Predicting Indoor Air and Dust Concentrations and Interpreting Measured Urinary Biomarkers

Mary Jacketti,[1,2] Robin Dodson,[3] Ruthann Rudel,[3] John Wambaugh,[2] Kristin Isaacs[2]

[1]Oak Ridge Institute for Science and Education (ORISE)
[2]U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure
[3]Silent Spring Institute

Society of Toxicology Annual Meeting
Enhancing Toxicology with Machine Learning
March 21, 2023

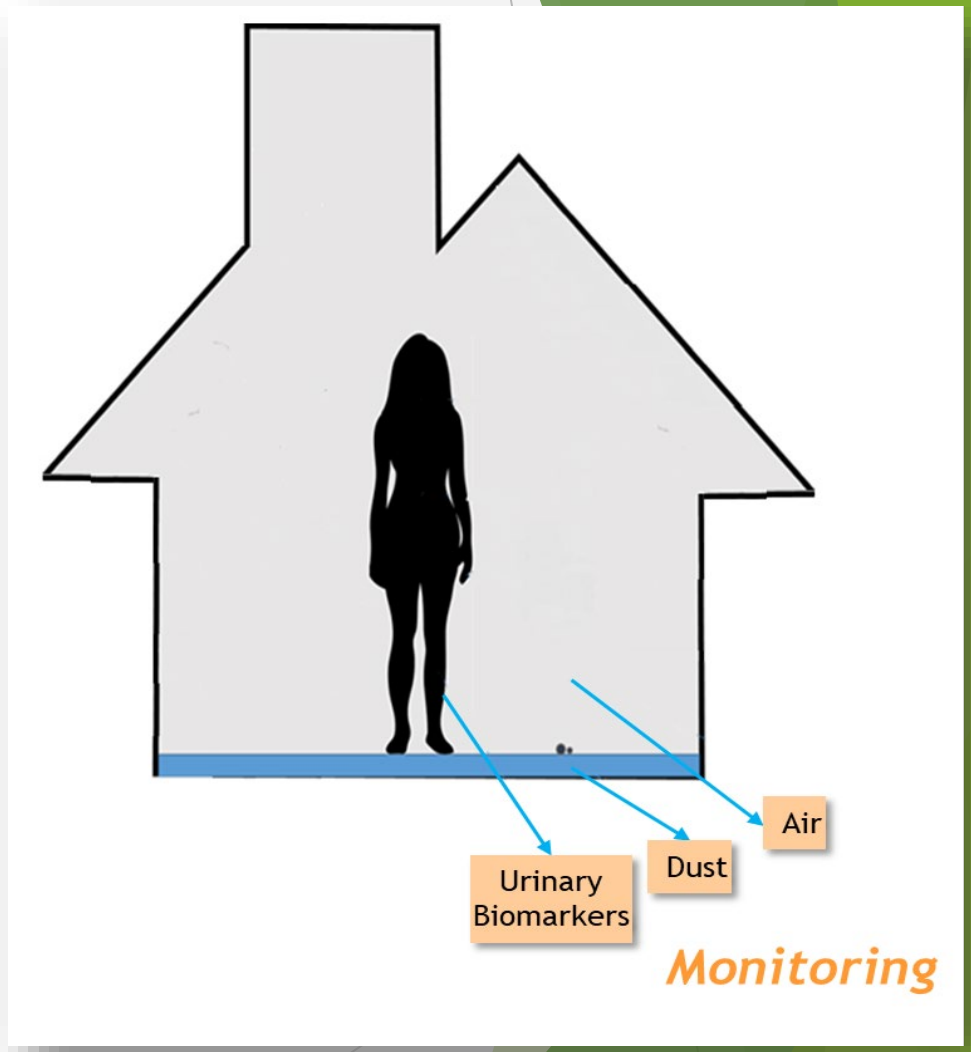
Office of Research and Development
Center for Computational Toxicology and Exposure

Disclaimer and Conflict of Interest Statement

- The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA
- The authors declare no conflict of interest

Introduction

- Exposure to many chemicals found indoors may cause potential risks to human health
- Many are identified as potential endocrine-disrupting compounds (EDCs)
- Measurement studies have collected environmental and biomarker data for potential EDCs, but there have been challenges in interpreting these data to derive meaningful conclusions regarding specific exposure pathways, routes, or processes
  - Gaps in exposure source or use information for measured chemicals
  - A high degree of censoring in environmental and biomarker measurements
  - Lack of exposure and partitioning models relevant to near field environments
- Indoor media (air, dust), and urinary biomarker concentrations for potential EDCs were collected in a study of 120 females age 60-80 (Household Exposure Study, Cape Cod, MA, Rudel et al. 2003)
- Data from US EPA's Exposure Forecasting (ExpoCast) project and new predictive models were used to assess sources of household exposure
  - Both forward and reverse modeling methods were used to characterize relationships among chemical source descriptors, near-field environmental measurements, and biomarker data



Data Limitations

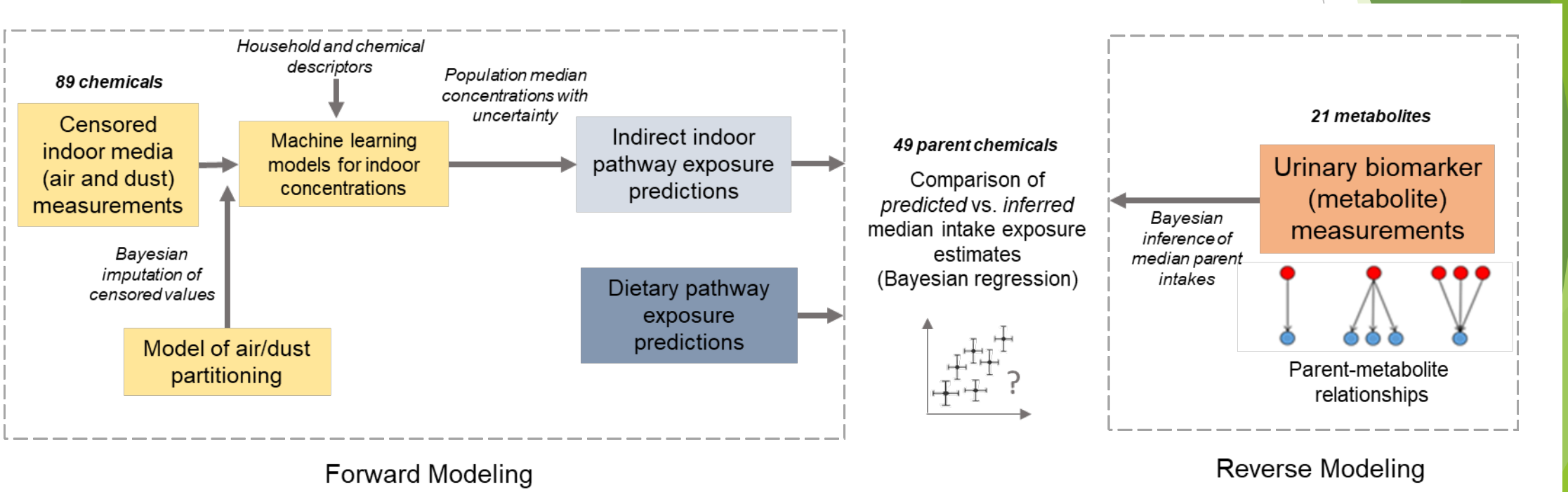
What we have:

1	Chemicals include phthalates, pesticides, PCBs, alkylphenols
2	78% of residential media measurements (across 89 chemicals) are below LOD
3	54% of urinary biomarker measurements (across 21 chemicals) are below LOD

What we need:

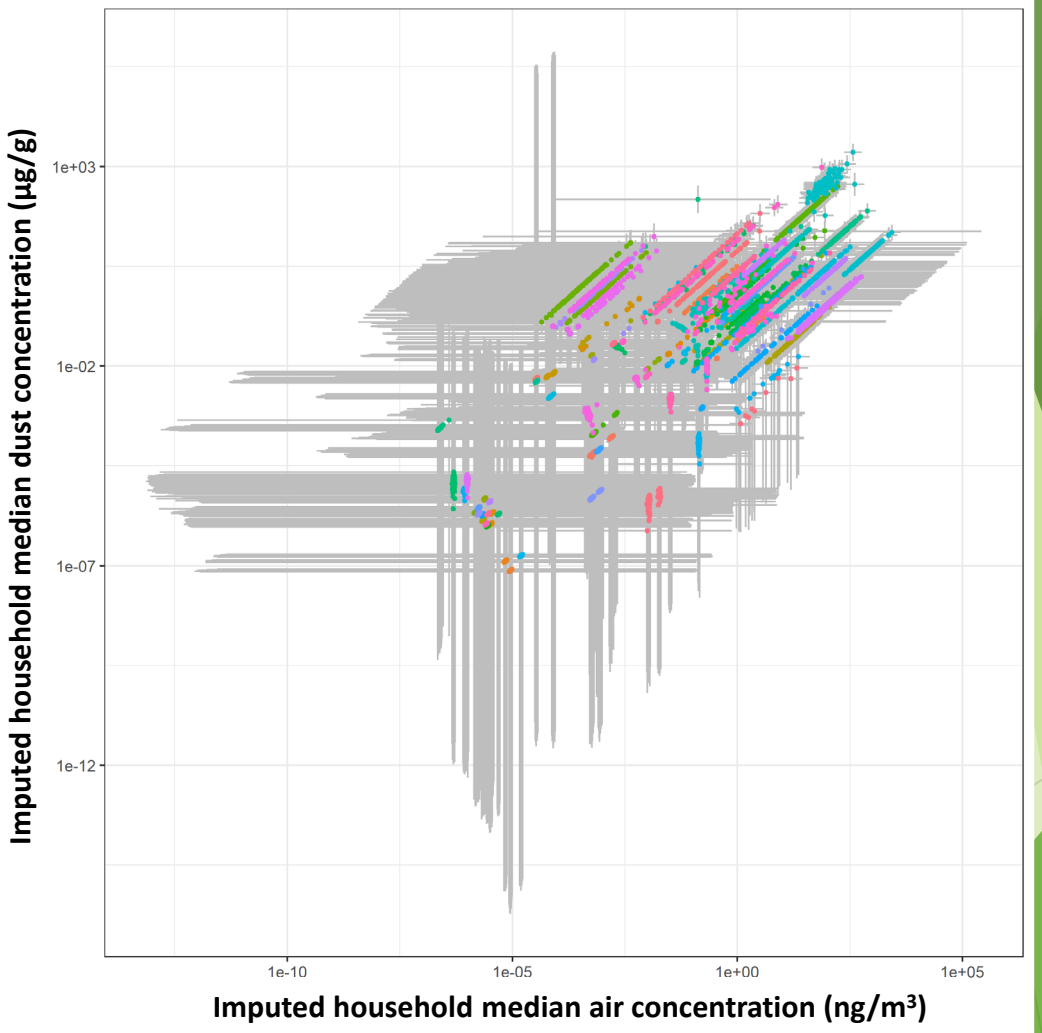
1	Air and dust concentrations for every chemical in every household from Cape Cod study
2	Concentrations for all urine metabolites and inferences of parent chemical intake
3	Air and dust concentrations for parent chemicals

Modeling Framework



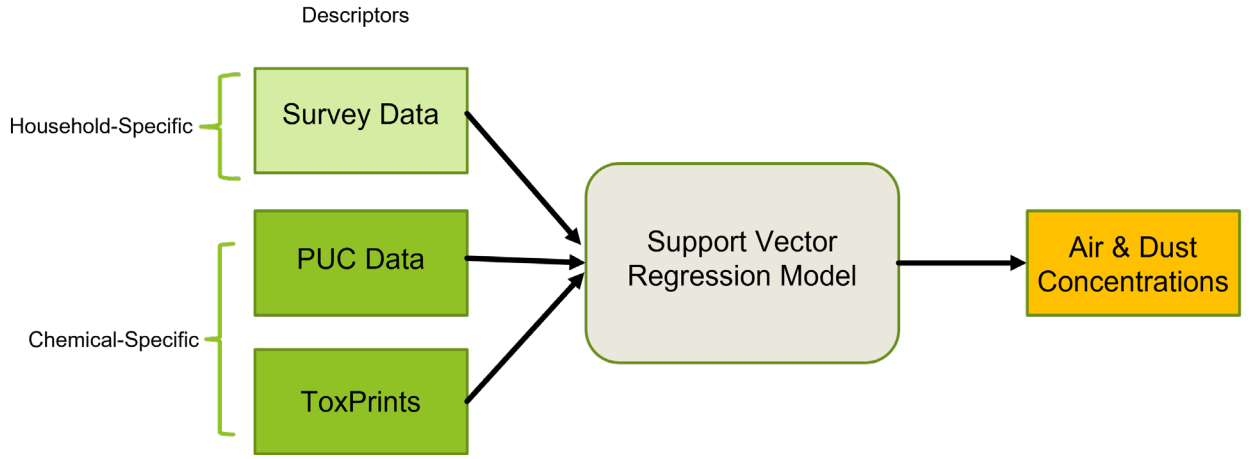
Partitioning Model for Air and Dust Concentrations

- Gas-phase chemicals will readily partition between airborne particles, surfaces, and house dust
- Model of Weschler and Nazaroff (2010) and latent variable estimation approach was used to create an estimate for every chemical in every household
- Bayesian analysis and Markov Chain Monte Carlo to sample from the posterior distribution for the air and dust geometric means to characterize the uncertainties in these estimates



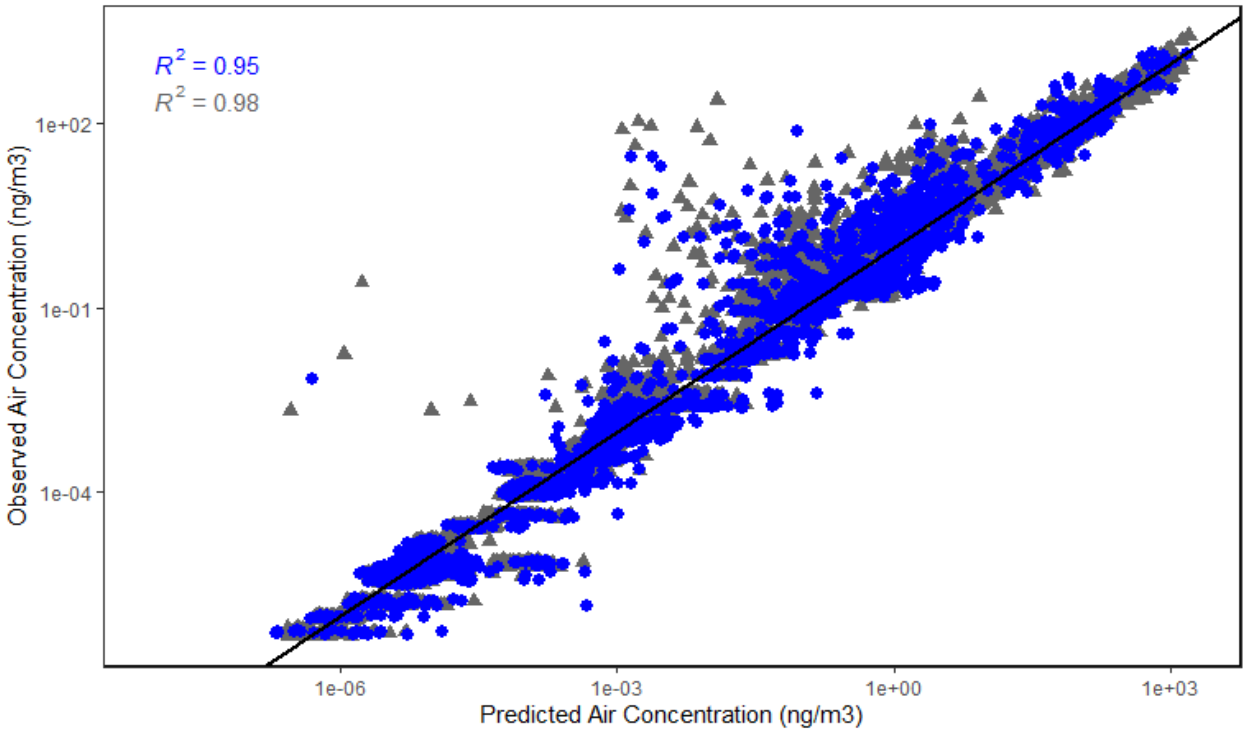
Model for Predicting Air and Dust Concentrations

- Chemicals were assigned 276 descriptors
  - Structural descriptors (Toxprint Chemotypes, Yang et al. 2015)
  - Presence in Product Use Categories (PUCs) from EPA's Chemicals and Products Database (Dionisio et al. 2018)
  - Household survey data (house age, use of products, min-max normalization)
- 80% of data used for training and 20% used for testing
- Support Vector Regression (SVR) models trained using radial basis function (RBF) kernel and 10-fold cross-validation
- Y-randomization conducted to make sure model not built on chance correlation

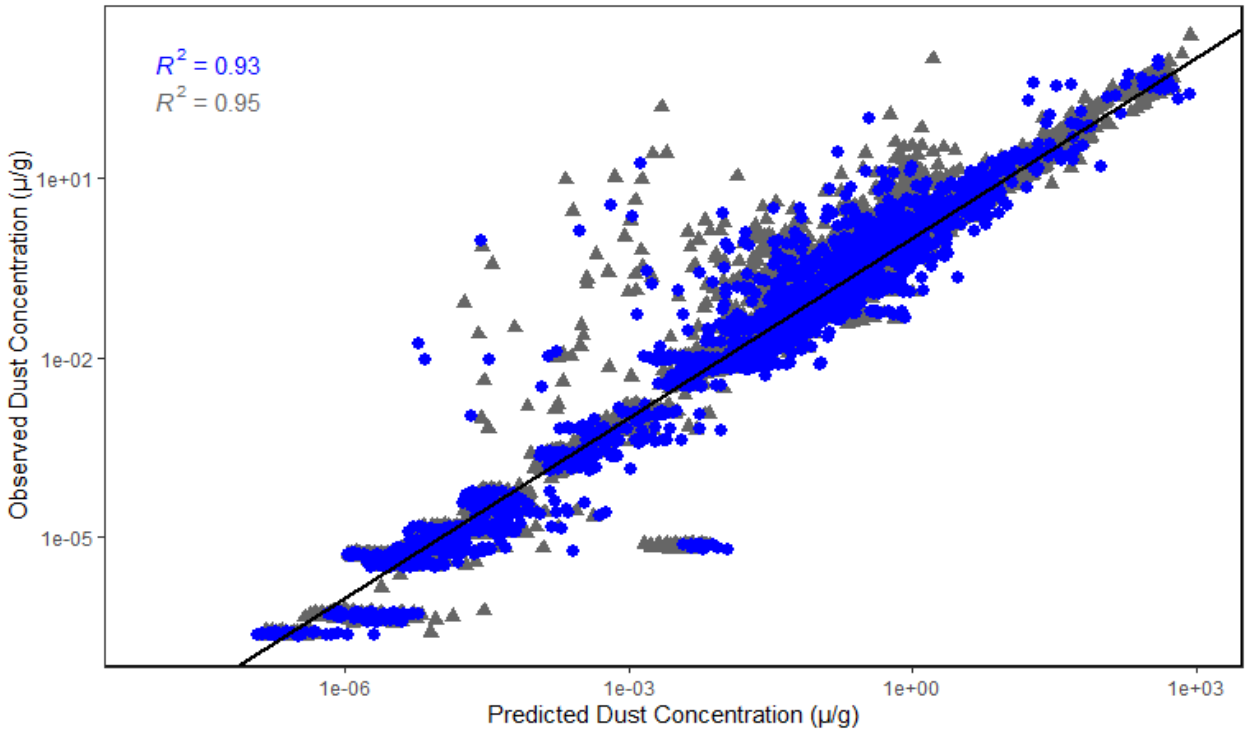


Model Performance: Training vs. Testing

Air



Dust



Training
Testing

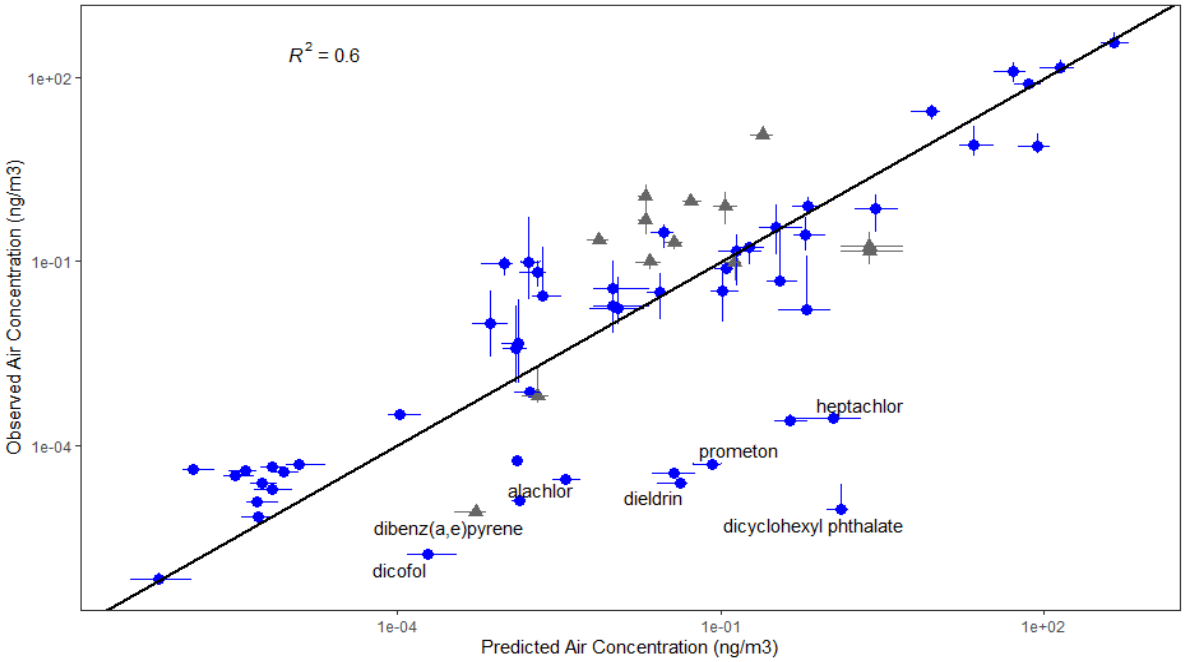
External Validation Datasets

- Indoor air and dust concentrations of 76 SVOCs from 50 homes in an exposure study in northern California (Dodson et al. 2015)
  - Estimates generated for every chemical in every household
  - 52 of the chemicals also measured in Cape Cod study
- Chemical concentrations in indoor air and dust were extracted from EPA's Multimedia Monitoring Database (Isaacs et al. 2022)
  - Median of the reported concentrations for each chemical was compared to version of models built without household-specific descriptors
  - 25 chemicals indoor air and 75 chemicals in dust, 16 and 36 of which were also measured in Cape Cod
- Only chemicals found in model domain were used for comparison

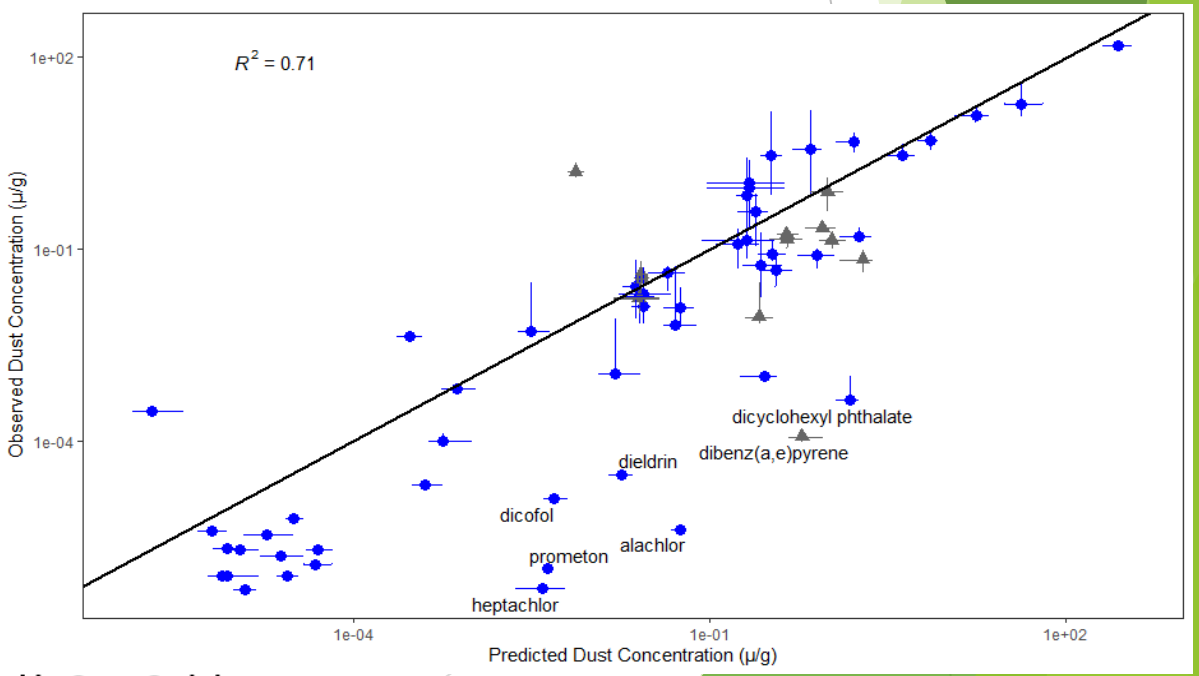
External Validation: California Dataset

- Model is overpredicting several historic use pesticides such as dieldrin, dicofol, & heptachlor
- Pesticides were applied widely in Cape Cod for tree pests, cranberry bogs, other agriculture, and mosquito control on wetlands
- Almost all of the median concentration of phthalates were lower than those measured in Cape Cod

Air



Dust

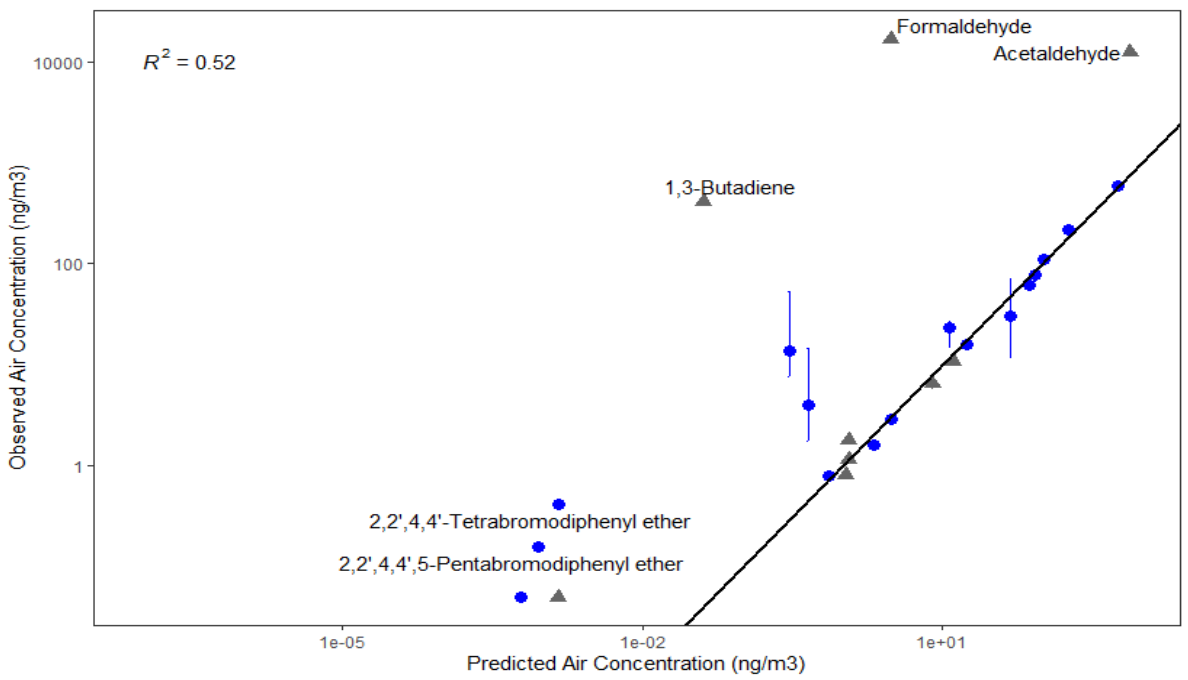


- **Chemical in Cape Cod dataset**
- **Chemical not in Cape Cod dataset**

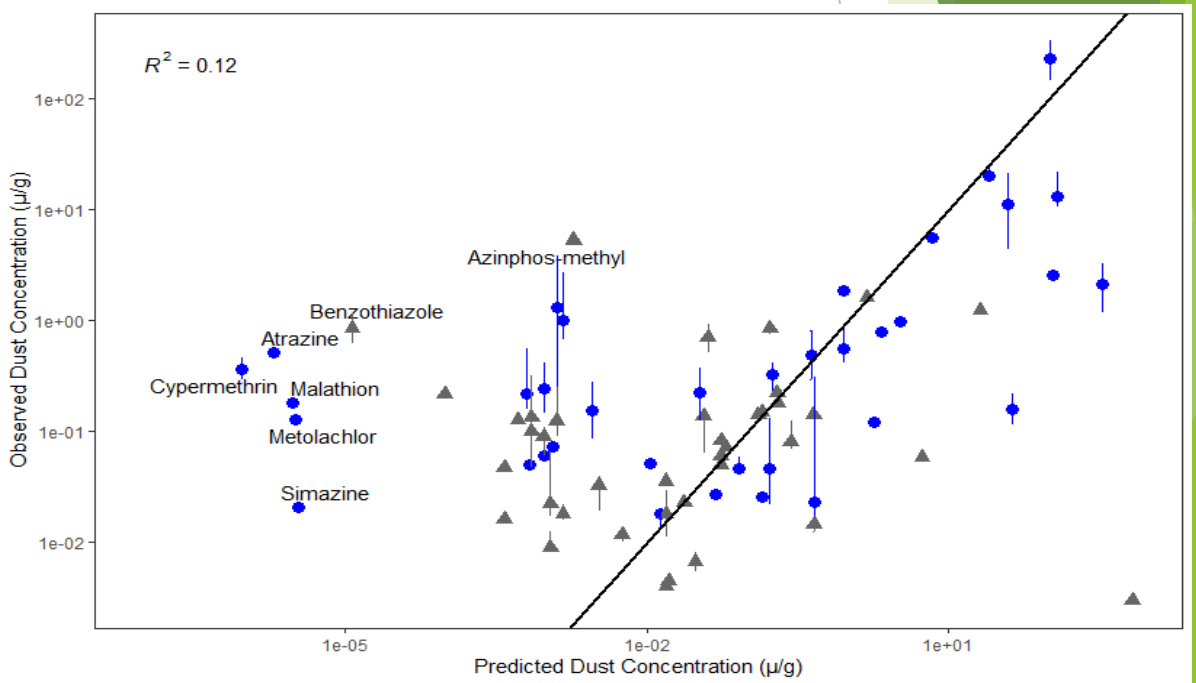
External Validation: MMDB Dataset

- Model is underpredicting several chemicals for both air and dust
- Concentrations for both come from studies where concentrations are higher than average (e.g., in homes near crop fields)

Air



Dust



- **Chemical in Cape Cod dataset**
- **Chemical not in Cape Cod dataset**

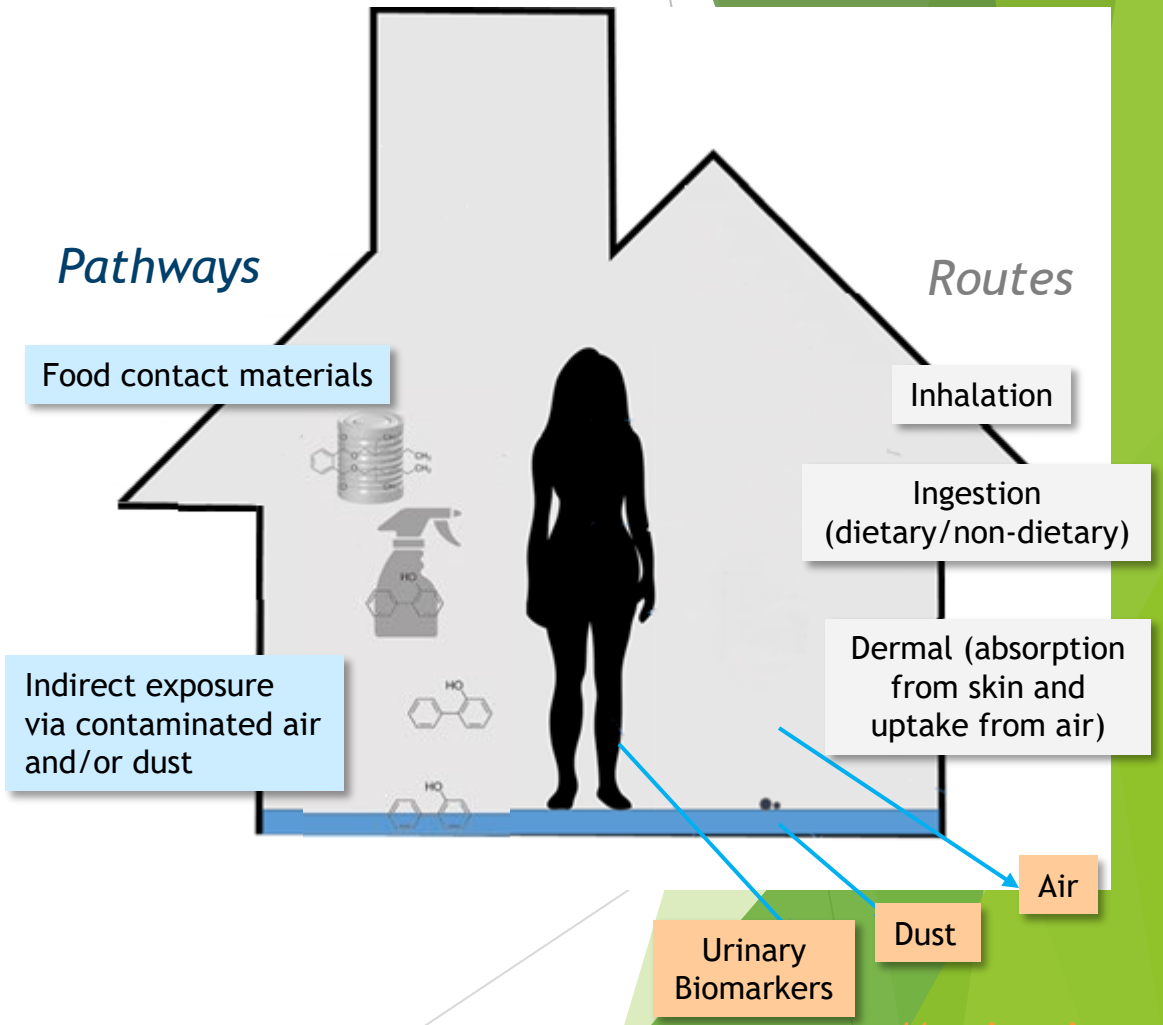
Inferred vs. Predicted Exposures

Inferred:

- Bayesian analysis via Markov Chain Monte Carlo was used to infer population median intake exposures for parent chemicals from creatinine-adjusted urinary metabolite concentrations using *a priori* knowledge of parent-metabolite relationships and a steady-state assumption (Wambaugh et al., 2014)
- 21 metabolites were mapped to 49 parent chemicals

Predicted:

- Air and dust concentrations for parents without measurements were estimated from SVR models
- Indirect indoor exposures were estimated the using high-throughput model of Little et al. (2012)
- These chemicals have exposure pathways beyond indoor air and dust
  - Food contact exposures from Biryol et al. (2017)



Explaining Inferred Exposures

- Both measured/predicted air and dust concentrations were significantly correlated with inferred exposures ($r=0.59$, $p<0.001$)
- Associated indoor exposure estimates were predictive of inferences ($R^2=0.33$, $p<0.001$)
- Examined both indoor and food contact exposures using similar approach as EPA's Systematic Empirical Evaluation of Models (Ring et al., 2018)
  - Bayesian aggregate model for inferred exposure

Inferred intake exposure

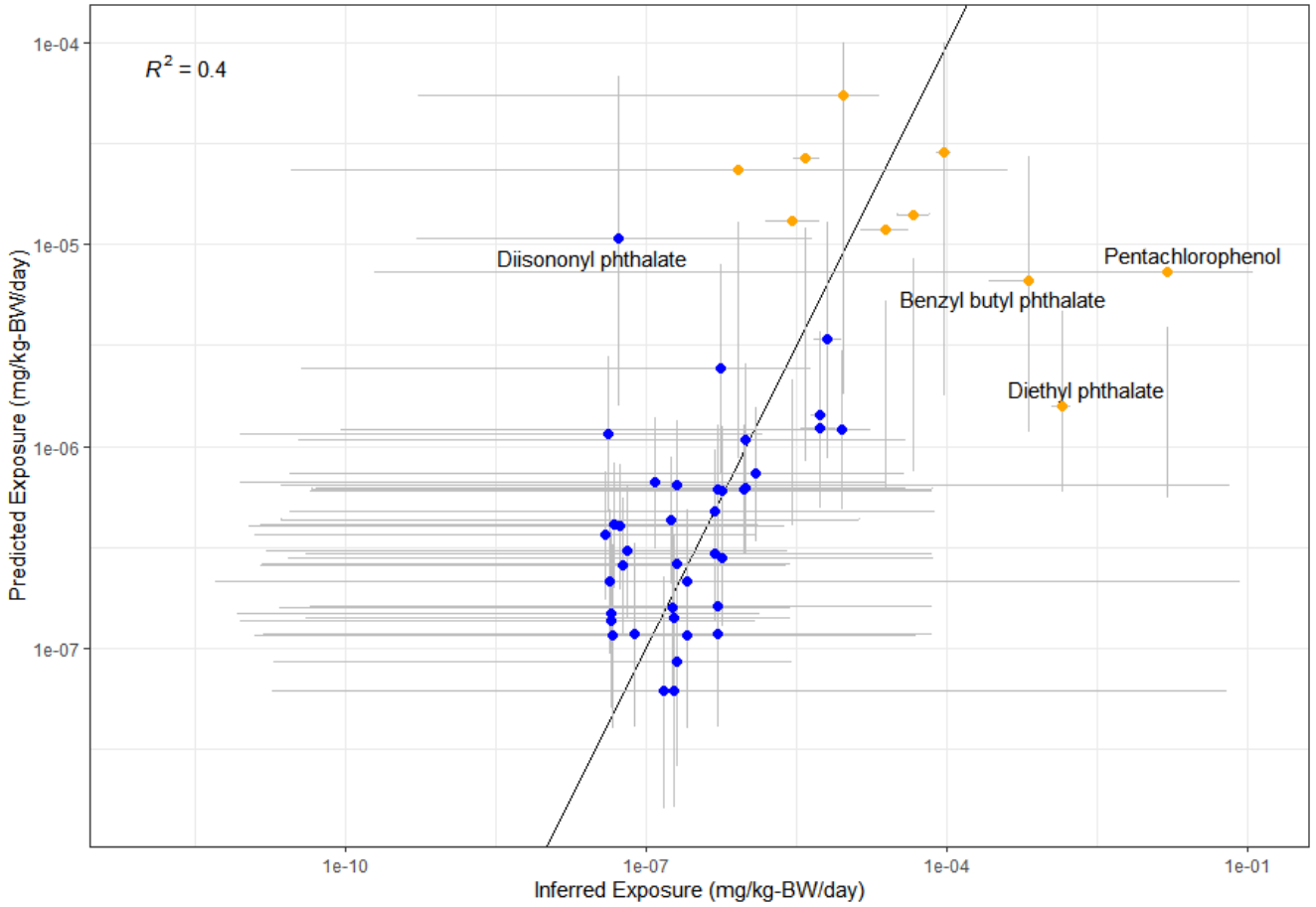
food contact exposure estimate

Indirect indoor (air/dust) exposure estimate

$$\log(E_{inf}) = A + FCS(B_{FCS}\log(E_{FCS})) + B_{ind}\log(E_{ind}) + \varepsilon$$

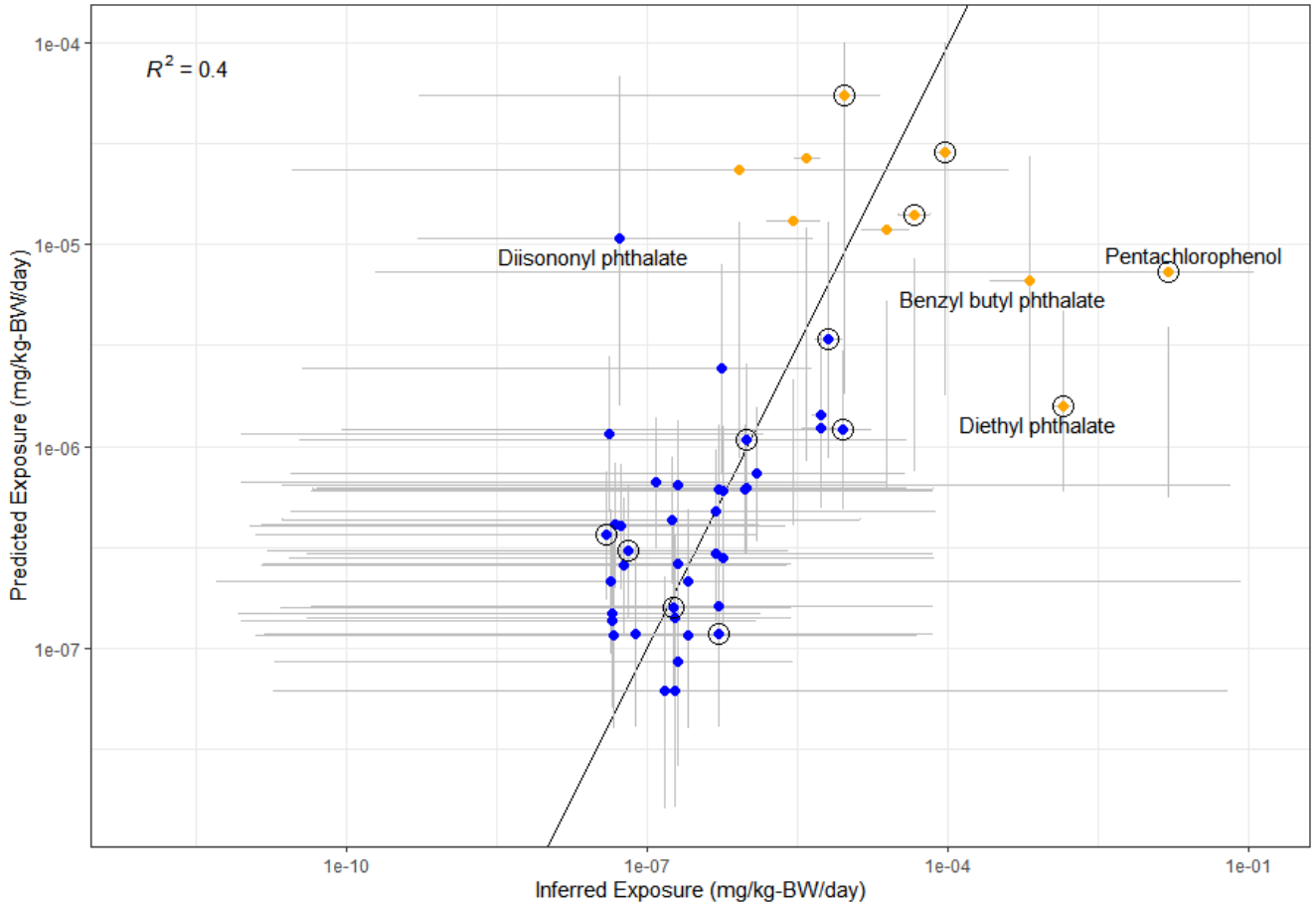
binary variable which is equal to 1 when a chemical has a food contact pathway in addition to an indoor exposure pathway

Bayesian Aggregate Model for Exposures



- Food contact pathway
- No food contact pathway

Bayesian Aggregate Model for Exposures



- Food contact pathway
- No food contact pathway

- These chemicals have been studied or explicitly found in drinking water in MMDB
- Other sources (e.g., other dietary sources) may be uncharacterized

Summary

- Air and dust concentrations can be predicted when utilizing product use information, molecular structure descriptors, and survey data on household characteristics, and predictions are reasonably reliable in different geographical locations and when survey data are not available
  - Model can be used to inform design of monitoring studies and high-throughput prediction of near-field exposures
  - Geospatial patterns in ambient (far-field) sources could be considered in the future for improving prediction of media concentrations
- Forward and reverse modeling approaches can aid in the interpretation of measured urinary biomarkers via inference of corresponding intake of parent chemicals, thereby elucidating and quantifying exposure pathway contributions for chemicals and chemical classes
  - High-throughput models should be developed for uncharacterized exposure pathways (e.g., drinking water)

Collaborators

Silent Spring Institute

Robin Dodson
Ruthann Rudel

US EPA

Derya Biryol (former ORISE)
Victoria Hull
Kristin Isaacs
Mary Jacketti (ORISE)
Risa Sayre
Woody Setzer
Zach Stanfield
John Wambaugh

References

- Biryol D, Nicolas CI, Wambaugh J, Phillips K, Isaacs K. High-throughput dietary exposure predictions for chemical migrants from food contact substances for use in chemical prioritization. Environ Int. 2017 Nov;108:185-194.
- Dionisio, Kathie L., Katherine Phillips, Paul S. Price, Chris Grulke, Antony Williams, Derya Biryol, Tao Hong, and Kristin K. Isaacs. 2018. 'The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products', Scientific Data, 5: 1-9.
- Dodson RE, Camann DE, Morello-Frosch R, Brody JG, Rudel RA. Semivolatile organic compounds in homes: strategies for efficient and systematic exposure measurement based on empirical and theoretical factors. Environ Sci Technol. 2015 Jan 6;49(1):113-22.
- Isaacs KK, Wall JT, Williams AR, Hobbie KA, Sobus JR, Ulrich E, Lyons D, Dionisio KL, Williams AJ, Grulke C, Foster CA, McCoy J, Bevington C. A harmonized chemical monitoring database for support of exposure assessments. Sci Data. 2022 Jun 16;9(1):314.
- Little JC, Weschler CJ, Nazaroff WW, Liu Z, Cohen Hubal EA. Rapid methods to estimate potential exposure to semivolatile organic compounds in the indoor environment. Environ Sci Technol. 2012 Oct 16;46(20):11171-8.
- Ring CL, Arnot JA, Bennett DH, Egeghy PP, Fantke P, Huang L, Isaacs KK, Jolliet O, Phillips KA, Price PS, Shin HM, Westgate JN, Setzer RW, Wambaugh JF. Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. Environ Sci Technol. 2019 Jan 15;53(2):719-732.
- Rudel RA, Camann DE, Spengler JD, Korn LR, Brody JG. Phthalates, alkylphenols, pesticides, polybrominated diphenyl ethers, and other endocrine-disrupting compounds in indoor air and dust. Environ Sci Technol. 2003 Oct 15;37(20):4543-53.
- Weschler CJ, Nazaroff WW, SVOC Partitioning between the gas phase and settled dust indoors. Atmospheric Environment. 2010. 44(30): 3609-3620.
- Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J, Terfloth L, Arvidson K, Richard A, Worth A, Rathman J. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. J Chem Inf Model. 2015 Mar 23;55(3):510-28.