



# Playing with data at EPA— ToxCast, ExpoCast, HTTK, and the Exposome

*John Wambaugh*  
*National Center for Computational Toxicology*  
*Office of Research and Development*  
*U.S. Environmental Protection Agency*

**June 14, 2016**

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

# Introduction

The timely characterization of the human and ecological risk posed by thousands of existing and emerging commercial chemicals is a critical challenge facing EPA in its mission to protect public health and the environment

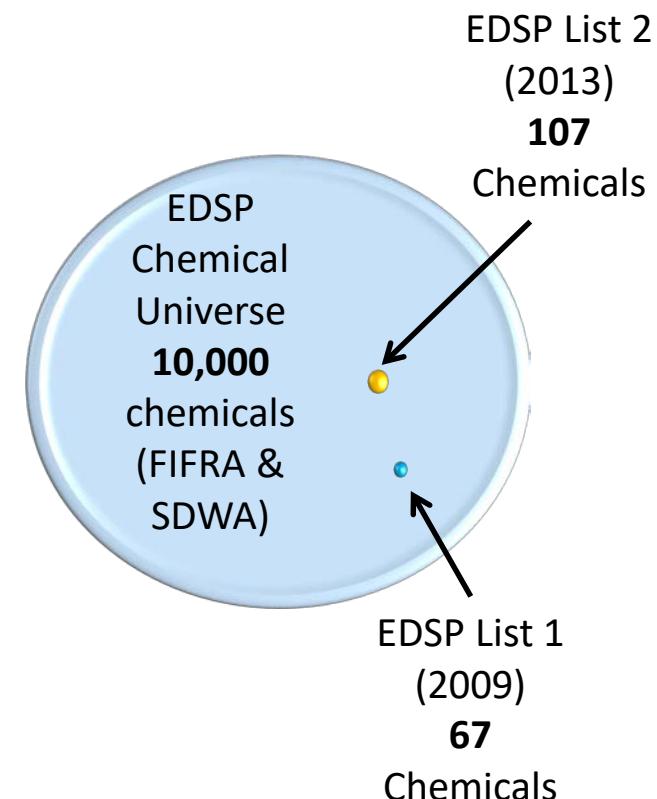


November 29, 2014

# Scale of the Problem

- Park *et al.* (2012): At least 3221 chemicals in humans, many appear to be exogenous

Endocrine Disruptor Screening Program (EDSP) Chemical List	Number of Compounds
Conventional Active Ingredients	838
Antimicrobial Active Ingredients	324
Biological Pesticide Active Ingredients	287
Non Food Use Inert Ingredients	2,211
Food Use Inert Ingredients	1,536
Fragrances used as Inert Ingredients	1,529
Safe Drinking Water Act Chemicals	3,616
<b>TOTAL</b>	<b>10,341</b>



So far 67 chemicals have completed testing and an additional 107 are being tested

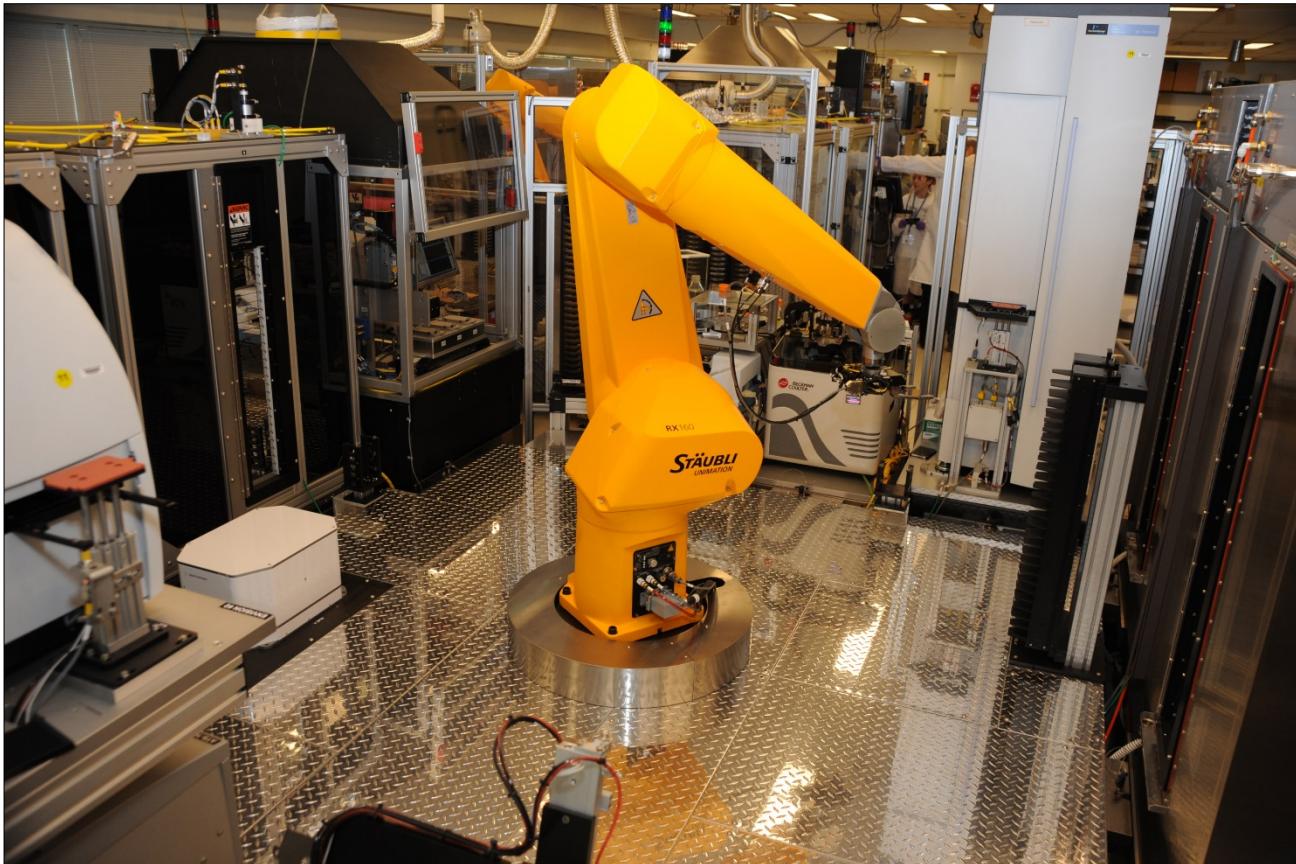
# High-Throughput Bioactivity



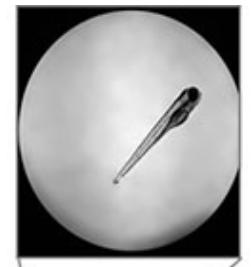
- **Tox21:** Examining >10,000 chemicals using ~50 assays intended to identify interactions with biological pathways (Schmidt, 2009)

- **EPA Toxicity Forecaster (ToxCast):**

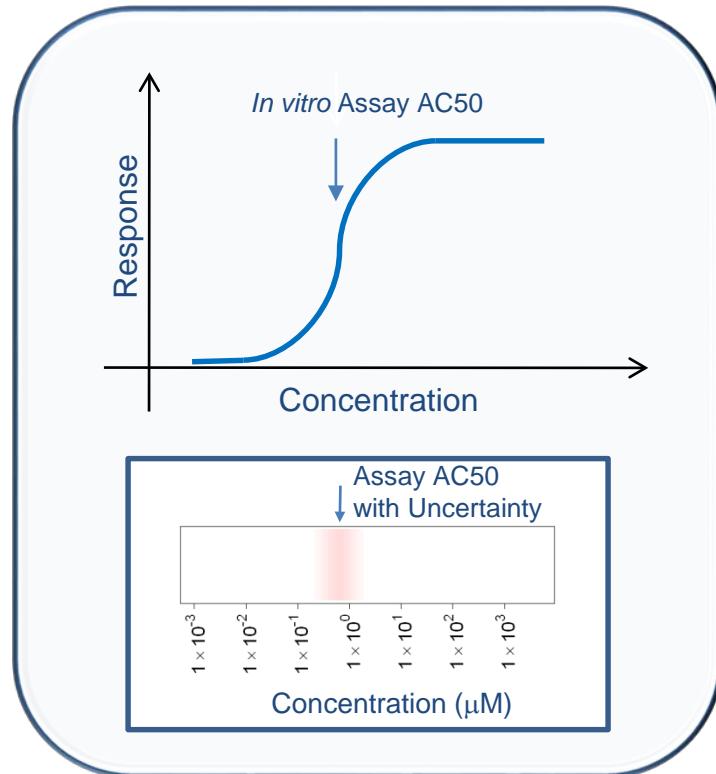
For a subset (>3000) of Tox21 chemicals run >1000 additional assay endpoints (Judson et al., 2010)



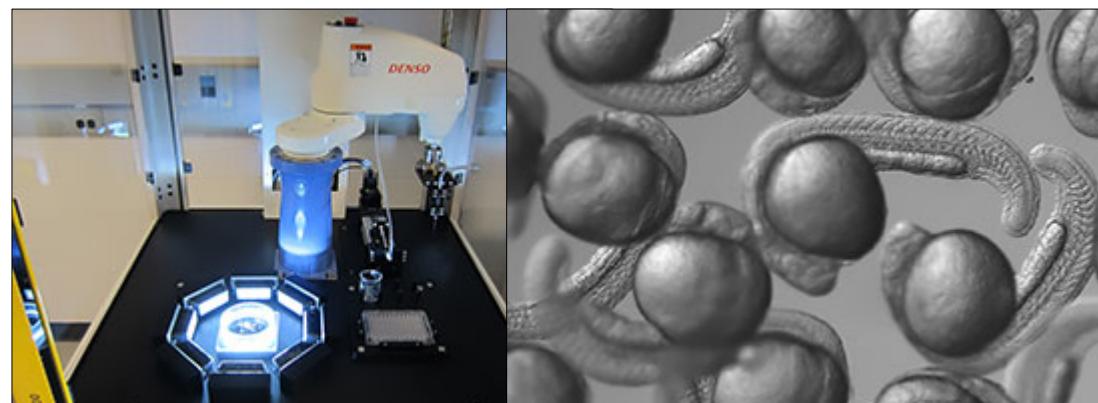
# High-Throughput Bioactivity



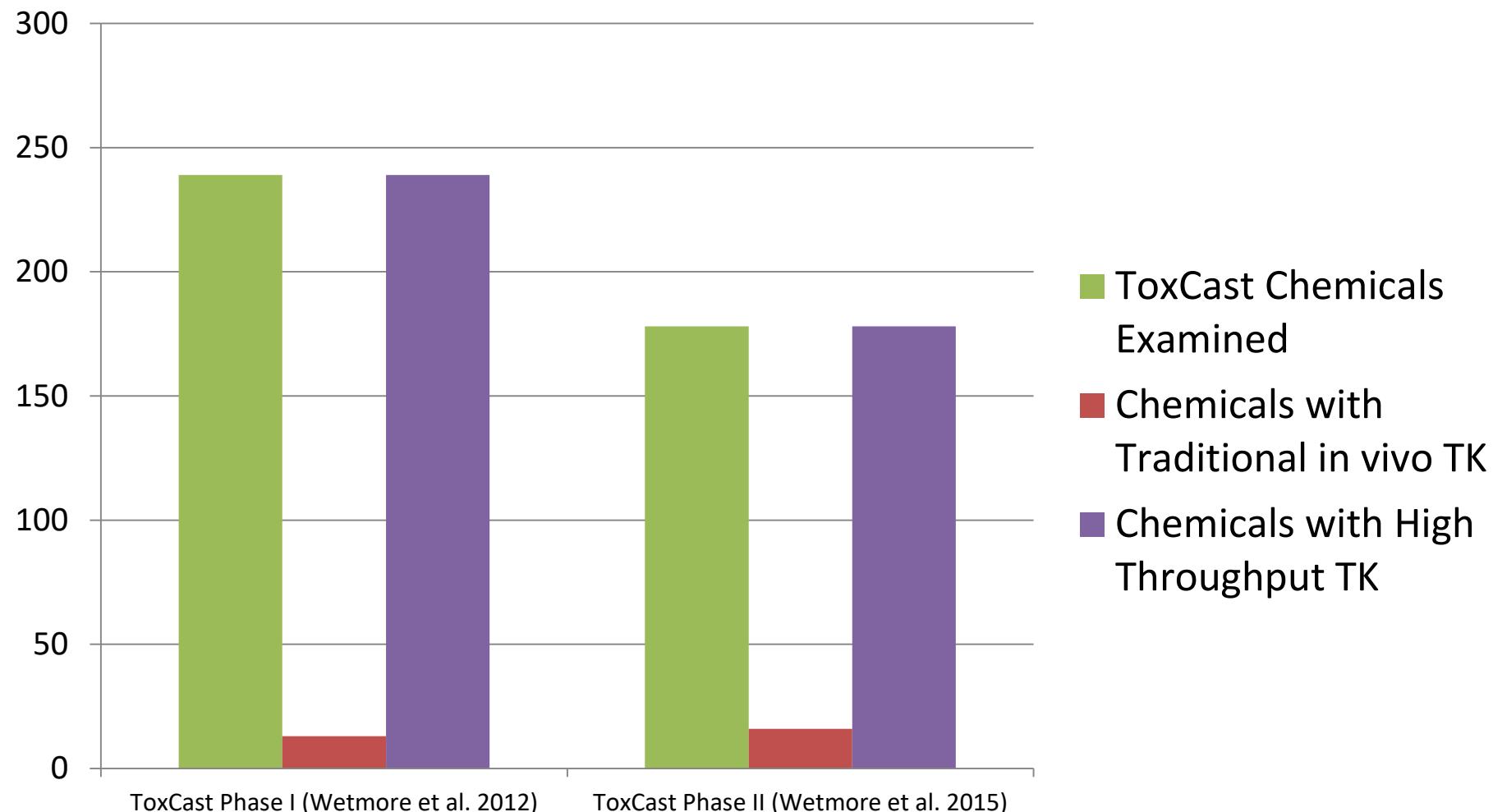
- Most assays conducted in dose-response format (identify 50% activity concentration – AC50 – and efficacy if data described by a Hill function)



- All data is made public:  
<http://actor.epa.gov/dashboard/>
- New datasets being added continuously (for example, 1060 chemicals tested by Truong et al., 2014 zebrafish assay)



# The Need for *In Vitro* Toxicokinetics



- Studies like Wetmore et al. (2012), addressed the need for TK data using *in vitro* methods

# High-Throughput Toxicokinetics

cran - Package httk X

https://cran.r-project.org/web/packages/httk/index.html

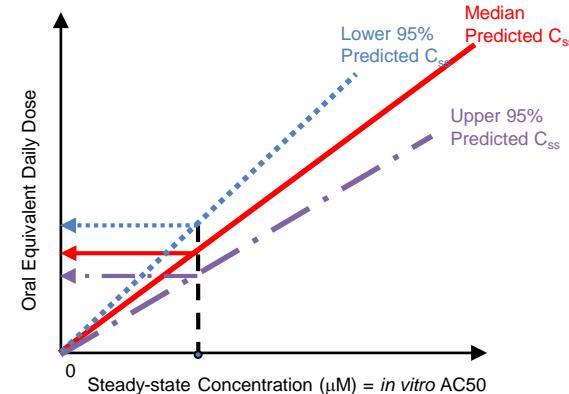
Bookmarks DSSToxViewer Journal Selector, tar... Journal / Author Name Selection of GC-MS... Other bookmarks

**httk: High-Throughput Toxicokinetics**

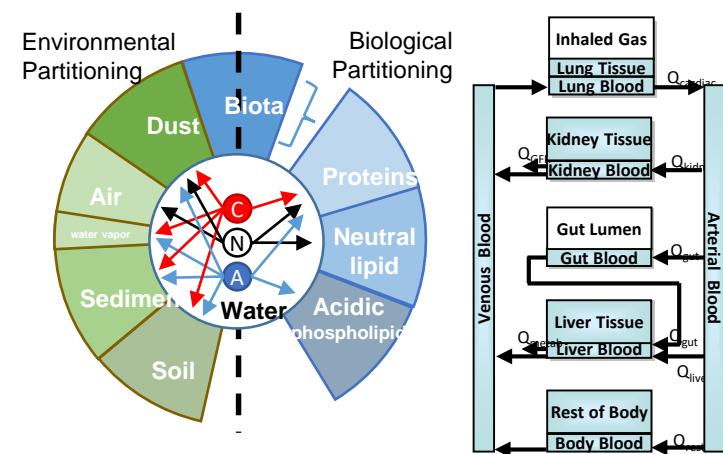
Functions and data tables for simulation and statistical analysis of chemical toxicokinetics ("TK") using data obtained from relatively high throughput, *in vitro* studies. Both physiologically-based ("PBTK") and empirical (e.g., one compartment) "TK" models can be parameterized for several hundred chemicals and multiple species. These models are solved efficiently, often using compiled (C-based) code. A Monte Carlo sampler is included for simulating biological variability and measurement limitations. Functions are also provided for exporting "PBTK" models to "SBML" and "JARNAC" for use with other simulation software. These functions and data provide a set of tools for *in vitro*-*in vivo* extrapolation ("IVIVE") of high throughput screening data (e.g., ToxCast) to real-world exposures via reverse dosimetry (also known as "RTK").

Version: 1.4  
Depends: R ( $\geq$  2.10)  
Imports: deSolve, msm  
Suggests: ggplot2  
Published: 2016-02-03  
Author: John Wambaugh and Robert Pearce, Schmitt method implementation by Jimena Davis, dynamic model adapted from code by R. Woodrow Setzer, Rabbit parameters from Nisha Sipes  
Maintainer: John Wambaugh <wambaugh.john at epa.gov>  
License: GPL-3

**"httk" R Package**  
543 Chemicals to date  
Lead programmer Robert Pearce  
Pearce *et al.* Journal Statistical Software (in press)



Open source *In Vitro-In Vivo* Extrapolation and Physiological-based Toxicokinetics



<https://cran.r-project.org/web/packages/httk/>

Can access from the R GUI: "Packages" then "Install Packages"

# “Exposure Science in the 21<sup>st</sup> Century”

2012 National Research Council  
report:

- New tools needed for screening and prioritization of chemicals for targeted toxicity testing
- New, focused exposure assessments or monitoring studies needed
- Better quantification of population vulnerability needed

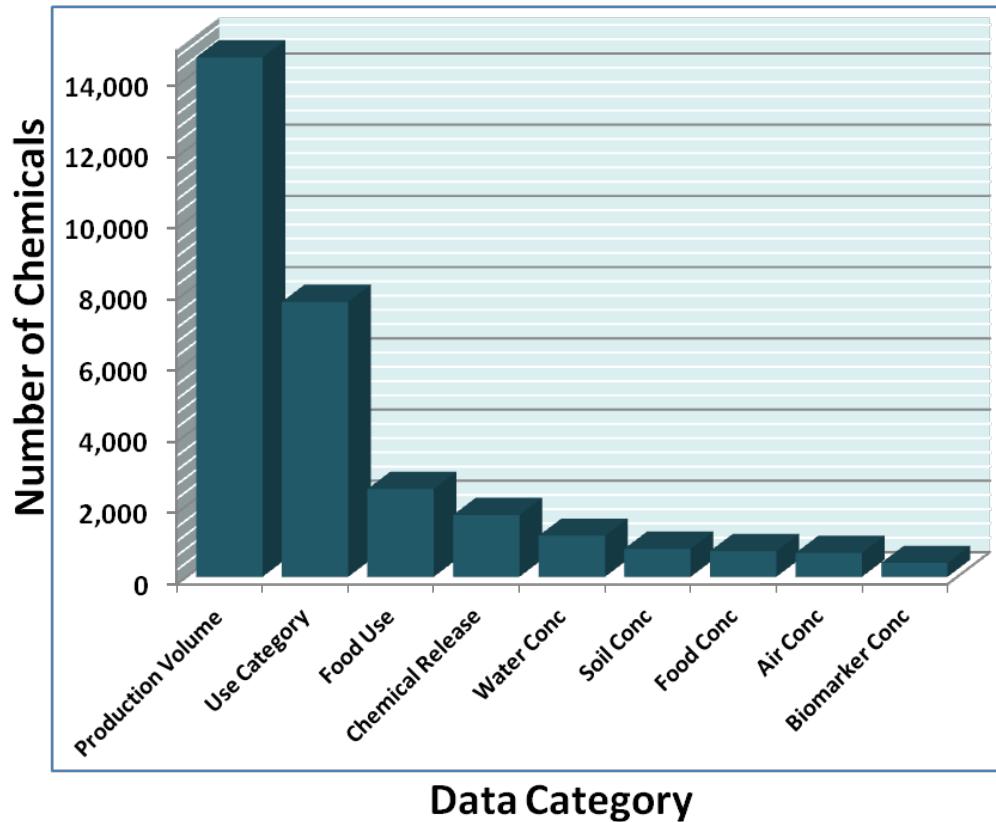
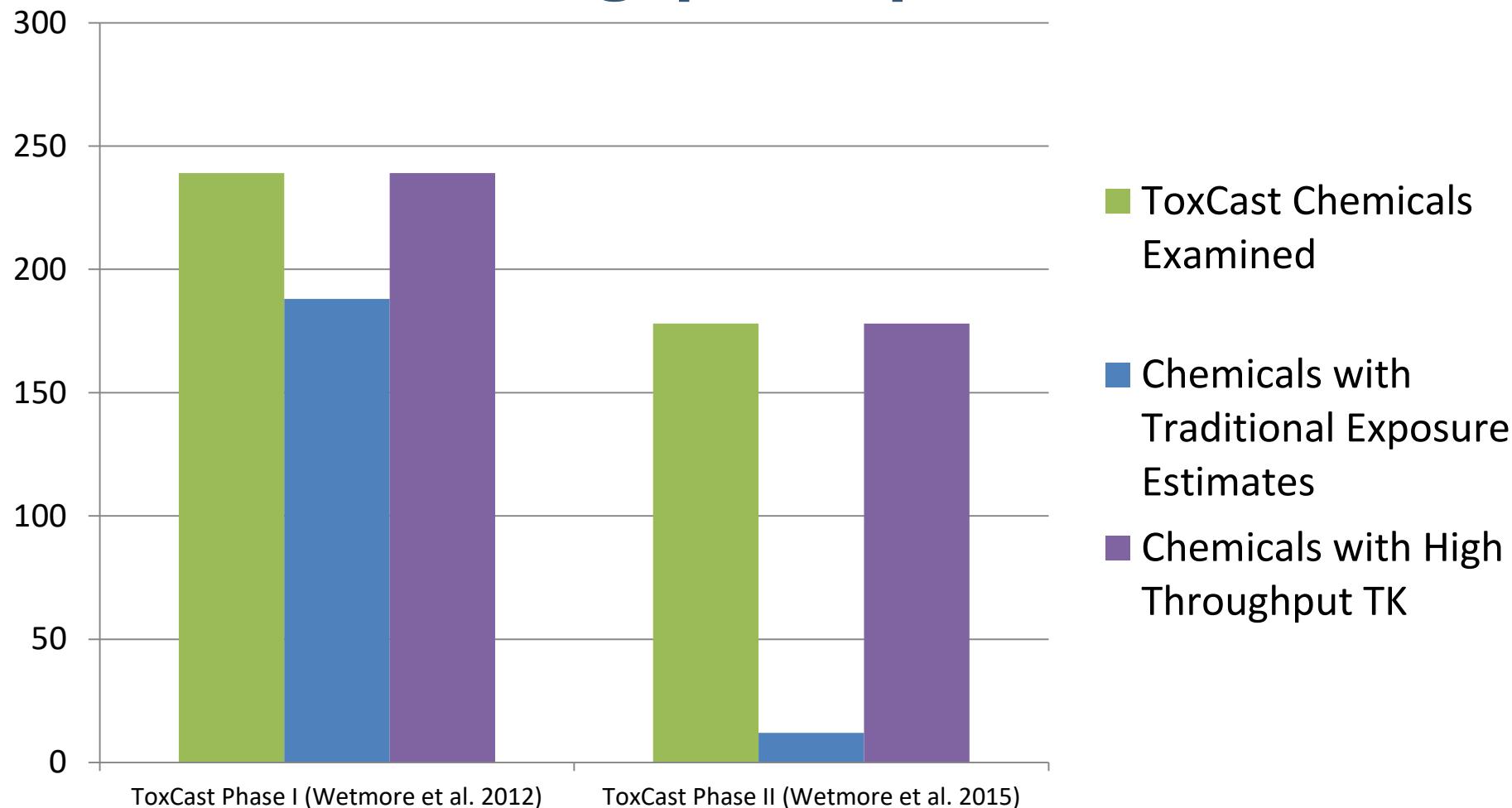


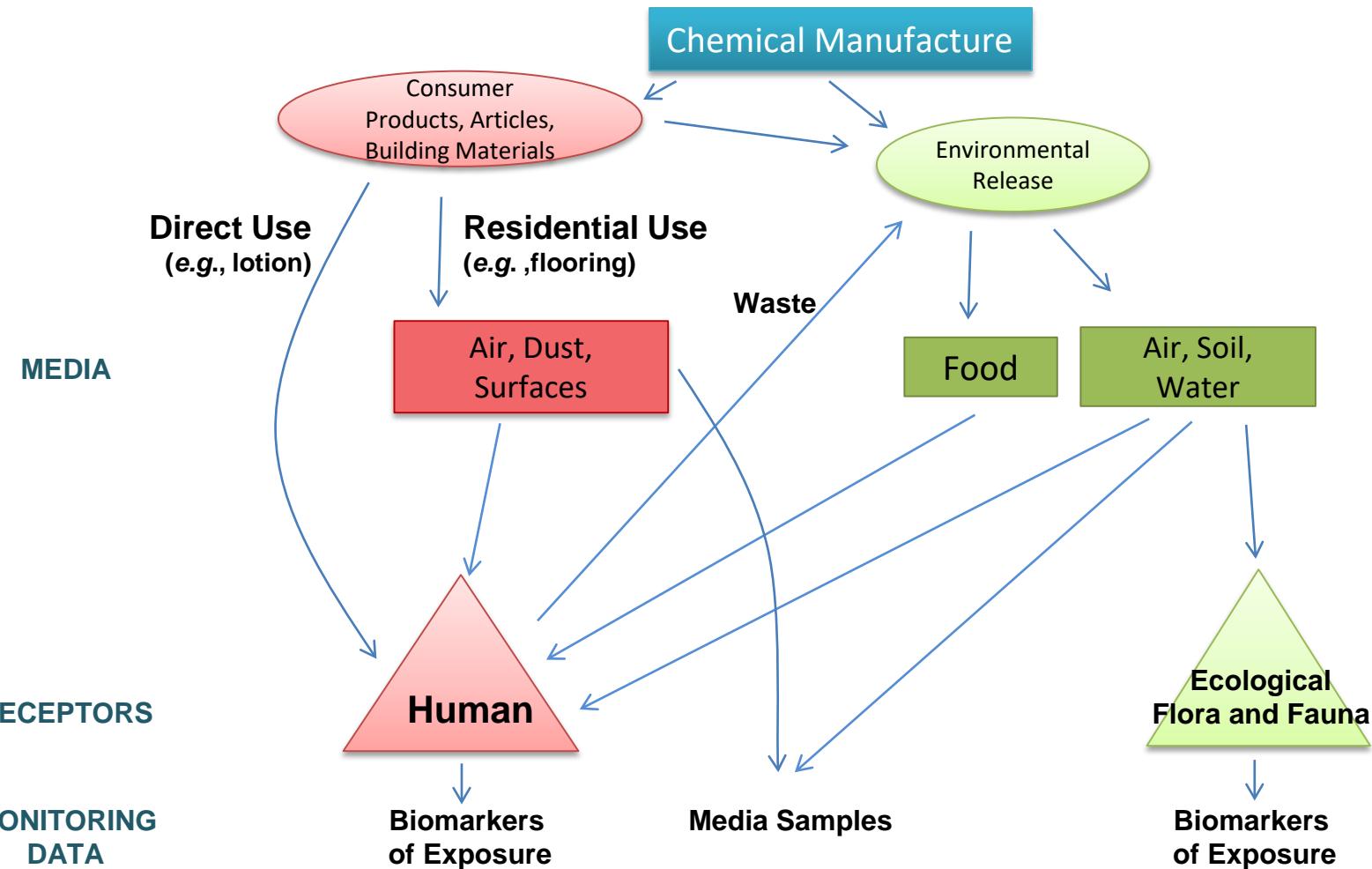
Figure from Egeghy et al. (2012),  
“The exposure data landscape for manufactured chemicals”

# The Need for High Throughput Exposure

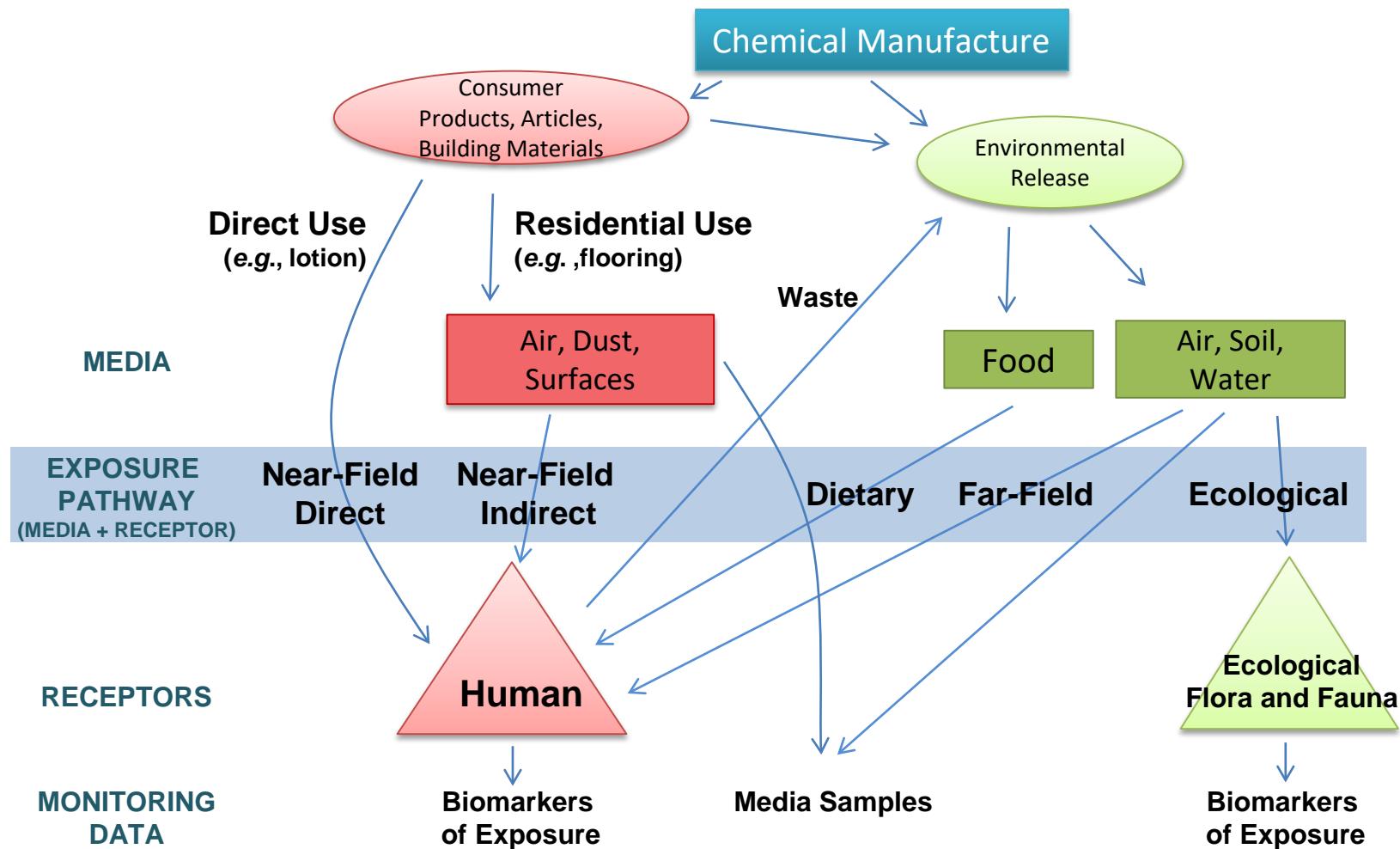


- Egeghy et al. (2012) – Most chemicals lack exposure data

# Exposure Space



# Exposure Pathways



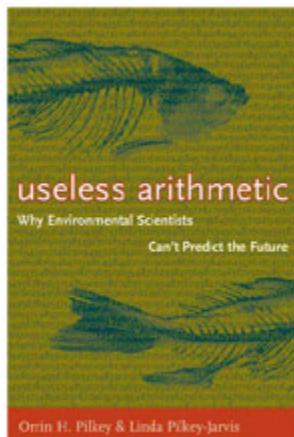
# The Exposure Event is Often Unobservable



- The exposure pathway is the actual interaction of the receptor and media, e.g. consuming potato chips
- For humans in particular, these events are often unobserved and for many reasons (including ethics and privacy) may remain unobservable
  - *Did you eat the serving size or the whole bag of potato chips?*
- **Either predict** exposure using data and models up-stream of the exposure event
- **Or infer** exposure pathways from down-stream data, especially biomarkers of exposure

# How To Deal with Uncertainty

- 1) Think probabilistically: ExpoCast evaluates model performance systematically across as many chemicals (and chemistries) as possible
- 2) Forecasts change : Today's forecast reflects the best available data today but we must accept that new data and new models will cause predictions to be revised
- 3) Look for consensus: We evaluate as many models and predictors/predictions as possible



Orrin Pilkey &  
Olinda Pilkey-Jarvis (2007)

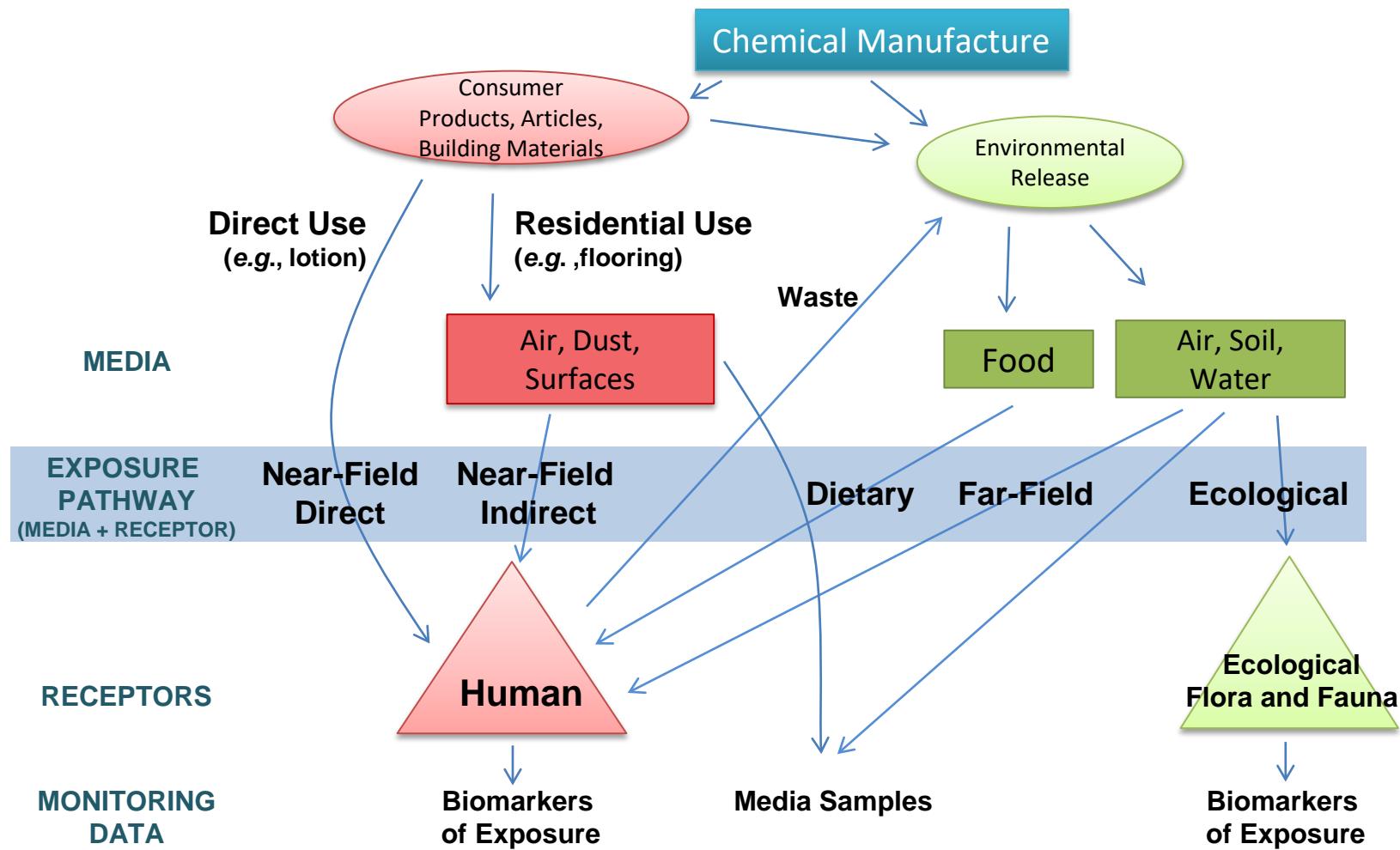
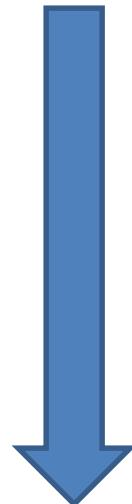


*the signal and the noise  
 and the noise and the noise  
 and the noise and the noise  
 why so many and predictions fail—but some don't the  
 and the noise and the noise and the noise  
 nate silver noise noise and the noise*

Nate Silver (2012)

# Forward Modeling of Exposure Pathways

## Data and Models



# Forward Predicting Exposure

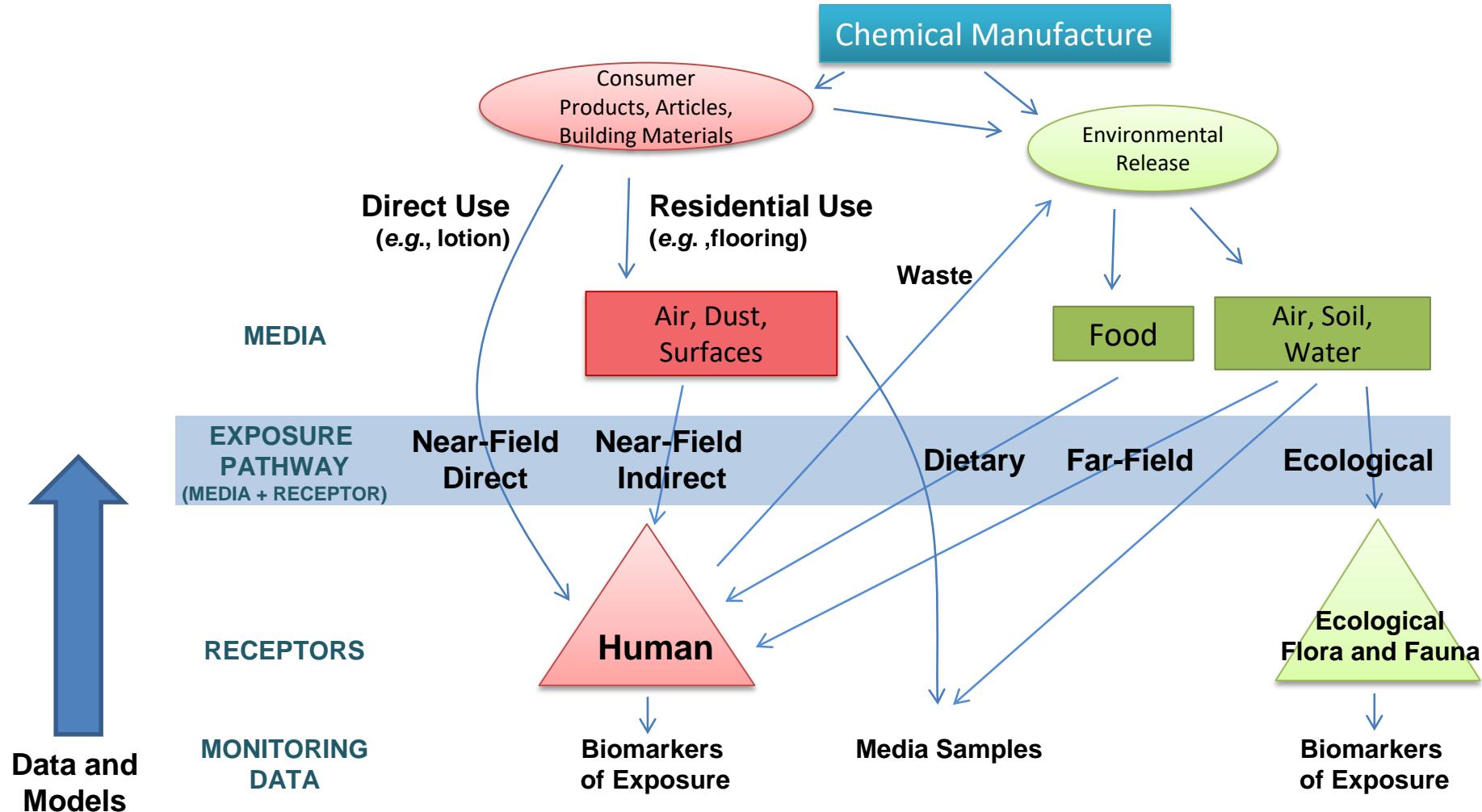
Estimated  
Value

Measured  
Value

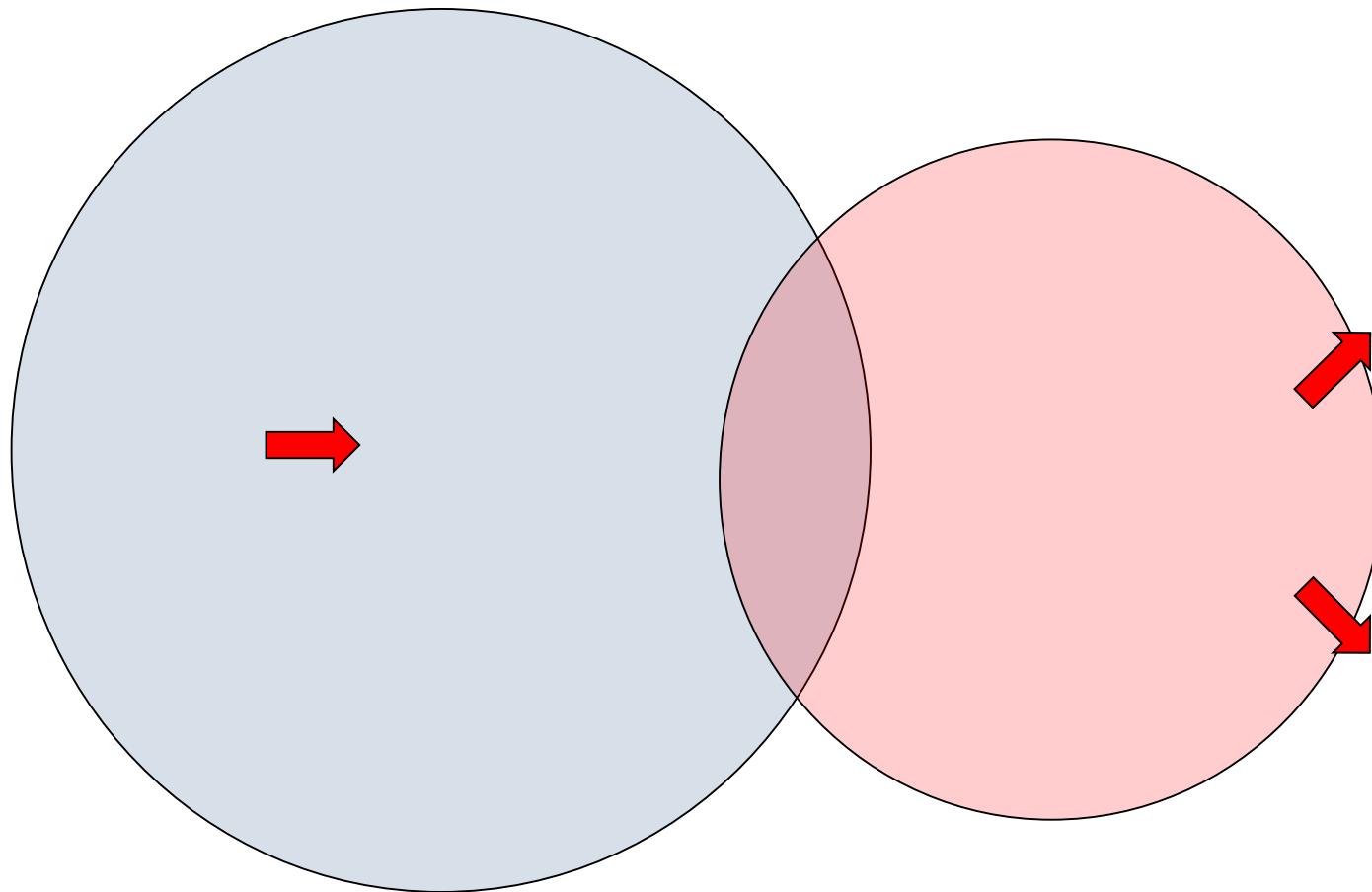
Empirical Model

Mechanistic Model

# Inference of Exposure Pathways

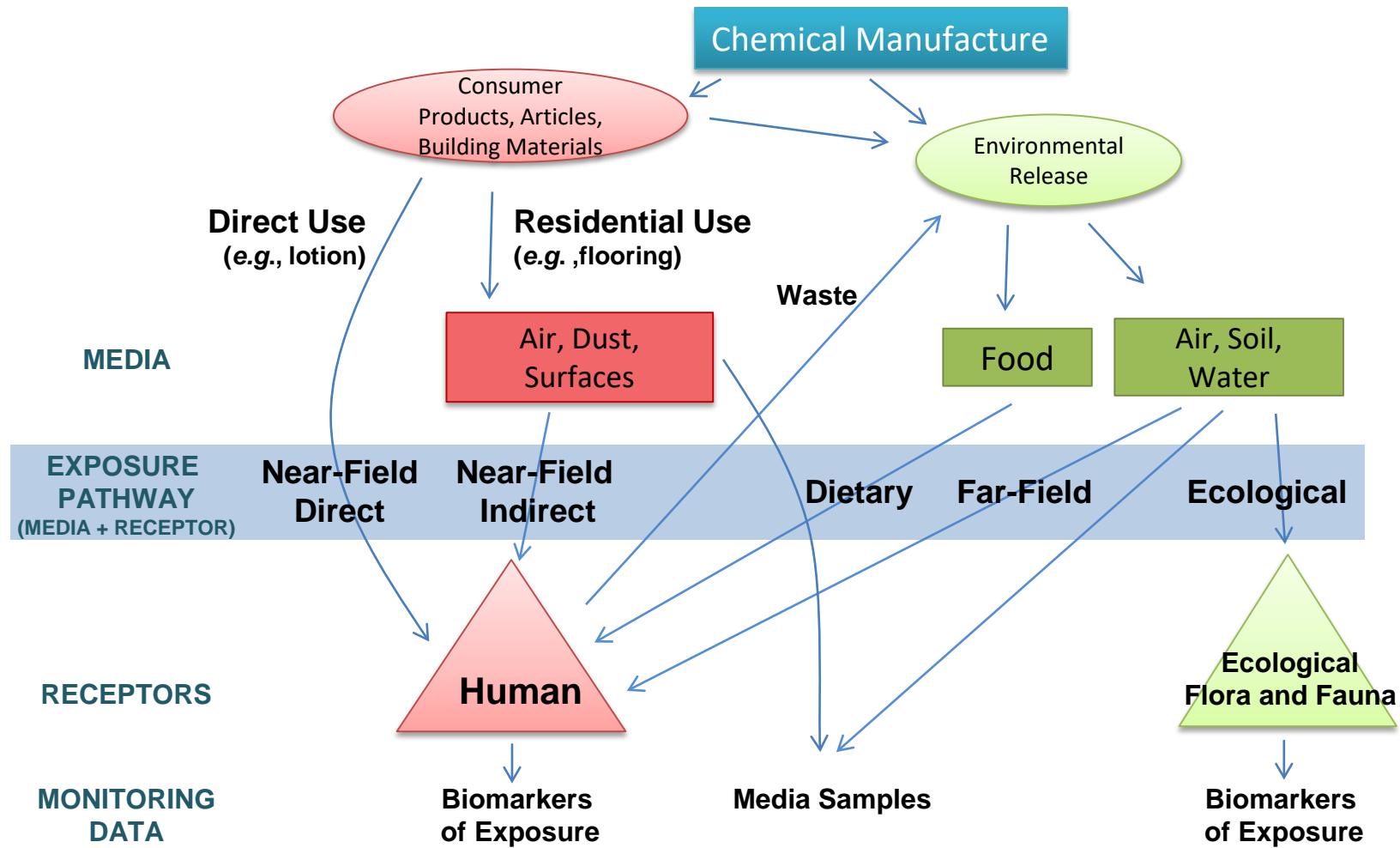
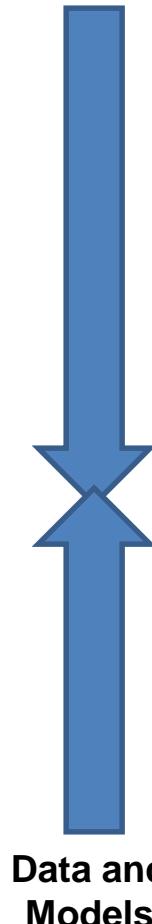


# Inferring Exposure

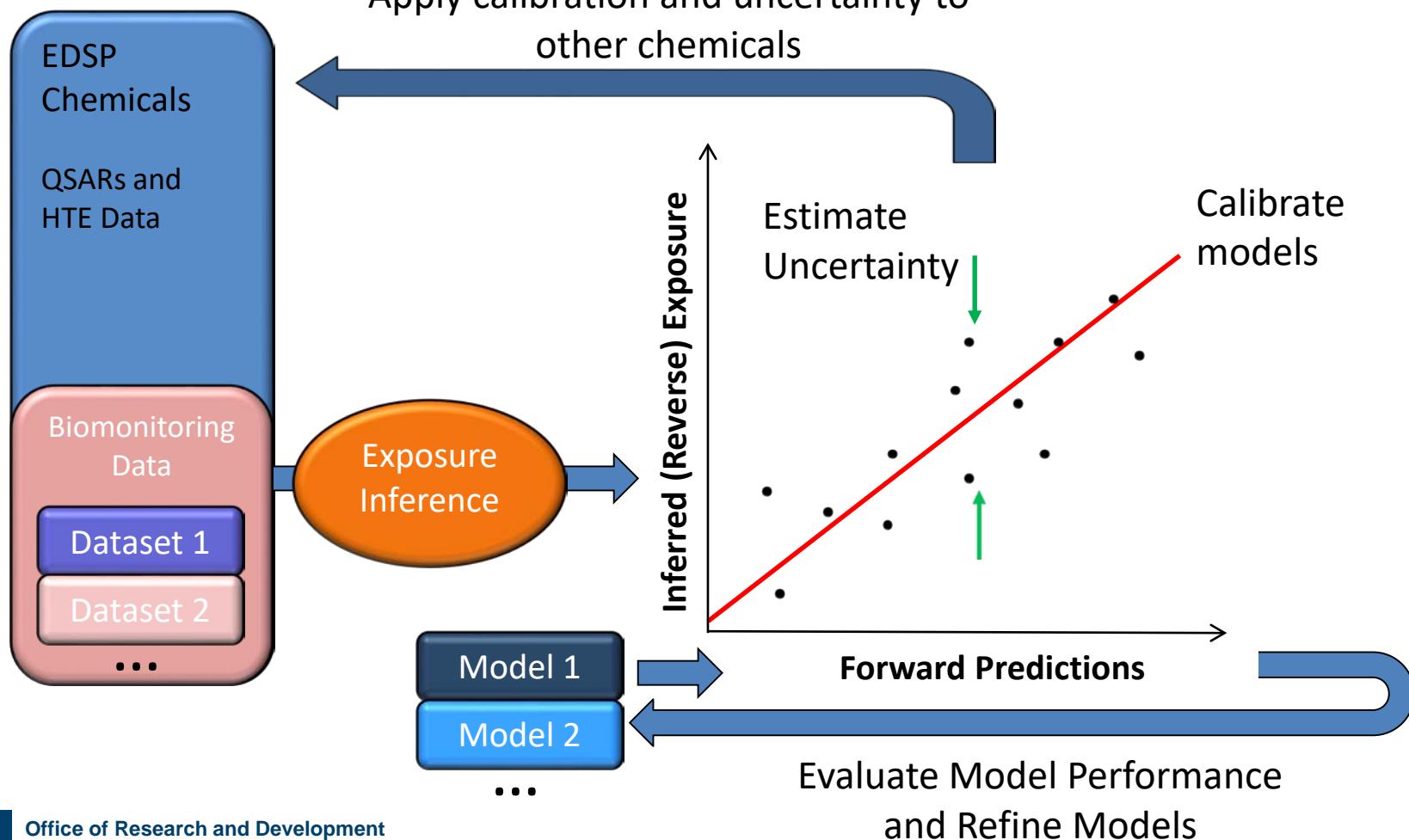


# Evaluation of Forward Predictions with Inferred Exposure

## Data and Models



# Systematic Empirical Evaluation of Models



# Exposures Inferred from NHANES

- Annual survey, data released on 2-year cycle.
- Different predictive models provide different chemical-specific predictions
  - Some models may do a better job form some chemical classes than others overall, so we want to evaluate performance against monitoring data
- Separate evaluations can be done for various demographics

National Health and Nutrition Examination Survey



# Mapping Putative Parent Chemicals to NHANES Analytes

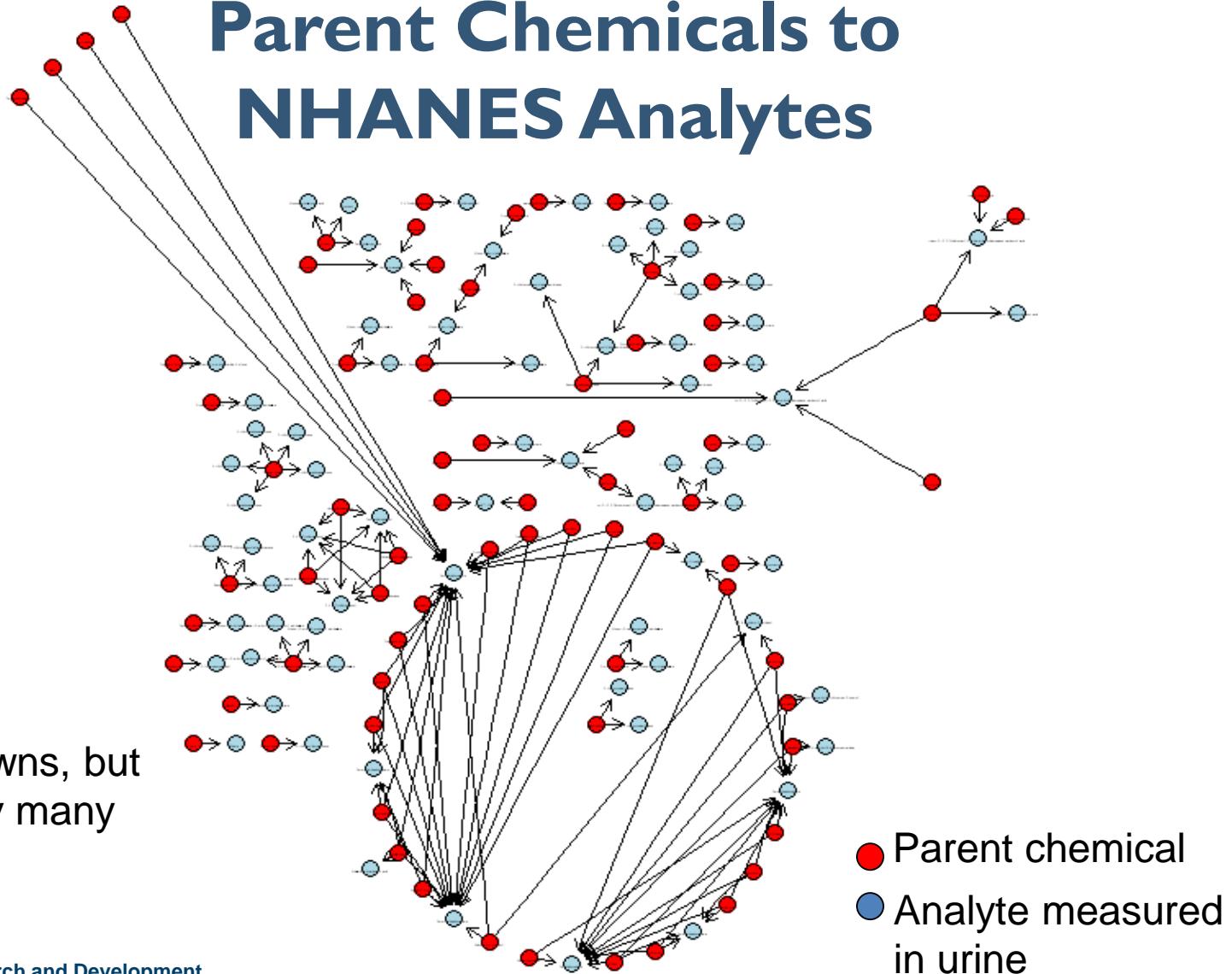


Figure from Wambaugh *et al.* (2013)

# Data Availability for Evaluating Model Predictions

- CDC NHANES urine data has been thoroughly analyzed (although other data sources exist)
- Many chemicals in the NHANES data set had median conc. below the limit of detection (LoD)
  - Most chemicals >LoD not high production volume
- 106 chemicals inferred from urine to date
- Dozens more expected with serum/blood model

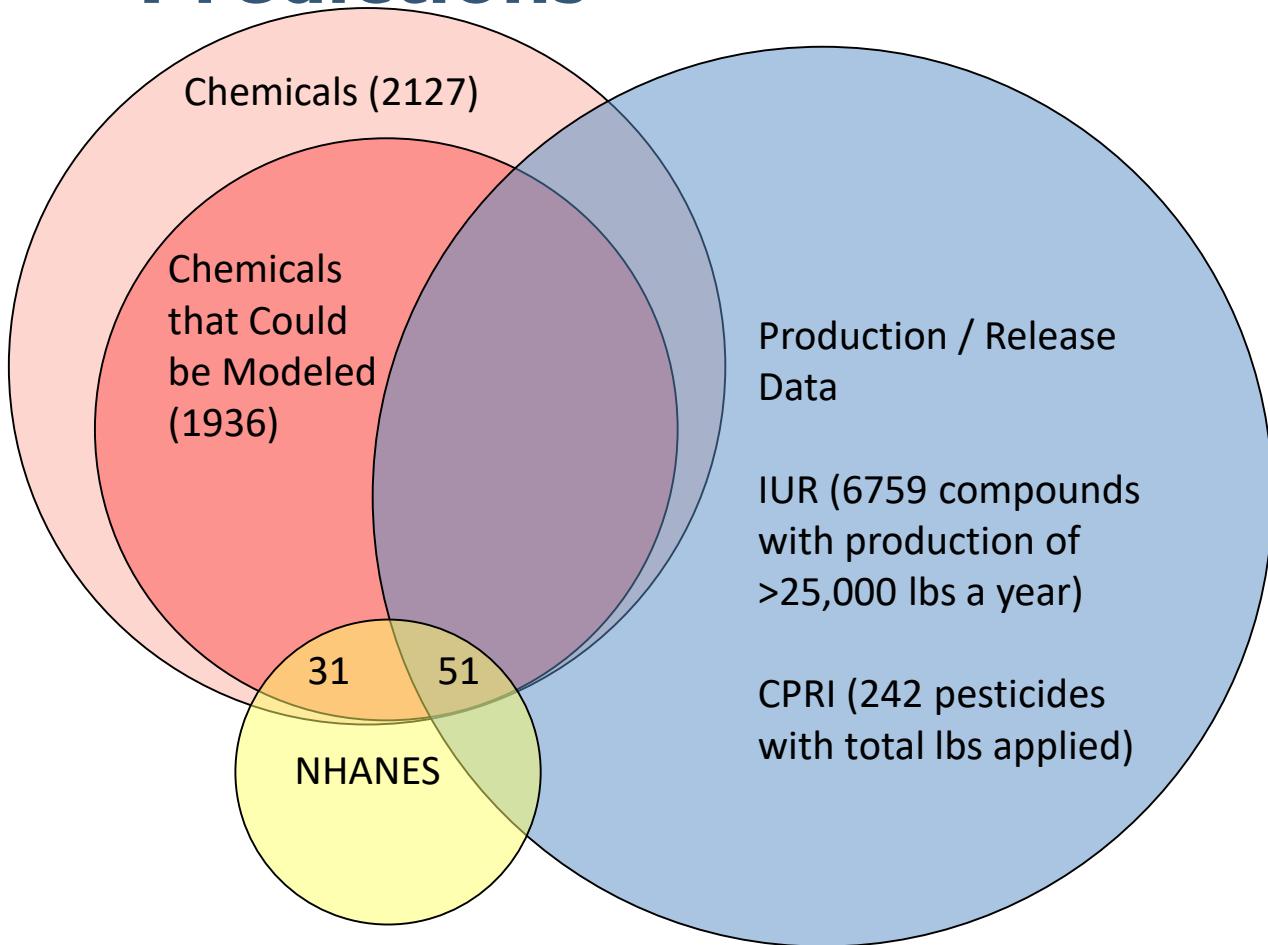
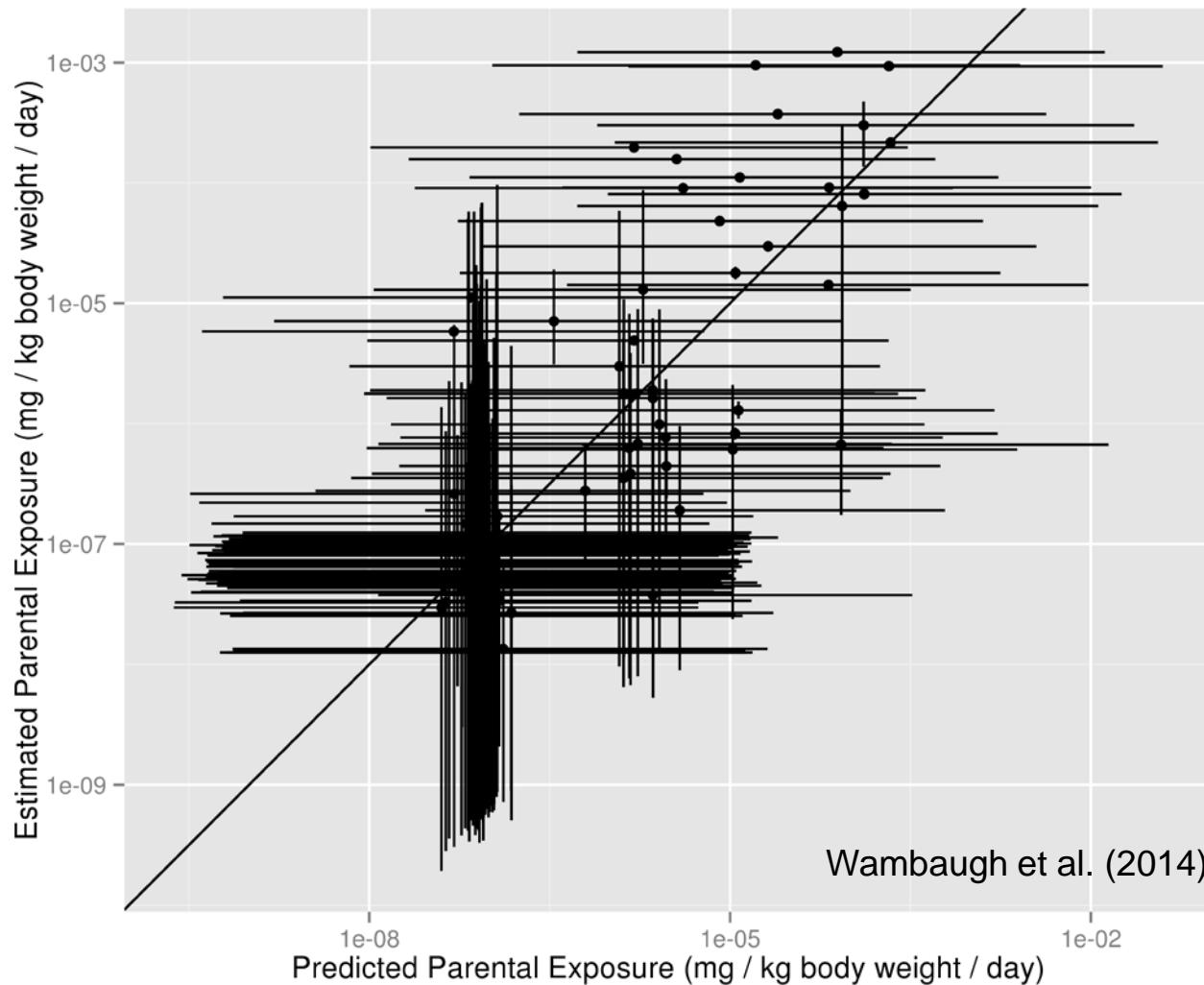


Figure from Wambaugh *et al.* (2013)

# Predicting NHANES exposure rates



CDC National Health and Nutrition Survey (NHANES) is a rolling survey to characterize biometrics for U.S. population, including exposure

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

# Predicting Exposure

Wambaugh et al. (2014)

$R^2 \approx 0.5$  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:

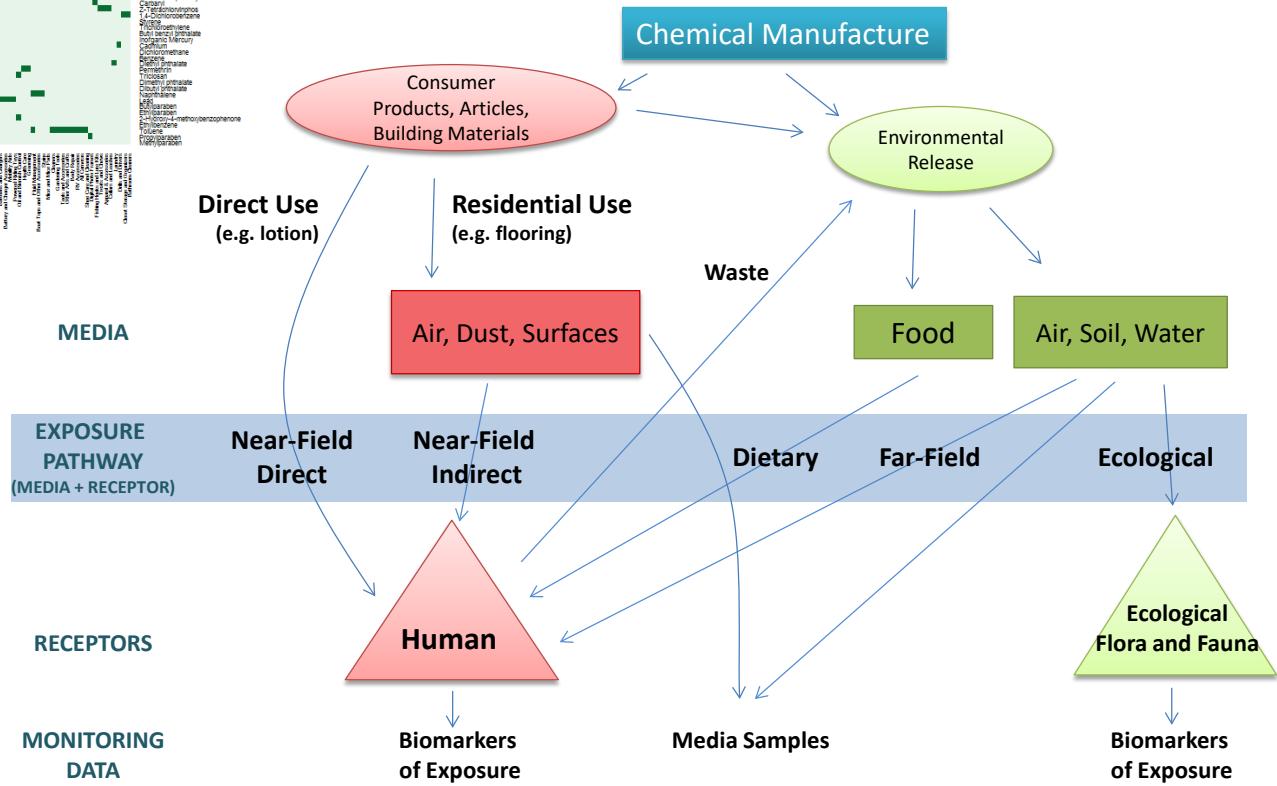
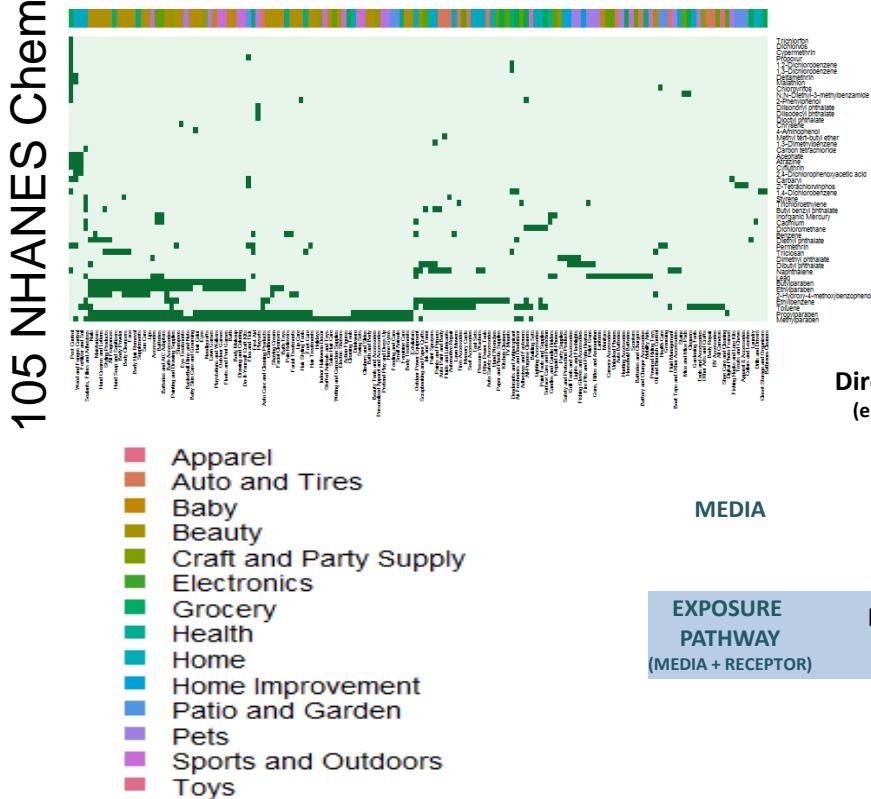
- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume

# Chemical Use Identifies Relevant Pathways

Heuristic	Description	Inferred NHANES Chemical Exposures (106)	<u>Number of Chemicals</u> Full Chemical Library ( 7784)
ACToR "Consumer use & Chemical/Industrial Process use"	Chemical substances in consumer products (e.g., toys, personal care products, clothes, furniture, and home-care products) that are also used in industrial manufacturing processes. Does not include food or pharmaceuticals.	37	683
ACToR "Chemical/Industrial Process use with no Consumer use"	Chemical substances and products in industrial manufacturing processes that are not used in consumer products. Does not include food or pharmaceuticals	14	282
ACToR UseDB "Pesticide Inert use"	Secondary ( <i>i.e.</i> , non-active) ingredients in a pesticide which serve a purpose other than repelling pests. Pesticide use of these ingredients is known due to more stringent reporting standards for pesticide ingredients, but many of these chemicals appear to be also used in consumer products	16	816
ACToR "Pesticide Active use"	Active ingredients in products designed to prevent, destroy, repel, or reduce pests (e.g., insect repellants, weed killers, and disinfectants).	76	877
TSCA IUR 2006 Total Production Volume	Sum total (kg/year) of production of the chemical from all sites that produced the chemical in quantities of 25,000 pounds or more per year. If information for a chemical is not available, it is assumed to be produced at <25,000 pounds per year.	106	7784

# Chemical Use Identifies Relevant Pathways

>2000 chemicals with Material Safety Data Sheets (MSDS) in CPCPdb (Goldsmith *et al.*, 2014)



# Chemical Use Information for >30,000 Chemicals

- Chemical-Product Categories (CPcat) database maps many different types of use information and ontologies onto each other
- Includes CPCPdb (Goldsmith, et al., 2014) with information on >2000 products from major retailors
- Largest single database has coarsest information: ACToR UseDB

**Table: Hits per use category for a given chemical**

CASRN	Category 1	Category 2	...	Category 12
65277-42-1	0	10	...	1
50-41-9	31	7	...	3
...	...	...	...	...



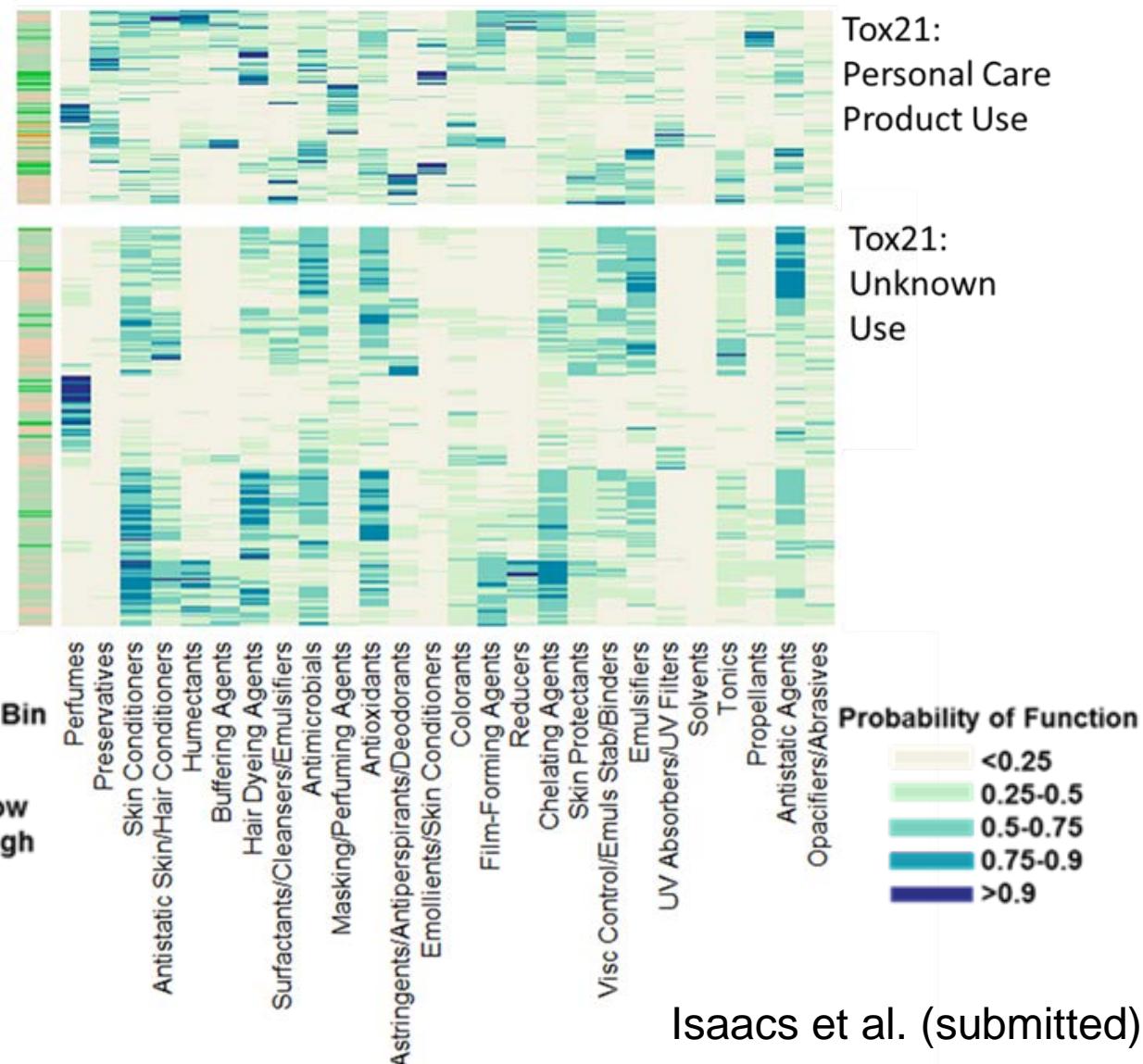
**Binary matrix**

CASRN	Category 1	Category 2	...	Category 12
65277-42-1	0	1	...	0
50-41-9	1	1	...	0
...	...	...	...	...

12 Chemical Use Categories
Antimicrobials
Chemical Industrial Process
Consumer
Dyes and Colorants
Fertilizers
Food Additive
Fragrances
Herbicides
Personal Care Products
Pesticides
Petrochemicals
Other

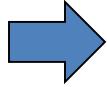
# Predicting Chemical Constituents

- Unfortunately CPCPdb does not cover every chemical-product combination (~2000 chemicals, but already >8000 in Tox21)
- We are now using machine learning to fill in the rest
- We can predict functional use and weight fraction for thousands of chemicals



# Targeted vs. Non- Targeted Screening

- Roughly half of the 106 chemicals with urine biomarkers in CDC NHANES were below the limit of detection yet Park et al. (2012) found evidence of thousands
  - Differences in sensitivity
  - Differences in screening method – targeted vs. non-targeted screening



# Targeted vs. Non- Targeted Screening

- Need to take into account transformation (e.g., metabolism)
- Need to control for background (e.g., endogenous chemicals)



# Targeted vs. Non- Targeted Screening

- When we do a targeted screen for a particular analyte, you typically gain accuracy and precision (and quantification) but are deliberately focusing on only part of the story



# Appropriate Skepticism for Non-Targeted Analysis and Suspect Screening

***“As chemists we are obliged to accept the assignment of barium to the observed activity, but as nuclear chemists working very closely to the field of physics we cannot yet bring ourselves to take such a drastic step, which goes against all previous experience in nuclear physics. It could be, however, that a series of strange coincidences has misled us.”***

Hahn and Strassmann (1938)

# Appropriate Skepticism for Non-Targeted Analysis and Suspect Screening

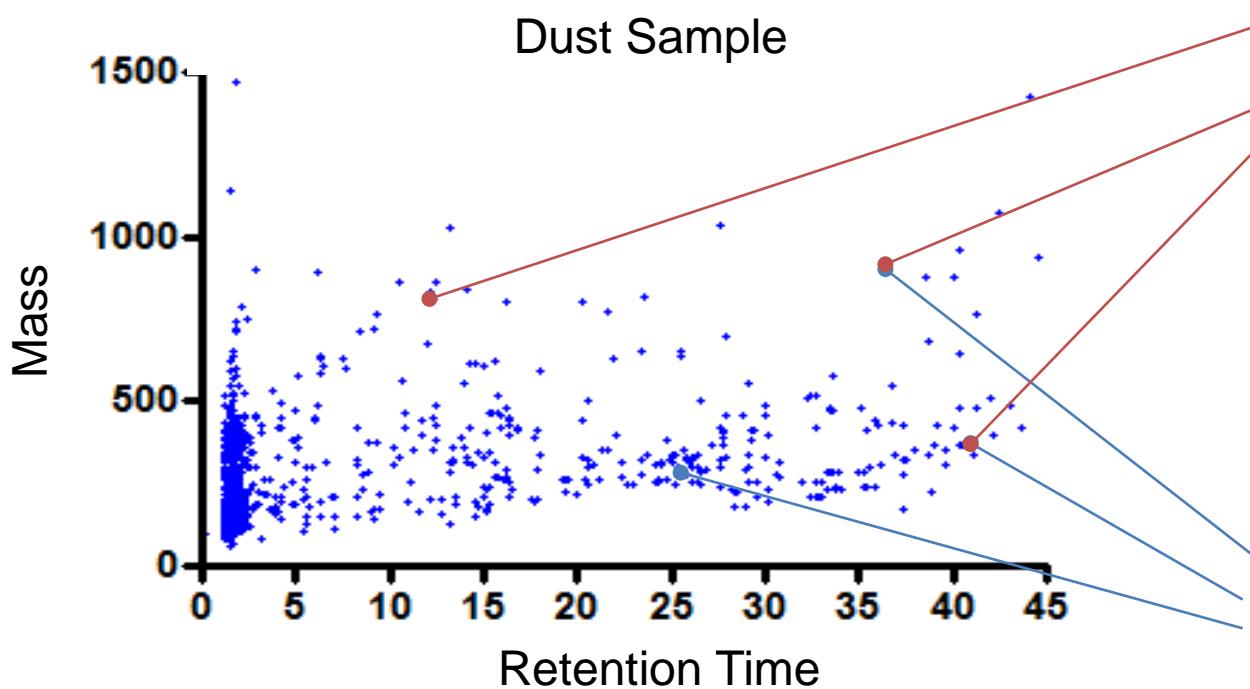
***“As chemists we are obliged to accept the assignment of barium to the observed activity, but as nuclear chemists working very closely to the field of physics we cannot yet bring ourselves to take such a drastic step, which goes against all previous experience in nuclear physics. It could be, however, that a series of strange coincidences has misled us.”***

Hahn and Strassmann (1938)

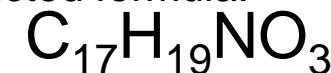
1944 Nobel Prize in Chemistry for “discovery of the fission of heavy nuclei”

# Suspect Screening in House Dust

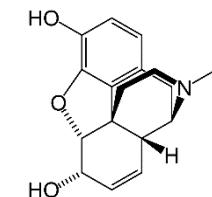
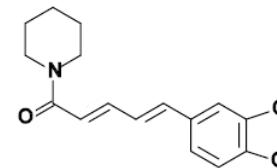
947 Peaks in an American Health Homes  
Dust Sample



Liquid chromatography peaks corresponds to a chemical with an accurate mass and predicted formula:



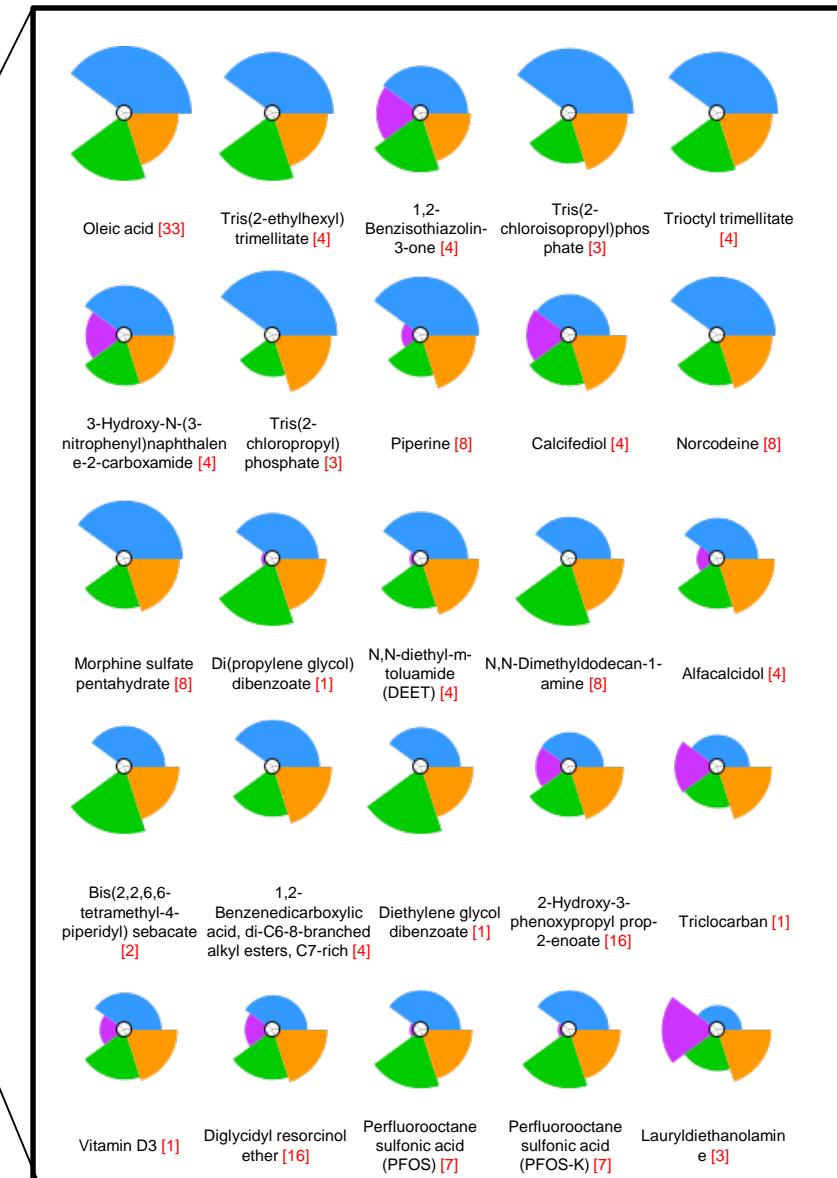
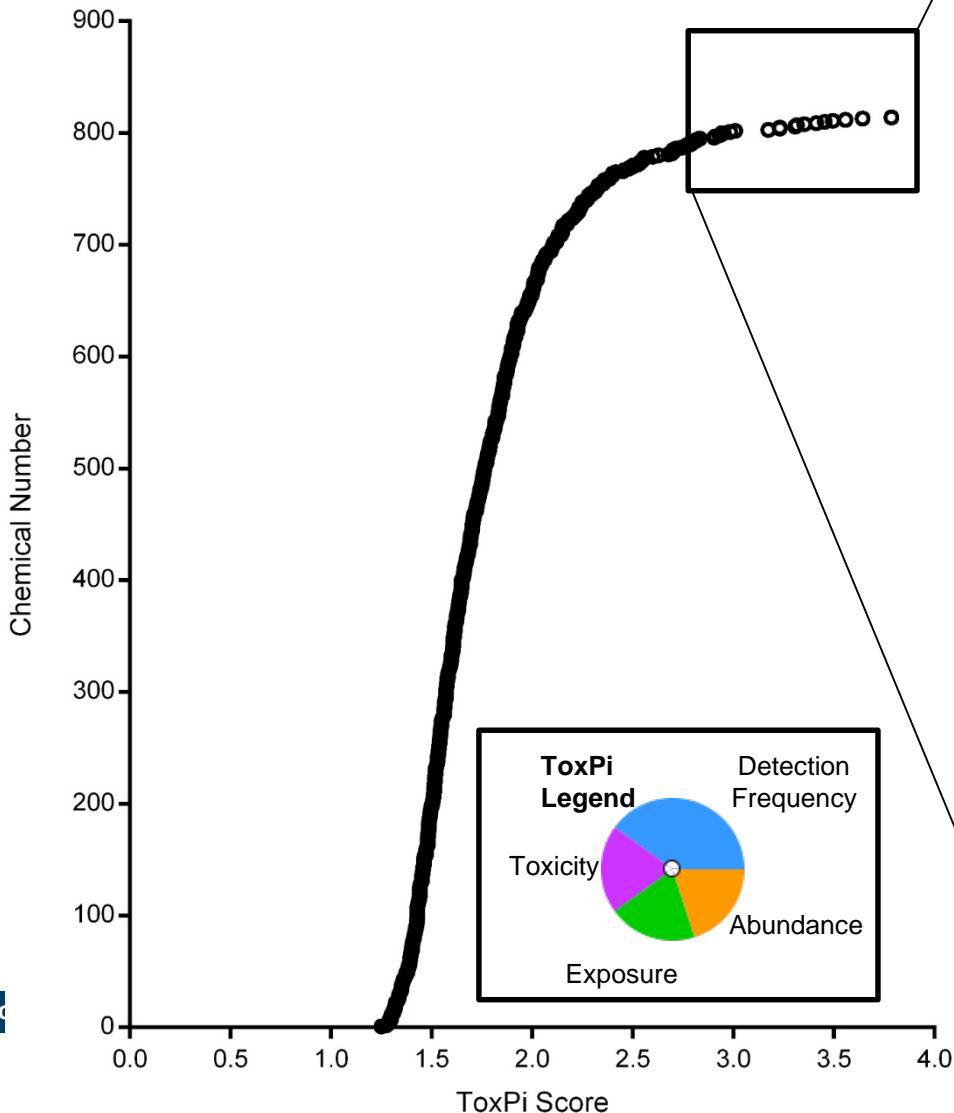
Multiple chemicals can have the same mass and formula:



Is chemical A present, chemical B, or both?

We are now expanding our identity libraries using reference samples of ToxCast chemicals

# Suspect Screening



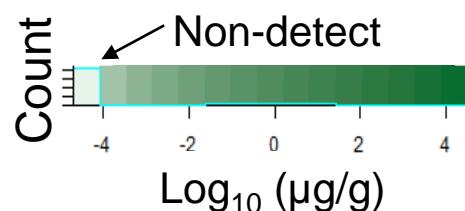
See Rager et al., Environment International (2016)

# ExpoCast Consumer Product Scan

In Preparation: Katherine Phillips et al.

“Product Deformulation to Identify Exposure Pathways for ToxCast Chemicals”

A total of 3803 unique chemical signatures were observed in test objects. Of these, 1597 were associated with a tentative chemical identification and 41 had confirmed chemical identities.



Test objects consisted of five arbitrary products in each of the following twenty categories:

- |                   |                          |                    |
|-------------------|--------------------------|--------------------|
| ■ Air freshener   | ■ Fabric upholstery      | ■ Shaving cream    |
| ■ Baby soap       | ■ Glass cleaners         | ■ Shower curtain   |
| ■ Carpet          | ■ Hand soap              | ■ Skin lotion      |
| ■ Carpet padding  | ■ Indoor house paint     | ■ Sunscreen        |
| ■ Cereals         | ■ Lipstick               | ■ Toothpaste       |
| ■ Cotton clothing | ■ Plastic children's toy | ■ Vinyl upholstery |
| ■ Deodorant       | ■ Shampoo                |                    |

- Flame Retardant
- Common Chemical
- ToxCast Chemical
- Estrogen Active

The chemicals found in a cotton shirt

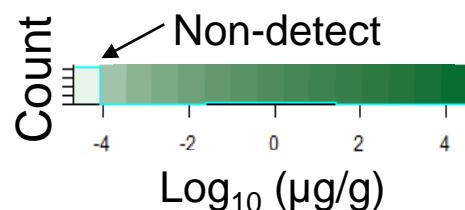


# ExpoCast Consumer Product Scan

In Preparation: Katherine Phillips et al.

“Product Deformulation to Identify Exposure Pathways for ToxCast Chemicals”

A total of 3803 unique chemical signatures were observed in test objects. Of these, 1597 were associated with a tentative chemical identification and 41 had confirmed chemical identities.



Test objects consisted of five arbitrary products in each of the following twenty categories:

- |   |  |   |
|---|--|---|
| ■ Air freshener<br>■ Baby soap<br>■ Carpet<br>■ Carpet padding<br>■ Cereals<br>■ Cotton clothing<br>■ Deodorant | ■ Fabric upholstery<br>■ Glass cleaners<br>■ Hand soap<br>■ Indoor house paint<br>■ Lipstick<br>■ Plastic children's toy | ■ Shaving cream<br>■ Shower curtain<br>■ Skin lotion<br>■ Sunscreen<br>■ Toothpaste<br>■ Vinyl upholstery |
|---|--|---|

- Flame Retardant
- Common Chemical
- ToxCast Chemical
- Estrogen Active

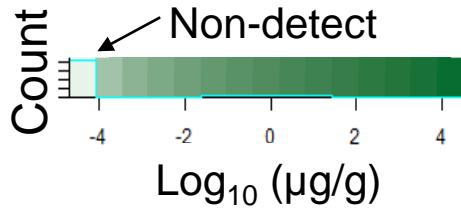
- ◀ Dark green is a high concentration
- ◀ Light green is not detected

# ExpoCast Consumer Product Scan

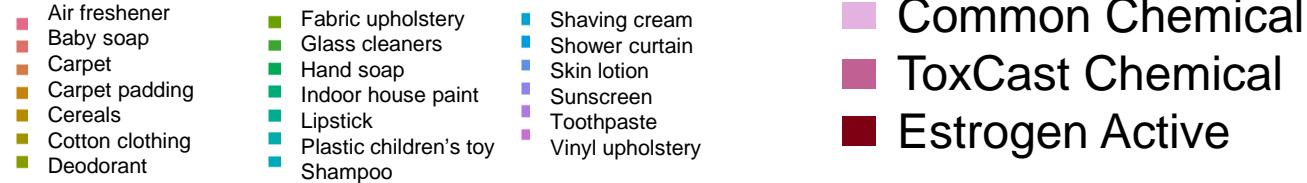
In Preparation: Katherine Phillips et al.

“Product Deformulation to Identify Exposure Pathways for ToxCast Chemicals”

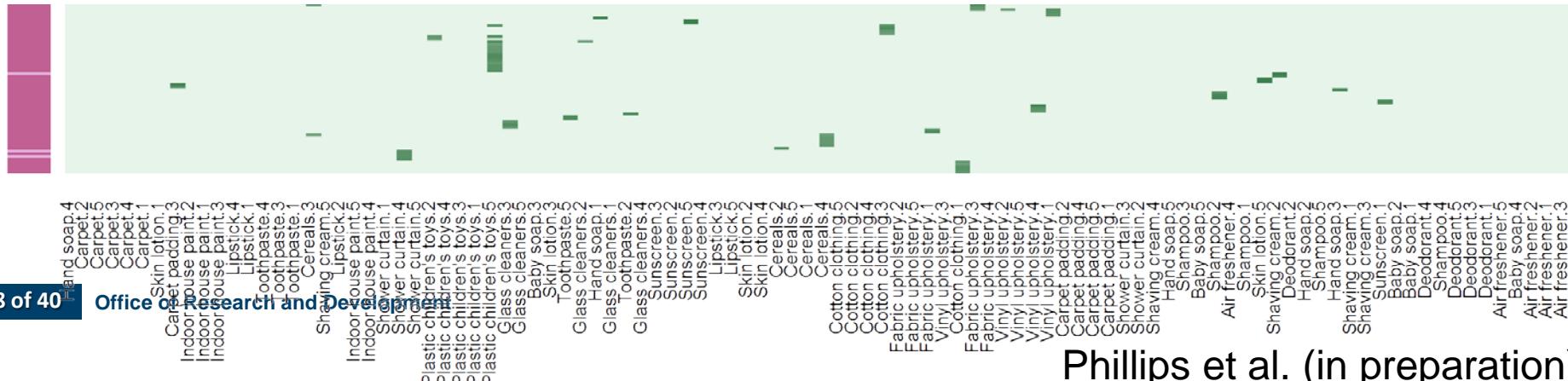
A total of 3803 unique chemical signatures were observed in test objects. Of these, 1597 were associated with a tentative chemical identification and 41 had confirmed chemical identities.



Test objects consisted of five arbitrary products in each of the following twenty categories:

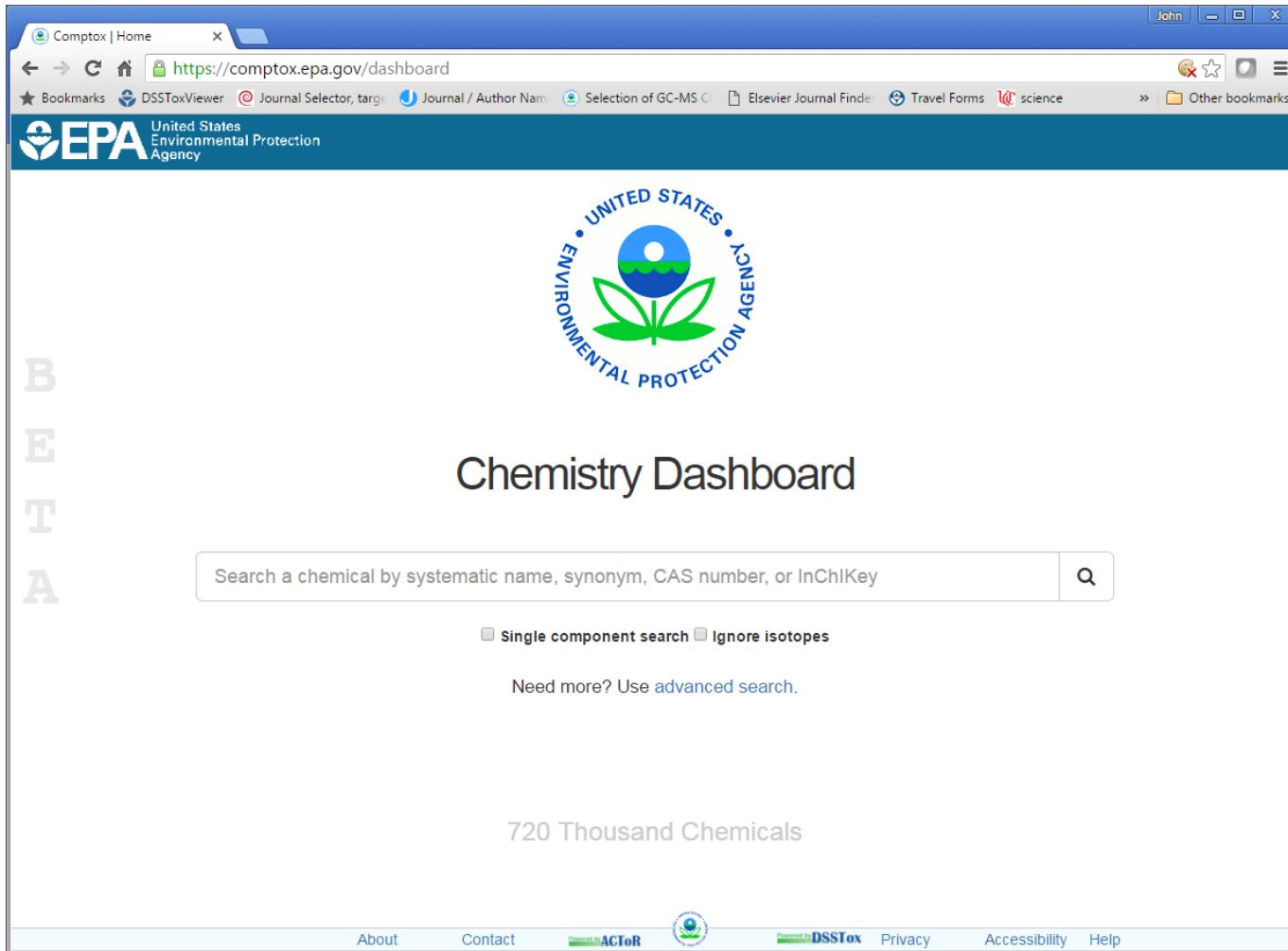


## 100 Consumer Products and Articles of Commerce



# A Google for Chemicals

<http://comptox.epa.gov/dashboard/>



The screenshot shows a web browser window titled "Comptox | Home". The address bar contains the URL "https://comptox.epa.gov/dashboard". The page itself is the "Chemistry Dashboard" of the EPA's Comptox system. It features the official United States Environmental Protection Agency logo at the top center. Below the logo, the text "Chemistry Dashboard" is displayed. A search bar is present with the placeholder text "Search a chemical by systematic name, synonym, CAS number, or InChIKey". To the right of the search bar is a magnifying glass icon. Underneath the search bar are two checkboxes: "Single component search" and "Ignore isotopes". Below these checkboxes is a link "Need more? Use advanced search.". At the bottom of the dashboard, the text "720 Thousand Chemicals" is visible. The footer of the page includes links for "About", "Contact", "Powered by ACToR" (with a small green circular logo), "Powered by DSSTox" (with a small blue circular logo), "Privacy", "Accessibility", and "Help". On the far left edge of the dashboard area, the letters "B", "E", "T", and "A" are partially visible, likely representing the word "BETA".

Examples:

“Bisphenol A”

“C17H19NO3”

# Conclusions

- The exposure pathway is the actual interaction of the receptor and media, and this event is often confounded by various sources of uncertainty
- Can use a combination of forward modeling and reverse inference from biomarkers of exposure to examine plausible exposures – there is a great need to carefully consider predictive performance with statistical analysis
- Monitoring is tricky, and there are trade offs between the precision of targeted monitoring for specific chemicals and non-targeted screening for all exogenous chemicals
  - Expanded monitoring data (exposure surveillance) allows evaluation of model predictions
    - Are chemicals missing that we predicted would be there?
    - Are there unexpected chemicals?
- There are low levels of thousands of chemicals present in the metabolome, relating these to exposures and health effects is an important unsolved problem

## Chemical Safety for Sustainability (CSS) Rapid Exposure and Dosimetry (RED) Project

### NCCT

Chris Grulke  
Richard Judson  
Dustin Kapraun\*  
Robert Pearce\*  
Ann Richard  
Woody Setzer  
Rusty Thomas  
**John Wambaugh**  
Antony Williams

### NRMRL

Yirui Liang\*  
Xiaoyu Liu

### NHEERL

Christopher Ecklund  
Jane Ellen Simmons  
Marina Evans  
Mike Hughes

**\*Trainees**

### NERL

Craig Barber  
Derya Biryol\*  
Namdi Brandon\*  
Peter Egeghy  
Brandall Ingle\*  
**Kristin Isaacs**  
Seth Newton  
Katherine Phillips  
Paul Price  
Mark Strynar  
Jon Sobus  
Mike Tornero-Velez  
Elin Ulrich  
Dan Vallero  
Barbara Wetmore

**Lead CSS Matrix Interface:**  
John Kenneke (NERL)

# Collaborators

Arnot Research and Consulting  
Jon Arnot  
Battelle Memorial Institute  
Anne Louise Sumner  
Anne Gregg  
Chemical Computing Group  
Rocky Goldsmith  
National Institute for Environmental Health Sciences (NIEHS) National Toxicology Program  
Mike Devito  
Steve Ferguson  
Nisha Sipes  
Netherlands Organisation for Applied Scientific Research (TNO)  
Sieto Bosgra  
North Carolina Central University  
Chantel Nicolas  
Research Triangle Institute  
Timothy Fennell  
ScitoVation  
Harvey Clewell  
Silent Spring Institute  
Robin Dodson  
Southwest Research Institute  
Alice Yau  
Kristin Favela  
Summit Toxicology  
Lesa Aylward  
Tox Strategies  
Caroline Ring  
University of California, Davis  
Deborah Bennett  
Hyeong-Moo Shin  
University of Michigan  
Olivier Jollivet  
University of North Carolina, Chapel Hill  
Alex Tropsha

# References

- Arnot, Jon A., et al. "Screening Level Risk Assessment Model for Chemical Fate and Effects in the Environment." *Environmental Science and Technology*, 40: 2316-2323 (2006)
- Cohen Hubal, Elaine A., et al. "Exposure science and the U.S. EPA National Center for Computational Toxicology." *Journal of Exposure Science and Environmental Epidemiology*, 20: 231-236 (2008)
- Dionisio, Kathie L., et al. "Exploring Consumer Exposure Pathways and Patterns of Use for Chemicals in the Environment." *Toxicology Reports* (2015)
- Edwards, Stephen, and Julian Preston "Systems biology and mode of action based risk assessment." *Toxicological Sciences* 106: 312-318 (2008)
- Egeghy, Peter P., et al. "The exposure data landscape for manufactured chemicals." *Science of the Total Environment* 414: 159-166 (2012)
- Gallagher, Jane, et al. "Mechanistic Indicators of Childhood Asthma (MICA) Study: piloting an integrative design for evaluating environmental health." *BMC Public Health* 11: 344 (2011)
- Goldsmith, M-R., et al. "Development of a consumer product ingredient database for chemical exposure screening and prioritization." *Food and chemical toxicology* 65 (2014): 269-279.
- Isaacs, Kristin K., et al. "SHEDS-HT: An Integrated Probabilistic Exposure Model for Prioritizing Exposures to Chemicals with Near-Field and Dietary Sources." *Environmental Science and Technology* 48.21 (2014): 12750-12759.
- LaKind, Judy S., and Daniel Q. Naiman. "Daily intake of bisphenol A and potential sources of exposure: 2005–2006 National Health and Nutrition Examination Survey." *Journal of Exposure Science and Environmental Epidemiology* 21.3 (2010): 272-279.
- National Research Council. "Exposure Science in the 21st Century: A Vision and a Strategy." (2012).
- 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)
- Pilkey, Orrin H., and Linda Pilkey-Jarvis. *Useless arithmetic: why environmental scientists can't predict the future*. Columbia University Press, 2007.
- Rosenbaum, Ralph K., et al. "USEtox—the UNEP-SETAC toxicity model: recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment" *International Journal of Life Cycle Assessment* 13: 532-546 (2006)
- Silver, Nate. *The signal and the noise: Why so many predictions fail—but some don't*. Penguin, 2012.
- Sobus, Jon R., et al. "A biomonitoring framework to support exposure and risk assessments." *Science of the Total Environment* 409.22 (2011): 4875-4884.
- Tan, Yu-Mei, et al. "Reconstructing human exposures using biomarkers and other “clues”." *Journal of Toxicology and Environmental Health, Part B* 15.1 (2012): 22-38.
- 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).
- Wild, Christopher P., "Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology." *Cancer Epidemiology, Biomarkers and Prevention*, 14, 2487 (2005)