United States Environmental Protection Agency

The Systematic Empirical Evaluation of Models (SEEM) framework:

High Throughput Exposure Science for Chemical Decision Making

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

Computational Toxicology Community of Practice Webinar

February 28, 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)

November 29, 2014

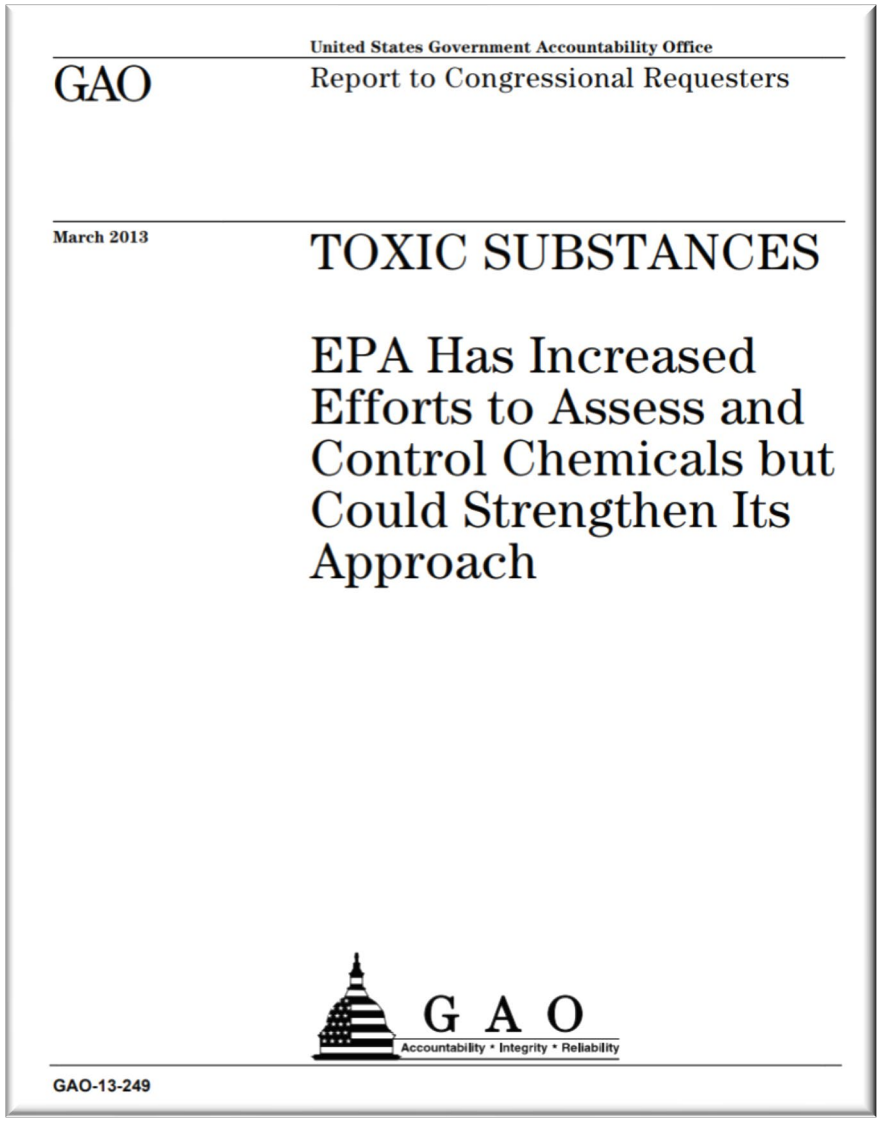
Chemical Regulation in the United States

- Most other chemicals, ranging from industrial waste to dyes to packing materials, are covered by the Toxic Substances Control Act (TSCA)

- Thousands of chemicals on the market were "grandfathered" in without assessment

  Judson et al. (2009), Egeghy et al. (2012), Wetmore et al. (2015)

"Tens of thousands of chemicals are listed with the Environmental Protection Agency (EPA) for commercial use in the United States, with an average of 600 new chemicals listed each year."
U.S. Government Accountability Office


United States Government Accountability Office

GAO

Report to Congressional Requesters

March 2013

TOXIC SUBSTANCES

EPA Has Increased Efforts to Assess and Control Chemicals but Could Strengthen Its Approach

G A O

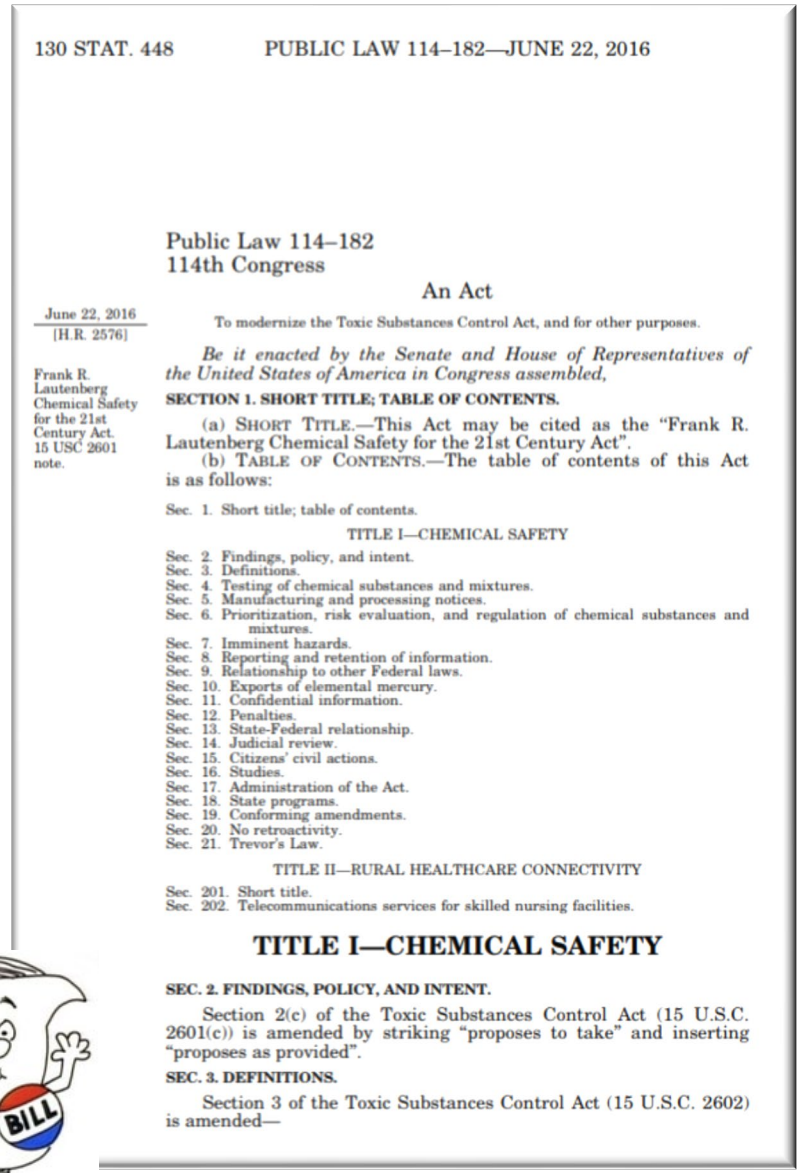
Accountability • Integrity • Reliability

GAO-13-249

March, 2013

Chemical Regulation in the United States

- TSCA was updated in June, 2016 to allow more rapid evaluation of chemicals (Frank R. Lautenberg Chemical Safety for the 21st Century Act)
- New approach methodologies (NAMs) are being considered to inform prioritization of chemicals for testing and evaluation (Kavlock et al., 2018)
- EPA has released a "A Working Approach for Identifying Potential Candidate Chemicals for Prioritization" (September, 2018)



130 STAT. 448 PUBLIC LAW 114–182—JUNE 22, 2016

Public Law 114–182
114th Congress

An Act

To modernize the Toxic Substances Control Act, and for other purposes.

June 22, 2016
[H.R. 2576]

Frank R. Lautenberg Chemical Safety for the 21st Century Act.
15 USC 2601 note.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Frank R. Lautenberg Chemical Safety for the 21st Century Act".

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:

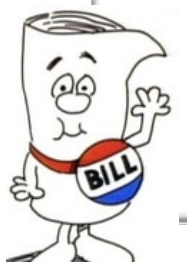
TITLE I—CHEMICAL SAFETY

SEC. 2. FINDINGS, POLICY, AND INTENT.

Section 2(c) of the Toxic Substances Control Act (15 U.S.C. 2601(c)) is amended by striking "proposes to take" and inserting "proposes as provided".

SEC. 3. DEFINITIONS.

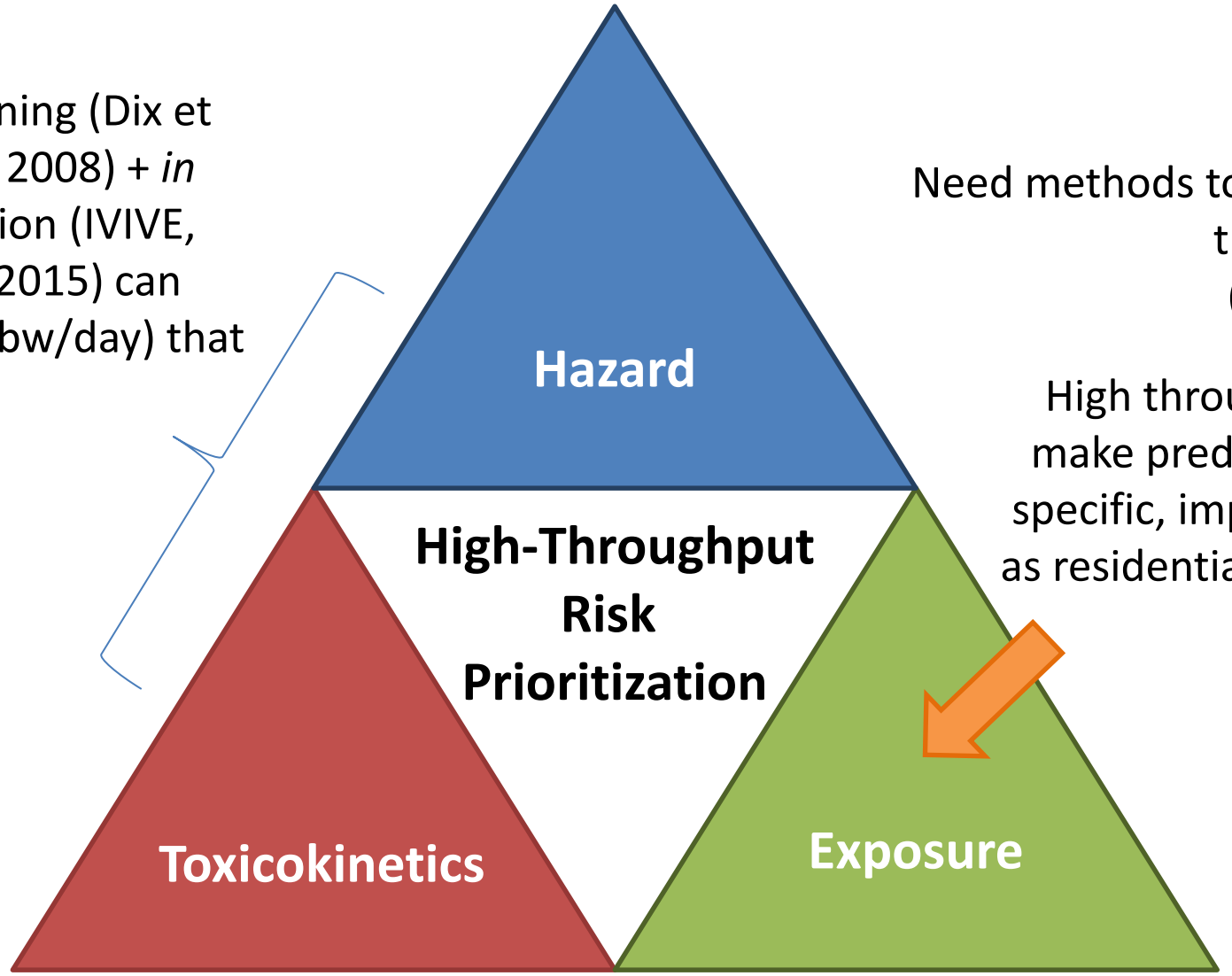
Section 3 of the Toxic Substances Control Act (15 U.S.C. 2602) is amended—



June 22, 2016

Risk = Hazard x Exposure

High throughput screening (Dix et al., 2006, Collins et al., 2008) + *in vitro-in vivo* extrapolation (IVIVE, Wetmore et al., 2012, 2015) can predict a dose (mg/kg bw/day) that might be adverse

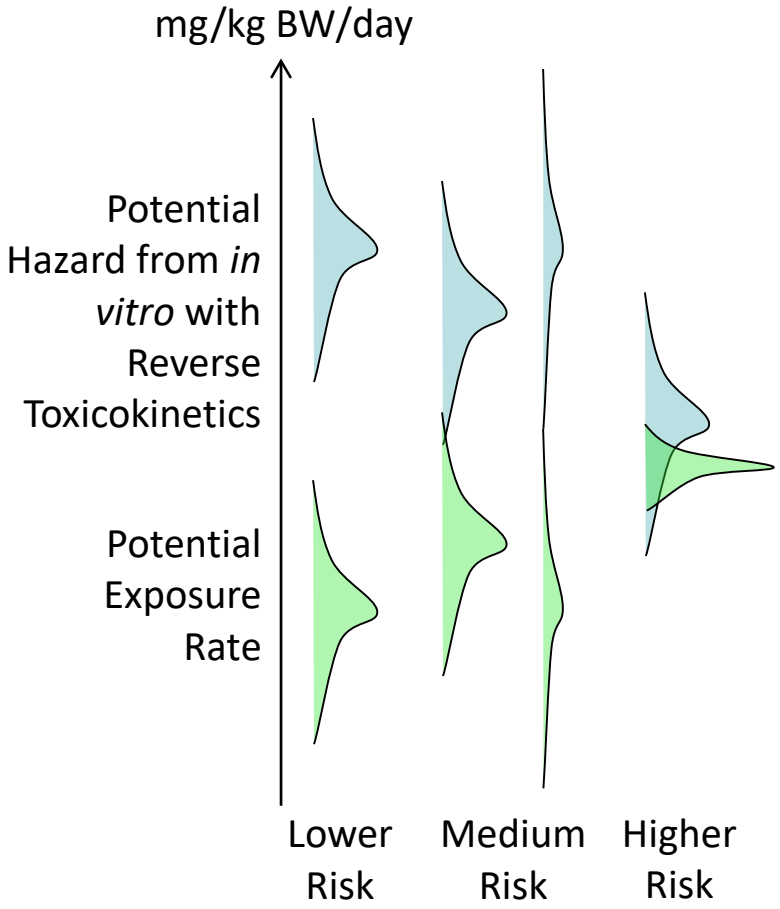


Need methods to forecast exposure for thousands of chemicals (Wetmore et al., 2015)

High throughput models exist to make predictions of exposure via specific, important pathways such as residential product use and diet

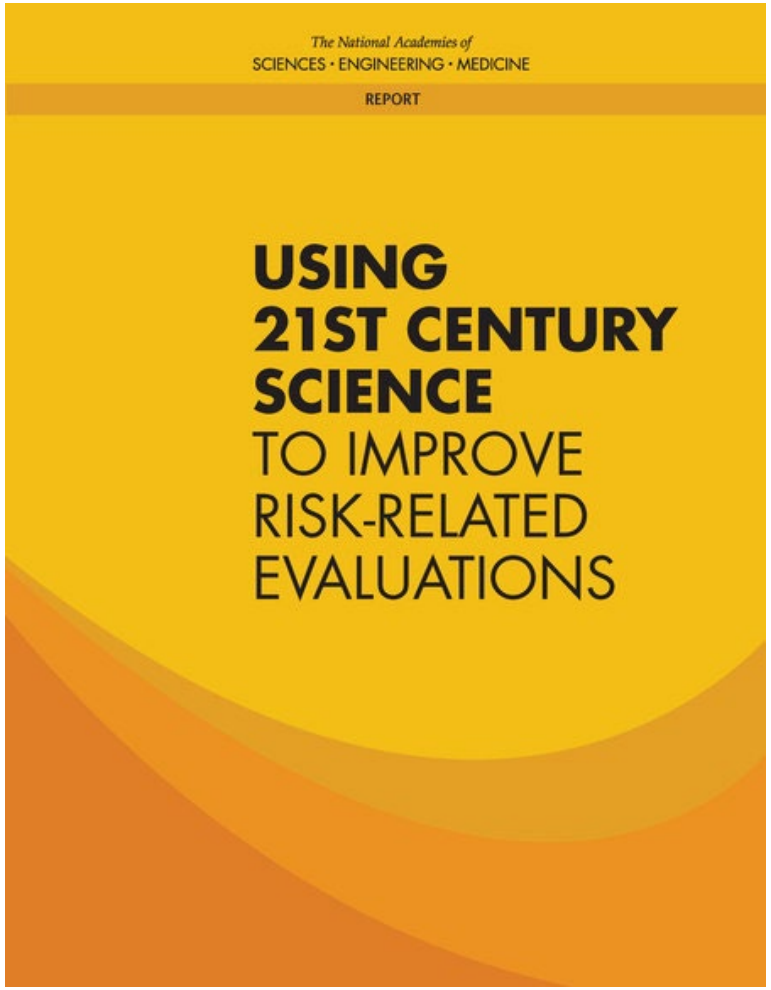
NRC (1983)

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…"

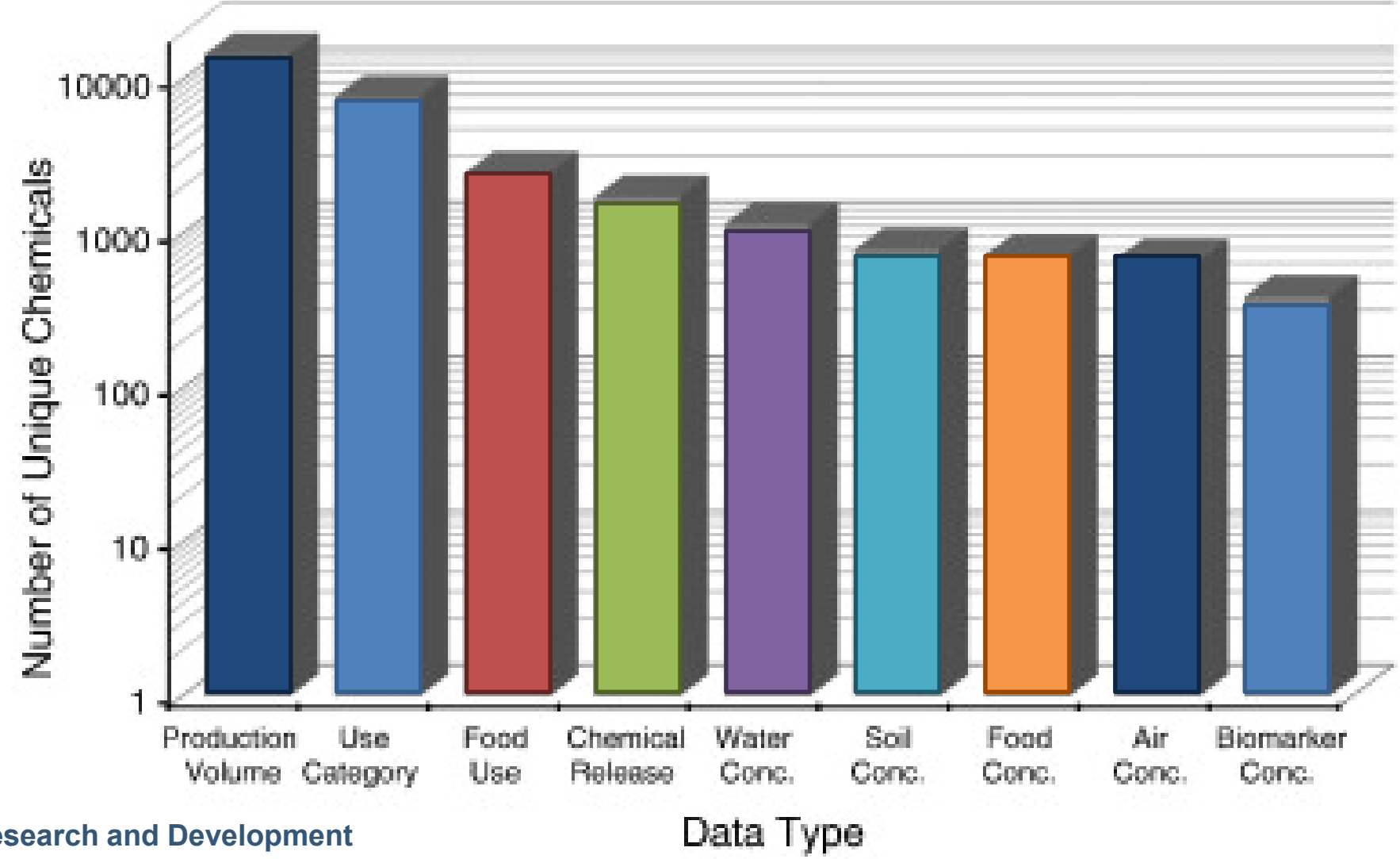
"…The committee sees the potential for the application of **computational exposure science** to be highly valuable and credible for comparison and **priority-setting among chemicals in a risk-based context**."



January, 2017

Limited Available Data for Exposure Estimation

Most chemicals lack public exposure-related data beyond production volume (Egeghy et al., 2012)



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**

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

Forecasting Exposure is a Systems Problem

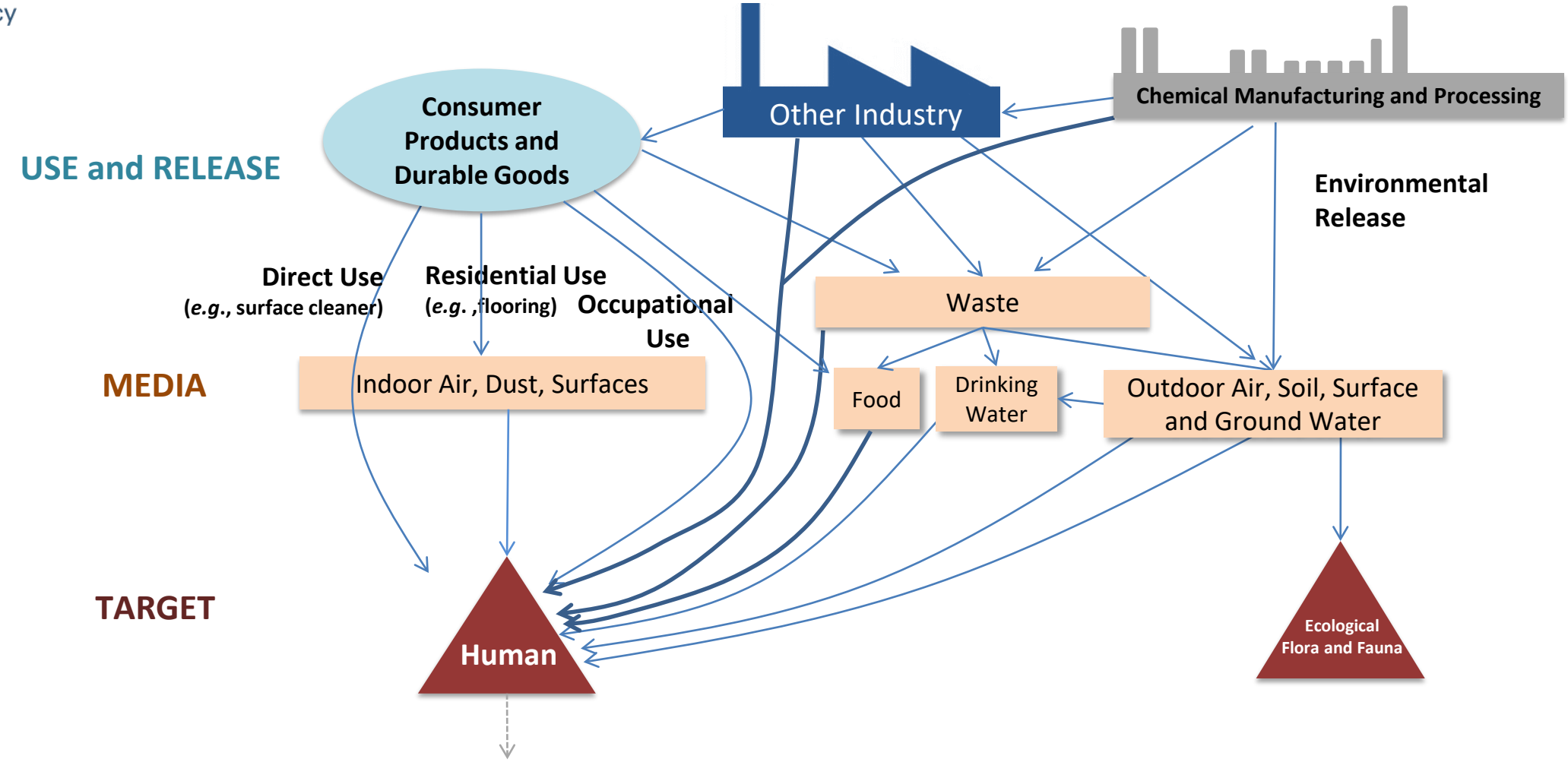
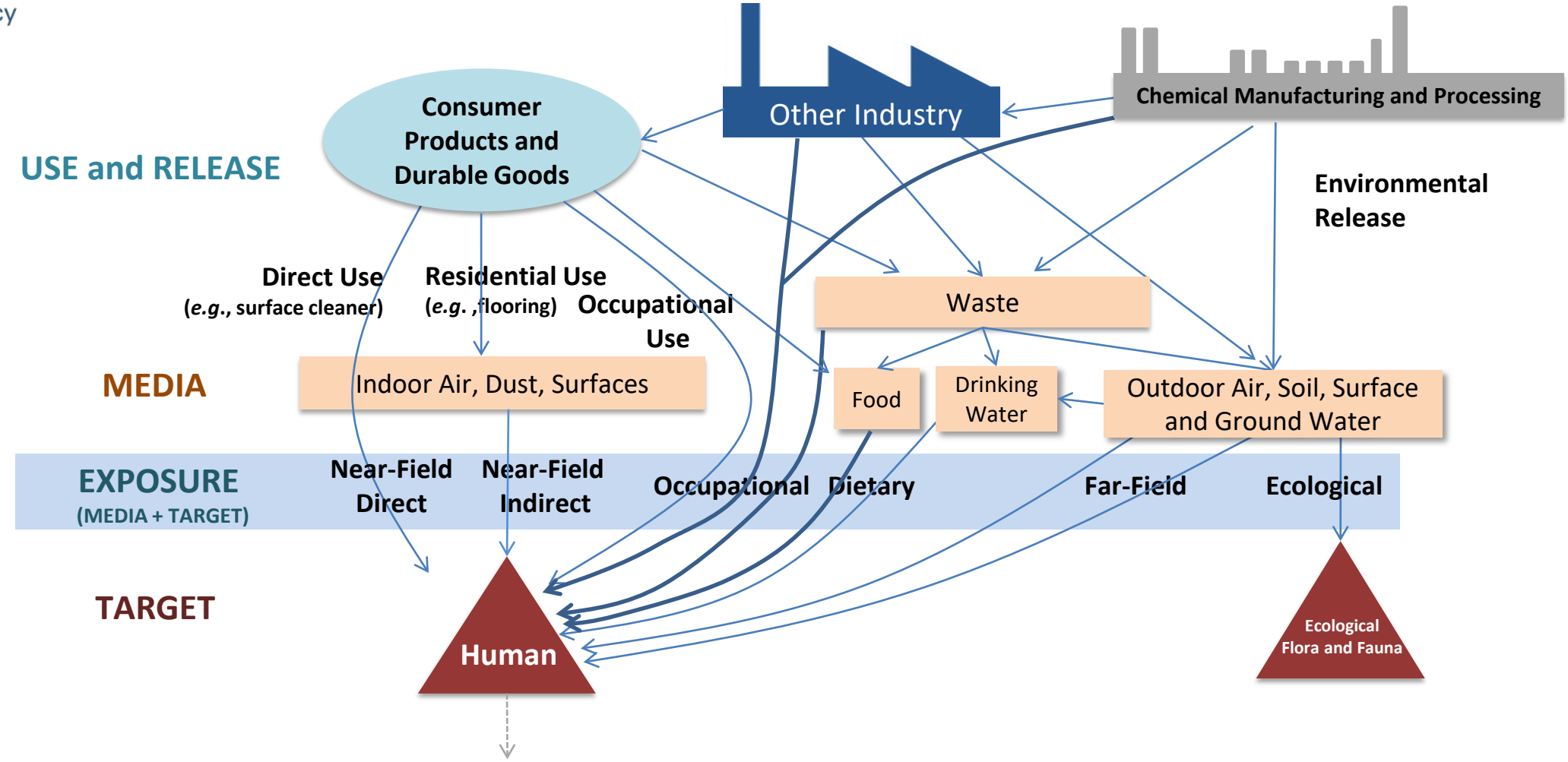


Figure from Kristin Isaacs

Source-based Exposure Pathways



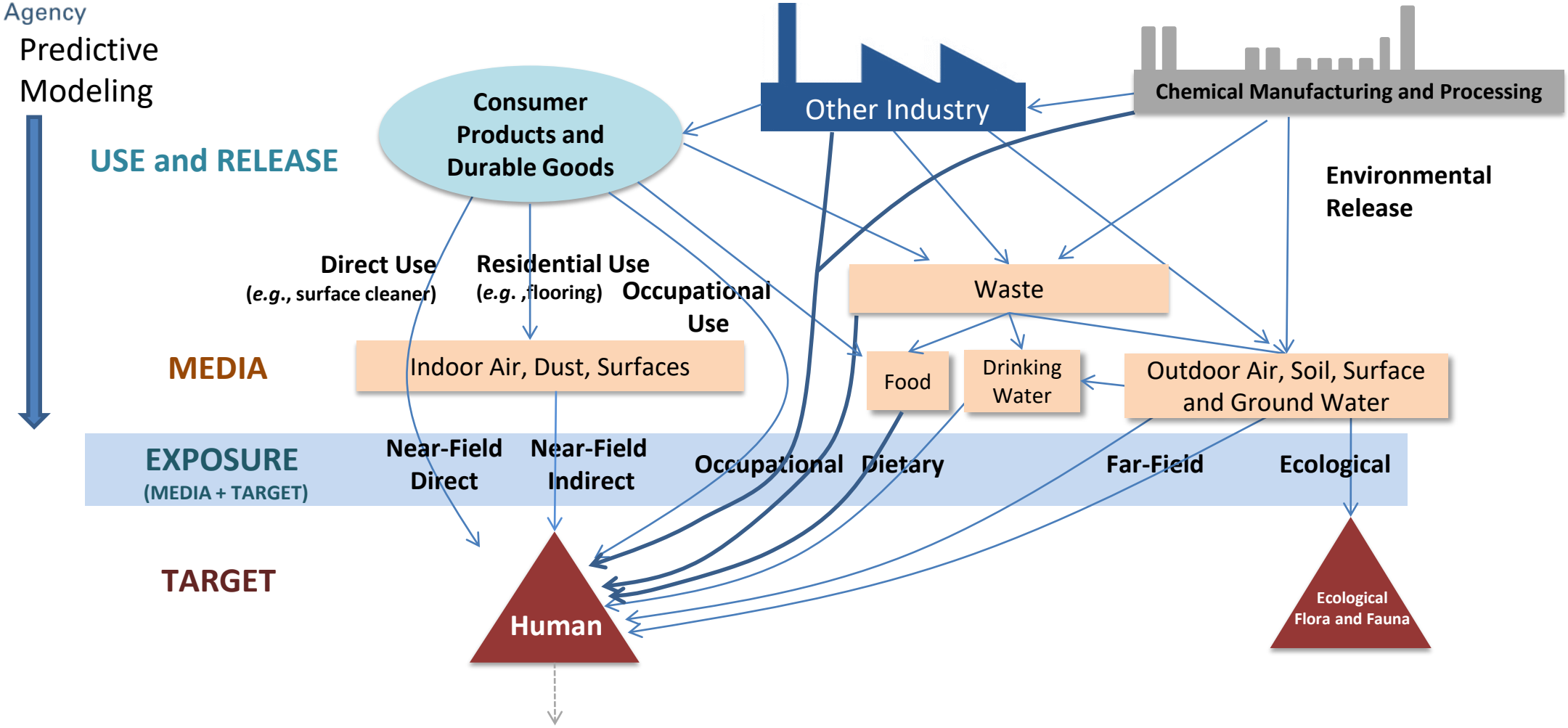
The exposure event itself is often unobservable

Figure from Kristin Isaacs

Models to Predict Exposure

United States Environmental Protection Agency

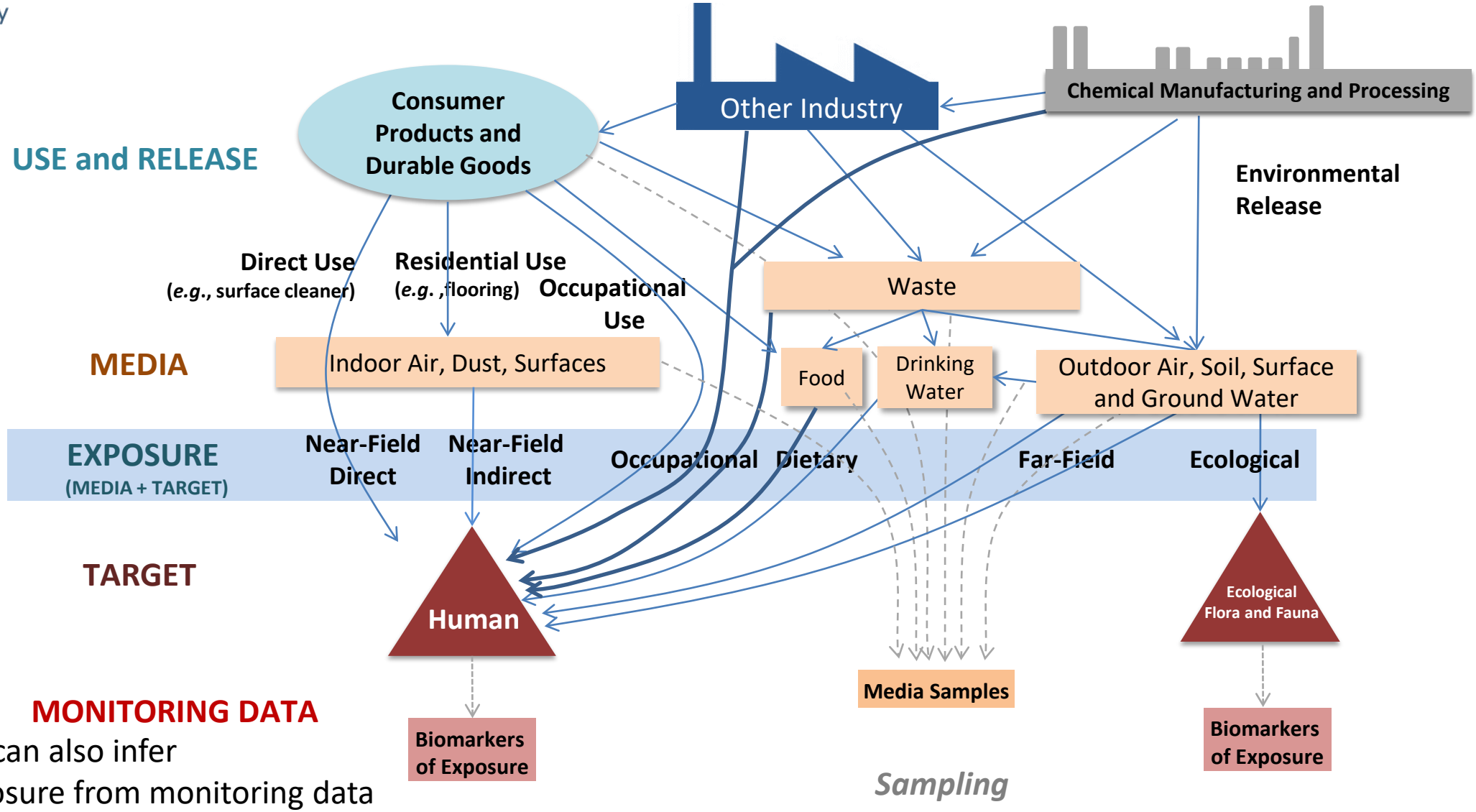
Predictive Modeling



We can try to predict exposure by describing the process leading to exposure

Figure from Kristin Isaacs

Monitoring Data



MONITORING DATA

We can also infer exposure from monitoring data

Figure from Kristin Isaacs

Models to Infer Exposure

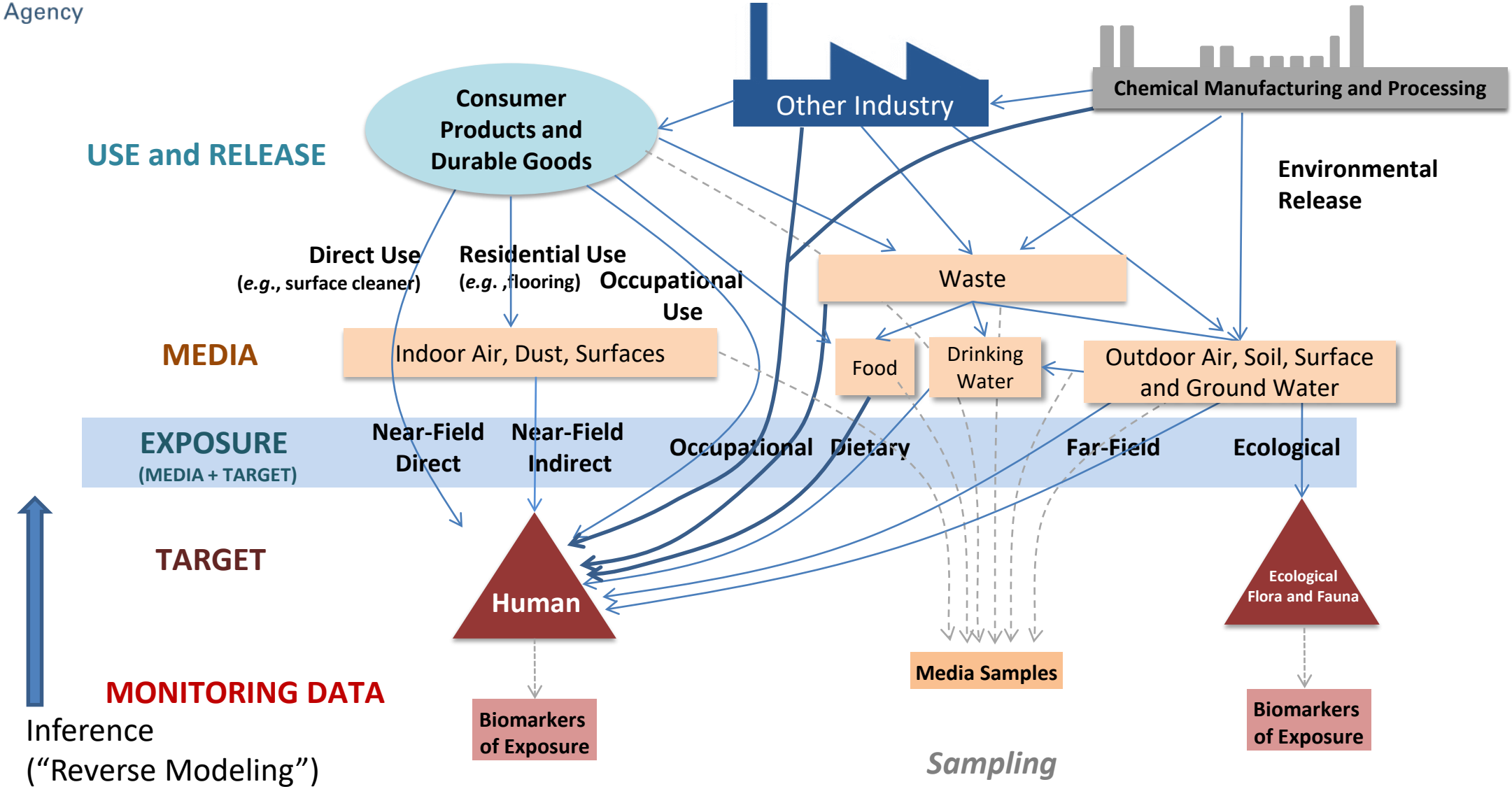


Figure from Kristin Isaacs

Evaluating Models with Monitoring Data

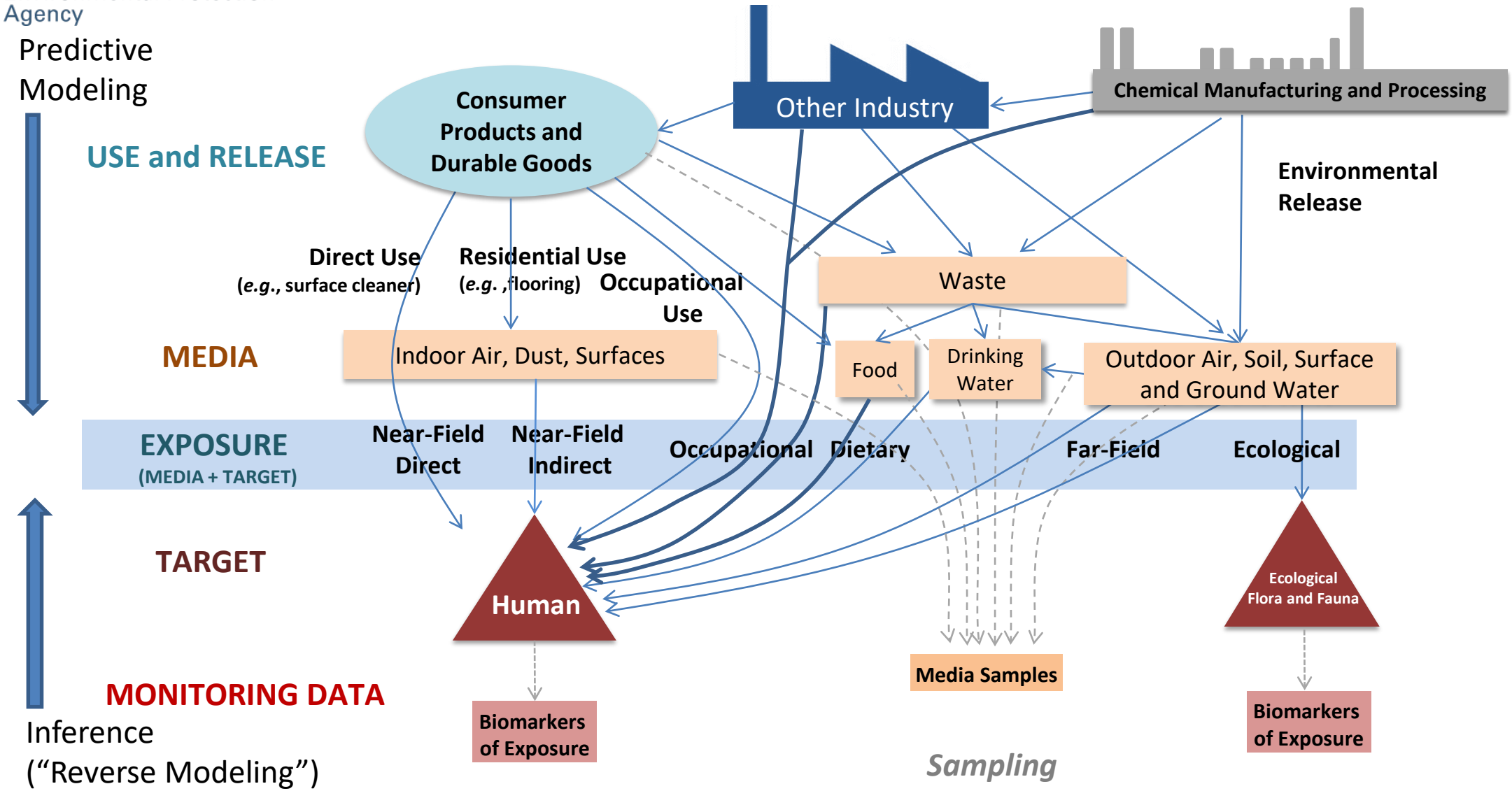
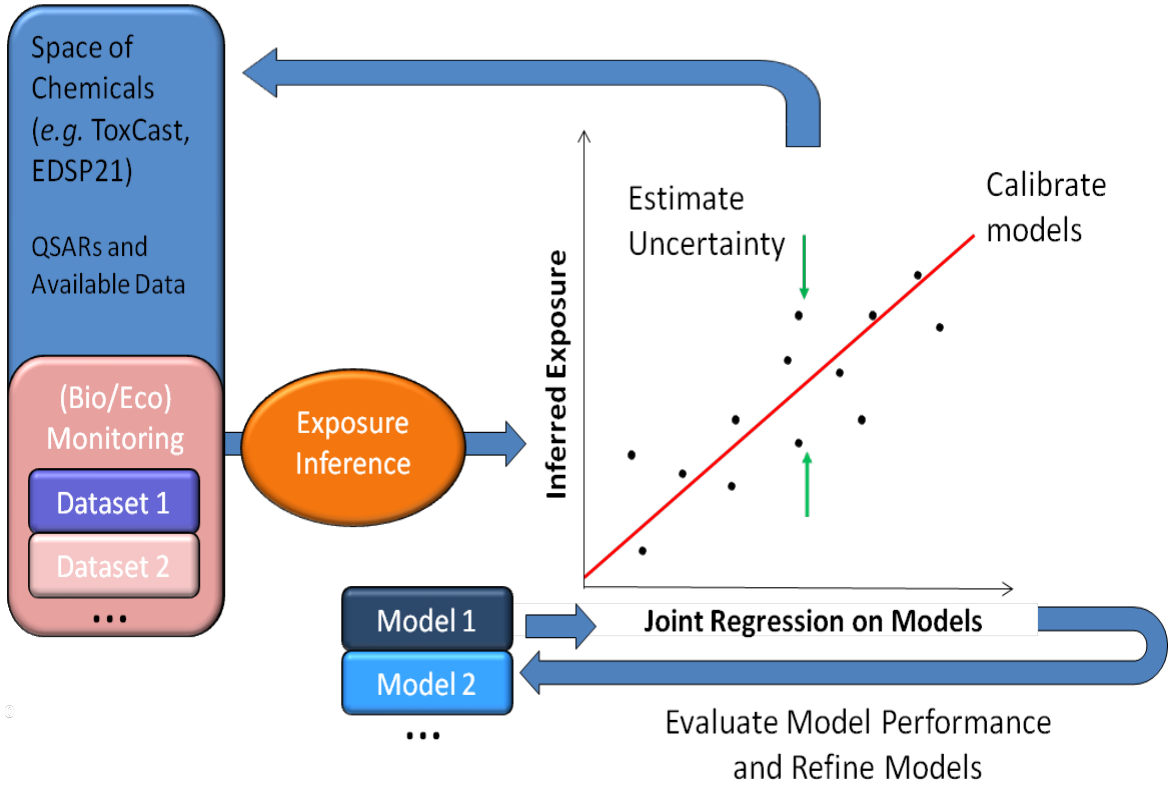


Figure from Kristin Isaacs

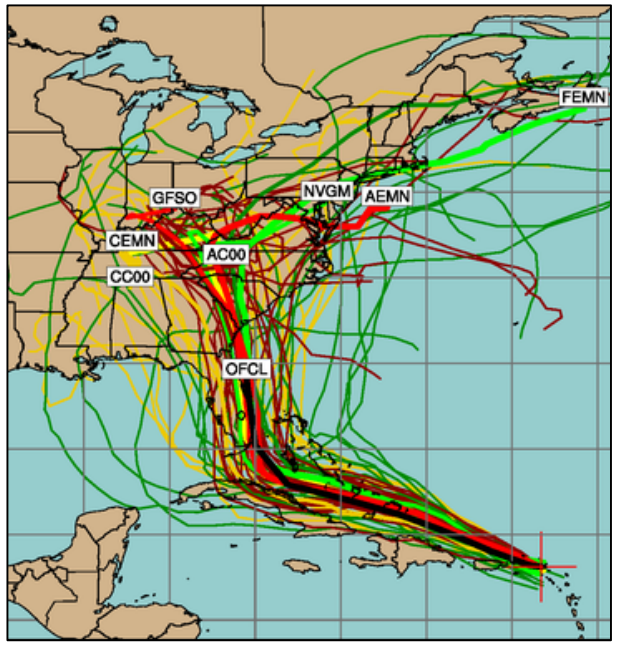
Consensus Exposure Predictions with the SEEM Framework

- 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?

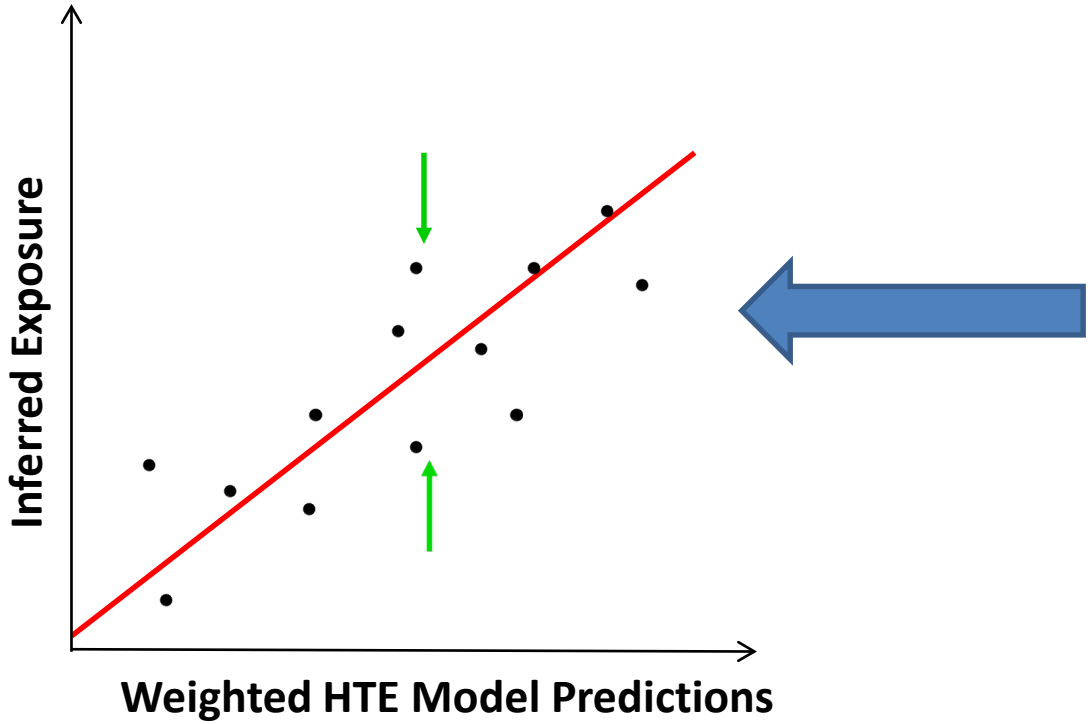
Hurricane Path Prediction is an Example of Integrating Multiple Models



SEEM is a Linear Regression

Multiple regression models:

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

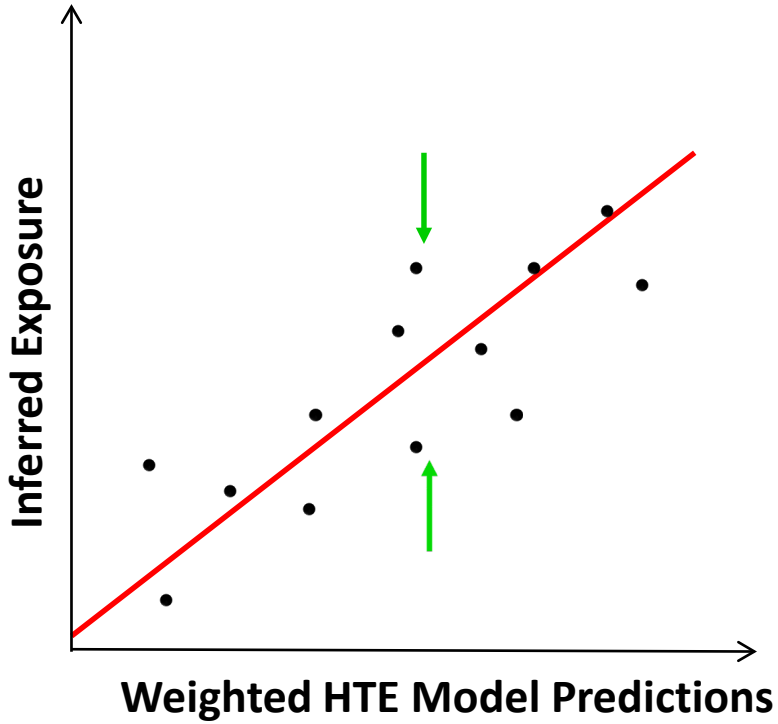


$\varepsilon \sim N(0, \sigma^2)$
Residual error, unexplained by the regression model

SEEM is a Linear Regression

Multiple regression models:

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



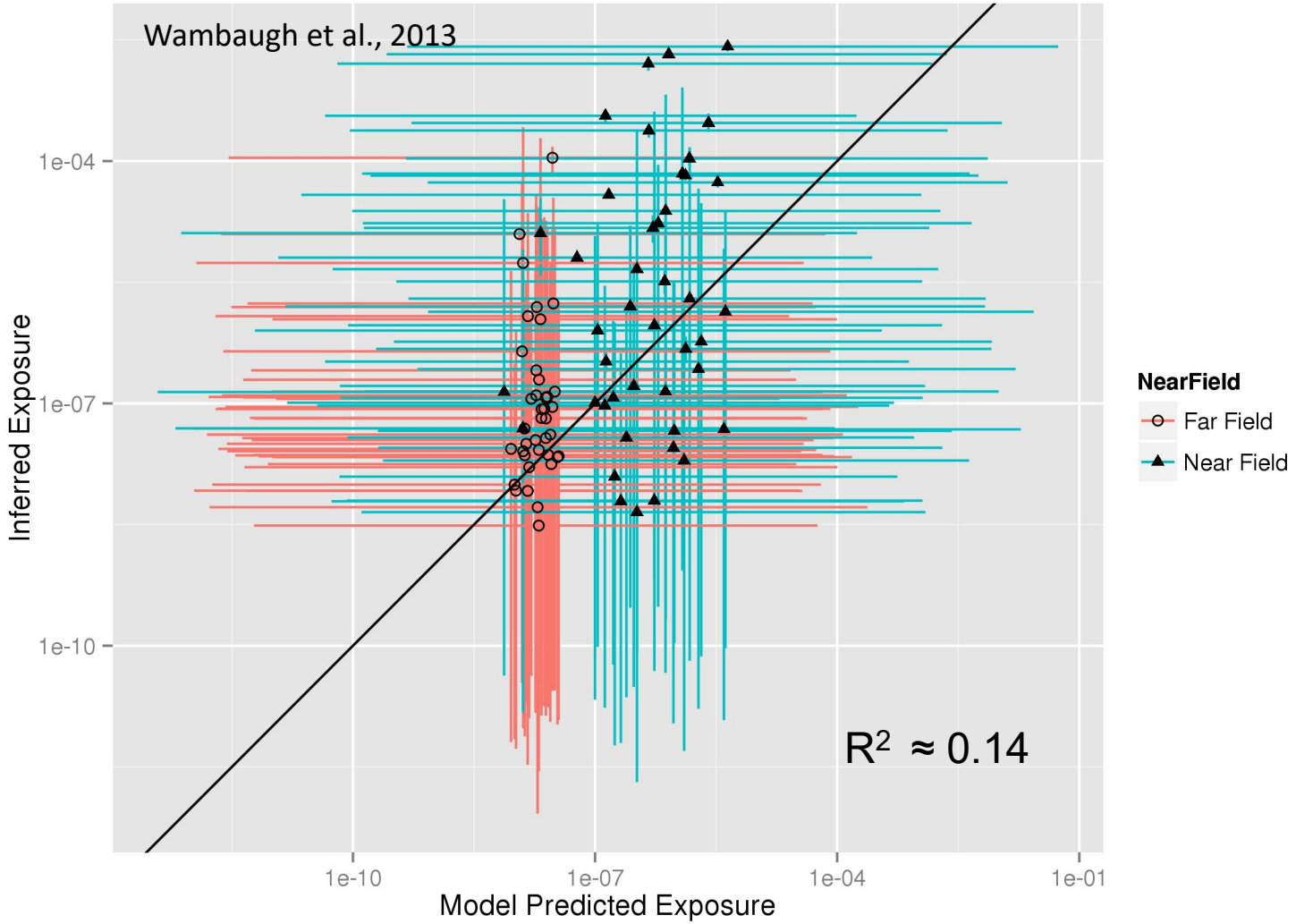
Not all models have predictions for all chemicals

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

What do we do for the rest?

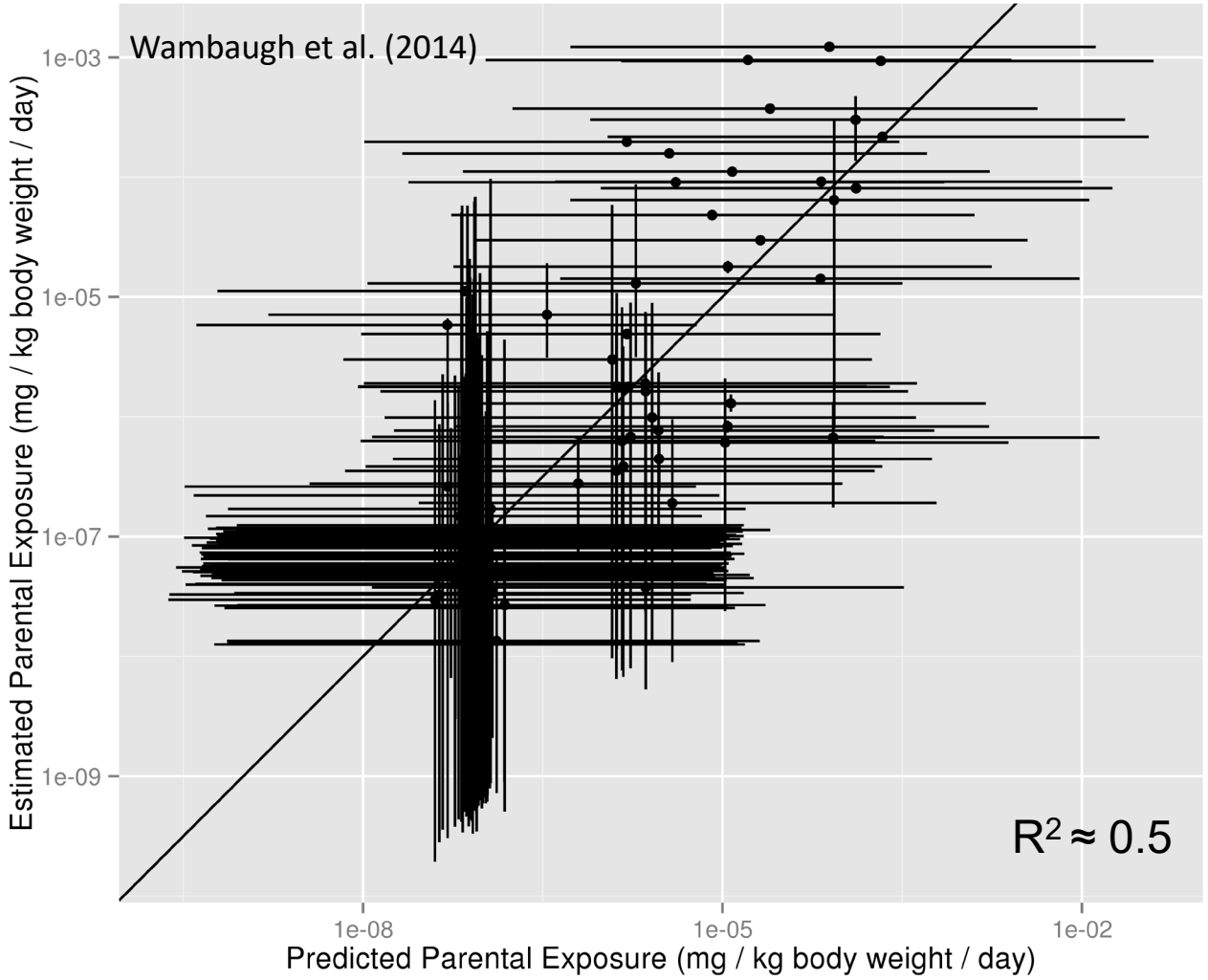
- Assign the average value?
- Zero?

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

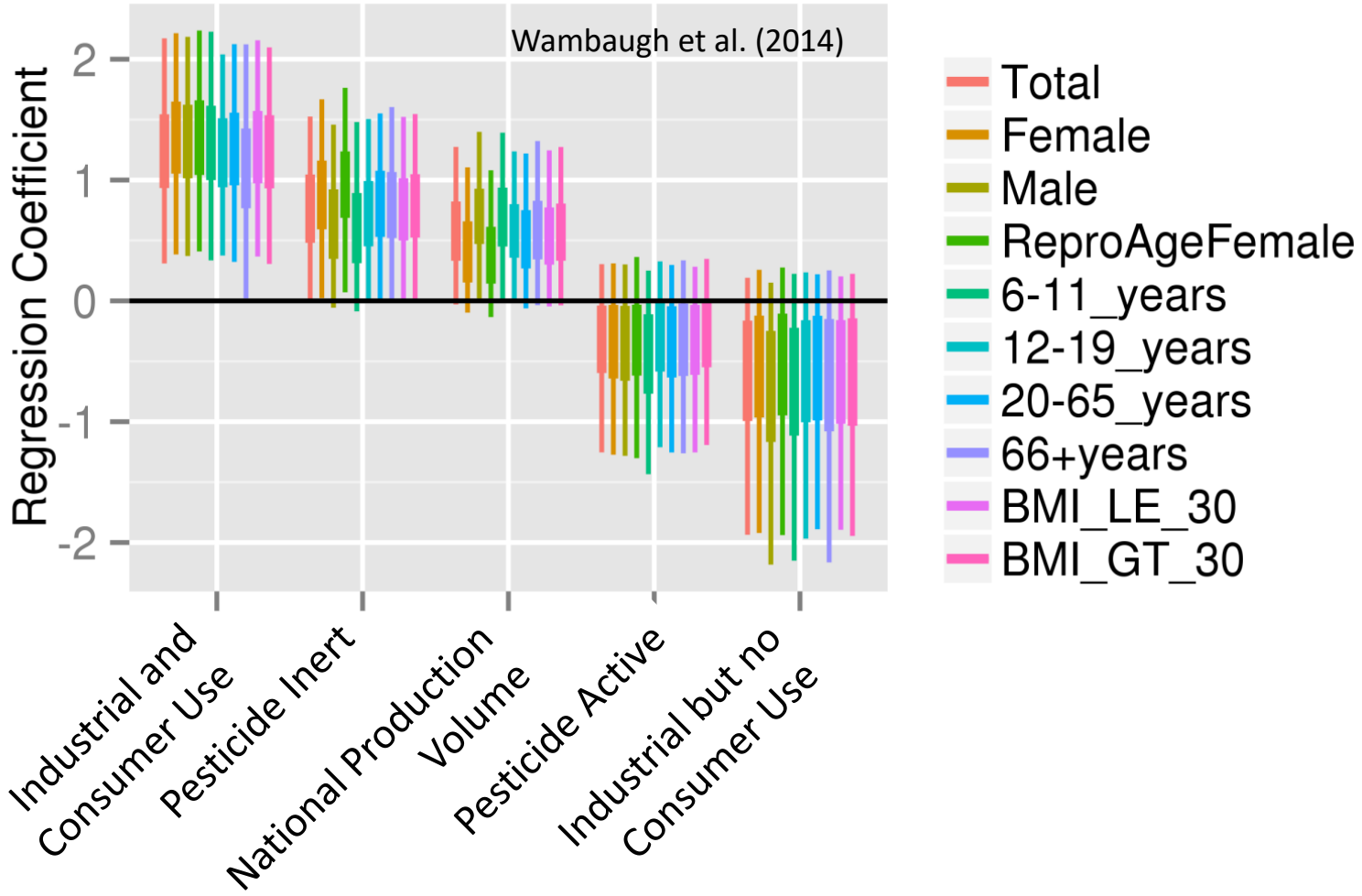


$R^2 \approx 0.5$ indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates

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

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†

What Do We Know About Chemical Use? Chemicals and Products Database

United States Environmental Protection Agency


Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Development of a consumer product ingredient database for chemical exposure screening and prioritization

M.-R. Goldsmith [a,*], C.M. Grulke [a], R.D. Brooks [b], T.R. Transue [c], Y.M. Tan [a], A. Frame [a,e], P.P. Egeghy [a], R. Edwards [d], D.T. Chang [a], R. Tornero-Velez [a], K. Isaacs [a], A. Wang [a,e], J. Johnson [a], K. Holm [a], M. Reich [f], J. Mitchell [g], D.A. Vallero [a], L. Phillips [a], M. Phillips [a], J.F. Wambaugh [a], R.S. Judson [a], T.J. Buckley [a], C.C. Dary [a]

MSDS Data

Occurrence and quantitative chemical composition

Broad "index" of chemical uses


Contents lists available at ScienceDirect

Toxicology Reports

journal homepage: www.elsevier.com/locate/toxrep

Exploring consumer exposure pathways and patterns of use for chemicals in the environment

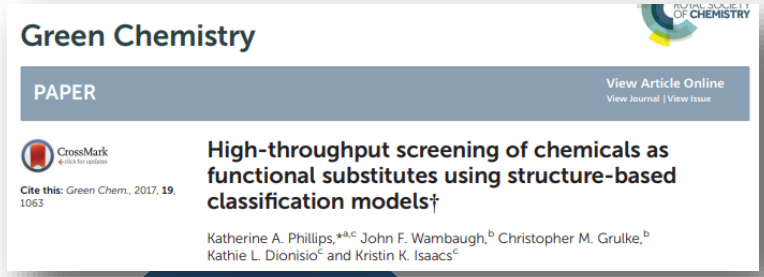
Kathie L. Dionisio [a], Alicia M. Frame [b,1], Michael-Rock Goldsmith [a,2], John F. Wambaugh [b], Alan Liddell [c,3], Tommy Cathey [d], Doris Smith [b], James Vail [b], Alexi S. Ernstoff [e], Peter Fantke [e], Olivier Jolliet [f],

CPCat

CPDat

EPA United States Environmental Protection Agency Chemistry Dashboard

https://comptox.epa.gov/dashboard


Green Chemistry

PAPER

View Article Online

View Journal | View Issue

Cite this: Green Chem., 2017, 19, 1063

High-throughput screening of chemicals as functional substitutes using structure-based classification models†

Katherine A. Phillips,*[a,c] John F. Wambaugh,[b] Christopher M. Grulke,[b] Kathie L. Dionisio[c] and Kristin K. Isaacs[c]

Functional Use Data

The roles that chemicals serve in products


Journal of Exposure Science and Environmental Epidemiology (2018) 28, 216–222

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved 1559-0631/18

www.nature.com/jes

ORIGINAL ARTICLE

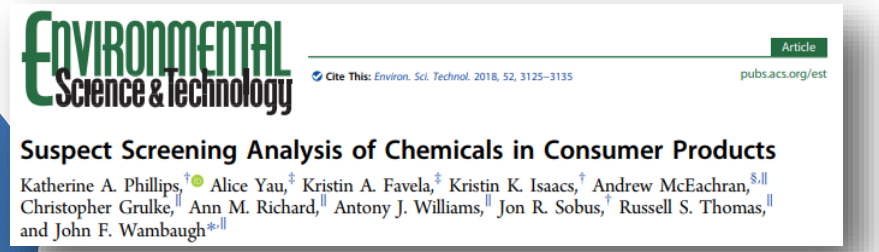
Consumer product chemical weight fractions from ingredient lists

Kristin K. Isaacs[1], Katherine A. Phillips[1], Derya Biryol[1,2], Kathie L. Dionisio[1] and Paul S. Price[1]

Ingredient Lists

Occurrence data

Measured Data


Environmental Science & Technology

Article

Cite This: Environ. Sci. Technol. 2018, 52, 3125−3135

pubs.acs.org/est

Suspect Screening Analysis of Chemicals in Consumer Products

Katherine A. Phillips,[†] Alice Yau,[‡] Kristin A. Favela,[‡] Kristin K. Isaacs,[†] Andrew McEachran,[§,∥] Christopher Grulke,[∥] Ann M. Richard,[∥] Antony J. Williams,[∥] Jon R. Sobus,[†] Russell S. Thomas,[∥] and John F. Wambaugh*[,∥]

Measurement of chemicals in consumer products

Slide from Kristin Isaacs

Ring et al., 2019

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

High-Throughput Stochastic Human Exposure and Dose Simulation Model (SHEDS-HT)

- We treat models like related assays and look for consensus while considering model appropriateness
- High-throughput model for simulating population exposures to chemical in consumer products via multiple product types, scenarios, and routes
- R package, code, and default input files for consumer products (derived from CPDat) are available:

Package 'ShedsHT'

September 9, 2016

Title To run the SHEDS-HT screening model for estimating human exposure to chemicals.

Version 0.1.1

Author Kristin Isaacs [aut, cre]

Maintainer Kristin Isaacs <isaacs.kristin@epa.gov>

Environmental Science & Technology

Article

pubs.acs.org/est

SHEDS-HT: An Integrated Probabilistic Exposure Model for Prioritizing Exposures to Chemicals with Near-Field and Dietary Sources

Kristin K. Isaacs,*,† W. Graham Glen,‡ Peter Egeghy,† Michael-Rock Goldsmith,§,○ Luther Smith,‡ Daniel Vallero,† Raina Brooks,∥ Christopher M. Grulke,⊥,○ and Halûk Özkaynak†

†U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, North Carolina 27709, United States

‡Alion Science and Technology, 1000 Park Forty Plaza Suite 200, Durham, North Carolina 27713, United States

§Chemical Computing Group, Suite 910, 1010 Sherbrooke Street West, Montreal, QC H3A 2R7, Canada

∥Student Services Contractor at U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, North Carolina 27709, United States

⊥Lockheed Martin, 109 T.W. Alexander Drive, Research Triangle Park, North Carolina 27709, United States

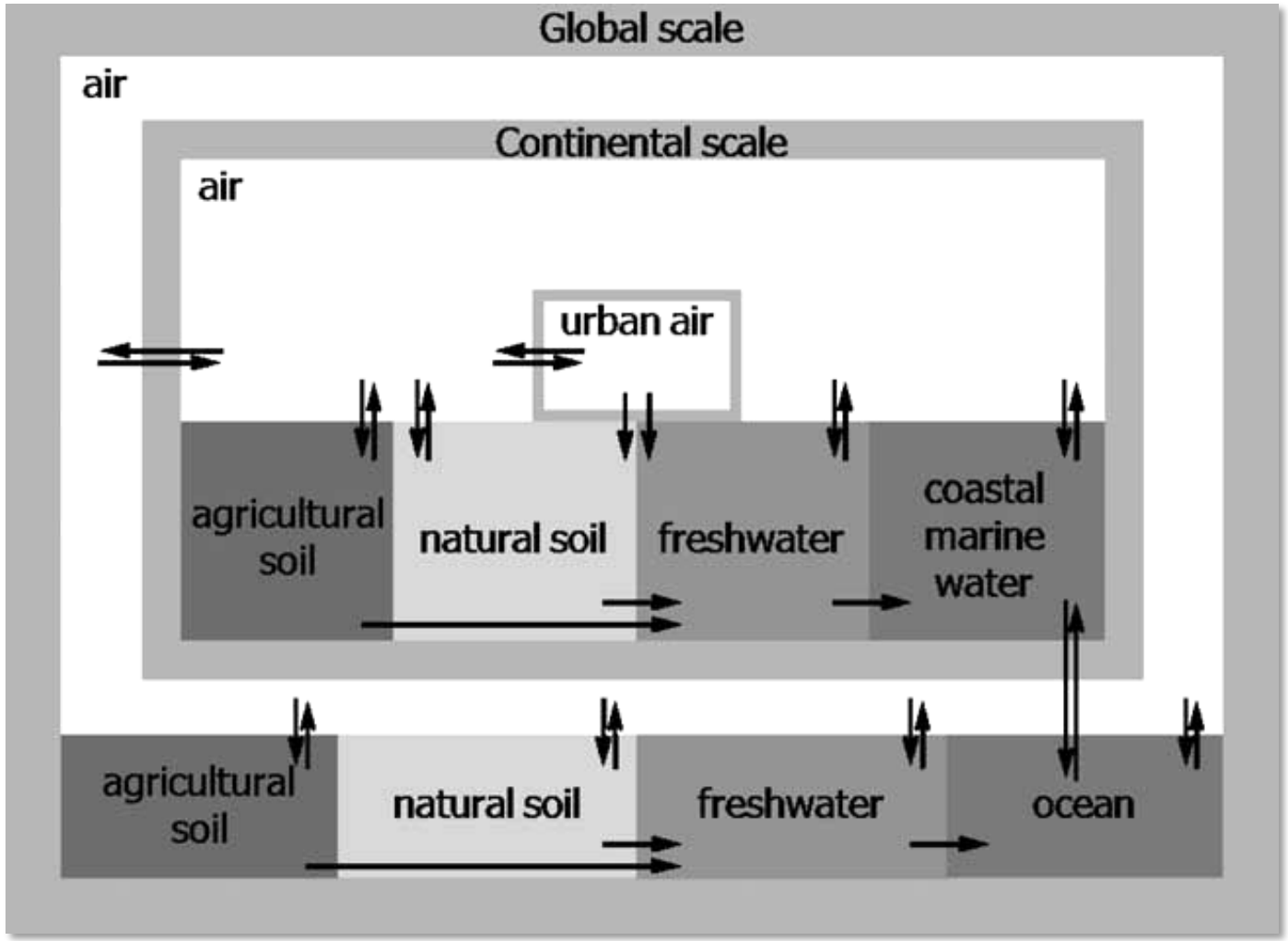
Supporting Information

https://github.com/HumanExposure/SHEDSHTRPackage

Isaacs *et al.* 2014

USEtox

- We treat models like related assays and look for consensus while considering model appropriateness
- United Nations Environment Program (UNEP) and Society for Environmental Toxicology and Chemistry (SETAC) toxicity model Version 2.0
- USEtox is a global scientific consensus fate, exposure and effect model
- USEtox consists of a set of nested environmental compartments at indoor, urban, continental, and global scale.

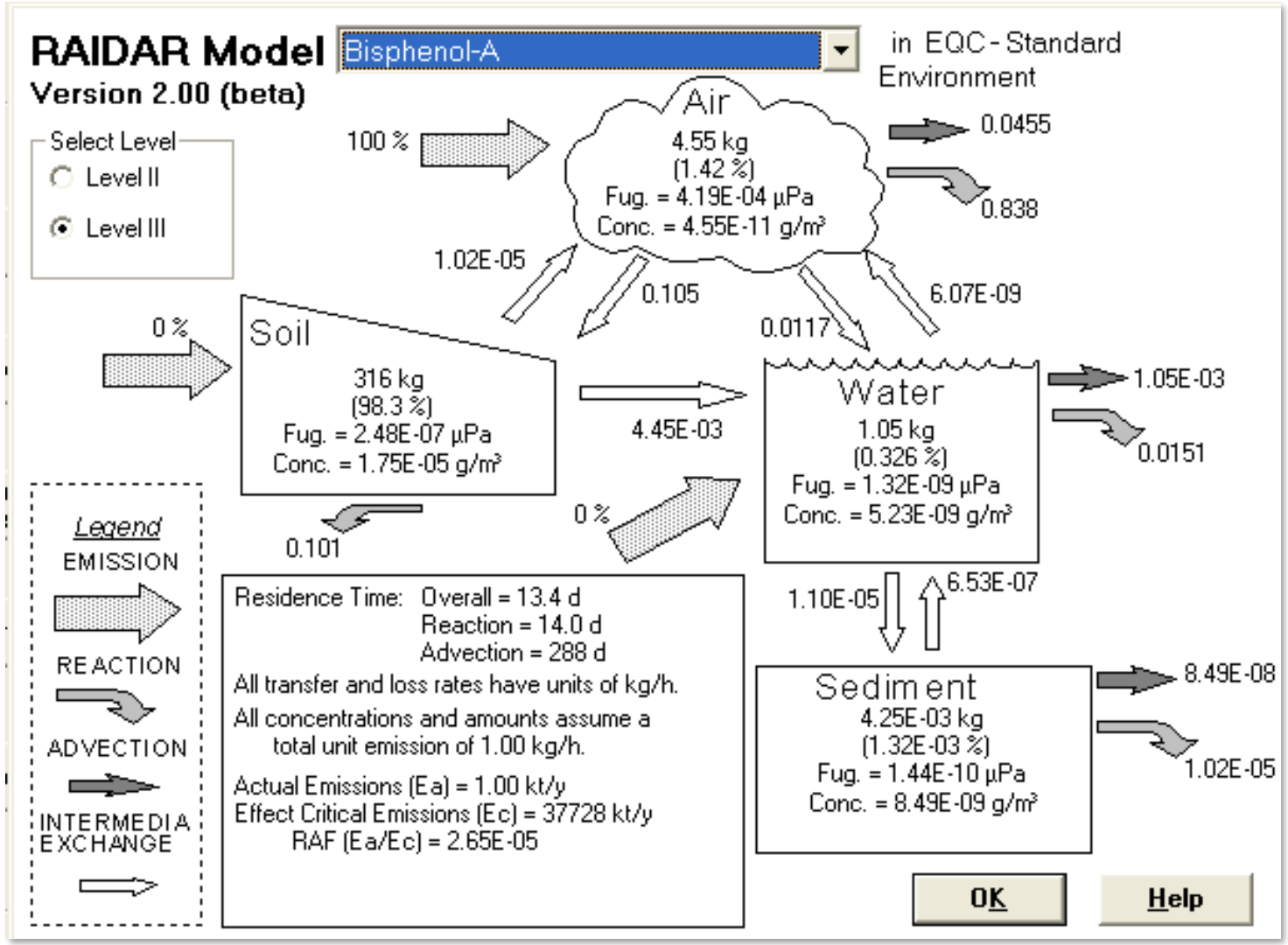


Rosenbaum *et al.* 2008

RAIDAR

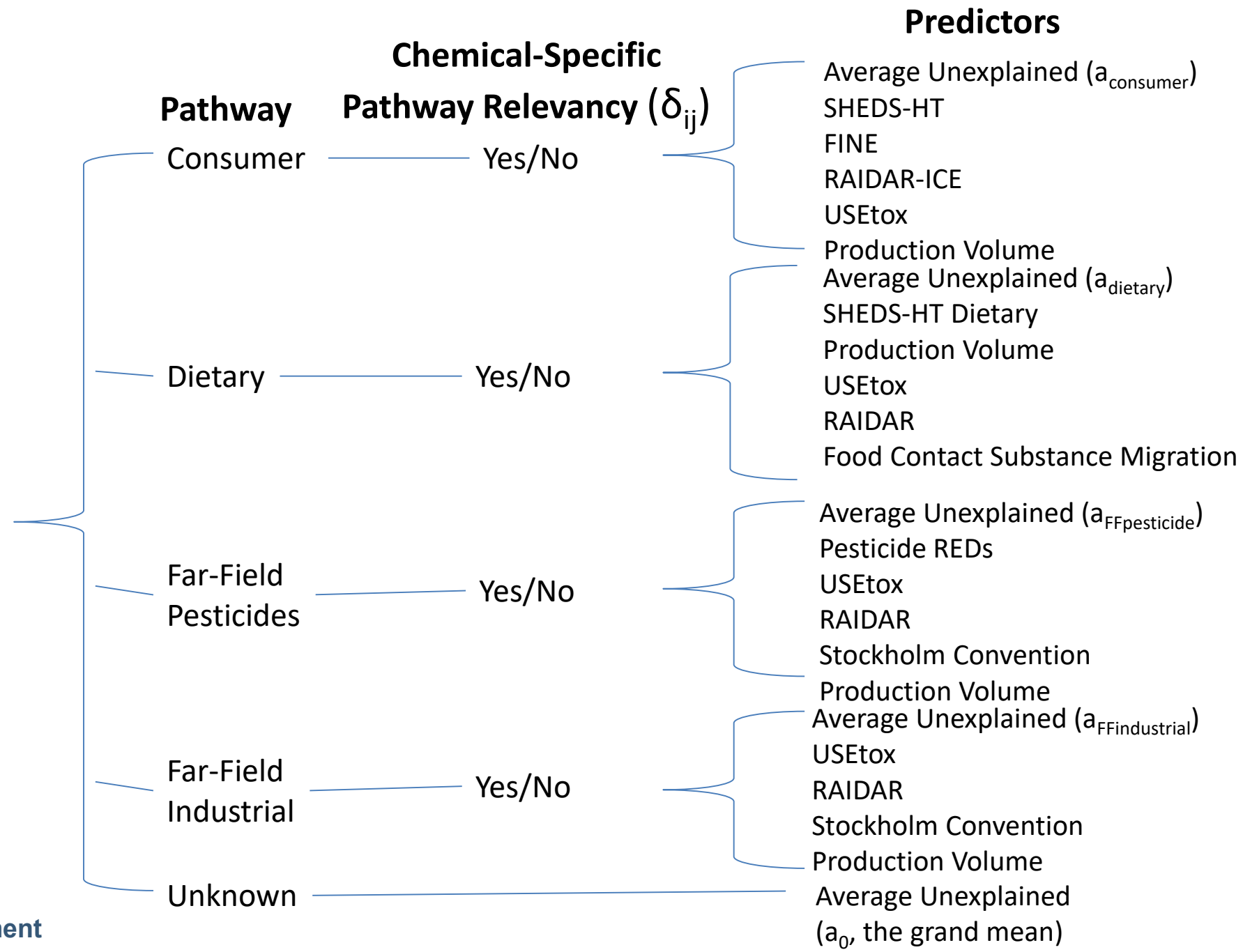
- We treat models like related assays and look for consensus while considering model appropriateness

- The Risk Assessment IDentification And Ranking (RAIDAR) model is an environmental fate and transport model linked with food web bioaccumulation models for representative ecological and agricultural targets and humans



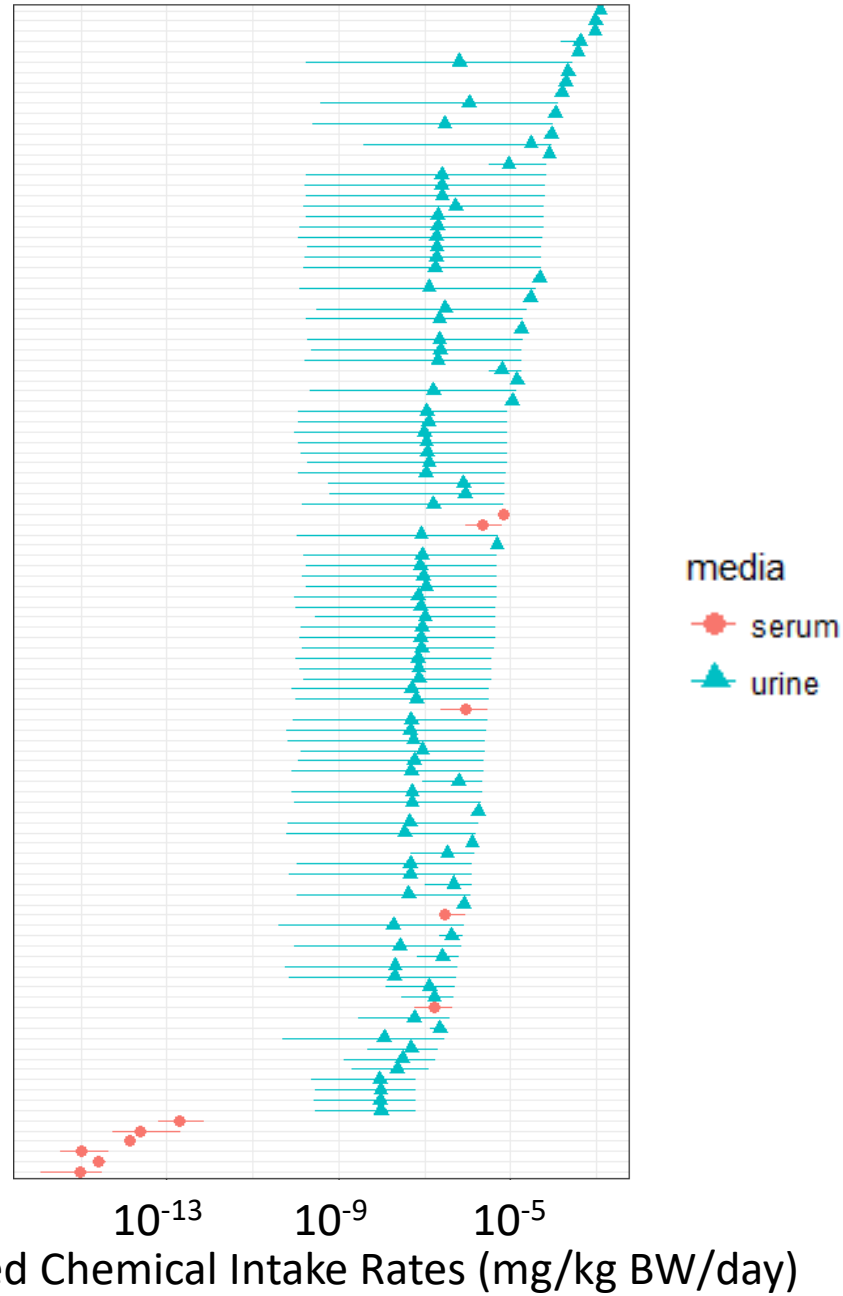
Arnot *et al.* 2006

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)

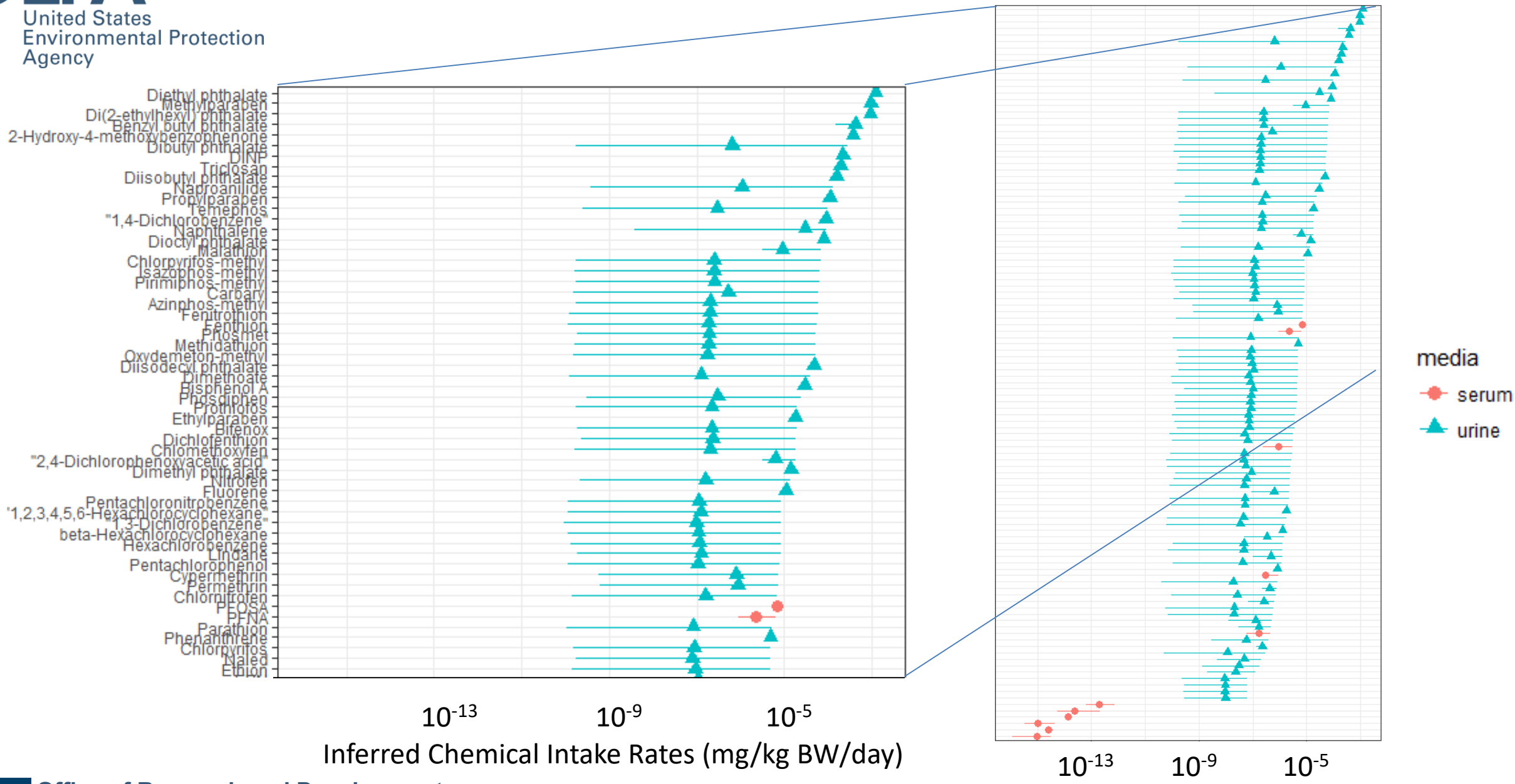
Reverse Dosimetry (Tan et al., 2006)

- 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



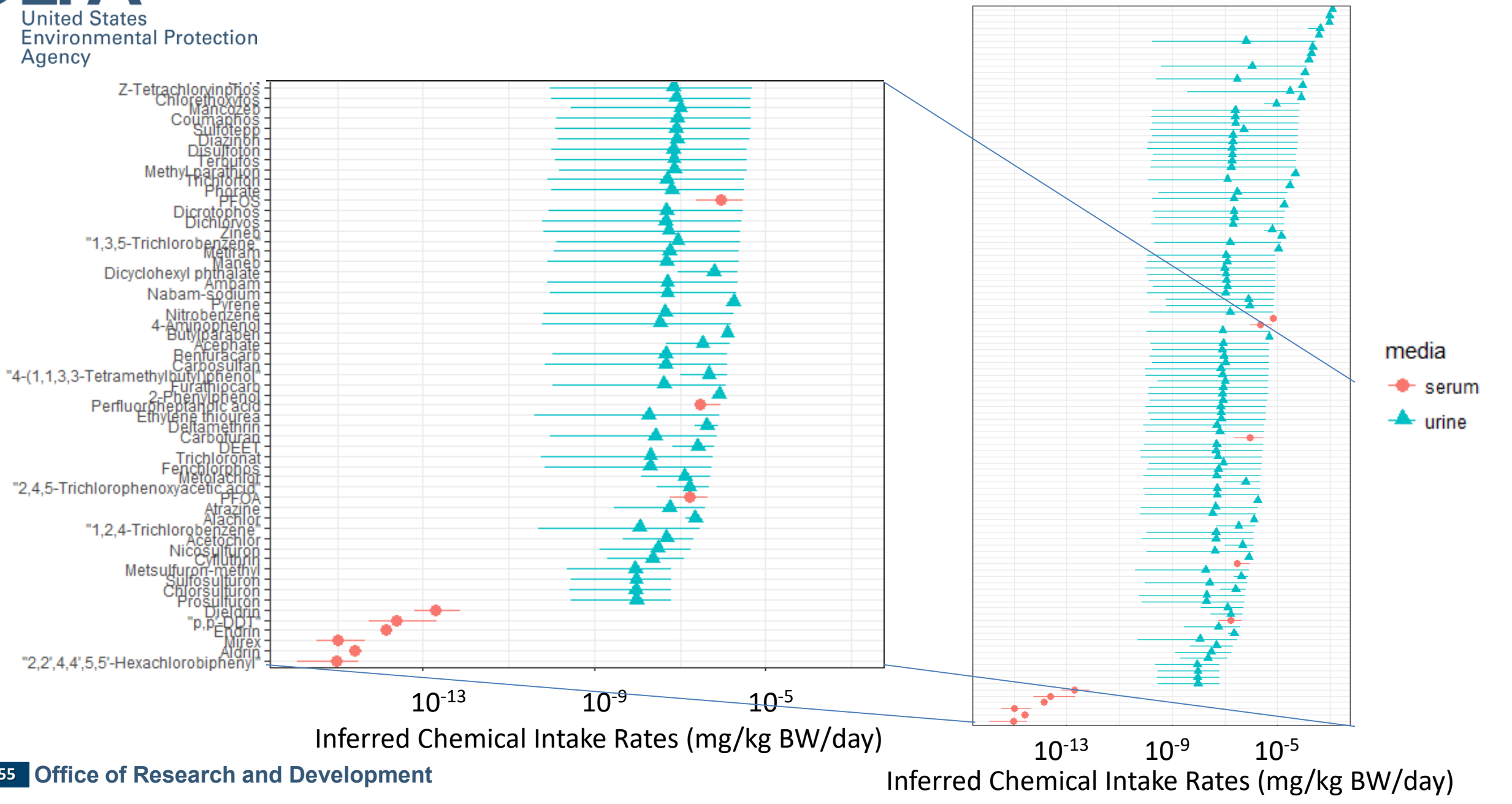
Reverse Dosimetry (Tan et al., 2006)

United States Environmental Protection Agency

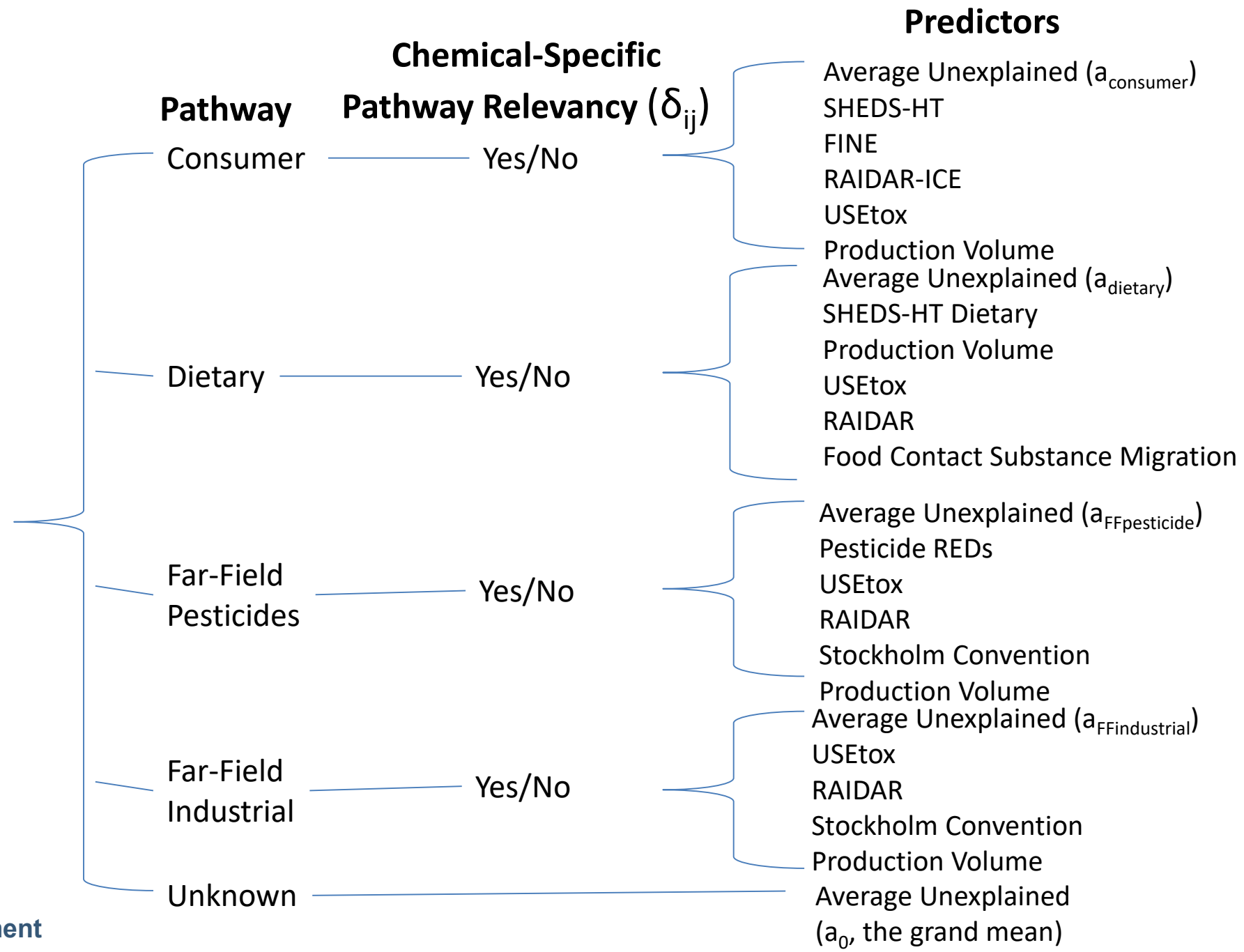


Reverse Dosimetry (Tan et al., 2006)

United States Environmental Protection Agency



SEEM3

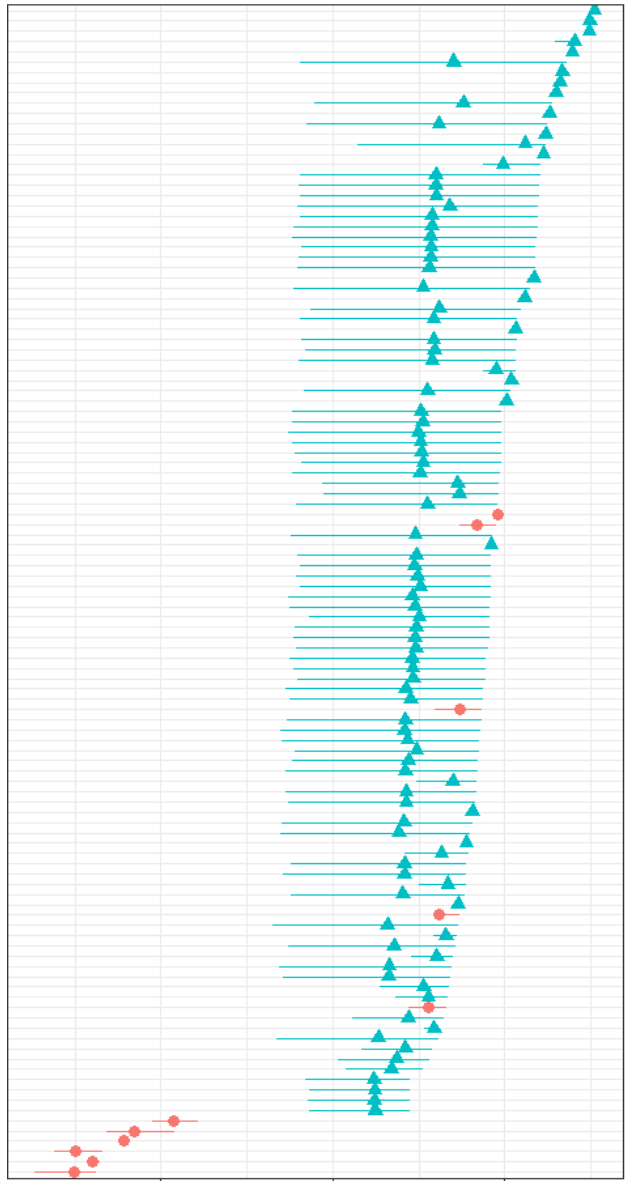

Total Chemical Intake Rate (mg/ kg BW/ day)
Pathway
Chemical-Specific Pathway Relevancy (δ_{ij})
Predictors
Consumer
Yes/No
Average Unexplained ($a_{consumer}$)
SHEDS-HT
FINE
RAIDAR-ICE
USEtox
Production Volume
Dietary
Yes/No
Average Unexplained ($a_{dietary}$)
SHEDS-HT Dietary
Production Volume
USEtox
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)

United States Environmental Protection Agency

Evaluation Data

Intake Rates Inferred from NHANES

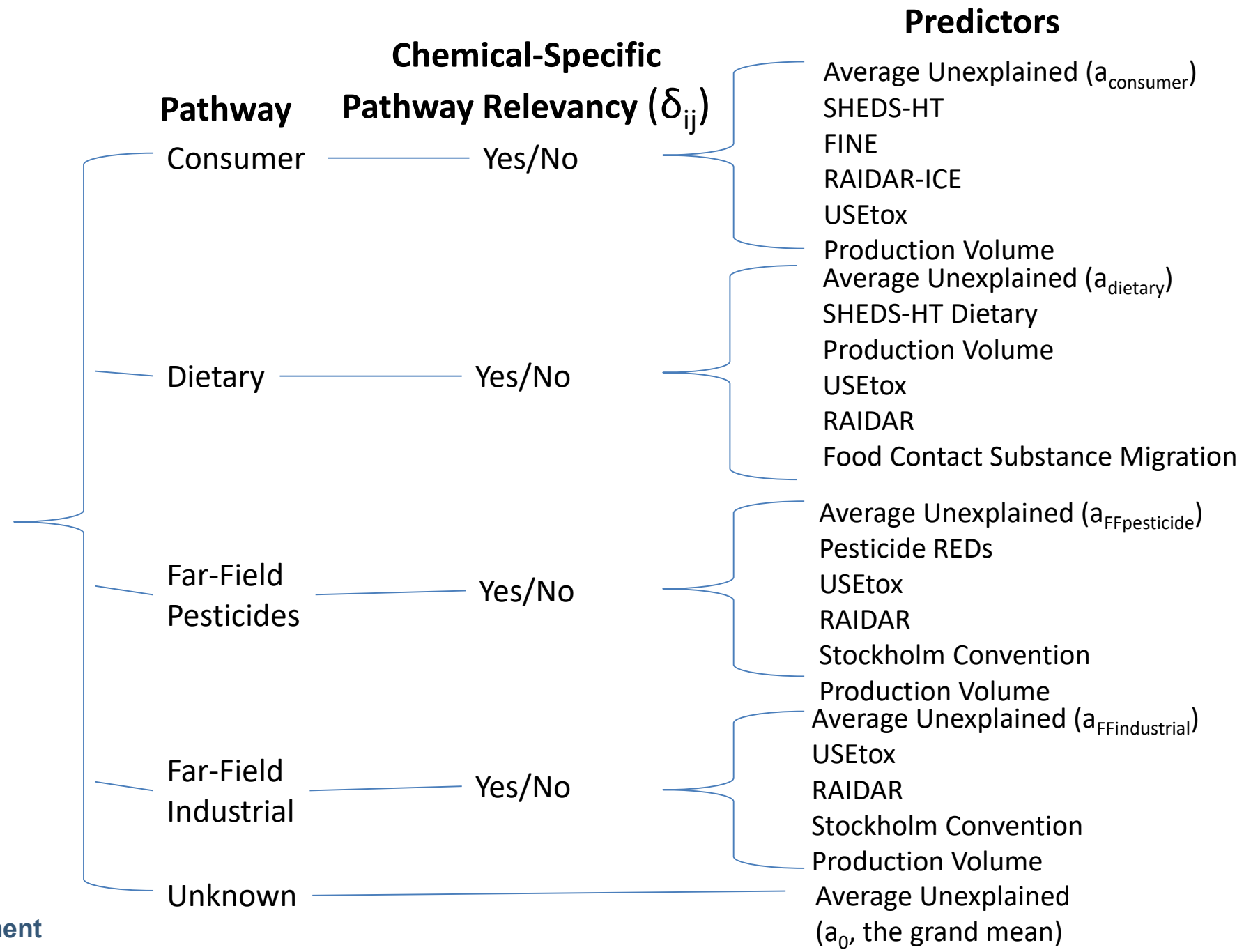
Total Chemical Intake Rate (mg/ kg BW/ day)



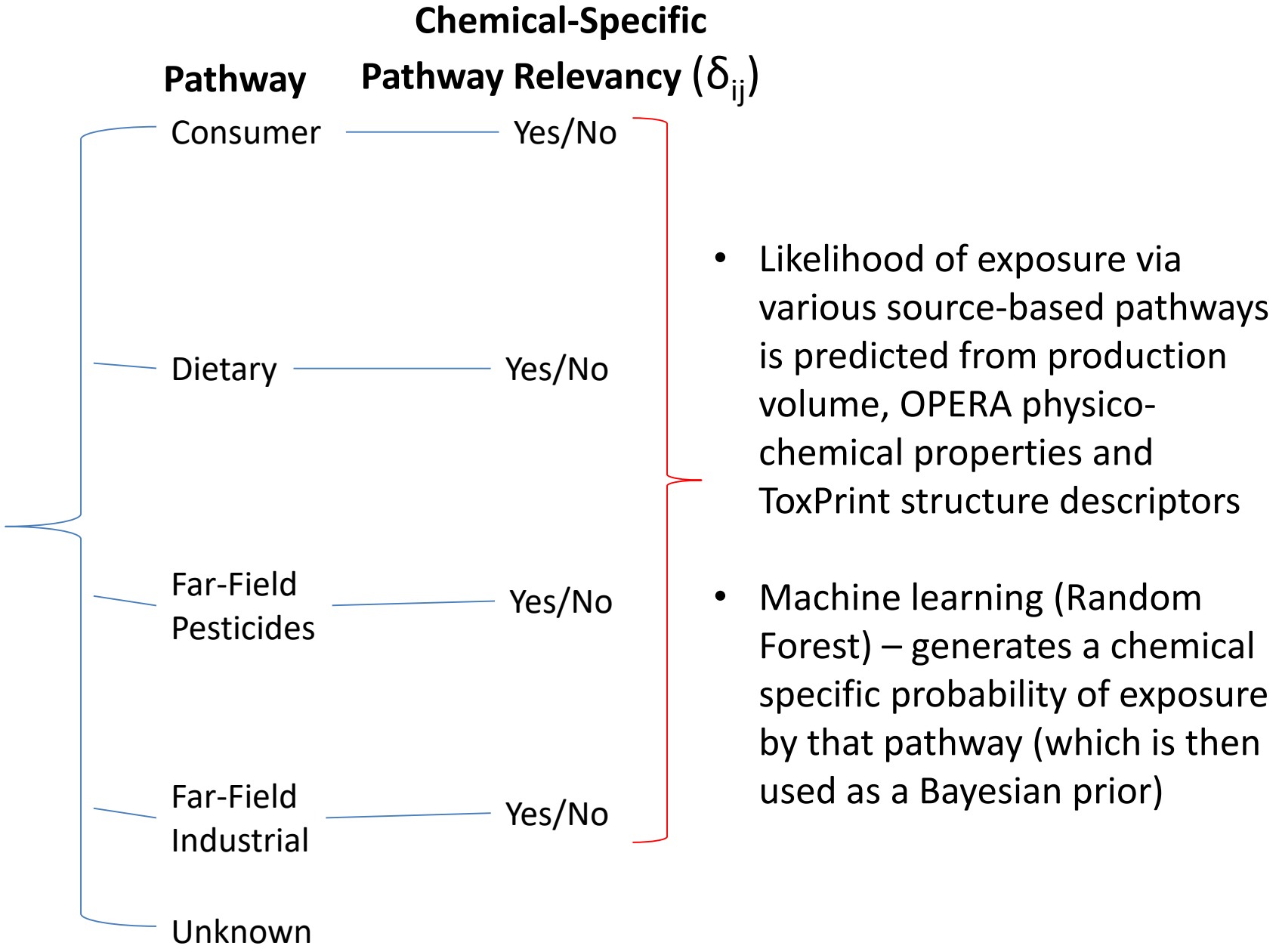
- NHANES urine (Wambaugh et al, 2014, Ring et al. 2017)
- NHANES serum/blood either using HTTK clearance (Pearce et al., 2017)
  - Literature clearance estimates were used for methodologically challenging chemicals not suited to HTTK

Ring et al., 2019

SEEM3

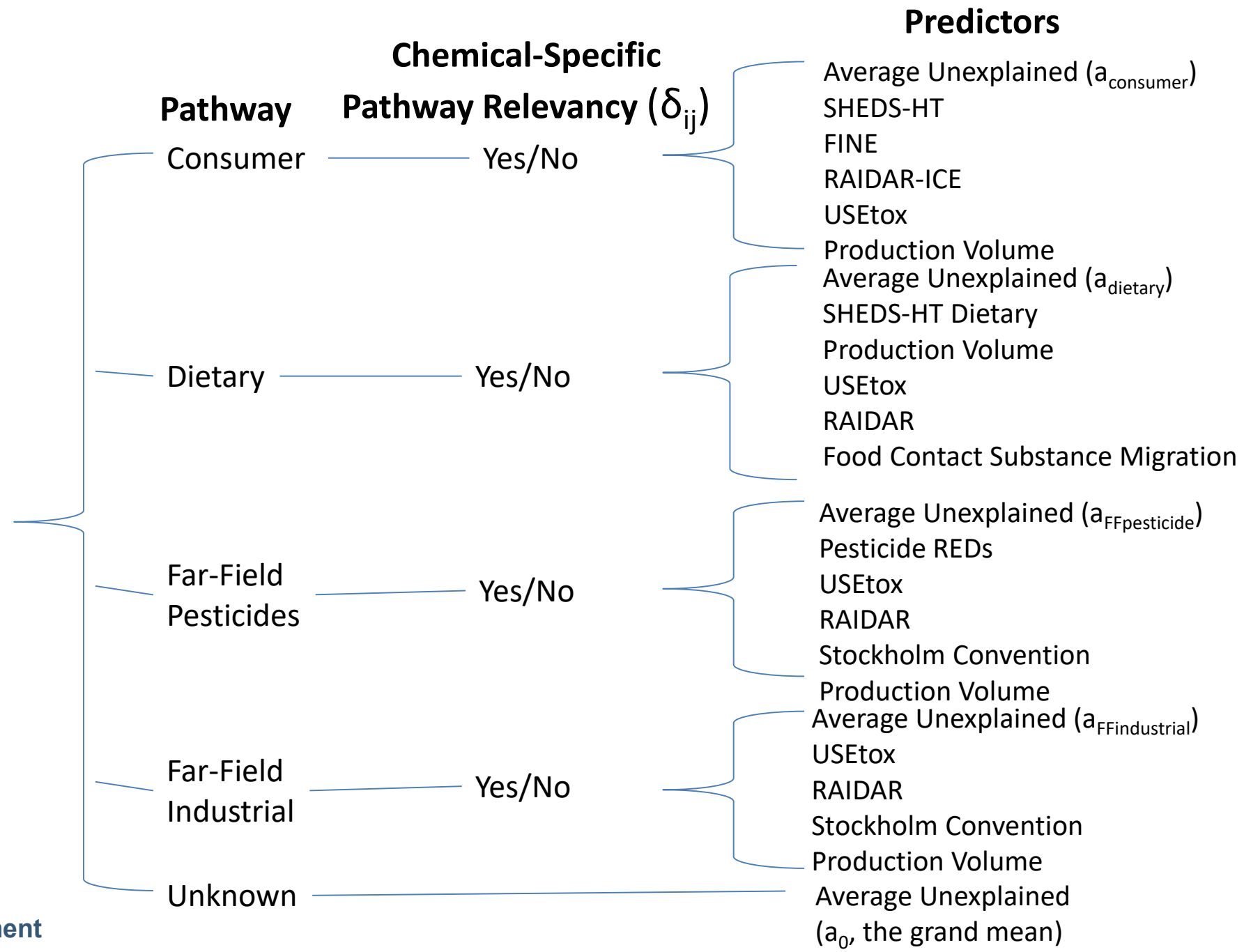

Total Chemical Intake Rate (mg/ kg BW/ day)
Pathway
Chemical-Specific Pathway Relevancy (δ_{ij})
Predictors
Consumer
Yes/No
Average Unexplained ($a_{consumer}$)
SHEDS-HT
FINE
RAIDAR-ICE
USEtox
Production Volume
Dietary
Yes/No
Average Unexplained ($a_{dietary}$)
SHEDS-HT Dietary
Production Volume
USEtox
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)

SEEM3



- Likelihood of exposure via various source-based pathways is predicted from production volume, OPERA physico-chemical properties and ToxPrint structure descriptors
- Machine learning (Random Forest) – generates a chemical specific probability of exposure by that pathway (which is then used as a Bayesian prior)

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)

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

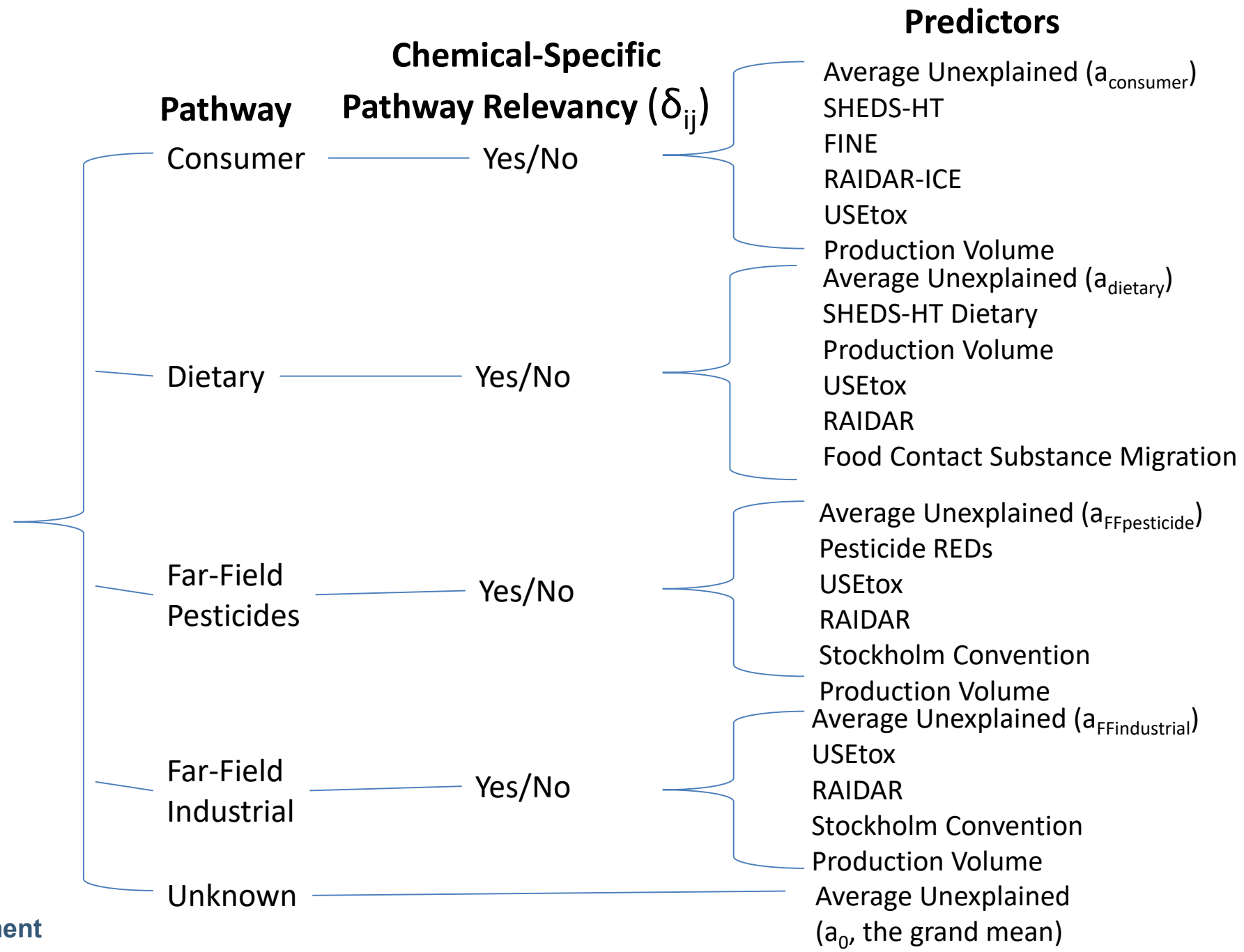
Bayesian analysis via Markov Chain Monte Carlo assigns each chemical either a "Yes" or "No" according to predicted probability

If the pathway is no for a chemical, nothing is added to the intake rate

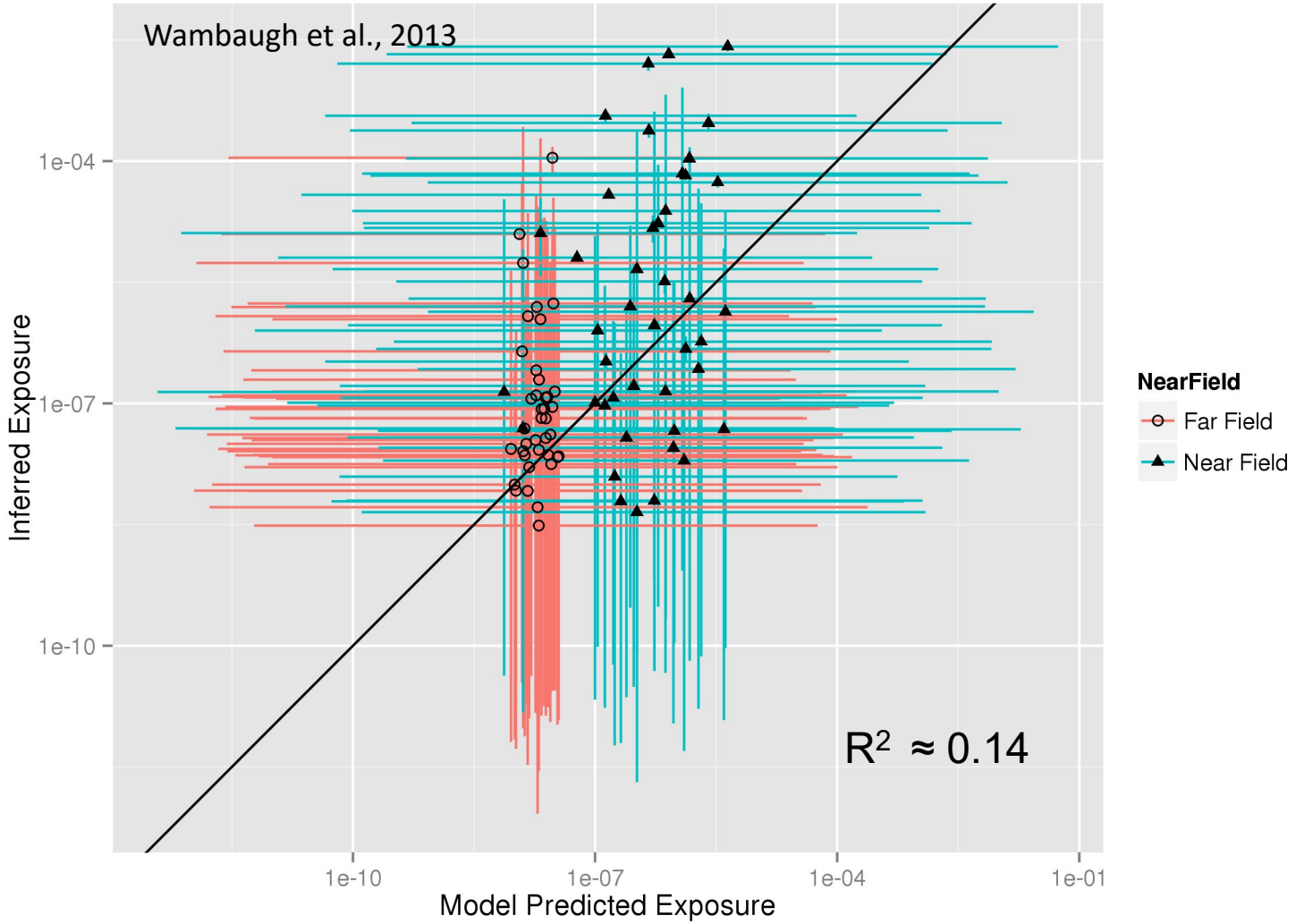
Linear regression is used to estimate the average unexplained exposure (intercept) and loadings (slopes, or predictive ability) for each model

Model predictions are centered at zero – **if there is no prediction for a chemical, the average value "zero" is added**

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)

First Generation SEEM



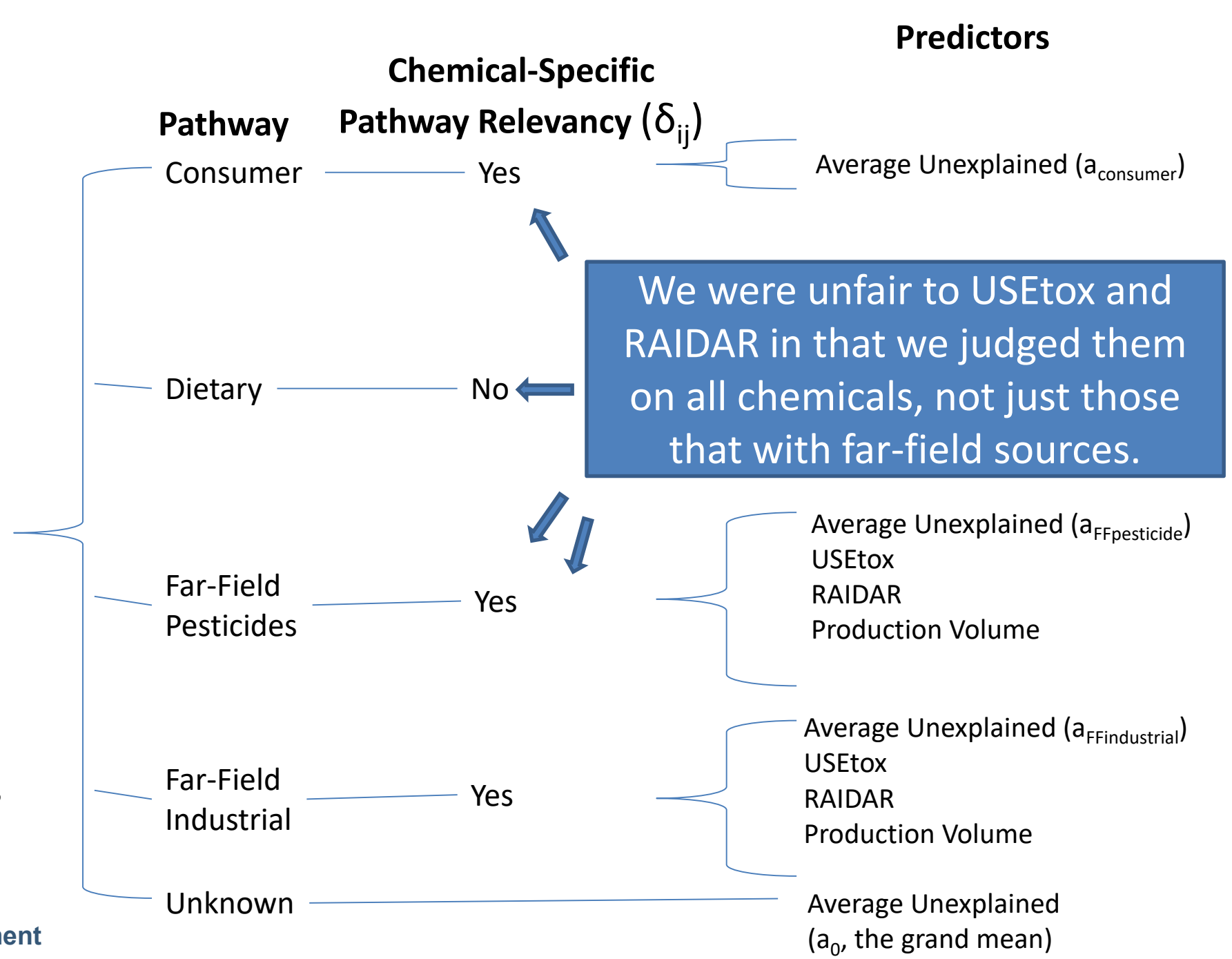
- Those chemicals with "near-field" – proximate, in the home, sources of exposure – had much higher rates of exposure than those with sources outside the home (Wallace et al., 1986)
- The only available "high throughput exposure models in 2013 were for far-field sources

United States Environmental Protection Agency

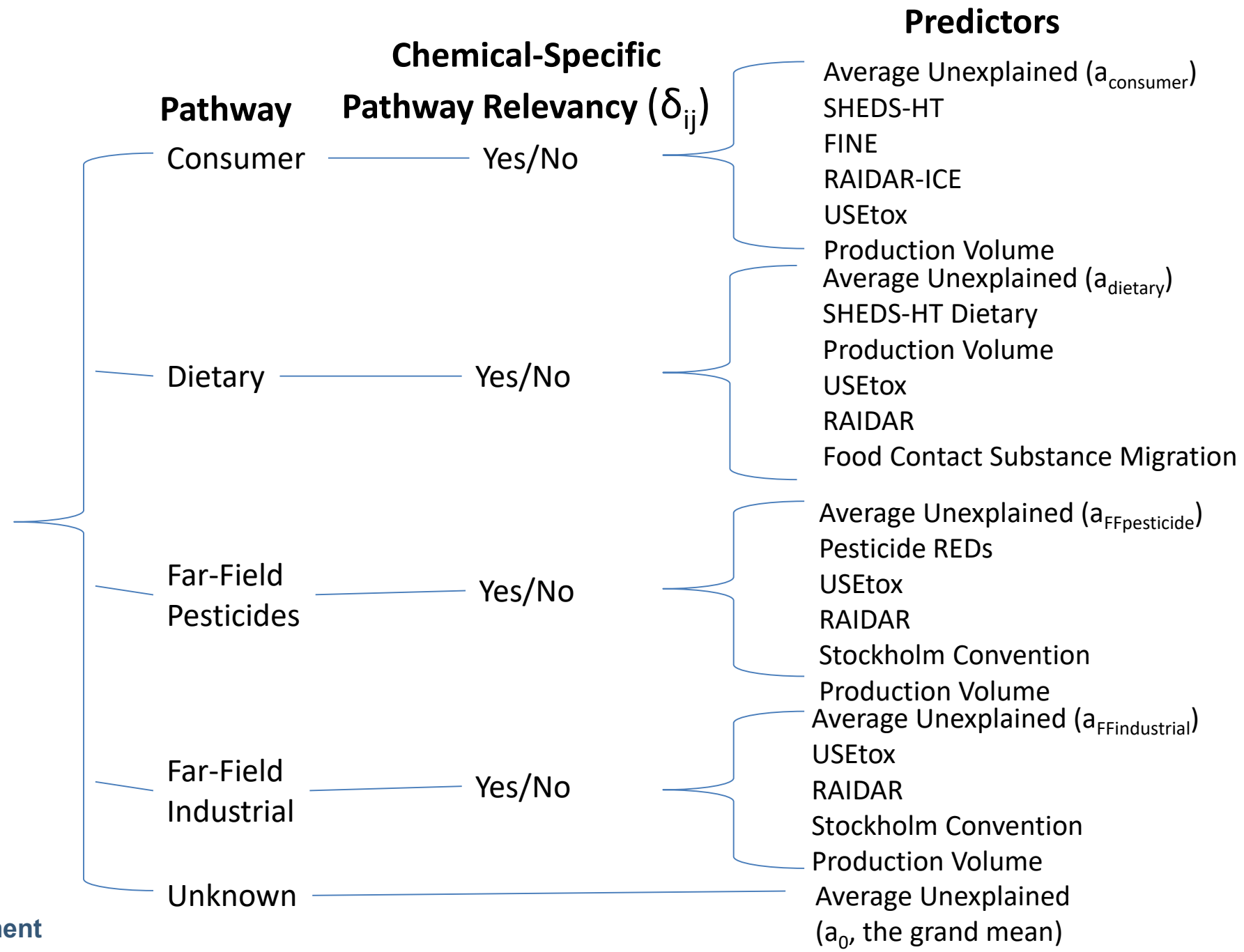
SEEM I

Total Chemical Intake Rate (mg/ kg BW/ day)

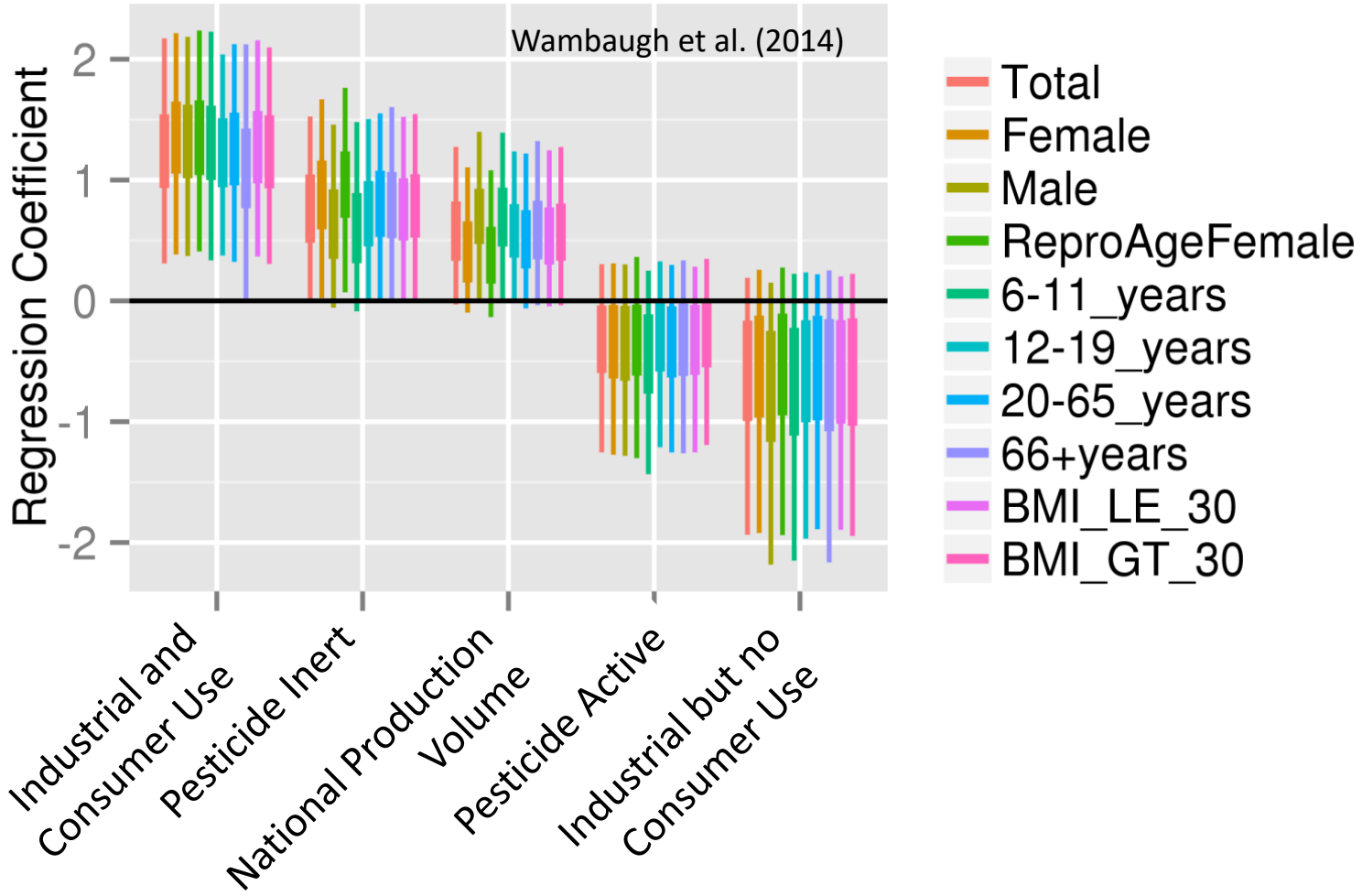
The 3rd Gen. SEEM framework incorporates the previous models



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)

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

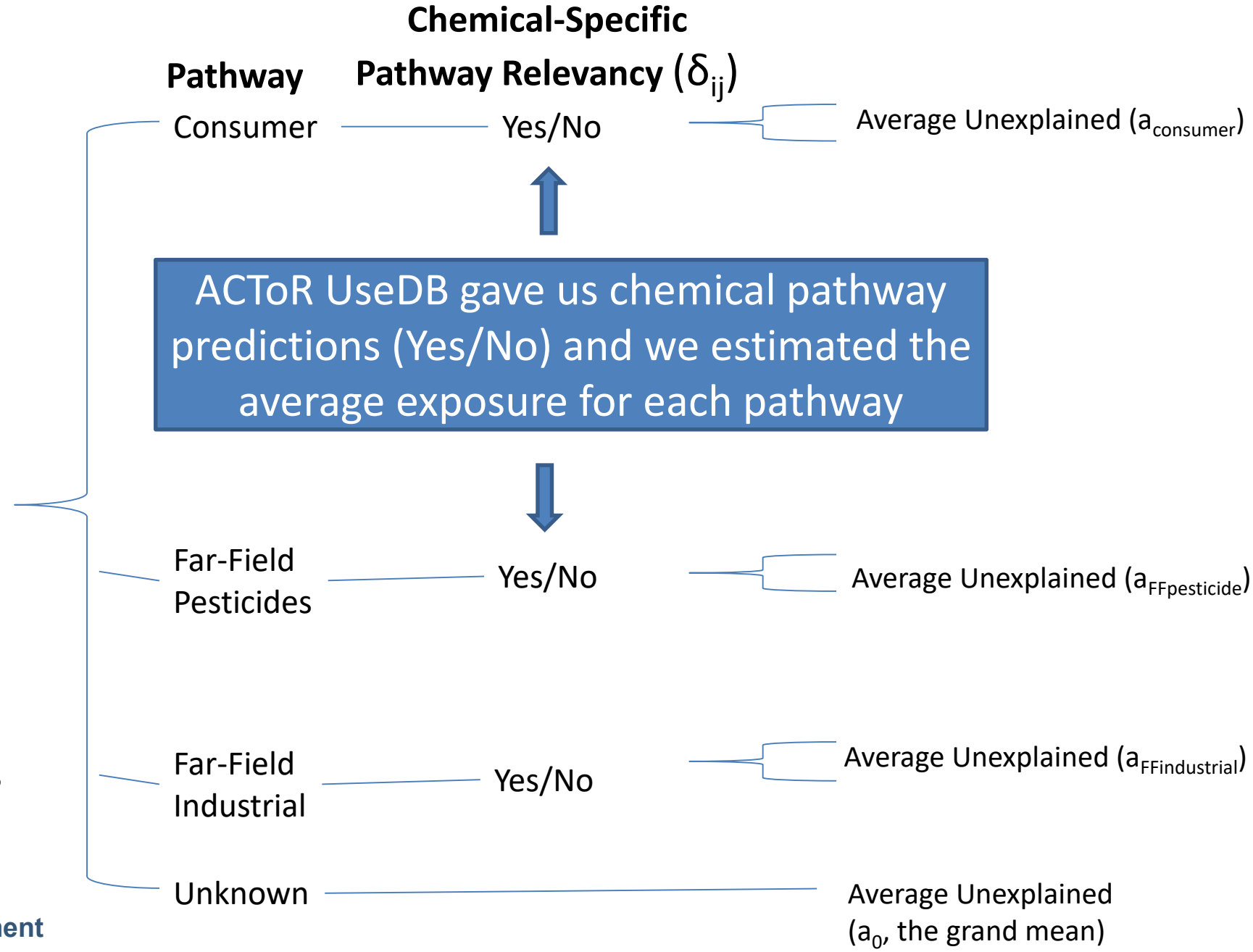
Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

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

SEEM2

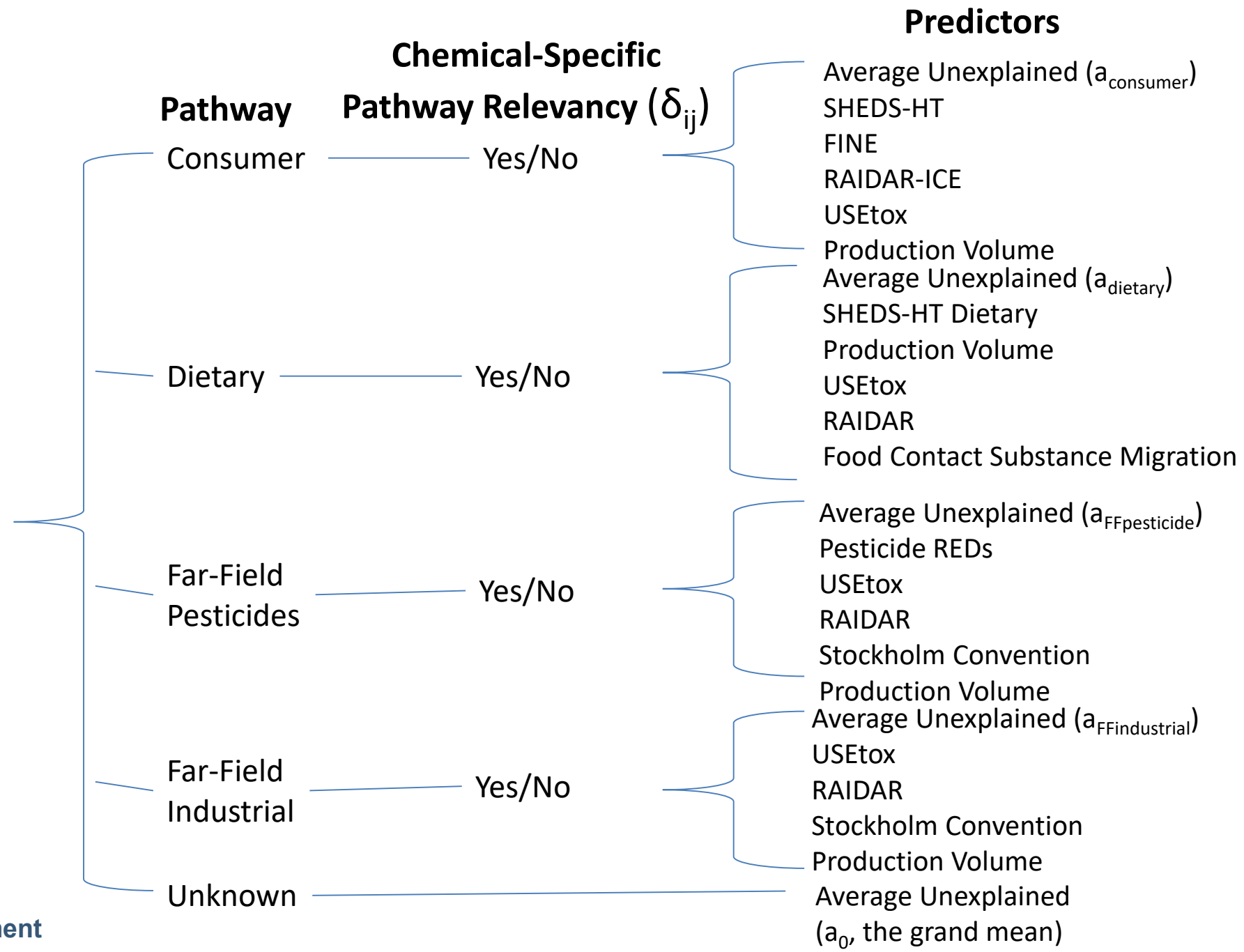
Predictors

Chemical-Specific Pathway Relevancy (δ_{ij})



The 3rd Gen. SEEM framework incorporates the previous models

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)

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

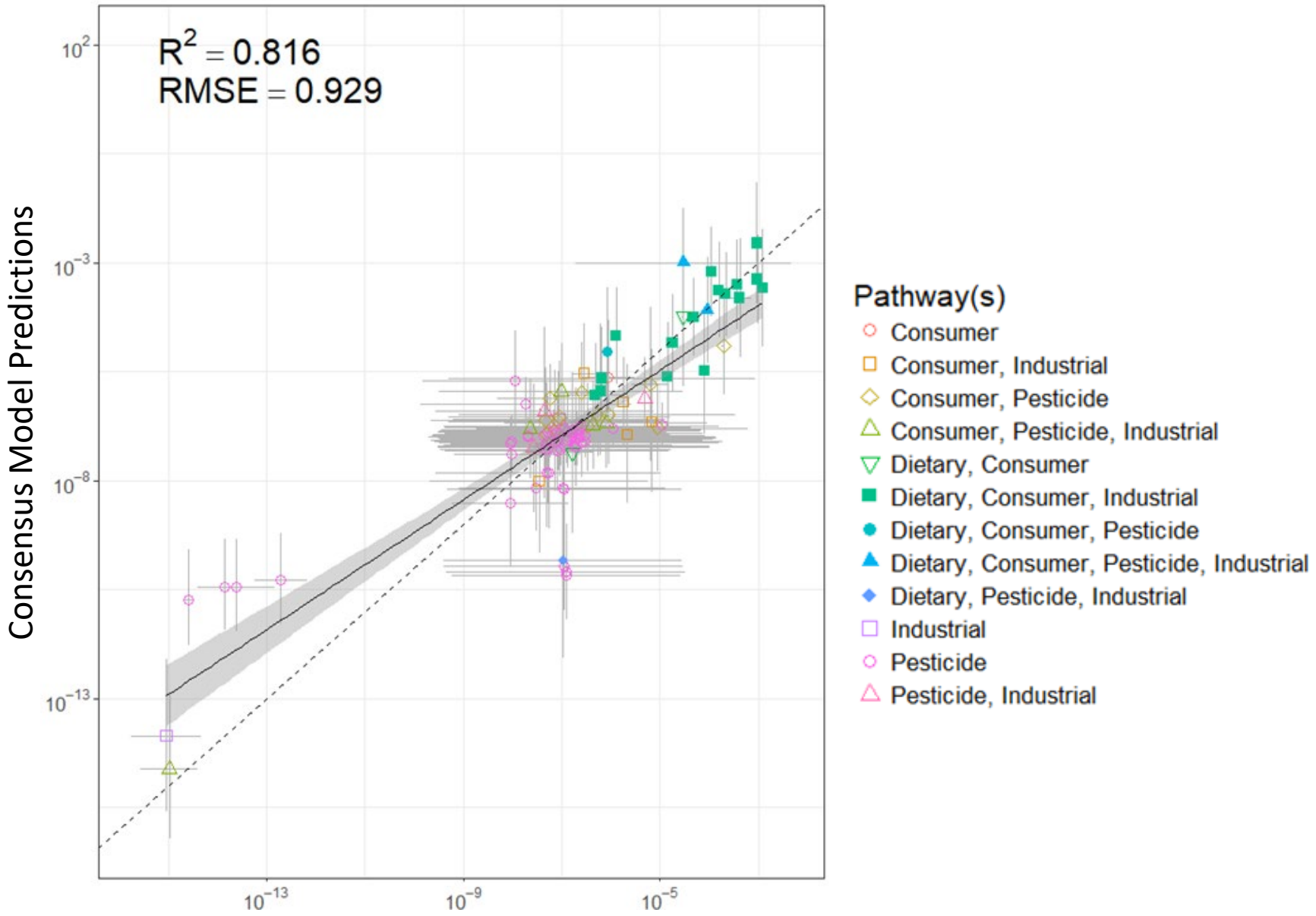
Most Important Predictors (Scaled Gini Impurity, Louppe et al., 2013)

	Normalized Gini impurity importance			
	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

Ring et al., 2019

Pathway-Based Consensus Modeling of NHANES

- 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

Estimated Model Parameters

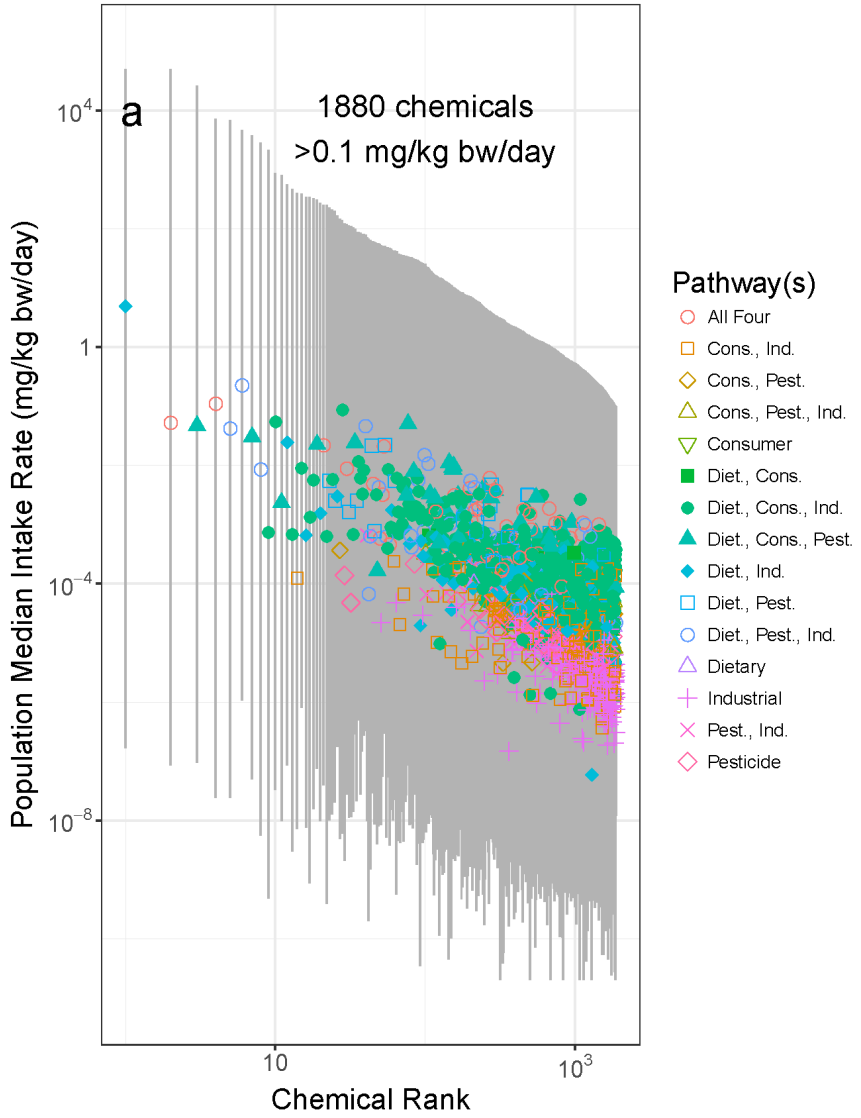
- Median parameter estimates from multivariate regression
- Standard deviation is reported in parentheses
- Statistically association indicated in bold

	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)**

Ring et al., 2019

Consensus Modeling of Median Chemical Intake

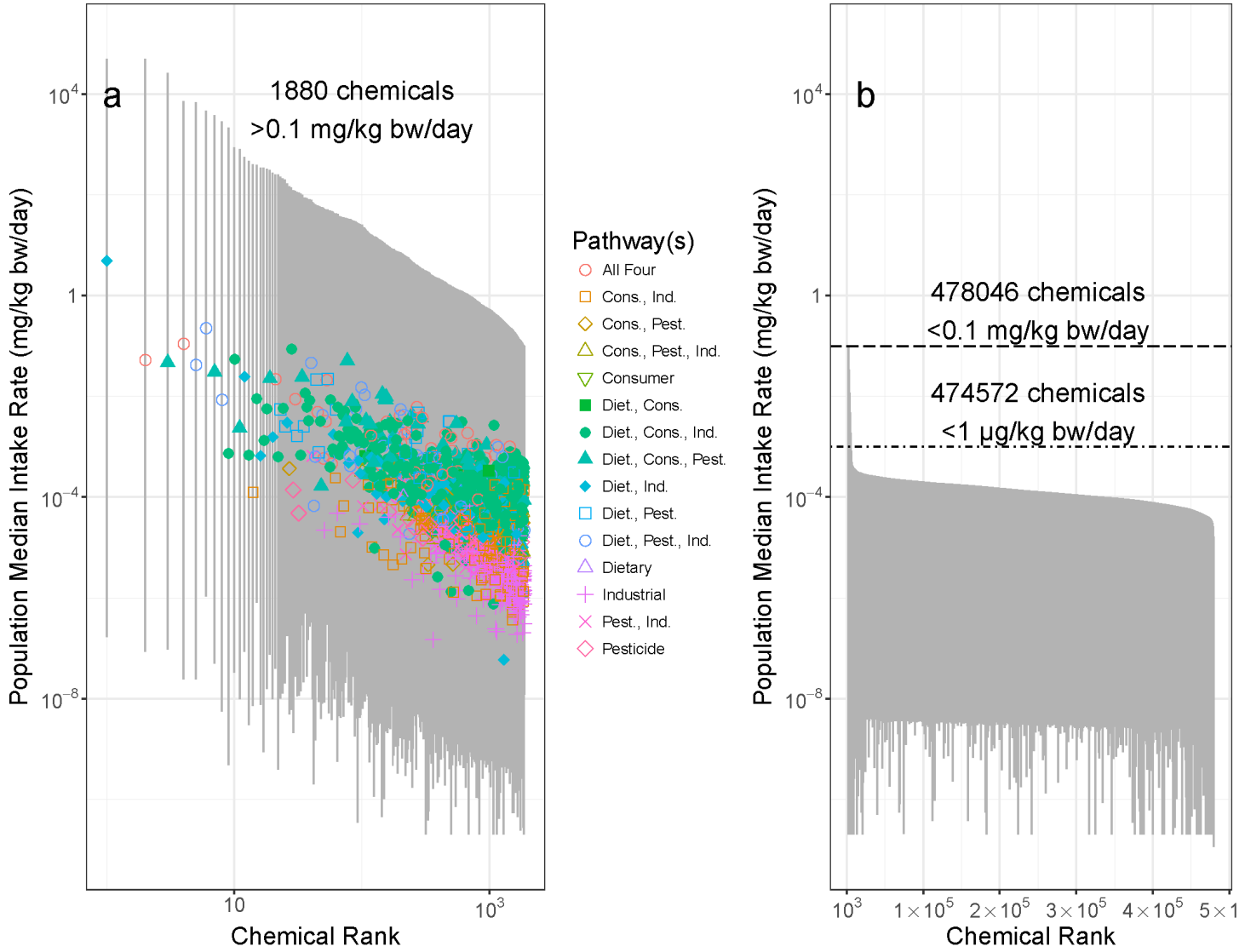
- We extrapolate to predict relevant pathway(s), median intake rate, and credible interval for each of 479,926 chemicals
- Of 687,359 chemicals evaluated, 30% have less than a 50% probability for exposure via any of the four pathways and are considered outside the “domain of applicability”
- This approach 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

Ring et al. (2019)

Environmental Science & Technology

Article

Cite This: Environ. Sci. Technol. 2019, 53, 719−732

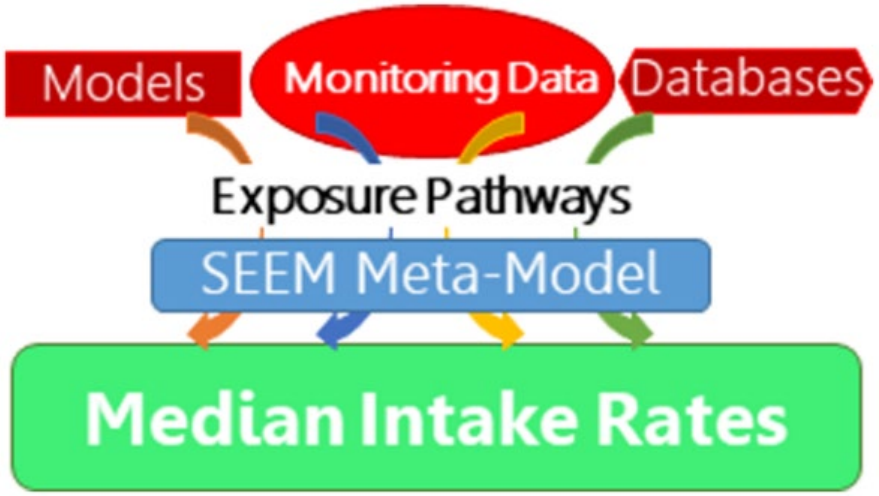
pubs.acs.org/est

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

Caroline L. Ring,[†,§,∞] Jon A. Arnot,[∥,⊥,#] Deborah H. Bennett,[▽] Peter P. Egeghy,[‡] Peter Fantke,[○] Lei Huang,[◆] Kristin K. Isaacs,[‡] Olivier Jolliet,[◆] Katherine A. Phillips,[‡] Paul S. Price,[‡] Hyeong-Moo Shin,[¶] John N. Westgate,[∥,○] R. Woodrow Setzer,[†] and John F. Wambaugh*[,†]

Table 1. Partial Technical Glossary

term	explanation
ExpoCast (Exposure Forecasting) Project	an ongoing U.S. Environmental Protection Agency project to develop new methods, data, and models for high-throughput exposure assessment (i.e., thousands of chemicals)[11,89]
exposure predictor	in this analysis "exposure predictor" refers to both the predictions of specific exposure models as well as other exposure-related information
exposure pathway	"the course an agent [chemical] takes from the source [environmental release] to the target [human]."[14] In this analysis we use the simple term "pathway" to represent the totality of paths that a chemical may follow from a particular source to reach a person
grand mean	the overall mean of a regression. In this analysis, the grand mean a_0 describes the average intake rate inferred from NHANES in contrast to the pathway-specific means[90]
intake	"the process by which an agent [chemical] crosses an outer exposure surface [some portion of an individual] of a target [human] without passing an absorption barrier, i.e. through ingestion or inhalation"[14]
intake rate	daily average intake (mg/kg body weight/day)
meet-in-the-middle	an approach in which predictions from models that make predictions from upstream data (e.g., activity) are compared with models that make inferences from downstream data (e.g., biomarkers), 18
near-field/far-field sources	"near-field" sources are proximate, indoor sources such as consumer product use in domestic settings, while "far-field" sources are distal with exposure mediated by environmental fate and transport[56,71,87]
random forest algorithm	a machine learning approach in which an ensemble of decision trees is used to make probabilistic predictions[29]
Systematic Empirical Evaluation of Models (SEEM)	SEEM is a consensus modeling method for exposure model evaluation and calibration. SEEM uses a meet-in-the-middle approach to calibrate high-throughput exposure predictors with intake rates inferred from biomonitoring data[20,21]



Haven't Had Enough?

SOT 2019 Sunrise Mini-Course SR02

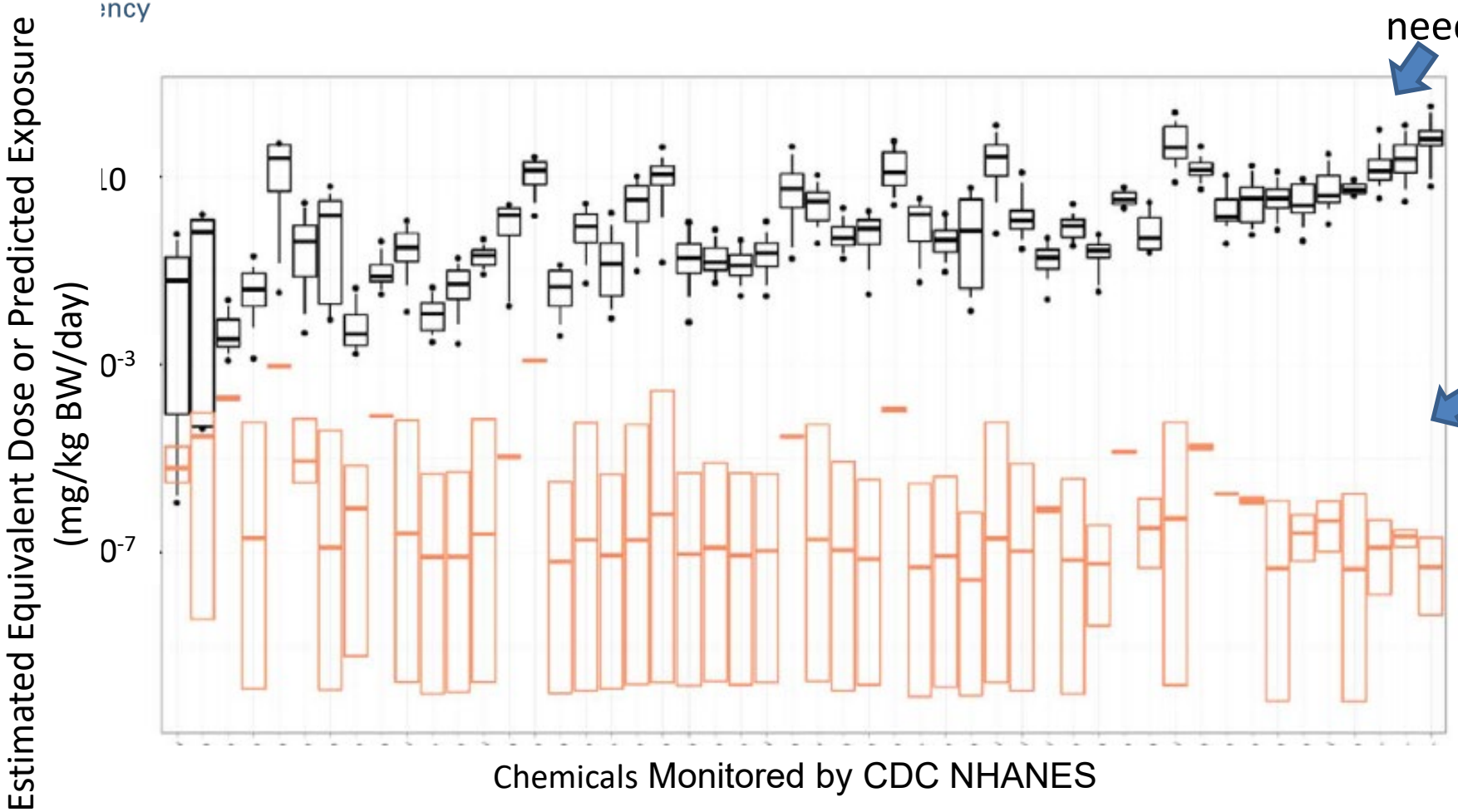
"Publicly Available Exposure Tools to Inform the Toxic Substances Control Act"

7 AM Sunday morning

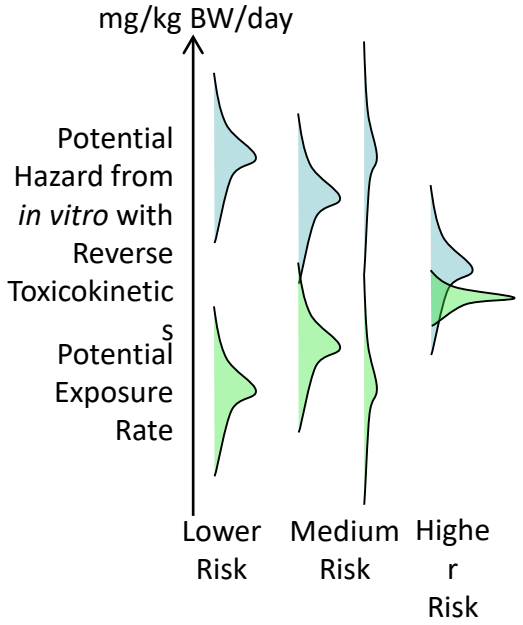
you can register in person at the SOT meeting

ToxCast + ExpoCast Provide NAMs for Chemical Prioritization

ToxCast + HTTK can estimate doses needed to cause bioactivity



Exposure intake rates can be inferred from biomarkers (Wambaugh et al., 2014)

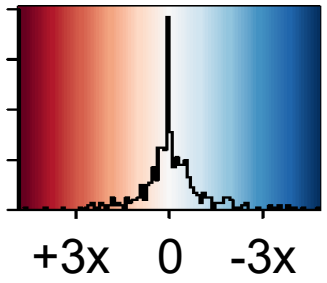


Ring *et al.* (2017)

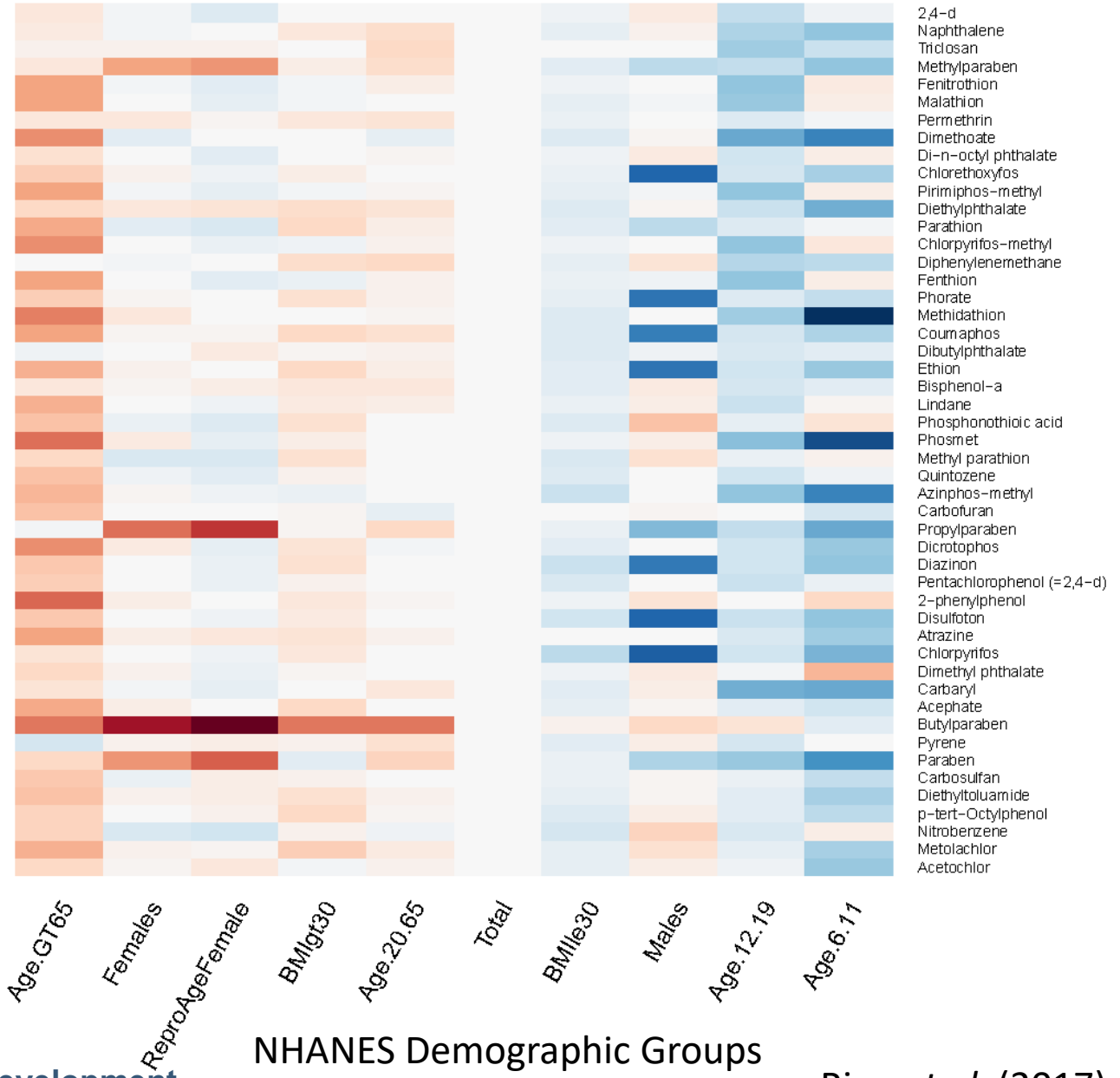
SEEM2: Life-stage and Demographic Specific Predictions

United States Environmental Protection Agency

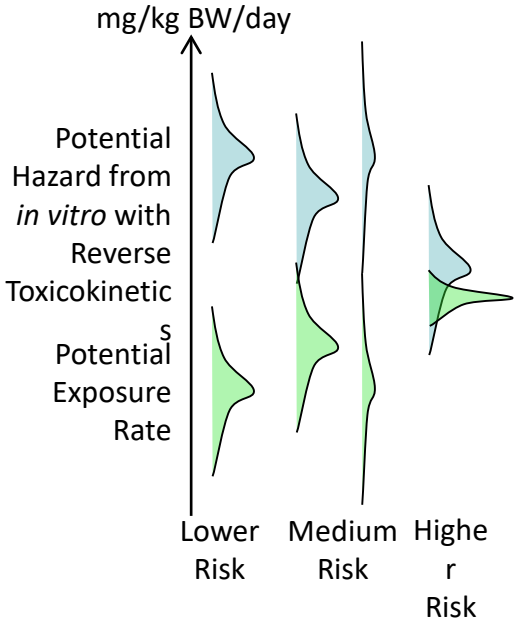
Change in Activity : Exposure Ratio



- To date SEEM3 predictions are only available for overall ("Total") U.S. Population



- We can calculate margin between bioactivity and exposure for specific populations
- Based on variation in toxicokinetics and exposure



Ring *et al.* (2017)

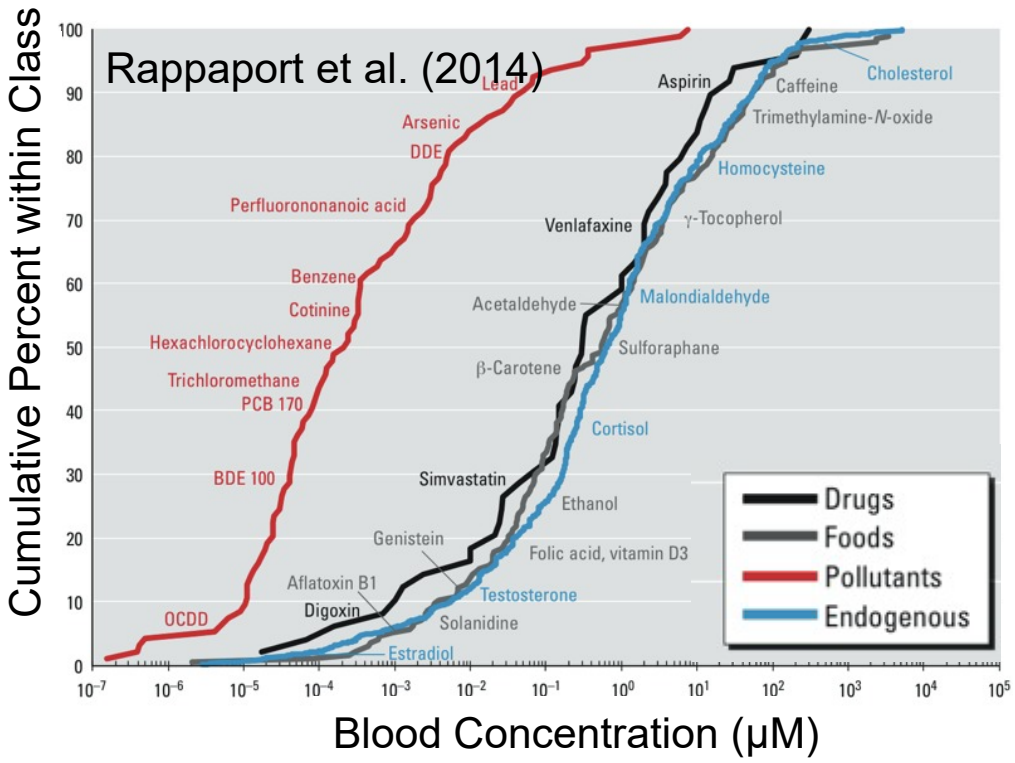
Conclusions

- 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



United States Environmental Protection Agency

Final Thought

"Scientists should resist the demand to describe any model, no matter how good, as validated. Rather than talking about strategies for validation, we should be talking about means of evaluation."
Naomi Oreskes

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

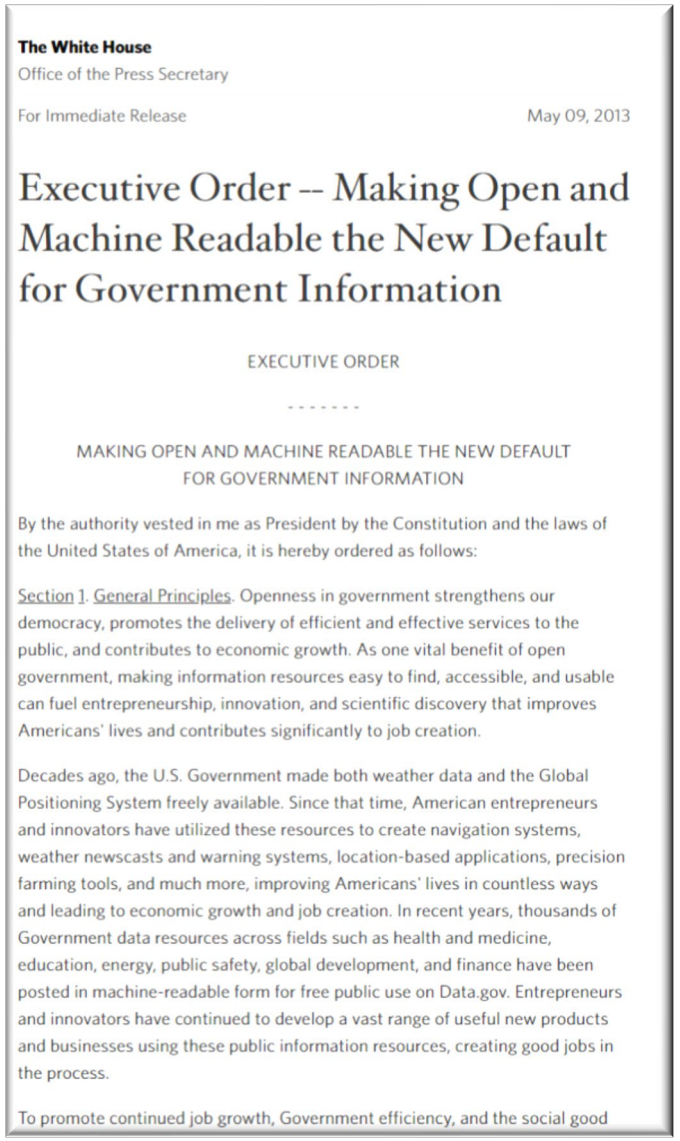
- Arnot, J. A.; et al., Develop Sub-Module for Direct Human Exposures to Consumer Products. Technical Report for the U.S. Environmental Protection Agency; ARC Arnot Research & Consulting, Inc.: Toronto, ON, Canada, 2014.
- Bennett, D. H.; Furtaw, E. J., Fugacity-based indoor residential pesticide fate 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.

Open and Machine Readable Modeling

- "Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public." Burwell et al. (2013):

- "…this includes using machine-readable and open formats…" Burwell et al. (2013):

- Machine learning models based on chemical structure and physico-chemical properties predict whether or not each pathway is relevant to a library of over 680,000 chemicals,
  - Each individual model prediction will also be made available



The White House
Office of the Press Secretary

For Immediate Release May 09, 2013

Executive Order -- Making Open and Machine Readable the New Default for Government Information

EXECUTIVE ORDER

- - - - - - -

MAKING OPEN AND MACHINE READABLE THE NEW DEFAULT FOR GOVERNMENT INFORMATION

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. General Principles. Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation.

Decades ago, the U.S. Government made both weather data and the Global Positioning System freely available. Since that time, American entrepreneurs and innovators have utilized these resources to create navigation systems, weather newscasts and warning systems, location-based applications, precision farming tools, and much more, improving Americans' lives in countless ways and leading to economic growth and job creation. In recent years, thousands of Government data resources across fields such as health and medicine, education, energy, public safety, global development, and finance have been posted in machine-readable form for free public use on Data.gov. Entrepreneurs and innovators have continued to develop a vast range of useful new products and businesses using these public information resources, creating good jobs in the process.

To promote continued job growth, Government efficiency, and the social good

May 9, 2013