EPA
United States
Environmental Protection
Agency

# Using Data Science for Chemical Safety at the U.S. EPA

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

**November 15, 2017**

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

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

# EPA Office of Research and Development

United States Environmental Protection Agency

- The Office of Research and Development (ORD) is the scientific research arm of EPA
  - 558 peer-reviewed journal articles in 2016
- Research is conducted by ORD's three national laboratories, four national centers, and two offices
  - Includes **National Center for Computational Toxicology** and **National Exposure Research Laboratory**
- 14 facilities across the country
- Six research programs
  - Includes **Chemical Safety for Sustainability**
- Research conducted by a combination of Federal scientists; contract researchers; and postdoctoral, graduate student, and post-baccalaureate trainees


ORD Facility in
Research Triangle Park, NC

# Chemical Regulation in the United States

- Park *et al.* (2012): At least 3221 chemicals 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

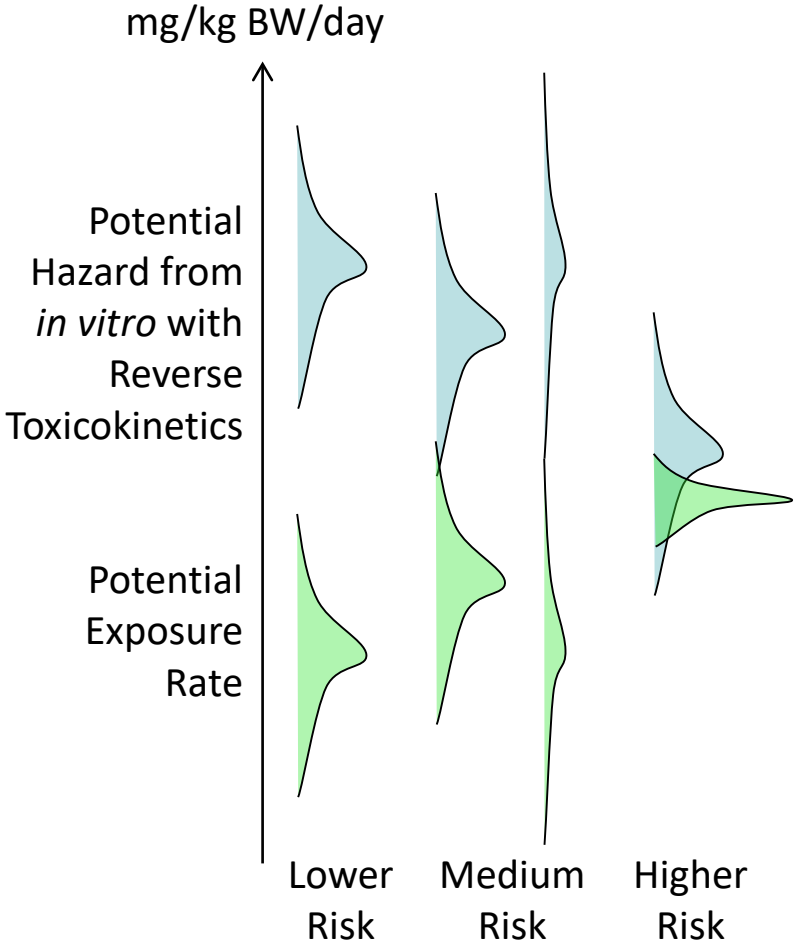
# Chemical Regulation in the United States

- Most other chemicals, ranging from industrial waste to dyes to packing materials are covered by the recently updated Toxic Substances Control Act (TSCA)
  - Thousands of chemicals on the market were either “grandfathered” in or were allowed without experimental assessment of hazard, toxicokinetics, or exposure
  - Thousands of new chemical use submissions are made to the EPA every year
  - **Methods are being developed to prioritize these existing and new chemicals for testing**


November 29, 2014

# Chemical Risk = Hazard + Exposure

- National Research Council (1983) identified chemical risk as a function of both inherent hazard and exposure
- To address thousands of chemicals, we need to use “high throughput methods” to prioritize those chemicals most worthy of additional study
- **High throughput risk prioritization** needs:
  1. high throughput **hazard** characterization (from HTT project)
  2. high throughput **exposure** forecasts
  3. high throughput **toxicokinetics** (*i.e.*, dosimetry) linking hazard and exposure

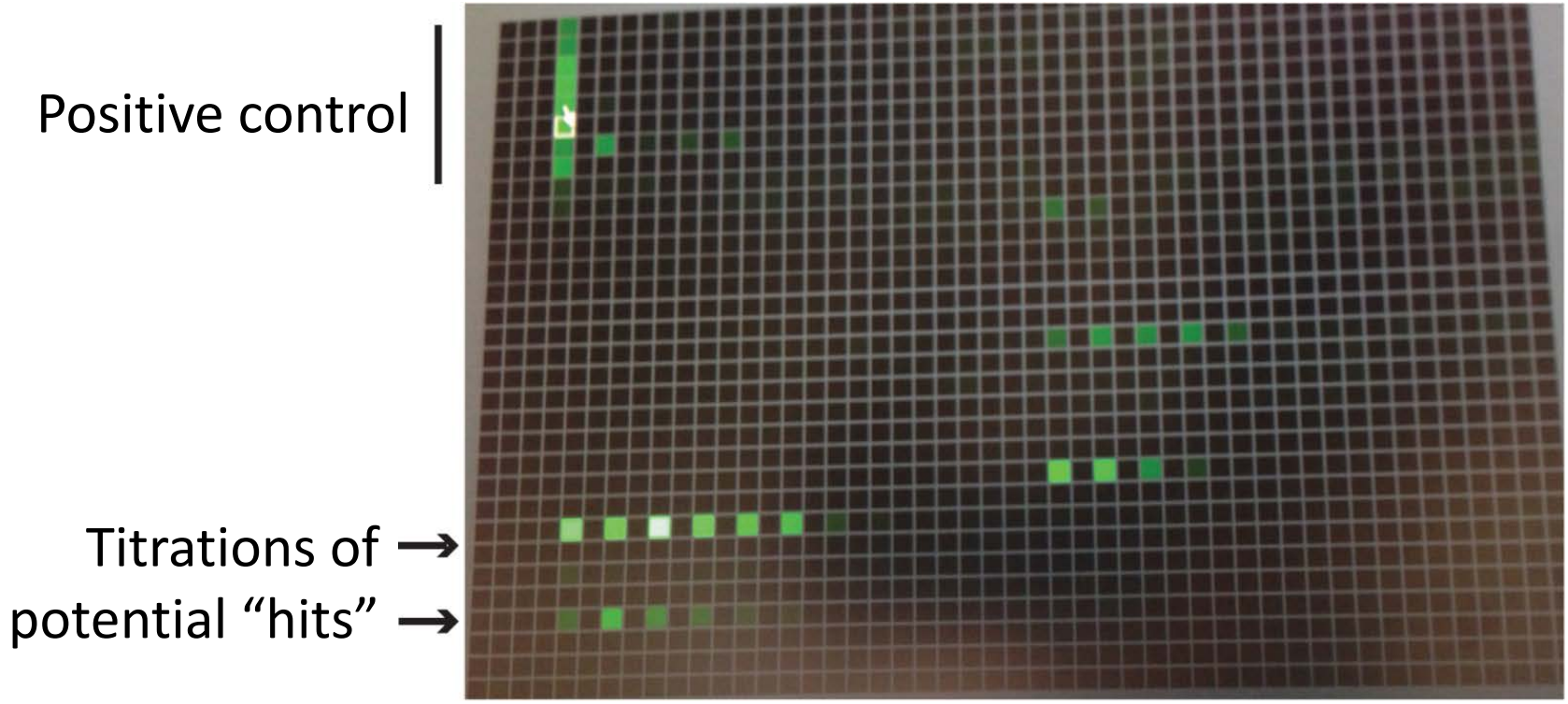


# High-throughput Screening

Hertzberg and Pope (2000):

- "New technologies in high-throughput screening have significantly increased throughput and reduced assay volumes"

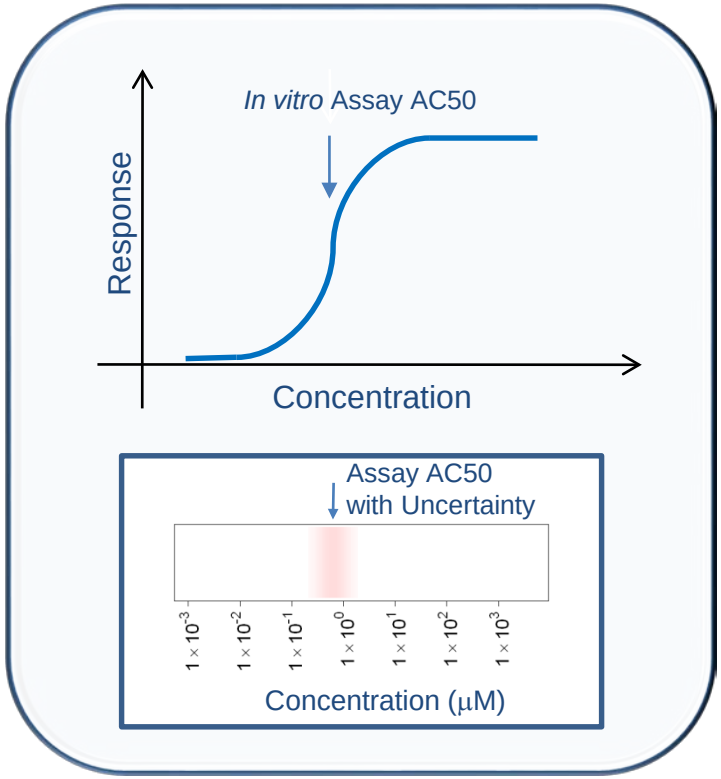
- "Key advances over the past few years include new fluorescence methods, detection platforms and liquid-handling technologies."




Kaewkhaw et al. (2016)

# High-Throughput Bioactivity

- We might estimate points of departure *in vitro* using high throughput screening (HTS)
- **Tox21**: Examining >8,000 chemicals using ~50 assays intended to identify interactions with biological pathways (Schmidt, 2009)
- **ToxCast**: For a subset (>2000) of Tox21 chemicals ran >1100 additional assays (Kavlock *et al.*, 2012)
- Most assays conducted in dose-response format (identify 50% activity concentration – AC50 – and efficacy if data described by a Hill function, Filer *et al.*, 2016)
- All data is public: http://comptox.epa.gov/dashboard/



# 2017 National Academies Report

EPA United States Environmental Protection Agency

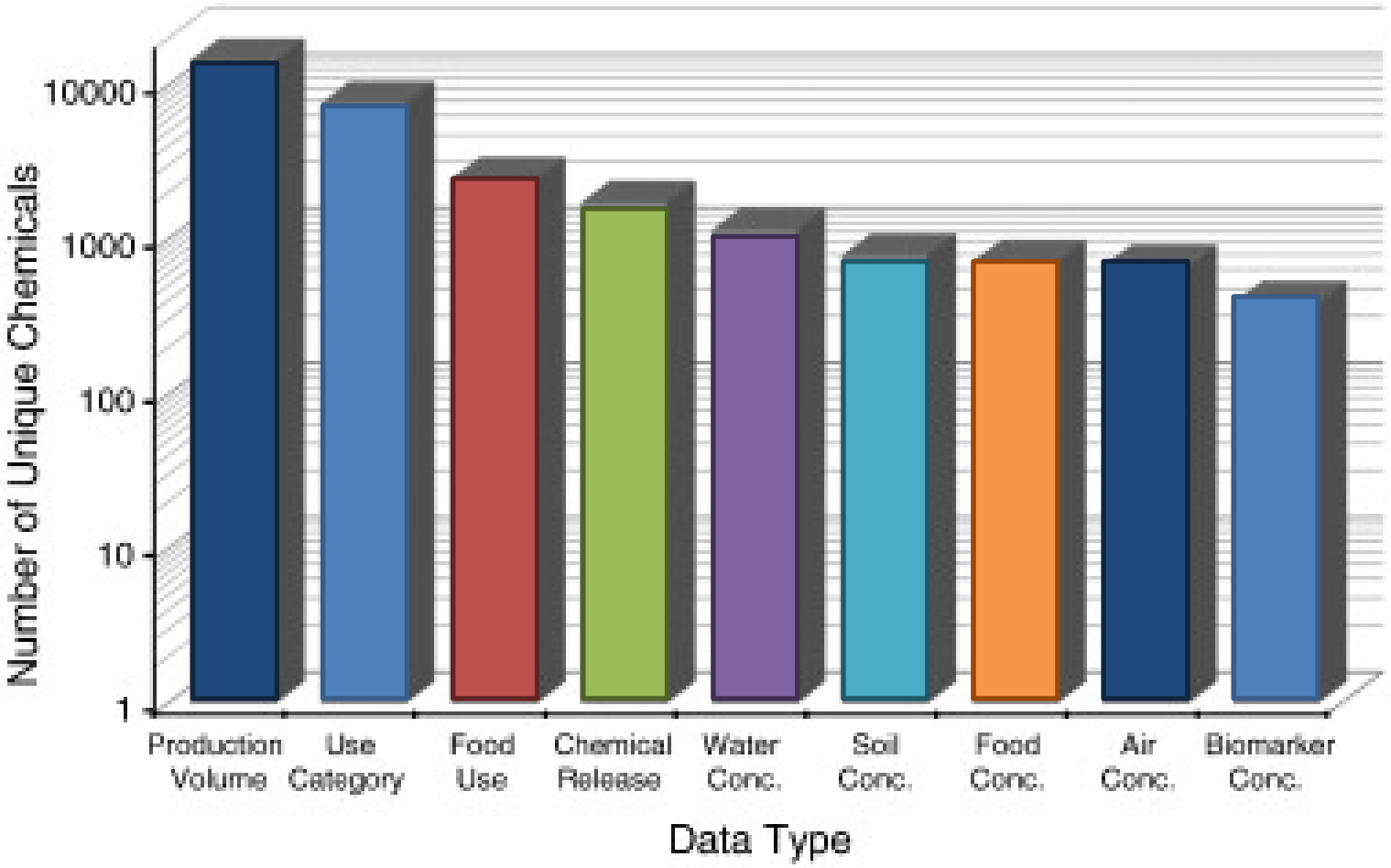




NAS (2017)

"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 [ExpoCast] have enabled first-tier risk-based rankings of chemicals on the basis of margins of exposure"

# Limited Available Data for Exposure Estimations




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

# CDC NHANES

What do we know about exposure?

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**

# Kapraun et al. (2017) EHP

- Targeted analytical chemistry used to quantitate concentration of specific chemicals in urine
  - Samples must be divided up for each chemical tested
  - NHANES cohort divided up to allow enough sample for testing all chemicals

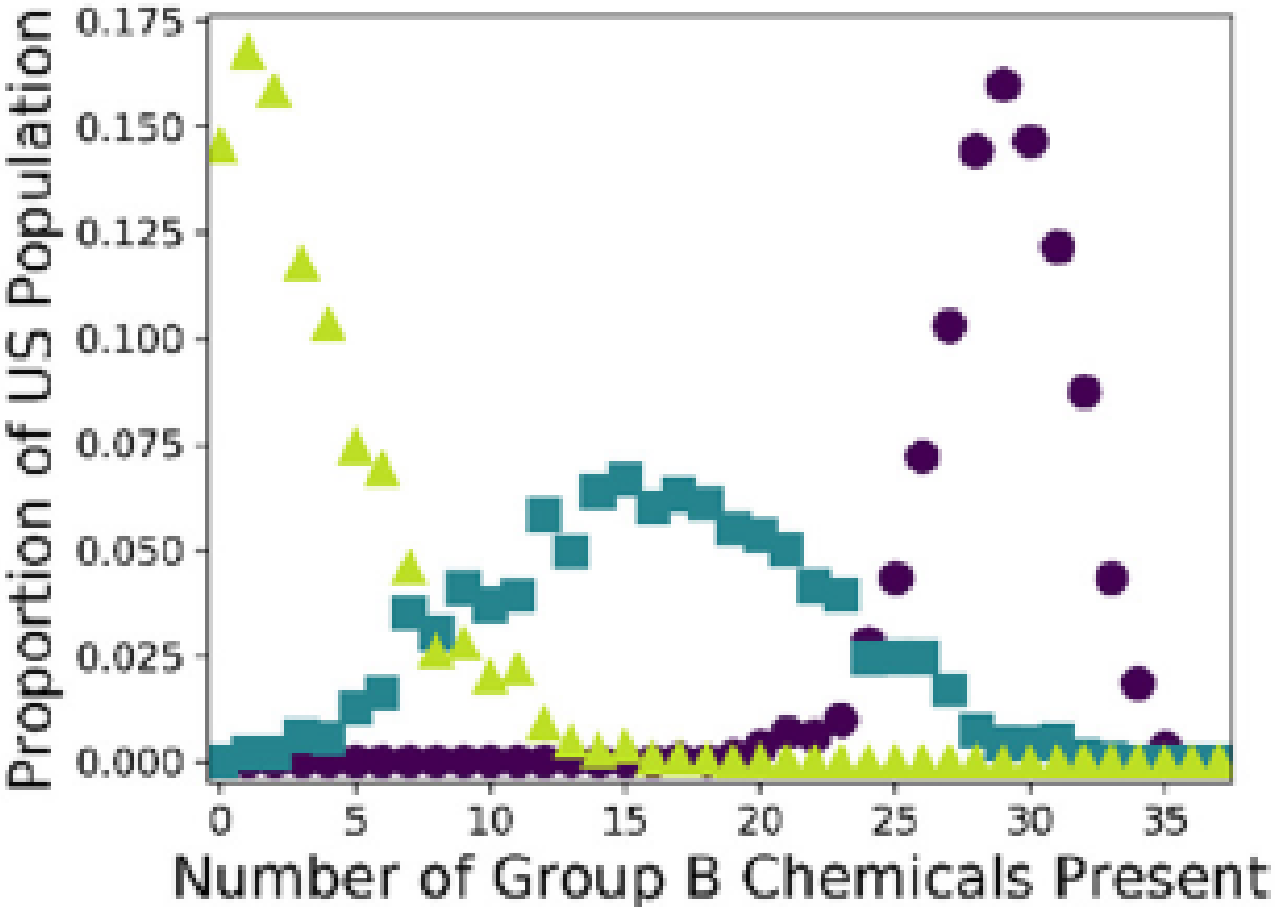
**Table 4.** Summary information for each of the National Health and Nutrition Examination Survey (NHANES) 2009–2010 subsamples.

| Category | Subsample A | Subsample B | Subsample C |
|---|---|---|---|
| Number of subjects | 2,741 | 2,736 | 2,132 |
| Number of chemicals | 29 | 37 | 40 |
| Maximum weight | 476,883.0 | 426,061.1 | 413,068.1 |
| Minimum weight | 14,002.7 | 13,975.1 | 12,659.3 |
| Sum of weights | 258,281,689.4 | 272,911,664.0 | 226,021,580.6 |
| Records needed | 18,445.1 | 19,528.5 | 17,854.1 |

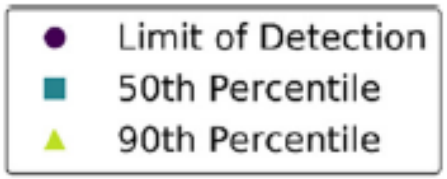
- We will focus on “Sub-sample B” PAHs, Phenols, Pesticides, and Phthalates

# Co-Occurrence of Chemicals in Individuals

The number of chemicals (out of 37) “present” in individuals depends upon where you set the limit




Ideally we would use some sort of chemical toxicity informed point of departure but don’t have that for all chemicals

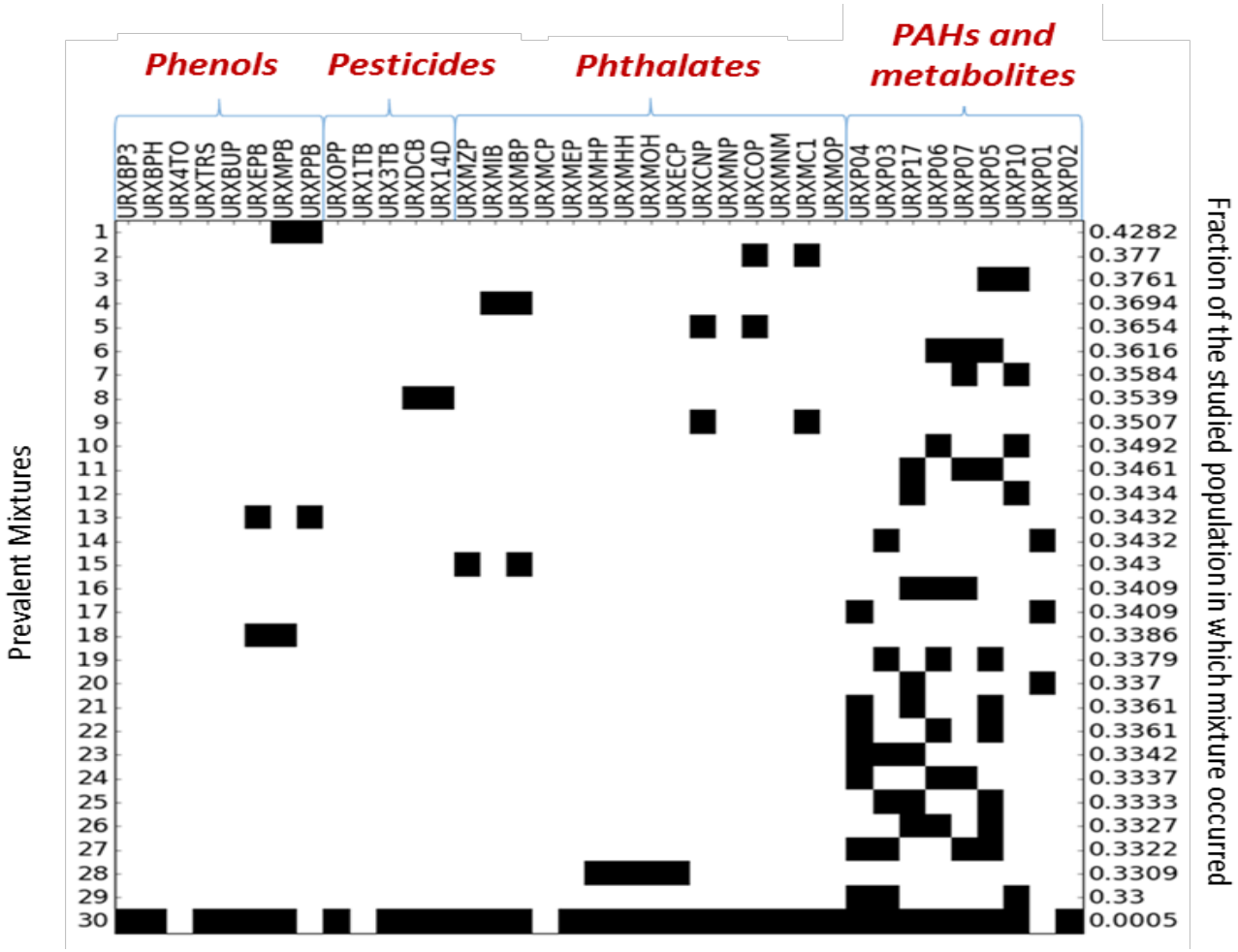





Kapraun et al. (2017)

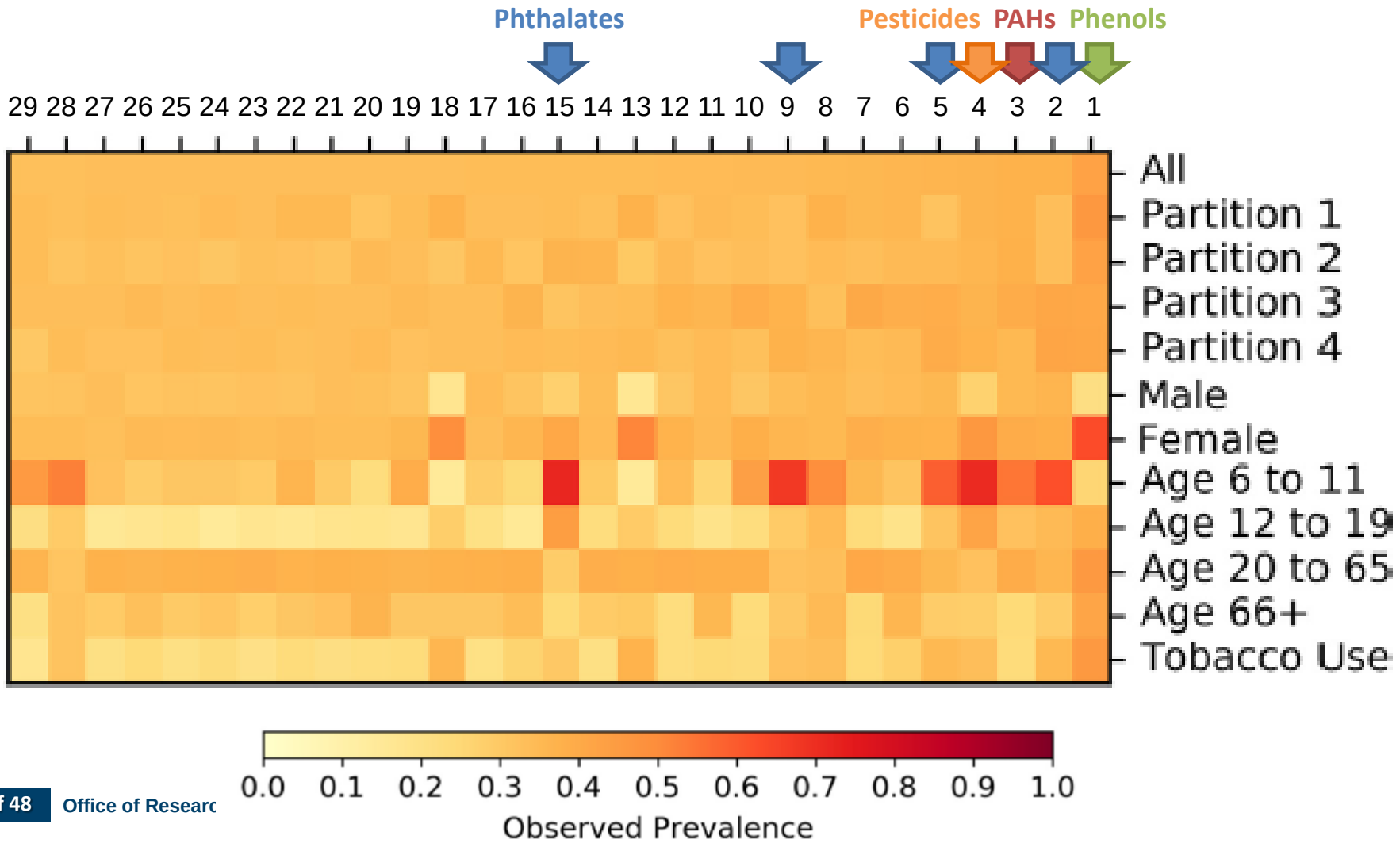
# Identifying Prevalent Mixtures

- We are using data-mining methods (frequent itemset mining or FIM, Borgelt, 2012) to identify combinations of items (chemicals) that co-occur together within samples from same individual
- Used total population median concentration as threshold for “presence”
- Identified a few dozen mixtures present in >30% of U.S. population




Kapraun et al. (2017)

# Demographic-Specific Prevalence of Combinations

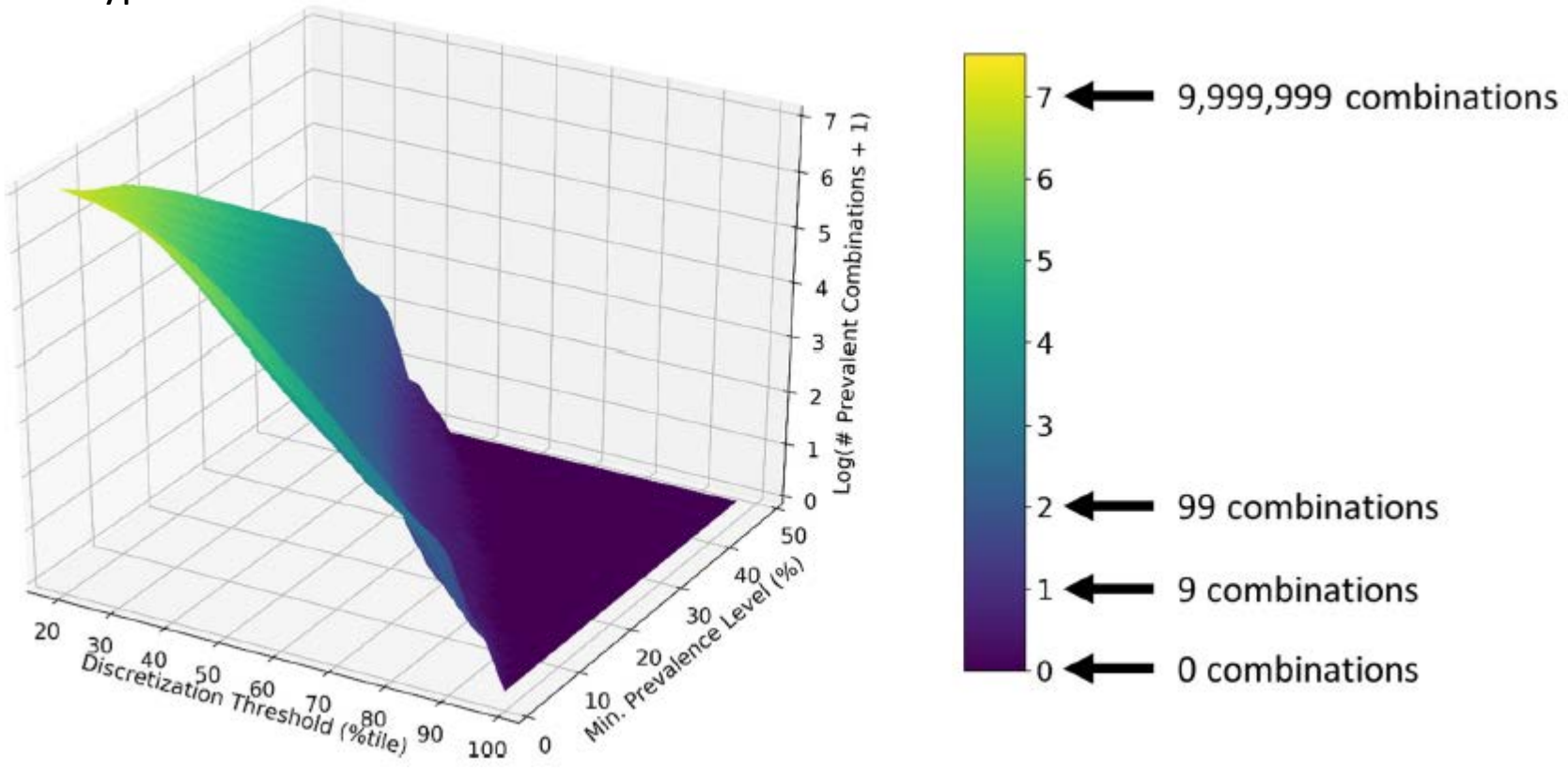





Kapraun et al. (2017)

# A Testable Number of Combinations

While high throughput screening (HTS) allows thousands of tests, there are millions of hypothetical combinations

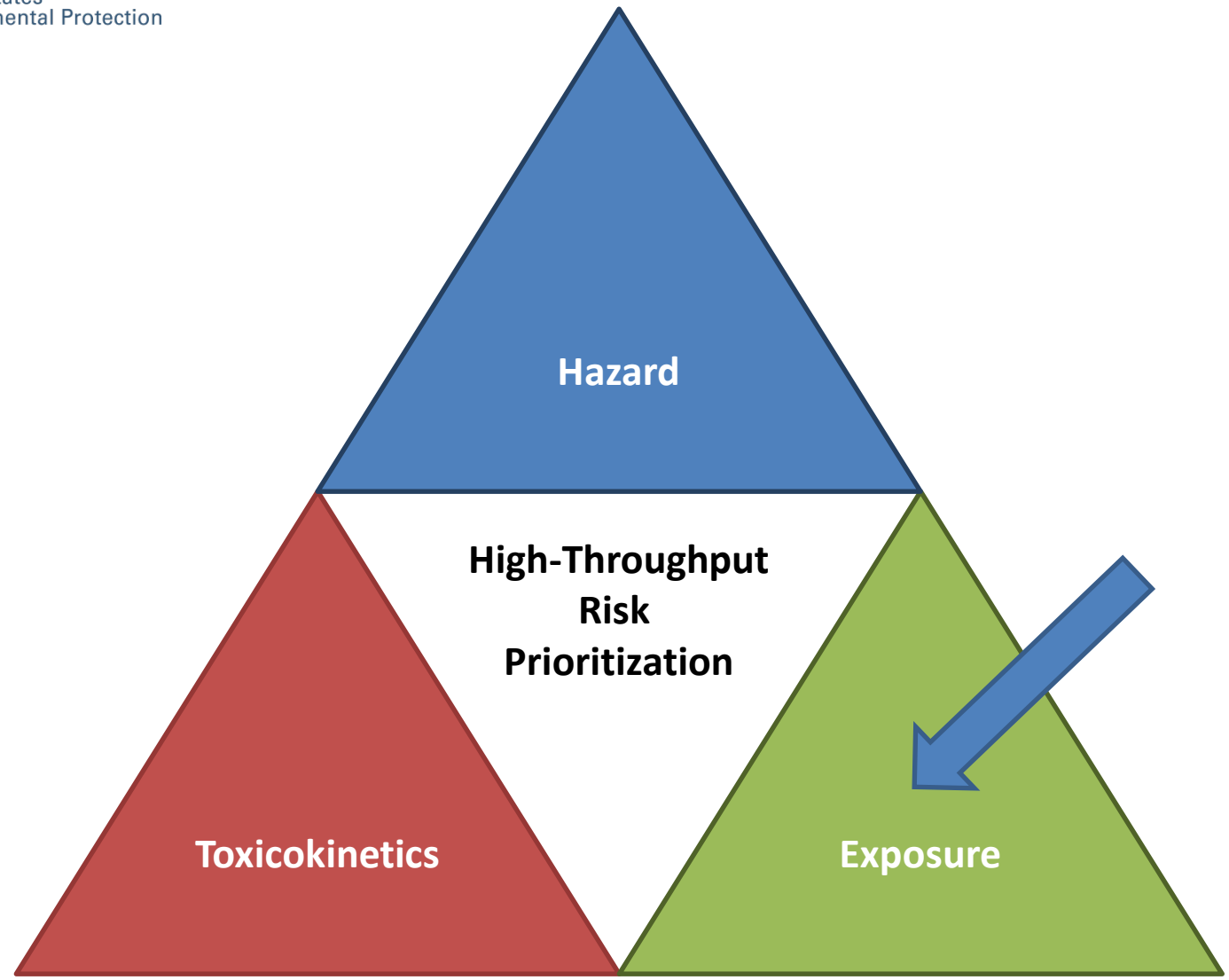



"Exposure based priority setting" (NAS, 2017) allows identification of most important mixtures to test

Kapraun et al. (2017)

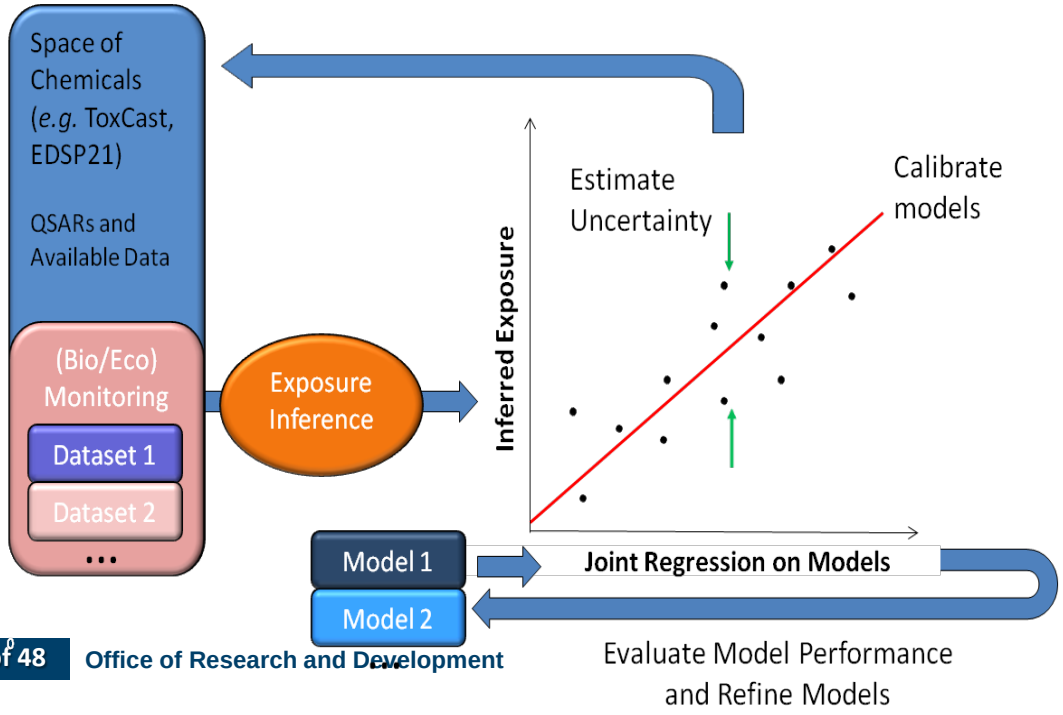
# New Exposure Data and Models

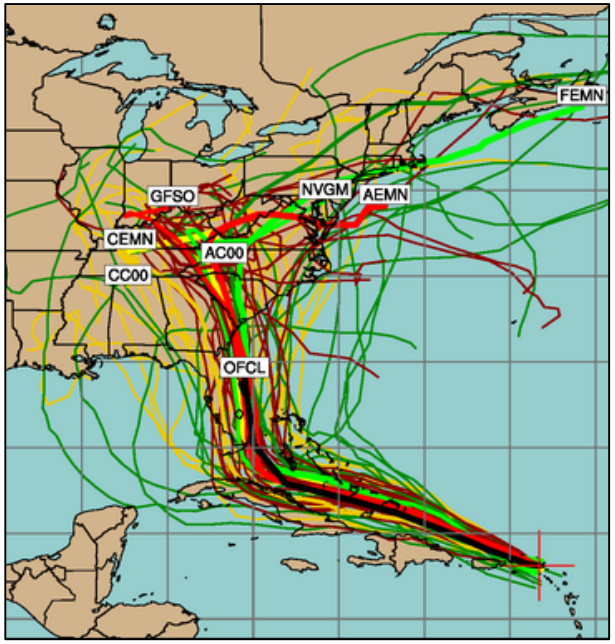
United States Environmental Protection Agency



# Consensus Exposure Predictions with the SEEM Framework

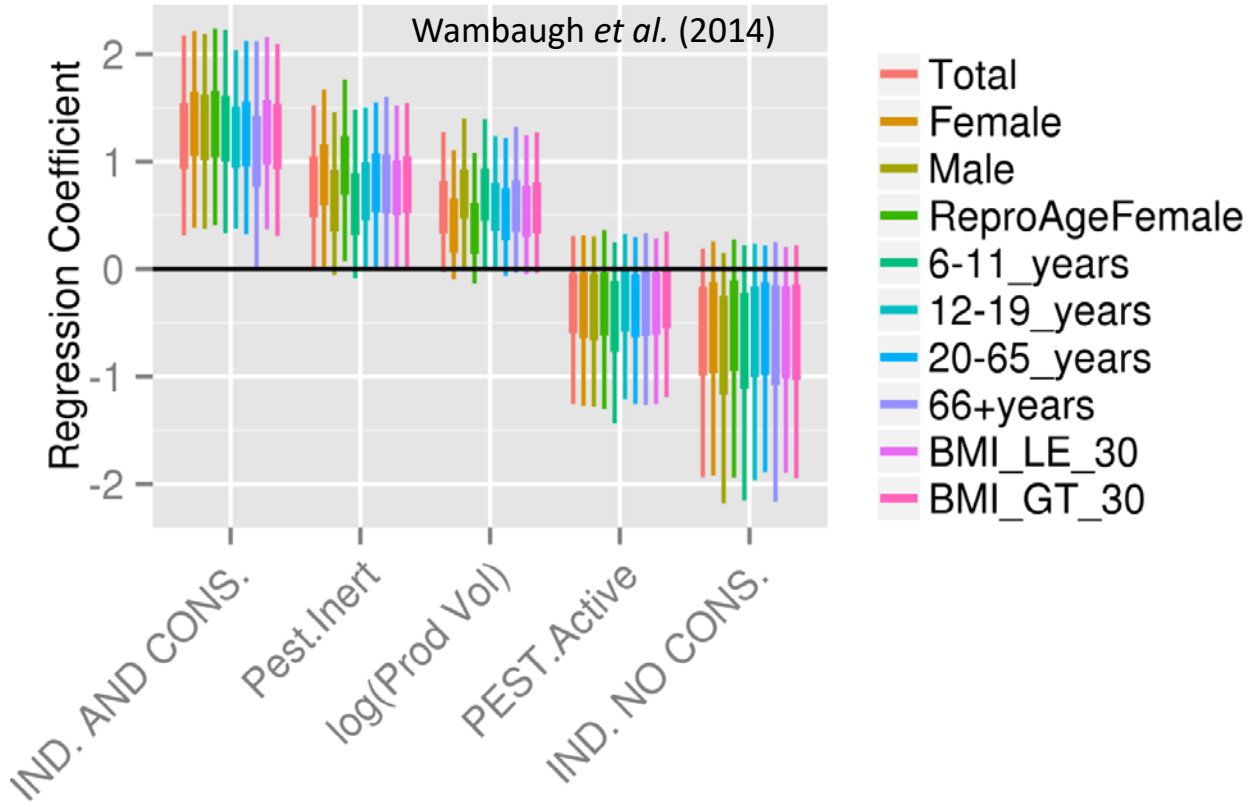
- We incorporate multiple models into consensus predictions for 1000s of chemicals within the **Systematic Empirical Evaluation of Models (SEEM) framework** (Wambaugh et al., 2013, 2014)
- We evaluate/calibrate predictions with available monitoring data
- This provides information similar to a sensitivity analysis: What models are working? What data are most needed? This is an iterative process.
- To date we have relied on median U.S. population exposure rates only




Integrating Multiple Models

# Heuristics of Exposure




Five descriptors explain roughly 50% of the chemical to chemical variability in median NHANES exposure rates

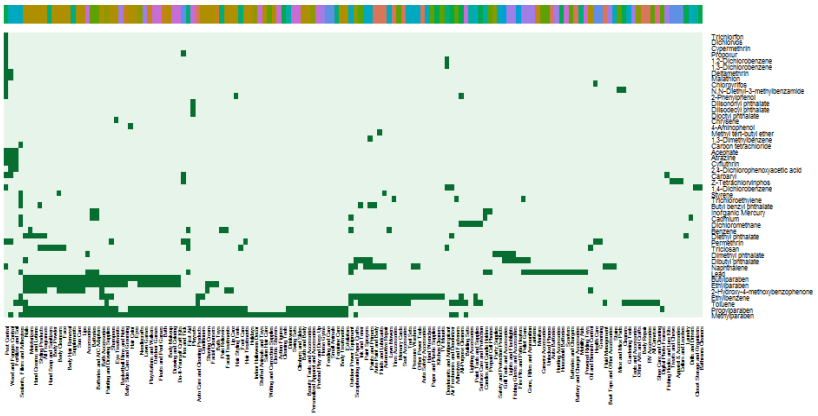
Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

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

# Chemical Use Identifies Relevant Pathways

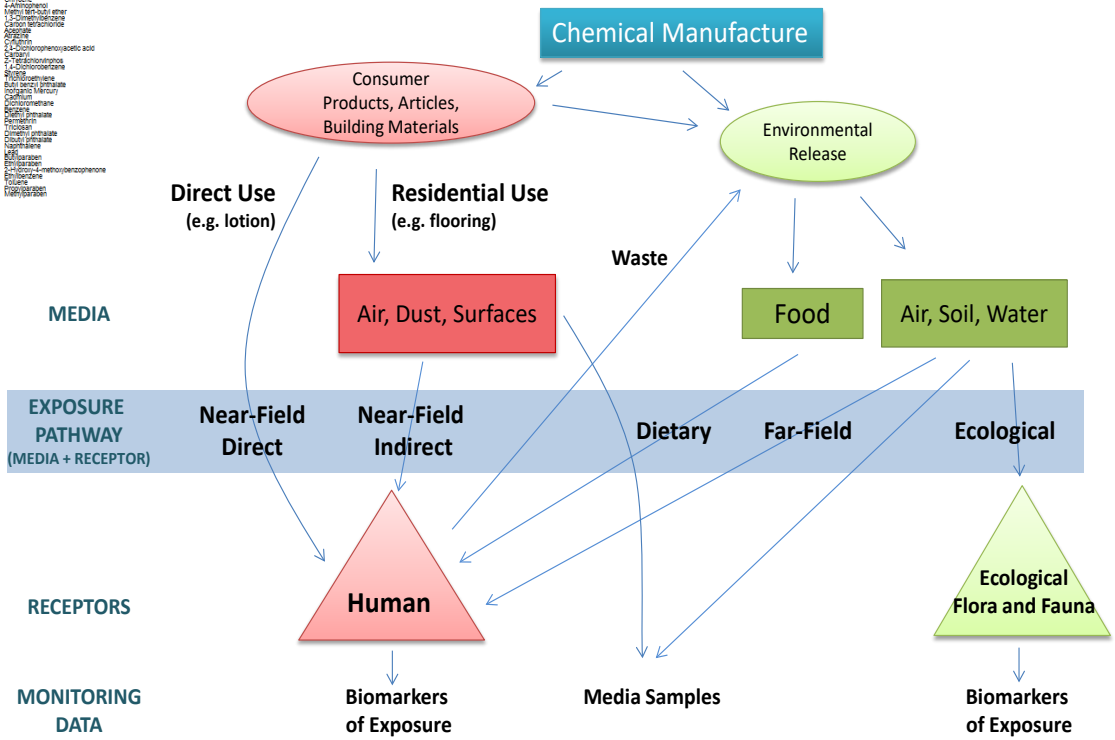
>2000 chemicals with Material Safety Data Sheets (MSDS) in CPCPdb (Goldsmith *et al.*, 2014)

106 NHANES Chemicals



- Apparel
- Auto and Tires
- Baby
- Beauty
- Craft and Party Supply
- Electronics
- Grocery
- Health
- Home
- Home Improvement
- Patio and Garden
- Pets
- Sports and Outdoors
- Toys

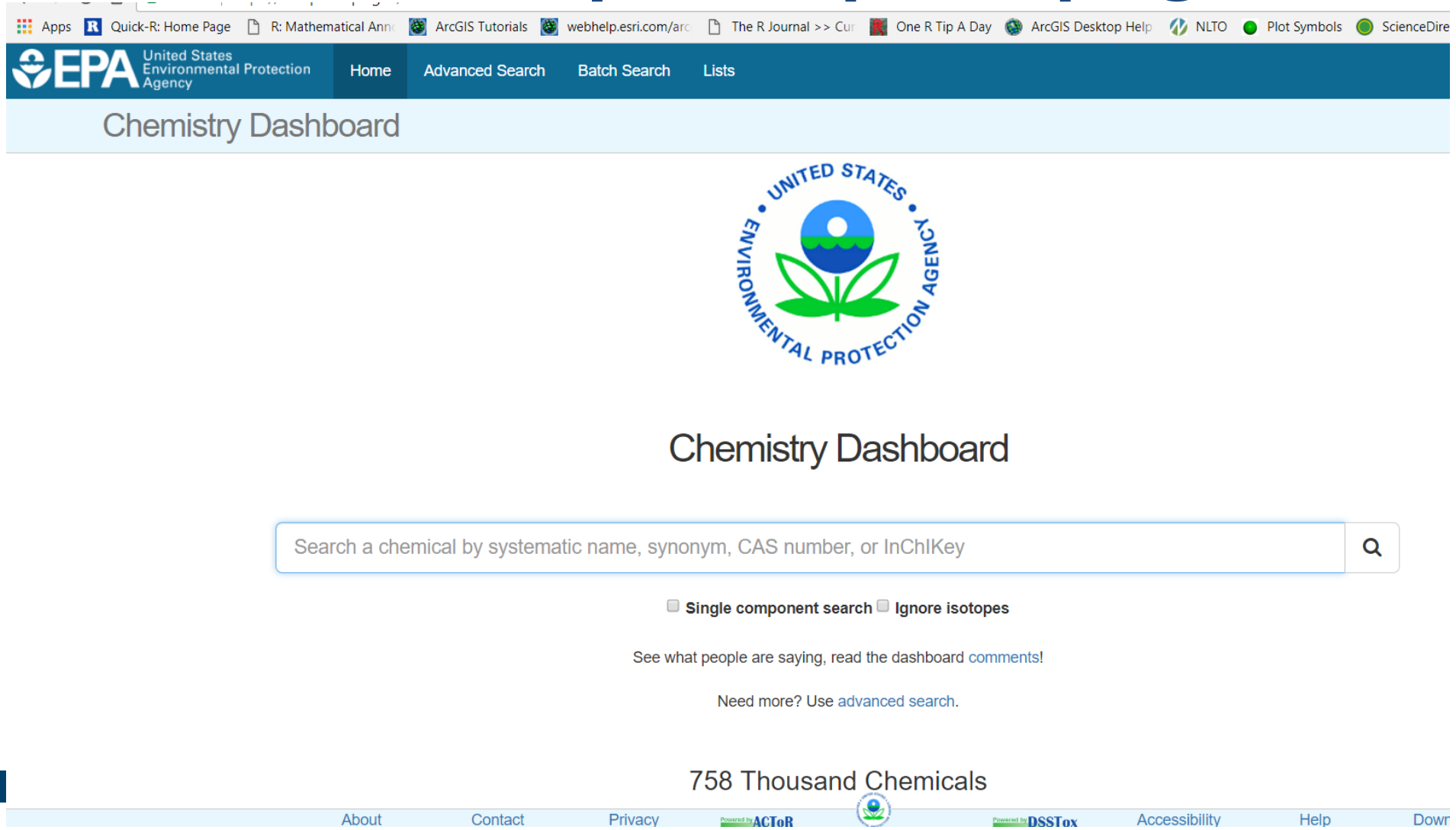
Some pathways have much higher average exposures!



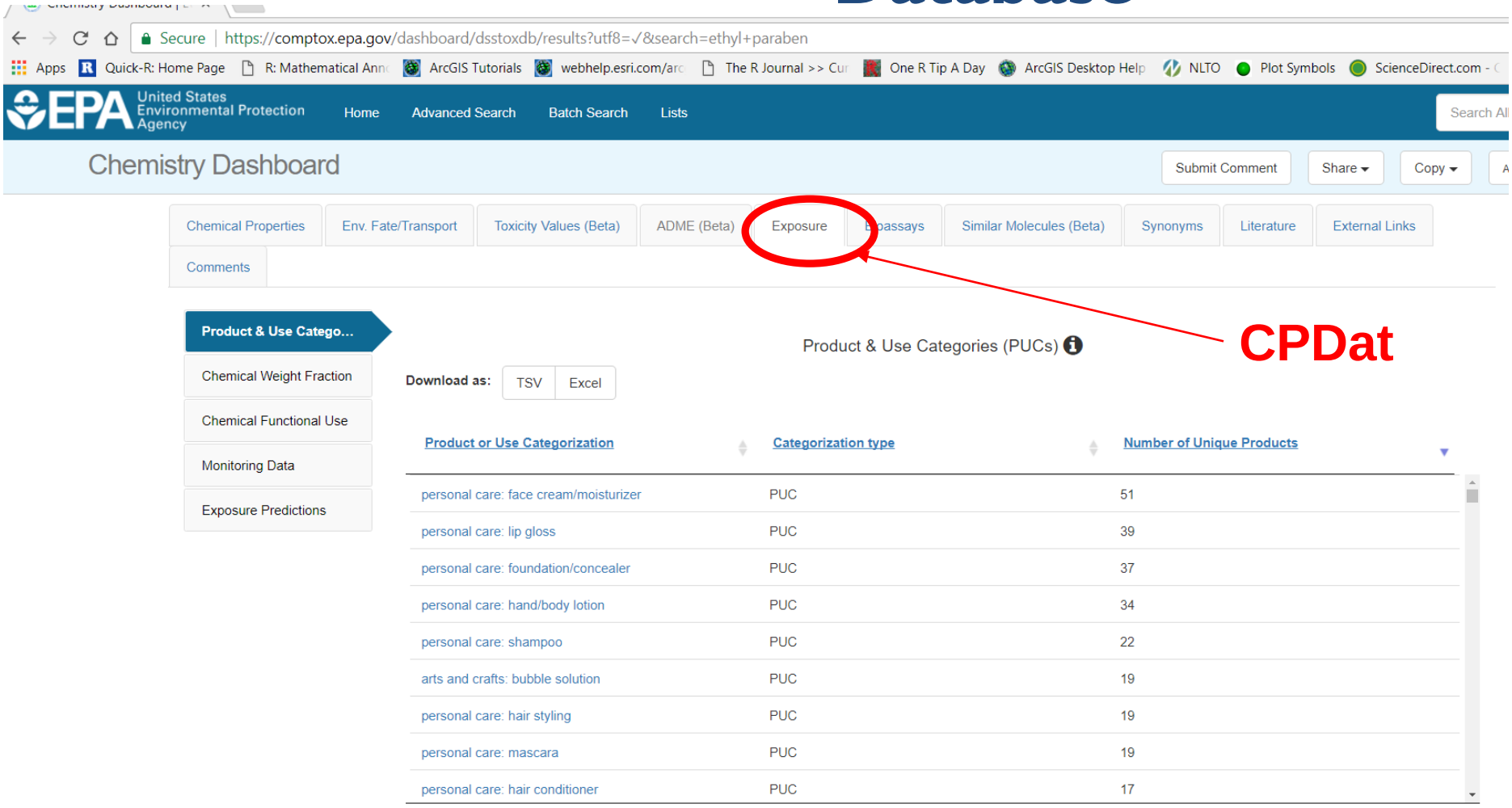



Near field sources have been known to be important at least since 1987 – see Wallace, *et al.*

# The Chemistry Dashboard
# http://comptox.epa.gov/

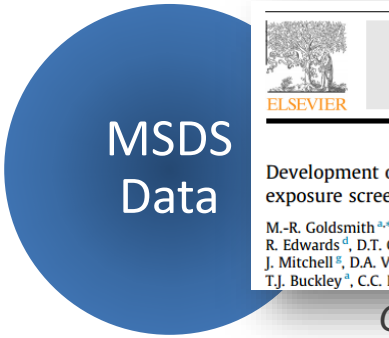


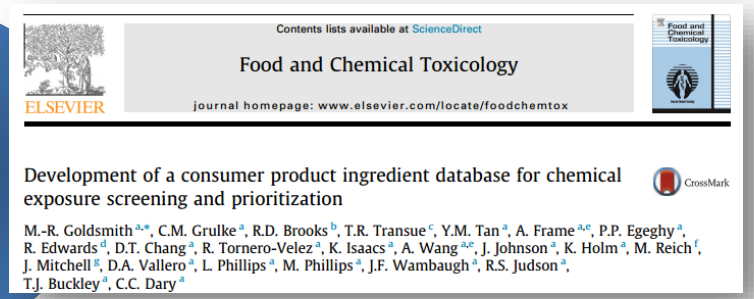
# Chemicals and Products Database



# 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,c], 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]

*Occurrence and quantitative chemical composition*


Contents lists available at ScienceDirect

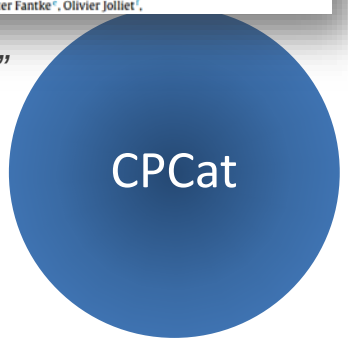
Toxicology Reports

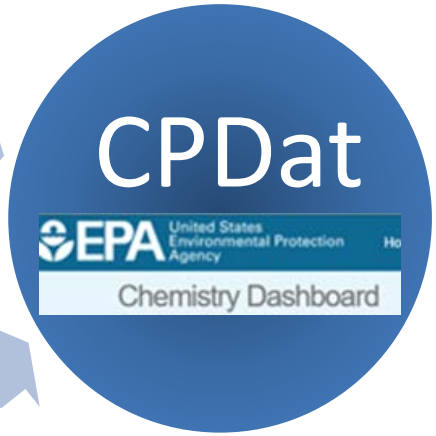
journal homepage: www.elsevier.com/locate/toxrep

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],

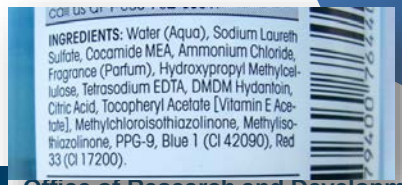
*Broad "index" of chemical uses*






*Occurrence data*


INGREDIENTS: Water (Aqua), Sodium Laureth Sulfate, Cocamide MEA, Ammonium Chloride, Fragrance (Parfum), Hydroxypropyl Methylcellulose, Tetrasodium EDTA, DMDM Hydantoin, Citric Acid, Tocopheryl Acetate [Vitamin E Acetate], Methylchloroisothiazolinone, Methylisothiazolinone, PPG-9, Blue 1 (CI 42090), Red 33 (CI 17200).


*Targeted and non-targeted measurement of chemicals in consumer products*

Also available as R Package

Slide from Kristin Isaacs

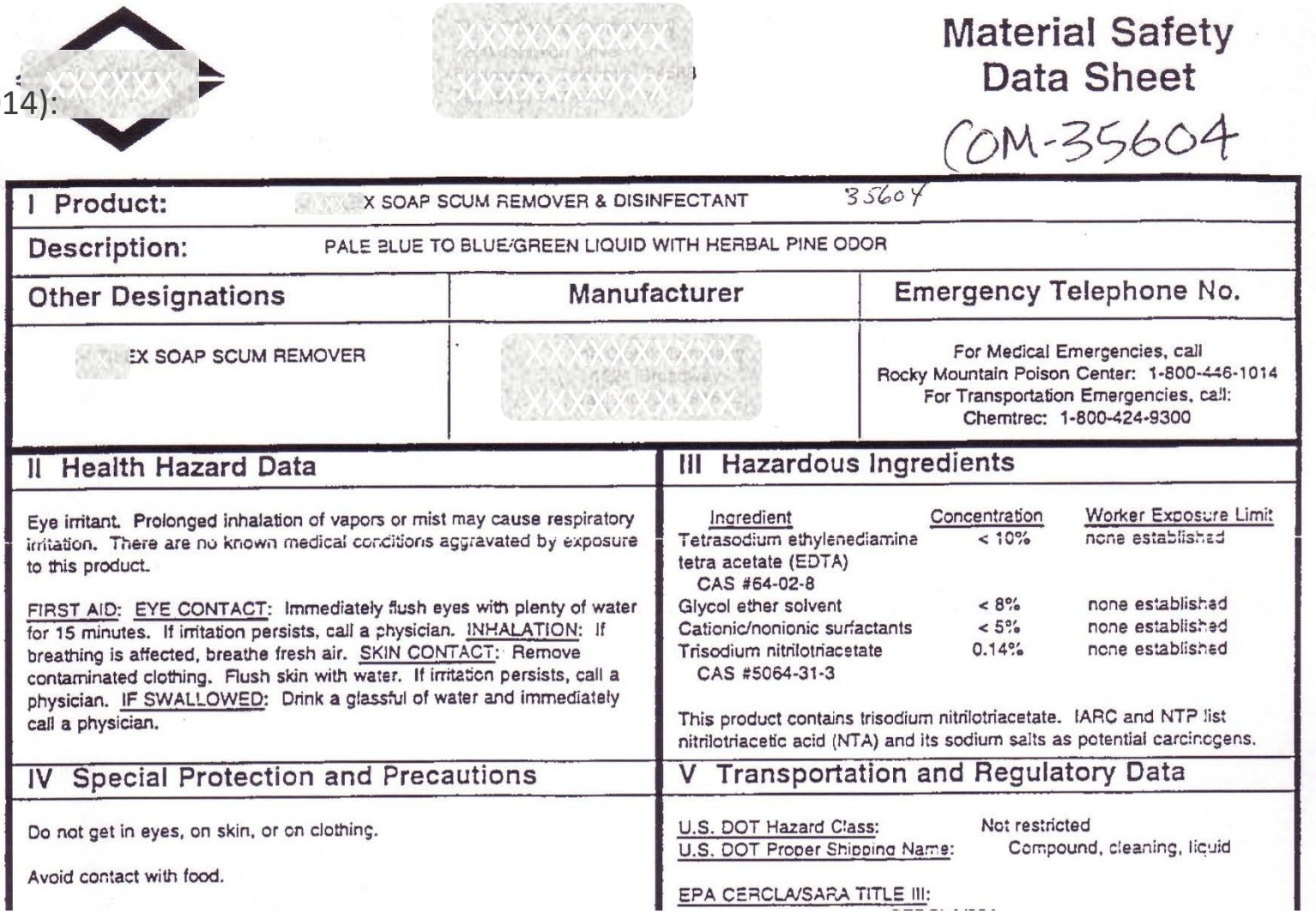
# Material Safety Data Sheets

United States Environmental Protection Agency

Goldsmith et al. (2014):

- ~20,000 product-specific Material Safety Data Sheets (MSDS) curated
- ~2,400 chemicals

Product-specific uses determined using web spider to click through categories (e.g., home goods, bath soaps, baby) to find each product



Material Safety Data Sheet

COM-35604

| I Product: | X SOAP SCUM REMOVER & DISINFECTANT 35604 | |
|---|---|---|
| Description: | PALE BLUE TO BLUE/GREEN LIQUID WITH HERBAL PINE ODOR | |
| Other Designations | Manufacturer | Emergency Telephone No. |
| EX SOAP SCUM REMOVER | | For Medical Emergencies, call Rocky Mountain Poison Center: 1-800-446-1014 For Transportation Emergencies, call: Chemtrec: 1-800-424-9300 |

**II Health Hazard Data**

Eye irritant. Prolonged inhalation of vapors or mist may cause respiratory irritation. There are no known medical conditions aggravated by exposure to this product.

FIRST AID: EYE CONTACT: Immediately flush eyes with plenty of water for 15 minutes. If irritation persists, call a physician. INHALATION: If breathing is affected, breathe fresh air. SKIN CONTACT: Remove contaminated clothing. Flush skin with water. If irritation persists, call a physician. IF SWALLOWED: Drink a glassful of water and immediately call a physician.

**III Hazardous Ingredients**

| Ingredient | Concentration | Worker Exposure Limit |
|---|---|---|
| Tetrasodium ethylenediamine tetra acetate (EDTA) CAS #64-02-8 | < 10% | none established |
| Glycol ether solvent | < 8% | none established |
| Cationic/nonionic surfactants | < 5% | none established |
| Trisodium nitrilotriacetate CAS #5064-31-3 | 0.14% | none established |

This product contains trisodium nitrilotriacetate. IARC and NTP list nitrilotriacetic acid (NTA) and its sodium salts as potential carcinogens.

**IV Special Protection and Precautions**

Do not get in eyes, on skin, or on clothing.

Avoid contact with food.

**V Transportation and Regulatory Data**

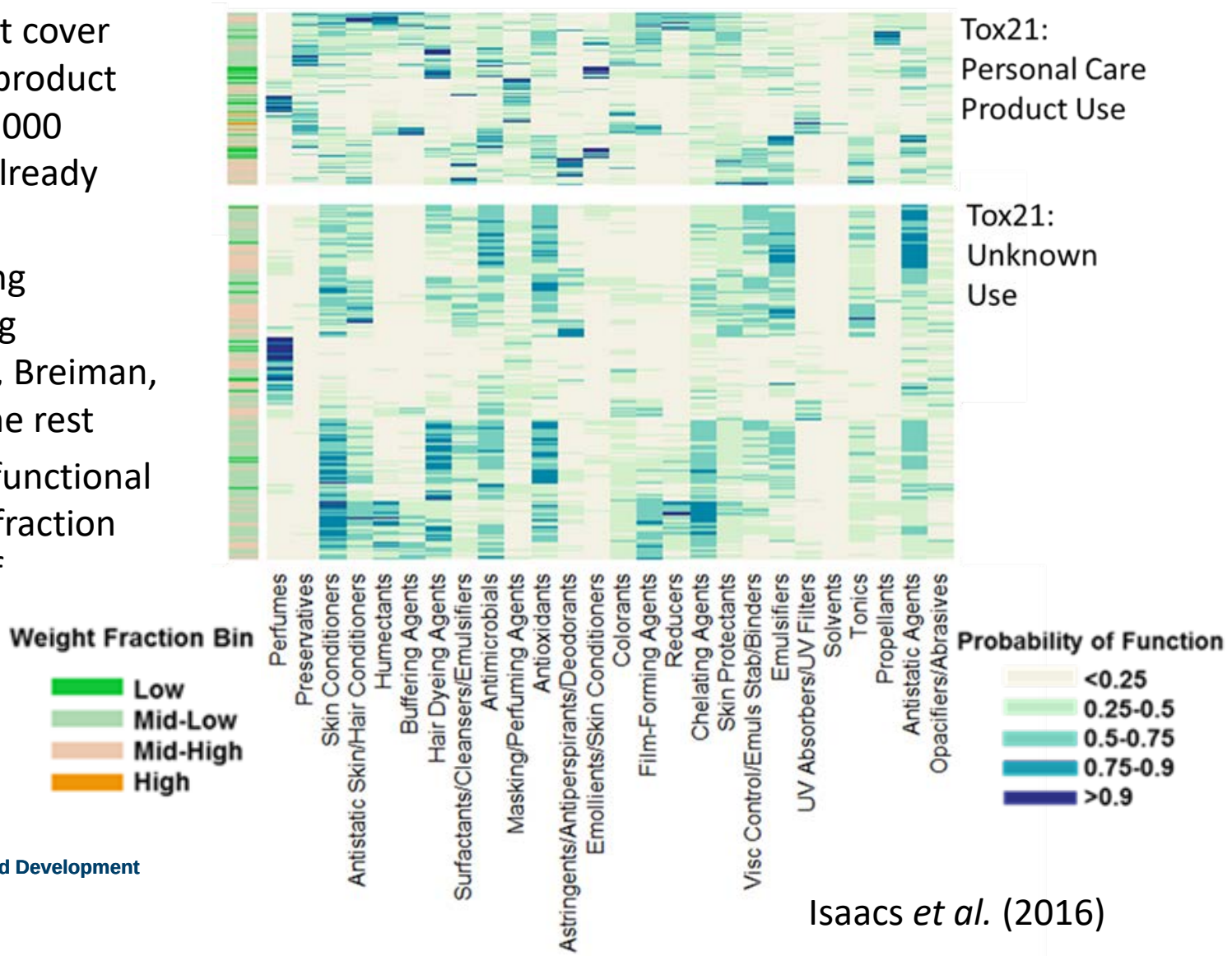
U.S. DOT Hazard Class: Not restricted

U.S. DOT Proper Shipping Name: Compound, cleaning, liquid

EPA CERCLA/SARA TITLE III:

# Predicting Chemical Constituents

- CPCPdb does not cover every chemical-product combination (~2000 chemicals, but already >8000 in Tox21)
- We are now using machine learning (Random Forest, Breiman, 2001) to fill in the rest
- We can predict functional use and weight fraction for thousands of chemicals

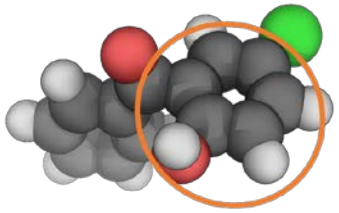


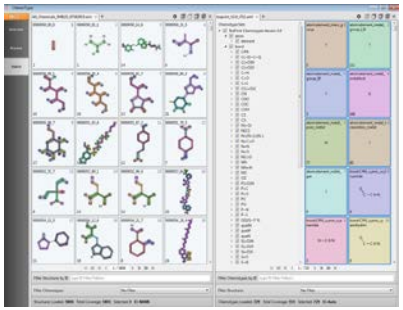

Isaacs *et al.* (2016)

# Predicting Function Based on Structure

Chemical Structure and Property Descriptors

Use Database (FUSE)






Prediction of Of Potential Alternatives from Chemical Libraries

**Random Forest Based Classification Models (Breiman, 2001)**

Each functional model evaluated on the basis of balanced accuracy, 5-fold CV, and Y-randomization classification errors

Phillips et al. (2017)

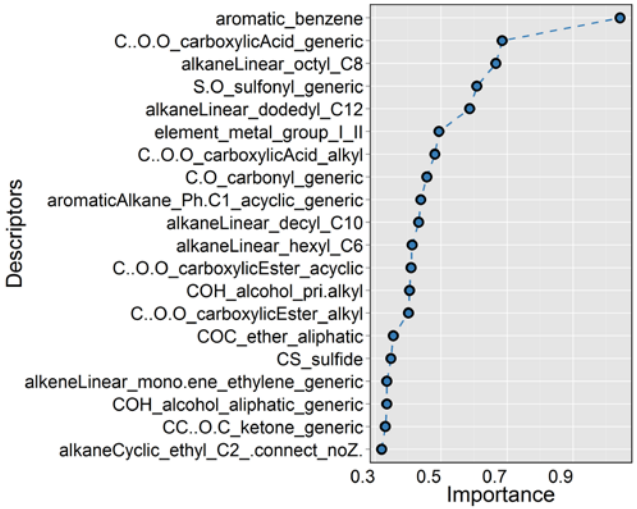
# Understanding Use Predictions

United States Environmental Protection Agency

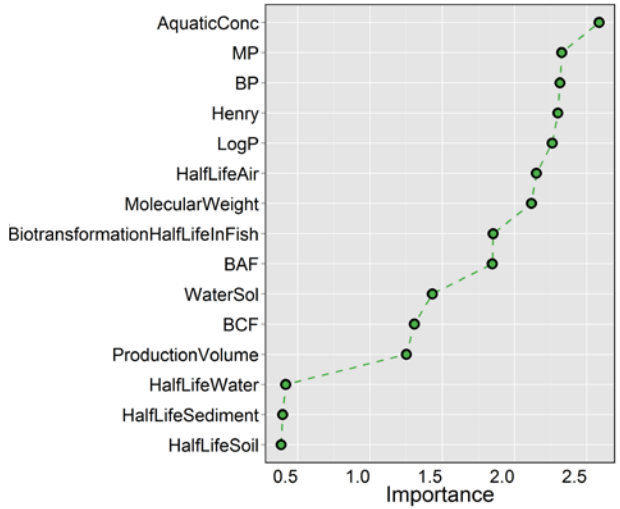
- Each functional model evaluated on the basis of balanced accuracy, 5-fold CV, and Y-randomization classification errors

**Random Forest Importance for Viscosity Controller Functional Use (Failed Model)**

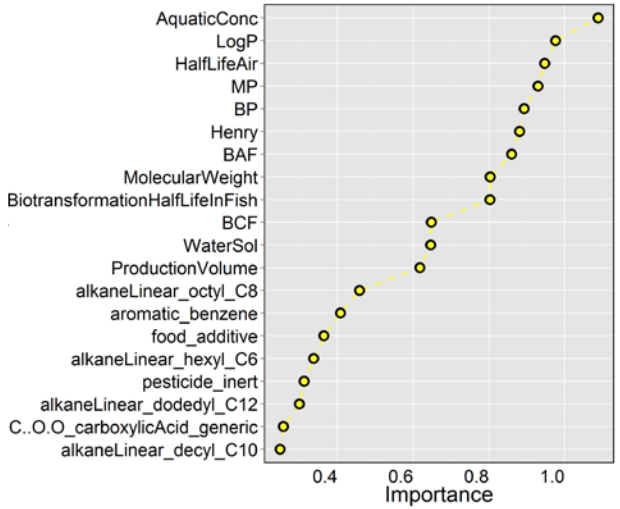
ToxPrint




EPI Suite




ToxPrint + EPI Suite

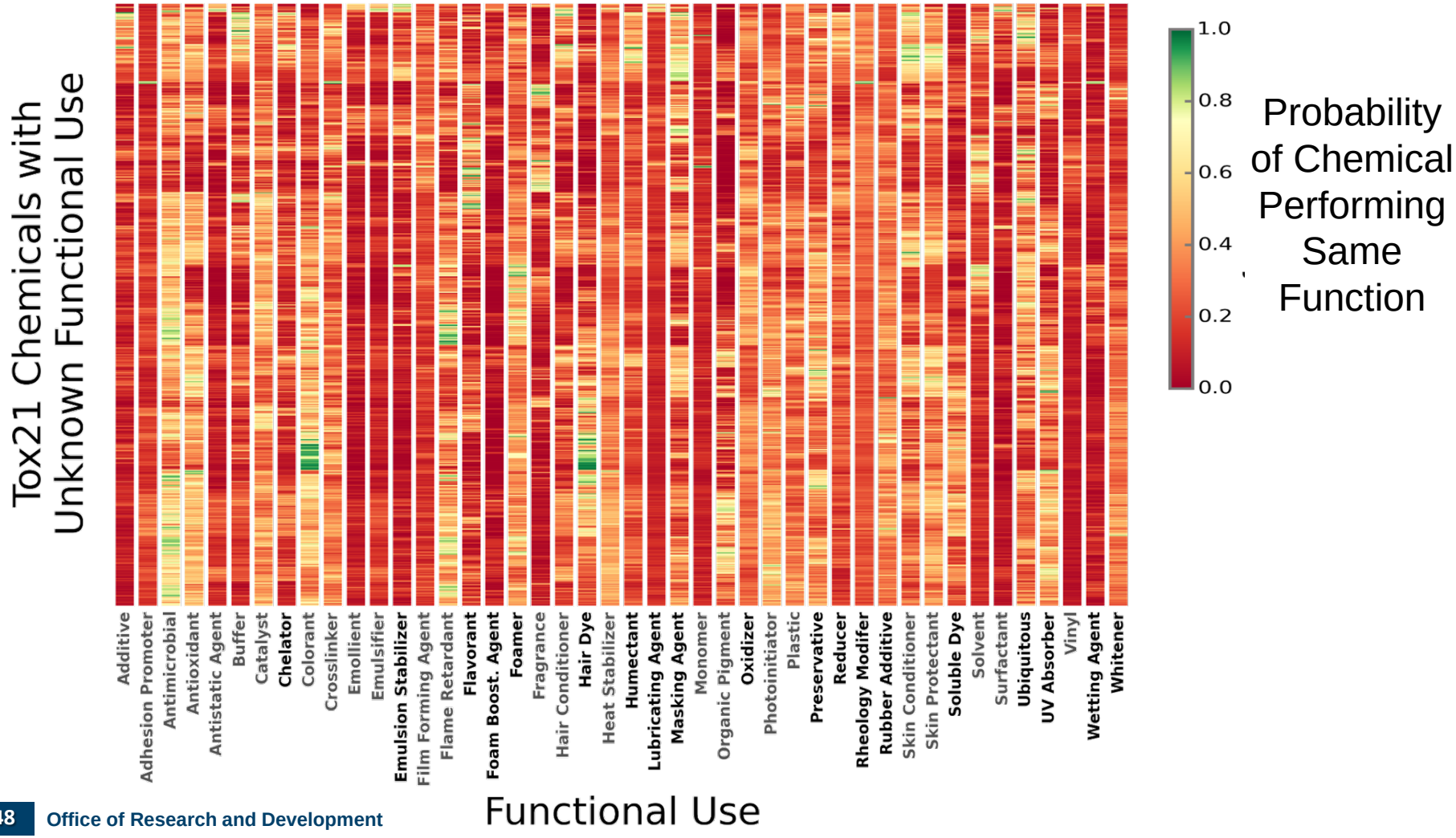



Viscosity controllers can be used to ***thicken*** or ***thin out*** mixtures of chemicals..



Phillips et al. (2017)

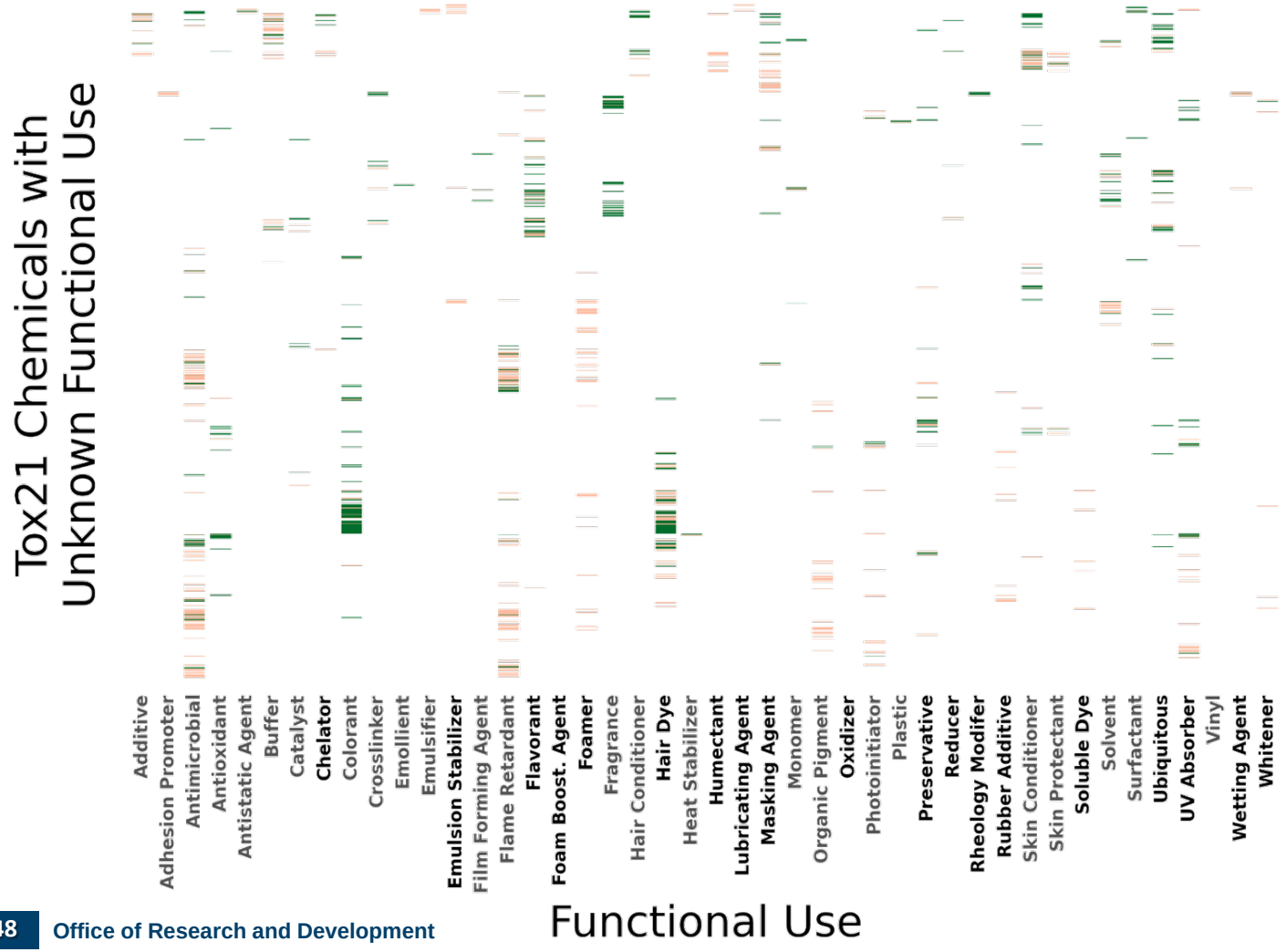
# Screening for Alternatives By Function and Bioactivity






Phillips et al. (2017)

# Screening for Alternatives By Function and Bioactivity






Phillips et al. (2017)

# Non-Targeted and Suspect-Screening Analysis

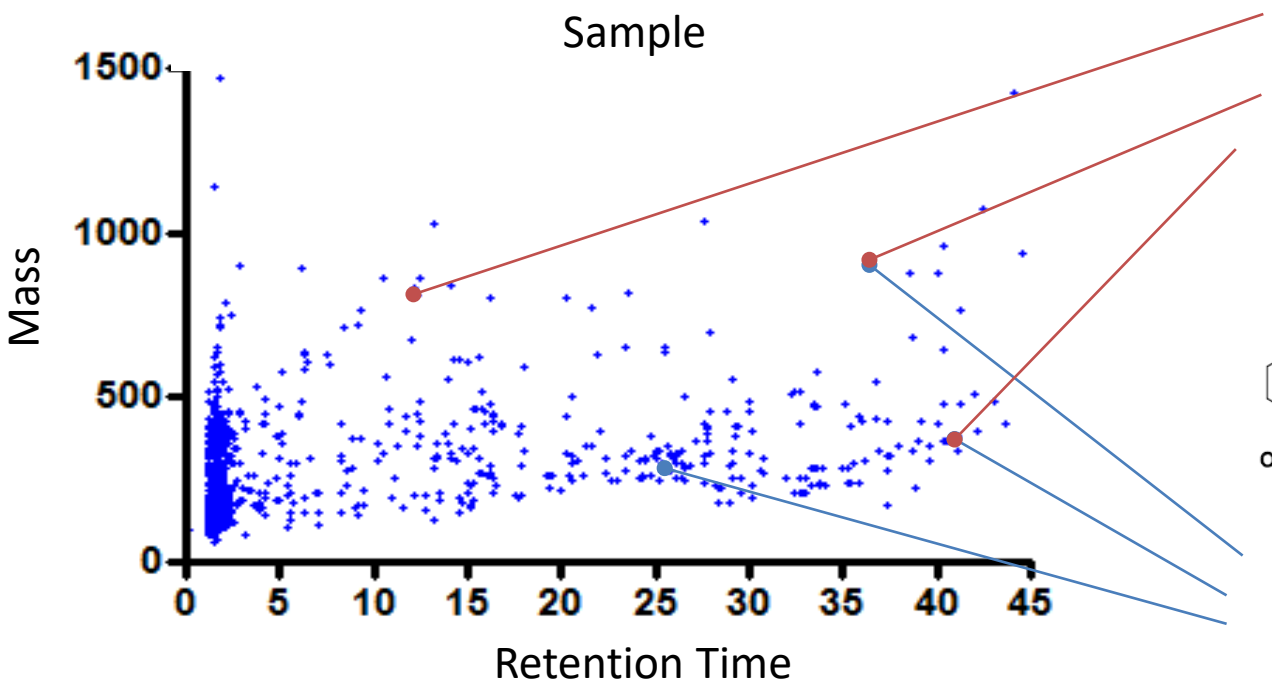
- Models present one way forward, but new analytic techniques may also allow insight in to chemicals composition of products and the greater environment
- EPA is coordinating a comparison of non-targeted screening workflows used by leading academic and government groups (led by Jon Sobus and Elin Ulrich)
  - Examining house dust, human plasma, and silicone wristbands (O'Connell, et al., 2014)
  - Similar to NORMAN Network (Schymanski et al., 2015) analysis of water
- Published analysis on house dust (Rager et al., 2016)

*"I'm searching for my keys."*

- 100 consumer products from a major U.S. retailer were analyzed, tentatively identifying 1,632 chemicals, 1,445 which were not in EPA's database of consumer product chemicals (Phillips *et al., submitted*)

# Suspect Screening Example: House Dust

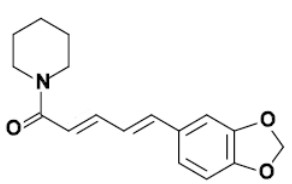
947 Peaks in an American Health Homes Dust Sample

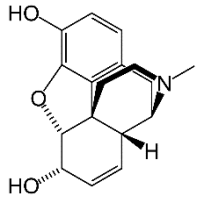



Each peak corresponds to a chemical with an accurate mass and predicted formula:

$C_{17}H_{19}NO_3$

Multiple chemicals can have the same mass and formula:




Is chemical A present, chemical B, both, or some other chemical (neither)?

We are expanding our reference libraries using ToxCast chemicals to enable greater numbers and better accuracy of confirmed chemicals



See Rager et al., (2016)

# 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."***

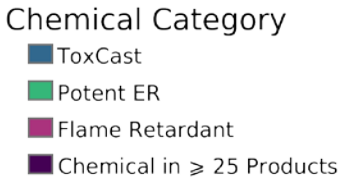
Hahn and Strassmann (1938)

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

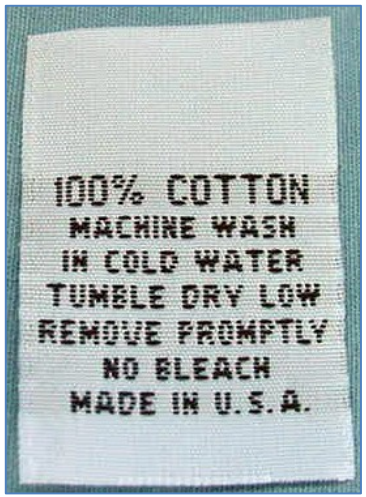
# Measuring Chemicals in Household Items

United States Environmental Protection Agency






The chemicals found in a cotton shirt



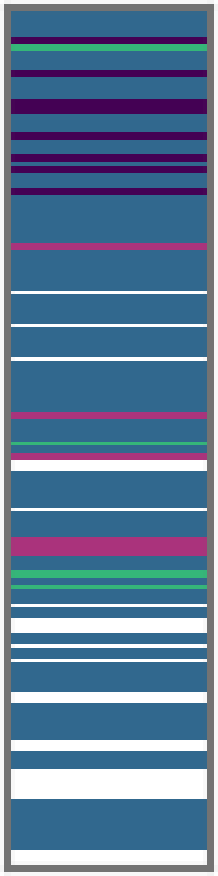



Phillips et al. (submitted)

# Measuring Chemicals in Household Items

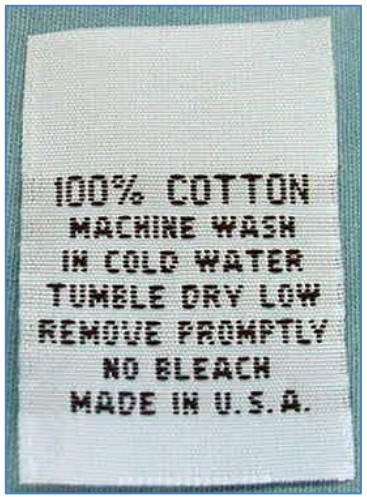
United States Environmental Protection Agency


126 Confirmed



⬅ Chemicals that are present

Chemical Category
- ToxCast
- Potent ER
- Flame Retardant
- Chemical in ⩾ 25 Products



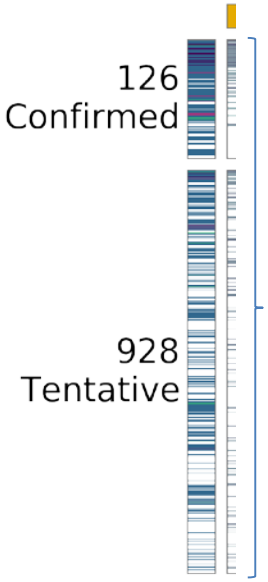

⬅ Chemicals that are absent (but found in other products)



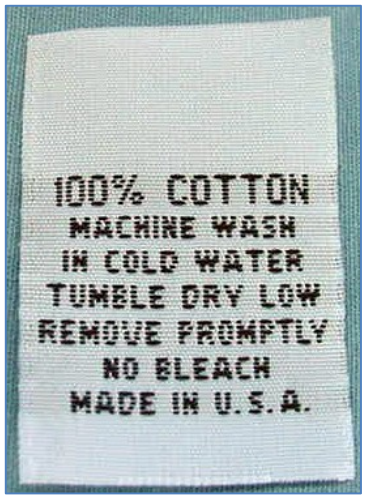
Phillips et al. (submitted)

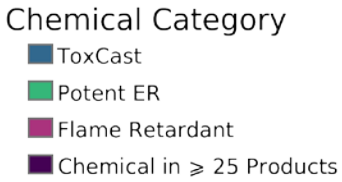
# Measuring Chemicals in Household Items

United States Environmental Protection Agency


The chemicals found in a cotton shirt



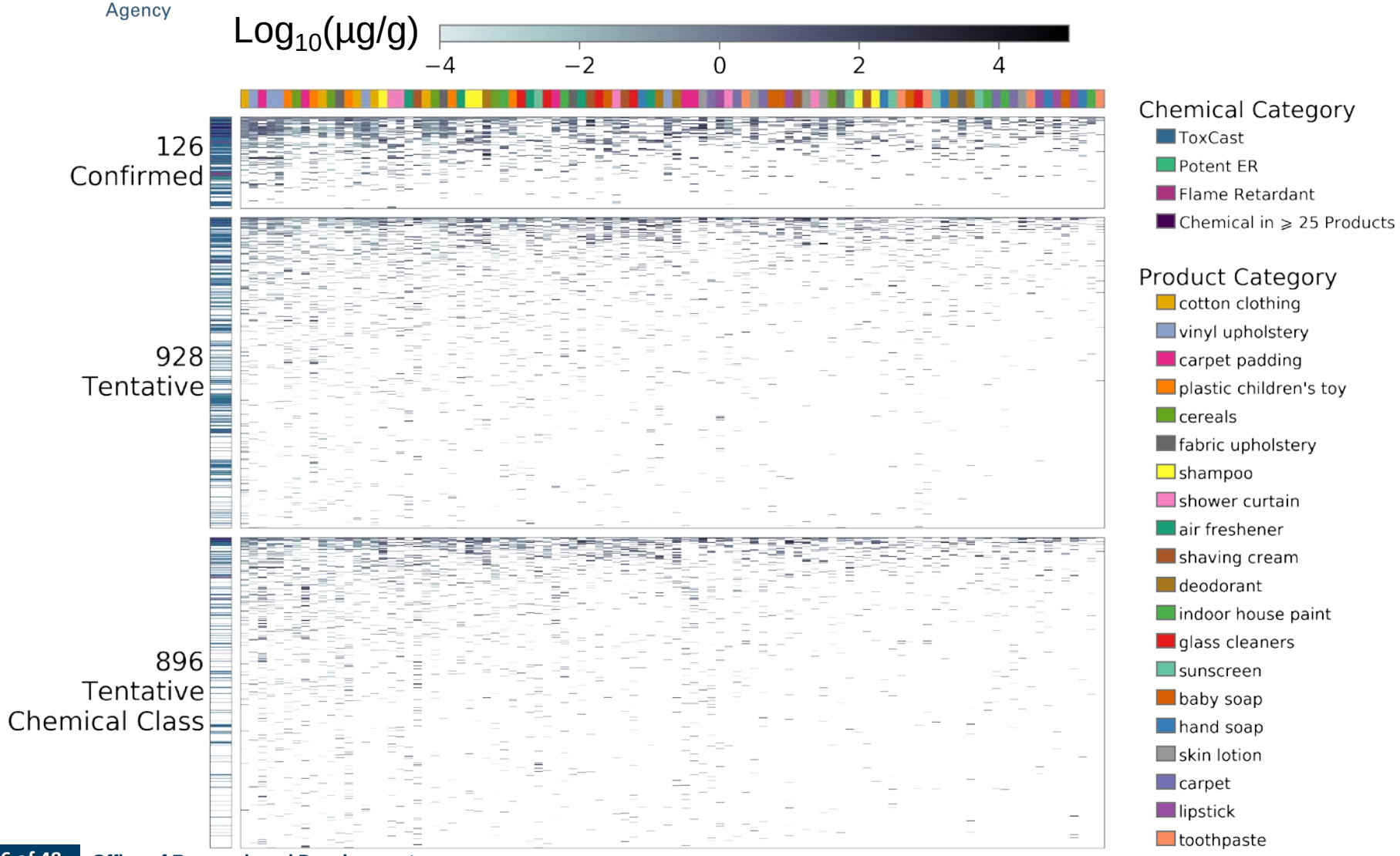






Phillips et al. (submitted)

# Measuring Chemicals in Household Items

United States Environmental Protection Agency

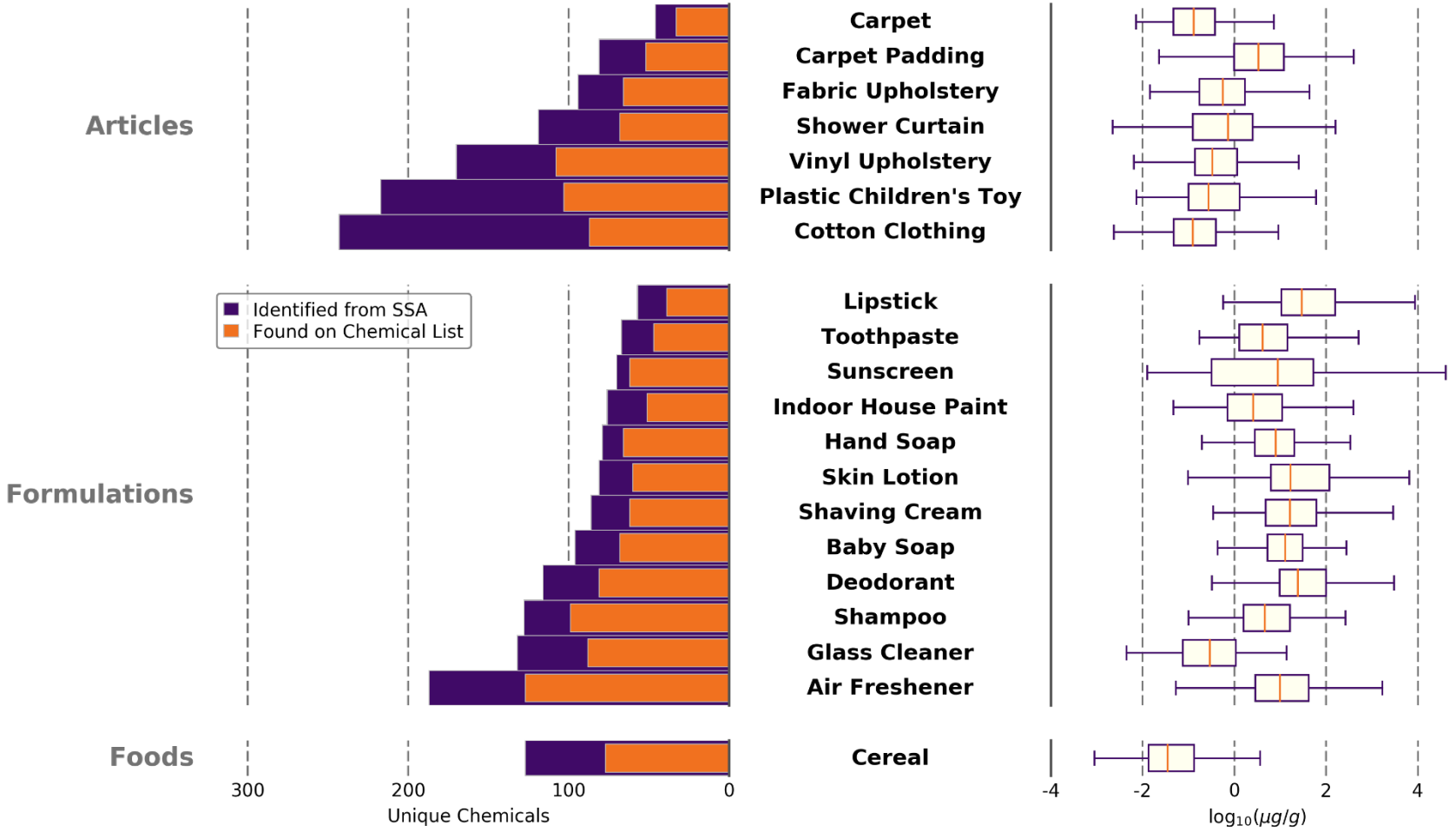





Phillips et al. (submitted)

# Product Scan Summary

Of 1,632 chemicals confirmed or tentatively identified, 1,445 were not present in CPCPdb (Goldsmith, et al., 2015)

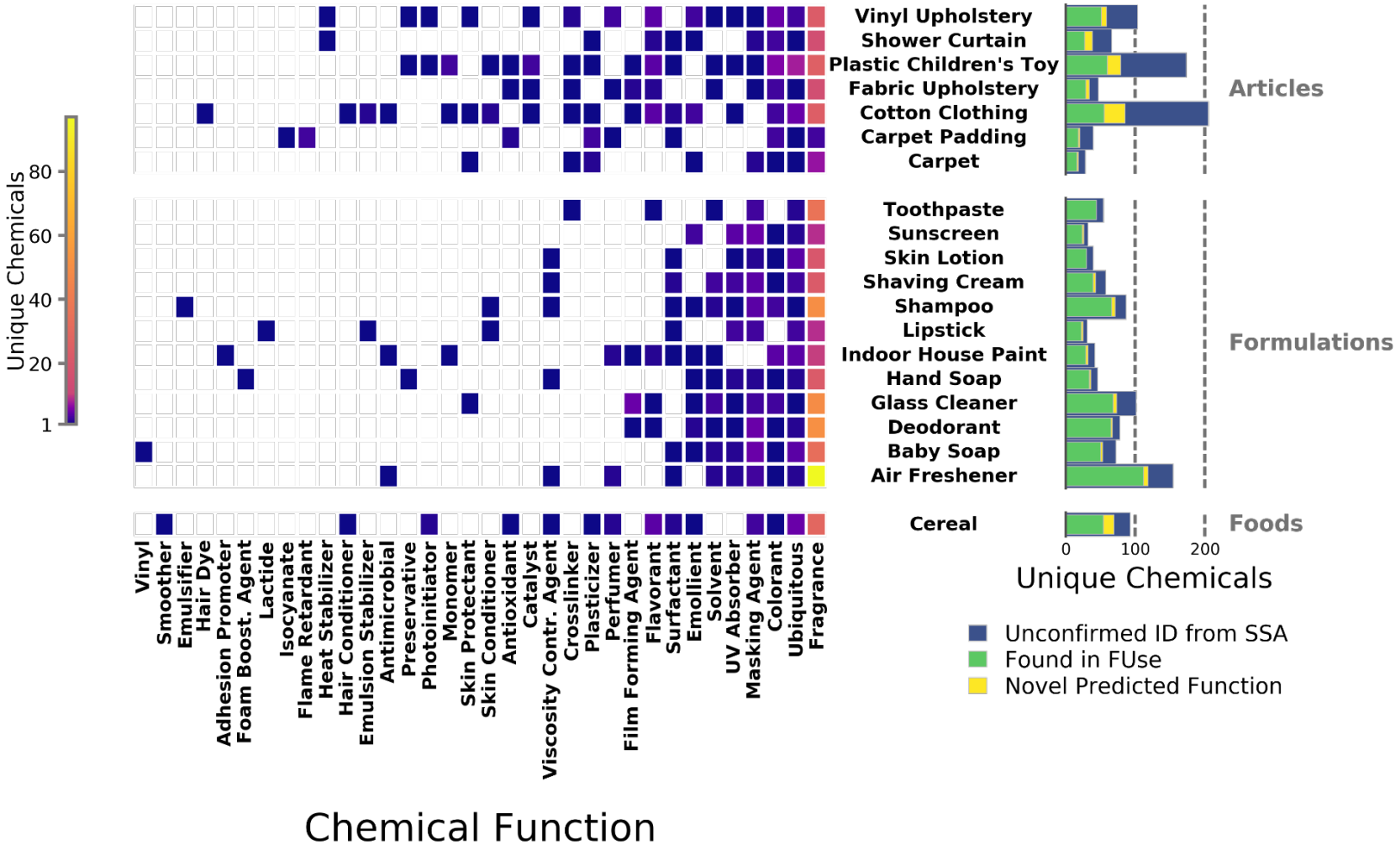





Phillips et al. (submitted)

# Predicting Chemical Function

Using the methods of Phillips *et al.*, (2017):




Phillips et al. (submitted)

# Caveats to Non-Targeted Screening

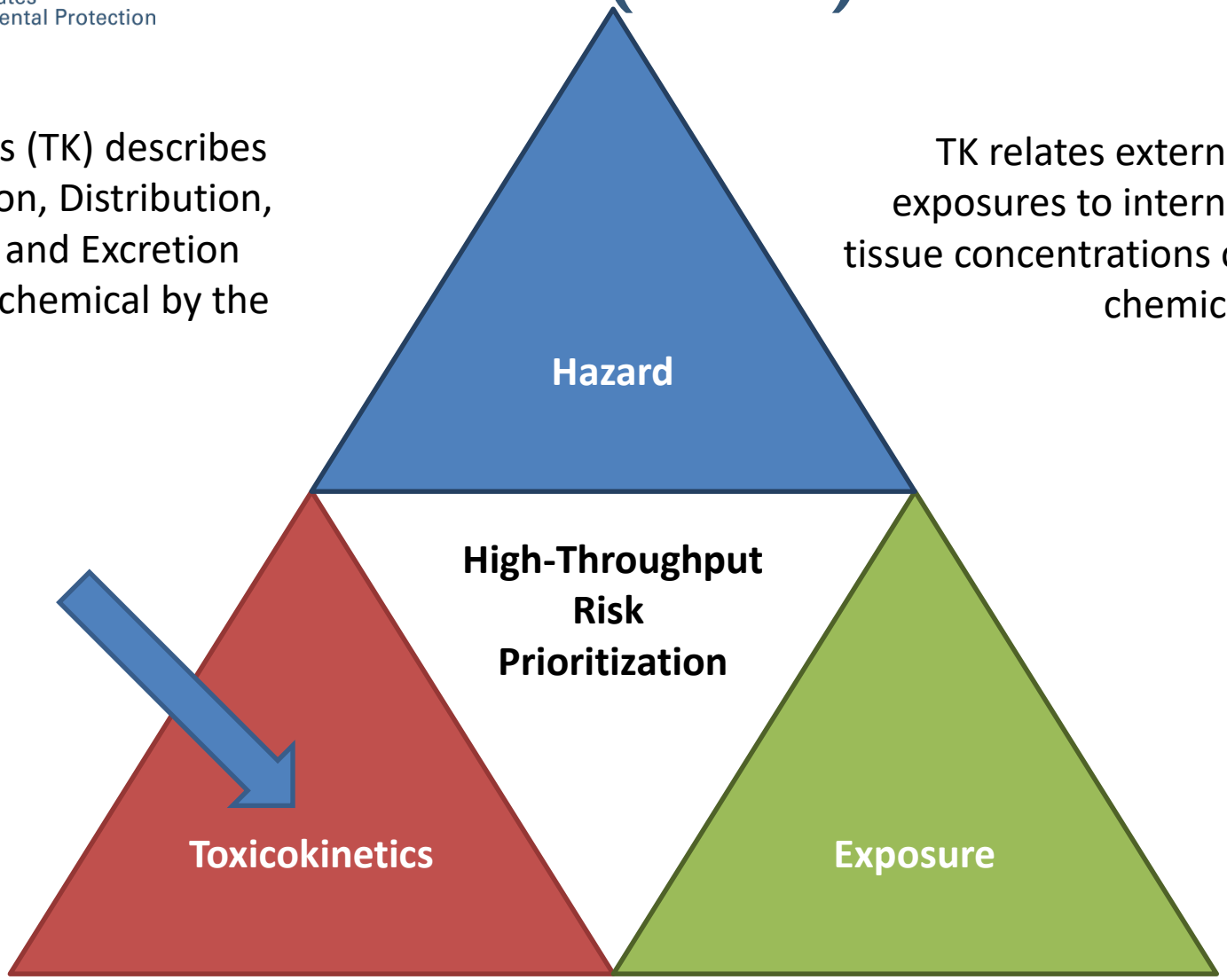
- **Chemical presence in an object does not mean that exposure occurs**
- **Only some chemical identities are confirmed, *most are tentative***
  - Can use formulation predictor models as additional evidence
- **Chemical presence in an object does not necessarily mean that it is bioavailable**
  - Can build emission models
- **Small range for quantitation leads to underestimation of concentration**
- **Product de-formulation caveats:**
  - Samples are being homogenized (e.g., grinding) and are extracted with a solvent (dichloro methane, DCM)
  - Only using one solvent (DCM, polar) and one method GCxGC-TOF-MS
  - Varying exposure intimacy, from carpet padding to shampoo to cereal
- **Exposure alone is not risk, need hazard data**

# High Throughput Toxicokinetics (HTTK)

United States Environmental Protection Agency

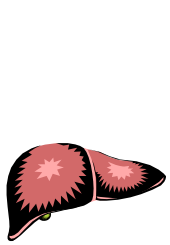
Toxicokinetics (TK) describes the Absorption, Distribution, Metabolism, and Excretion (ADME) of a chemical by the body

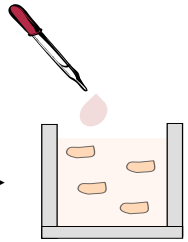
TK relates external exposures to internal tissue concentrations of chemical



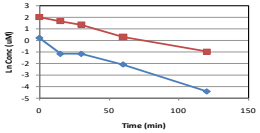
# High-Throughput Toxicokinetics (HTTK)

- **Most chemicals do not have TK data** – we use *in vitro* HTTK methods adapted from pharma to fill gaps
- In drug development, HTTK methods estimate therapeutic doses for clinical studies – predicted concentrations are typically on the order of values measured in clinical trials (Wang, 2010)





Human Hepatocytes (10 donor pool)



Intrinsic Hepatic Clearance ($Cl_{int}$)

**Measurements require chemical-specific methods for concentration**

*In Vitro - In Vivo* Extrapolation

Predicted Plasma Concentrations

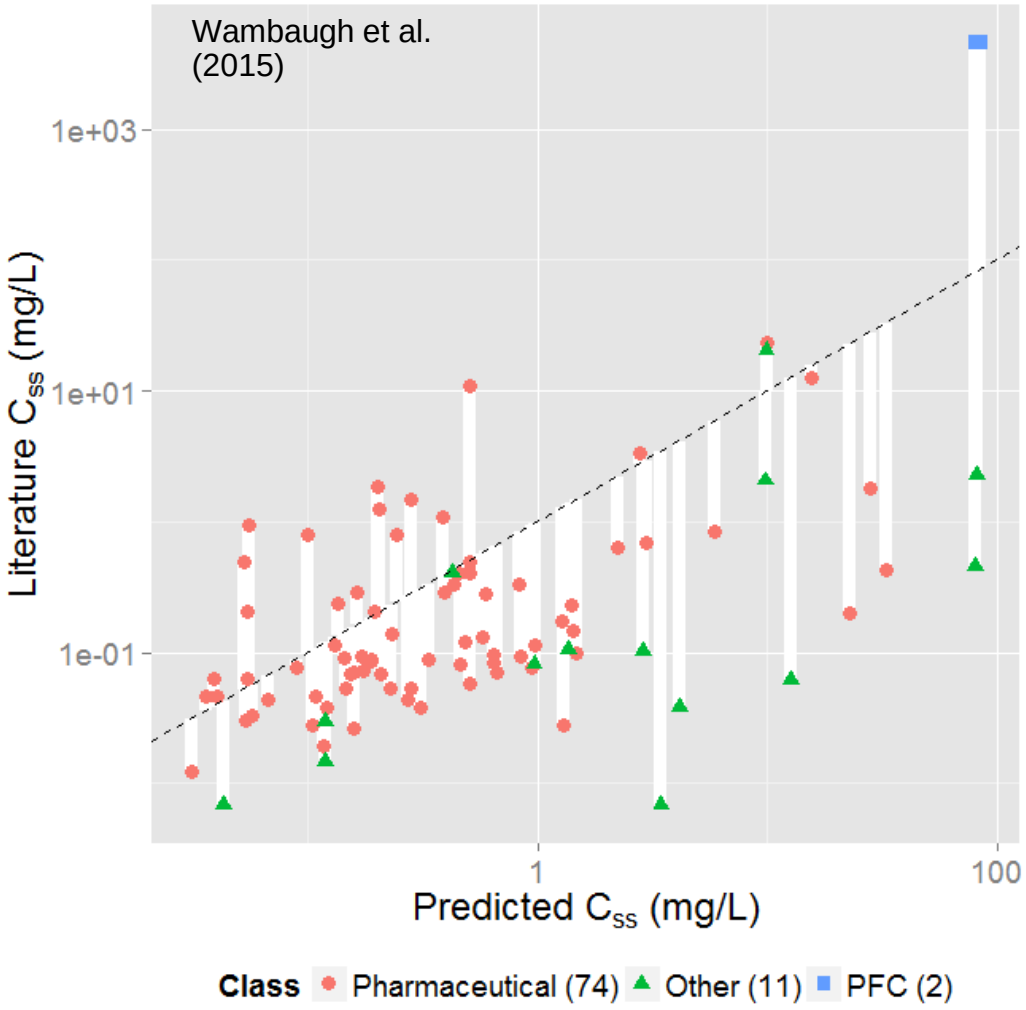
Human Plasma (6 donor pool)

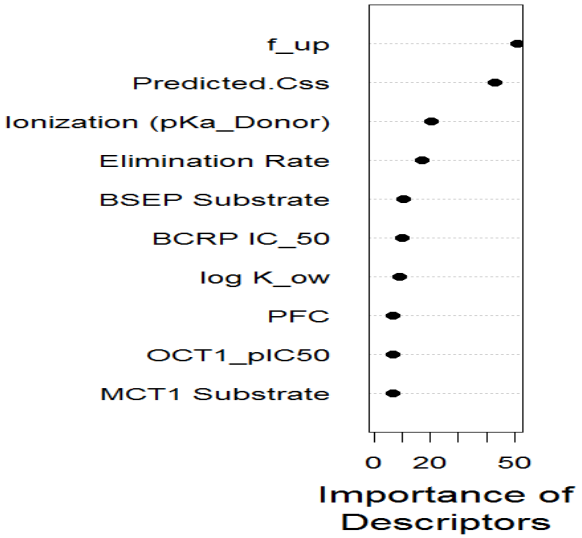
Plasma Protein Binding ($f_{up}$)

Rotroff et al. (2010) 35 chemicals
Wetmore et al. (2012) +204 chemicals
Wetmore et al. (2015) +163 chemicals

Figure from Barbara Wetmore

# Evaluating Predictions of Steady-State Plasma Concentration ($C_{ss}$)



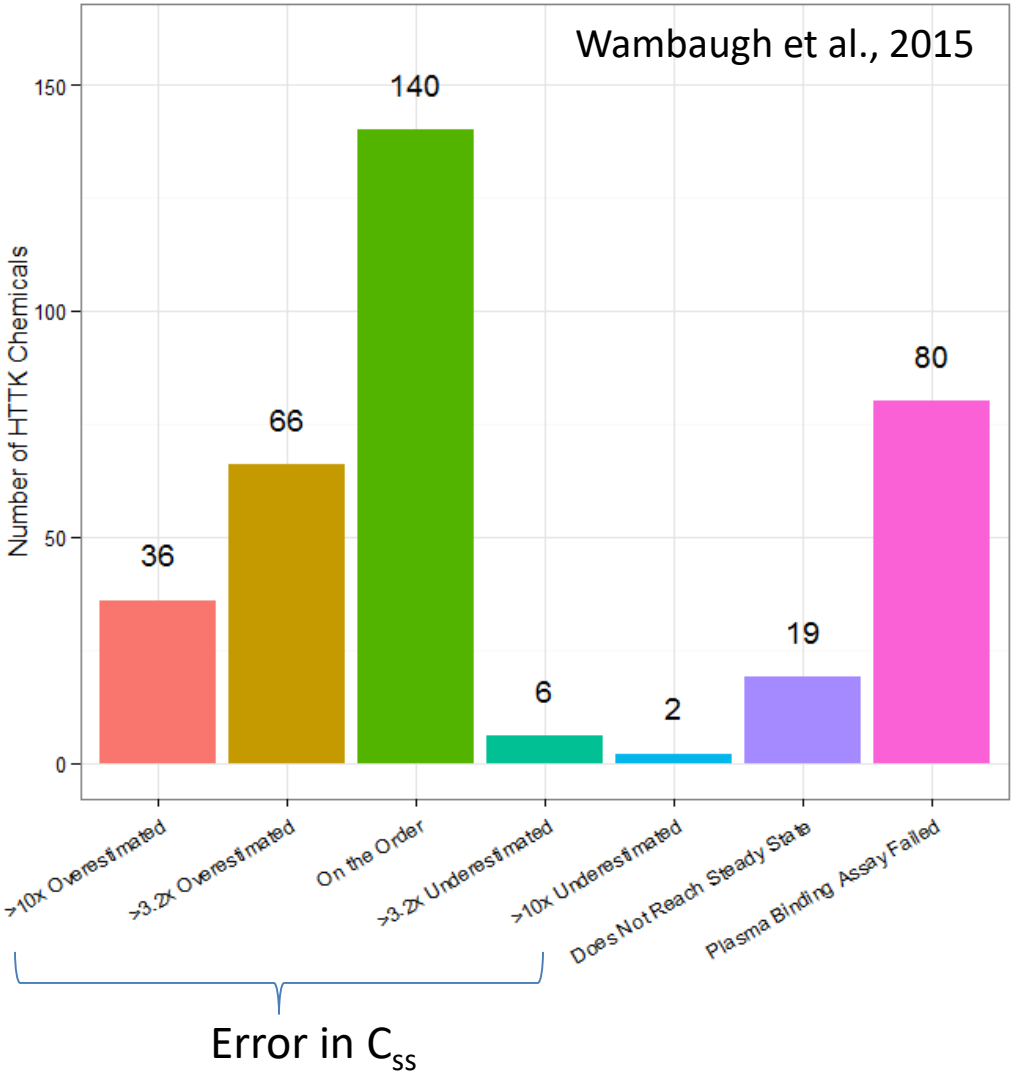




- When we compare the $C_{ss}$ predicted from *in vitro* HTTK with *in vivo* $C_{ss}$ values determined from the literature we find limited correlation ($R^2$ ~0.34)
- The dashed line indicates the identity (perfect predictor) line:
  - Over-predict for 65
  - Under-predict for 22
- The white lines indicate the discrepancy between measured and predicted values (the residual)

# Predicting Error in HTTK Predictions

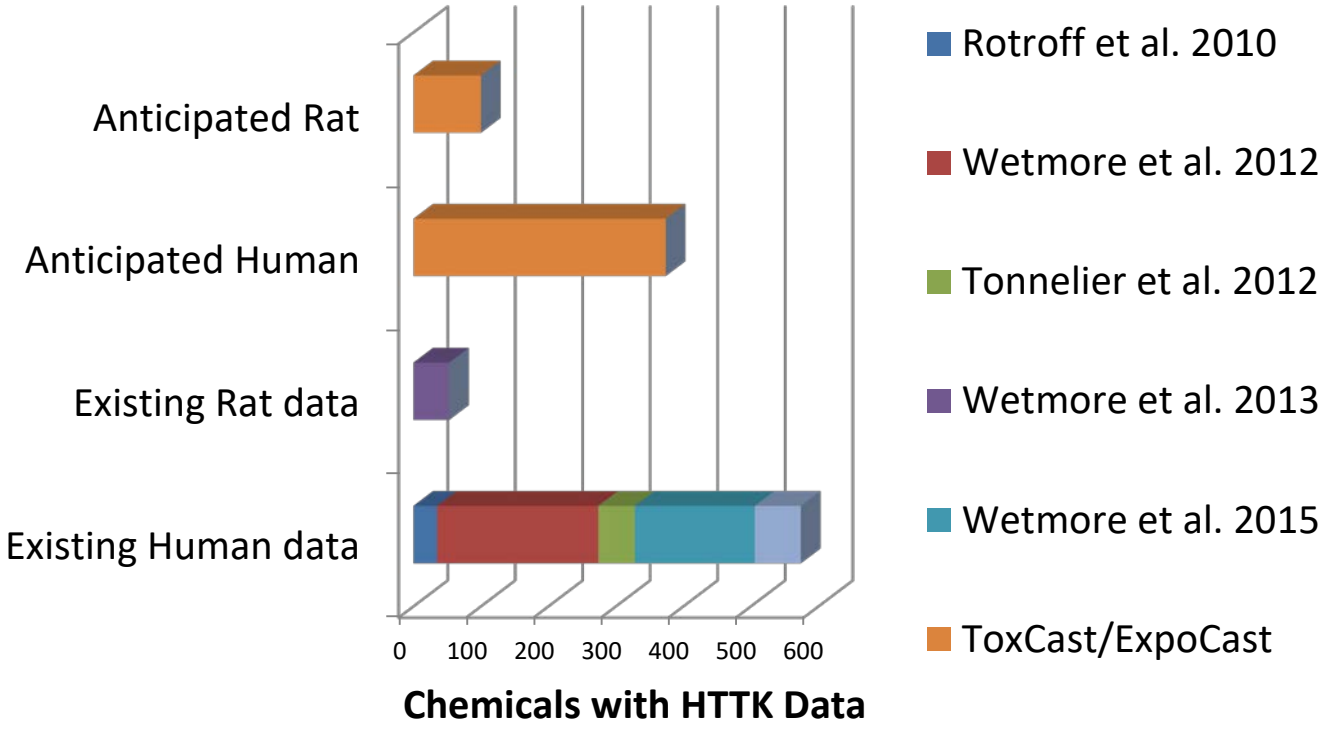
United States Environmental Protection Agency

- For most compounds in the environment there will be no clinical trials
- Uncertainty must be well characterized
  - We compare to *in vivo* data to get **empirical estimates of HTTK uncertainty**
  - Any approximations, omissions, or mistakes should work to increase the estimated uncertainty when evaluated systematically across chemicals
- Through comparison to *in vivo* data, a cross-validated (Random Forest, Breiman, 2001) predictor of success or failure of HTTK has been constructed
- We also have categories for chemicals that do not reach steady-state or for which plasma binding assay fails



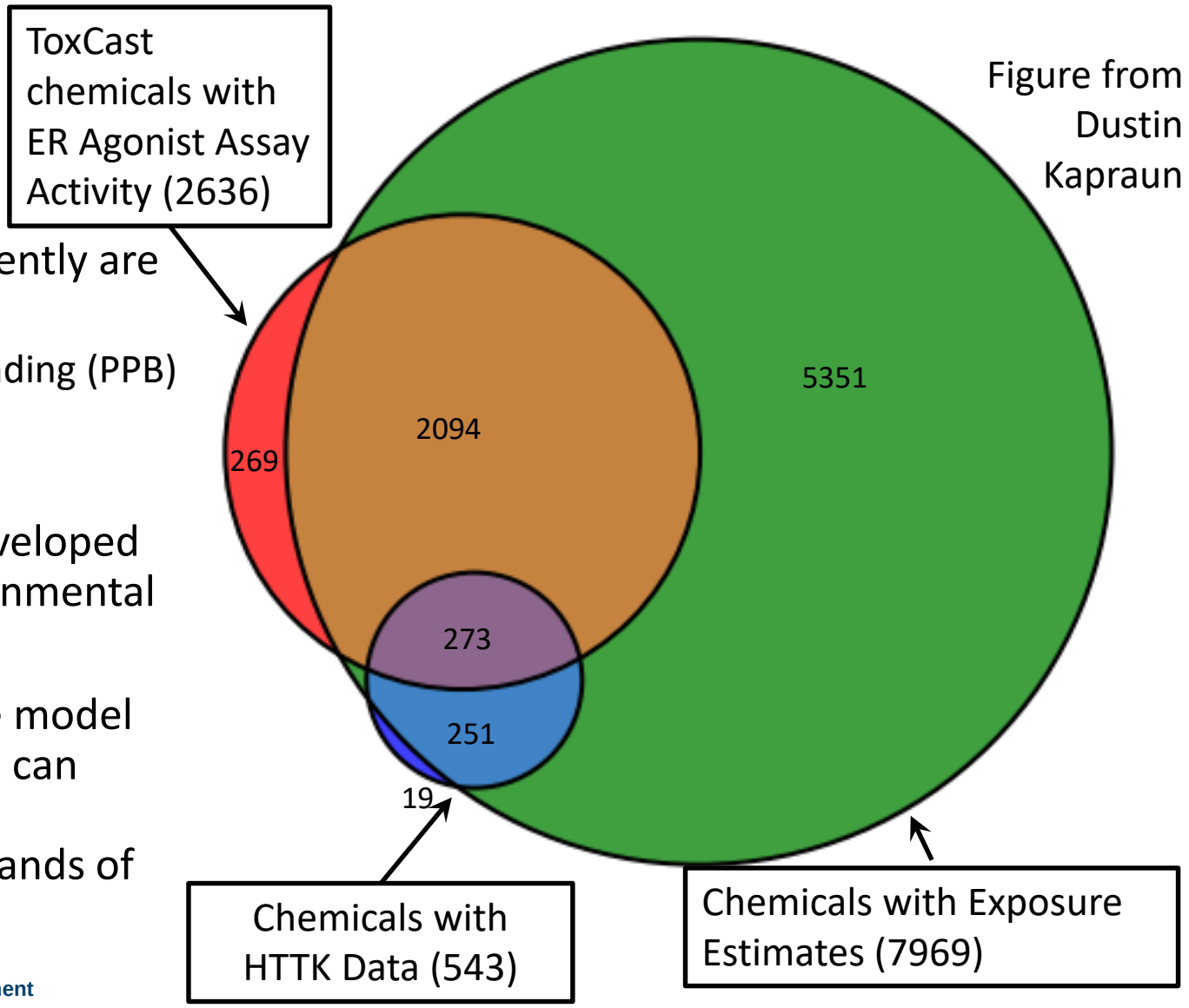
# Chemicals with HTTK Data

- Measurement of *in vitro* clearance and binding both require chemical-specific analytical chemistry methods – these can be difficult to develop
- Methods are appropriate for chemicals that are soluble, non-volatile only



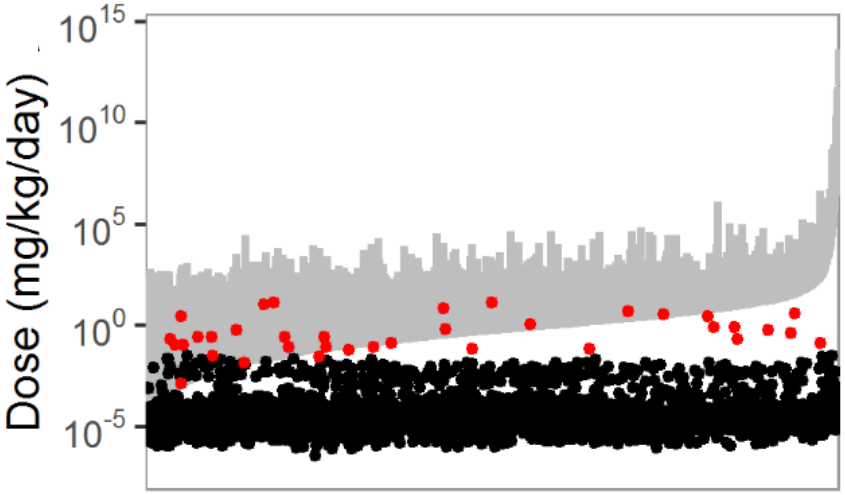
# Predicting Critical TK Parameters

- Two parameters currently are key to HTTK model:
  - Plasma protein binding (PPB)
  - Hepatic clearance (metabolism)
- Ingle *et al.* (2016) developed PPB model for environmental chemicals
- If a hepatic clearance model can be developed we can provide tentative TK predictions for thousands of more chemicals

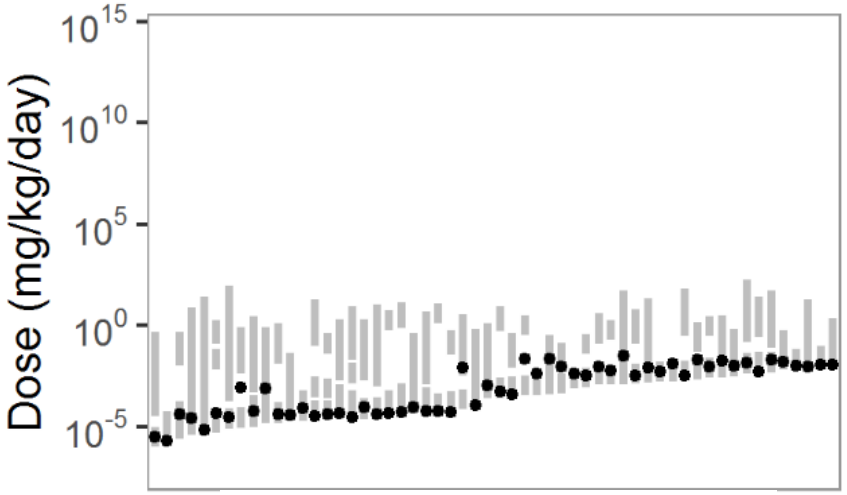


# Using Predicted HTTK for Risk Prioritization

Sipes et al. used Simulations Plus ADMET Predictor to make *in silico* predictions of metabolism and protein binding:




Doses ranges for all 3925 Tox21 compounds eliciting a 'possible'-to-'likely' human *in vivo* interaction alongside estimated daily exposure



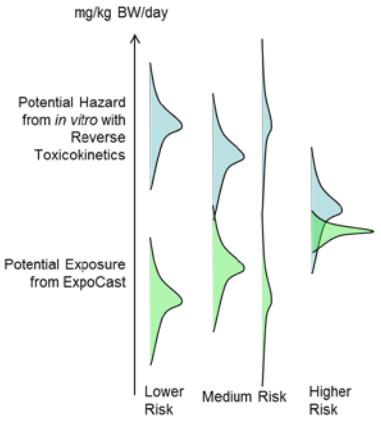

56 compounds with potential *in vivo* biological interaction at or above estimated environmental exposures

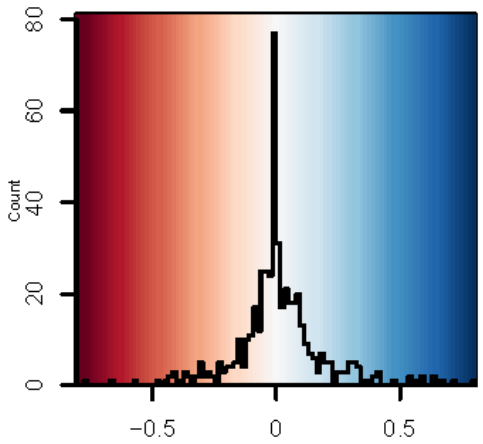


Sipes et al., (2017)

# Life-stage and Demographic Specific Predictions

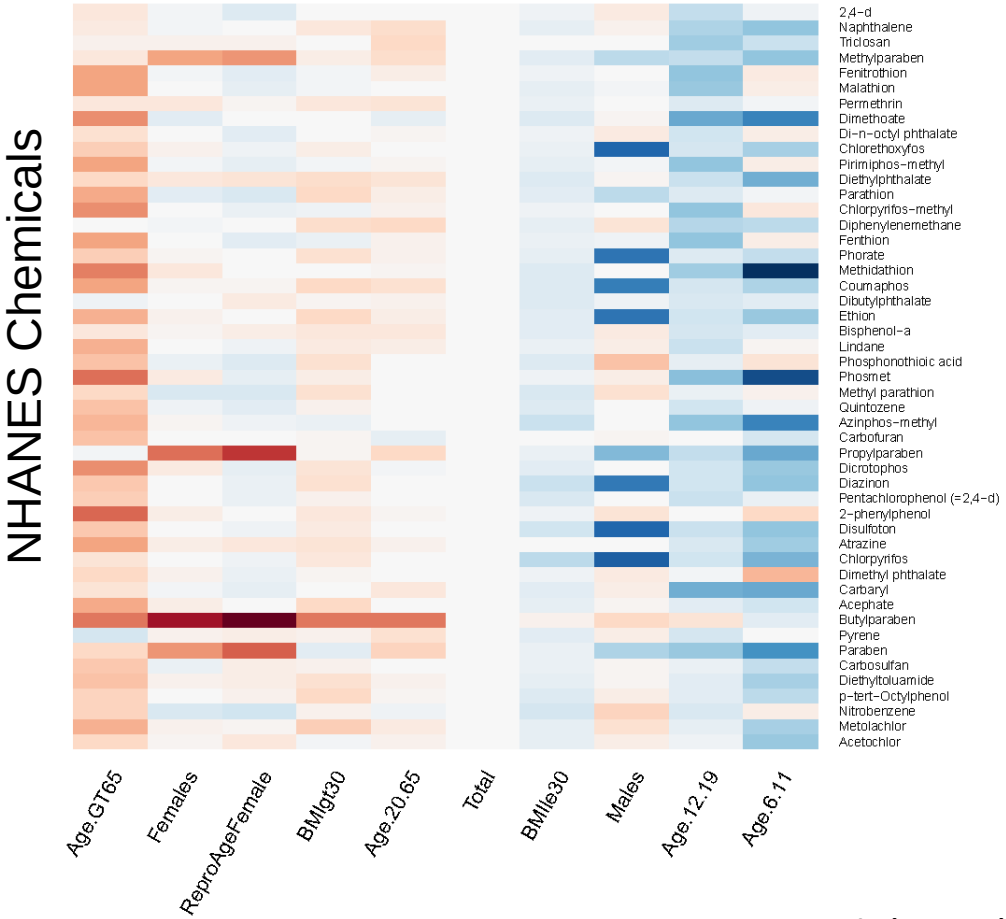
- Can calculate margin between bioactivity and exposure for specific populations




Change in Risk Relative to Total Population

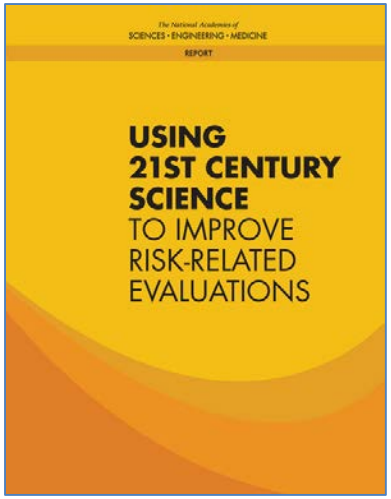
**Change in Activity:Exposure Ratio**




Ring *et al*. (2017)

# Conclusions

- We would like to know more about the risk posed by thousands of chemicals in the environment – which ones should we start with?
  - High throughput screening (HTS) provides a path forward for identifying potential hazard
- Using big data analytics we can identify priority combinations of chemicals
- Using *in vitro* methods developed for pharmaceuticals, we can relatively efficiently predict TK for large numbers of chemicals, but we are limited by analytical chemistry




National Academy of Sciences, January, 2017:
"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…"

EPA United States Environmental Protection Agency

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


**NCCT**
Chris Grulke
Greg Honda*
Richard Judson
Andrew McEachran*
Robert Pearce*
Ann Richard
Risa Sayre*
Woody Setzer
Rusty Thomas
**John Wambaugh**
Antony Williams

**NRMRL**
Yirui Liang*
Xiaoyu Liu

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

**NERL**
Craig Barber
Namdi Brandon*
Peter Egeghy
Hongtai Huang*
Brandall Ingle*
**Kristin Isaacs**
Seth Newton
Katherine Phillips
Paul Price
Jeanette Reyes*
Jon Sobus
John Streicher*
Mark Strynar
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
Barbara Wetmore

*Trainees

**Lead CSS Matrix Interfaces:**
John Kenneke (NERL)
John Cowden (NCCT)


## 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
**Research Triangle Institute**
Timothy Fennell
**ScitoVation**
Harvey Clewell
Kamle Mansouri
Chantel Nicolas
**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 Jolliet
**University of North Carolina, Chapel Hill**
Alex Tropsha

United States Environmental Protection Agency

# References


- Borgelt, C. (2012). Frequent item set mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(6), 437-456.
- Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
- Breyer, Stephen. Breaking the vicious circle: Toward effective risk regulation. Harvard University Press, 2009
- Dionisio, Kathie L., et al. "Exploring Consumer Exposure Pathways and Patterns of Use for Chemicals in the Environment." Toxicology Reports (2015)
- Egeghy, P. P., et al. (2012). The exposure data landscape for manufactured chemicals. Science of the Total Environment, 414, 159-166.
- Filer, Dayne L.. "The ToxCast analysis pipeline: An R package for processing and modeling chemical screening data." US Environmental Protection Agency: http://www. epa. gov/ncct/toxcast/files/MySQL% 20Database/Pipeline_Overview. pdf (2014)
- Goldsmith, M. R., et al. (2014). Development of a consumer product ingredient database for chemical exposure screening and prioritization. Food and chemical toxicology, 65, 269-279.
- Hahn, O., & Straßmann, F. (1938). Über die Entstehung von Radiumisotopen aus Uran durch Bestrahlen mit schnellen und verlangsamten Neutronen. Naturwissenschaften, 26(46), 755-756.
- Hertzberg, R. P., & Pope, A. J. (2000). High-throughput screening: new technology for the 21st century. Current opinion in chemical biology, 4(4), 445-451.
- Ingle, Brandall L., et al. "Informing the Human Plasma Protein Binding of Environmental Chemicals by Machine Learning in the Pharmaceutical Space: Applicability Domain and Limits of Predictability." Journal of Chemical Information and Modeling 56.11 (2016): 2243-2252.
- Kaewkhaw, R., et al. (2016). Treatment paradigms for retinal and macular diseases using 3-D retina cultures derived from human reporter pluripotent stem cell linestreatment design using PSC-Derived 3-D retina cultures. Investigative ophthalmology & visual science, 57(5), ORSFl1-ORSFl11.
- Kapraun, Dustin et al., "A Method for Identifying Prevalent Chemical Combinations in the US Population," Environmental Health Perspectives, 2017
- Kavlock, Robert, et al. "Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management." Chemical research in toxicology 25.7 (2012): 1287-1302.
- National Academies of Sciences, Engineering, and Medicine. (2017). Using 21st century science to improve risk-related evaluations. National Academies Press.
- National Research Council. (1983). Risk Assessment in the Federal Government: Managing the Process Working Papers. National Academies Press.
- National Research Council. (2007) Toxicity testing in the 21st century: a vision and a strategy. National Academies Press.
- O'Connell, S. G., Kincl, L. D., & Anderson, K. A. (2014). Silicone wristbands as personal passive samplers. Environmental science & technology, 48(6), 3327-3335.
- 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, et al. "httk: R Package for High-Throughput Toxicokinetics." Journal of Statistical Software, 2017
- Phillips, Katherine A., et al. "High-throughput screening of chemicals as functional substitutes using structure-based classification models." Green Chemistry (2017).
- Rager, J. et al. (2016). Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. Environment international, 88, 269-280.
- Ring, Caroline, et al., "Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability", Environment International, 2017
- Rotroff, Daniel, et al., (2010) "Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening." Tox. Sciences 117(2), 348-58
- Schymanski, E. L., et al. (2015). Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Analytical and bioanalytical chemistry, 407(21), 6237-625
- Schmidt, Charles W. "TOX 21: new dimensions of toxicity testing." Environmental health perspectives 117.8 (2009): A348.
- Sipes, Nisha, et al. (2017) "An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library", Environmental Science and Technology
- Tonnelier, A., Coecke, S., & Zaldívar, J. M. (2012). Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model. Archives of toxicology, 86(3), 393-403.
- 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).
- Wang, Y.-H. (2010). "Confidence Assessment of the Simcyp Time-Based Approach and a Static Mathematical Model in Predicting Clinical Drug-Drug Interactions for Mechanism-Based CYP3A Inhibitors." Drug Metabolism and Disposition 38(7), 1094-1104
- 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 population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicological sciences 2014;142:210-224
- 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.