United States Environmental Protection Agency

Fusing Exposure Data and Models to Reduce Uncertainty

John F. Wambaugh

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

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

https://orcid.org/0000-0002-4024-534X

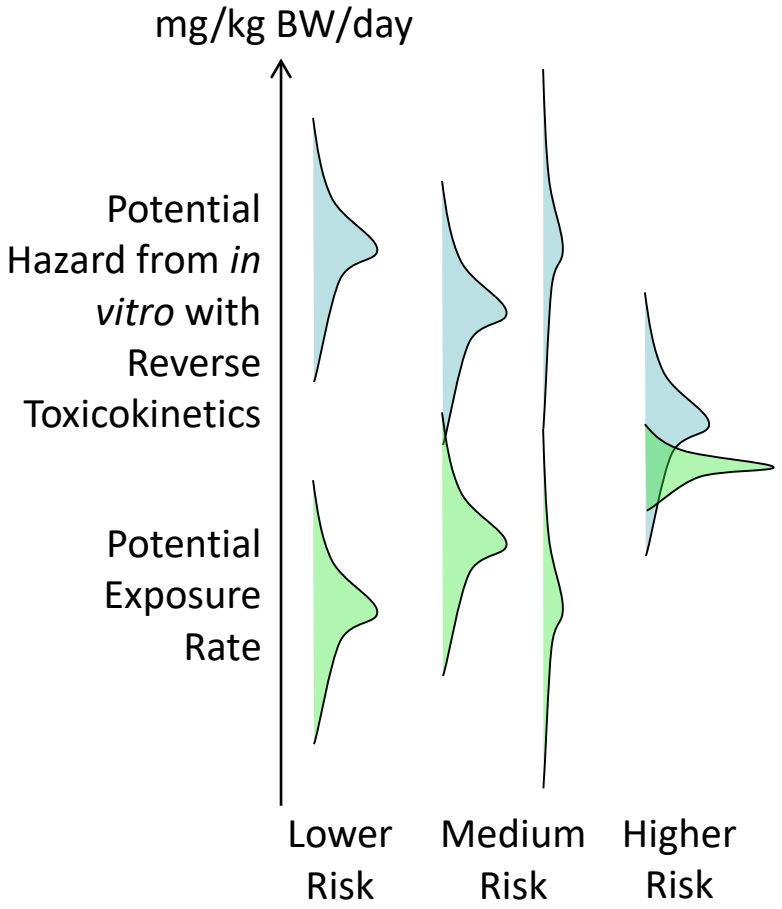
ToxForum
Determining Relevant Low-Level Chemical Exposures for Safety Assessments of Consumer Products
May 20, 2019

Chemical Regulation in the United States

- Park *et al.* (2012): At least 3221 chemical signatures in pooled human blood samples, many appear to be exogenous
- A tapestry of laws covers the chemicals people are exposed to in the United States (Breyer, 2009)
- Different testing requirements exist for food additives, pharmaceuticals, and pesticide active ingredients (NRC, 2007)
- Most other chemicals, ranging from industrial waste to dyes to packing materials, are covered by the Toxic Substances Control Act (TSCA)

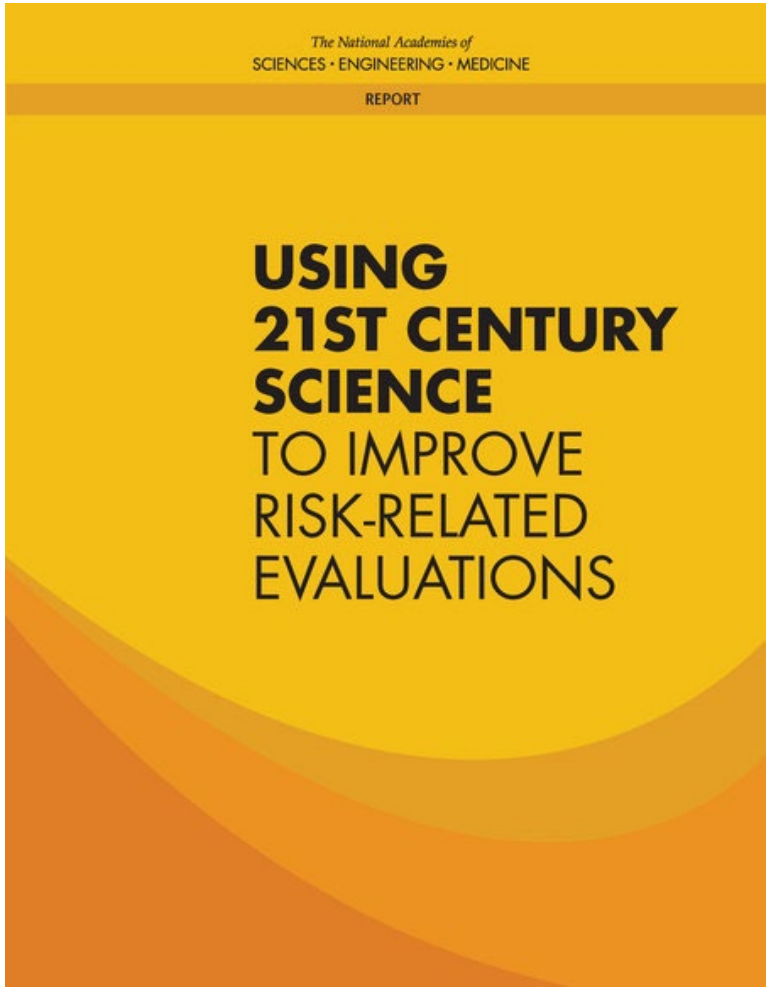
November 29, 2014

Risk Assessment in the 21st Century



"Translation of high-throughput data into risk-based rankings is an important application of exposure data for chemical priority-setting. Recent advances in high-throughput toxicity assessment, notably the ToxCast and Tox21 programs… and in high-throughput computational exposure assessment… have enabled first-tier risk-based rankings of chemicals on the basis of margins of exposure…"

"…The committee sees the potential for the application of **computational exposure science** to be highly valuable and credible for comparison and **priority-setting among chemicals in a risk-based context**."



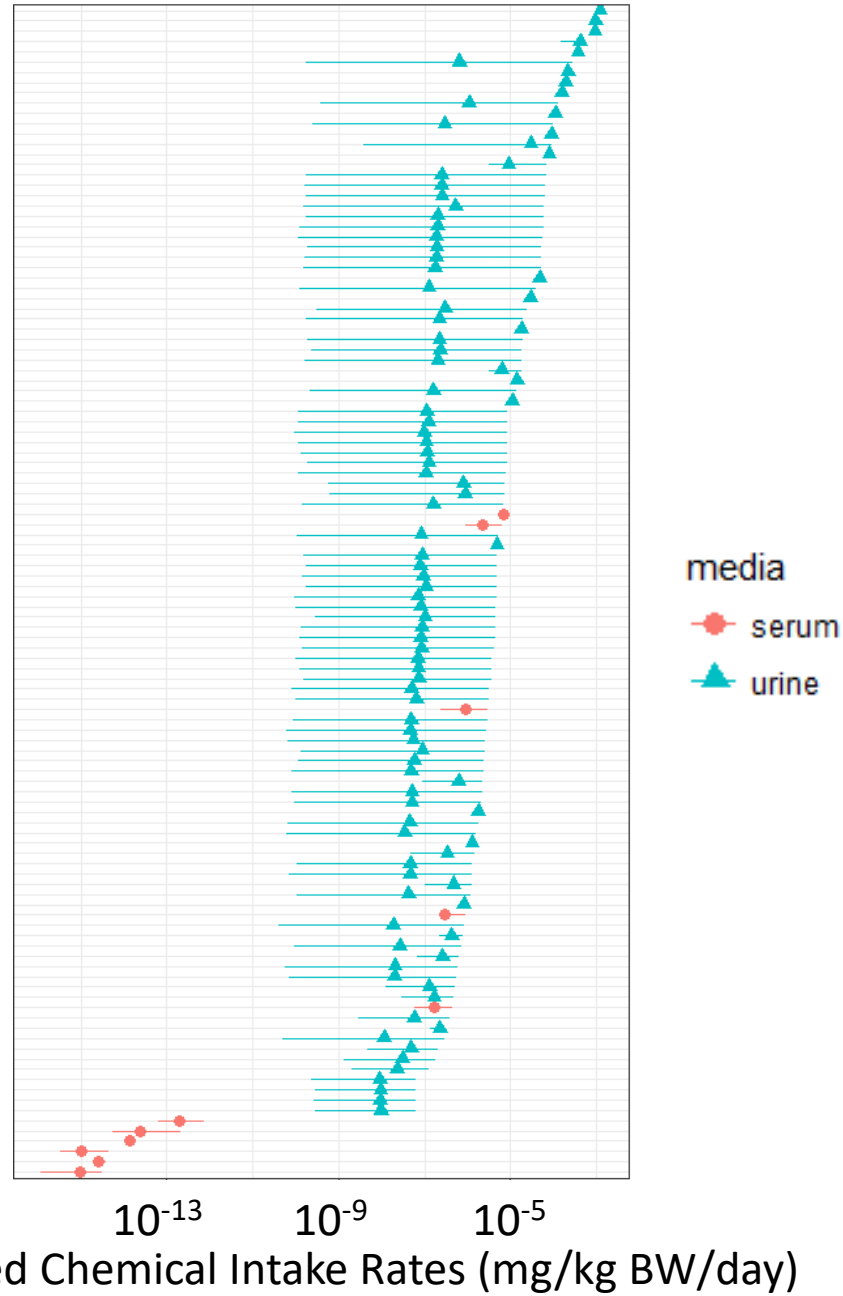
January, 2017

What Do We Know About Exposure? Biomonitoring Data

- Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey (NHANES) provides an important tool for monitoring public health
- Large, ongoing CDC survey of US population: demographic, body measures, medical exam, biomonitoring (health and exposure), …
- Designed to be representative of US population according to census data
- Data sets publicly available (http://www.cdc.gov/nchs/nhanes.htm)
- Includes measurements of:
  - Body weight
  - Height
  - **Chemical analysis of blood and urine**

Reverse Dosimetry (Tan et al., 2006)

- Median chemical intake rates (mg / kg body weight /day) were inferred from:
  - NHANES urine (Wambaugh et al, 2014, Ring et al. 2017)
  - NHANES serum/blood either using HTTK clearance (Pearce et al., 2017)
    - Literature clearance estimates were used for methodologically challenging chemicals not suited to HTTK



What Do We Know About Exposure? Exposure Models

- Human chemical exposures can be coarsely grouped into "**near field**" sources that are close to the exposed individual (consumer or occupational exposures) "**far-field"** scenarios wherein individuals are exposed to chemicals that were released or used far away (ambient exposure) (Arnot *et al.*, 2006).

- A model captures knowledge and a hypothesis of how the world works (MacLeod *et al.*, 2010)

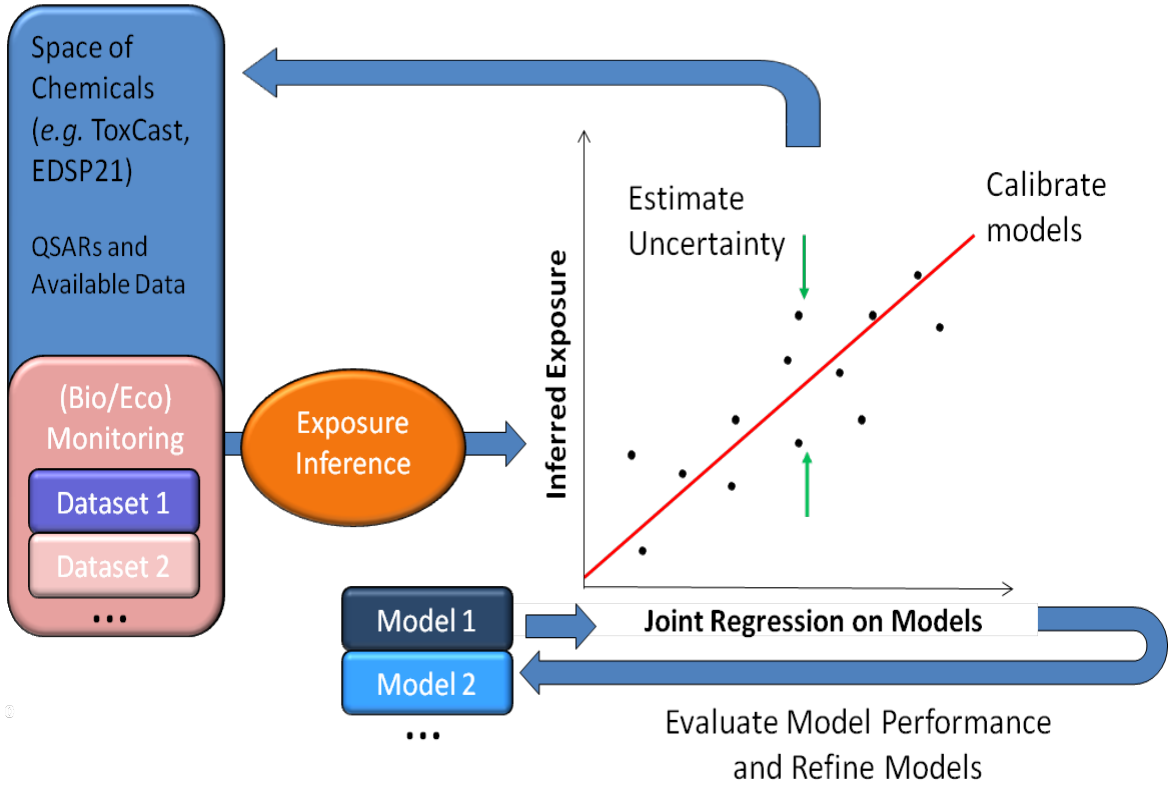
- EPA's EXPOsure toolBOX (EPA ExpoBox) is a toolbox created to assist individuals from within government, industry, academia, and the general public with assessing exposure
  - Includes many, many models

  https://www.epa.gov/expobox

"Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations… The only question of interest is 'Is the model illuminating and useful?'" George Box

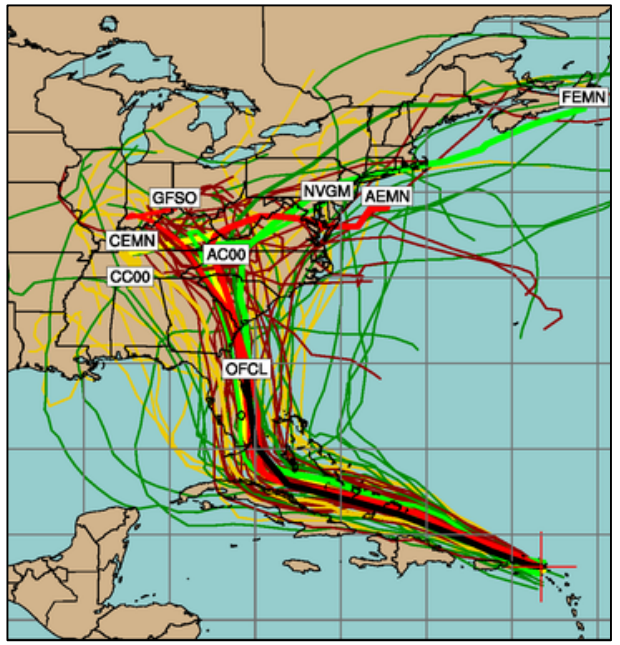
Consensus Exposure Predictions with the SEEM Framework

- Different exposure models incorporate **knowledge**, **assumptions**, and **data** (MacLeod et al., 2010)
- We incorporate multiple models into consensus predictions for 1000s of chemicals within the **Systematic Empirical Evaluation of Models (SEEM)** (Wambaugh et al., 2013, 2014)



- Evaluation is similar to a sensitivity analysis: What models are working? What data are most needed?

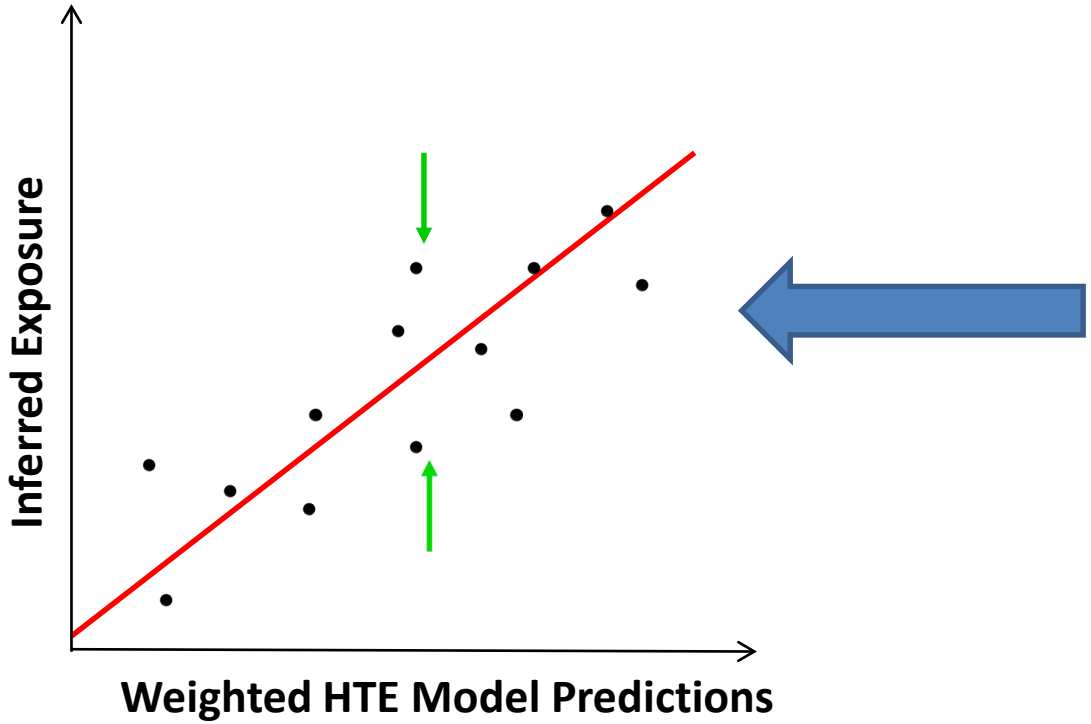
Hurricane Path Prediction is an Example of Integrating Multiple Models



SEEM is a Linear Regression

Multiple regression models:

Log(Parent Exposure) = a + m * log(Model Prediction) + b* Near Field + ε

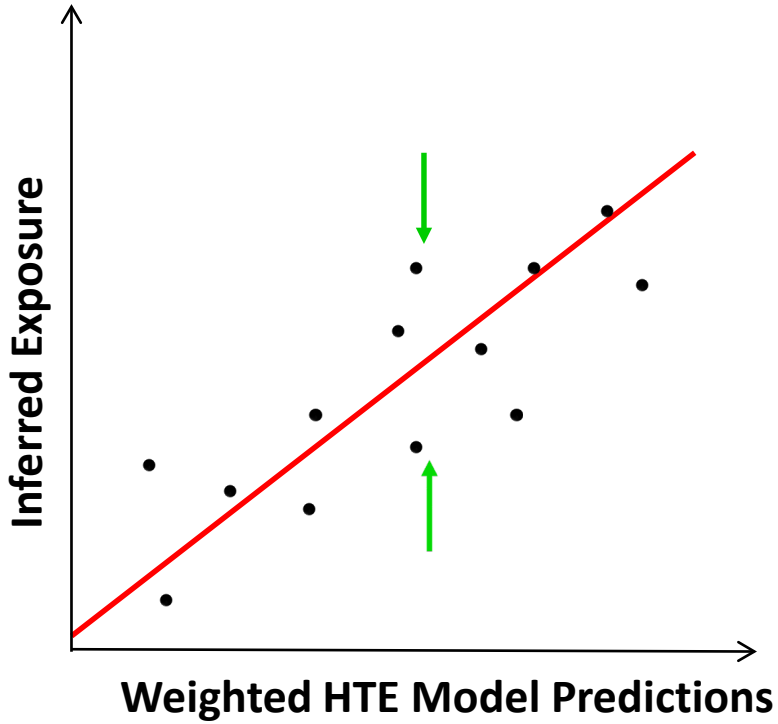


$\varepsilon \sim N(0, \sigma^2)$
Residual error, unexplained by the regression model

SEEM is a Linear Regression

Multiple regression models:

$$\text{Log(Parent Exposure)} = a + m * \log(\text{Model Prediction}) + b * \text{Near Field} + \varepsilon$$



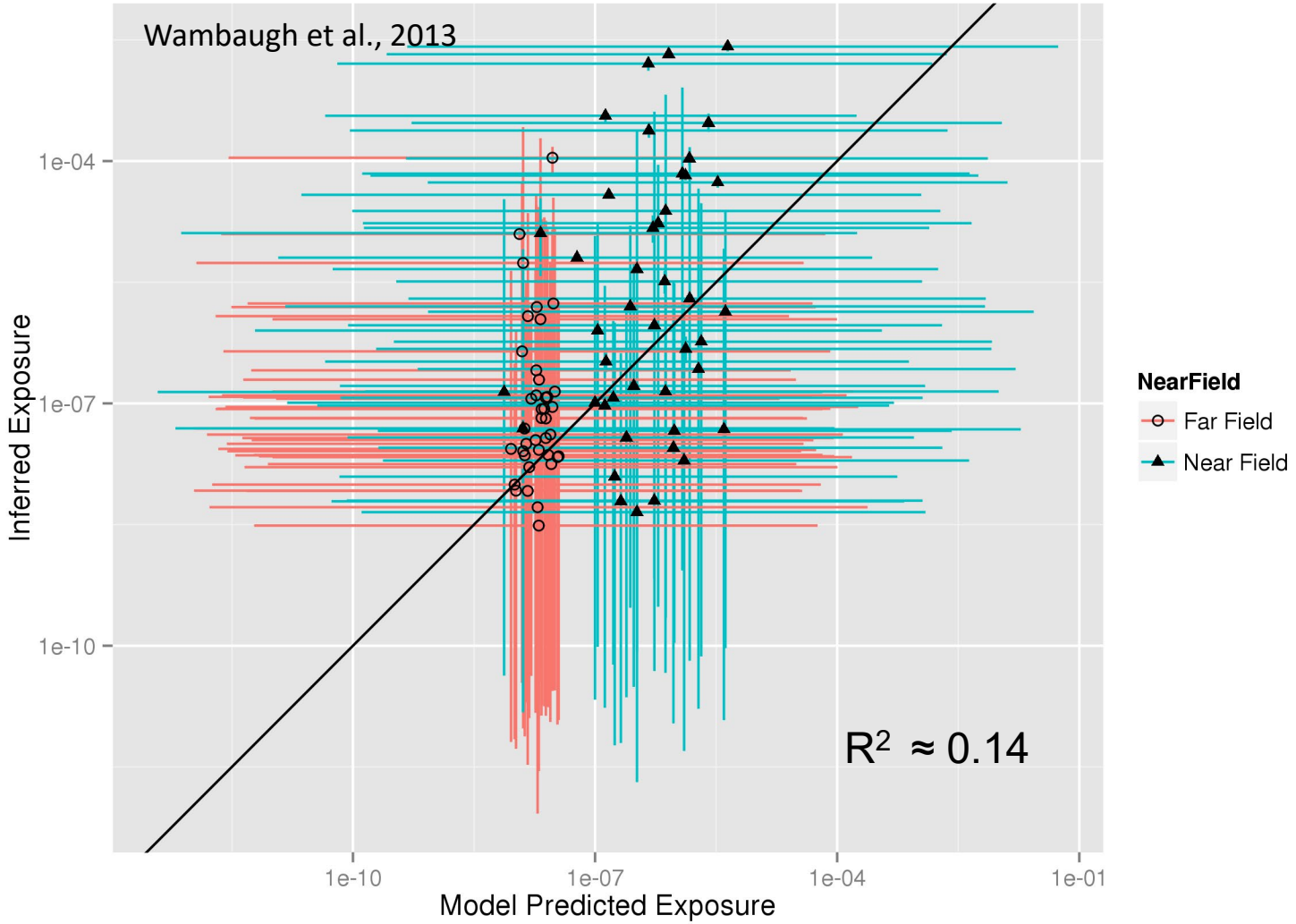
Not all models have predictions for all chemicals

- We can run SHEDS-HT (Isaacs et al., 2014) for ~2500 chemicals

What do we do for the rest?

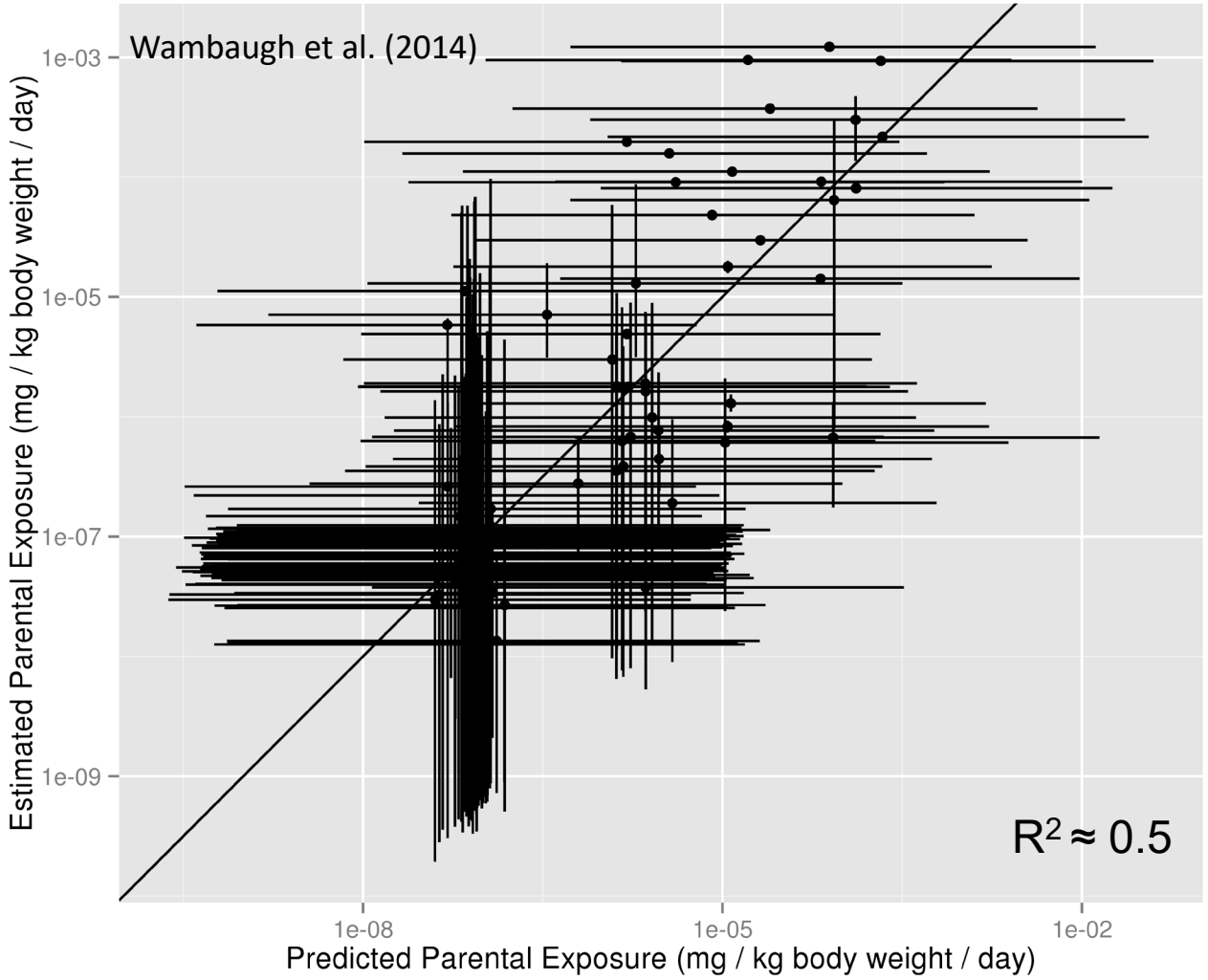
- Assign the average value?
- Zero?

First Generation SEEM



- Those chemicals with "near-field" – proximate, in the home, sources of exposure – had much higher rates of exposure than those with sources outside the home (Wallace et al., 1986)
- The only available "high throughput exposure models in 2013 were for far-field sources

Second Generation SEEM

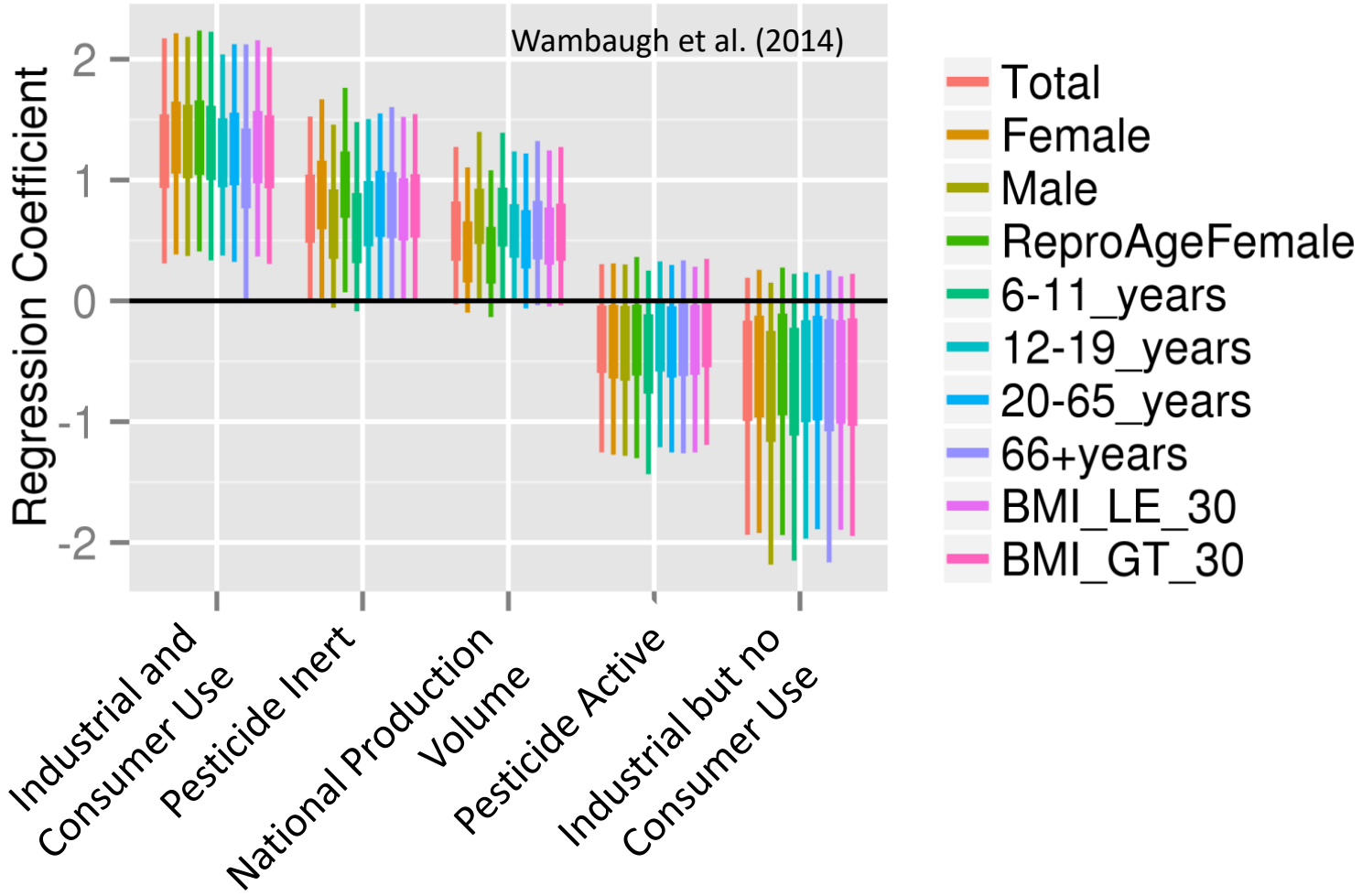


$R^2 \approx 0.5$ indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates

Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume

Heuristics of Exposure



$R^2 \approx 0.5$ indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates

Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume

Collaboration on High Throughput Exposure Predictions

Jon Arnot, Deborah H. Bennett, Peter P. Egeghy, Peter Fantke, Lei Huang, Kristin K. Isaacs, Olivier Jolliet, Hyeong-Moo Shin, Katherine A. Phillips, Caroline Ring, R. Woodrow Setzer, John F. Wambaugh, Johnny Westgate

Predictor	Reference(s)	Chemicals Predicted	Pathways
EPA Inventory Update Reporting and Chemical Data Reporting (CDR) (2015)	US EPA (2018)	7856	All
Stockholm Convention of Banned Persistent Organic Pollutants (2017)	Lallas (2001)	248	Far-Field Industrial and Pesticide
EPA Pesticide Reregistration Eligibility Documents (REDs) Exposure Assessments (Through 2015)	Wetmore et al. (2012, 2015)	239	Far-Field Pesticide
United Nations Environment Program and Society for Environmental Toxicology and Chemistry toxicity model (USEtox) Industrial Scenario (2.0)	Rosenbaum et al. (2008)	8167	Far-Field Industrial
USEtox Pesticide Scenario (2.0)	Fantke et al. (2011, 2012, 2016)	940	Far-Field Pesticide
Risk Assessment IDentification And Ranking (RAIDAR) Far-Field (2.02)	Arnot et al. (2008)	8167	Far-Field Pesticide
EPA Stochastic Human Exposure Dose Simulator High Throughput (SHEDS-HT) Near-Field Direct (2017)	Isaacs (2017)	7511	Far-Field Industrial and Pesticide
SHEDS-HT Near-field Indirect (2017)	Isaacs (2017)	1119	Residential
Fugacity-based INdoor Exposure (FINE) (2017)	Bennett et al. (2004), Shin et al. (2012)	645	Residential
RAIDAR-ICE Near-Field (0.803)	Arnot et al., (2014), Zhang et al. (2014)	1221	Residential
USEtox Residential Scenario (2.0)	Jolliet et al. (2015), Huang et al. (2016,2017)	615	Residential
USEtox Dietary Scenario (2.0)	Jolliet et al. (2015), Huang et al. (2016), Ernstoff et al. (2017)	8167	Dietary

Knowledge of Exposure Pathways Limits High Throughput Exposure Models

"In particular, the assumption that 100% of [quantity emitted, applied, or ingested] is being applied to each individual use scenario is a very conservative assumption for many compound / use scenario pairs."

Environmental Science & Technology

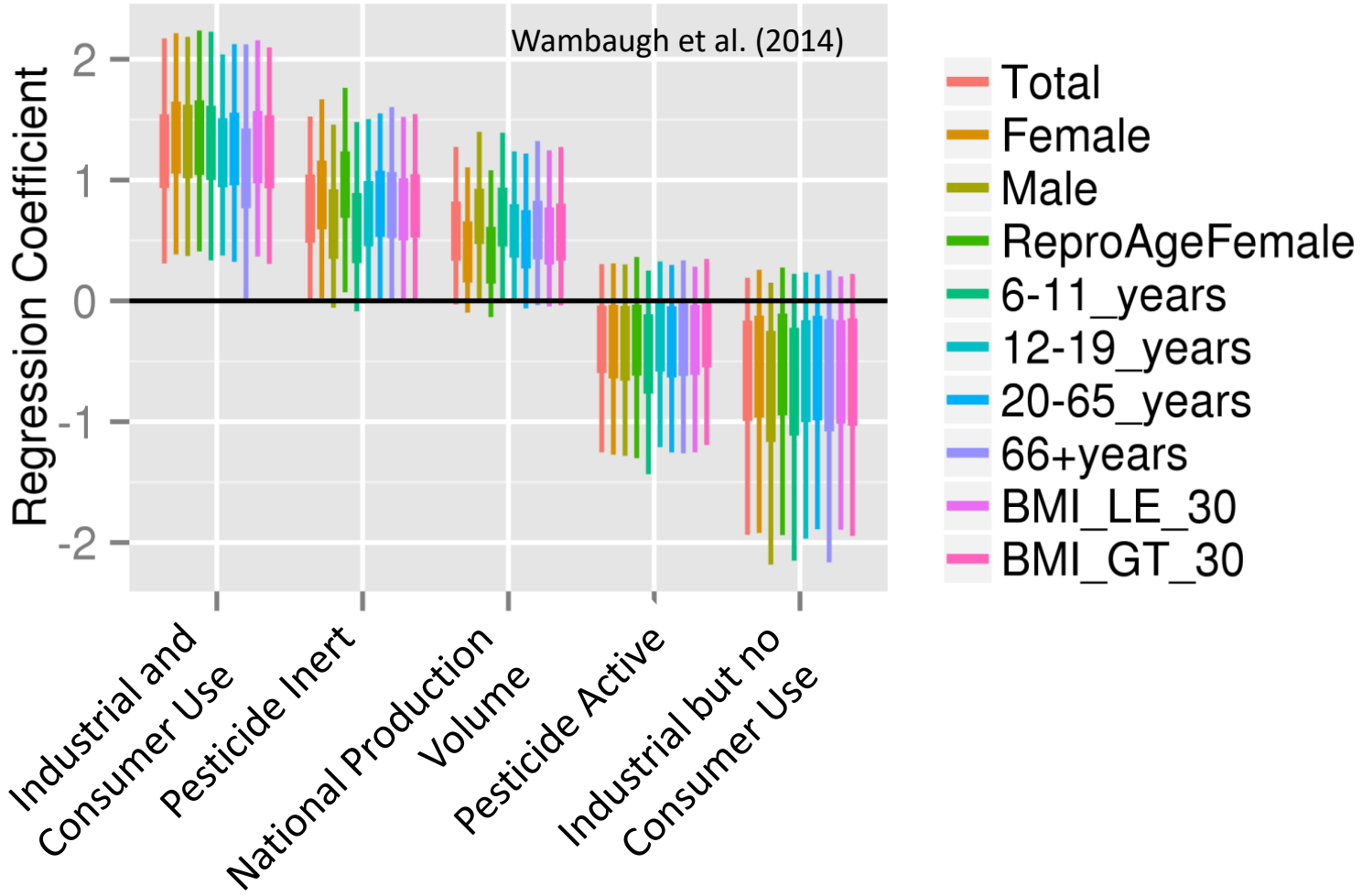
Article

pubs.acs.org/est

Risk-Based High-Throughput Chemical Screening and Prioritization using Exposure Models and in Vitro Bioactivity Assays

Hyeong-Moo Shin,[*,†] Alexi Ernstoff,[‡,§] Jon A. Arnot,[∥,⊥,#] Barbara A. Wetmore,[▽] Susan A. Csiszar,[§] Peter Fantke,[‡] Xianming Zhang,[○] Thomas E. McKone,[◆,¶] Olivier Jolliet,[§] and Deborah H. Bennett[†]

Heuristics of Exposure

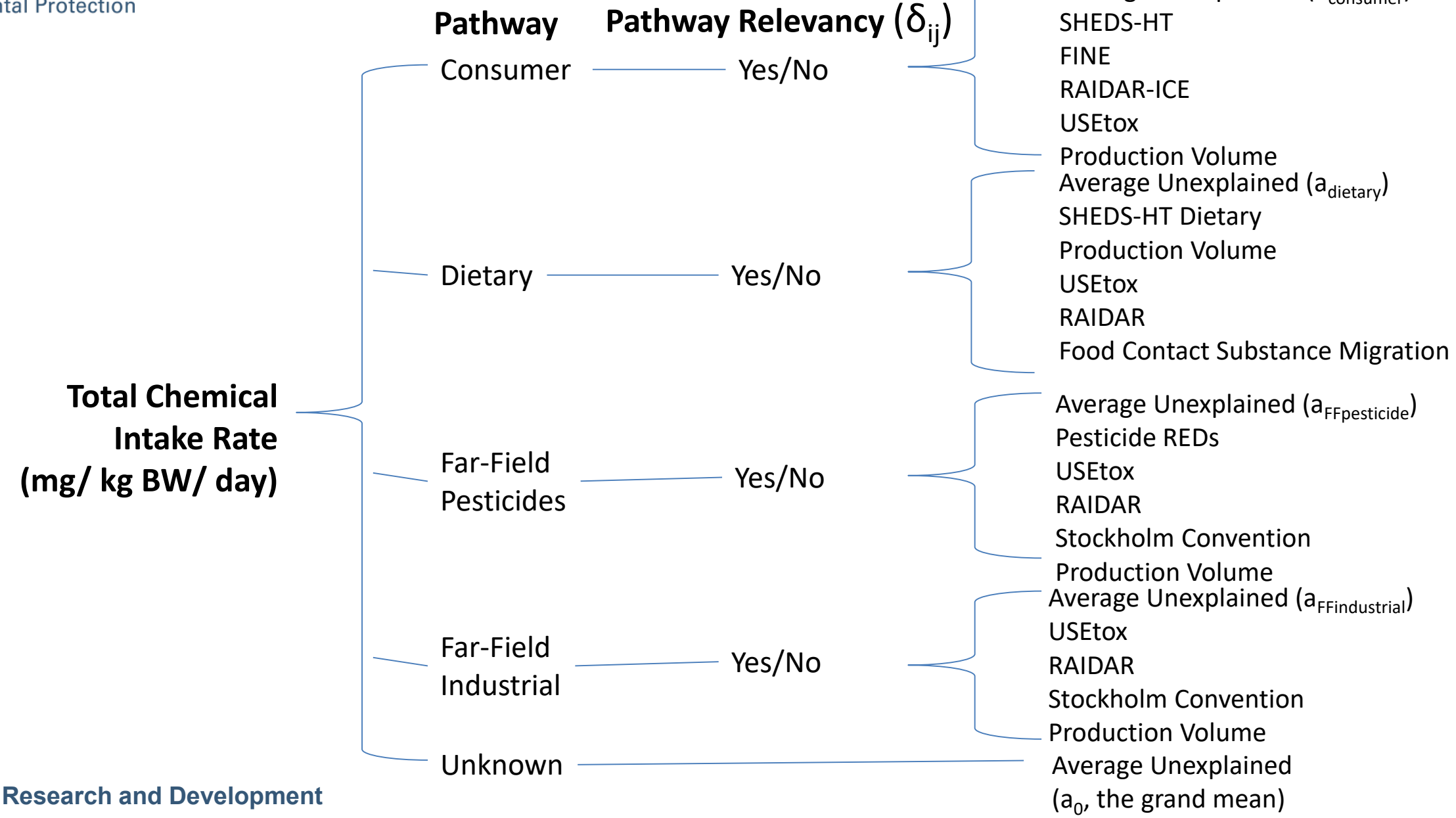


$R^2 \approx 0.5$ indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates

Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume

SEEM3

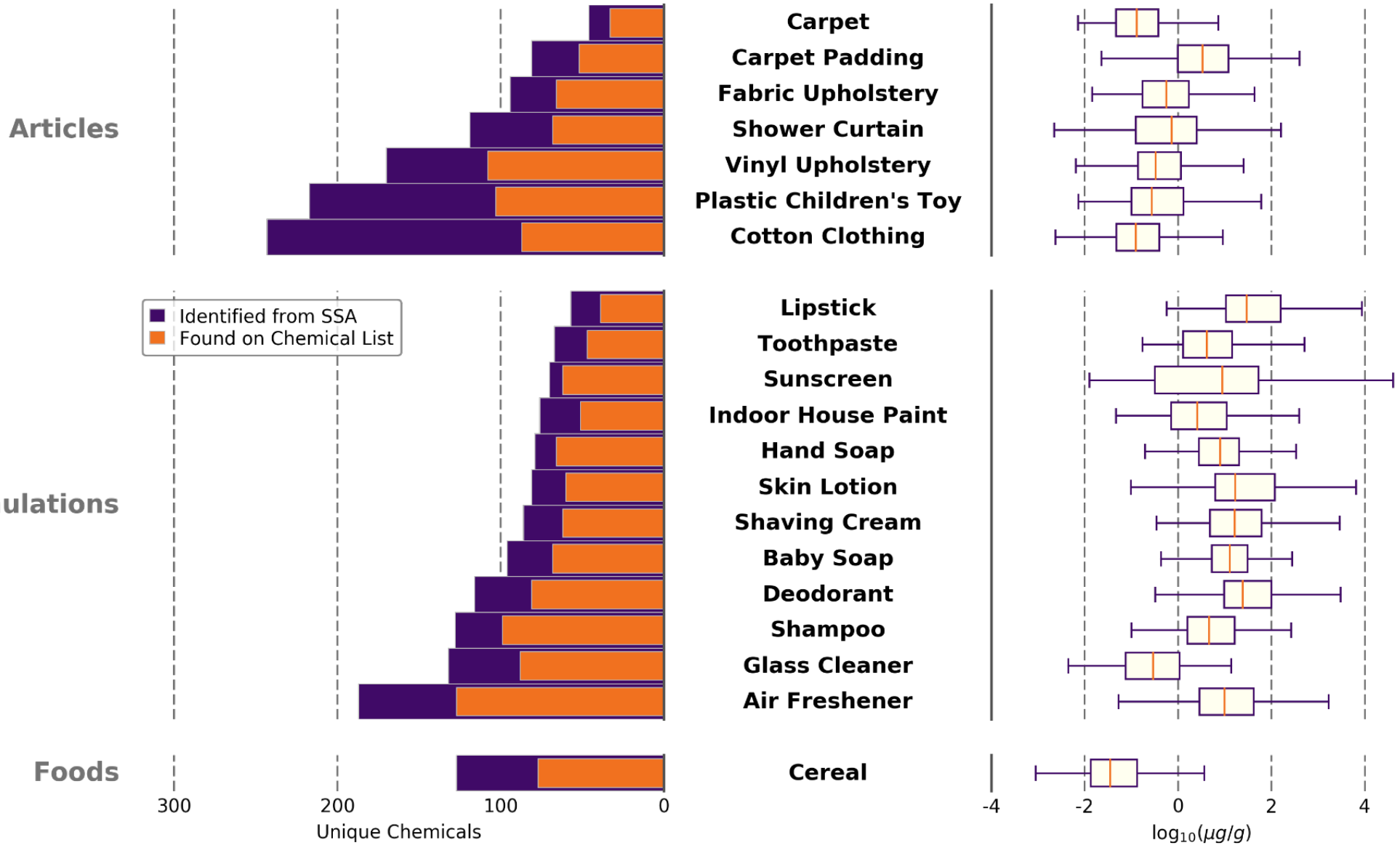

Total Chemical Intake Rate (mg/ kg BW/ day)
Pathway
Chemical-Specific Pathway Relevancy (δ_{ij})
Predictors
Consumer
Yes/No
Average Unexplained ($a_{consumer}$)
SHEDS-HT
FINE
RAIDAR-ICE
USEtox
Production Volume
Dietary
Yes/No
Average Unexplained ($a_{dietary}$)
SHEDS-HT Dietary
Production Volume
USEtox
RAIDAR
Food Contact Substance Migration
Far-Field Pesticides
Yes/No
Average Unexplained ($a_{FFpesticide}$)
Pesticide REDs
USEtox
RAIDAR
Stockholm Convention
Production Volume
Far-Field Industrial
Yes/No
Average Unexplained ($a_{FFindustrial}$)
USEtox
RAIDAR
Stockholm Convention
Production Volume
Unknown
Average Unexplained (a_0, the grand mean)

Developing Pathway-Specific Chemical Data

ExpoCast household item pilot study analyzed 5 examples each of 20 diverse household items.

Of 1,632 chemicals confirmed or tentatively identified, 1,445 were not present in CPCPdb

This gives us positive reference chemicals – negatives even harder



Phillips *et al. (2018)*

Predicting Exposure Pathways

We use the method of Random Forests to relate chemical structure and properties to exposure pathway

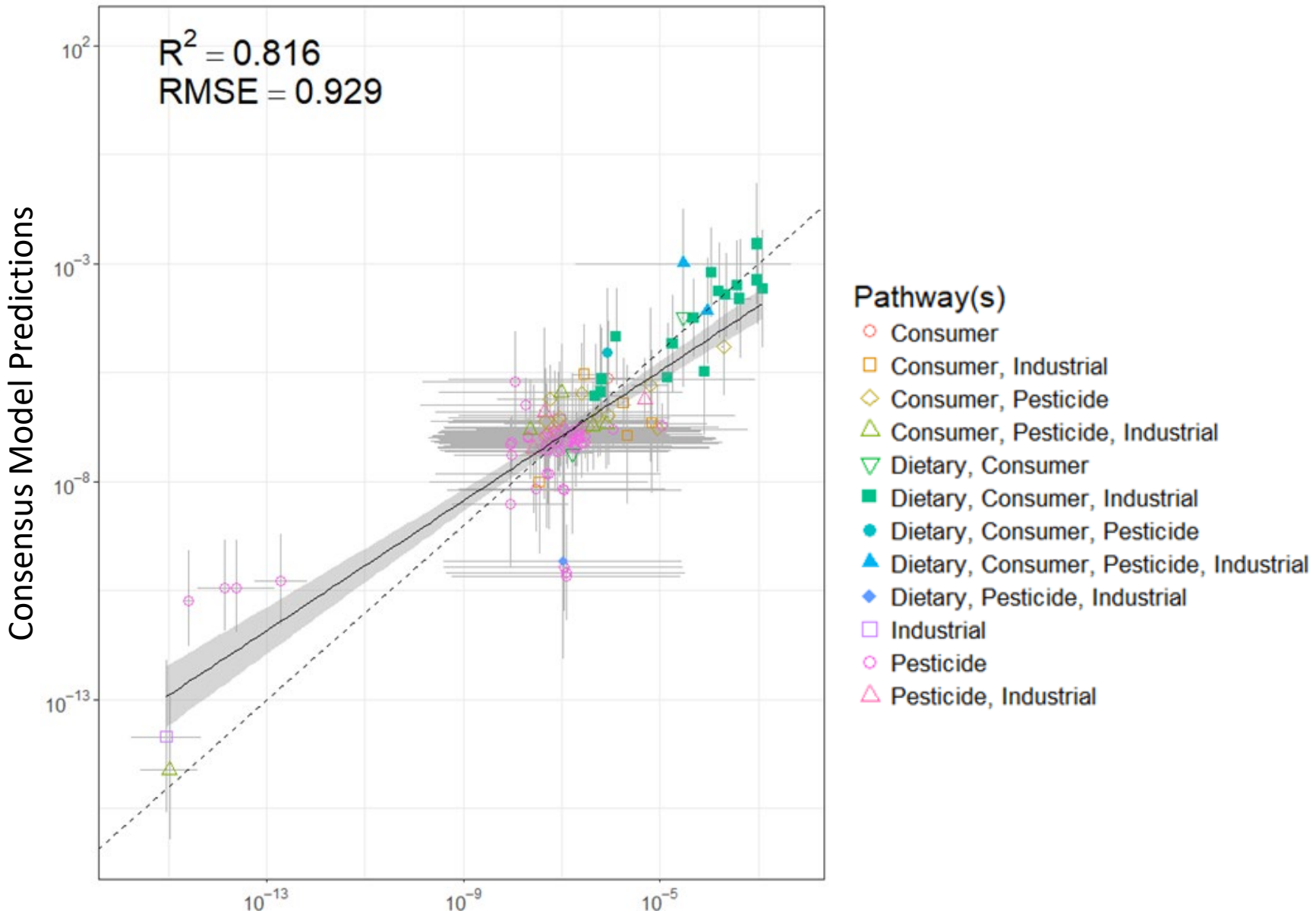
	NHANES Chemicals	Positives	Negatives	OOB Error Rate	Positives Error Rate	Balanced Accuracy	Sources of Positives	Sources of Negatives
Dietary	24	2523	8865	27	32	73	FDA CEDI, ExpoCast, CPDat (Food, Food Additive, Food Contact), NHANES Curation	Pharmapendium, CPDat (non-food), NHANES Curation
Near-Field	**49**	1622	567	26	24	74	CPDat (consumer_use, building_material), ExpoCast, NHANES Curation	CPDat (Agricultural, Industrial), FDA CEDI, NHANES Curation
Far-Field Pesticide	**94**	1480	6522	21	36	80	REDs, Swiss Pesticides, Stockholm Convention, CPDat (Pesticide), NHANES Curation	Pharmapendium, Industrial Positives, NHANES Curation
Far Field Industrial	**42**	5089	2913	19	16	81	CDR HPV, USGS Water Occurrence, NORMAN PFAS, Stockholm Convention, CPDat (Industrial, Industrial_Fluid), NHANES Curation	Pharmapendium, Pesticide Positives, NHANES Curation

Ring et al., 2019

Pathway-Based Consensus Modeling of NHANES

United States Environmental Protection Agency

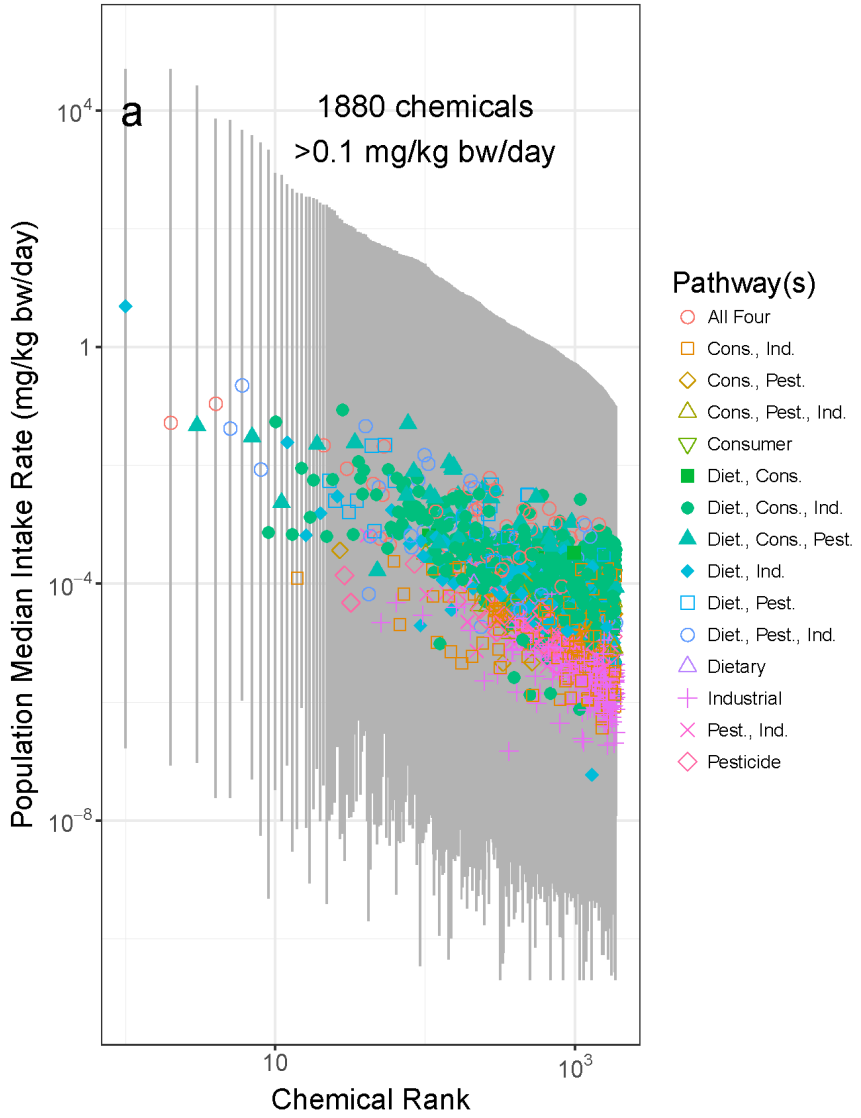
- Machine learning models were built for each of four exposure pathways
- Pathway predictions can be used for large chemical libraries
- Use prediction (and accuracy of prediction) as a prior for Bayesian analysis
- Each chemical may have exposure by multiple pathways



Ring et al., 2019

Consensus Modeling of Median Chemical Intake

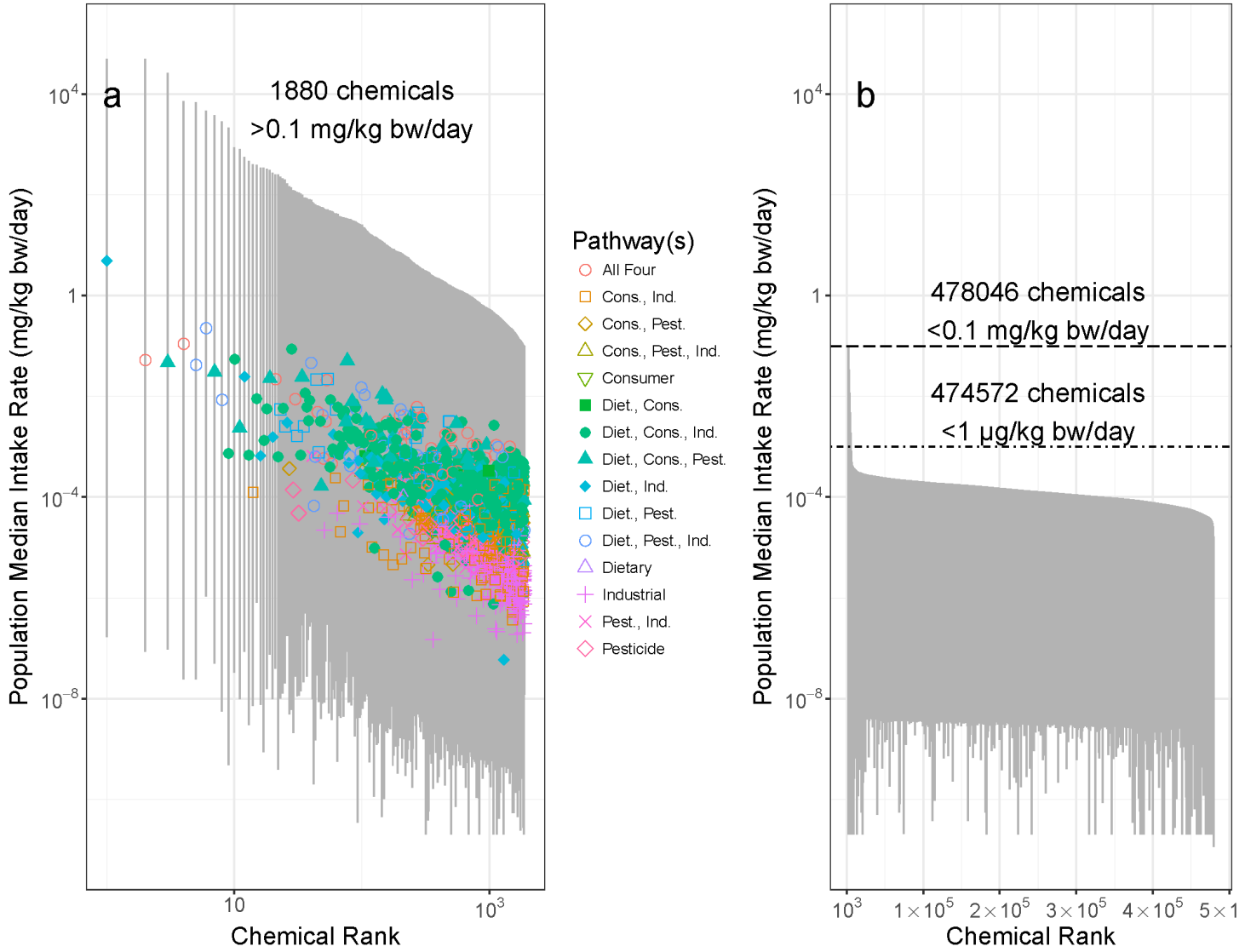
- We extrapolate to predict relevant pathway(s), median intake rate, and credible interval for each of 479,926 chemicals
- Of 687,359 chemicals evaluated, 30% have less than a 50% probability for exposure via any of the four pathways and are considered outside the “domain of applicability”
- This approach identifies 1,880 chemicals for which the median population intake rates may exceed 0.1 mg/kg bodyweight/day.



Ring et al., 2019

Consensus Modeling of Median Chemical Intake

- We extrapolate to predict relevant pathway(s), median intake rate, and credible interval for each of 479,926 chemicals
- Of 687,359 chemicals evaluated, 30% have less than a 50% probability for exposure via any of the four pathways and are considered outside the "domain of applicability"
- This approach identifies 1,880 chemicals for which the median population intake rates may exceed 0.1 mg/kg bodyweight/day.
- There is 95% confidence that the median intake rate is below 1 µg/kg BW/day for 474,572 compounds.



Ring et al., 2019

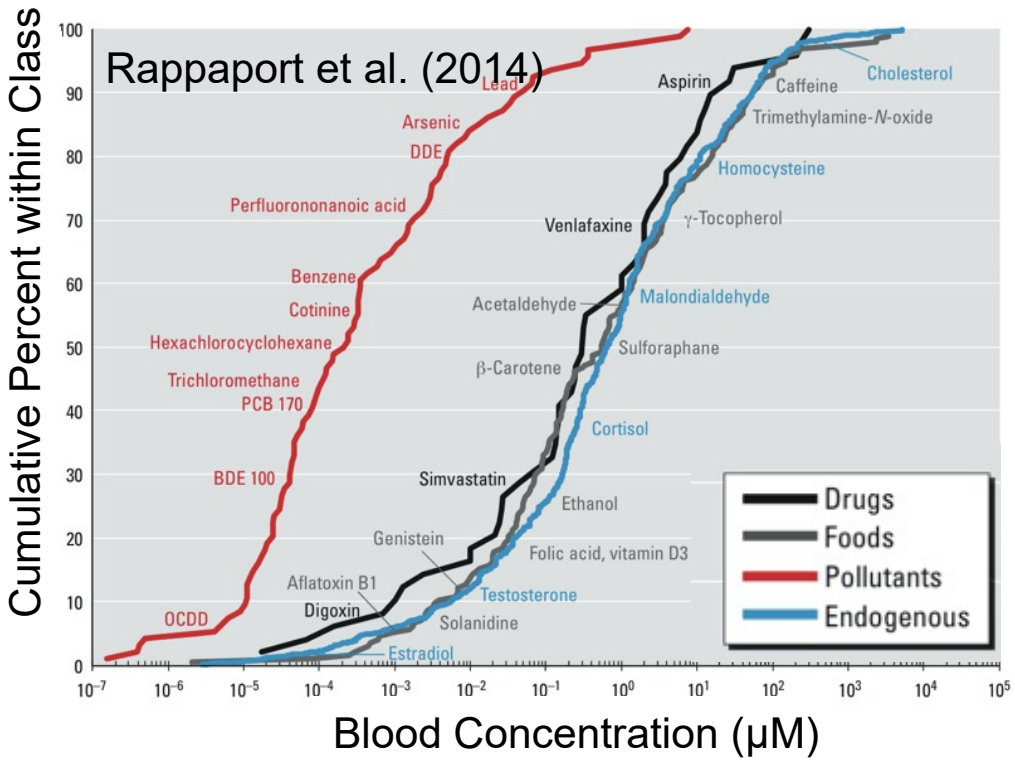
Conclusions

- We can make chemical-specific estimates of intake rate for hundreds of thousands of chemical
  - Only predicting median intake rate (and even that has large uncertainty)
  - Synthesizing as many models and other data as we can find

- Models incorporate Knowledge, Assumptions and Data (Macleod, et al., 2010)
  - The trick is to know which model to use and when
  - Machine learning models allow educated guesses

- We are using existing chemical data to predict pathways
  - Not all chemicals fit within the domain of applicability
  - Need better training data for random (non-targeted analysis of environmental media needed)

- Eventually we have got to go beyond NHANES
  - Current evaluation based upon 114 chemicals
  - Non-targeted analysis of blood may eventually be possible



ExpoCast Project (Exposure Forecasting)

United States Environmental Protection Agency

NCCT
Chris Grulke
Greg Honda*
Richard Judson
Ann Richard
Risa Sayre*
Mark Sfeir*
Rusty Thomas
John Wambaugh
Antony Williams

NRMRL
Xiaoyu Liu

NHEERL
Linda Adams
Christopher Ecklund
Marina Evans
Mike Hughes
Jane Ellen Simmons
Tamara Tal

NERL
Cody Addington*
Namdi Brandon*
Alex Chao*
Kathie Dionisio
Peter Egeghy
Hongtai Huang*
Kristin Isaacs
Ashley Jackson*
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Seth Newton
Katherine Phillips
Paul Price
Jeanette Reyes*
Randolph Singh*
Marci Smeltz
Jon Sobus
John Streicher*
Mark Strynar
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
Barbara Wetmore

***Trainees**

Collaborators

Arnot Research and Consulting
Jon Arnot
Johnny Westgate
Institut National de l'Environnement et des Risques (INERIS)
Frederic Bois
Integrated Laboratory Systems
Kamel Mansouri
National Toxicology Program
Mike Devito
Steve Ferguson
Nisha Sipes
Ramboll
Harvey Clewell
ScitoVation
Chantel Nicolas
Silent Spring Institute
Robin Dodson
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Aylward
Technical University of Denmark
Peter Fantke
Tox Strategies
Caroline Ring
Miyoung Yoon
Unilever
Beate Nicol
Cecilie Rendal
Ian Sorrell
United States Air Force
Heather Pangburn
Matt Linakis
University of California, Davis
Deborah Bennett
University of Michigan
Olivier Jolliet
University of Texas, Arlington
Hyeong-Moo Shin

United States Environmental Protection Agency

References

- Arnot, J. A.; et al., Develop Sub-Module for Direct Human Exposures to Consumer Products. Technical Report for the U.S. Environmental Protection Agency; ARC Arnot Research & Consulting, Inc.: Toronto, ON, Canada, 2014.
- Bennett, D. H.; Furtaw, E. J., Fugacity-based indoor residential pesticide fate model. Environmental Science & Technology 2004, 38, (7), 2142-2152.
- Breyer, Stephen. Breaking the vicious circle: Toward effective risk regulation. Harvard University Press, 2009
- Burwell, Sylvia M., et al. "Memorandum for the Heads of Executive Departments and Agencies: Open Data Policy--Managing Information as an Asset." (2013).
- Collins, Francis S., George M. Gray, and John R. Bucher. "Transforming environmental health protection." *Science (New York, NY)* 319.5865 (2008): 906.
- Dix, David J., et al. "The ToxCast program for prioritizing toxicity testing of environmental chemicals." *Toxicological Sciences* 95.1 (2006): 5-12.
- Egeghy, P. P., et al. (2012). The exposure data landscape for manufactured chemicals. Science of the Total Environment, 414, 159-166.
- Ernstoff, A. S., et al.., High-throughput migration modelling for estimating exposure to chemicals in food packaging in screening and prioritization tools. Food and Chemical Toxicology 2017, 109, 428-438.
- Huang, Lt al.., A review of models for near-field exposure pathways of chemicals in consumer products. Science of The Total Environment 2017, 574, 1182-1208.
- Huang, L.; Jolliet, O., A parsimonious model for the release of volatile organic compounds (VOCs) encapsulated in products. Atmospheric Environment 2016, 127, 223-235.
- Jolliet, O. et al. Defining Product Intake Fraction to Quantify and Compare Exposure to Consumer Products. Environmental Science & Technology 2015, 49, (15), 8924-8931.
- Kavlock, Robert J., et al. "Accelerating the pace of chemical risk assessment." Chemical research in toxicology 31.5 (2018): 287-290.
- MacLeod, Matthew, et al. "The state of multimedia mass-balance modeling in environmental science and decision-making." (2010): 8360-8364
- McEachran, Andrew D., Jon R. Sobus, and Antony J. Williams. "Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard." *Analytical and bioanalytical chemistry* 409.7 (2017): 1729-1735.
- National Research Council. (1983). Risk Assessment in the Federal Government: Managing the Process Working Papers. National Academies Press.
- Obama, B. H. "Executive Order 13642: Making Open and Machine Readable the New Default for Government Information. Washington, DC: Office of the Executive." (2013).
- Park, Youngja, H., et al. "High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring." Toxicology 295:47-55 (2012)
- Pearce, Robert G., et al. "Httk: R package for high-throughput toxicokinetics." Journal of statistical software 79.4 (2017): 1.
- Rager, Julia E., et al. "Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring." *Environment international* 88 (2016): 269-280.
- Rappaport, Stephen M., et al. "The blood exposome and its role in discovering causes of disease." Environmental health perspectives 122.8 (2014): 769-774.
- Ring, Caroline L., et al. "Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability." Environment International 106 (2017): 105-118.
- Ring, Caroline L., et al. "Consensus Modeling of Median Chemical Intake for the US Population Based on Predictions of Exposure Pathways." Environmental science & technology 53.2 (2018): 719-732
- Shin, H.-M.; McKone, T. E.; Bennett, D. H., Intake Fraction for the Indoor Environment: A Tool for Prioritizing Indoor Chemical Sources. Environmental Science & Technology 2012, 46, (18), 10063-10072.
- Shin, Hyeong-Moo, et al. "Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays." Environmental science & technology 49.11 (2015): 6760-6771.
- Sobus, Jon R., et al. "Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA." *Journal of exposure science & environmental epidemiology* (2017): 1.
- Tan, Yu-Mei, Kai H. Liao, and Harvey J. Clewell III. "Reverse dosimetry: interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling." Journal of Exposure Science and Environmental Epidemiology 17.7 (2007): 591.
- Wallace et al., "The TEAM Study: Personal exposures to toxic substances in air, drinking water, and breath of 400 residents of New Jersey, North Carolina, and North Dakota ." Environmental Research 43: 209-307 (1987)
- Wambaugh, John F., et al. "High-throughput models for exposure-based chemical prioritization in the ExpoCast project." Environmental science & technology 47.15 (2013): 8479-848.
- Wambaugh, John F., et al. "High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals." Environmental science & technology (2014).
- Wetmore, Barbara A., et al. "Integration of dosimetry, exposure and high-throughput screening data in chemical toxicity assessment." Toxicological Sciences (2012): kfr254.
- Wetmore, Barbara A., et al. "Incorporating High-Throughput Exposure Predictions with Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing." Toxicological Sciences 148.1 (2015): 121-136.
- Zhang, X.; Arnot, J. A.; Wania, F., Model for screening-level assessment of near-field human exposure to neutral organic chemicals released indoors. Environmental science & technology 2014, 48, (20), 12312-12319.