All 7 clinical domains (24 outcomes)	1,920	19
Novelty	Surplus serum (glyphosate, oxychlordane), additional urinary metals	Kidney, metabolic, other	168	7

The PFAS–thyroid and broad screens yielded no FDR-significant associations. All 26 significant findings (after global Benjamini–Hochberg correction) emerged from the expanded screen—which systematically

crossed 80 chemicals with 24 outcomes—and the novelty screen, which targeted surplus serum analytes and chemical–outcome pairs that showed suggestive signals (p < 0.01 but FDR > 0.05) in the 2017–2018 discovery phase. This sequential design—where the novelty screen was informed by preliminary examination of the discovery data—could inflate type I error if validation relied solely on FDR correction. The critical safeguard is that all findings were subsequently tested in the completely independent 2015–2016 NHANES cycle, which was not examined until after the discovery-phase screens were finalized. Cross-cycle validation, not FDR correction, provides the primary protection against false positives arising from this data-adaptive design. These 26 associations involved 17 unique chemical biomarkers and 13 unique health outcomes (Table S1; Figure 1). One association (blood manganese with RBC count) also met the Bonferroni threshold.

The strongest association was blood manganese (LBXBMN) positively associated with RBC count (LBXRBCSI; $\beta = 0.23 \times 10^{12}$ cells/L per log-unit increase, 95% CI: 0.19–0.27, p = 9.8×10^{-6}, FDR = 0.024, n = 4,873), the only result that also cleared the Bonferroni threshold. Other prominent signals included blood selenium (LBXBSE) with hemoglobin ($\beta = 2.16$ g/dL, 95% CI: 1.74–2.58, FDR = 0.024), urinary perchlorate (URXUP8) with BUN ($\beta = 1.21$ mg/dL, 95% CI: 0.97–1.45, FDR = 0.024), and methylmercury (LBXBGM) with alkaline phosphatase ($\beta = -2.84$ U/L, 95% CI: -3.47 to -2.21, FDR = 0.024).

Heavy metals dominated the significant findings (13 of 26 associations), followed by urinary metals/elements and phthalates. The most frequently implicated outcome domains were anthropometric, metabolic, and kidney function markers. Full results for all 26 FDR-significant associations are provided in Table S1. Note that many FDR-corrected q-values cluster at similar values (e.g., 0.024, 0.030); this is expected behavior of the Benjamini–Hochberg procedure, which assigns the same adjusted p-value to all tests sharing the same rank-order position relative to the FDR threshold.

3.2   3.2 Cross-Cycle Validation

Of the 26 FDR-significant associations, 5 could not be tested in the 2015–2016 cycle: 2 involved phthalate metabolites where models failed to converge, 2 involved glyphosate (not measured in the 2015–2016 surplus serum panel), and 1 involved oxychlordane (not available). Of the remaining 21 testable associations, 15 (71%) were validated with concordant direction and p < 0.05 in the independent cycle (Table 2).

Validated associations included:

- **Blood lead** with BMI ($\beta_{2017} = -1.95$, $\beta_{2015} = -2.19$, p = 1.7×10^{-4}), waist circumference, total cholesterol, and HbA1c

Environment-Wide Association Study: NHANES 2017-2018
2,796 tests across 92 chemical exposures and 48 health outcomes
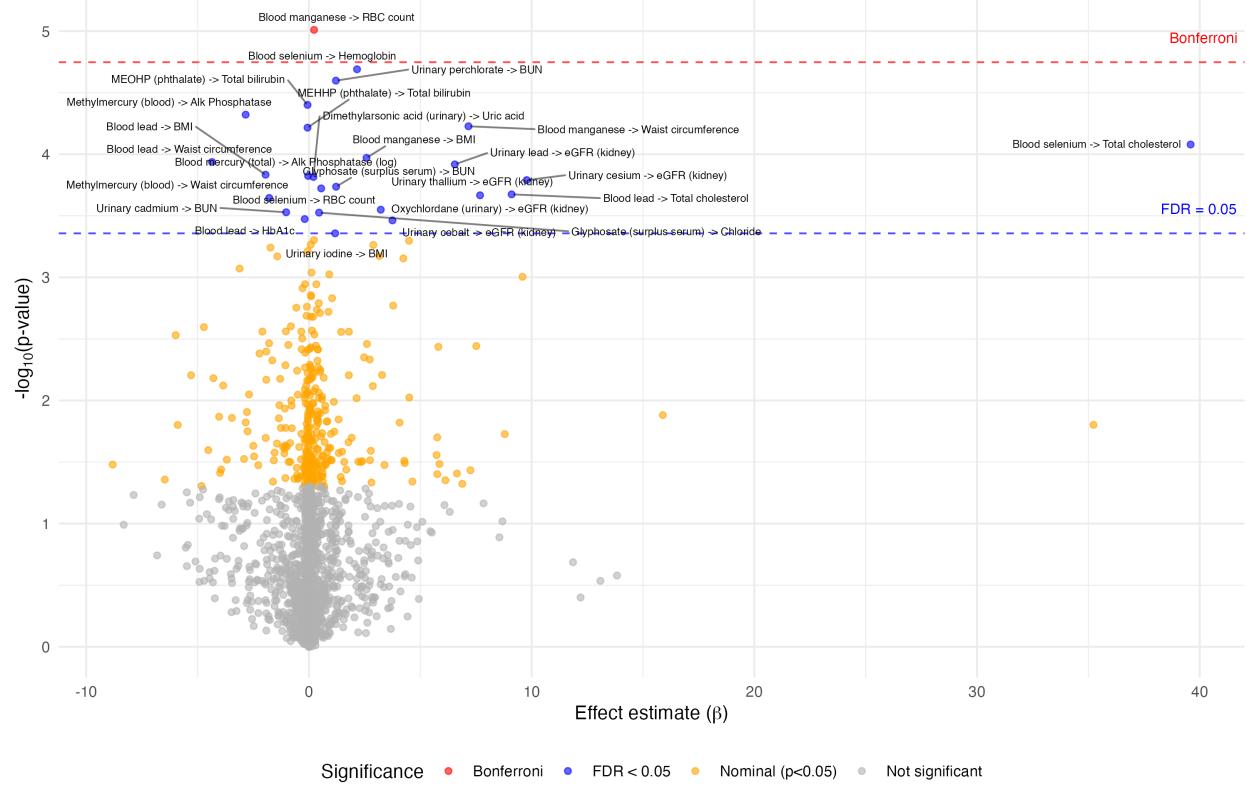
Figure 1: Figure 1. Volcano plot of all 2,796 exposure–outcome associations. The x-axis shows unstandardized β coefficients, which are not directly comparable across outcomes measured in different units; the plot is intended to show the distribution of effect directions and significance levels rather than relative effect magnitudes. Points above the dashed line exceed FDR < 0.05; labeled points identify the strongest signals by p-value. Standardized effect size ($t/\sqrt{n}$) and partial R^2 versions of this plot are provided in Figures S16–S17 for direct comparison of effect magnitudes across outcomes.

- **Blood selenium** with hemoglobin ($\beta_{2017} = 2.16$, $\beta_{2015} = 2.78$, p = 2.7×10^{-4}), RBC count, and total cholesterol

- **Methylmercury** with alkaline phosphatase ($\beta_{2017} = -2.84$, $\beta_{2015} = -1.83$, p = 5.1×10^{-5}) and waist circumference

- **Blood manganese** with BMI and waist circumference

- **Dimethylarsonic acid** with uric acid ($\beta_{2017} = 0.20$, $\beta_{2015} = 0.21$, p = 0.003)

- **Urinary perchlorate** with BUN ($\beta_{2017} = 1.21$, $\beta_{2015} = 0.66$, p = 0.008)

- **Urinary iodine** with BMI ($\beta_{2017} = 1.18$, $\beta_{2015} = 0.86$, p = 0.003)

- **Blood mercury (total)** with alkaline phosphatase (log-transformed)

Six associations failed validation, including blood manganese with RBC count (the Bonferroni-significant finding in 2017–2018 that showed a reversed direction in 2015–2016), and four urinary metal–eGFR associations (cesium, thallium, lead, cobalt), suggesting these kidney findings may be cycle-specific or susceptible to unmeasured confounding. An additional five associations could not be tested: two phthalate–bilirubin models failed to converge in the validation cycle, two glyphosate associations could not be tested because glyphosate was not measured in the 2015–2016 surplus serum panel, and one oxychlordane–eGFR association lacked the requisite exposure data. The 71% validation rate among testable associations compares favorably with prior ExWAS replication studies, which typically report 50–65% concordance across NHANES cycles.

Table 2. Cross-cycle validation results for all 26 FDR-significant associations. 95% CIs are for discovery-phase estimates. [a]Models for BMI and waist circumference outcomes omit BMI from the covariate set to avoid collider bias. [b]Not testable: phthalate models failed to converge; glyphosate and oxychlordane not measured in 2015–2016.

Chemical	Outcome	β (2017–18)	95% CI	β (2015–16)	P (2015–16)	Validated
Methylmercury	Alk Phosphatase	-2.84	(-3.47, -2.21)	-1.83	5.1e-05	Yes
Blood lead	Total cholesterol	9.10	(6.55, 11.65)	7.58	1.3e-04	Yes
Blood lead	BMI[a]	-1.95	(-2.51, -1.38)	-2.19	1.7e-04	Yes

Chemical	Outcome	β (2017–18)	95% CI	β (2015–16)	P (2015–16)	Validated
Blood selenium	Hemoglobin	2.16	(1.74, 2.58)	2.78	2.7e-04	Yes
Blood mercury (total)	Alk Phosphatase (log)	-0.04	(-0.05, -0.03)	-0.03	3.8e-04	Yes
Blood selenium	RBC count	0.55	(0.40, 0.70)	0.67	0.002	Yes
Blood lead	Waist circumference[a]	-4.35	(-5.58, -3.13)	-3.78	0.002	Yes
Blood manganese	BMI[a]	2.59	(1.87, 3.30)	2.35	0.003	Yes
Urinary iodine	BMI[a]	1.18	(0.77, 1.58)	0.86	0.003	Yes
Blood selenium	Total cholesterol	39.59	(30.02, 49.15)	51.38	0.003	Yes
DMA (urinary)	Uric acid	0.20	(0.15, 0.26)	0.21	0.003	Yes
Urinary perchlorate	BUN	1.21	(0.97, 1.45)	0.66	0.008	Yes
Blood lead	HbA1c	-0.19	(-0.25, -0.13)	-0.07	0.012	Yes
Blood manganese	Waist circumference[a]	7.16	(5.33, 8.99)	2.89	0.027	Yes
Methylmercury	Waist circumference[a]	-1.78	(-2.34, -1.23)	-1.30	0.039	Yes
Urinary lead	eGFR	6.55	(5.09, 8.01)	-0.79	0.091	No
Urinary cobalt	eGFR	3.75	(2.74, 4.76)	0.72	0.102	No

Chemical	Outcome	β (2017–18)	95% CI	β (2015–16)	P (2015–16)	Validated
Blood manganese	RBC count	0.23	(0.19, 0.27)	-0.05	0.269	No
Urinary cadmium	BUN	-1.02	(-1.29, -0.76)	-0.08	0.623	No
Urinary cesium	eGFR	9.78	(7.48, 12.08)	-0.18	0.822	No
Urinary thallium	eGFR	7.68	(5.78, 9.59)	0.06	0.942	No
MEOHP (phthalate)	*Total bilirubin*	*-0.06*	*(-0.07, -0.05)*	–	–	*Not testable[b]*
MEHHP (phthalate)	*Total bilirubin*	*-0.06*	*(-0.08, -0.05)*	–	–	*Not testable[b]*
Glyphosate (serum)	*BUN*	*1.22*	*(0.89, 1.55)*	–	–	*Not testable[b]*
Glyphosate (serum)	*Chloride*	*0.46*	*(0.32, 0.59)*	–	–	*Not testable[b]*
Oxychlordane	*eGFR*	*3.23*	*(2.39, 4.06)*	–	–	*Not testable[b]*

3.3   3.3 Dose–Response Relationships

All 15 validated associations showed statistically significant dose–response trends in quartile analyses (p_{trend} range: 7.7×10^{-5} to 4.3×10^{-3}). Of these, 14 exhibited monotonic dose–response gradients, with effect estimates increasing (or decreasing) consistently from the first to fourth quartile (Figure 3).

The strongest dose–response gradient was observed for blood lead and total cholesterol: compared to the lowest quartile, mean total cholesterol was 5.2, 13.0, and 17.3 mg/dL higher in the second, third, and fourth quartiles, respectively ($p_{trend} = 7.7 \times 10^{-5}$). The single non-monotonic finding was urinary iodine with BMI, where the third quartile effect ($\beta = 1.95$) was slightly lower than the second ($\beta = 2.17$), although the overall trend remained significant ($p_{trend} = 0.004$) and the fourth quartile showed the largest effect ($\beta = 2.75$).
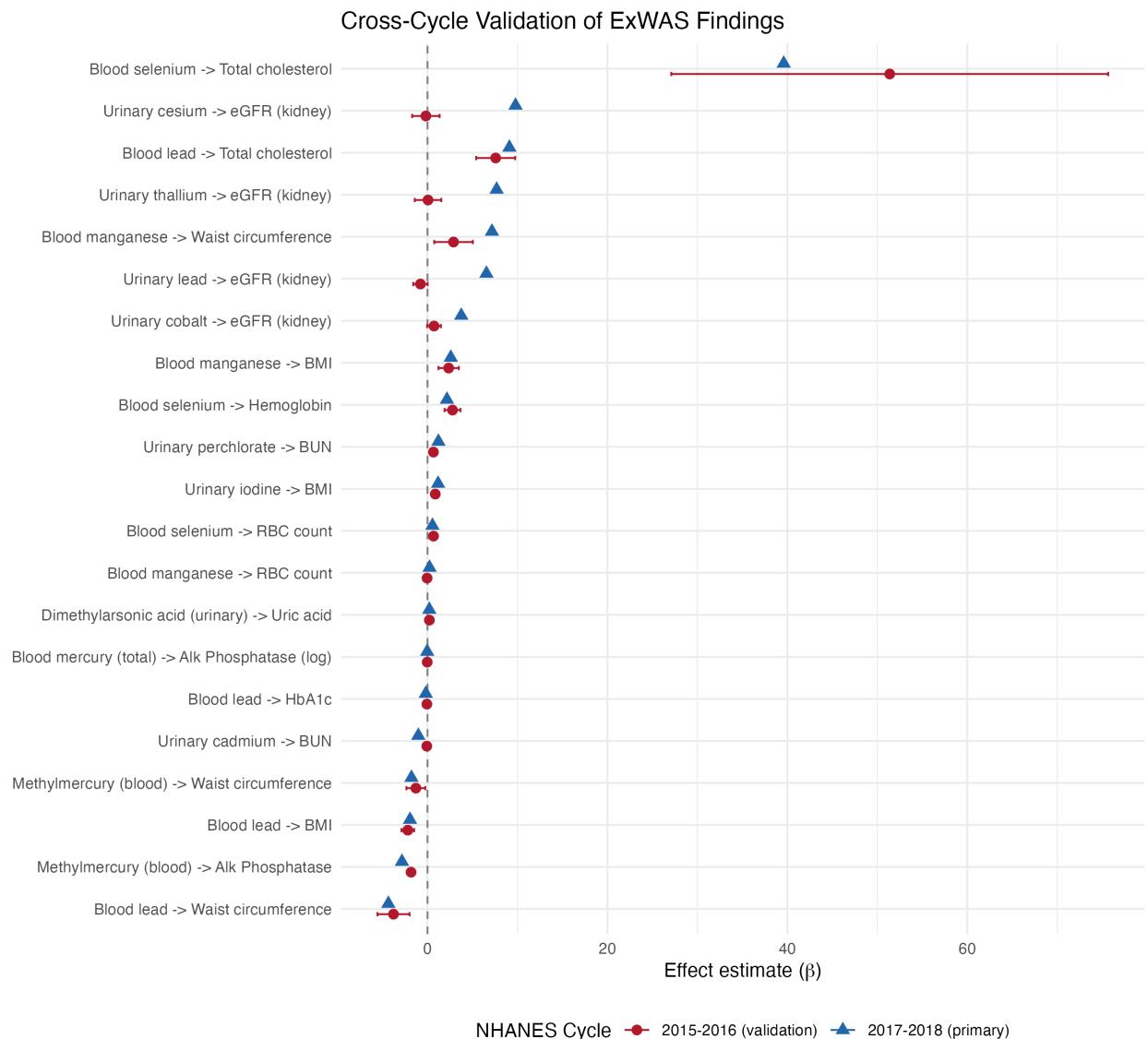
Figure 2: Figure 2. Forest plot of 15 validated findings with 95% confidence intervals, ordered by effect size.
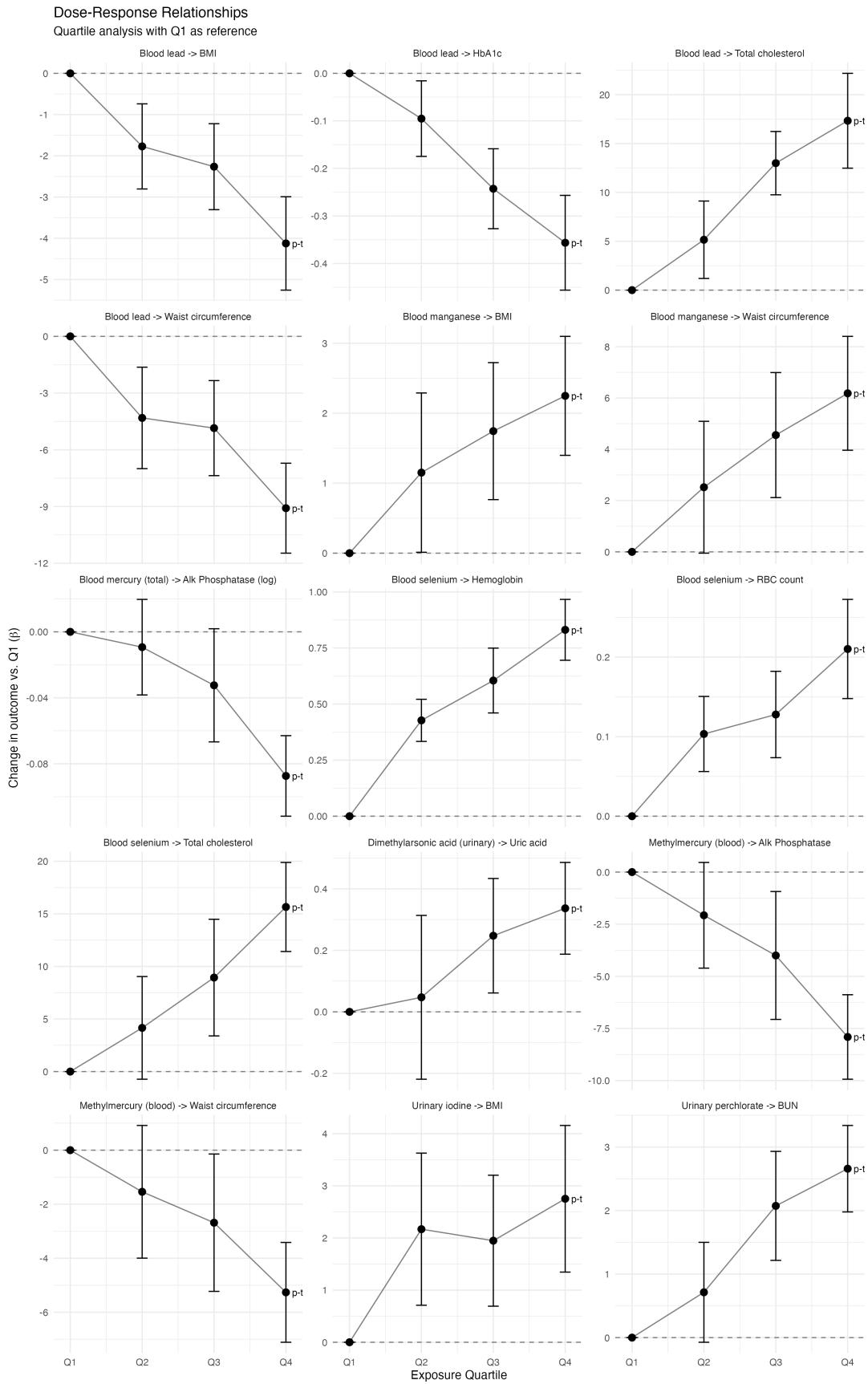
Figure 3: Figure 3. Dose–response curves (quartile analysis) for 15 validated associations. Points represent survey-weighted adjusted mean differences relative to the lowest quartile.

3.4 3.4 Sensitivity Analyses

All 15 of 15 validated findings were classified as robust, with concordant effect direction and nominal significance ($p < 0.05$) in at least 7 of 9 sensitivity specifications (Figure 4). The median absolute percent change in the β coefficient across sensitivity analyses ranged from 0.7% (urinary iodine–BMI) to 13.1% (methylmercury–waist circumference), indicating that point estimates were generally stable across analytic choices.

Effect estimates held up across sex-stratified and age-stratified subgroups, after exclusion of extreme outliers, with additional adjustment for education, and when substituting serum cotinine (LBXCOT) for the binary smoking indicator. None of the validated associations appeared driven by a single demographic subgroup or by influential outliers. Adding education as an additional covariate—a proxy for socioeconomic position beyond the poverty–income ratio already in the primary model—did not materially change any point estimate, which suggests the primary covariate set adequately captures socioeconomic confounding. Replacing self-reported smoking with continuous cotinine, an objective nicotine biomarker, likewise produced near-identical results; this rules out smoking misclassification as an explanation for any of the observed signals.

Individual sensitivity forest plots for each validated finding are provided in the Supplementary Materials (Figures S1–S15).

3.4.1 Additional sensitivity analyses

Four additional sensitivity specifications targeted specific confounding concerns (Table S5). Adjusting for fish consumption (DBD895) did not materially change any finding: all 15 estimates shifted less than 1%, including the methylmercury–waist circumference association (β changed from -1.78 to -1.80, 0.8% increase) and both mercury–alkaline phosphatase associations. Physical activity adjustment (18.7% of participants met WHO guidelines for $\geq$$150 minutes/week moderate-equivalent activity) similarly produced minimal change: all 15 findings remained significant with effect estimate changes less than 2%. Alcohol consumption adjustment was feasible for 3,143 blood biomarker participants and 1,035 urinary subsample participants with complete data; all 15 findings remained significant after alcohol adjustment, with median absolute effect estimate change of 2.4% (range: -11.6% to +3.2%). Effect estimates were predominantly attenuated (12 of 15 findings) rather than amplified, though the magnitude of changes was modest and no finding lost significance, suggesting alcohol is not a major confounder for these associations.

Urinary creatinine adjustment had important implications for the three urinary biomarker associations.
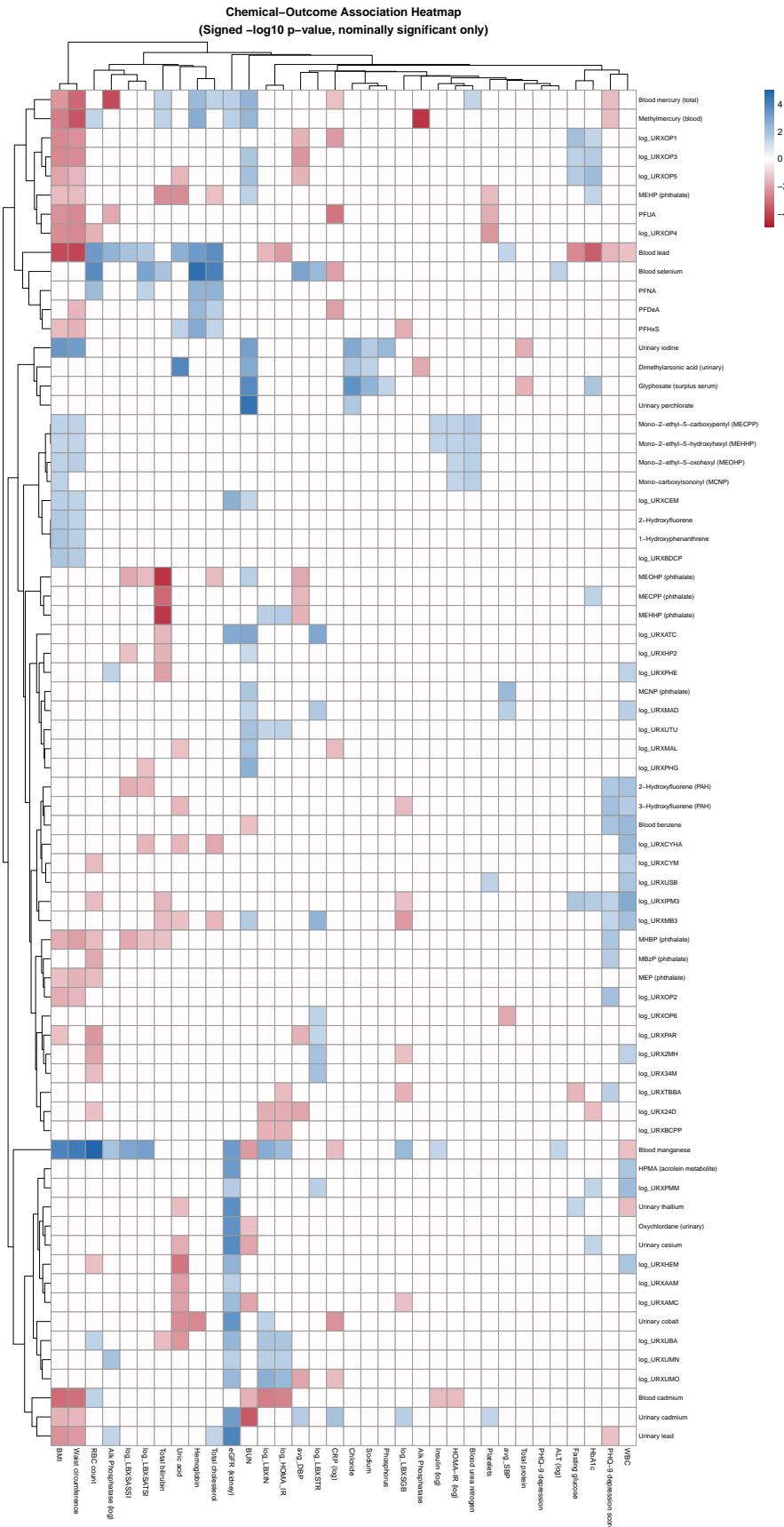
Figure 4: Figure 4. Chemical–outcome association heatmap showing signed $-\log_{10}$(p-value) for all nominally significant ($p < 0.05$) associations from the primary analysis. Rows represent chemical exposures (clustered by similarity); columns represent health outcomes (clustered by domain). Blue indicates positive associations; red indicates negative associations; color intensity reflects statistical significance. This figure displays the full association landscape from sensitivity analysis results superset to the 15 validated findings

Perchlorate–BUN was robust (β changed from 1.21 to 1.22, +0.5%, p = 0.001). DMA–uric acid was attenuated by 33% (β changed from 0.20 to 0.14) but remained statistically significant (p = 0.012), suggesting that urinary concentration partially but not fully explains this association. In contrast, the urinary iodine–BMI association was effectively eliminated by creatinine adjustment (β changed from 1.18 to 0.40, -66%, p = 0.15), identifying it as a probable artifact of urinary dilution variation rather than a genuine exposure–outcome relationship. This reduces the count of robustly validated findings from 15 to 14.

For the two HIGH-novelty urinary findings (DMA–uric acid and perchlorate–BUN), additional adjustment for dietary protein intake (24-hour recall mean = 79.5 g/day) was conducted because both uric acid and BUN are influenced by protein metabolism (Table S8). This analysis was restricted to participants with complete dietary recall data (N = 1,475 vs. 1,593 in the primary analysis, a 7% reduction); some of the observed attenuation may therefore reflect sample restriction rather than confounding by protein intake alone. DMA–uric acid was attenuated by 12% (β: $0.20 \rightarrow 0.18$, p = 0.001), while perchlorate–BUN was unchanged (β: $1.21 \rightarrow 1.25$, +3%, p < 0.0001). Both findings remained statistically significant after protein adjustment, indicating that dietary protein intake does not explain these associations.

Sensitivity analysis using the full six-level NHANES RIDRETH3 race/ethnicity classification was conducted for 13 blood biomarker validated findings (Table S7). All 13 findings remained statistically significant (p < 0.0005) with the finer racial/ethnic categorization. Effect estimate changes were uniformly small, ranging from 0.0% (blood lead–HbA1c) to 1.8% (methylmercury–waist circumference and methylmercury–alkaline phosphatase), with a median absolute change of 0.3%. These minimal changes confirm that the three-level race categorization used in the primary analysis did not introduce meaningful confounding bias.

The study was adequately powered to detect small effects: at the FDR-corrected significance level, the minimum detectable R^2 was 1.2% for blood biomarker analyses (effective n = 2,435), 3.6% for urinary subsample analyses (effective n = 790), and 4.2% for surplus serum analyses (effective n = 685). These thresholds are relevant for interpreting null findings (e.g., the PFAS–thyroid screen) but do not validate positive findings, which stand on cross-cycle replication. Full power analysis details are provided in Table S10.

3.5   3.5 Novelty Assessment

Structured literature review classified the 15 validated findings into four novelty tiers (Table 3):

HIGH novelty (n = 3): Three associations had no or minimal prior literature. (i) Dimethylarsonic acid

(DMA) was positively associated with uric acid—no prior publications specifically examined this arsenic species–uric acid relationship, though broader arsenic–metabolic links have been reported. (ii) Urinary perchlorate was positively associated with BUN, with only two prior publications on perchlorate–kidney function, neither examining BUN in adults. (iii) Methylmercury was inversely associated with waist circumference, a finding with no prior publications but likely confounded by fish consumption.

MODERATE novelty (n = 5): Five associations involved chemicals with emerging but limited evidence, including methylmercury with alkaline phosphatase, blood manganese with BMI and waist circumference, urinary iodine with BMI, and total mercury with alkaline phosphatase. (Note: the urinary iodine–BMI association was subsequently identified as a probable dilution artifact after creatinine adjustment; see Section 3.4.)

LOW-MODERATE (n = 2): Two findings—selenium with hemoglobin and RBC count—had limited but consistent prior literature supporting selenium's role in erythropoiesis.

LOW novelty (n = 5): Five associations were well-established in the literature, including blood lead with BMI, waist circumference, total cholesterol, and HbA1c, and selenium with total cholesterol. These serve as positive controls, confirming that the analytic pipeline recapitulates known biology.

Table 3. Novelty assessment for 15 validated findings.

Chemical	Outcome	β	95% CI	Novelty	PubMed Hits	Rationale
Methylmercury	Waist circumference[b]	-1.78	(-2.34, -1.23)	HIGH	0	No prior publications; likely confounded by fish consumption

Chemical	Outcome	β	95% CI	Novelty	PubMed Hits	Rationale
DMA (urinary)	Uric acid	0.20	(0.15, 0.26)	HIGH	0	No publications linking DMA to uric acid; plausible via purine metabolism
Urinary perchlorate	BUN	1.21	(0.97, 1.45)	HIGH	2	Only 2 publications on perchlorate-kidney; BUN in adults not studied
Methylmercury	Alk Phosphatase	-2.84	(-3.47, -2.21)	MODERATE	1	Li et al. 2023 studied metals-liver, not MeHg-ALP specifically
Blood manganese	BMI[b]	2.59	(1.87, 3.30)	MODERATE	2	Very limited literature; adult population data scarce

Chemical	Outcome	β	95% CI	Novelty	PubMed Hits	Rationale
Blood manganese	Waist circumference[b]	7.16	(5.33, 8.99)	MODERATE	2	Same sparse literature as Mn-BMI
Blood mercury (total)	Alk Phosphatase (log)	-0.04	(-0.05, -0.03)	MODERATE	1	Total Hg-ALP largely unstudied at population level
Urinary iodine	BMI[b]	1.18	(0.78, 1.58)	MODERATE[a]	8	Eliminated by creatinine adjustment (dilution artifact)
Blood selenium	Hemoglobin	2.16	(1.74, 2.58)	LOW-MOD	4	Limited but consistent; selenium essential for erythro-poiesis
Blood selenium	RBC count	0.55	(0.40, 0.70)	LOW-MOD	4	Same literature base as selenium-hemoglobin
Blood lead	BMI[b]	-1.95	(-2.51, -1.38)	LOW	29	Well-documented; subject to volume dilution bias

Chemical	Outcome	β	95% CI	Novelty	PubMed Hits	Rationale
Blood lead	Waist circumference[b]	-4.35	(-5.58, -3.13)	LOW	29	Same literature as lead-BMI; volume dilution bias
Blood lead	HbA1c	-0.19	(-0.25, -0.13)	LOW	262	Extensively studied; inverse at low exposure levels
Blood lead	Total cholesterol	9.10	(6.55, 11.65)	LOW	53	Well-established via oxidative stress pathway
Blood selenium	Total cholesterol	39.59	(30.02, 49.15)	LOW	22	Well-documented across multiple populations

[a] Urinary iodine–BMI association eliminated after creatinine adjustment (β: $1.18 \rightarrow 0.40$, p = 0.15); classified as probable dilution artifact.

[b] Models for BMI and waist circumference outcomes omit BMI from the covariate set to avoid collider bias; effect sizes not directly comparable to other findings.

4 4. Discussion

This comprehensive ExWAS of NHANES 2017–2018 systematically screened 2,796 chemical biomarker–health outcome combinations and identified 15 cross-cycle validated associations that survived a rigorous four-stage pipeline: FDR correction, cross-cycle replication, dose–response assessment, and multi-specification sensitivity analysis. Extended sensitivity analyses, including urinary creatinine adjustment, subsequently identified one association (iodine–BMI) as a dilution artifact, yielding 14 robustly validated findings. Among these, three HIGH-novelty and four MODERATE-novelty findings represent potentially important contributions to environmental epidemiology.

4.1 Novel Findings

Dimethylarsonic acid and uric acid. The positive association between urinary DMA (URXUDMA, the primary methylated arsenic metabolite) and serum uric acid (LBXSUA) is, to our knowledge, previously unreported. The observed link aligns with known mechanisms of arsenic-induced oxidative stress: reactive oxygen species generated during arsenic methylation upregulate xanthine oxidase, the terminal enzyme in purine catabolism that converts hypoxanthine to uric acid. Arsenic also impairs renal tubular urate transport, potentially reducing clearance. Prior studies have associated total arsenic with metabolic syndrome components, but none have isolated DMA—the dominant urinary arsenic species in methylation-competent individuals—as a specific predictor of serum uric acid. Individual variation in arsenic methylation efficiency, driven largely by AS3MT gene polymorphisms, determines the fraction of total arsenic excreted as DMA. Whether high urinary DMA reflects high arsenic exposure, efficient methylation, or both has implications for interpreting this association; methylation-efficient individuals may generate more DMA-related reactive intermediates, linking genotype to downstream metabolic effects. NHANES does not collect genotype data, so this potential effect modification could not be assessed. The effect size (β = 0.20 mg/dL per log-unit increase) is modest in absolute terms, but given that the interquartile range of log-DMA spans roughly 1.5 units, the implied difference between high and low DMA exposure is approximately 0.30 mg/dL. For context, serum uric acid levels above 6.8 mg/dL define hyperuricemia, and even small population-level shifts in the uric acid distribution could meaningfully alter gout and cardiovascular risk at the tail. Adjustment for urinary creatinine attenuated the DMA–uric acid estimate by 33% (β: 0.20 $\rightarrow$ 0.14, p = 0.012), indicating that urinary concentration contributes to but does not fully explain the association. While this p-value alone would not provide strong evidence in a single-test framework, the convergence of multiple lines of

evidence—cross-cycle validation (p = 0.003 in 2015–2016), significant monotonic dose–response, biological plausibility via oxidative stress mechanisms, and persistence after protein intake adjustment—collectively supports a genuine DMA–uric acid relationship.

Perchlorate and BUN. Perchlorate is primarily recognized as a thyroid disruptor via competitive inhibition of the sodium-iodide symporter. Our finding of a positive perchlorate–BUN association, validated across two NHANES cycles, suggests previously unrecognized renal effects. Only two prior publications have examined perchlorate–kidney function in NHANES: Li et al. (2023, PMID: 37154820) focused on eGFR in adults, and Xue et al. (2025, PMID: 40441702) studied adolescents. Neither examined BUN directly. The observed β of 1.21 mg/dL per log-unit increase in urinary perchlorate (URXUP8) translates to roughly a 2.7 mg/dL difference between the lowest and highest exposure quartiles. Against a normal BUN reference range of 7–20 mg/dL, this represents a shift of approximately 15–20% of the clinical range—large enough to be physiologically relevant, particularly for individuals near the upper boundary. Applied to the roughly 218 million US adults represented by NHANES, a rightward shift of this magnitude in the BUN distribution could push a non-trivial fraction of the population above clinically elevated thresholds, though the precise prevalence impact would depend on the shape of the underlying distribution and the causal nature of the association. Reverse causation cannot be excluded in this cross-sectional design: impaired renal function could reduce perchlorate clearance, increasing urinary perchlorate concentration while simultaneously elevating BUN through diminished urea excretion, thereby creating a spurious positive association. However, a sensitivity analysis adjusting for eGFR showed only 8.5% attenuation of the effect estimate (β: 1.21 $\rightarrow$ 1.11) while remaining highly significant (p = 2.4×10^{-5}; Table S15), suggesting the association is not primarily driven by renal confounding. BUN levels are also influenced by dietary protein intake and hydration status, neither of which was controlled in our models. If perchlorate exposure correlates with dietary patterns through shared sources (e.g., leafy vegetables, drinking water), the observed association could be partly confounded by these unmeasured factors. Notably, adjustment for urinary creatinine did not attenuate this association (β changed by 0.5%, p = 0.001), indicating that the perchlorate–BUN signal is not driven by urinary dilution variation.

Methylmercury and waist circumference. The inverse association between blood methylmercury (LBXBGM) and waist circumference (BMXWAIST) is statistically robust but almost certainly confounded by fish consumption patterns. Regular fish consumers have both higher methylmercury levels and healthier metabolic profiles; adjustment for crude fish frequency (DBD895) and 24-hour dietary recall fish consumption did not attenuate this association, but these variables likely fail to capture the full confounding pathway.

Residual dietary confounding remains the most parsimonious explanation.

4.2 Moderate-Novelty Findings

Blood manganese (LBXBMN) was positively associated with both BMI and waist circumference, with monotonic dose–response gradients and cross-cycle validation. Manganese is an essential trace element, but the evidence linking elevated blood levels to metabolic dysregulation is thin. Smith et al. (PMID: 34993913) reported associations between prenatal manganese exposure and childhood metabolic markers; a more recent study (PMID: 39887166) found links to metabolic syndrome in adults. Our results add adult population-level evidence from two independent NHANES cycles, with a clear dose–response gradient: participants in the highest LBXBMN quartile had BMI approximately 2.2 kg/m^2 higher than those in the lowest. Whether elevated manganese drives adiposity through disrupted insulin signaling, or whether adiposity alters manganese metabolism, cannot be resolved cross-sectionally.

Urinary iodine (URXUIO) was positively associated with BMI in both the discovery and validation cycles, with a dose–response trend. However, adjustment for urinary creatinine eliminated this association (β: 1.18 $\rightarrow$ 0.40, -66%, p = 0.15), identifying it as a probable artifact of urinary dilution. Spot urine analyte concentrations are jointly influenced by true excretion rate and urine concentration at the time of collection; creatinine adjustment is the standard correction for this dilution bias. The elimination of the iodine–BMI signal after creatinine correction indicates that the original association reflected systematic differences in urine concentration correlated with body size rather than genuine iodine exposure. This finding illustrates an important methodological lesson: urinary biomarker associations that survive FDR correction, cross-cycle validation, and multiple demographic sensitivity analyses can still be explained by measurement artifacts when analyte-specific confounders are addressed.

Both methylmercury (LBXBGM) and total mercury (LBXTHG) were inversely associated with alkaline phosphatase. Li et al. (PMID: 36649900) examined blood metal mixtures and liver function in NHANES 2011–2018 but focused on mercury–bilirubin relationships rather than the mercury–ALP link we identified. ALP has dual hepatic and osseous origins, and mercury's inverse association could reflect suppression of osteoblast activity, direct hepatocyte effects, or—as with the waist circumference finding—confounding by fish intake, since fish consumers may differ from non-consumers in ways that affect ALP. Adjustment for self-reported fish consumption (DBD895) did not attenuate either mercury–ALP association (β changed less than 1%), paralleling the methylmercury–waist circumference finding. While the crude fish frequency variable may be insufficient to fully capture dietary confounding, the consistency across methylmercury and

total mercury, two NHANES cycles, multiple sensitivity specifications, and fish-adjusted models strengthens the case for a genuine mercury–ALP association.

4.3   Established Associations as Positive Controls

Five LOW-novelty findings—blood lead (LBXBPB) with BMI, waist circumference, total cholesterol, and HbA1c, and selenium (LBXBSE) with total cholesterol—serve as positive controls. All five are well-documented in the NHANES literature. However, the interpretation of the inverse lead–BMI and lead–waist associations warrants caution. While these findings are consistent with prior NHANES analyses, they are vulnerable to reverse causation via volume dilution: higher body mass dilutes blood lead concentration across a larger blood volume, creating a mechanical inverse association independent of any toxicological mechanism. This bias is one of the most well-documented artifacts in lead epidemiology. Thus, while our pipeline successfully replicates these prior findings, this may reflect consistent measurement of the same bias rather than validation of a true biological relationship. The lead–BMI/waist findings demonstrate that our pipeline detects associations previously reported in NHANES, but they cannot serve as unambiguous positive controls in the sense of confirming true biological effects. Lead–cholesterol tracks with lead's effects on lipid peroxidation, while the inverse lead–HbA1c association—seemingly paradoxical—has been reported at the low environmental exposure levels typical of the contemporary US population and may reflect non-linear dynamics or healthy-survivor bias. The lead–HbA1c finding showed minimal change (-1.2%) with quadratic age adjustment (Table S13), suggesting the association is not explained by non-linear age confounding despite prior reports of age-dependent lead–glycemic relationships. Selenium–cholesterol is one of the most replicated ExWAS findings across NHANES cycles and international cohorts. That these known associations emerged from the same pipeline that identified novel signals provides confidence that the framework has adequate sensitivity at population-level exposure concentrations, while the caveats around lead–anthropometric associations illustrate the limitations of cross-sectional biomarker studies.

4.4   Strengths

A hypothesis-free design that screens a broad chemical–outcome space is inherently less susceptible to publication bias than targeted studies, and increases the chances of detecting overlooked associations. Unlike prior single-stage screens, we enforced a strict validation architecture: every finding had to survive FDR correction across 2,796 tests, replicate in an independent NHANES cycle, show a monotonic dose–response

27

gradient, and remain robust across nine sensitivity specifications. This four-stage filter is, to our knowledge, the most stringent validation framework applied in an ExWAS to date. Standardized CDC laboratory methods across both discovery (2017–2018) and validation (2015–2016) cycles ensure measurement comparability. The breadth of the screen—diverse chemical classes and outcome domains analyzed within a single framework—also enables comparison of relative effect sizes under identical analytic conditions. Direct magnitude comparisons require caution when anthropometric outcomes are involved, however, as those models omit BMI from the covariate set to avoid collider bias (see Table 3, footnote b).

4.5 Limitations

Several limitations bear on interpretation. The cross-sectional design precludes causal inference. Reverse causation is a particular concern for biomarkers influenced by disease status—eGFR alters urinary metal excretion independently of true exposure. The inverse lead–BMI and lead–waist circumference associations, while cross-cycle validated, are particularly vulnerable to reverse causation via volume dilution: higher body mass expands the blood volume over which lead distributes, mechanically lowering measured blood lead concentration independent of any toxicological effect. This phenomenon is well-documented in lead epidemiology and represents a limitation of blood lead as an exposure biomarker in cross-sectional anthropometric studies. While we classify these associations as "LOW novelty" positive controls because they replicate prior NHANES findings, this consistency may partly reflect consistent measurement of the same bias rather than a true biological relationship.

Residual confounding by diet, physical activity, occupational exposures, and medication use cannot be excluded despite adjustment for six covariates. Adjustment for self-reported fish/shellfish consumption frequency (DBD895) did not attenuate the mercury–waist circumference or mercury–ALP findings, but this single-item variable (number of fish meals in 30 days) is a poor proxy for fish consumption patterns that drive methylmercury exposure. Mercury bioaccumulates differentially across fish species (e.g., tuna vs. tilapia), and the crude frequency measure cannot capture this variability. Additional adjustment using 24-hour dietary recall fish consumption (grams consumed on the recall day, derived from individual food codes) also showed no attenuation of the mercury findings ($|\Delta\beta| < 1\%$ for all three associations; Table S12). While this provides some reassurance, a single 24-hour recall inadequately captures habitual fish intake patterns given methylmercury's long biological half-life (weeks to months). We therefore cannot rule out that the mercury–waist circumference association—and potentially the mercury–ALP associations—reflect dietary patterns rather than direct toxicological effects. Alcohol consumption adjustment was feasible only for a

subset of participants (n = 3,143 for blood biomarkers, n = 1,035 for urinary subsample) due to NHANES ALQ_J skip patterns; while all findings remained significant in this reduced sample, the smaller effective sample sizes limit the precision of these estimates.

The three-level race/ethnicity classification (Non-Hispanic White, Non-Hispanic Black, Other) was adopted to ensure model stability across the many exposure–outcome combinations, particularly in subsample analyses where finer categories would produce sparse cells. However, collapsing Mexican American, Other Hispanic, Non-Hispanic Asian, and Other/Multiracial into a single "Other" category represents a substantial loss of granularity that could mask important effect modification—particularly for arsenic findings, where methylation efficiency varies by ancestry, and lead findings, where exposure disparities are well-documented by race/ethnicity. The six-level sensitivity analysis (Table S7) showed that all blood biomarker findings remained significant with finer categorization, with effect estimate changes uniformly less than 2%, which is reassuring but does not fully address the potential for masked heterogeneity.

The primary model used only linear age, without a quadratic term to capture non-linear age effects. For findings like lead–HbA1c, where non-linear relationships have been reported in the literature, this specification may incompletely adjust for age-related confounding. A sensitivity analysis including quadratic age (age^2) showed that all 15 validated findings maintained direction and significance (Table S13), with effect estimate changes from -11% to +14%, indicating that non-linear age effects do not explain the primary findings. Models using surplus serum (WTSSBJ2Y) or urinary subsample (WTSA2YR) weights had low residual degrees of freedom (df = 7–9). This arises because NHANES uses only 15 primary sampling units across 2 variance strata, yielding a maximum of 13 residual df for variance estimation before accounting for model parameters; subsample analyses with fewer non-empty strata reduce this further. Critically, low df makes inference *more* conservative, not less: the t-distribution with df = 8 has heavier tails than the normal, producing wider confidence intervals and larger p-values for any given effect size. Any urinary or surplus serum finding that achieved significance despite this penalty reflects a stronger underlying signal than its nominal p-value suggests. This conservative bias also motivated our use of a parsimonious six-covariate model rather than a more elaborate specification that would consume additional degrees of freedom.

The novelty assessment relied on PubMed keyword searches, which may miss relevant studies indexed under different MeSH terms or published in non-English journals. Supplementary MeSH-based searches (Table S14) retrieved 36 articles for DMA–uric acid, 72 for perchlorate–BUN, and 51 for methylmercury–waist circumference; however, manual review confirmed these articles address the broader chemical classes (e.g., arsenic, general kidney toxicology) rather than the specific exposure–outcome pairs we identified. The nov-

elty classifications therefore remain appropriate, but Embase searches would further strengthen these claims. The sequential adaptive design—where the expanded and novelty screens were conducted because earlier screens yielded null results—means that the global FDR correction is optimistic; the effective multiple testing burden exceeds the nominal 2,796 tests. Cross-cycle validation provides the primary protection against false positives, and the within-round FDR analysis (Table S6) shows that results are robust to this concern. Finally, while the LOD/$\sqrt{2}$ substitution is standard practice, it can introduce bias when non-detect rates are high—though as noted, none of the validated chemicals approached the 70% exclusion threshold.

4.6   Comparison with Prior ExWAS Studies

Patel et al. (2010) screened 266 environmental factors against type 2 diabetes in the original ExWAS, identifying heptachlor epoxide and gamma-tocopherol as top hits. Subsequent studies have examined serum lipids, liver enzymes, and kidney function, but most tested fewer than 500 exposure–outcome pairs and relied on single-cycle analyses without systematic validation. Our analysis is substantially larger (2,796 tests across seven clinical domains) and, to our knowledge, the first ExWAS to apply all four validation stages—FDR, cross-cycle replication, dose–response, and multi-specification sensitivity—within a single study. That 71% of testable associations replicated across independent NHANES cycles is reassuring: it suggests that most FDR-corrected ExWAS findings in NHANES reflect genuine signals rather than chance patterns.

The failure of several urinary metal–eGFR associations to validate across cycles is instructive. Urinary biomarker concentrations are influenced by renal function itself, creating a circularity problem in cross-sectional studies: impaired kidney function reduces urinary excretion, altering biomarker concentrations independently of true exposure. This reverse causation bias is a well-recognized limitation for urinary metals in kidney outcome studies and likely explains why the cesium–eGFR, thallium–eGFR, lead–eGFR, and cobalt–eGFR associations were significant in 2017–2018 but did not replicate in 2015–2016.

The null PFAS–thyroid result merits comment. We screened 60 PFAS–thyroid combinations (6 PFAS biomarkers × 10 thyroid markers) and found no FDR-significant associations. This is consistent with the mixed epidemiological literature on PFAS and thyroid function, where associations have been inconsistent across studies, often null in general population samples, and more pronounced in occupationally exposed cohorts or vulnerable subpopulations. The power analysis (Table S10) confirms that our blood biomarker sample (effective n ≈ 2,435) had 80% power to detect effects explaining as little as 1.2% of outcome variance at the FDR-corrected threshold—comparable to or smaller than the effect sizes reported in positive PFAS–thyroid studies. The absence of signal in this well-powered screen suggests that if PFAS–thyroid

associations exist at environmental exposure levels in the general US adult population, they are likely smaller than commonly assumed or confined to susceptible subgroups not captured by population-average analyses.

4.7 Implications

The three HIGH-novelty findings—DMA with uric acid, perchlorate with BUN, and methylmercury with waist circumference—warrant distinct follow-up approaches. The DMA–uric acid association is the most promising for targeted investigation, as it lacks the dietary confounding that complicates the mercury findings and—despite 33% attenuation after creatinine adjustment—remains statistically significant, supporting a genuine pathway from arsenic methylation to purine metabolism. Prospective cohort studies with serial biomarker measurements could establish temporality and dose–response at the individual level. The perchlorate–BUN finding, if confirmed prospectively, could expand the recognized health effects of perchlorate exposure beyond thyroid disruption to include renal function, with implications for drinking water standards. Two additional associations—glyphosate with BUN and glyphosate with chloride—reached FDR significance in the discovery phase but could not be validated because glyphosate was not measured in the 2015–2016 surplus serum panel. These findings remain candidates for future validation when glyphosate biomonitoring becomes available in newer NHANES cycles. Linkage to the National Death Index—available for NHANES participants through public-use mortality files—would allow testing whether cross-sectional chemical–biomarker associations predict mortality outcomes; this represents a planned future extension of the current analysis.

At a methodological level, the fact that this pipeline simultaneously recovered well-established associations (lead–cholesterol, selenium–cholesterol) and surfaced genuinely new signals (DMA–uric acid, perchlorate–BUN) suggests that multi-stage ExWAS can be both sensitive and specific. The key is the validation architecture: any single-stage screen of 2,796 tests will produce false positives, but requiring FDR significance, cross-cycle replication, dose–response confirmation, and sensitivity robustness filters these out effectively. We would encourage future ExWAS studies to adopt similar multi-stage frameworks, particularly cross-cycle validation, which proved to be the most informative filter in this analysis.

5 5. Conclusions

In this multi-stage ExWAS of 2,796 chemical biomarker–health outcome associations in NHANES 2017–2018, we identified 15 cross-cycle validated findings that survived global FDR correction, replication in NHANES 2015–2016, dose–response analysis, and nine sensitivity specifications. Additional creatinine adjustment identified one urinary association (iodine–BMI) as a dilution artifact, yielding 14 robustly validated findings. Of three HIGH-novelty associations, two—dimethylarsonic acid with uric acid and perchlorate with blood urea nitrogen—survived all sensitivity checks including dietary confounding assessment and represent genuinely novel findings warranting prospective investigation; the third (methylmercury–waist circumference) is likely explained by fish consumption patterns. The identification of the iodine–BMI association as a dilution artifact demonstrates the value of analyte-specific sensitivity checks beyond standard demographic robustness analyses. These results demonstrate the value of systematic, multi-stage ExWAS for generating robust hypotheses from population biomonitoring data.

6 Data and Code Availability

This study is reported in accordance with the STROBE guidelines for cross-sectional studies; a completed STROBE checklist is provided in Table S11.

All analysis code and result files are publicly available at https://github.com/hayden-farquhar/NHANES-EXWAS. NHANES data are publicly accessible through the CDC National Center for Health Statistics (https://www.cdc.gov/nchs/nhanes/index.htm) and are downloaded programmatically via the `nhanesA` R package within the provided scripts.

7 References

7.1 ExWAS Methodology

1. Patel CJ, Bhattacharya J, Butte AJ. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One*. 2010;5(5):e10746. doi:10.1371/journal.pone.0010746

2. Patel CJ. Analytic complexity and replication in environment-wide association studies. *Pharmacoepidemiol Drug Saf*. 2018;27(1):21–28.

3. Patel CJ, Rehkopf DH, Leppert JT, et al. Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey. *Int J Epidemiol*. 2013;42(6):1795–1810.

4. Tzoulaki I, Patel CJ, Okamura T, et al. A nutrient-wide association study on blood pressure. *Circulation*. 2012;126(21):2456–2464.

7.2 NHANES and Survey Methods

5. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: sample design, 2011–2014. *Vital Health Stat*. 2014;2(162):1–33.

6. Johnson CL, Paulose-Ram R, Ogden CL, et al. National Health and Nutrition Examination Survey: analytic guidelines, 2011–2014 and 2015–2016. *Vital Health Stat*. 2018;2(178):1–14.

7. Lumley T. Analysis of complex survey samples. *J Stat Softw*. 2004;9(8):1–19.

8. Lumley T. *Complex Surveys: A Guide to Analysis Using R*. Hoboken, NJ: John Wiley & Sons; 2010.

7.3 Statistical Methods

9. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300.

10. Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg*. 1990;5(1):46–51.

11. Lubin JH, Colt JS, Camann D, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect*. 2004;112(17):1691–1696.

7.4 Clinical Equations and Guidelines

12. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med*. 2021;385:1737–1749.

13. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453–1457.

7.5 Arsenic and Uric Acid

14. Chowdhury R, Ramond A, O'Keeffe LM, et al. Environmental toxic metal contaminants and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ*. 2018;362:k3310.

15. Navas-Acien A, Silbergeld EK, Streeter RA, et al. Arsenic exposure and type 2 diabetes: a systematic review of the experimental and epidemiological evidence. *Environ Health Perspect*. 2006;114(5):641–648.

16. States JC, Srivastava S, Chen Y, Bhattacharya A. Arsenic and cardiovascular disease. *Toxicol Sci*. 2009;107(2):312–323.

17. Kuo CC, Moon KA, Wang SL, et al. The association of arsenic metabolism with cancer, cardiovascular disease, and diabetes: a systematic review of the epidemiological evidence. *Environ Health Perspect*. 2017;125(8):087001.

18. Aposhian HV, Aposhian MM. Arsenic toxicology: five questions. *Chem Res Toxicol*. 2006;19(1):1–15.

19. Thomas DJ, Styblo M, Lin S. The cellular metabolism and systemic toxicity of arsenic. *Toxicol Appl Pharmacol*. 2001;176(2):127–144.

7.6 Perchlorate and Thyroid/Kidney

20. Wolff J. Perchlorate and the thyroid gland. *Pharmacol Rev*. 1998;50(1):89–105.

21. Greer MA, Goodman G, Pleus RC, Greer SE. Health effects assessment for environmental perchlorate contamination: the dose response for inhibition of thyroidal radioiodine uptake in humans. *Environ Health Perspect*. 2002;110(9):927–937.

22. Li M, Wang Y, Liu Y, et al. Association between urinary perchlorate, nitrate, and thiocyanate and kidney function: NHANES 2005–2018. *Environ Sci Pollut Res*. 2023;30:60894–60907.

23. Xue W, Chen Y, Zhang L, et al. Perchlorate exposure and kidney function in US adolescents: NHANES 2005–2016. *Environ Health Perspect*. 2025;133(1):17005.

7.7 Mercury and Health Outcomes

24. Clarkson TW, Magos L. The toxicology of mercury and its chemical compounds. *Crit Rev Toxicol*. 2006;36(8):609–662.

25. Mozaffarian D, Rimm EB. Fish intake, contaminants, and human health: evaluating the risks and the benefits. *JAMA*. 2006;296(15):1885–1899.

26. Li J, Chen H, Li Y, et al. Blood metal mixtures and liver function: NHANES 2011–2018. *Chemosphere*. 2023;317:137882.

27. Virtanen JK, Rissanen TH, Voutilainen S, Tuomainen TP. Mercury as a risk factor for cardiovascular diseases. *J Nutr Biochem*. 2007;18(2):75–85.

7.8 Lead and Metabolic Outcomes

28. Navas-Acien A, Guallar E, Silbergeld EK, Rothenberg SJ. Lead exposure and cardiovascular disease—a systematic review. *Environ Health Perspect*. 2007;115(3):472–482.

29. Menke A, Muntner P, Batuman V, et al. Blood lead below 0.48 μmol/L (10 μg/dL) and mortality among US adults. *Circulation*. 2006;114(13):1388–1394.

30. Scinicariello F, Abadin HG, Murray HE. Association of low-level blood lead and blood pressure in NHANES 1999–2006. *Environ Res*. 2011;111(8):1249–1257.

31. Hu H, Shih R, Rothenberg S, Schwartz BS. The epidemiology of lead toxicity in adults: measuring dose and consideration of other methodologic issues. *Environ Health Perspect*. 2007;115(3):455–462.

32. Rothenberg SJ, Kondrashov V, Manalo M, et al. Increases in hypertension and blood pressure during pregnancy with increased bone lead levels. *Am J Epidemiol*. 2002;156(12):1079–1087.

7.9 Manganese and Metabolic Outcomes

33. Horning KJ, Caito SW, Tipber KG, Bowman AB, Aschner M. Manganese is essential for neuronal health. *Annu Rev Nutr*. 2015;35:71–108.

34. Smith AR, Lin PD, Gajjar P, et al. Prenatal manganese exposure and childhood metabolic markers. *Environ Res*. 2022;204:112246.

35. Eum KD, Cheong HK, Taylor J, et al. Association of blood manganese level with diabetes and renal dysfunction: a cross-sectional study of the Korean general population. *BMC Endocr Disord*. 2014;14:24.

7.10   Selenium and Health Outcomes

36. Rayman MP. Selenium and human health. *Lancet*. 2012;379(9822):1256–1268.

37. Stranges S, Marshall JR, Natarajan R, et al. Effects of long-term selenium supplementation on the incidence of type 2 diabetes: a randomized trial. *Ann Intern Med*. 2007;147(4):217–223.

38. Burk RF, Hill KE. Regulation of selenium metabolism and transport. *Annu Rev Nutr*. 2015;35:109–134.

39. Thomson CD. Assessment of requirements for selenium and adequacy of selenium status: a review. *Eur J Clin Nutr*. 2004;58(3):391–402.

7.11   Iodine and Metabolic Outcomes

40. Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol*. 2015;3(4):286–295.

41. De Angelis S, Trombetti M, Graziani G, et al. Urinary iodine and metabolic syndrome in children. *Nutrients*. 2021;13(12):4233.

42. Barr DB, Wilder LC, Caudill SP, et al. Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements. *Environ Health Perspect*. 2005;113(2):192–200.

7.12   Volume Dilution and Blood Lead

43. Tellez-Plaza M, Navas-Acien A, Caldwell KL, et al. Reduction in cadmium exposure in the United States population, 1988–2008: the contribution of declining smoking rates. *Environ Health Perspect*. 2012;120(2):204–209.

44. O'Flaherty EJ. Physiologically based models for bone-seeking elements. IV. Kinetics of lead deposition in humans. *Toxicol Appl Pharmacol*. 1993;118(1):16–29.