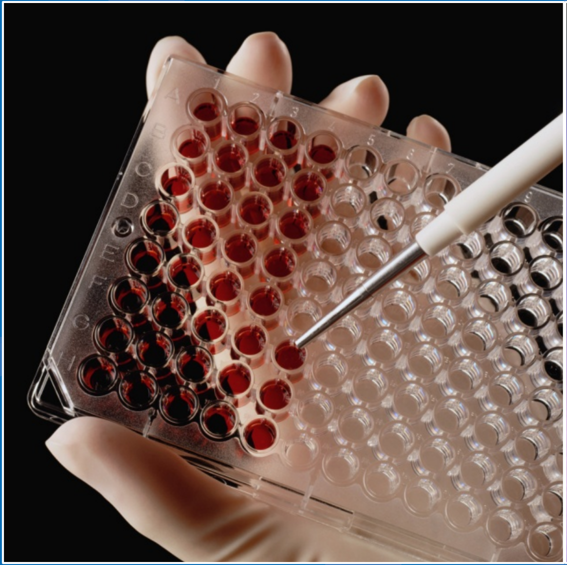
High Throughput Exposure Prediction for the ExpoCast Project

John Wambaugh
U.S. EPA, National Center for Computational Toxicology



August 23, 2012

Introduction

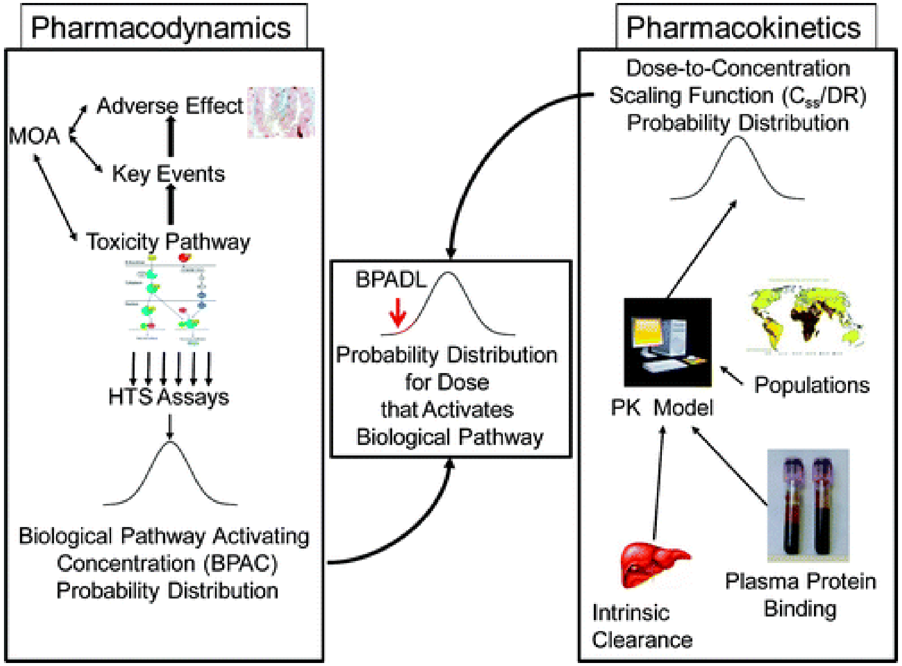
There are thousands of environmental chemicals, many without enough data for evaluation

Risk is the product of hazard and exposure

High throughput *in vitro* methods beginning to bear fruit on hazard for many of these chemicals

Methods exist for approximately converting these *in vitro* results to daily doses needed to produce similar levels in a human (Wetmore *et al.* (2011))

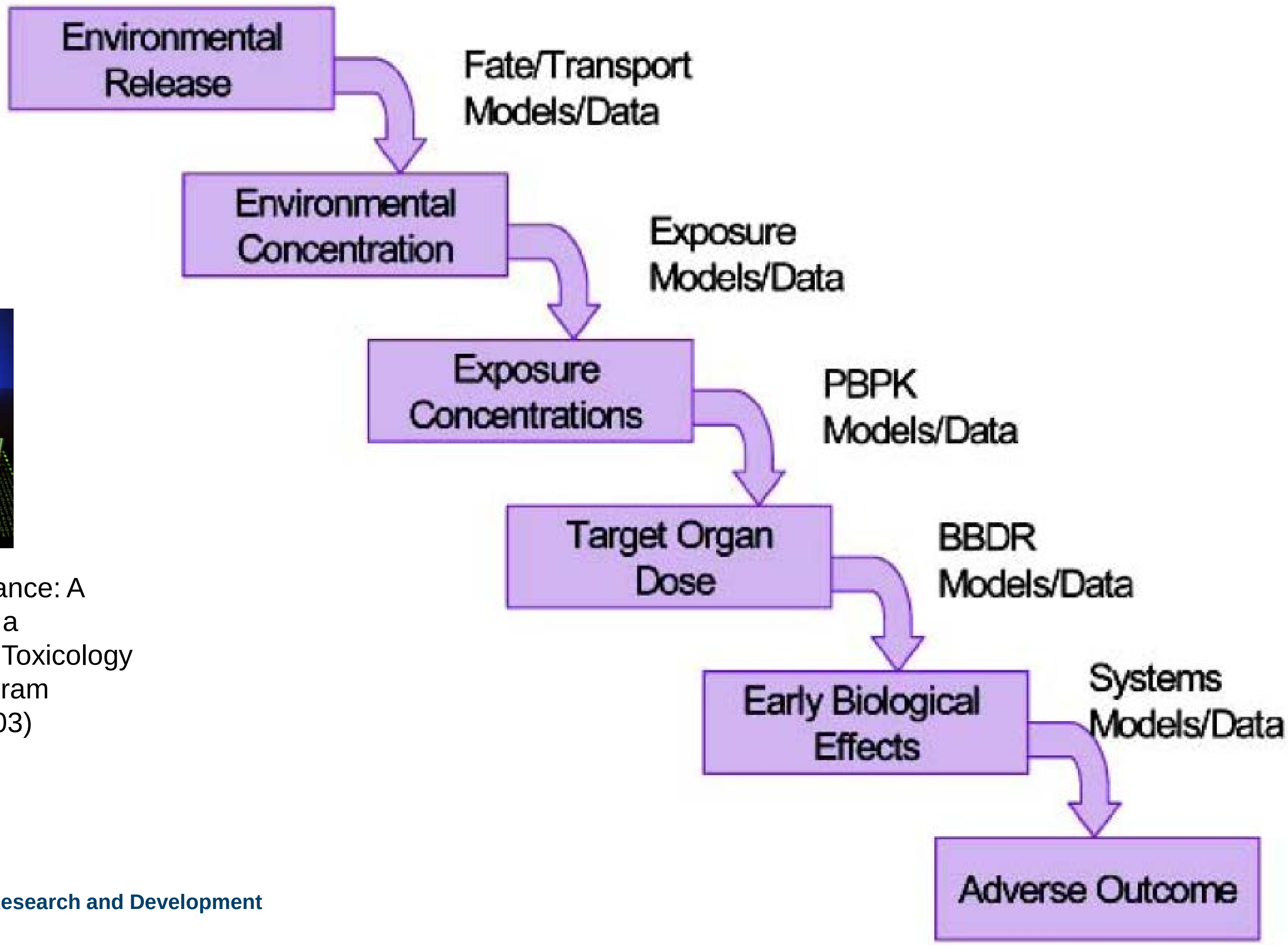
Without a similar capacity for exposure cannot place risk early into prioritizations

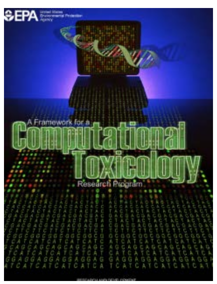


Judson *et al.*, (2011) "Estimating Toxicity-Related Biological Pathway Altering Doses for High-throughput Chemical Risk Assessment" Chemical Research in Toxicology **24** 451-462

EPA United States Environmental Protection Agency

Source-to-Outcome Continuum

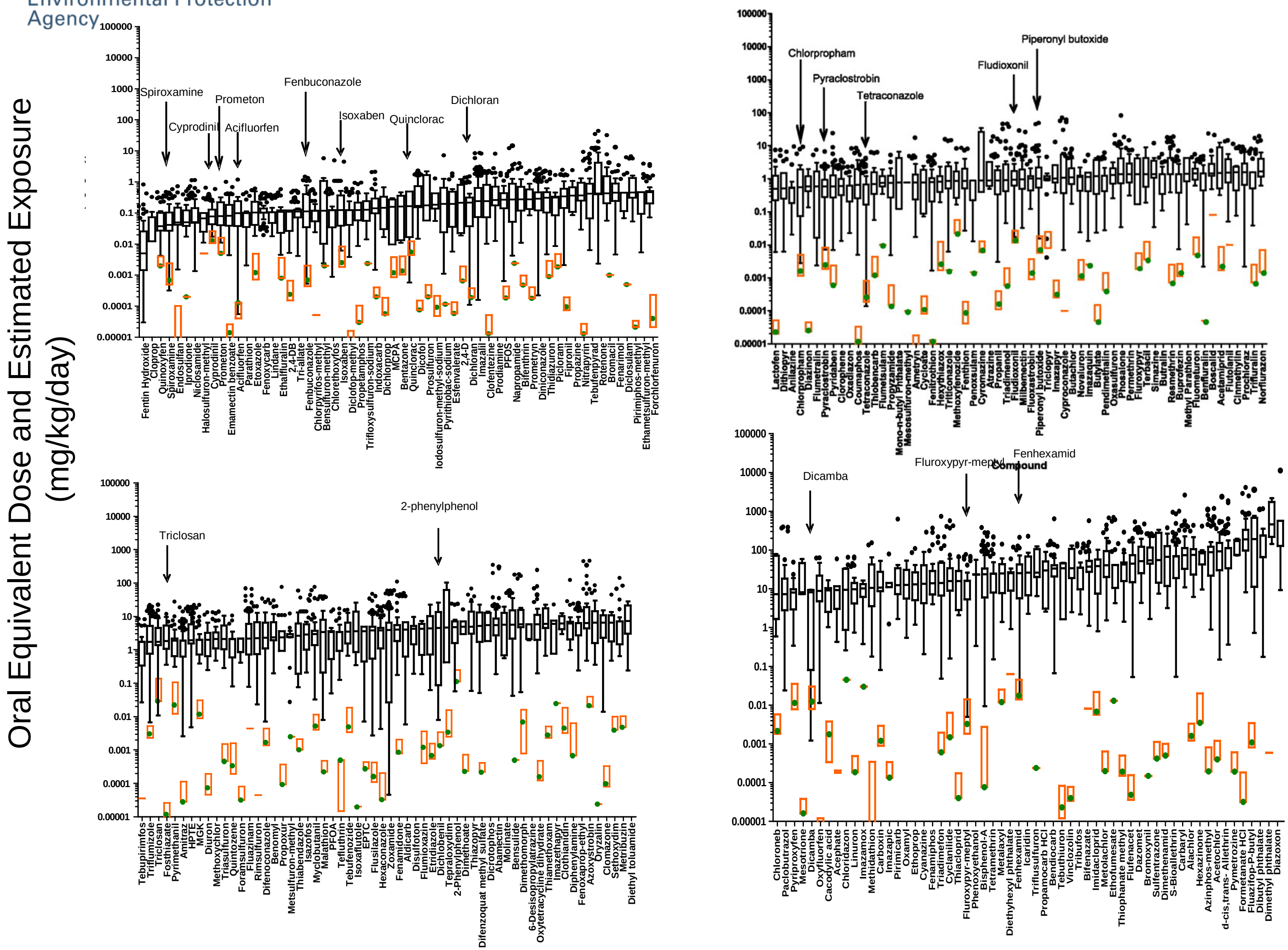




Strategic Guidance: A Framework for a Computational Toxicology Research Program (November 2003)

Oral Doses Equivalent to ToxCast Concentrations

United States Environmental Protection Agency



Wetmore et al. *Tox. Sci (2011)*

High Throughput Exposure Prioritization

United States Environmental Protection Agency

Goal: A high-throughput exposure approach to use with the ToxCast chemical hazard identification.

Proof of Concept: Using off-the-shelf models capable of quantitatively predicting exposure determinants in a high throughput (1000s of chemicals) manner

To date have found only fate and transport models to have sufficient throughput

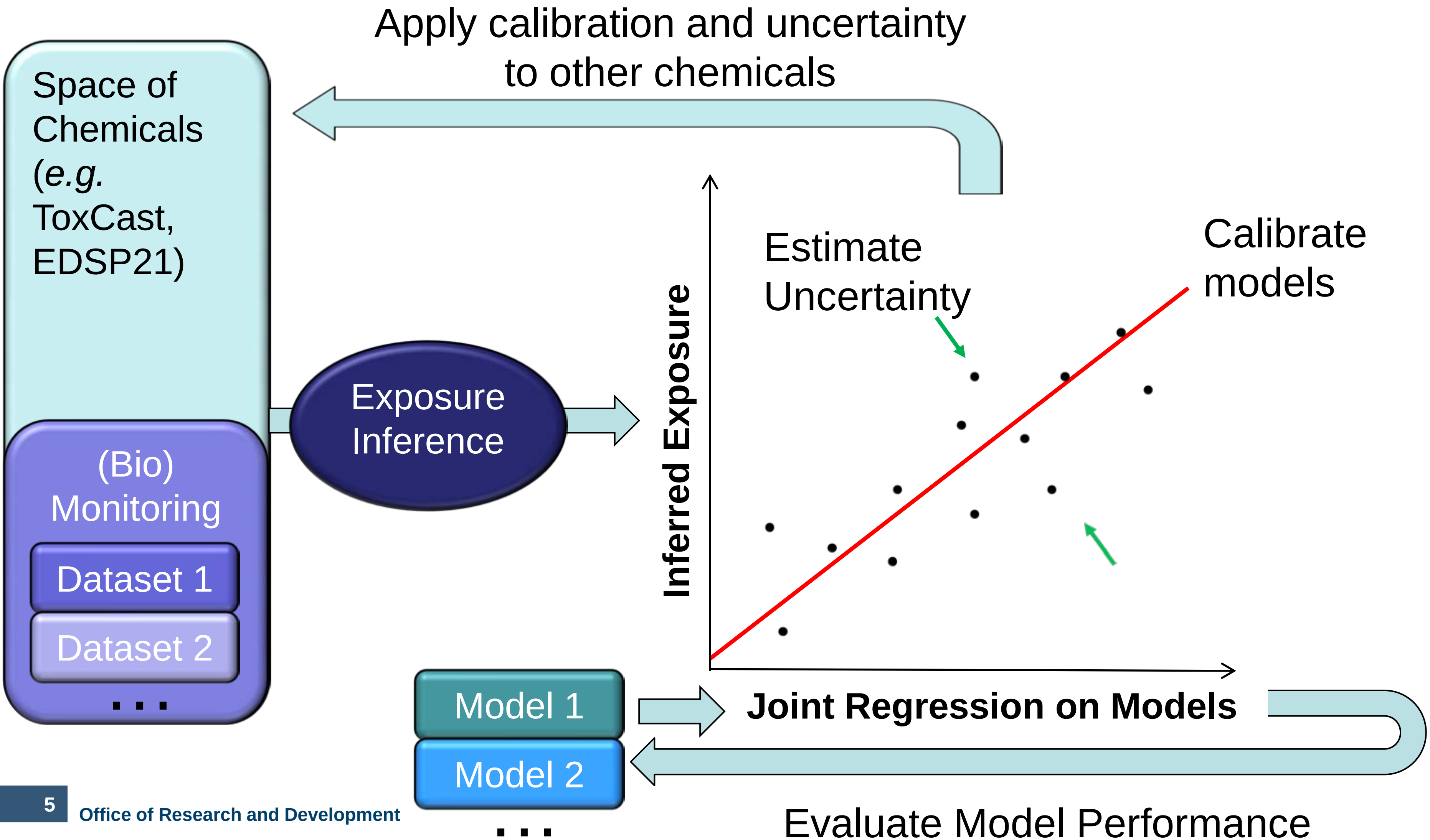
These models predict the contribution from manufacture and industrial use to overall exposure rapidly and efficiently

Applying and developing new high throughput models of consumer use and indoor exposure

Environmental Fate and Transport

Consumer Use and Indoor Exposure

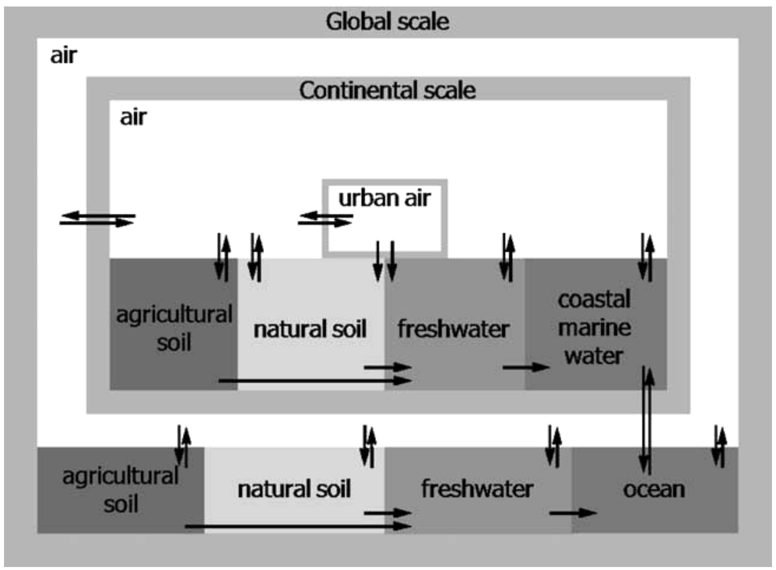
Framework for High Throughput Exposure Screening



Off the Shelf Models

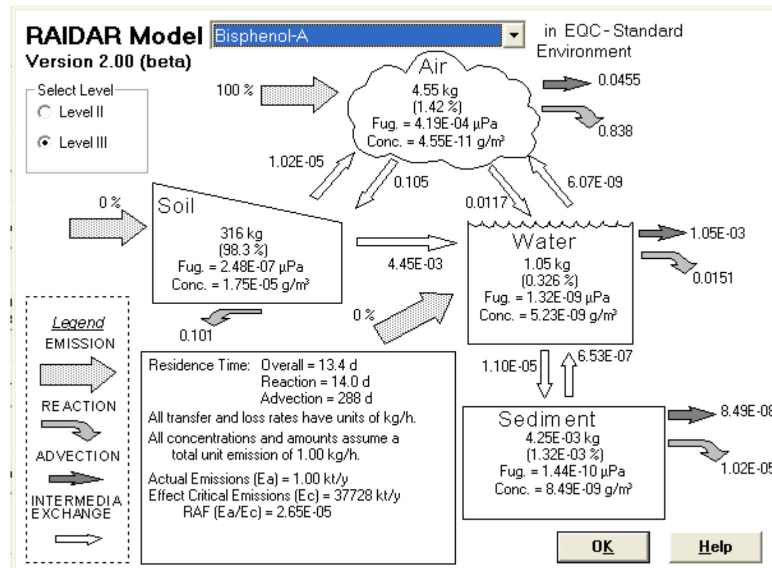
Treat different models like related high-throughput assays

USEtox



United Nations Environment Program and Society for Environmental Toxicology and Chemistry toxicity model Version 1.01
Rosenbaum *et al.* 2008

RAIDAR

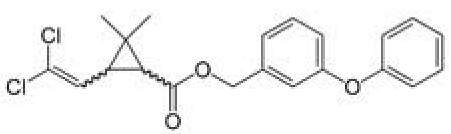


Risk Assessment IDentification And Ranking model Version 2.0
Arnot *et al*. 2006

Parameterizing the Models

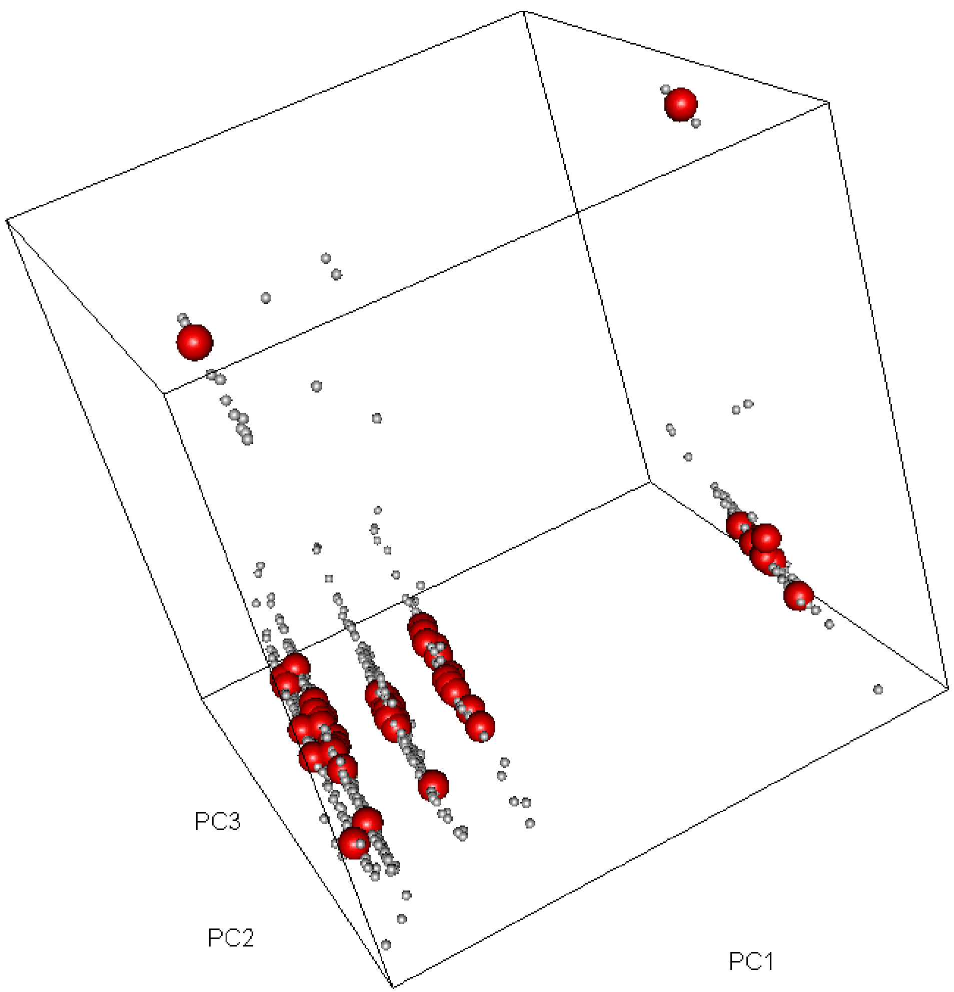
Model parameters mostly predicted from structure (SMILES)

Cl/C(Cl)=C/C3C(C(=O)OCc2cccc(Oc1ccccc1)c2)C3(C)C



Variable	Description	Unit	Source	Default	QSAR	USEtox	RAIDAR
Chemical Name			ToxCast			Yes	Yes
CAS			ToxCast			Yes	Yes
MW	Molecular Weight	g/mol	ToxCast			Yes	Yes
Data Temperature		Degrees C		25			Yes
K_{OW}	Octanol:Water Partition Coefficient	1	Episuite		Yes	Yes	Yes
Koc	Organic Carbon:Water Partition Coefficient	L/kg	USEtox QSAR		Yes	Yes	
K_H25C	Henry's Law Coefficient (25 degrees C)	Pa*M^3/mol	Episuite		Yes	Yes	Yes
Pvap25	Vapor Pressure (25 degrees C)	Pa	Episuite		Yes	Yes	Yes
Sol25	Solubility (25 degrees C)	mg/L	Episuite		Yes	Yes	Yes
K_{DOC}	Dissolved Organic Carbon:Water Partition Coefficient	L/kg	USEtox QSAR		Yes	Yes	
$kdeg_A$	Degradation Rate in Air	1/s	Episuite		Yes	Yes	Yes
$kdeg_W$	Degradation Rate in Water	1/s	Episuite		Yes	Yes	Yes
$kdeg_{Sd}$	Degradation Rate in Sediment	1/s	Episuite		Yes	Yes	Yes
$kdeg_{Sl}$	Degradation Rate in Soil	1/s	Episuite		Yes	Yes	Yes
$kdeg_{biota}$	Degradation Rate in Biota	1/s		2.40E+12			Yes
$kdeg_{human}$	Degradation Rate in Humans	1/s		2.40E+12			Yes
pKa	Acid Dissociation Constant	1	QikProp		Yes		Yes
BAF	Bioaccumulation Factor	L/kg	EpiSuite		Yes	Yes	
LC50	Average Log Aquatic 50% Lethal Concentration	Log(mg/L)	EcoSAR		Yes	Yes	

Chemical Landscape



Range of physico-chemical properties for the 1600 chemicals evaluated

Principal component one: half-life in soil, water, and sediment

Principal component two: octonol-water partition coefficient (logP)

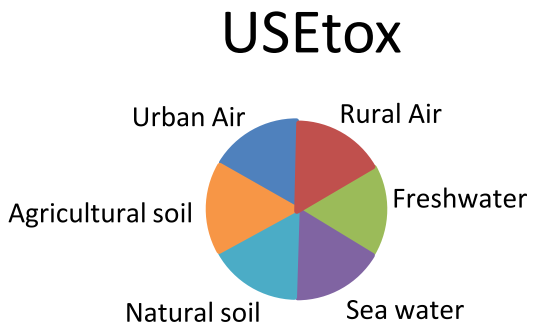
Principal component three: half-life in air

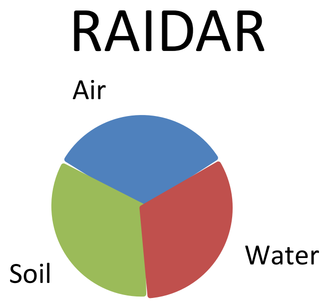
Larger spheres are those for which NHANES data was available

Partitioning Release into the Environment

Models predict fate depending upon release profile (Level III Fugacity Model)

Release profile can be chemical-specific, class-specific, or default depending on data





Estimated behavior/consumption can in turn yield human and ecological prediction

Partitioning Release into the Environment

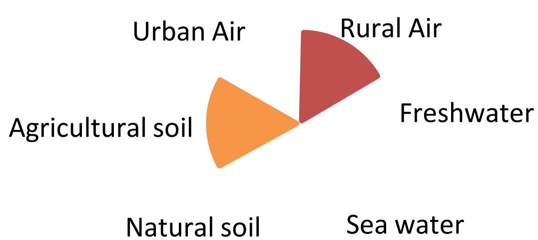
If we have the data then we would use it, but we don't

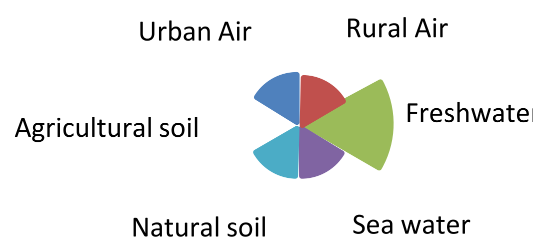
Assuming an "average" release profile

Food-use Pesticide

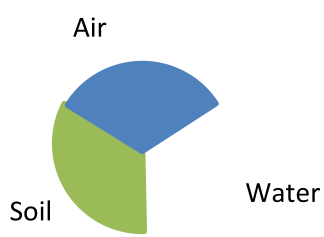
TSCA / Industrial

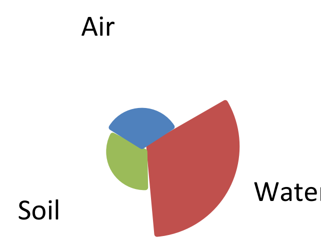
USEtox





RAIDAR

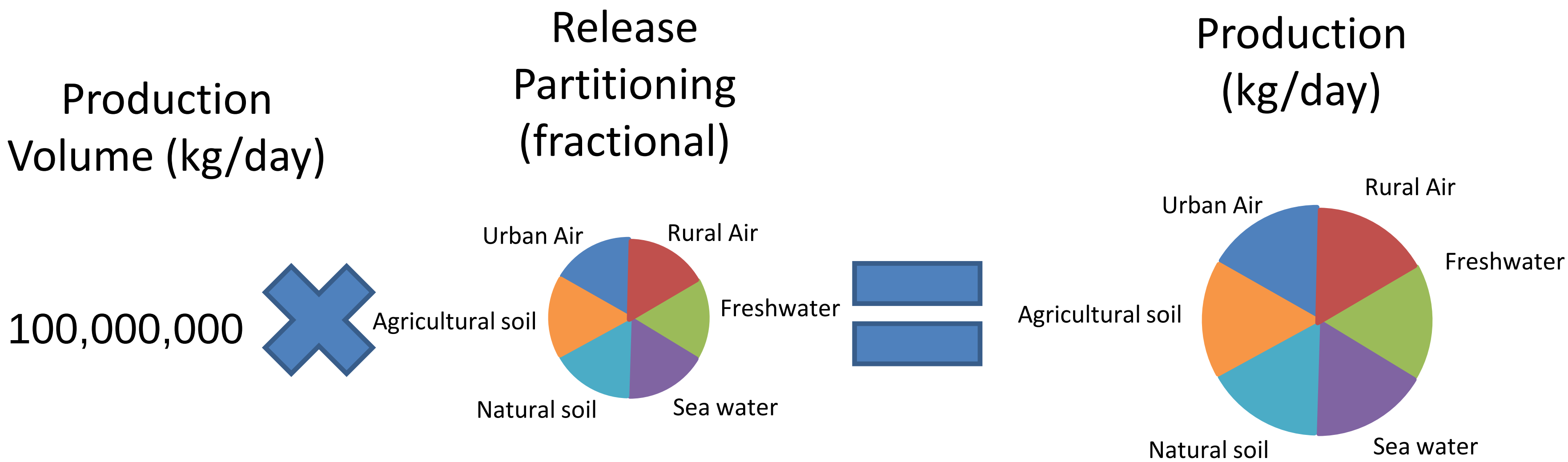




Production Volume is an Overall Multiplier

Using EPA Toxic Substances Control Act (TSCA) Chemical Data Reporting (CDR) Rule (Formerly Inventory Update Reporting – IUR) data for production volumes

Crop Protection Research Institute has data on many pesticides (which are heavily favored in ToxCast Phase I) although the data is old (2002)

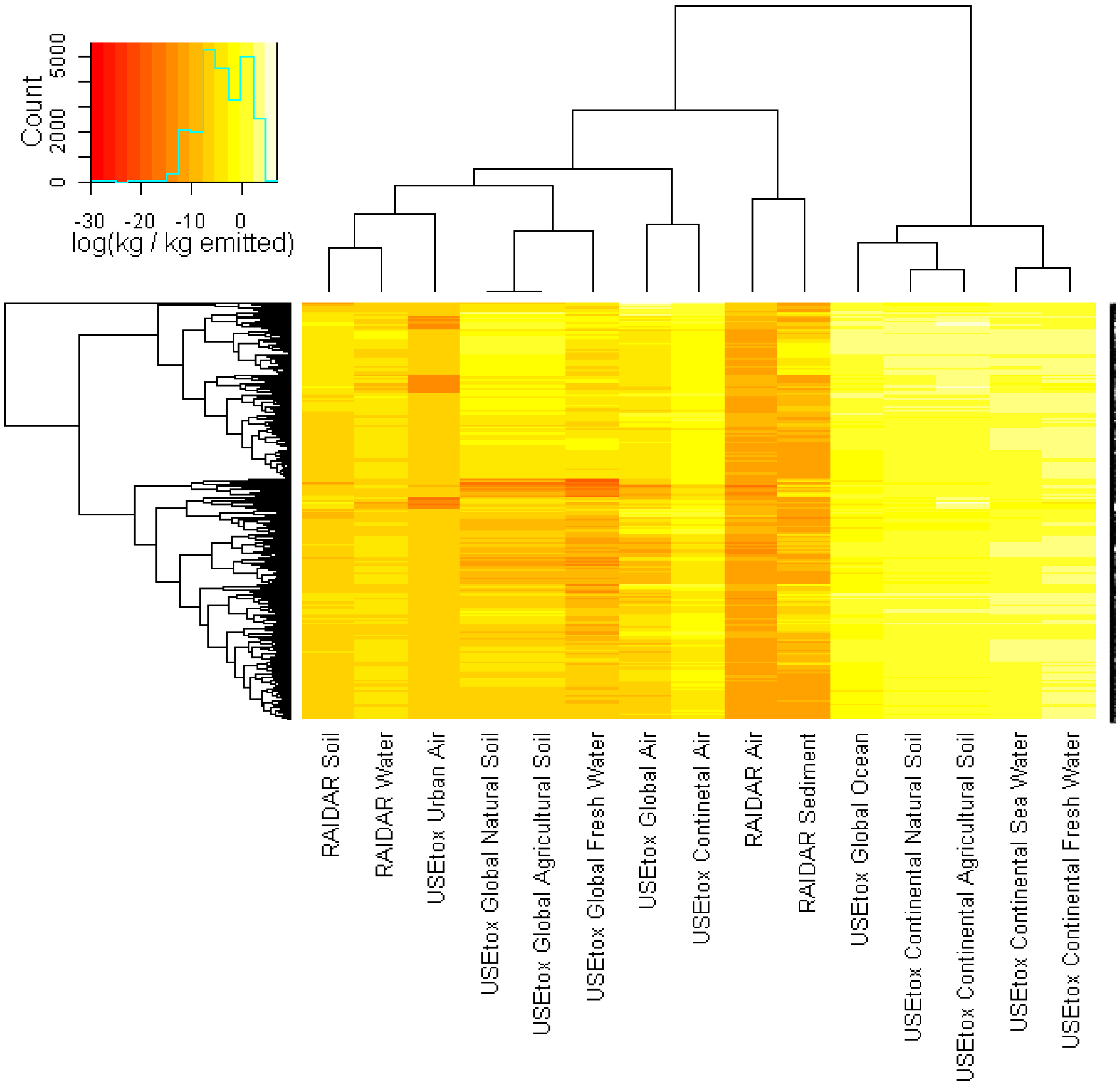


High Throughput Fate Predictions

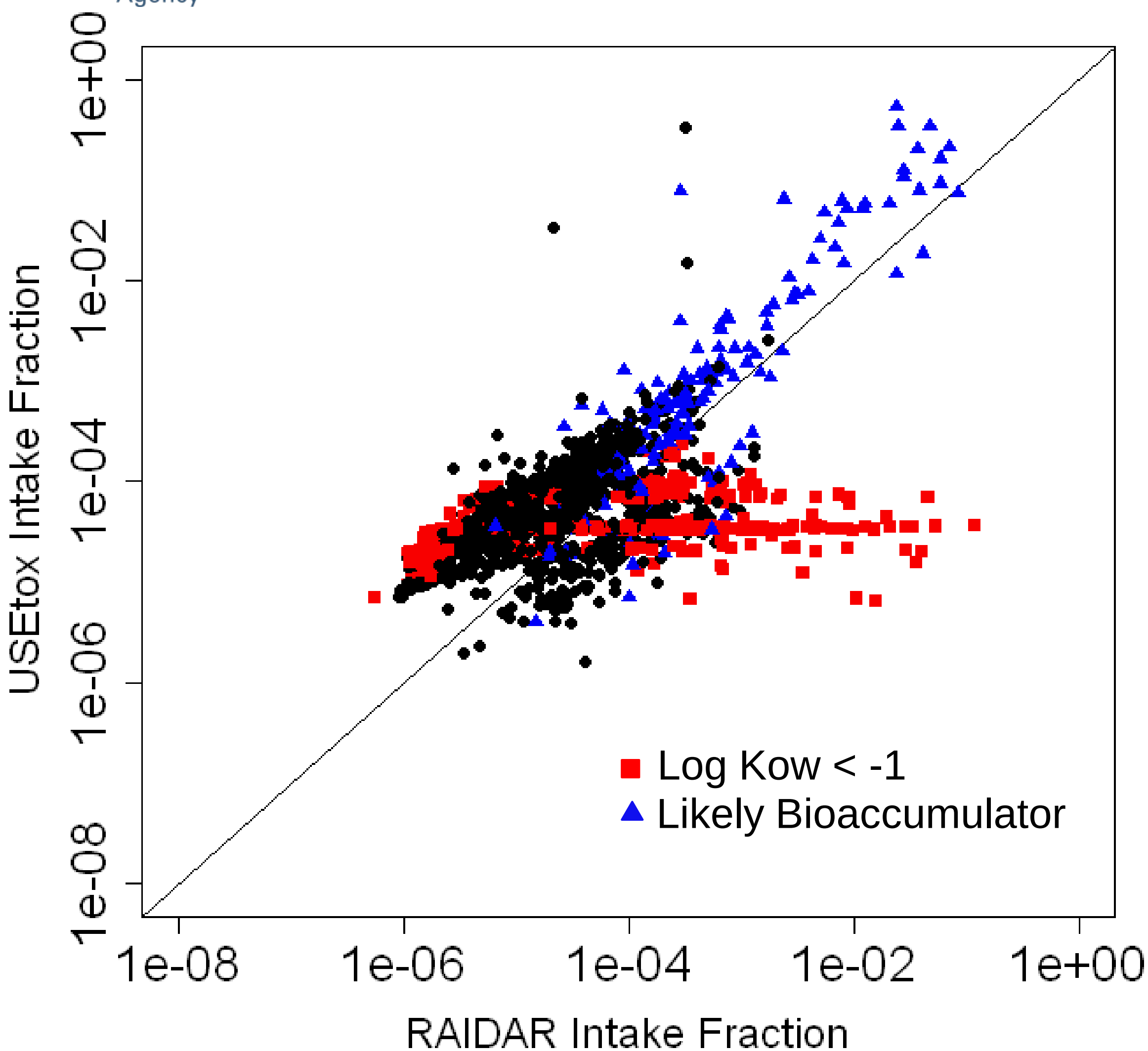
Clustering 1678 chemicals by the media into which they partition most

Could infer behavior of understudied chemicals from similar, well-known counterparts – “fate read-across”

Fate predictions not terribly consistent



Population Exposure from Environmental Media



Both models assume exposure scenarios that relate environmental media to food and inhalation exposure

Can calculate intake fraction (population exposure in kg per kg emitted)

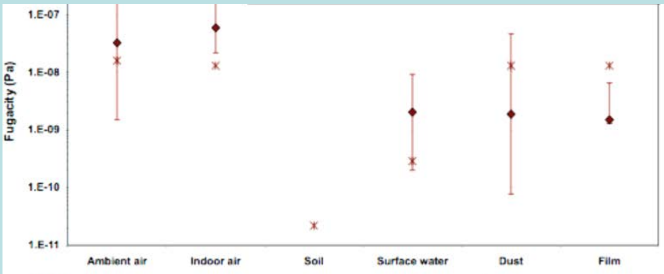
General agreement for most chemicals – putative bioaccumulators predicted to be highest

Issue with accumulation in plants causes larger predictions for RAIDAR in some cases

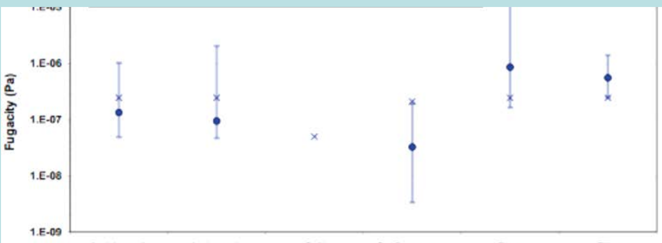
Literature Ground-truthing Efforts

United States Environmental Protection Agency

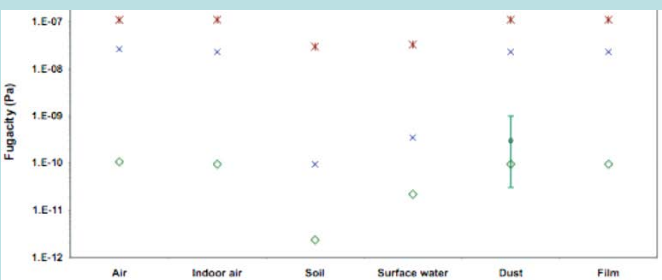
Chlorpyrifos



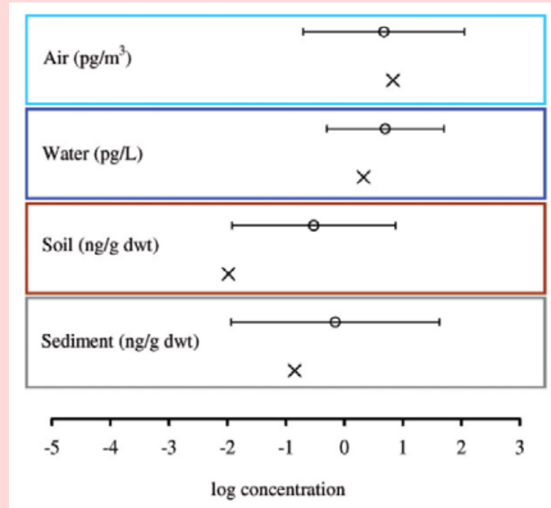
Diazinon



Dimethoate, Malathion, Oxydemeton-methyl

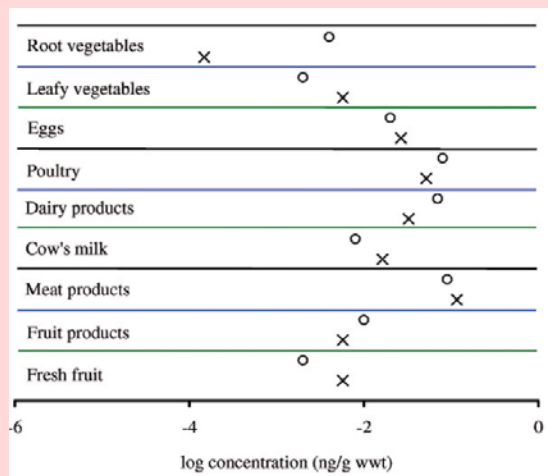


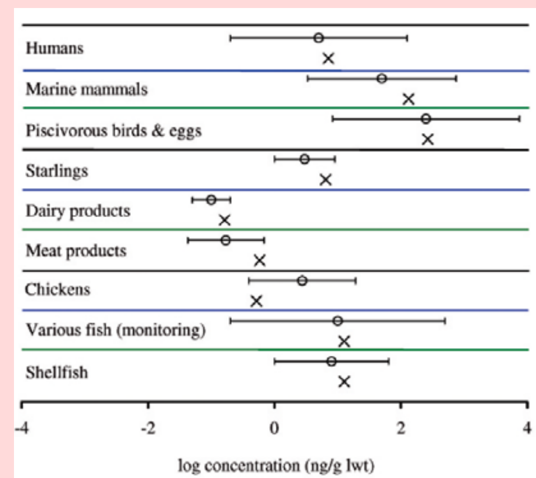
McKone et al. *(2007)*



Cowan-Ellsbury et al. *(2009)*

PBDE99





o Measured

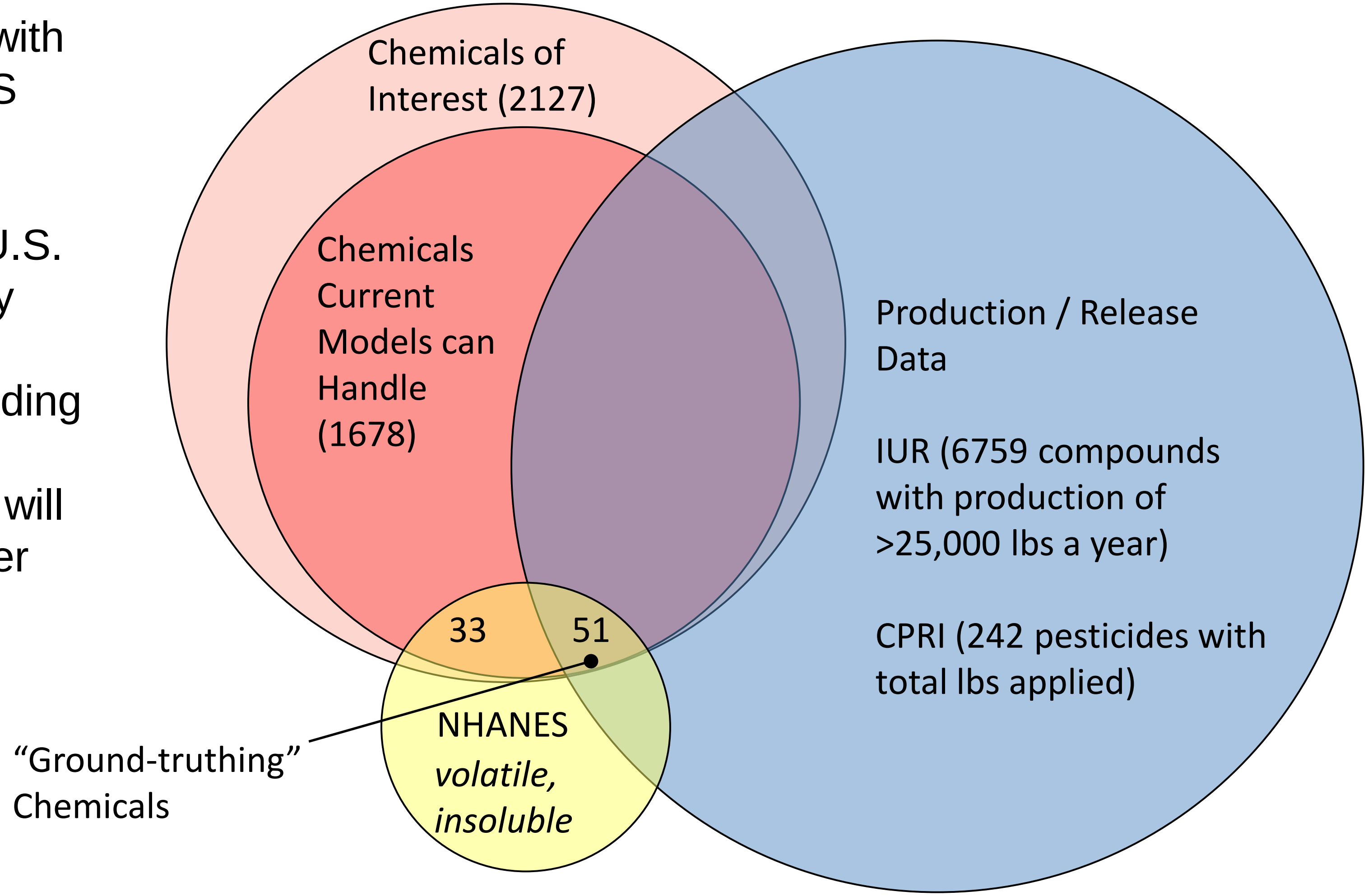
x Model Predicted

Data Availability for Model Predictions and Ground-truthing

Ground-truth with CDC NHANES urine data

Focusing on U.S. median initially

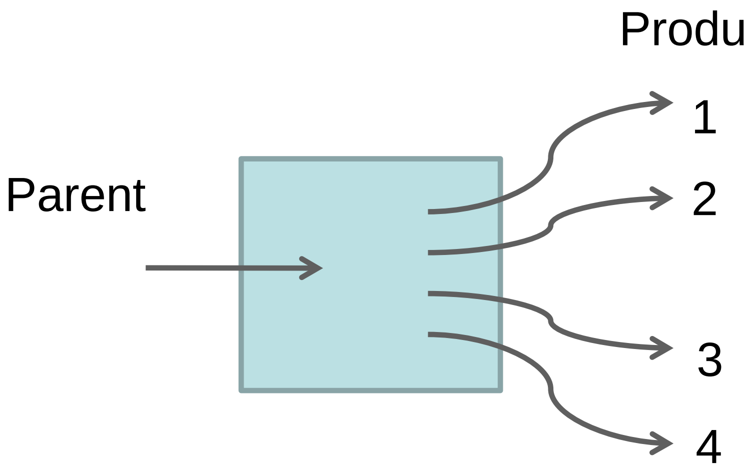
Capable of adding population variability, but will need consumer use models



Linking NHANES Urine Data and Exposure

Steady-state assumption

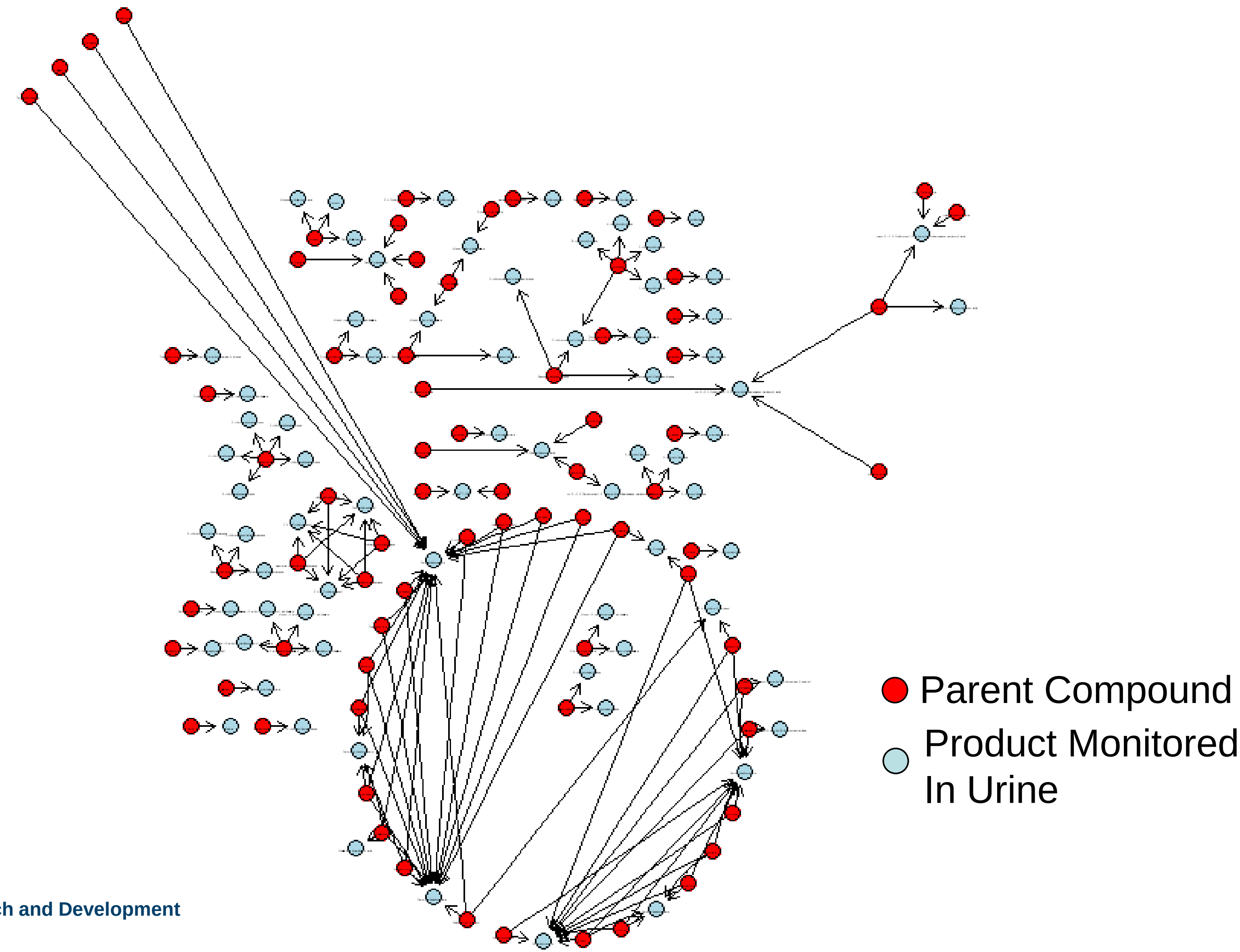
$$(\mathrm{mg/kg/day})_i = \frac{1}{70\ \mathrm{kg}} \frac{\mathrm{mg}_i}{\mathrm{g}_{\mathrm{creatine}}} * \frac{\mathrm{g}_{\mathrm{creatine}}}{\mathrm{day}}$$



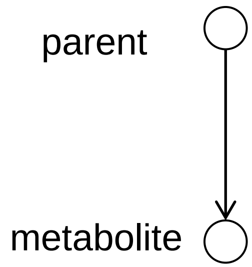
$$(\mathrm{mg/kg/day})_0 = \mathrm{MW}_0 \sum_i \frac{(\mathrm{mg/kg/day})_i}{\mathrm{MW}_i}$$

Lakind and Naiman (2008)

Mapping of NHANES Parents and Urine Products

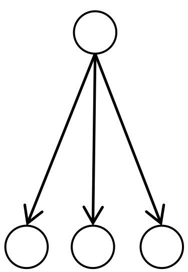


Degeneracy of an Exposure Biomarker



Unknowns: 2 (mass, bioavailability)

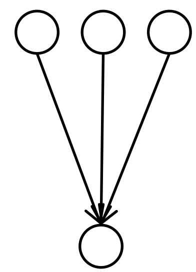
Knowns: 1 (mass)



Unknowns: 2

Unknowns: 3 (fractions),
Knowns: 1 (mass balance)

Knowns: 3 (mass)

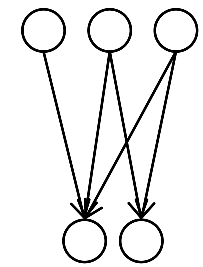


Unknowns: 6

Unknowns: 3 (fractions),
Knowns: 3 (mass balance)

Knowns: 1 (mass)

Bayesian Model for f_{ij}



The real situation may be even more complicated

Further complicated by limit of detection of NHANES chemicals – many chemicals that are checked for are below the LoD

However, we still can predict that N parent exposure are related to P=f*N urine products, and many f_{ij} are zero,

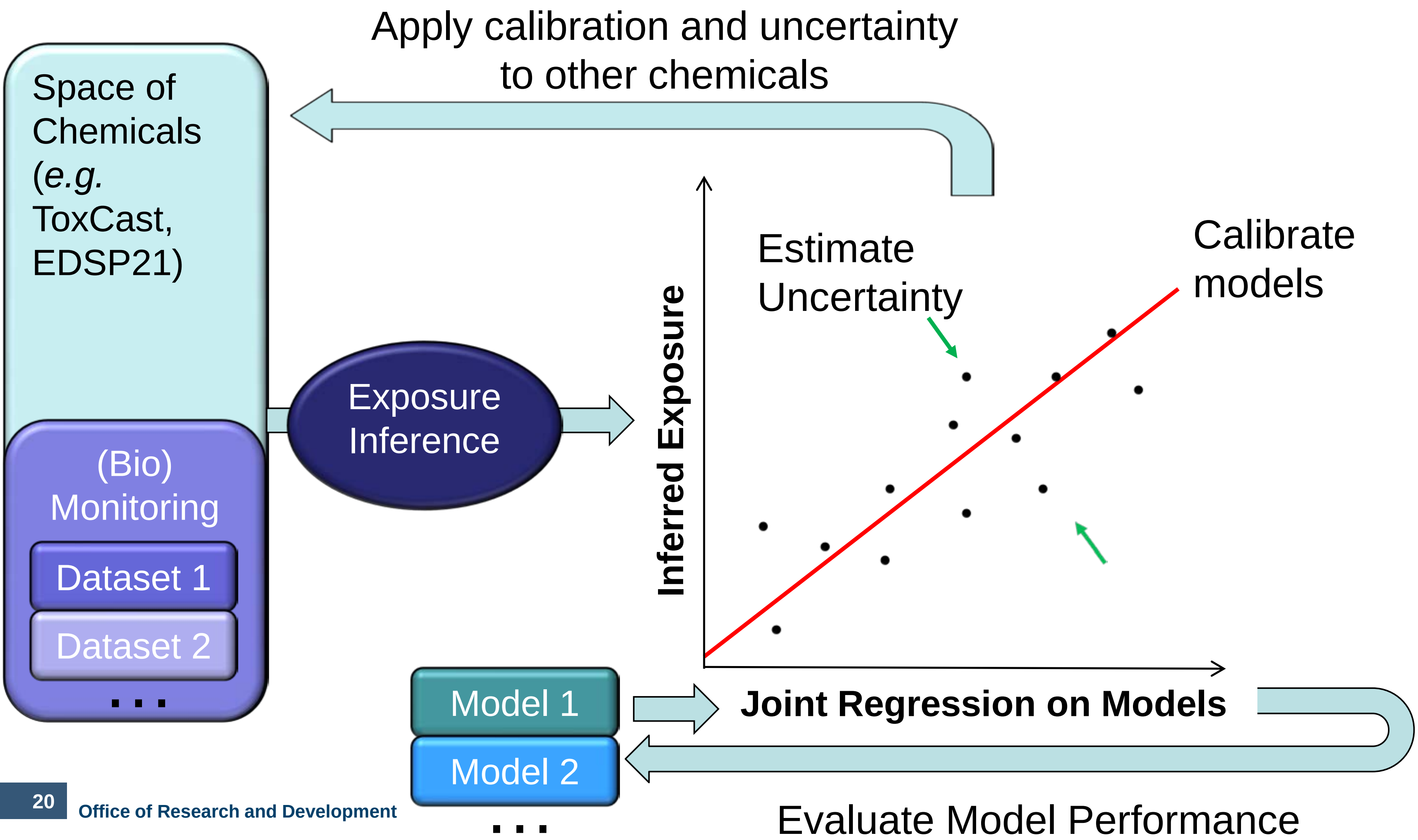
Use MCMC to explore range of possible parent predictions

Also incorporate uncertainty in production volume and use all quantiles of NHANES data

Unknown fraction f_{ij} for each urine product j due to parent i:

$$(\mathrm{mg/kg/day})_j = \mathrm{MW}_j \sum_i f_{ij} \frac{(\mathrm{mg/kg/day})_i}{\mathrm{MW}_i}$$

Framework for High Throughput Exposure Screening

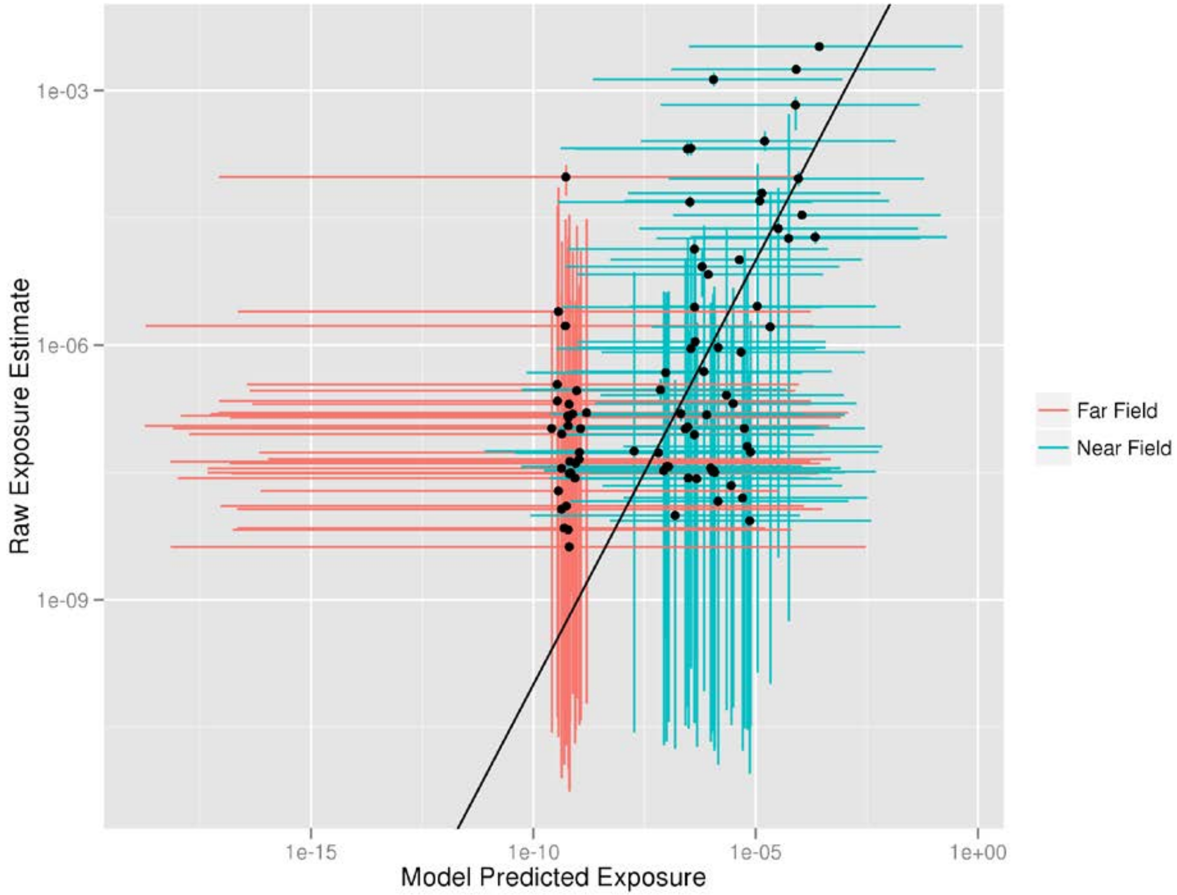


Calibrate ExpoCast Predictions to CDC NHANES Data

$$Y \sim b_1 + b_2 * N + m_2 \log(vu) + m_3 \log(vr)$$

Comparison between model predictions and biomonitoring data indicates correlation

Indoor/consumer use is critical: Compounds with near-field applications more than 100x greater



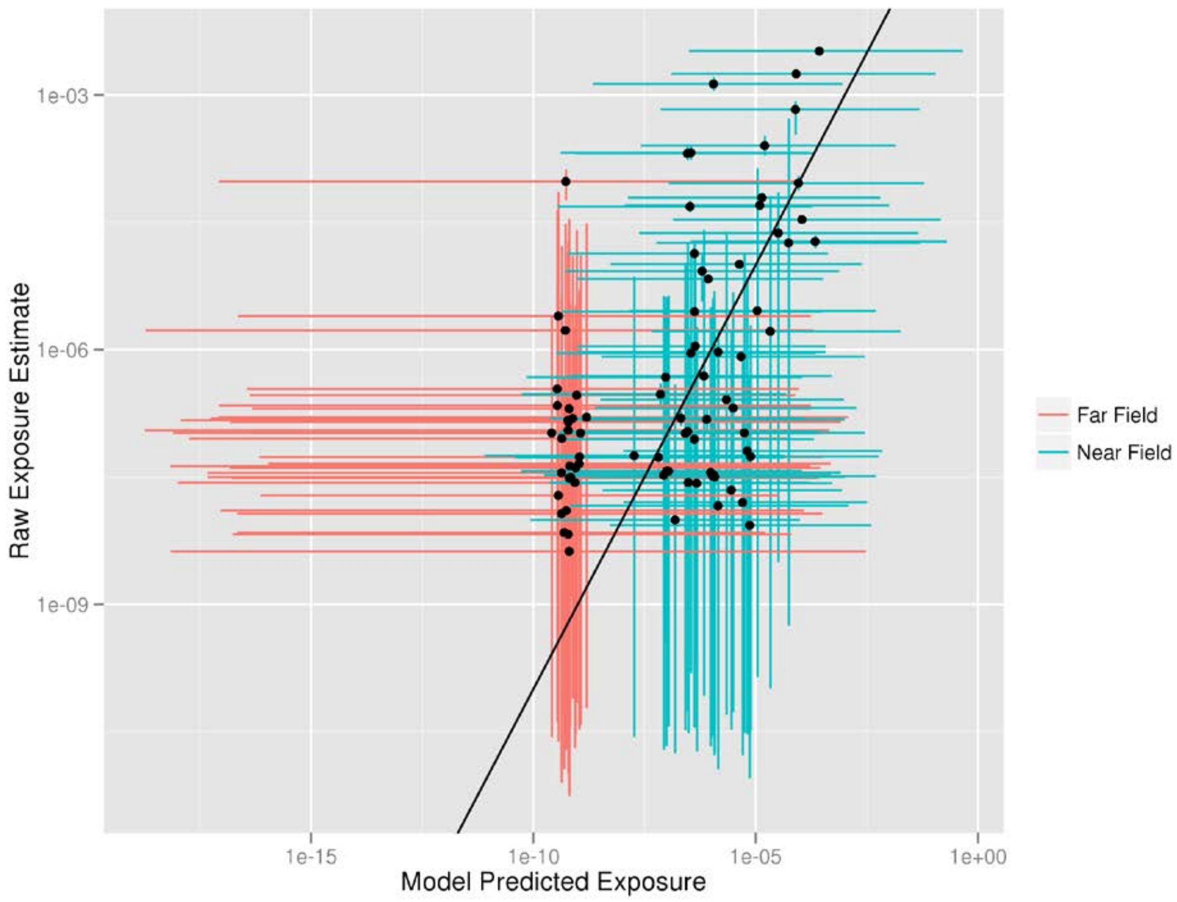
Use Categories from ACToR

$$Y \sim b_1 + b_2 * N + m_2 \log(vu) + m_3 \log(vr)$$

The sources for various chemical data were assigned to various use categories.

Chemicals from multiple sources were assigned to multiple categories.

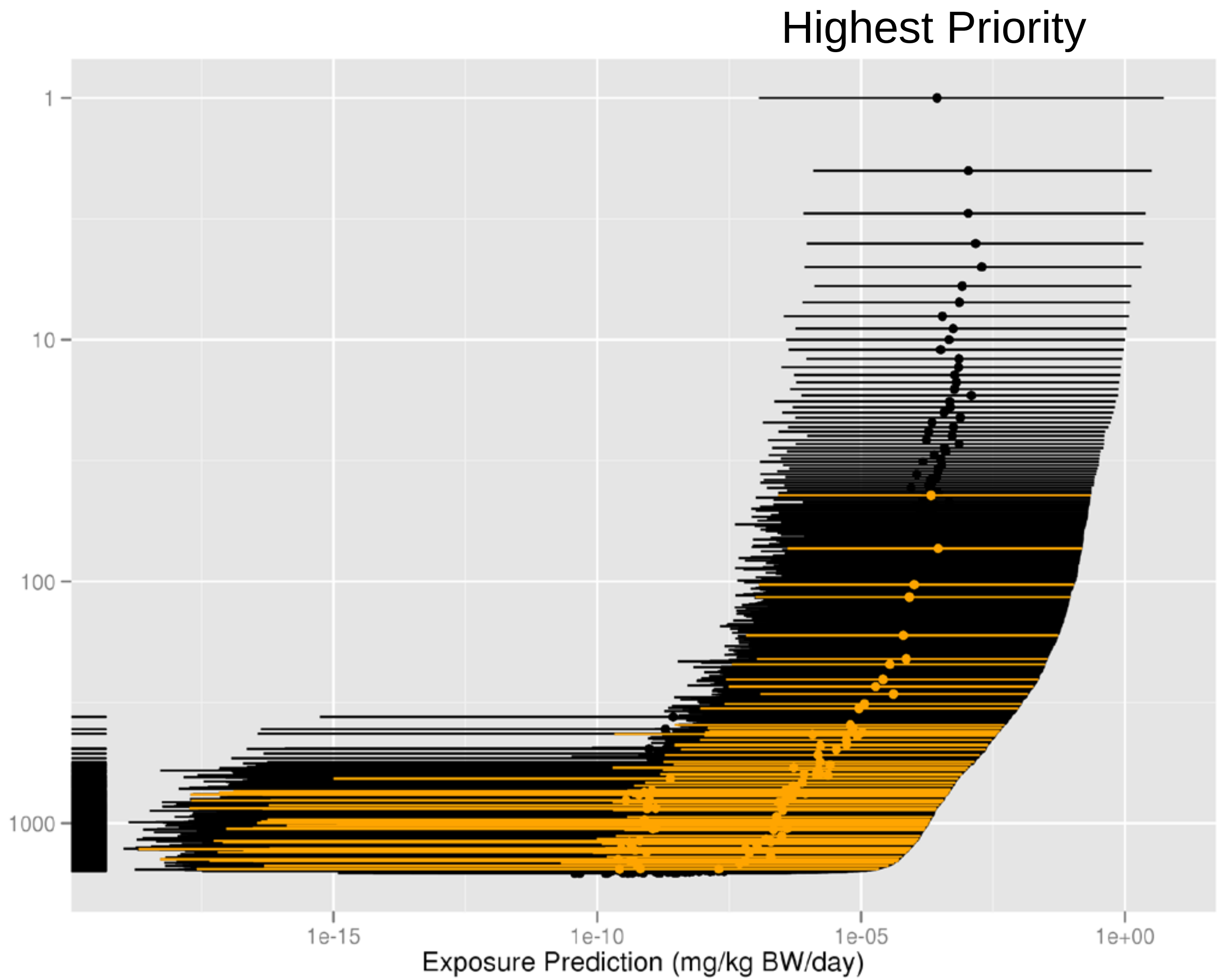
Four categories – personal care products, consumer use, fragrance, and food additive – were aggregated into a single “near field” indicator variable.



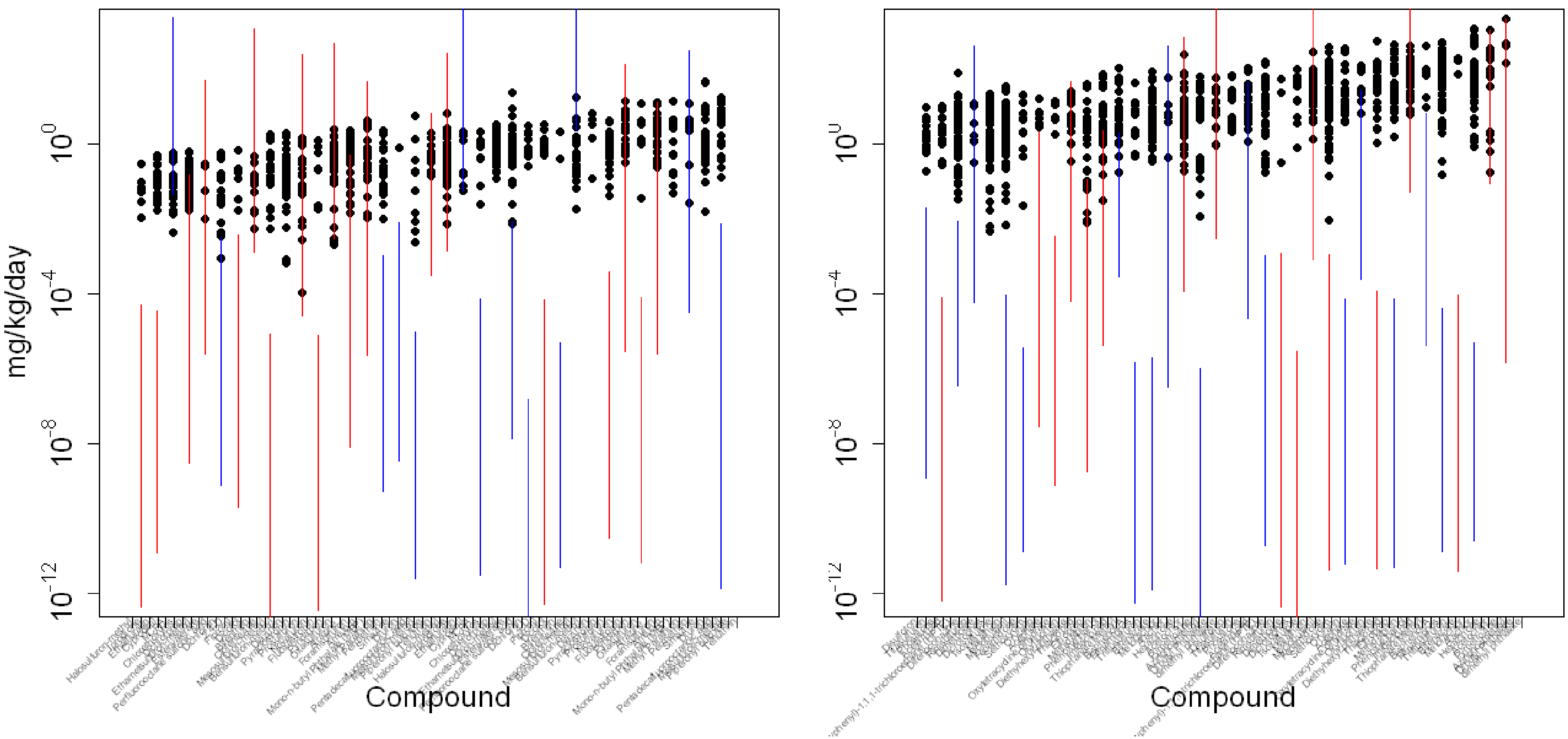
Exposure Prioritization from ExpoCast

Uncertainty of prediction indicated by the horizontal confidence interval from the empirical calibration to the NHANES data

Horizontal dotted line indicates the fiftieth percentile rank and the vertical dotted line indicates the cutoff between overlapping top-half and lower half confidence intervals



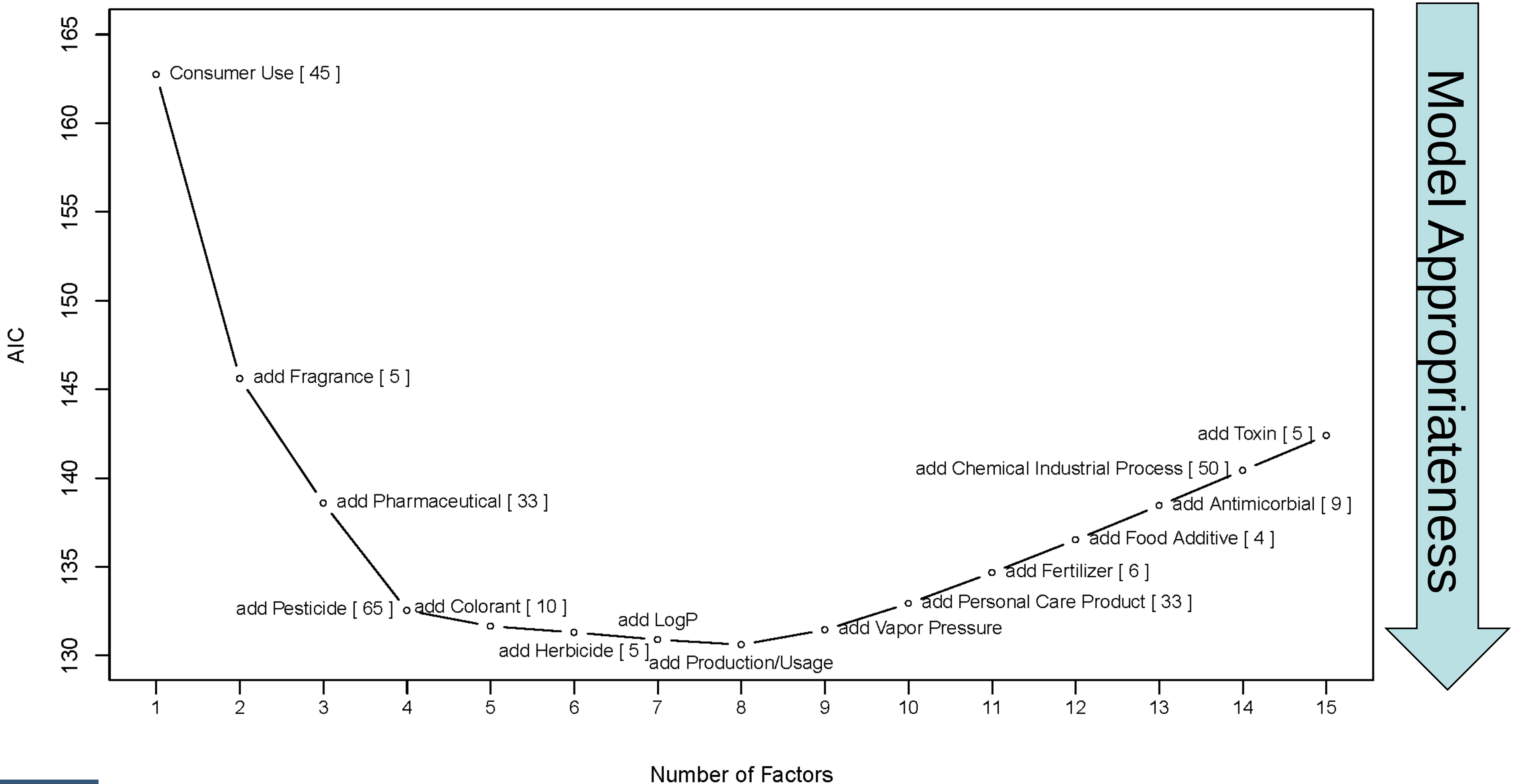
Wetmore et al. (2011)
ToxCast Oral Equivalents



Lines indicate ExpoCast prediction 95% CI
indoor/consumer use in red, blue otherwise

Statistical Near Field Model

Further investigating near field use determinants using expanded information



Conclusions

- ExpoCast can use environmental fate and transport models to make high-throughput exposure predictions
  - These prioritizations have been compared with CDC NHANES ground truth
  - This biomonitoring data gives empirical calibration and estimate of uncertainty
- Indoor/consumer use is a primary determinant
- Next steps:
  - HT models for exposure from consumer use and indoor environment
  - Use and evaluate these models as additional HT exposure assays
  - ORISE Postdoc position for high throughput modeling of nearfield indirect exposure (*e.g.* flooring, furniture)

U.S. E.P.A. Office of Research and Development ExpoCast Team

NCCT
Elaine Cohen-Hubal
David Dix
Alicia Frame
Sumit Gangwal
Richard Judson
Robert Kavlock
Thomas Knudsen
Stephen Little
Shad Mosher
James Rabinowitz
David Reif
Woody Setzer
Amber Wang

NERL
Peter Eghehy
Kathie Dionisio
Dan Vallero

University of Toronto at Scarborough
Jon Arnot

University of Michigan
Olivier Jolliet

USDA FSIS
Jade Mitchell-Blackwood