**Supporting Information:**

Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways

Caroline L. Ring1,3, Jon A. Arnot4,5,6, Deborah H. Bennett7, Peter P. Egeghy2, Peter Fantke8, Lei Huang9, Kristin K. Isaacs2, Olivier Jolliet9, Katherine A. Phillips2, Paul S. Price2, Hyeong-Moo Shin10, John N. Westgate4, R. Woodrow Setzer1, John F. Wambaugh1,\*

1National Center for Computational Toxicology, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina 27711

2National Exposure Research Laboratory, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina 27711

3Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee 37831

4ARC Arnot Research and Consulting, 36 Sproat Ave. Toronto, ON, Canada, M4M 1W4

5Department of Physical & Environmental Sciences, University of Toronto Scarborough

1265 Military Trail, Toronto, ON, Canada, M1C 1A4

6Department of Pharmacology and Toxicology, University of Toronto, 1 King's College Cir, Toronto, ON, Canada, M5S 1A8

7 Department of Public Health Sciences, University of California, Davis, California, 95616

8Quantitative Sustainability Assessment Division, Department of Management Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

9Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109

10Department of Earth and Environmental Sciences, University of Texas, Arlington, Texas, 76019

\*Corresponding Author: John F. Wambaugh

109 T.W. Alexander Dr., NC 27711, USA

Wambaugh.John@epa.gov

Phone: (919) 541-7641

**The supplemental materials accompanying this manuscript consist of this eight-page document, two figures, seven tables, and an appendix with the Bayesian inference model used with JAGS. Four of the tables are contained in separate files.**

# Supplemental Tables

**Table S1**: All predictors and predictions for chemicals with structures in the DSStox library (separate file: SupTable-all.chem.preds-2018-10-09.zip)

**Table S2**: Assignments of source-based aggregate pathways for chemicals monitored by the NHANES program (separate file: SupTable2-NHANES-ManualPathway-022218.xlsx )

**Table S3**: Persistent organic pollutants identified by the Stockholm Convention (separate file: SupTable3-StockholmConvention-011218.xlsx)

**Table S4**: Intake Rate Inferences from the NHANES biomonitoring data used to evaluate models. (separate file: SupTable4-NHANES-ExposureInferrences-2018-04-10.xlsx)

**Table S5**: Top twenty-five most “important” chemical descriptors. The most important descriptors for each model was characterized with Gini impurity importance1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Normalized Gini impurity importance1 | | | |
|  | Dietary | Near-Field | Far-Field Pesticide | Far Field Industrial |
| NCCT\_LogKAW | 1.00 | 0.88 | 1.00 | 1.00 |
| NCCT\_VP | 0.84 | 1.00 | 0.99 | 0.83 |
| NCCT\_MP | 0.94 | 0.95 | 0.89 | 0.81 |
| NCCT\_LogKOA | 0.85 | 0.89 | 0.90 | 0.89 |
| Structure\_MolWt | 0.86 | 0.89 | 0.91 | 0.69 |
| NCCT\_BP | 0.79 | 0.79 | 0.92 | 0.74 |
| NCCT\_HL | 0.72 | 0.69 | 0.87 | 0.58 |
| NCCT\_BIODEG | 0.74 | 0.53 | 0.85 | 0.65 |
| NCCT\_KOC | 0.72 | 0.60 | 0.88 | 0.48 |
| NCCT\_LogP | 0.73 | 0.58 | 0.80 | 0.50 |
| NCCT\_Csatw | 0.72 | 0.56 | 0.79 | 0.52 |
| NCCT\_AOH | 0.69 | 0.54 | 0.82 | 0.51 |
| NCCT\_WS | 0.69 | 0.54 | 0.80 | 0.53 |
| NCCT\_BCF | 0.69 | 0.56 | 0.79 | 0.46 |
| NCCT\_KM | 0.69 | 0.50 | 0.79 | 0.45 |
| bond.X.any.\_halide | 0.25 | 0.15 | 0.24 | 0.13 |
| NCCT\_RBiodeg | 0.28 | 0.18 | 0.13 | 0.14 |
| bond.CX\_halide\_aromatic.X\_generic | 0.16 | 0.10 | 0.25 | 0.15 |
| bond.C..O.N\_carboxamide\_generic | 0.13 | 0.10 | 0.11 | 0.20 |
| bond.CN\_amine\_aliphatic\_generic | 0.21 | 0.05 | 0.18 | 0.09 |
| bond.NC.O\_aminocarbonyl\_generic | 0.13 | 0.10 | 0.11 | 0.20 |
| ring.hetero\_.6.\_Z\_generic | 0.14 | 0.08 | 0.09 | 0.20 |
| ring.aromatic\_benzene | 0.15 | 0.09 | 0.12 | 0.14 |
| chain.alkaneLinear\_ethyl\_C2\_.connect\_noZ\_CN.4. | 0.14 | 0.12 | 0.10 | 0.06 |
| bond.C.O\_carbonyl\_generic | 0.11 | 0.09 | 0.11 | 0.10 |

**Table S6**: Median parameter estimates from multivariate regression. The median standard deviation are reported. If the median is roughly two standard deviations away from zero then the predictor has a statistically significant multivariate association (value indicated in bold).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Grand Mean (Unexplained)** | **Dietary** | **Residential** | **Far-Field Pesticide** | **Far-Field Industrial** |
| Pathway Mean | -0.291 (0.319) | **0.483 (0.292)** | **0.888 (0.26)** | **0.346 (0.302)** | -0.104 (0.228) |
| NHANES Chemicals | 0 | 22 | 45 | 88 | 34 |
| All Chemicals | 86.9% | 1.22% | 4.68% | 1.58% | 9.89% |
| SHEDS Direct |  |  | **0.187 (0.0635)** |  |  |
| SHEDS Indirect |  |  | 0.0405 (0.0688) |  |  |
| FINE |  |  | 0.0159 (0.0496) |  |  |
| Food Contact |  | **0.378 (0.134)** |  |  |  |
| REDS |  |  |  | 0.0287 (0.144) |  |
| RAIDAR |  |  |  | **-0.119 (0.0959)** | **-0.296 (0.142)** |
| RAIDAR.ICE |  |  | -0.0991 (0.161) |  |  |
| USETox Pest |  |  |  | **0.129 (0.0631)** |  |
| USETox Indust |  |  |  |  | **-0.29 (0.135)** |
| USETox Res |  |  | -0.0167 (0.117) |  |  |
| USETox Diet |  | **-0.599 (0.169)** |  |  |  |
| Production.Volume |  | 0.459 (0.252) | -0.152 (0.198) | **0.383 (0.126)** | -0.093 (0.162) |
| Stockholm |  |  |  | **-1.48 (0.256)** | **-1.94 (0.462)** |

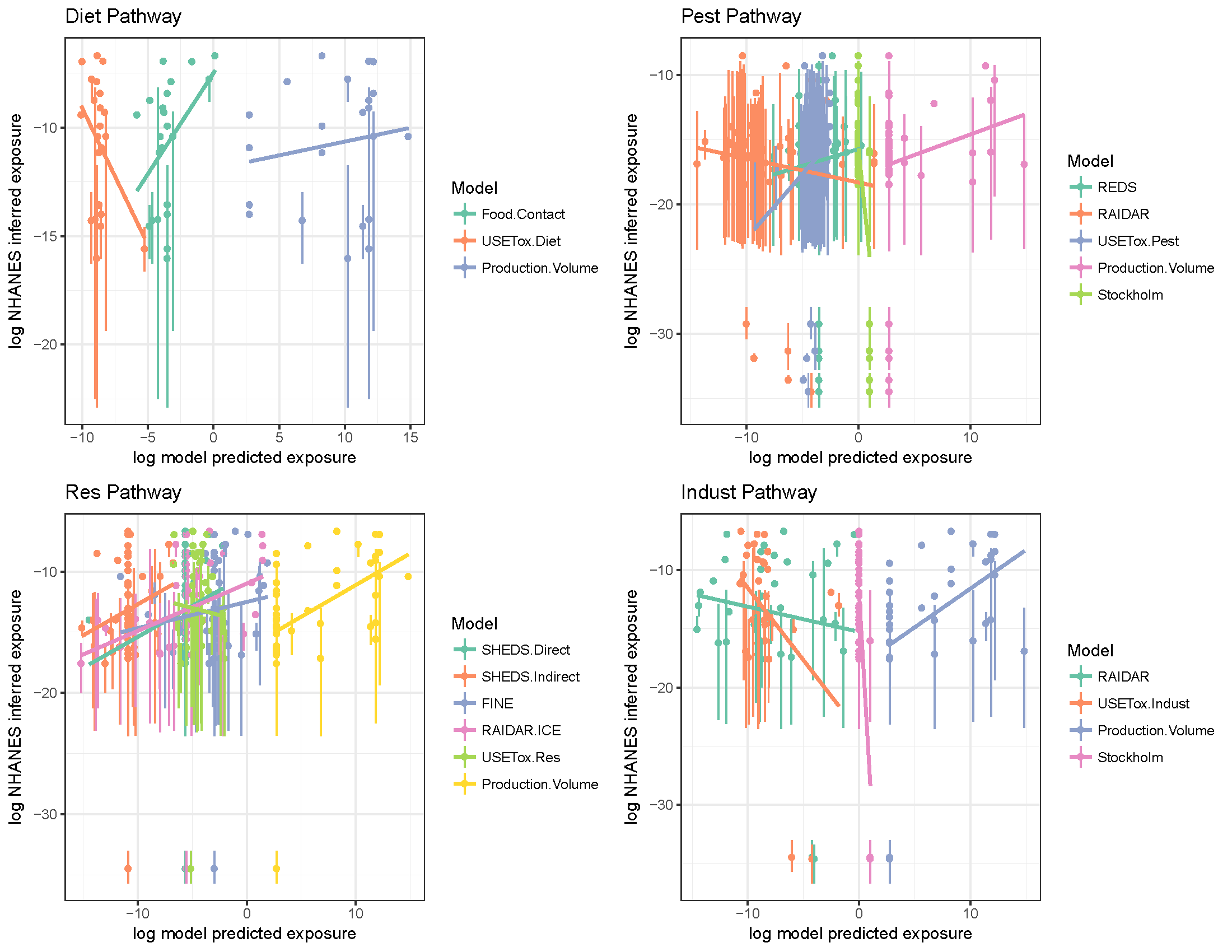
**Table S7**: Summary of univariate association of predictors with inferred intake rates for chemicals predicted by each model. The number of evaluation chemicals are the NHANES chemicals with both model predictions and are believed to correspond to the relevant pathway. The median and lower/upper limits of the 95% credible interval are reported. The column “Non-Zero” indicates whether the credible interval does not include zero and is therefore the model has a statistically significant univariate association.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Pathway | Model | Evaluation.Chems | Univariate | Lower95 | Upper95 | NonZero |
| Diet | Food.Contact | 17 | 0.28 | 0.05 | 0.53 | Y |
| Diet | USETox.Diet | 20 | -0.30 | -0.60 | 0.00 | Y |
| Diet | Production.Volume | 22 | 0.12 | -0.13 | 0.36 |  |
| Res | SHEDS.Direct | 32 | 0.18 | 0.05 | 0.31 | Y |
| Res | SHEDS.Indirect | 28 | 0.03 | -0.13 | 0.17 |  |
| Res | FINE | 43 | 0.10 | -0.03 | 0.22 |  |
| Res | RAIDAR.ICE | 29 | 0.40 | 0.14 | 0.65 | Y |
| Res | USETox.Res | 43 | -0.05 | -0.32 | 0.22 |  |
| Res | Production.Volume | 45 | 0.37 | 0.14 | 0.59 | Y |
| Pest | REDS | 22 | 0.18 | -0.13 | 0.48 |  |
| Pest | RAIDAR | 86 | -0.12 | -0.35 | 0.11 |  |
| Pest | USETox.Pest | 86 | 0.15 | -0.04 | 0.33 |  |
| Pest | Production.Volume | 88 | 0.23 | -0.02 | 0.49 |  |
| Pest | Stockholm | 88 | -1.77 | -2.32 | -1.21 | Y |
| Indust | RAIDAR | 32 | -0.14 | -0.62 | 0.34 |  |
| Indust | USETox.Indust | 32 | -0.67 | -1.11 | -0.24 | Y |
| Indust | Production.Volume | 34 | 0.51 | 0.14 | 0.87 | Y |
| Indust | Stockholm | 34 | -3.70 | -4.76 | -2.64 | Y |

# Supplemental Figures



**Figure S1**: The geometric mean intake rates inferred from NHANES urine2, 3 and serum 4, 5. Plot points indicate the geometric mean. The 95% confidence intervals indicated by horizontal lines depend upon three factors: 1) the distribution of the individuals in the population, 2) any degeneracy in the relationship between urine metabolites and parent chemicals (i.e., multiple metabolites mapping to multiple parents), and 3) the number of samples above the limit of quantitation.

****

**Figure S2**: Univariate comparison of model predictions vs. NHANES inferred intake rates on a per pathway basis.

# Appendix: JAGS Model

# obs contains the inferred mean log exposure rates for Nchem chemicals

# from Wambaugh et al. (2014)

# obs.sd contains the inferred standard deviation on log scale

# obs.lod contains the limit of detection on exposure rate inferrence

#

# a[Npath] is the average (log) exposure rate for a pathway

# a0 is the average exposure unexplained by the pathways analyzed

#

# wmatrix[Nweights] is the weight for a given combination of model and pathway

# wmodel[Nweights] identifies which model a given weight corresponds to

# wpathway[Newights] identifies a pathway a given weight correpsonds to

# w[Nmodel,Npath] is a matrix of mostly zeros, nonzero entries correspond to

# models that participate in a given pathway

#

# deltaprior[Nchem,Npath] is a matrix of probabilities that a chemical is subject

# to a given pathway (precomputed via method of Random Forests)

# delta[Nchem,Npath] is a Boolean matrix indicating whether or not a given

# chemical has exposure via a given pathway

#

# sigma[Npath] is a pathway-specific estimate of standard deviation

#

var R2, consensus[Nchem], a[Npath], wvector[Nnonzerow], means[Nchem,Npath], weightedpreds[Nchem, Nnonzerow], summean[Nchem], sumweightedpreds[Nchem], delta[Nchem, Npath]

model {

for (chem in 1:Nchem){

isAboveLOD[chem] ~ dinterval(exp(logobs[chem]), LOD[chem])

consensus[chem] <- a0 + summeans[chem]+sumweightedpreds[chem]

logobs[chem] ~ dnorm(consensus[chem], 1/(pow(sigma0,2)+pow(sigma[chem],2)))

sigma[chem] ~ dunif(0,100) #For missing/censored data only

for (path in 1:Npath){

# delta[chem, path] ~ dbern(deltap[chem, path])

# deltap[chem, path] ~ dbeta(alpha[chem, path], beta[chem, path])

means[chem,path] <- delta[chem, path]\*a[path]

}

for (i in 1:Nnonzerow)

{

weightedpreds[chem,i] <- delta[chem, wpath[i]]\*wvector[i]\*pred[chem,wmdl[i]] #model predicted exposures

}

summeans[chem] <- sum(means[chem,])

sumweightedpreds[chem] <- sum(weightedpreds[chem,])

} #end loop over chemicals

a ~ dmnorm(mu\_a, omega\_a)

# sample the model weights:

wvector ~ dmnorm(mu\_w, omega\_w)

#Calculate Bayesian R2 (Gelman, 2017) as we go along

meanconsensus <- mean(consensus)

meanresidual <- mean(logobs-consensus)

R2 <- sum(pow(consensus-meanconsensus,2)) / (sum(pow(consensus-meanconsensus,2)) + sum(pow(logobs-consensus-meanresidual,2)))

#Priors

sigma0 ~ dunif(0,100)

a0 ~ dnorm(0, 1/pow(a0\_sd,2))

#Hyperpriors

mu\_a ~ dmnorm(mu\_a\_mu, mu\_a\_omega)

omega\_a ~ dwish(omega\_a\_R, omega\_a\_k)

mu\_w ~ dmnorm(mu\_w\_mu, mu\_w\_omega)

omega\_w ~ dwish(omega\_w\_R, omega\_w\_k)

}

# References

1. Louppe, G.; Wehenkel, L.; Sutera, A.; Geurts, P. In *Understanding variable importances in forests of randomized trees*, Advances in neural information processing systems, 2013; 2013; pp 431-439.

2. Wambaugh, J. F.; Wang, A.; Dionisio, K. L.; Frame, A.; Egeghy, P.; Judson, R.; Setzer, R. W., High throughput heuristics for prioritizing human exposure to environmental chemicals. *Environmental Science and Technology* **2014,** *48*, (21), 12760-7.

3. Ring, C. L.; Pearce, R. G.; Setzer, R. W.; Wetmore, B. A.; Wambaugh, J. F., Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. *Environment International* **2017,** *106*, 105-118.

4. CDC National Health and Nutrition Examination Survey. http://www.cdc.gov/nchs/nhanes.htm

5. Pearce, R. G.; Setzer, R. W.; Strope, C. L.; Sipes, N. S.; Wambaugh, J. F., Httk: R package for high-throughput toxicokinetics. *Journal of Statistical Software* **2017,** *79*, (1), 1-26.

Disclaimer

The views expressed in this publication are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Reference to commercial products or services does not constitute endorsement.