Supplemental Information for *Bayesian Inference of Chemical Exposures from NHANES Urine Biomonitoring Data*

Zachary Stanfield1, R. Woodrow Setzer1, Victoria Hull1,2, Risa R. Sayre1, Kristin K. Isaacs1, John F. Wambaugh1,\*

1. Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 27711, United States
2. Oak Ridge Associated Universities (ORAU), Oak Ridge, Tennessee, 37830, United States

\* Corresponding Author

#

# JAGS Code for Bayesian Model

Below is the JAGS code used in the fitOnlyP function of the R package bayesmarker. The package can also be found on GitHub (github.com/USEPA/CompTox-HumanExposure-bayesmarker).

bayes\_model <- "

var mixmu[Nn], mixtau[Nn], mixpi[Nn];

data

{

# Turning the input standard errors into precisions for the estimated log geometric mean product

# concentrations

 for (i in 1:Mn)

 {

 tau.se[i] <- 1/(se[i]\*se[i])

 }

}

model

{

 ## +++++++++++++++++++++++++++++++++++++++++++++

 ## The heart of the model:

 ## Model the parent exposures

 ## +++++++++++++++++++++++++++++++++++++++++++++

 for (i in 1:N) {

 ## log of product of production volume and unit exposure

 lP[i] ~ dnorm(lPmu, tau.V)

 P[i]<- exp(lP[i])

 }

 ## hyperparameter for lP

 lPmu ~ dnorm(0, 0.001)

 ## code so that prior for sd(log(P[i])) is half-Cauchy(25)

 sd.dum ~ dt(0,1/625,1)

 sd.V <- abs(sd.dum)

 tau.V <- 1/pow(sd.V,2)

 ## ++++++++++++++++++++++++++++++++++++++++++++++

 ## Link the unobserved parent exposure to the observed

 ## (but censored) metabolite (w/some parent) urinary

 ## values.

 ##

 ## Model metabolite concentrations as (usually) unknown fractions

 ## of the parent.

 ##

 ## Phi[i,j] say what fraction of parent[i] goes to metabolite[j]

 ## observed values are lognormally distributed around U, with known

 ## precision, but censored. Some of the elements of Phi are known to be

 ## 0 or 1 (thanks to Jim Rabinowitz for some of the 1's).

 lU <- log(t(Phi) %\*% P)

 ## tau.se <- 1/(se \* se)

 for (j in 1:Mn) {

 ly[j] ~ dnorm(lU[j],tau.se[j])

 }

 ## for j > Mn, what we have is the number of observations > LOD, and the total sample size.

 ## so let Pralod[j] <- 1 - pnorm(lu[j], sd[j - Mn])

 ## Then ly[j] <- number above lod, and is binomial with mean Pralod[j] and n SS[j]

 for (j in (Mn+1):M) {

 Pralod[j - Mn] <- 1 - pnorm(lod[j - Mn], lU[j], tau[j - Mn])

 ly[j] ~ dbin(Pralod[j - Mn], SS[j - Mn])

 }

 ## The population lsds look to be mixtures of normal distributions for the data

 ## where moments seem well-estimated.

 ## Estimate mixmu, mixtau, mixpi externally, and input as data.

 for (j in 1:(M - Mn)) {

 lsd[j] ~ dnormmix(mixmu, mixtau, mixpi)

 tau[j] <- exp(-2\*lsd[j])

 }

 for (i in 1:NBranches)

 {

 phi[Bstart[i]:Bstop[i]] ~ ddirch(Alpha[Bstart[i]:Bstop[i]])

 }

 for (i in 1:Ndelta) {

 Phi[indx[i,1],indx[i,2]] <- phi[i]

 }

}

"

# Supplemental Figures



**Figure S1**. Depiction of general workflow for the bayesmarker R package. The input dataset is a single Excel file that has 3 sheets (Tables S1-S3; described in the purple box). These tables are created manually based on which metabolites/cohorts are of interest and how the biomonitoring data is organized, here showing National Health and Nutrition Examination Survey (NHANES). The NHANES survey weights are needed to translate from their oversampling procedure to represent the modern U.S. population. The package has 5 major functions (descriptions in green box). The final output is exposure estimates in mg/kg bodyweight/day of the parent chemicals of the measured metabolites.



**Figure S2.** Visualization of the influence of various potential contributors to uncertainty on inferred exposure estimates from urine biomonitoring data. Each data point represents a parent chemical in a single NHANES cohort for each of its metabolites measured in that cohort (so all unique parent-cohort-metabolite triplets). Uncertainty (on the y-axis) is the order of magnitude spanned by the 95% credible interval around the median exposure estimate (in other words, log10(upper 95% CI/lower 95% CI) = uncertainty), and the fraction below the LOD (on the x-axis) represents the number of measurements below the LOD divided by the total number of measurements for the relevant metabolite. Chemicals are stratified by chemical class, points are colored by NHANES cohort, point size represents the number of parent chemicals shared by the relevant metabolite, and the point shape indicates whether or not a chemical was directly measured in NHANES (in other words, a urine concentration was obtained by NHANES for that parent chemical).