United States Environmental Protection Agency

ExpoCast
exposure forecasting

High Throughput Exposure Models
and the Systematic Empirical Evaluation of Models (SEEM) Framework

John Wambaugh

US EPA CSS-HERA Board of Scientific Counselors
Chemical Safety Subcommittee Meeting

February 2-5, 2021

United States Environmental Protection Agency

Stakeholder Need as stated in research plan

Chemical exposure scenarios and pathways:

Chemical evaluations require information to estimate exposure via a variety of high-priority pathways, including scenario-specific data and models particular to consumer products and materials in the indoor environment, as well as occupational, ambient and ecological pathways.

Exposure Pathways

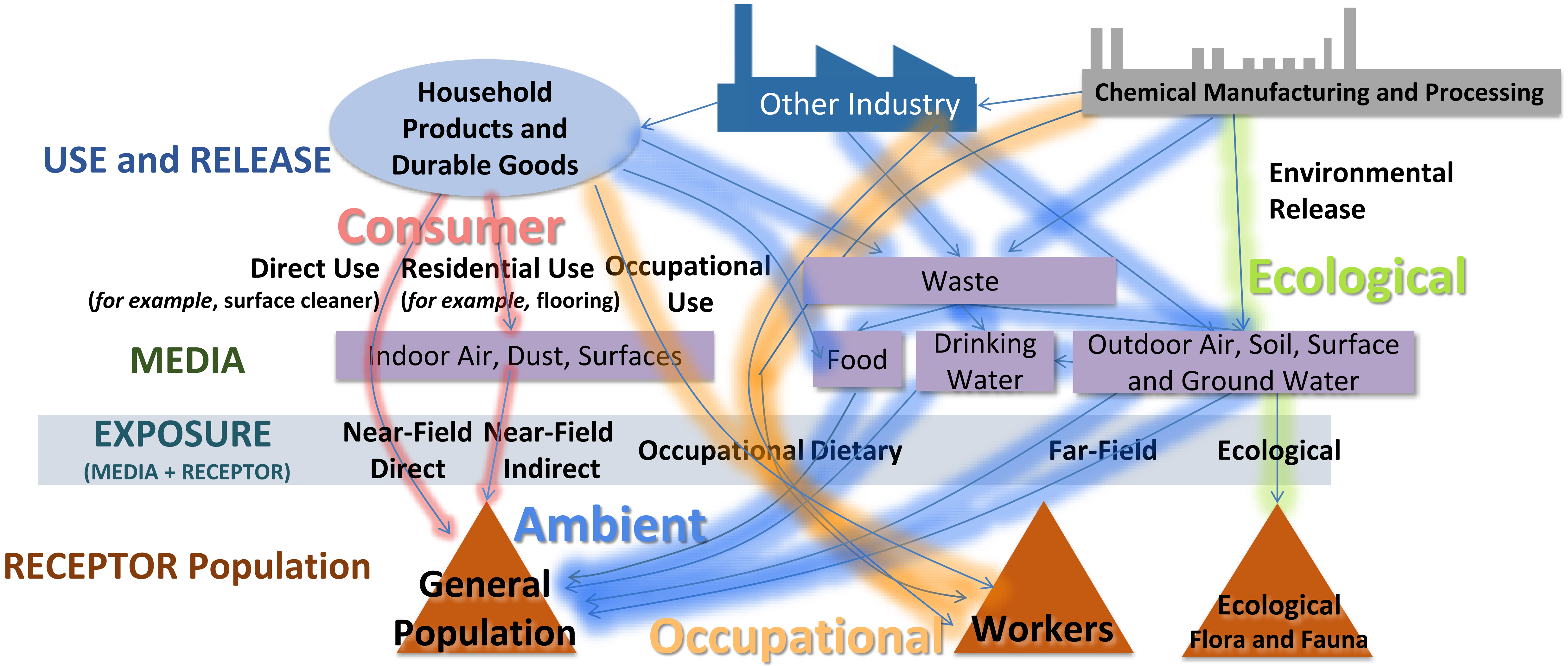


Figure from Kristin Isaacs

Properties of High-Throughput Exposure Models

1) Capable of handling **many chemicals with minimal descriptive information**
2) Cover **one or more relevant exposure routes**
3) **Allow for integration** with models for other pathways
4) Scientifically plausible
5) Allow for the assessment of interindividual and intraindividual variation in exposure
6) **Amenable to integration within statistical frameworks that quantify uncertainty**
7) **No more complicated than necessary**

ELSEVIER

Current Opinion in Toxicology

Available online 31 July 2019

In Press, Journal Pre-proof

New Approach Methodologies for Exposure Science

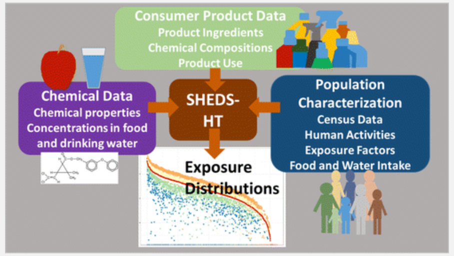
John F. Wambaugh [1], Jane C. Bare [2], Courtney C. Carignan [3], Kathie L. Dionisio [4], Robin E. Dodson [5, 6], Olivier Jolliet [7], Xiaoyu Liu [8], David E. Meyer [2], Seth R. Newton [4], Katherine A. Phillips [4], Paul S. Price [4], Caroline L. Ring [9], Hyeong-Moo Shin [10], Jon R. Sobus [4], Tamara Tal [11], Elin M. Ulrich [4], Daniel A. Vallero [4], Barbara A. Wetmore [4], Kristin K. Isaacs [4]

United States Environmental Protection Agency

Existing HT Models for Key Pathways

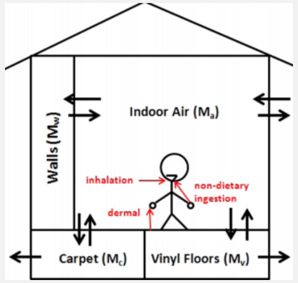
Consumer (Near-Field) Pathways

SHEDS-HT (Isaacs et al., 2014)



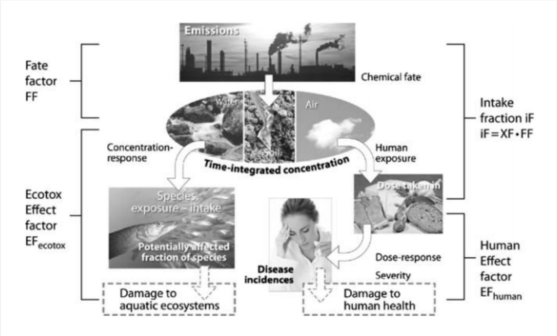
RAIDAR-ICE (Li et al., 2018)

FINE (Shin et al., 2015)

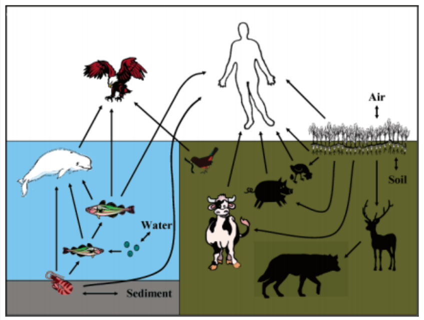


Ambient (Far-Field) Pathways

UseTox (Rosenbaum et al., 2008)

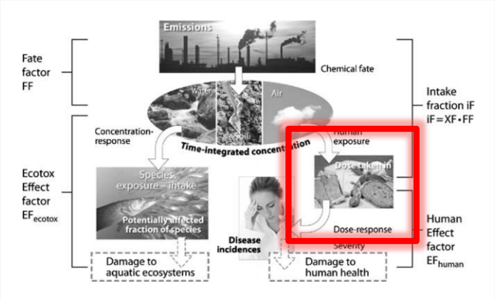


RAIDAR (Arnot et al., 2006, 2008)



Dietary Pathways

UseTox **(Rosenbaum et al. (2008)**



SHEDS-HT (Biryol et al., 2017)

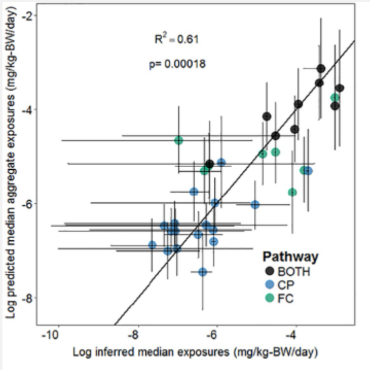
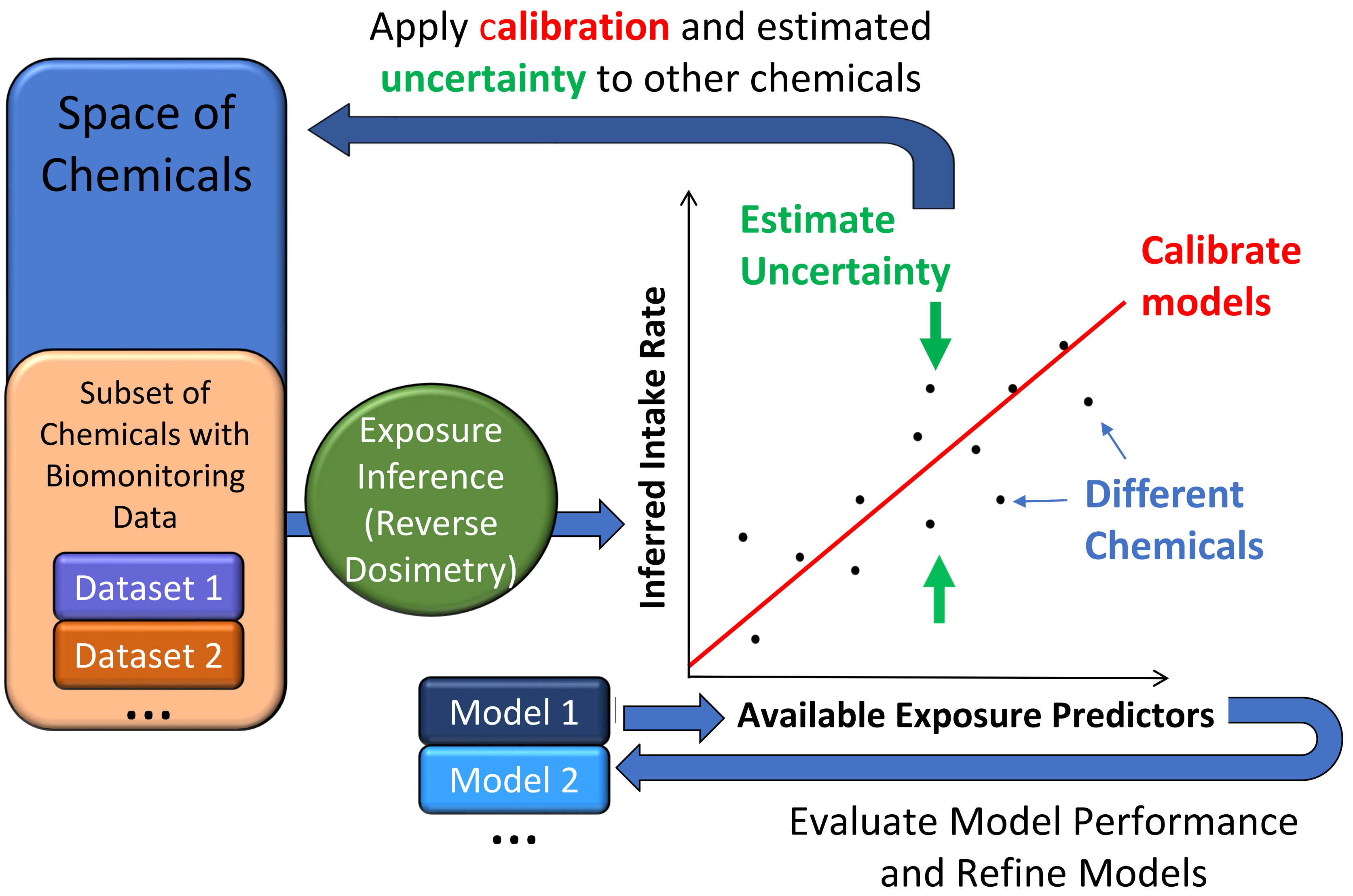


Figure from Kristin Isaacs

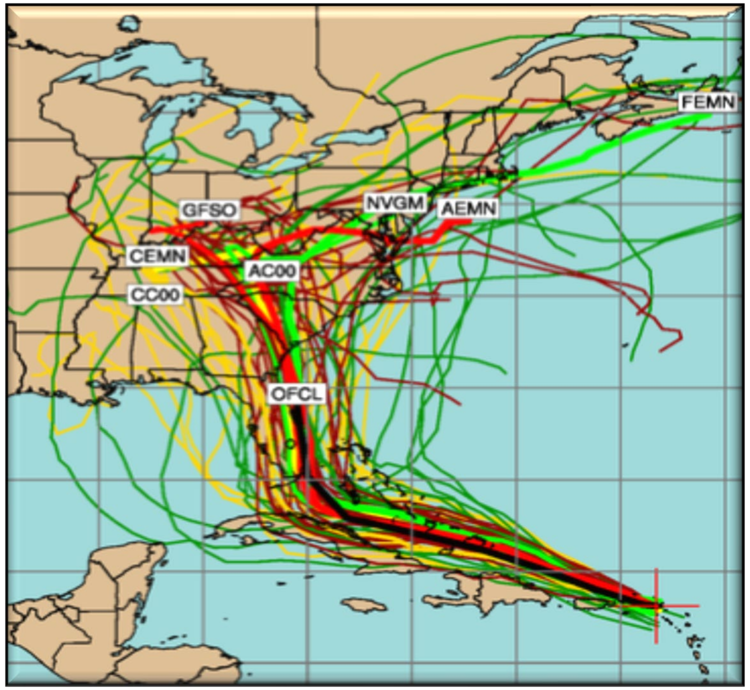
Consensus Exposure Predictions with the SEEM Framework

- Different exposure models incorporate **knowledge**, **assumptions**, and **data** (MacLeod et al., 2010)
- We incorporate multiple models (including SHEDS-HT, USEtox, RAIDAR) into consensus predictions for 1000s of chemicals within the **Systematic Empirical Evaluation of Models (SEEM)** (Wambaugh et al., 2013, 2014, Ring et al., 2019)
- Evaluation is like a sensitivity analysis: What models are working? What data are most needed?



Ensemble Predictions

- We can use ensemble methods to make more stable models and characterize uncertainty
- "Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a (weighted) vote of their predictions." Dietterich (2000)
- Ensemble systems have proven themselves to be very effective and extremely versatile in a broad spectrum of problem domains and real-world applications (Polikar, 2012)
- Ensemble learning techniques in the machine learning paradigm can be used to integrate predictions from multiple tools. Pradeep (2016)



Hurricane Path Prediction is an Example of Integrating Multiple Models

SEEM3 Collaboration

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	Pathway(s)
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

Ring et al. (2018)

SEEM3 Considers Pathway of Exposure

We organize models by the exposure pathways they cover

We calibrate predictors based on ability to explain median NHANES exposure rates

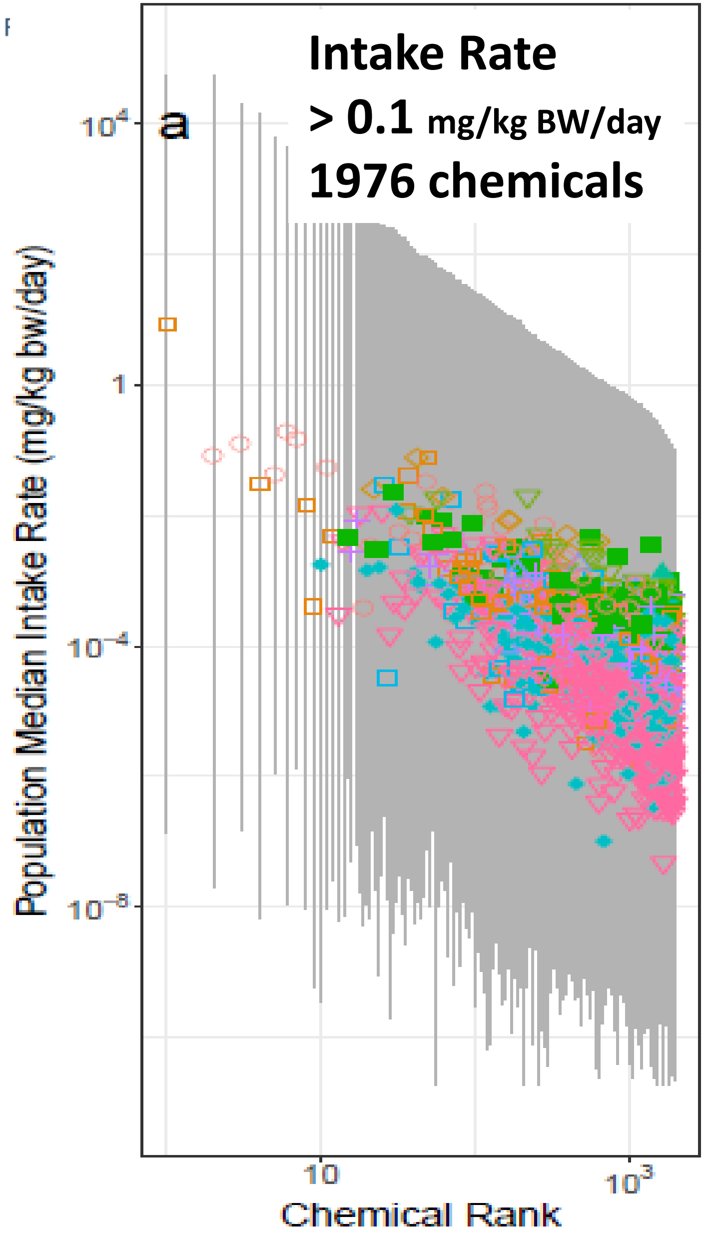
General Population Median Chemical Exposure (mg/kg BW/day)

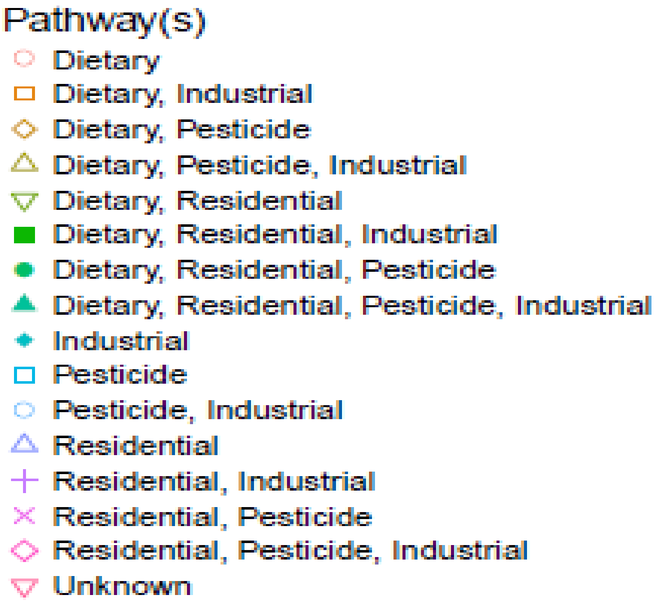
Pathway	Chemical-Specific Pathway Relevancy	Predictors
Residential	Yes/No	Average Unexplained Residential SHEDS-HT Direct Residential SHEDS-HT Indirect Residential USETox FINE RAIDAR-ICE Production Volume
Dietary	Yes/No	Average Unexplained Dietary SHEDS-HT Dietary Production Volume USETox
Far-Field Pesticides	Yes/No	Average Unexplained Pesticide Pesticide REDs USETox Stockholm Convention Production Volume
Far-Field Industrial	Yes/No	Average Unexplained Industrial USEtox RAIDAR Stockholm Convention Production Volume
Unknown		Average Unexplained Overall

Ring et al. (2018)

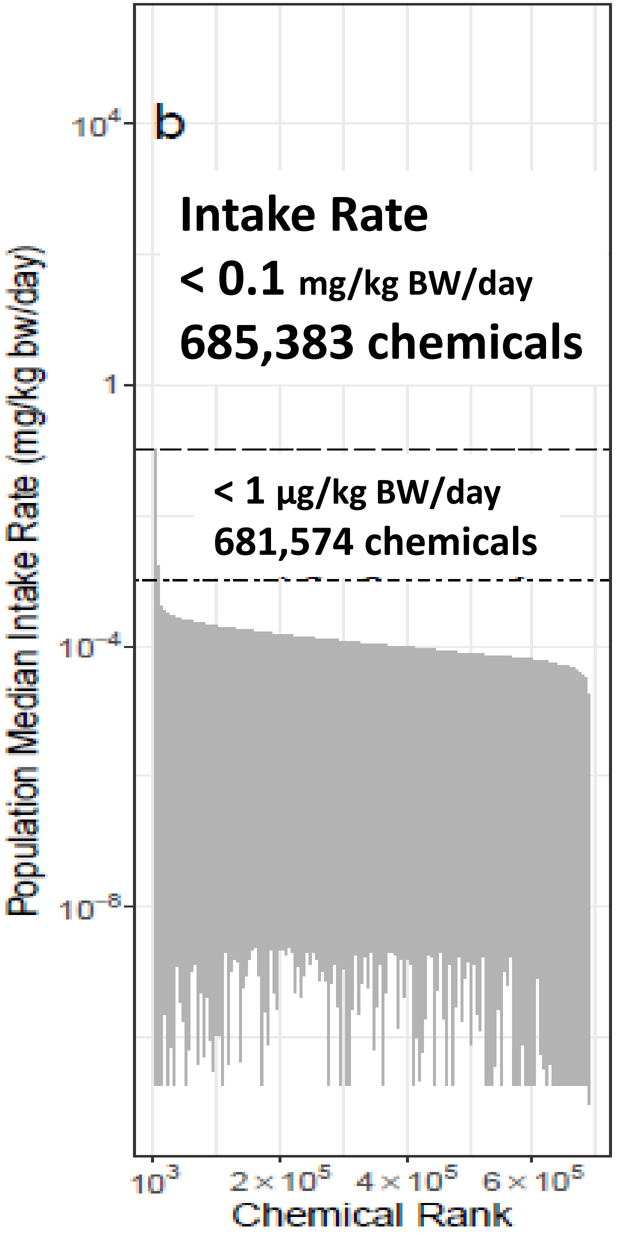
Consensus Modeling of Median Chemical Intake

United States Environmental Protection Agency



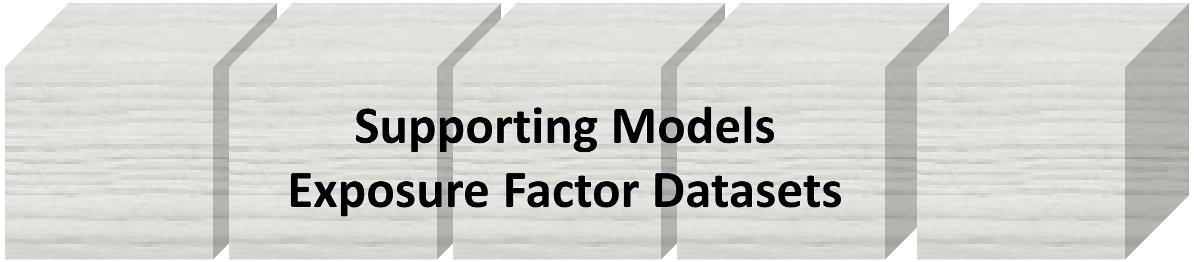


Of 687,359 chemicals evaluated, 30% have less than a 50% probability for exposure via any of the four pathways and are considered outside the "domain of applicability"



Ring et al. (2018)

ExpoCast SEEM Models: Required Building Blocks for the Output

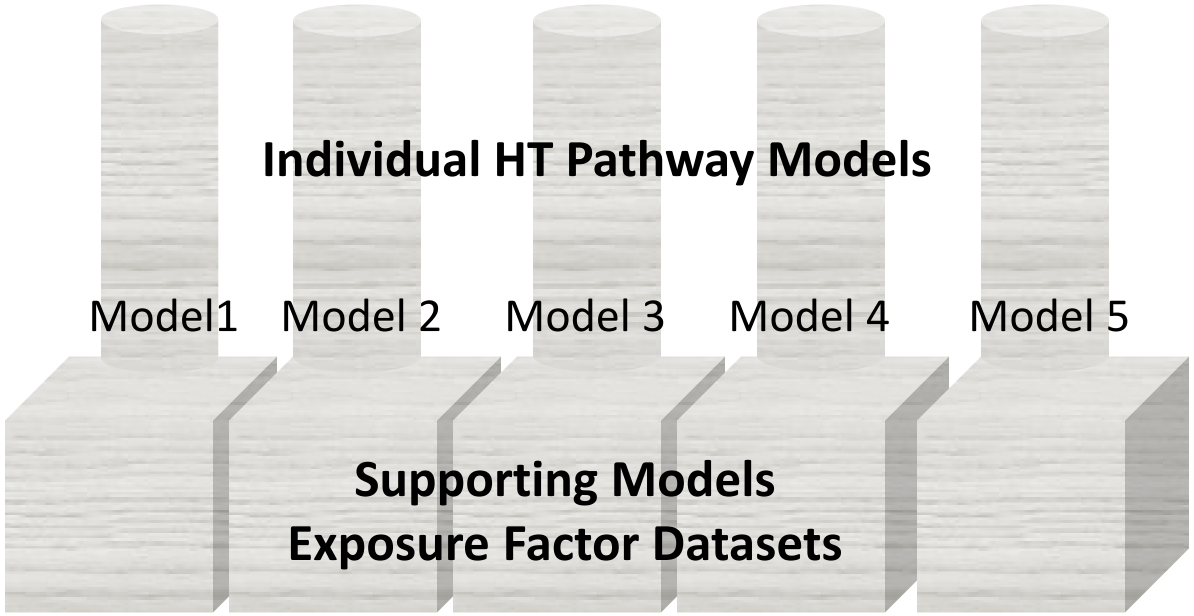


Machine-learning models for filling gaps from structure when no data are available

Composition and use/release data

Slide from Kristin Isaacs

ExpoCast SEEM Models: Required Building Blocks for the Output



for example, SHEDS-HT, HT ChemSteer, external models

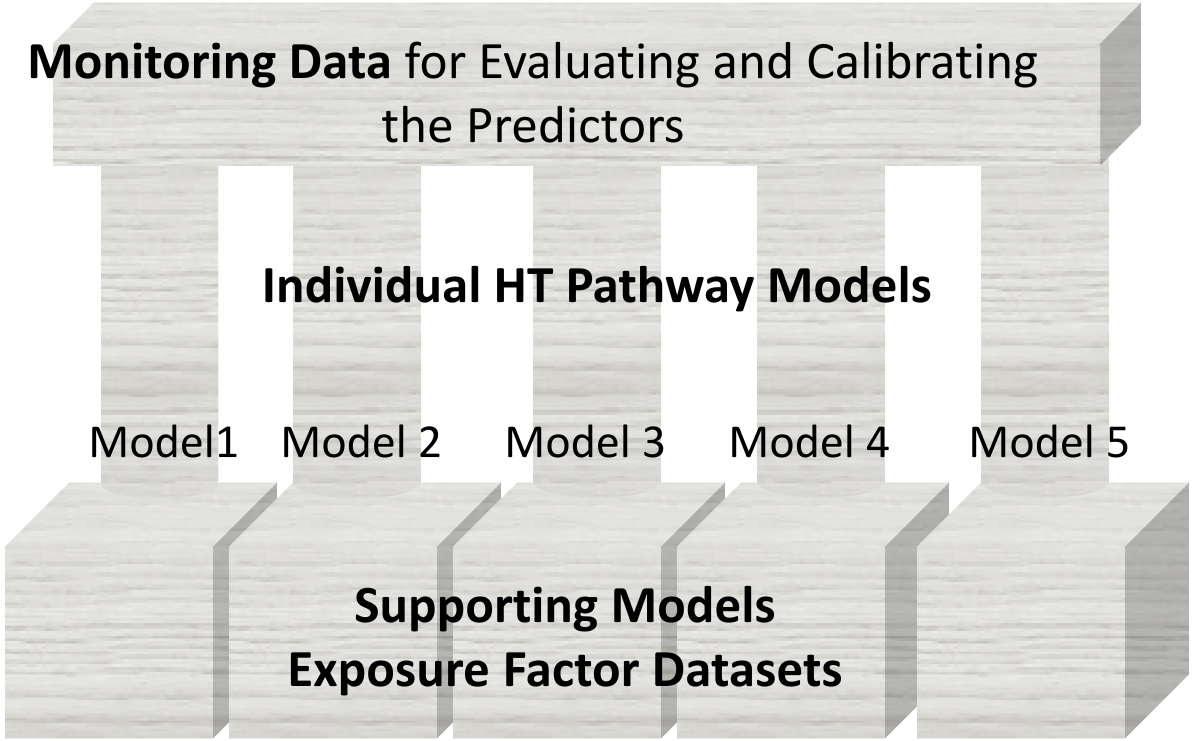
Machine-learning models for filling gaps from structure when no data are available

Composition and use/release data

Slide from Kristin Isaacs

United States Environmental Protection Agency

ExpoCast SEEM Models: Required Building Blocks for the Output



Including NHANES biomonitoring and USGS water datasets

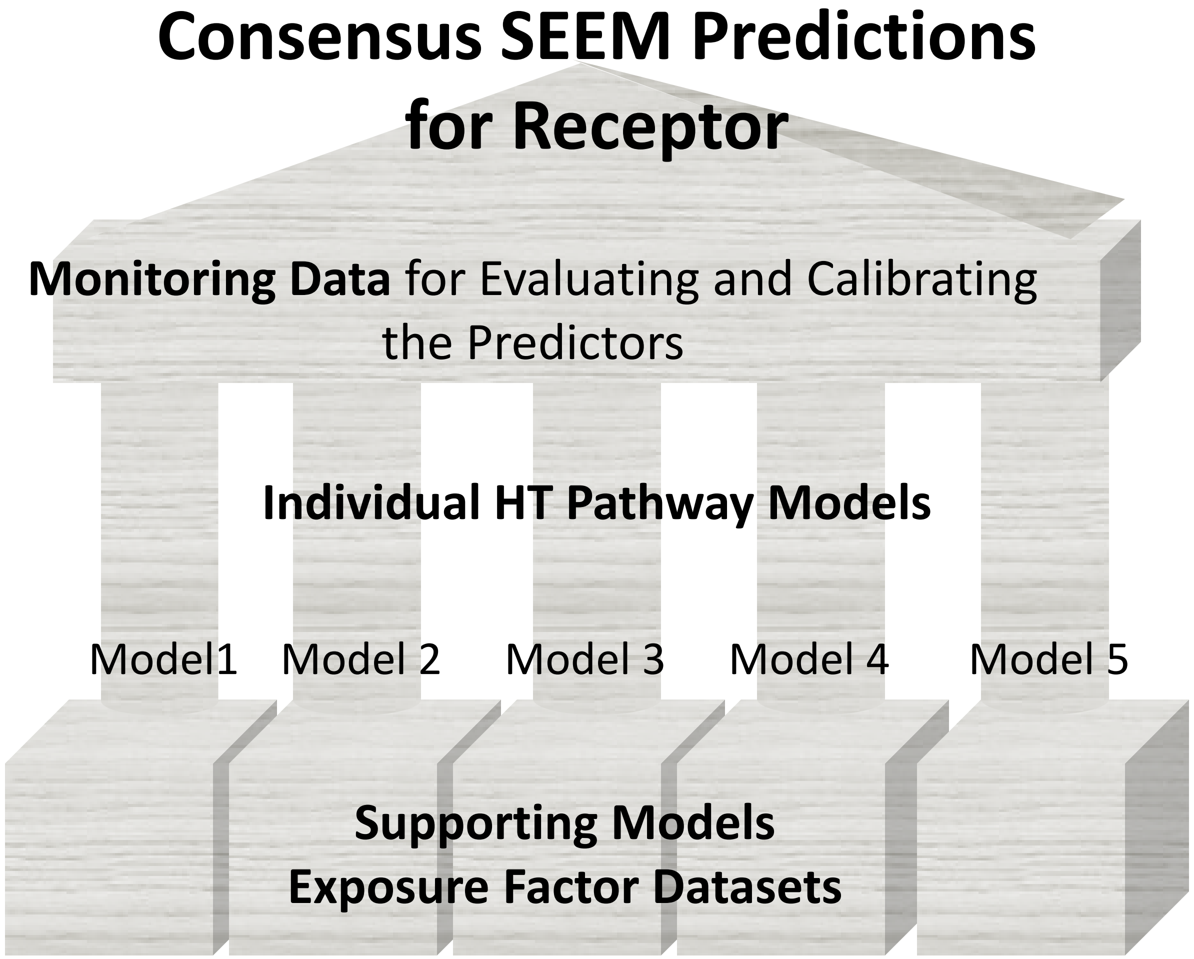
for example, SHEDS-HT, HT ChemSteer, external models

Machine-learning models for filling gaps from structure when no data are available

Composition and use/release data

Slide from Kristin Isaacs

ExpoCast SEEM Models: Required Building Blocks for the Output



Including NHANES biomonitoring and USGS water datasets

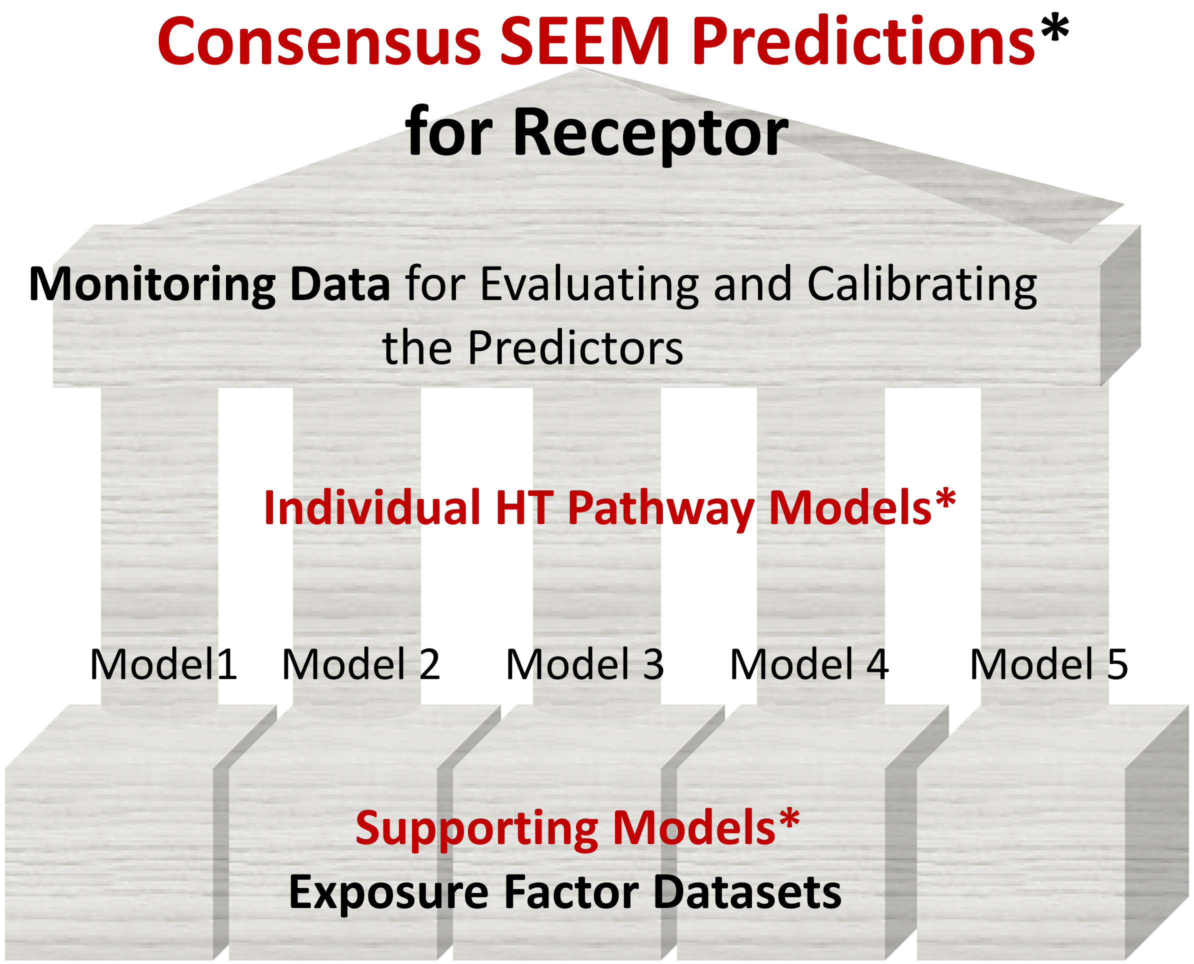
for example, SHEDS-HT, HT ChemSteer, external models

Machine-learning models for filling gaps from structure when no data are available

Composition and use/release data

Slide from Kristin Isaacs

ExpoCast SEEM Models: Required Building Blocks for the Output



***New Approach Methodologies for Exposure:**

Application to Real Decision Contexts

Including NHANES biomonitoring and USGS water datasets

for example, SHEDS-HT, HT ChemSteer, external models

Machine-learning models for filling gaps from structure when no data are available

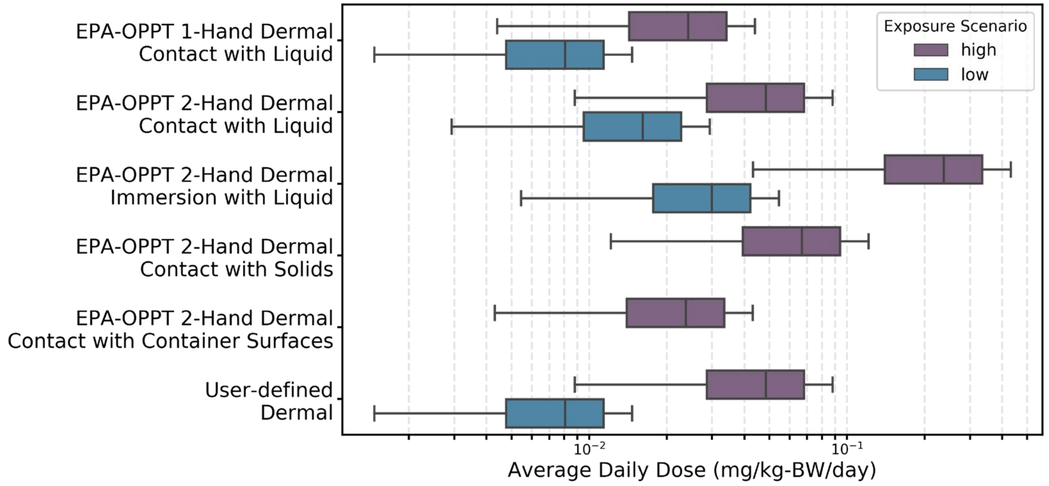
Composition and use/release data

Slide from Kristin Isaacs

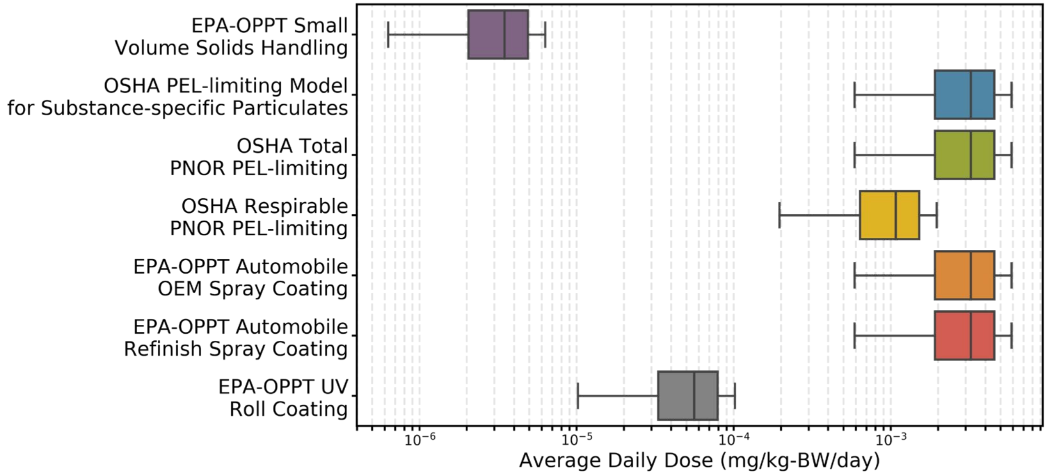
Formatting Occupational Exposure Models for HT Use

- We have developed consensus models for consumer and some ambient pathways, but ecological and occupational consensus models are ongoing
- Many predictors for these pathways exist, but they are not typically oriented for high throughput capacity, for example EPA's ChemSTEER (Chemical Screening Tool for Exposures and Environmental Releases)
- Command Line Occupational Exposure Tool (CLOET) a command line tool that allows use of ChemSTEER v3.0 in a high throughput manner
- Multiple scenarios for each model have been run and tested against ChemSTEER GUI to test for model fidelity.

Dermal Models



Inhalation Models



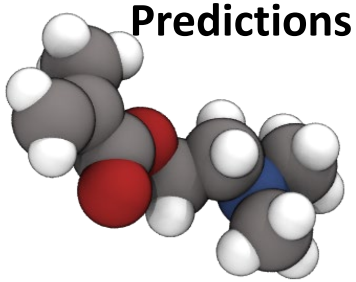
Concentrations were varied from 0.1 to 1

Slide from Katherine Phillips

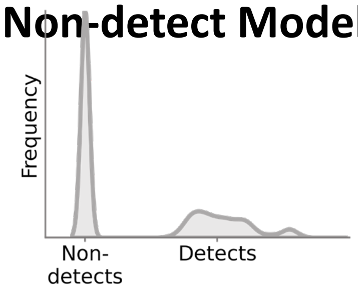
Two-Stage Occupational Exposure Model

- OSHA's chemical exposure health data set for air samples was used to build a two-stage model that predicts 1) if a chemical is likely to be detected in air and 2) what the likely concentration would be
- OPERA physicochemical property distributions across NAICS sector and subsectors are included as input distributions to the models in addition to the OSHA data
- Bayesian Hierarchical Regression allows us to organize our predictions (either detect/non-detect or concentration) by NAICS Sector and/or Subsector

OPERA Property Predictions

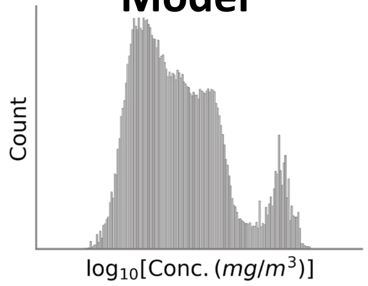


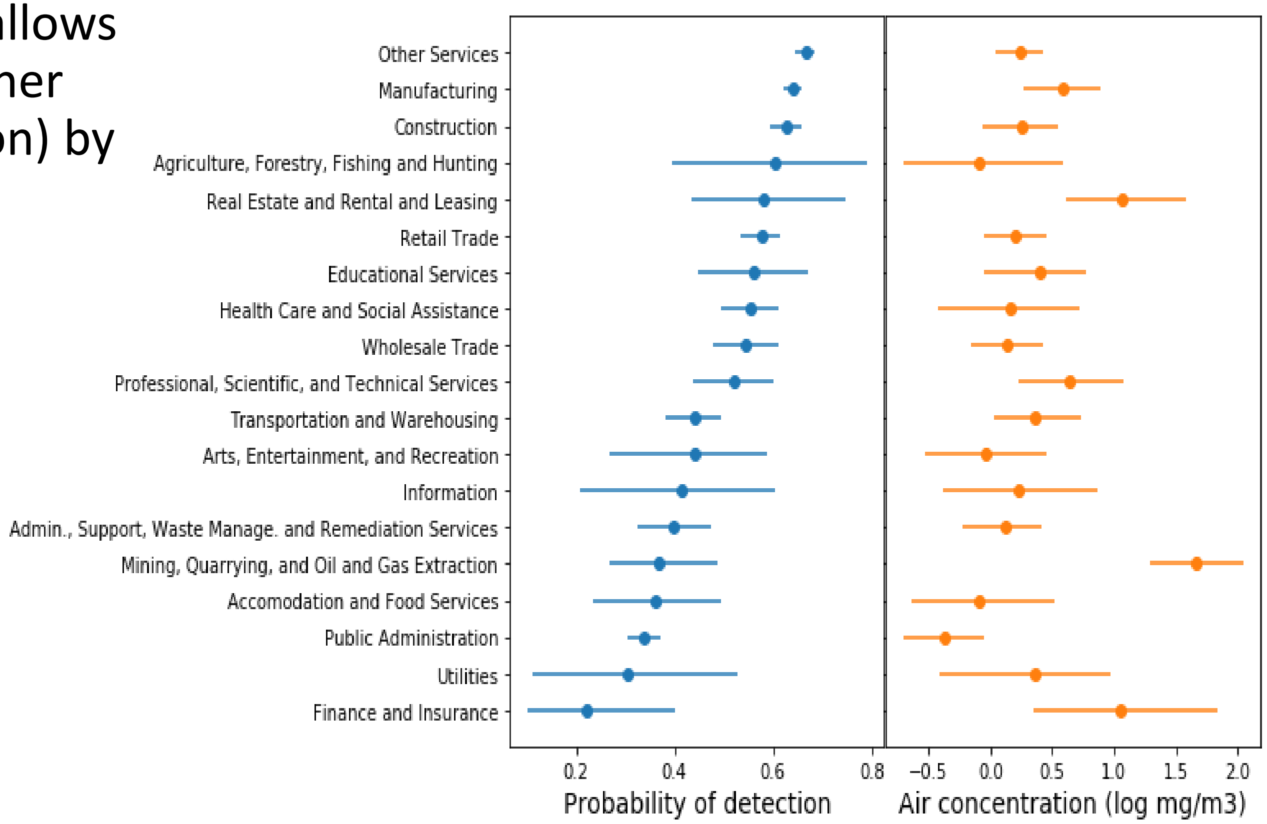
Detect / Non-detect Model



Non-detects

Air Concentration Model





Minucci et al, *in preparation*

Slide from Katherine Phillips

EcoSEEM Metamodel for Surface Water Chemical Concentrations

Chemical-specific information → Release (loading) and fate predictors → **EcoSEEM metamodel** ← USGS/EPA water monitoring data

(l) (p) (m) (y)

Sayre et al,
in preparation

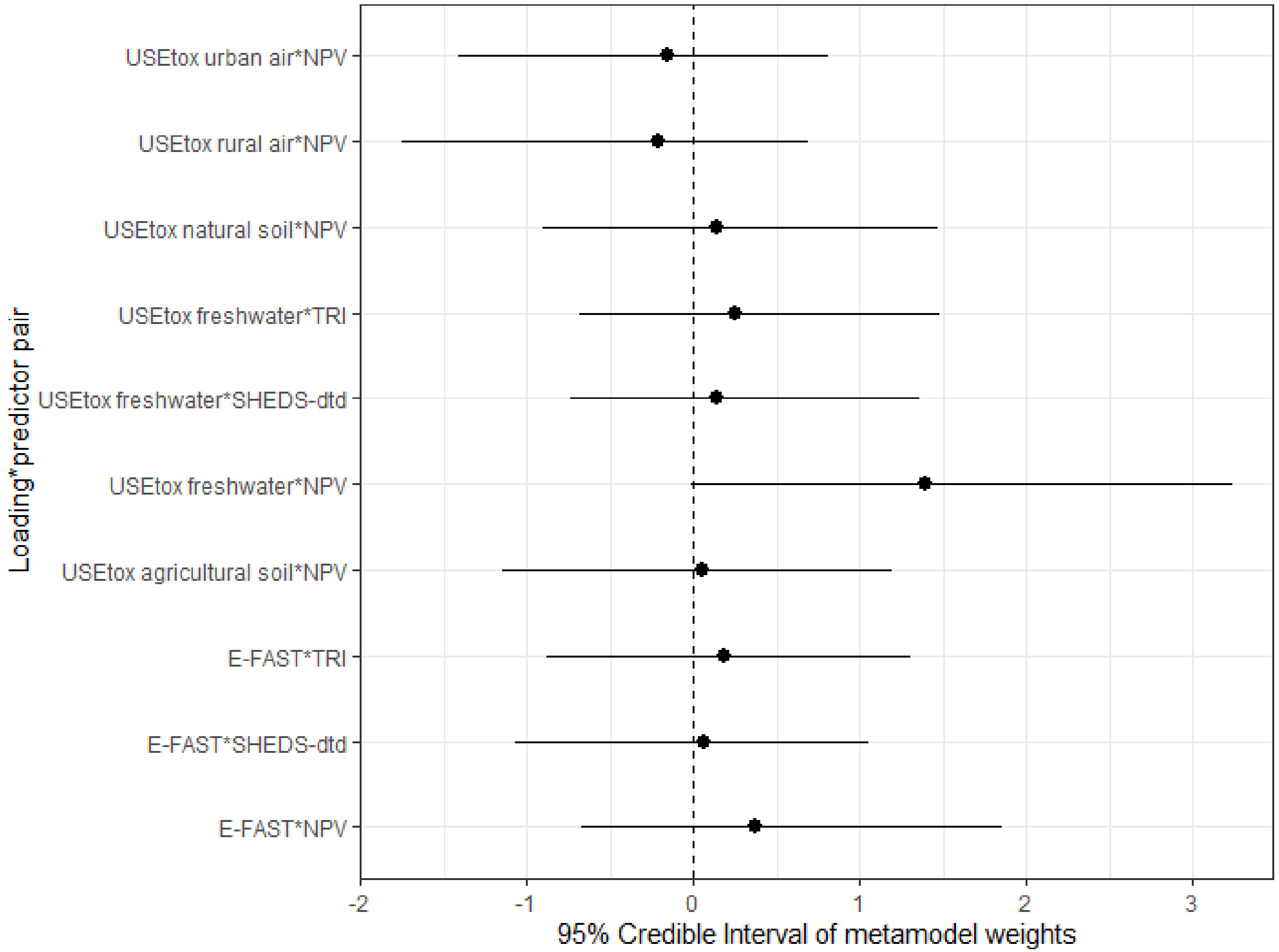
$$\ln y_i = m_0 + \sum_{j=1}^{n_j} \sum_{k=1}^{n_{kj}} m_{jk} \ln\left(l_{ji}\, p_{ki}\right)$$

Slide from Risa Sayre

EcoSEEM Evaluating Predictive Ability of HT Surface Water Models

- The strength of the correlation between each combination of release and fate model predictions and the observed water concentrations allows model calibration

- The most informative pair for bulk concentrations was USEtox freshwater model using loadings from NPV

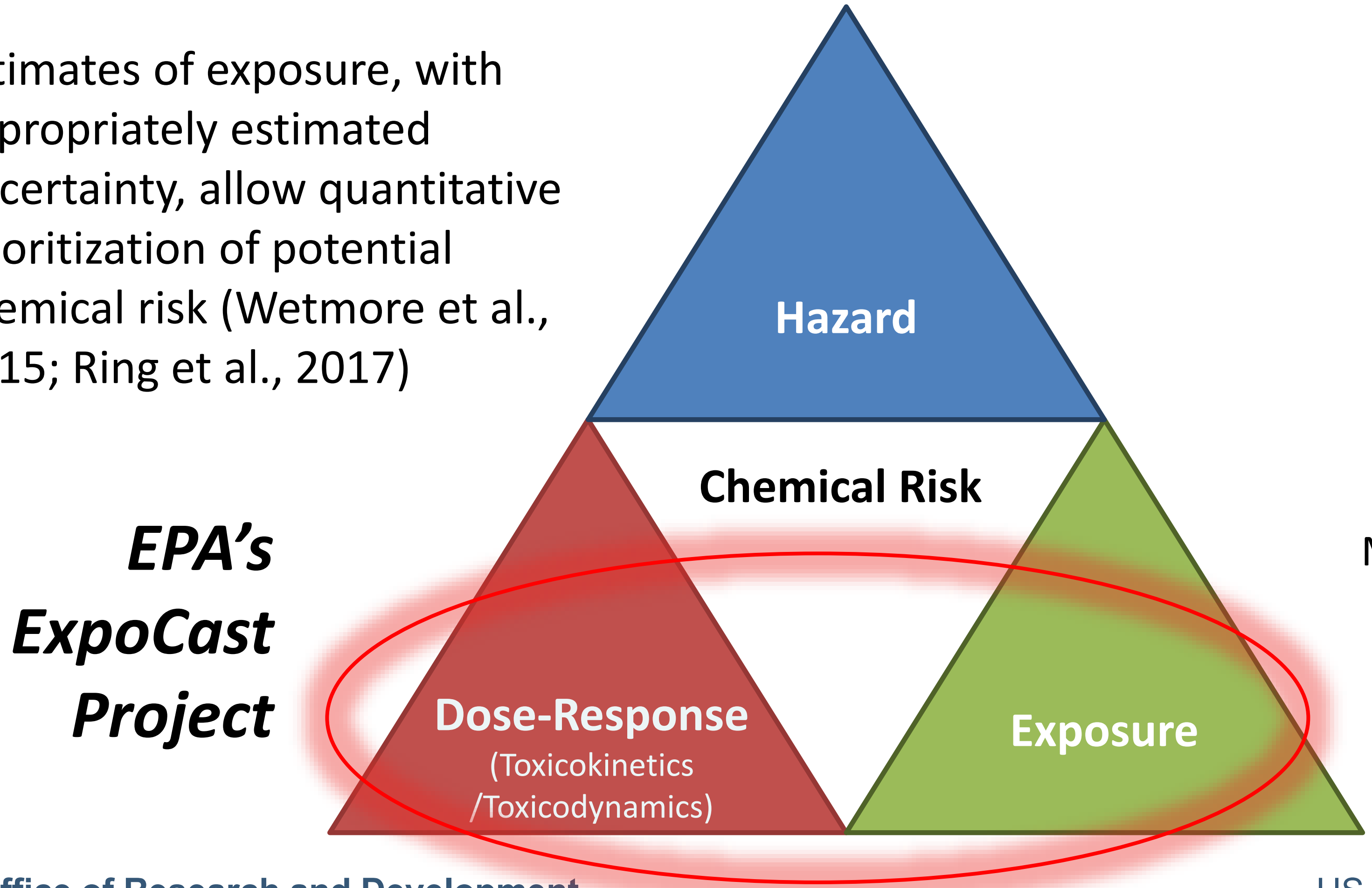


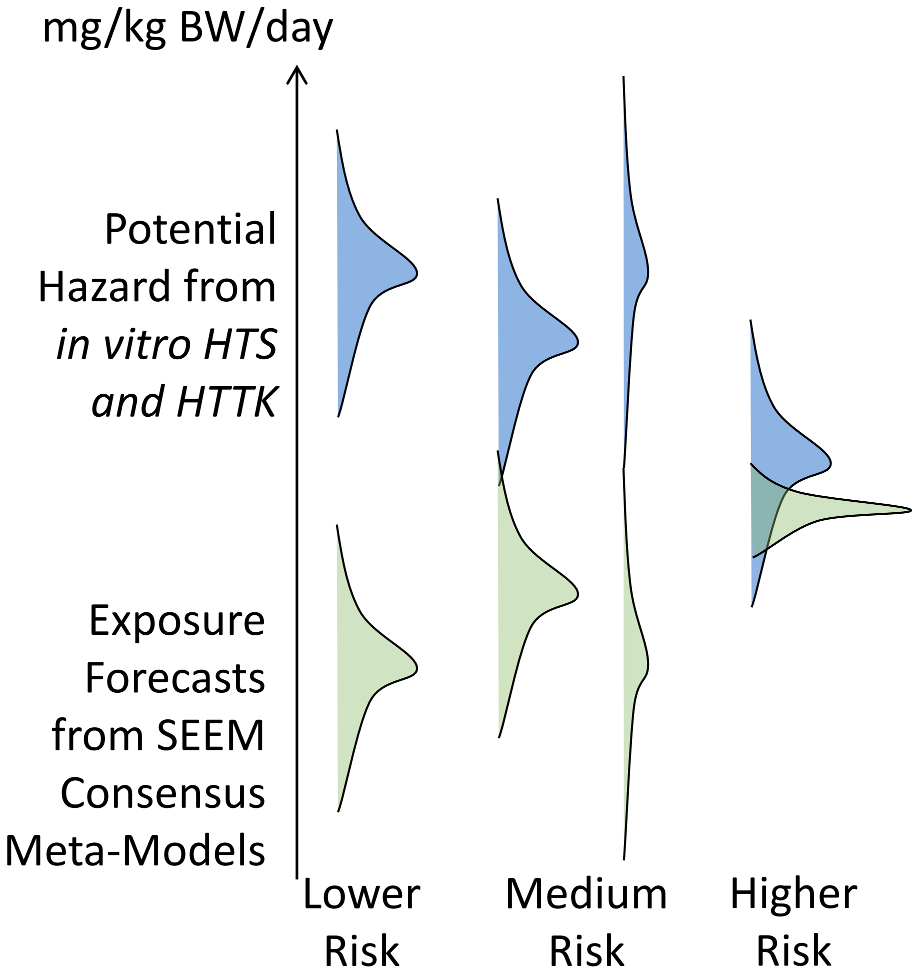
Sayre et al,
in preparation

Outlook

- SEEM metamodels have been developed for consumer and some ambient pathways (Ring et al., 2018) and ecological and occupational consensus models are in development

- Estimates of exposure, with appropriately estimated uncertainty, allow quantitative prioritization of potential chemical risk (Wetmore et al., 2015; Ring et al., 2017)





ExpoCast Project (Exposure Forecasting)

Center for Computational Toxicology and Exposure

Linda Adams
Lucas Albrecht*
Matthew Boyce*
Miyuki Breen*
Alex Chao
Daniel Dawson*
Mike Devito
Alex East*
Lindsay Eddy*
Christopher Eklund
Peter Egeghy
Marina Evans
Alex Fisher*
Rocky Goldsmith
Louis Groff*
Chris Grulke

Colin Guider*
Mike Hughes
Victoria Hull*
Kristin Isaacs
Richard Judson
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Seth Newton
Alli Phillips
Katherine Phillips
Paul Price
Tom Purucker
Ann Richard
Caroline Ring

Risa Sayre
Mark Sfeir*
Marci Smeltz*
Jon Sobus
Zach Stanfield*
Mike Tornero-Velez
Rusty Thomas
Elin Ulrich
Dan Vallero
Barbara Wetmore
John Wambaugh
Antony Williams

CEMM

Hongwan Li*
Xiaoyu Liu
Zachary Robbins*
Mark Strynar

*Trainees

Collaborators

Arnot Research and Consulting
Jon Arnot
Johnny Westgate
Integrated Laboratory Systems
Xiaoqing Chang
Shannon Bell
National Toxicology Program
Steve Ferguson
Kamel Mansouri
Ramboll
Harvey Clewell
Silent Spring Institute
Robin Dodson
Simulations Plus
Michael Lawless
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Aylward
Technical University of Denmark
Peter Fantke
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
University of Nevada
Li Li
University of North Carolina, Chapel Hill
Julia Rager
Marc Serre

References

EPA ORD Publications in Bold

Arnot, J.A, et al. 2006. Screening level risk assessment model for chemical fate and effects in the environment. Environ. Sci. Technol. 40: 2316-2323

Arnot, J.A.; Mackay D. 2008. Policies for chemical hazard and risk priority setting: Can persistence, bioaccumulation, toxicity and quantity information be combined? Environ. Sci. Technol. 42: 4648-4654. DOI: 10.1021/es800106g

Barber MC, et al (2017). "Developing and applying metamodels of high resolution process-based simulations for high throughput exposure assessment of organic chemicals in riverine ecosystems." Science of the Total Environment, 605, 471-481.

Bennett, D. H.; Furtaw, E. J., Fugacity-based indoor residential pesticide fate model. Environmental Science & Technology 2004, 38, (7), 2142-2152.

Biryol D, et al.. High-throughput dietary exposure predictions for chemical migrants from food contact substances for use in chemical prioritization. Environment international. 2017 Nov 1;108:185-94.

Dietterich, Thomas G. "Ensemble methods in machine learning." *International workshop on multiple classifier systems*. Springer, Berlin, Heidelberg, 2000. **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.**

Fantke P, Jolliet O. Life cycle human health impacts of 875 pesticides. The International Journal of Life Cycle Assessment. 2016 May 1;21(5):722-33.

Fantke P , et al.. Dynamic toxicity modelling based on the USEtox matrix framework. InSETAC Europe 25th Annual Meeting 2015 (pp. 33-34). SETAC Europe.

Huang, L , et 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.

Isaacs KK , et al.. SHEDS-HT: an integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources. Environmental science & technology. 2014 Nov 4;48(21):12750-9.

Jolliet, O , et al., P., Defining Product Intake Fraction to Quantify and Compare Exposure to Consumer Products. Environmental Science & Technology 2015, 49, (15), 8924-8931.

Lallas PL. The Stockholm Convention on persistent organic pollutants. The American Journal of International Law. 2001 Jul 1;95(3):692-708.

Li L, et al.. A model for risk-based screening and prioritization of human exposure to chemicals from near-field sources. Environmental science & technology. 2018 Nov 8;52(24):14235-44.

MacLeod, Matthew, et al. "The state of multimedia mass-balance modeling in environmental science and decision-making." (2010): 8360-8364.Minucci et al, *in preparation*

Minucci, Jeff et al. "High Throughput Model for Occupational Exposure"

Polikar, Robi. "Ensemble learning." Ensemble machine learning. Springer, Boston, MA, 2012. 1-34.

Pradeep, Prachi, et al. "An ensemble model of QSAR tools for regulatory risk assessment." Journal of cheminformatics 8.1 (2016): 1-9.

Rosenbaum RK, et al. USEtox—the UNEP-SETAC toxicity model: recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. The International Journal of Life Cycle Assessment. 2008 Nov 1;13(7):532.

Ring, Caroline L., et al. "Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability." *Environment international* 106 (2017): 105-118.

Ring CL, et al. Consensus modeling of median chemical intake for the US population based on predictions of exposure pathways. Environmental science & technology. 2018 Dec 5;53(2):719-32.

Sayre et al, *"Consensus Model for Predicting Chemical Surface Water Concentrations"*

Shin, H.-M., et al., 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

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).

Wambaugh JF, et al. New approach methodologies for exposure science. Current Opinion in Toxicology. 2019 Jun 1;15:76-92.

Wetmore, B. A., et al. Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing. Toxicol. Sci. 2015, 148 (1), 121–36.

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.