

# Endogenous Estrogen Metabolites and Gastric Cancer Risk Among Postmenopausal Women

Statistical Analysis and Results

*Isaac Zhao*

*01/10/19*

# Materials and Methods

## Study Design

Table 1: Demographics Table

	Cohort						Case Control					
	Iran		Korea (KMCC)		Germany		Korea (SNU)		Japan		Overall	
	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control
Sample Size (n)	43	81	29	54	10	11	146	137	15	15	243	298
Age (mean (sd))	62.3 (8.2)	62.5 (8)	63.8 (6.3)	61.7 (6.8)	63.7 (5.4)	65.7 (5.5)	64.5 (8)	57.3 (5.9)	72.1 (6.1)	71.9 (5.7)	64.5 (7.9)	60.6 (7.6)
BMI (mean (sd))	27.7 (5.7)	27.3 (5.5)	24.5 (3)	24.5 (3.6)	30 (5.5)	29.7 (4.6)	23.9 (3.2)	24.9 (3.3)	22.9 (3.4)	21 (2.4)	24.7 (4.3)	25.4 (4.4)
Ever smoked (n (%))	3 (7)	2 (2.5)	3 (10.3)	4 (7.4)	0	2 (18.2)	8 (5.5)	6 (4.4)	2 (13.3)	1 (6.7)	16 (6.6)	15 (5)
Ever drank alcohol (n (%))	0	0	6 (20.7)	11 (20.4)	2 (20)	5 (45.5)	30 (20.5)	47 (34.3)	2 (13.3)	3 (20)	40 (16.5)	66 (22.1)
Any educational degree (n (%))	1 (2.3)	10 (12.3)	19 (65.5)	33 (61.1)	10 (100)	10 (90.9)	130 (89)	122 (89.1)	NA	NA	160 (65.8)	175 (58.7)
Relative with gastric cancer (n (%))	1 (2.3)	1 (1.2)	NA	NA	0	1 (9.1)	28 (19.2)	15 (10.9)	1 (6.7)	5 (33.3)	30 (12.3)	22 (7.4)

Note:

NA = Data not available

Incident gastric cancer and two case-control studies of early-stage cancer were used for analysis. For the incident gastric cancer set, pre-diagnostic urine samples from all available postmenopausal (or age 60+ years) women diagnosed with gastric cancer and incidence-density matched controls from three prospective cohort studies (Golestan Cohort (Iran), Korean Multicenter Cancer Cohort, and ESTHER Cohort (Germany)) were tested. For the early-stage case-control gastric cancer set, urine samples from postmenopausal (or age 60+ years) women diagnosed with early-stage gastric cancer (AJCC clinical stages 1A [T1, N0, M0] or 1B [T1, N1, M0 or T2, N0, M0]) and 1:1 age-matched (+/- 5 years) controls from established case-control studies in Japan and Korea (Seoul Gastric Cancer Study) were tested.

Postmenopausal women with gastric cancer were matched by age to gastric cancer-free controls. Women who ever used post-menopausal hormone replacement were excluded since we were specifically interested in the effects of endogenous estrogens. Premenopausal women were also excluded since estrogen levels vary over the menstrual cycle, greatly complicating interpretation of measurements; in any case, gastric cancer is rare prior to age 50 years. On the other hand, restriction of the case-control set to stage 1 gastric cancer will limit the risk of reverse causality.

Urine specimens were collected at enrollment in prospective studies and pre-treatment in case-control studies and continuously cryopreserved at -70/-80 degrees Celcius until analysis.

Variables considered for covariate adjustment to estimate odds ratio effect sizes in the analysis portion were age, BMI, smoking, alcohol, education, and if the subject's relative had gastric cancer. Categories were simplified into never or ever to account for differences in how each variable was recorded for each study. This avoided ambiguous temporality of substance use for variables such as smoking and alcohol. Dividing into only two categories also reduced extreme sparsity sample size problems.

## Laboratory Assay

Stable isotope dilution liquid chromatography-tandem mass spectrometry (LC-TMS) was used at the NCI Laboratory of Proteomics and Analytical Technologies, MD to simultaneously measure the total concentration of 2 parent estrogens (estrone and estradiol) and 13 estrogen metabolites (2-hydroxyestrone, 2-methoxyestrone, 2-hydroxyestradiol, 2-methoxyestradiol, 2-hydroxyestrone-3-methyl ether, 4-hydroxyestrone, 4-methoxyestrone, 4-methoxyestradiol, 16 $\alpha$ -hydroxyestrone, 16-ketoestradiol, estriol, 17-epiestriol, and 16-epiestriol) in an aliquot of 500  $\mu$ L urine assay for each participant. In urine, parent estrogens and their metabolites are present primarily in conjugated form. Estrogen concentrations in spot urine samples were normalized to creatinine levels in order to adjust for variation in urinary volume.

# Results

## Covarying Relationships Between Metabolites

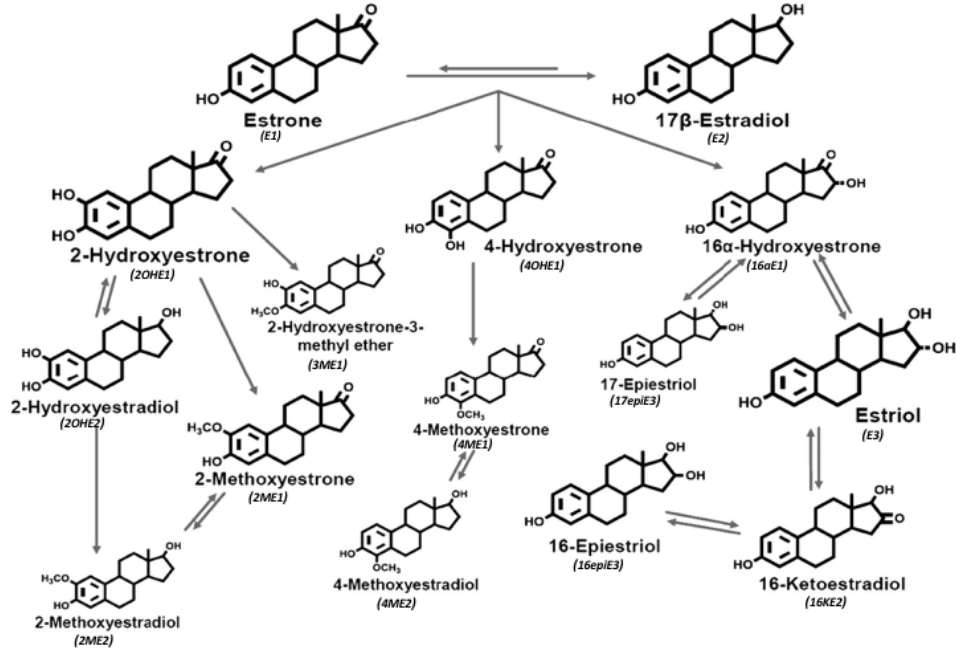


Figure 1: Estrogen Metabolite Hydroxylation Pathway Diagram

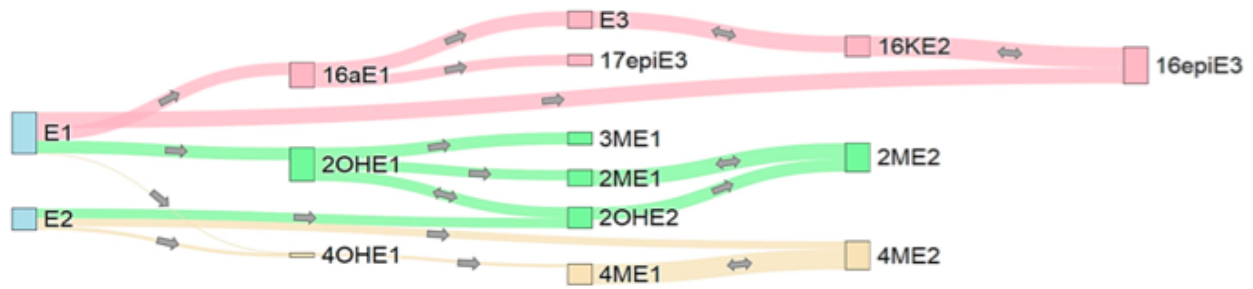


Figure 2: Metabolite correlations between direct relationships on hydroxylation group pathway. Line thickness proportional to spearman correlations between substrates and products on hydroxylation pathways (2-OH, green; 4-OH, tan; 16-OH, pink), ranging from 0.01 to 0.62.

Correlations between metabolites were relatively equal throughout the pathways except for the 4-Hydroxylation pathway connection with the 4OHE1 metabolite. The 16-Hydroxylation group showed slightly stronger correlations than the other hydroxylation groups.

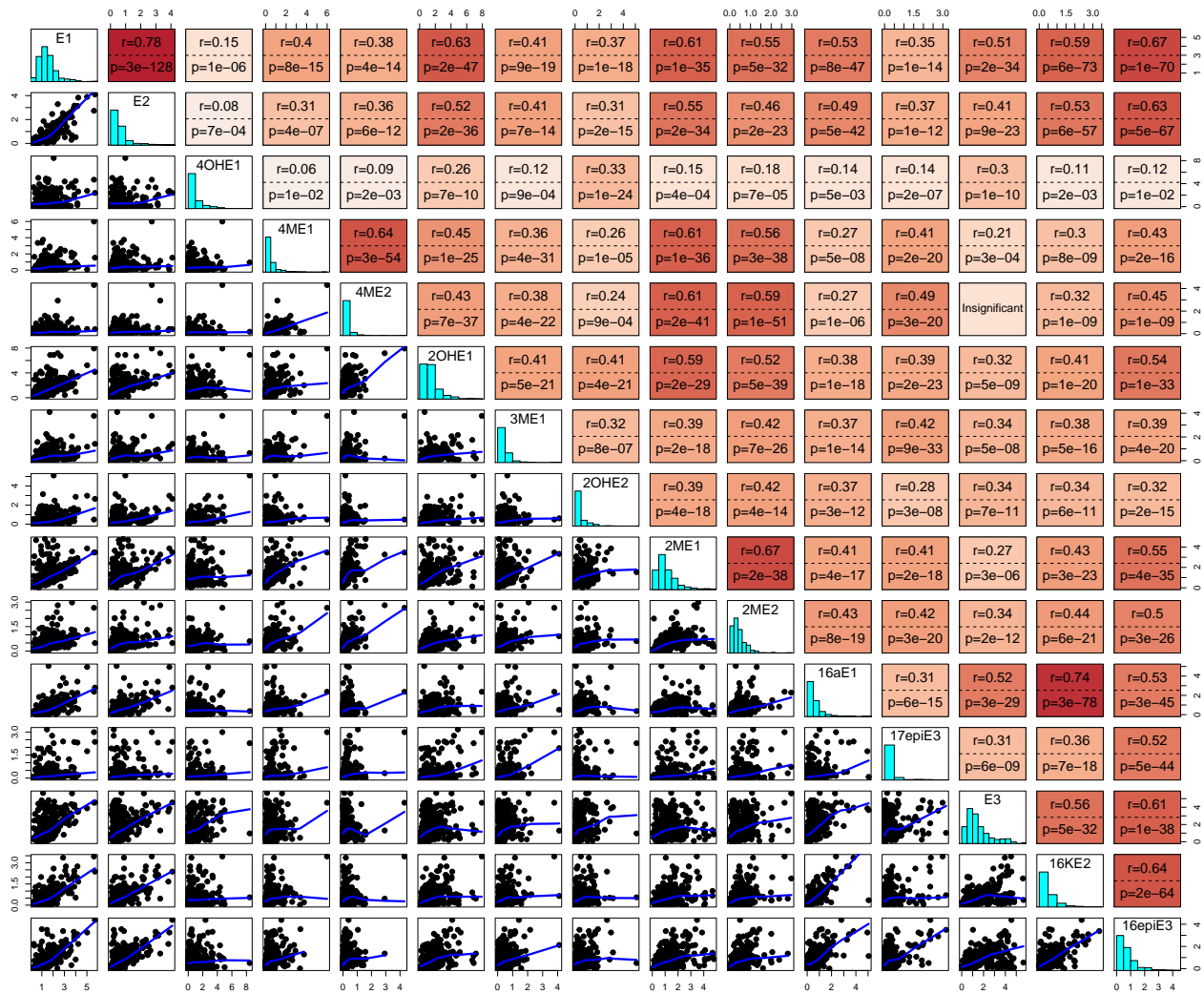


Figure 3: Metabolite correlation matrix. Redder regions indicate stronger positive spearman correlations while bluer indicate stronger inverse spearman correlations. Scatterplot and histogram on log scale.

The relationship between estrogen metabolites were computed using spearman correlations on log transformed concentrations. Spearman correlation was chosen to account for nonlinear relationships among the metabolites that are more resistant to outliers and influential points. Log transformations were used to better visualize the paired scatterplots due to highly skewed distributions.

The histograms show that the estrogen metabolite concentration distributions were still highly right tail skewed even after log transformations. The correlation matrix shows that all estrogens were positively associated with each other, even after adjusting for bonferonni corrections of significance levels. The strongest correlation was among parent estrogens E1 and E2, which was expected. The lowest smooth curve regression line with span 2/3 (in dark blue on the scatterplot) shows that on the log scale, there did not appear to be any major nonlinear higher order interactions (i.e. relationships were mostly linear on the log scale).

## Metabolite and Creatinine Missing Data

Table 2: Table of missing values for metabolites and creatinine

Estrogen	Missing (n)	Missing (%)
E1	25	1.600
E2	23	1.400
4OHE1	183	11.400
4ME1	34	2.100
4ME2	23	1.400
2OHE1	28	1.800
3ME1	24	1.500
2OHE2	61	3.800
2ME1	22	1.400
2ME2	23	1.400
16aE1	77	4.800
17epiE3	29	1.800
E3	54	3.400
16KE2	42	2.600
16epiE3	35	2.200
creatinine	20	1.200

The average percent missing values for estrogen metabolites was about 3%, with 4OHE1 having the highest percentage of missing values. Missing values for metabolites were caused by mis-alignment of co-eluting peaks of the LC-TMS device, which could not be fully amended even after correction. The missing values were not imputed due to the unknown relationship mis-measurement of the LC-TMS device had with the probability of a measurement to be un-recorded. In other words, it could not be determined if the distribution of missingness was independent of laboratory procedures. All estrogen batches were re-run to compare non-overlapping missing values, which determined that missingness did not appear to be correlated with how high or low the concentrations were. Thus, the standard approach of replacing all missing values with half lower limit of quantification (LLOQ) or lower limit of detection (LOD) would be inappropriate. Since missingness was determined to be independent of measured variables, complete cases (i.e. removing all missing values) was used. It should be noted that results that are borderline significant at the  $\alpha=0.05$  level may change if missing values were treated differently instead.

## Quality Control (QC) Analysis

Laboratory personnel was blinded to the case-control status of sample donors. A quality control (QC) set of 20 masked duplicate samples plus 4 additional laboratory control replicates from subjects with high available volumes representative of all studies was performed. Coefficient of variation (CV) and intraclass correlation coefficient (ICC) was calculated for log-transformed standardized estrogen metabolite and creatinine concentrations for the QC samples to assess within- and between-batch variations for assay reliability. Estrogen concentrations were log transformed to improve normality of distributions to meet assumptions necessary for the computation of variance and mean. A constant of 1 was added to concentrations before log transformations to ensure all measurements were positive values. The formula used to calculate ICC was  $ICC = \frac{\sigma_{bs}^2}{\sigma_{bs}^2 + \sigma_{bb}^2 + \sigma_{ws}^2}$  and the formula for the computation of CV was  $CV = \frac{\sigma_{ws}^2 + \sigma_{bb}^2}{\mu}$ ; where  $\sigma_{bs}^2$  = variance between subject,  $\sigma_{bb}^2$  = variance between batch, and  $\sigma_{ws}^2$  = variance within subject. The variance components were computed using a two stage multilevel model with varying intercept for ID and batch. The lower limit of quantitation for each analyte was 0.04 ng/mL.

## Laboratory Control QC Samples

Table 3: Laboratory control samples QC calculations

Estrogen	$\sigma_{bs}^2$	$\sigma_{bb}^2$	$\sigma_{ws}^2$	Mean	ICC (%)	CV (%)
E1	0.3	0.01	0.02	2.05	89.7	9
E2	0.39	0.01	0.02	1.17	94.4	12.9
4OHE1	6.62	1.69	0	2.27	79.7	57.3
4ME1	0	0.14	0.12	0.36	0	141.4
4ME2	0	0.01	0.01	0.11	20.5	114.2
2OHE1	0.59	0.77	0.75	1.53	28.1	80.4
3ME1	0.01	0.01	0.01	0.4	20.4	38.2
2OHE2	1.29	0.04	0.34	0.51	77.2	120.6
2ME1	0.12	0.1	0.35	1.31	21.2	51
2ME2	0.01	0.01	0.01	0.36	41.6	39.9
16aE1	0.43	0.01	0.04	1.08	90.5	19.8
17epiE3	0.37	0.01	0.02	0.58	92.8	29.1
E3	1.01	0.32	0.12	2.18	70	30.2
16KE2	0.51	0	0.03	1	94.4	17.2
16epiE3	0.35	0.09	0.11	1.12	63.4	40.2

Note: bs = between subject, bb = between batch, ws = within subject

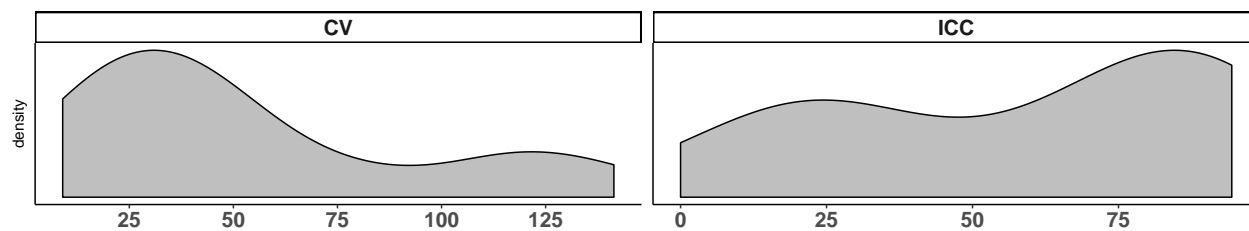


Figure 4: Distribution of CV and ICC for all metabolites for laboratory control QC samples

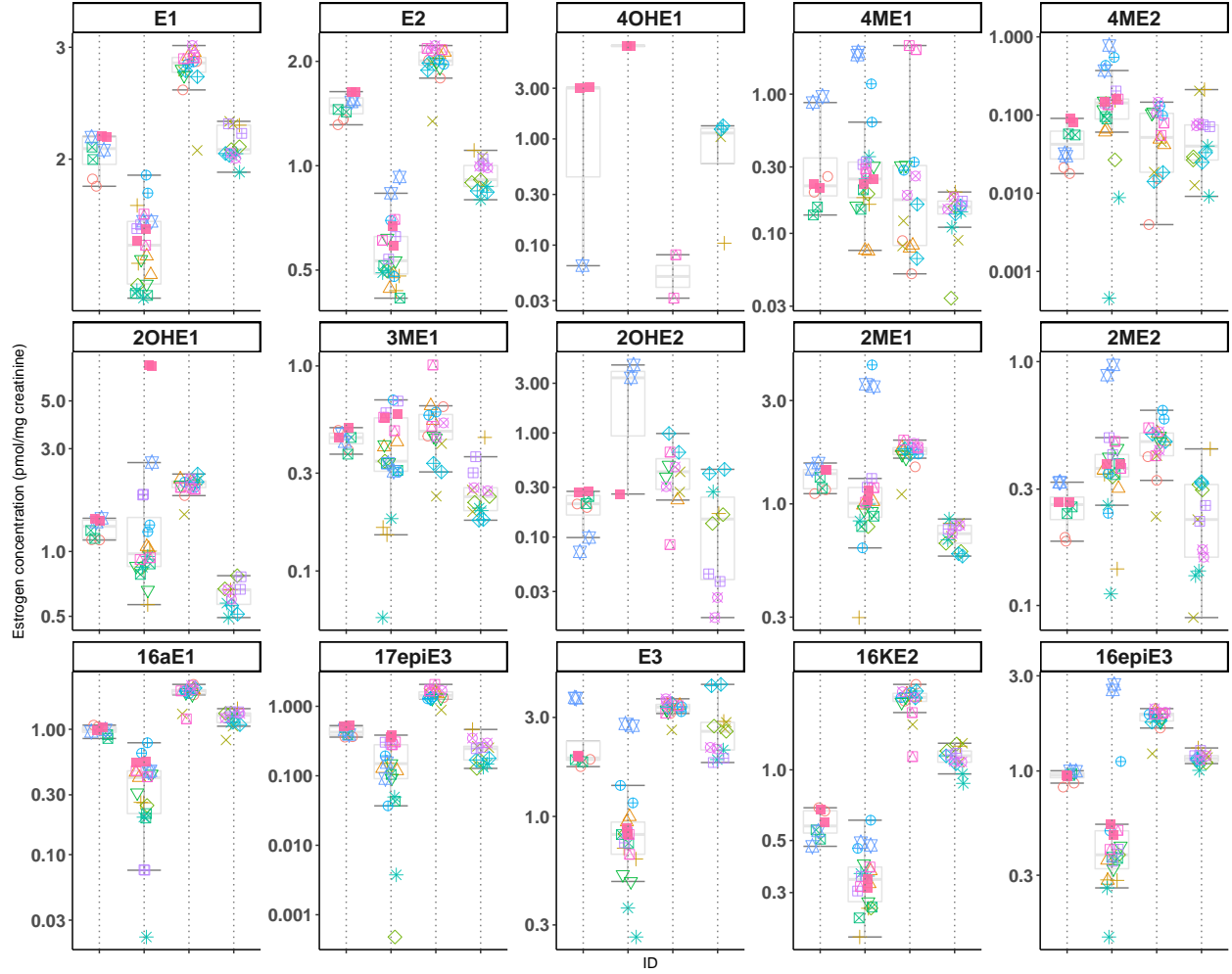


Figure 5: Four laboratory control QC sample IDs (x-axis) plotted against measured standardized concentrations. Batch identification indicated by color and shape.

The laboratory control QC samples consisted of 4 distinct urine sample types: 2 pre-menopausal ( $n=8$  and  $16$ ), 1 post-menopausal ( $n=22$ ), and 1 male urine ( $n=14$ ), totaling 60 observations. In order to assess within and between variation, four samples were allocated within in each batch, some of which had duplicate sample types in the same batch.

Most metabolites had ICCs  $\geq 75\%$  with most CVs  $\leq 50\%$ , indicating moderate reproducibility generalizable across all batches. The QC samples plotted on the log scale illustrated that a few measurements for the same sample were many fold different than their counterparts. Two batches were specifically problematic, often having measurements many fold higher than other batches across different estrogens. Concentrations of these batches were not adjusted downwards since the sample size of QC samples in these batches were not sufficient enough for justification. Within subject variation was so significant that when concentrations were classified into tertiles, many replicates were in different tertile categories.

## Korea SNU Duplicate QC Samples

Table 4: Korea SNU replicates QC calculations

Estrogen	$\sigma_{bs}^2$	$\sigma_{bb}^2$	$\sigma_{ws}^2$	Mean	ICC (%)	CV (%)
E1	0.21	0.06	0.13	1.35	52.6	32.2
E2	0.08	0.03	0.04	0.44	52.4	61.2
4OHE1	0.17	0.27	0.25	0.78	25.1	91.2
4ME1	0.25	0.06	0.1	0.63	61.6	63.2
4ME2	0.02	0	0.02	0.2	52.8	72.3
2OHE1	1.16	0.12	0.21	1.41	77.8	40.8
3ME1	0.04	0.03	0.1	0.29	24.8	121.8
2OHE2	0	0.15	0.03	0.26	0	159.2
2ME1	0.27	0	0.21	1.03	55.6	44.6
2ME2	0	0.06	0.04	0.41	3	79.6
16aE1	0.09	0.02	0.05	0.46	56.1	56.1
17epiE3	0.12	0.11	0.11	0.38	36.7	121.3
E3	0.36	1.51	0.51	1.69	15.2	84.1
16KE2	0.05	0.01	0.05	0.38	42.8	66.1
16epiE3	0.33	0.01	0.1	0.73	76.2	43.9

Note: bs = between subject, bb = between batch, ws = within subject

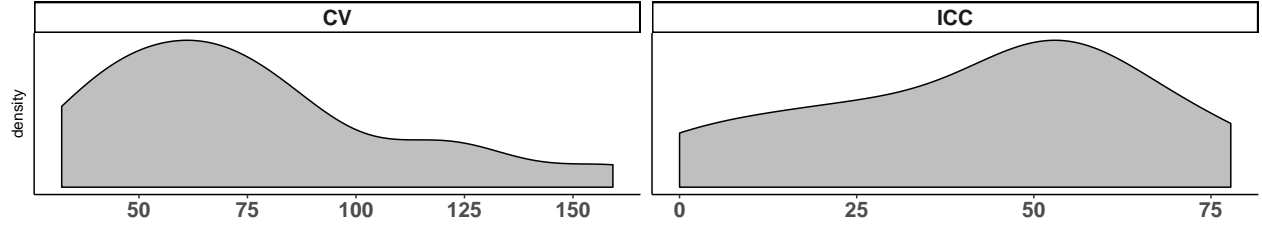


Figure 6: Distribution of CV and ICC for all metabolites for Korea SNU QC samples



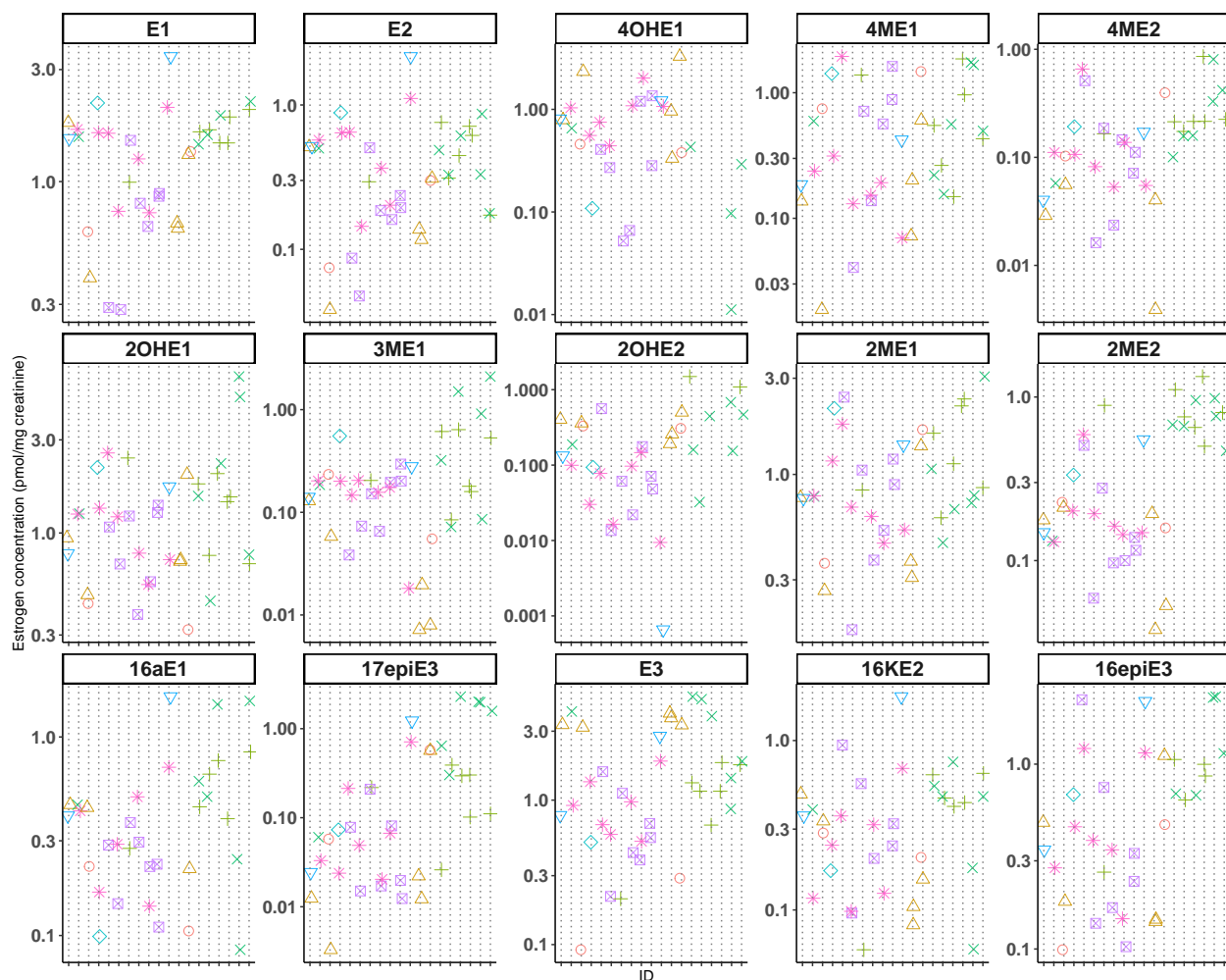


Figure 7: 19 Korea SNU QC sample IDs (x-axis) plotted against measured standardized concentrations. Batch identification indicated by color and shape.

The Korea SNU replicates consisted of 19 unique samples each with two measurements, totaling 38 observations. Most metabolites had ICCs  $\geq 50\%$  with most CVs  $\leq 60\%$ , indicating moderately inconsistent measurements. Replication results for Korea SNU duplicates was worse than over all batches assessed with laboratory control samples. Many measurements for the same sample were many fold different than their counterparts. Thus, it should be noted that metabolite measurements for Korea SNU may not be reproducible and caution is advised.

## Assessing Association between Covariates and Gastric Cancer

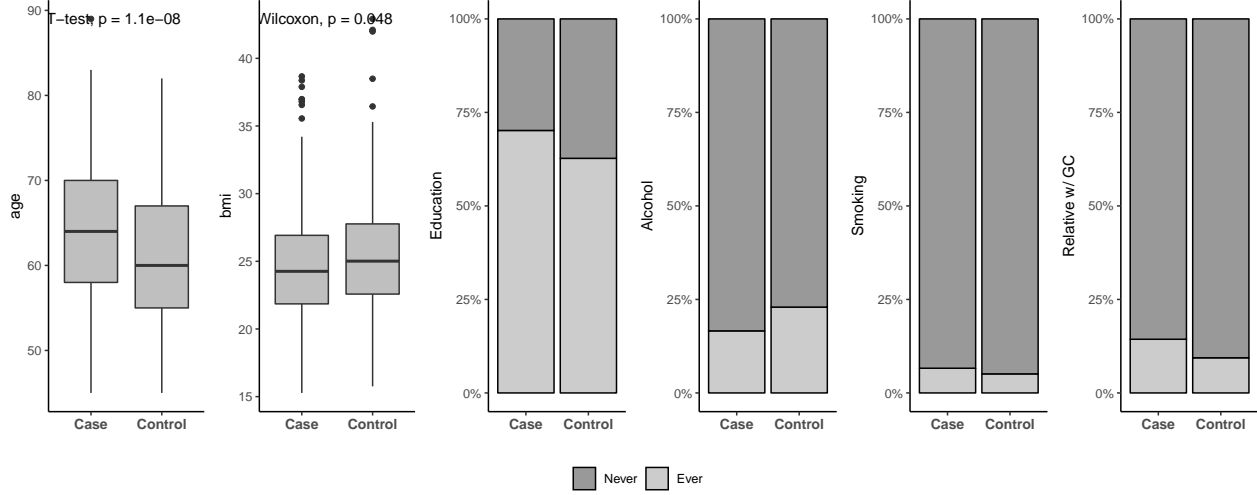


Table 5: Logistic regression comparing studies of origin to case

<i>Dependent variable:</i>	
case	
Iran	−0.633*** (−1.003, −0.263)
Korea (KMCC)	−0.622*** (−1.073, −0.170)
Germany	−0.095 (−0.952, 0.761)
Korea (SNU)	0.071 (−0.163, 0.305)
Japan	0.000 (−0.716, 0.716)
Observations	540
Log Likelihood	−364.357
Akaike Inf. Crit.	738.714

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Assessing Association between Estrogens and Gastric Cancer

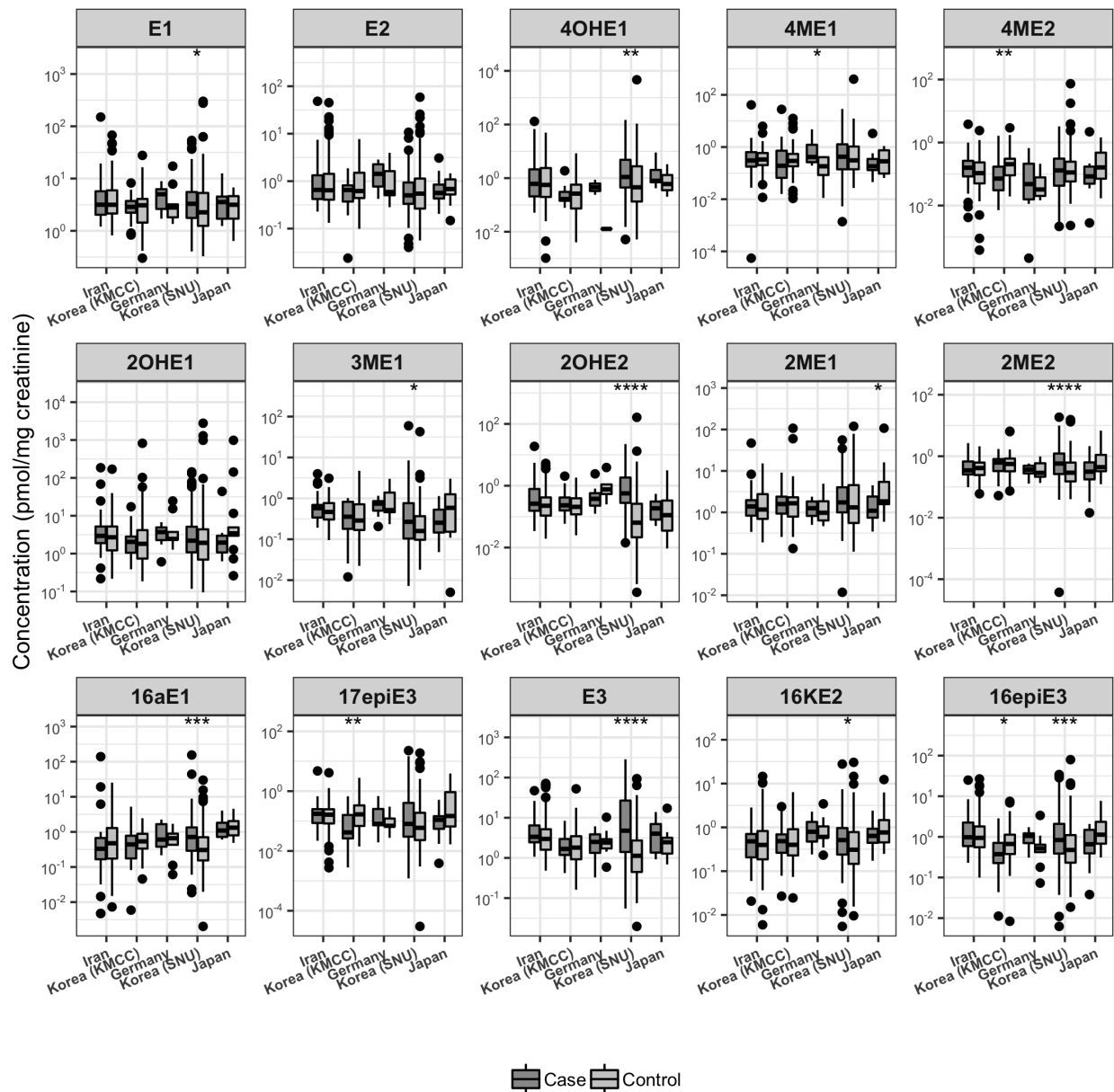


Figure 8: Study specific Wilcoxon-Mann-Whitney tests for cases vs. controls

Postmenopausal women with gastric cancer was compared with controls for both pre-diagnostic and early-stage case-control sample sets. Each study was analyzed separately to account for differences in the relationship between estrogens and gastric cancer by country of origin and study type. Each of the estrogen markers were separately analyzed to assess an association with gastric cancer. Wilcoxon-Mann-Whitney (also known as Mann-Whitney U) test was used to compare estrogen marker concentrations as a continuous value between cases and controls. Significance levels are indicated by stars above boxplots (\* :  $p \leq 0.05$ , \*\* :  $p \leq 0.01$ , \*\*\* :  $p \leq 0.001$ , \*\*\*\* :  $p \leq 0.0001$ ). According to the boxplot comparisons, over half of the metabolites had cases that had significantly higher concentrations for cases than controls at the  $\alpha=0.05$  level for Korea SNU. However, the three metabolites that were significant for Korea KMCC (4ME2, 17epiE3, 16epiE3) showed the

opposite association (i.e. controls had significantly higher levels of these metabolites compared to cases).

Table 6: Estrogen concentration median and tertile cutpoint

	Cohort			Case Control	
	Iran	Korea (KMCC)	Germany	Korea (SNU)	Japan
	Median (Tertile)	Median (Tertile)	Median (Tertile)	Median (Tertile)	Median (Tertile)
Estrone (E1)	3.15 (2.37, 4.31)	3.08 (1.68, 3.62)	2.79 (1.97, 3.11)	2.26 (1.53, 3.95)	3.16 (2.31, 4.21)
Estradiol (E2)	0.64 (0.49, 1.02)	0.62 (0.46, 1.21)	0.59 (0.53, 1.03)	0.55 (0.33, 0.83)	0.69 (0.56, 0.82)
4-Hydroxyestrone (4OHE1)	0.56 (0.36, 1.38)	0.26 (0.12, 0.4)	0.01 (0.01, 0.01)	0.45 (0.23, 1.24)	0.68 (0.36, 1.13)
4-Methoxyestrone (4ME1)	0.33 (0.24, 0.47)	0.29 (0.22, 0.41)	0.18 (0.14, 0.29)	0.31 (0.19, 0.69)	0.28 (0.13, 0.73)
4-Methoxyestradiol (4ME2)	0.11 (0.07, 0.18)	0.2 (0.12, 0.24)	0.03 (0.02, 0.06)	0.11 (0.07, 0.18)	0.18 (0.08, 0.38)
2-Hydroxyestrone (2OHE1)	2.7 (1.51, 4.26)	1.81 (1.06, 3.29)	2.44 (2.38, 2.9)	1.93 (0.9, 2.92)	3.43 (3.04, 4.67)
2-Hydroxyestrone-3-methyl ether (3ME1)	0.46 (0.36, 0.63)	0.29 (0.22, 0.62)	0.52 (0.46, 0.89)	0.16 (0.11, 0.25)	0.6 (0.33, 1.11)
2-Hydroxyestradiol (2OHE2)	0.22 (0.13, 0.36)	0.21 (0.14, 0.32)	0.78 (0.58, 1.05)	0.06 (0.03, 0.14)	0.12 (0.05, 0.2)
2-Methoxyestrone (2ME1)	1.17 (0.86, 1.93)	1.66 (0.93, 2.13)	0.98 (0.78, 1.49)	1.32 (0.71, 2.68)	1.84 (1.7, 3.64)
2-Methoxyestradiol (2ME2)	0.41 (0.29, 0.55)	0.56 (0.45, 0.65)	0.29 (0.22, 0.44)	0.29 (0.19, 0.45)	0.44 (0.37, 0.94)
16a-Hydroxyestrone (16aE1)	0.48 (0.29, 0.87)	0.54 (0.43, 0.66)	0.66 (0.55, 0.77)	0.31 (0.2, 0.51)	1.32 (0.82, 1.57)
17-Epiestriol (17epiE3)	0.16 (0.11, 0.22)	0.17 (0.1, 0.24)	0.07 (0.07, 0.12)	0.06 (0.03, 0.13)	0.15 (0.07, 0.21)
Estriol (E3)	2.97 (2.25, 4.68)	1.81 (1.13, 2.63)	2.41 (1.99, 2.79)	1.13 (0.74, 2.1)	2.46 (1.57, 3)
16-Ketoestradiol (16KE2)	0.4 (0.27, 0.65)	0.4 (0.25, 0.52)	0.6 (0.59, 0.84)	0.31 (0.2, 0.59)	0.76 (0.52, 0.84)
16-Epiestriol (16epiE3)	0.94 (0.72, 1.37)	0.67 (0.48, 0.94)	0.52 (0.46, 0.58)	0.48 (0.3, 0.81)	1.14 (0.88, 2.1)

*Note:*

Tertile (33% and 66% quantiles) determined by control subjects

Units in pmol/mg creatinine

Study-specific estrogen marker tertiles were defined by the concentration distribution in controls to account for the extremely skewed distributions and improve interpretability. Tertile categories were treated as a numeric entry rather than a factor type to assess overall linear associations (i.e.  $\beta$  coefficient for metabolite markers interpreted as risk difference from one tertile to the next) Changing the number of quantiles did not significantly affect results (i.e. splitting concentrations into quartiles or quintiles did not change linear associations by much). Korea SNU appeared to have lower estrogen concentrations than other studies, especially for the metabolites with high concentrations (E1, 2OHE1, E3).

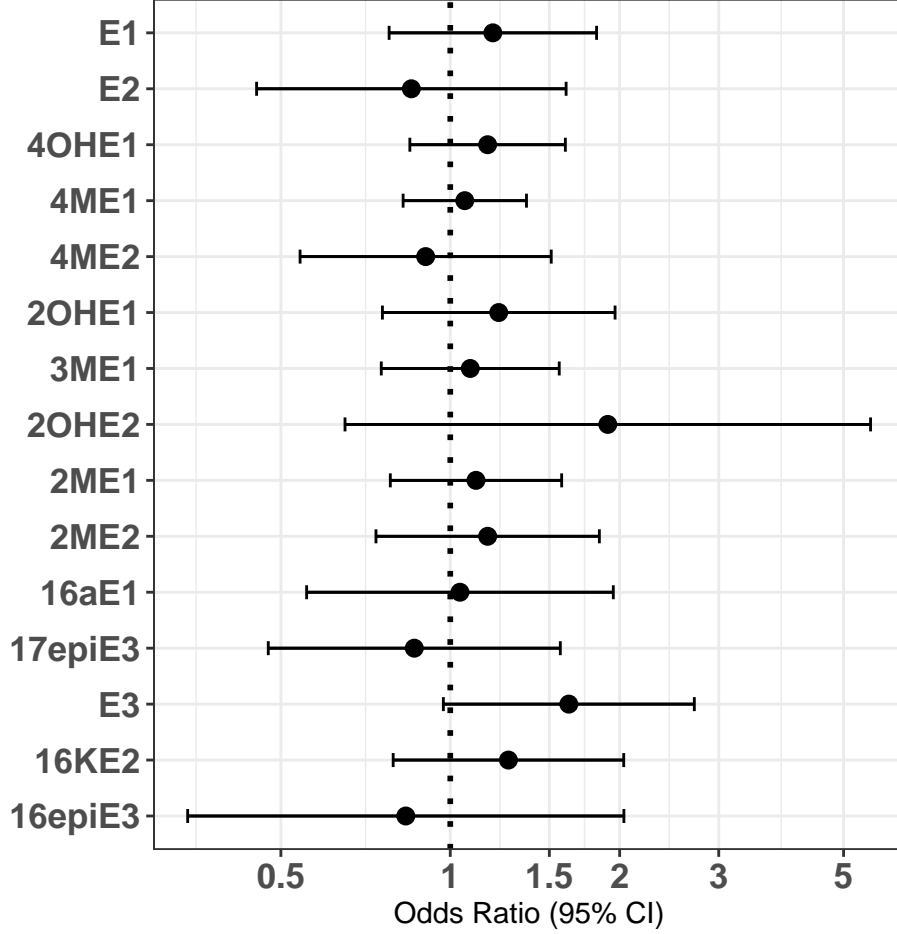


Figure 9: Forestplot of fully adjusted pooled random effect meta-analysis logistic regression model

Odds ratios (OR) and corresponding 95% confidence intervals (CI) were calculated using random effects maximum likelihood for each estrogen marker using multivariable unconditional logistic regression adjusted for age, body mass index, smoking, alcohol, education, and family history of gastric cancer as available. All studies could not be adjusted for all covariates due to . The model for Japan did not include education and Korea (KMCC) did not include family history of gastric cancer due to data not being collected for these covariates. # Iran = age + bmi, education # korea (KMCC) = age, bmi, smoking, education # Germany = age, bmi # Korea (SNU) = age, bmi, smoking, education, relativegc # Japan = age, bmi, relativegc

Multivariate logistic model used for each study:

- Iran:  $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 E_i + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3$
- Korea (KMCC):  $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 E_i + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_5 X_4$
- Germany:  $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 E_i + \beta_2 X_1 + \beta_3 X_2$
- Korea (SNU):  $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 E_i + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_5 X_4 + \beta_6 X_5$
- Japan:  $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 E_i + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_5$

Where  $\pi = Pr(Y = 1|E, X)$ ,  $E_i$  = metabolites 1-15 as numeric tertile entries,  $X_1$  = age (years),  $X_2$  = BMI,  $X_3$  = any educational degree,  $X_4$  = ever smoked, and  $X_5$  = relative with gastric cancer.

ORs were pooled by random effect meta-analysis separately for cohort and case-control studies to assess design-specific effect differences and then were combined to assess the overall associations. Markers with

statistically significant ordinal trends ( $p < 0.05$ ) were further examined for associations of each tertile with gastric cancer risk comparing to the lowest tertile. All p-values were two-sided. Given the exploratory nature of our study p-values were not corrected for multiple comparisons. All statistical analyses and data visualizations were performed using R package “metafor” for meta-analysis and ggplot2.

## Additional Analysis

**Table 1: Comparison of median estrogen metabolite concentrations reported from each study**

<i>Estrogen</i>	<i>Nomenclature</i>	<i>Constanza et al. 2018</i>	<i>Moore et al. 2016 (% Difference) [Absolute Difference]</i>	<i>Sampson et al. 2016 (% Difference) [Absolute Difference]</i>
E1	Estrone	3.02	2.04 (68%) [-0.98]	5.8 (192%) [2.78]
E2	Estradiol	0.57	0.42 (74%) [-0.15]	1.2 (211%) [0.63]
4OHE1	4-Hydroxyestrone	0.55	0.22 (40%) [-0.33]	0.7 (127%) [0.15]
4ME1	4-Methoxyestrone	0.33	0.05 (15%) [-0.28]	0.1 (30%) [-0.23]
4ME2	4-Methoxyestradiol	0.12	0.02 (17%) [-0.1]	0.1 (83%) [-0.02]
2OHE1	2-Hydroxyestrone	2.33	1.62 (70%) [-0.71]	4.9 (210%) [2.57]
3ME1	2-Hydroxyestrone-3-methyl ether	0.32	0.09 (28%) [-0.23]	0.3 (94%) [-0.02]
2OHE2	2-Hydroxyestradiol	0.24	0.38 (158%) [0.14]	1.2 (500%) [0.96]
2ME1	2-Methoxyestrone	1.43	0.36 (25%) [-1.07]	1.1 (77%) [-0.33]
2ME2	2-Methoxyestradiol	0.44	0.16 (36%) [-0.28]	0.5 (114%) [0.06]
16aE1	16a-Hydroxyestrone	0.53	0.49 (92%) [-0.04]	1.5 (283%) [0.97]
17epiE3	17-Epiestriol	0.11	0.12 (109%) [0.01]	0.4 (364%) [0.29]
E3	Estriol	2.42	1.83 (76%) [-0.59]	5.5 (227%) [3.08]
16KE2	16-Ketoestradiol	0.46	0.53 (115%) [0.07]	1.6 (348%) [1.14]
16epiE3	16-Epiestriol	0.72	0.2 (28%) [-0.52]	0.6 (83%) [-0.12]

*Note: Units in pmol/mg creatinine.*

Standardized estrogen concentrations agreed with Moore et al., 2016 and Sampson et al., 2016 which used similar datasets of the same demographis (Comparison table found in Other Analysis section below).

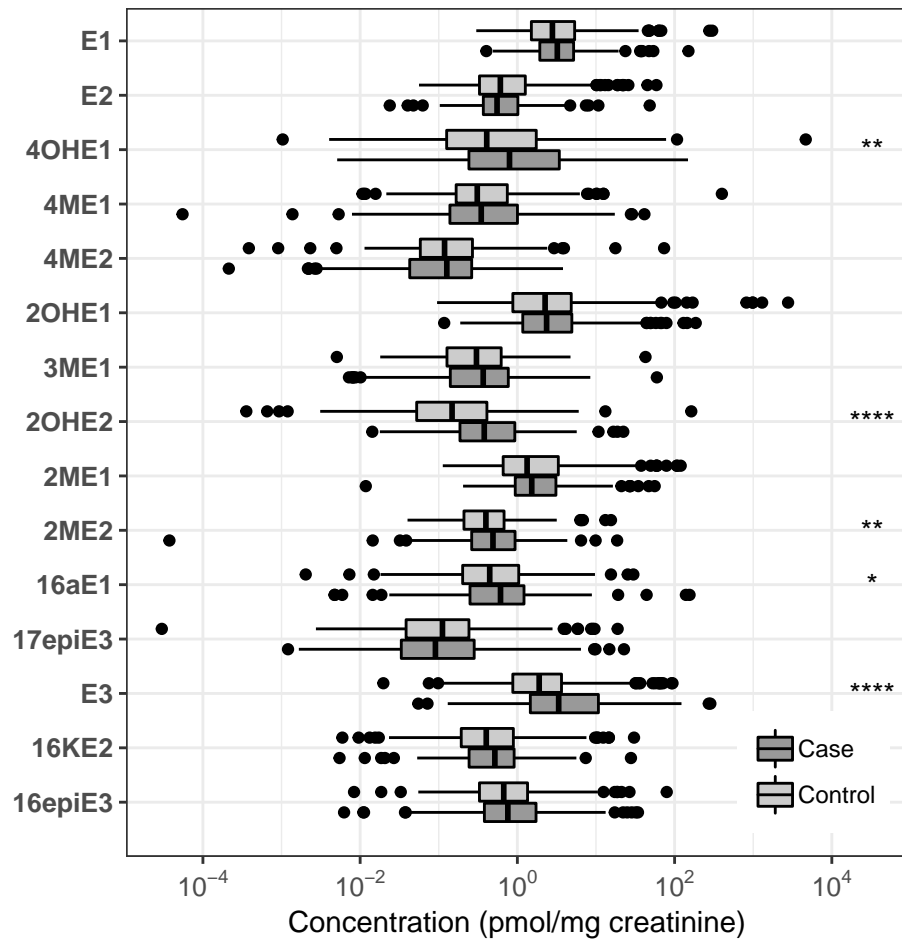


Figure 10: Pooled Wilcoxon-Mann-Whitney tests for cases vs. controls