March 2, 2020

Exposure-based Chemical Priority Setting in the 21st Century

John Wambaugh
Center for Computational Toxicology and Exposure
Office of Research and Development
U.S. Environmental Protection Agency
wambaugh.john@epa.gov

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

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

US EPA Office of Research and Development

- The Office of Research and Development (ORD) is the scientific research arm of EPA
  - 562 peer-reviewed journal articles in 2018

- Research is conducted by ORD's four national centers, and three offices organized to address:
  - Public health and env. assessment; comp. tox. and exposure; env. measurement and modeling; and env. solutions and emergency response.

- 13 facilities across the United States

- Research conducted by a combination of Federal scientists (including uniformed members of the **Public Health Service**); 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 chemical signatures in pooled human blood samples, many appear to be exogenous

- A tapestry of laws covers the chemicals people are exposed to in the United States (Breyer, 2009)

- Chemical safety testing is primarily for food additives, pharmaceuticals, and pesticide active ingredients (NRC, 2007)
  - Different levels depending on category

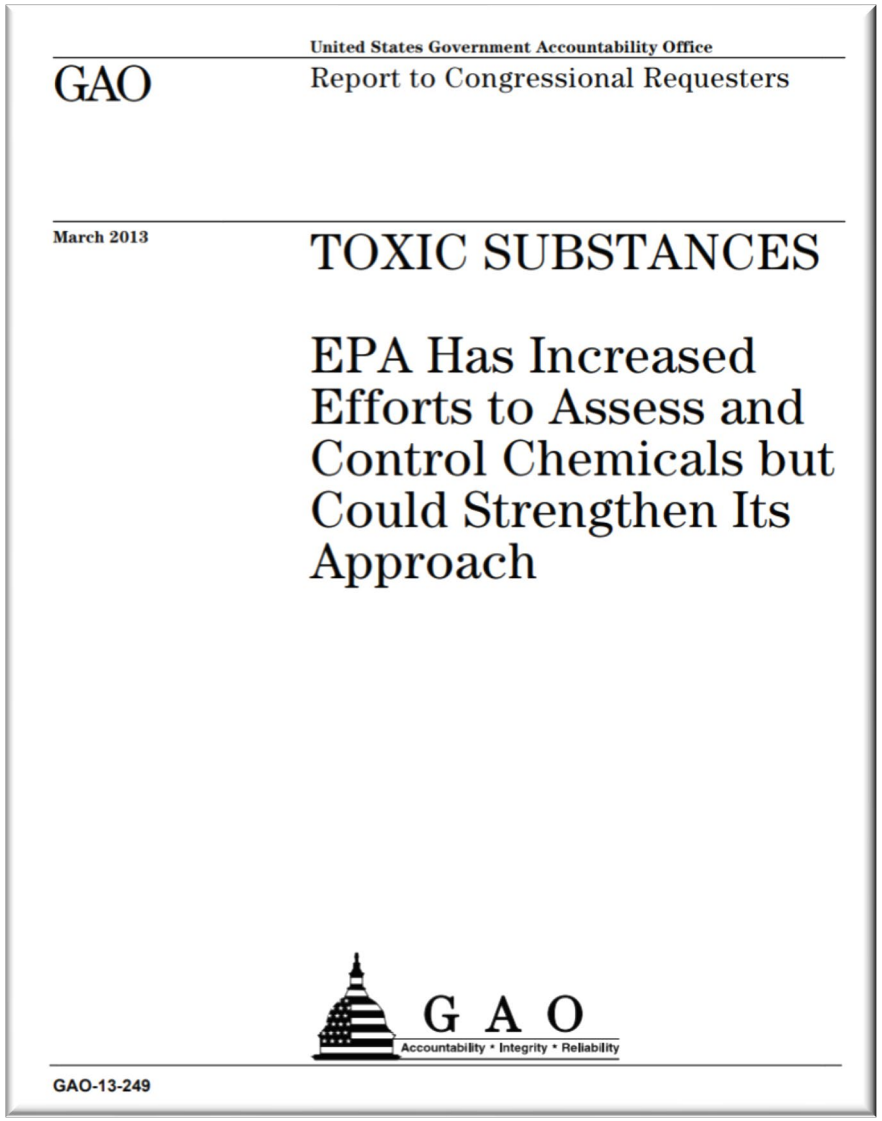
November 29, 2014

Chemical Regulation in the United States

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

- Thousands of chemicals on the market were "grandfathered" in without assessment
  Judson et al. (2009), Egeghy et al. (2012), Wetmore et al. (2015)

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


United States Government Accountability Office
GAO
Report to Congressional Requesters

March 2013

TOXIC SUBSTANCES

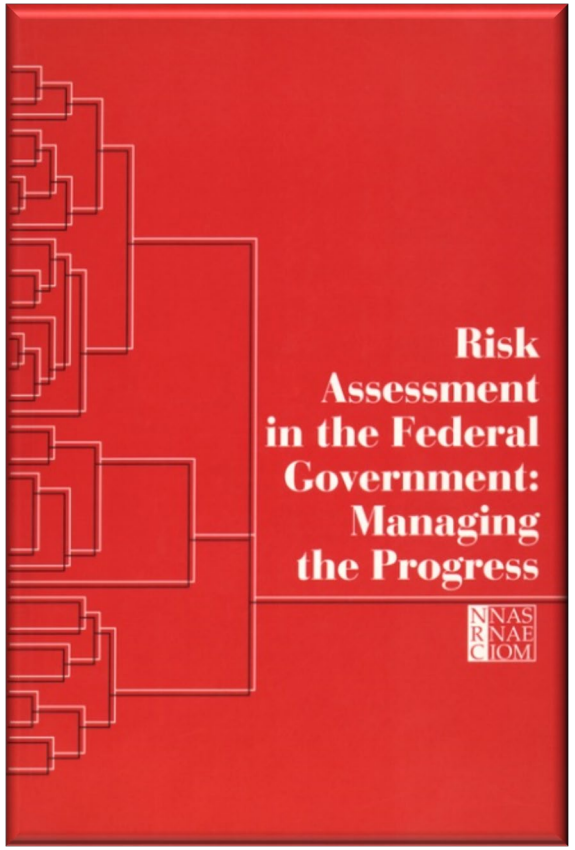
EPA Has Increased Efforts to Assess and Control Chemicals but Could Strengthen Its Approach

G A O
Accountability • Integrity • Reliability

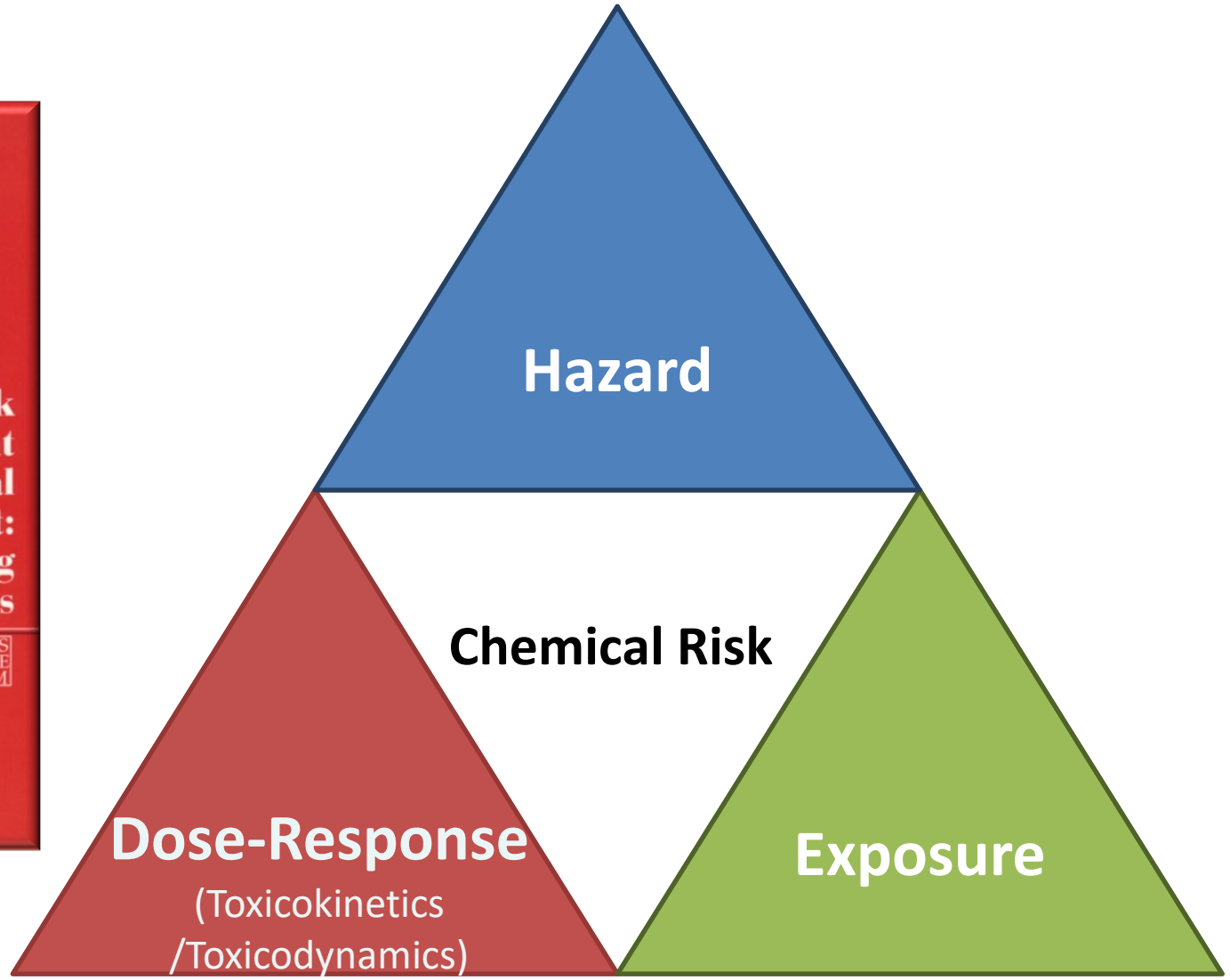
GAO-13-249

March, 2013

Three Components for Chemical Risk



NRC (1983)



The National Academy of Sciences, Engineering and Medicine (1983) outlined three components for determining chemical risk.

Toxic Substances Control Act (TSCA)

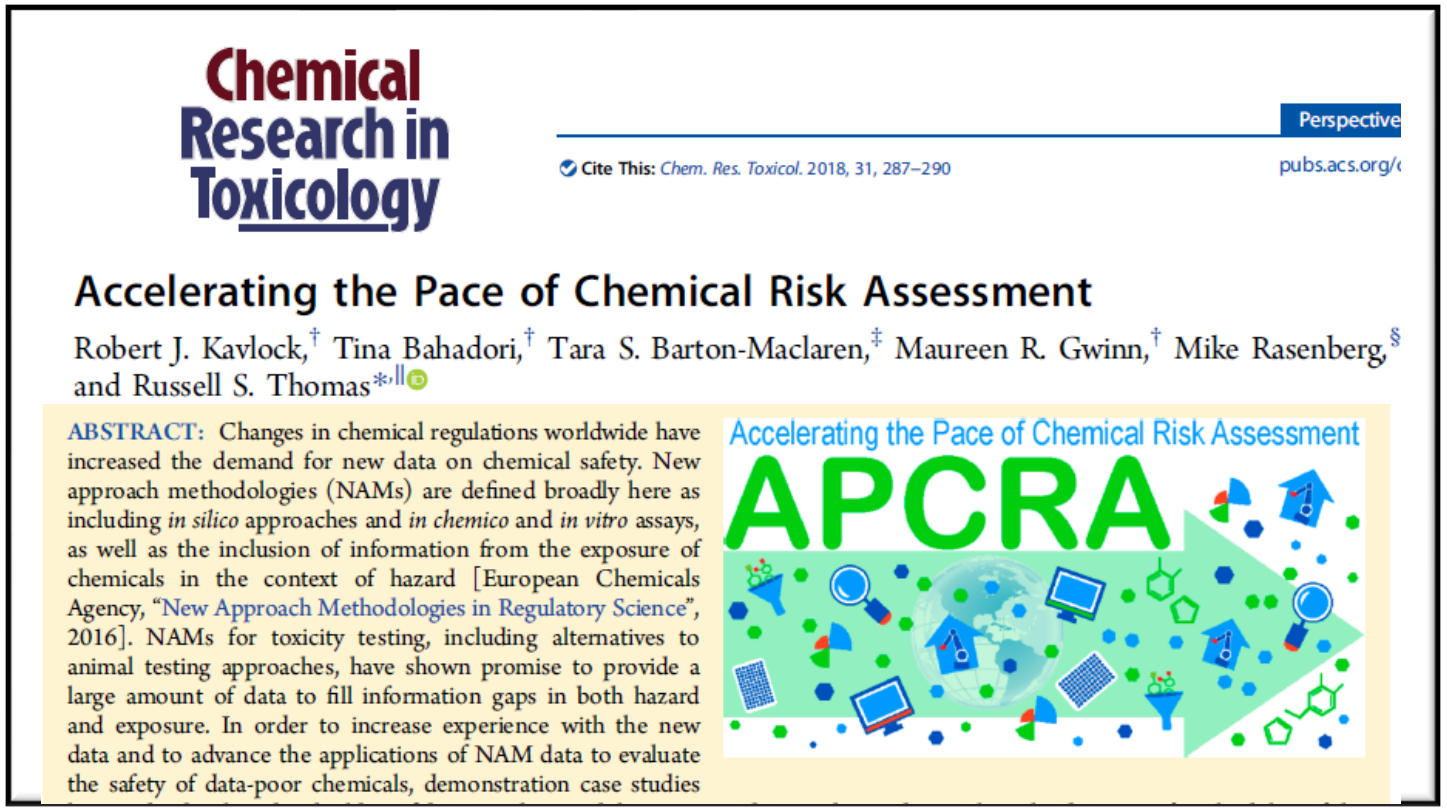
- TSCA was updated in June, 2016 to allow more rapid evaluation of chemicals (Frank R. Lautenberg Chemical Safety for the 21st Century Act)
- New approach methodologies (NAMs) are being considered to inform prioritization of chemicals for testing and evaluation (Kavlock et al., 2018)
- EPA has released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

Schmidt, C. W. (2016). TSCA 2.0: A new era in chemical risk management", Environmental Health Perspectives, A182-A186.

New Approach Methodologies (NAMs)

- There are roughly 10,000 TSCA-relevant chemicals in commerce
  - Traditional methods are too resource-intensive to address all of these

- NAMs include:
  - High throughput screening (ToxCast)
  - High throughput exposure estimates (ExpoCast)
  - High throughput toxicokinetics (HTTK)



Chemical Research in Toxicology

Perspective

Cite This: *Chem. Res. Toxicol.* 2018, 31, 287−290

pubs.acs.org/

Accelerating the Pace of Chemical Risk Assessment

Robert J. Kavlock,[†] Tina Bahadori,[†] Tara S. Barton-Maclaren,[‡] Maureen R. Gwinn,[†] Mike Rasenberg,[§] and Russell S. Thomas*[,∥]

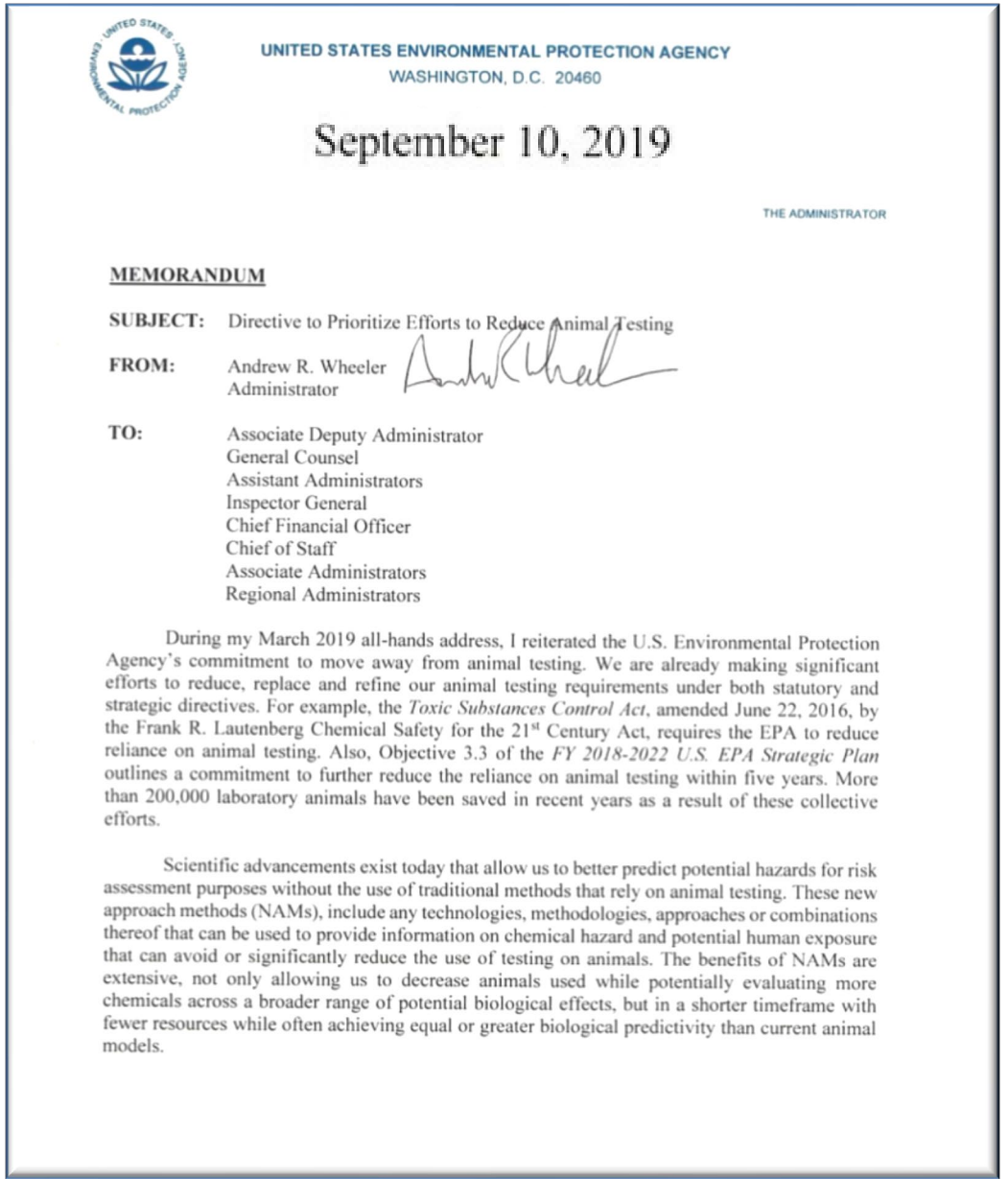
ABSTRACT: Changes in chemical regulations worldwide have increased the demand for new data on chemical safety. New approach methodologies (NAMs) are defined broadly here as including *in silico* approaches and *in chemico* and *in vitro* assays, as well as the inclusion of information from the exposure of chemicals in the context of hazard [European Chemicals Agency, "New Approach Methodologies in Regulatory Science", 2016]. NAMs for toxicity testing, including alternatives to animal testing approaches, have shown promise to provide a large amount of data to fill information gaps in both hazard and exposure. In order to increase experience with the new data and to advance the applications of NAM data to evaluate the safety of data-poor chemicals, demonstration case studies

- TSCA Proof of concept: Examine ~200 chemicals with ToxCast, ExpoCast and HTTK
  - HTTK was rate limiter on number of chemicals
  - *"A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA"*

Replacing Animal Testing with NAMs

- Administrator of the EPA: "To aggressively pursue a reduction in animal testing, I am directing leadership and staff in the Office of Chemical Safety and Pollution Prevention and the Office of Research and Development [ORD] to prioritize … the reduction of animal testing while ensuring protection of human health and the environment."

- "These new approach methods (NAMs), include any technologies, methodologies, approaches or combinations thereof that can be used to provide information on chemical hazard and potential human exposure that can avoid or significantly reduce the use of testing on animals"
  - NAMs for filling information gaps for decision-making
  - integrating data steams into chemical risk assessment
  - making the information publicly available



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

September 10, 2019

THE ADMINISTRATOR

MEMORANDUM

SUBJECT: Directive to Prioritize Efforts to Reduce Animal Testing

FROM: Andrew R. Wheeler
Administrator

TO: Associate Deputy Administrator
General Counsel
Assistant Administrators
Inspector General
Chief Financial Officer
Chief of Staff
Associate Administrators
Regional Administrators

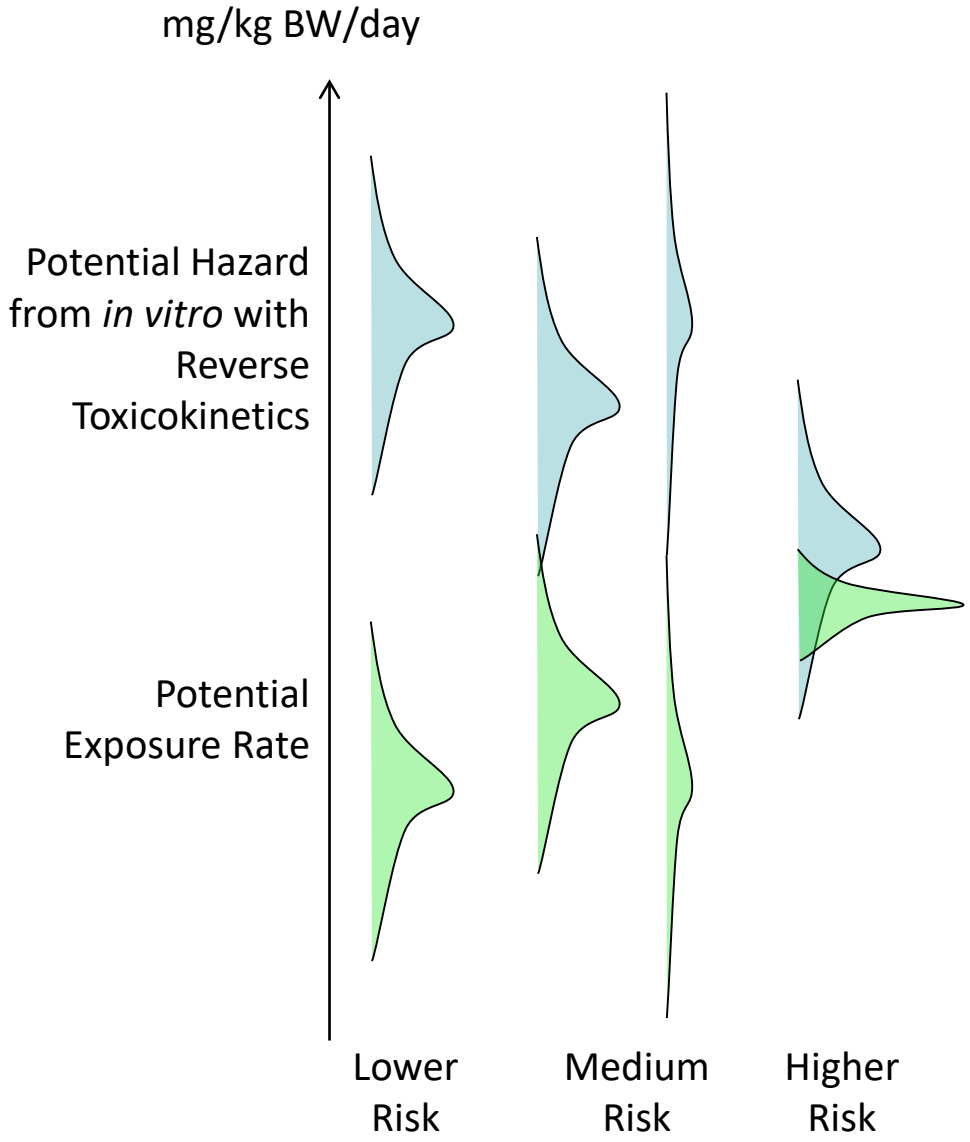
During my March 2019 all-hands address, I reiterated the U.S. Environmental Protection Agency's commitment to move away from animal testing. We are already making significant efforts to reduce, replace and refine our animal testing requirements under both statutory and strategic directives. For example, the *Toxic Substances Control Act*, amended June 22, 2016, by the Frank R. Lautenberg Chemical Safety for the 21st Century Act, requires the EPA to reduce reliance on animal testing. Also, Objective 3.3 of the *FY 2018-2022 U.S. EPA Strategic Plan* outlines a commitment to further reduce the reliance on animal testing within five years. More than 200,000 laboratory animals have been saved in recent years as a result of these collective efforts.

Scientific advancements exist today that allow us to better predict potential hazards for risk assessment purposes without the use of traditional methods that rely on animal testing. These new approach methods (NAMs), include any technologies, methodologies, approaches or combinations thereof that can be used to provide information on chemical hazard and potential human exposure that can avoid or significantly reduce the use of testing on animals. The benefits of NAMs are extensive, not only allowing us to decrease animals used while potentially evaluating more chemicals across a broader range of potential biological effects, but in a shorter timeframe with fewer resources while often achieving equal or greater biological predictivity than current animal models.

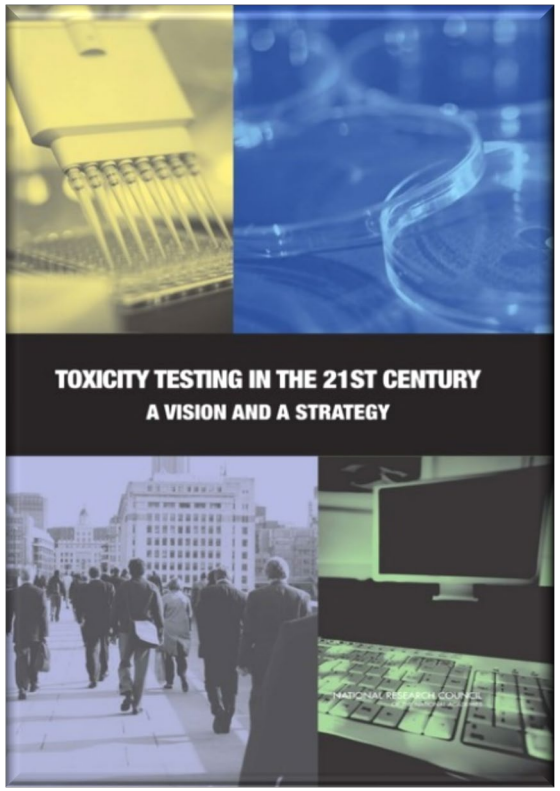
Chemical Risk = Hazard x Exposure

- The U.S. National Research Council (1983) identified chemical risk as a function of both inherent hazard and exposure

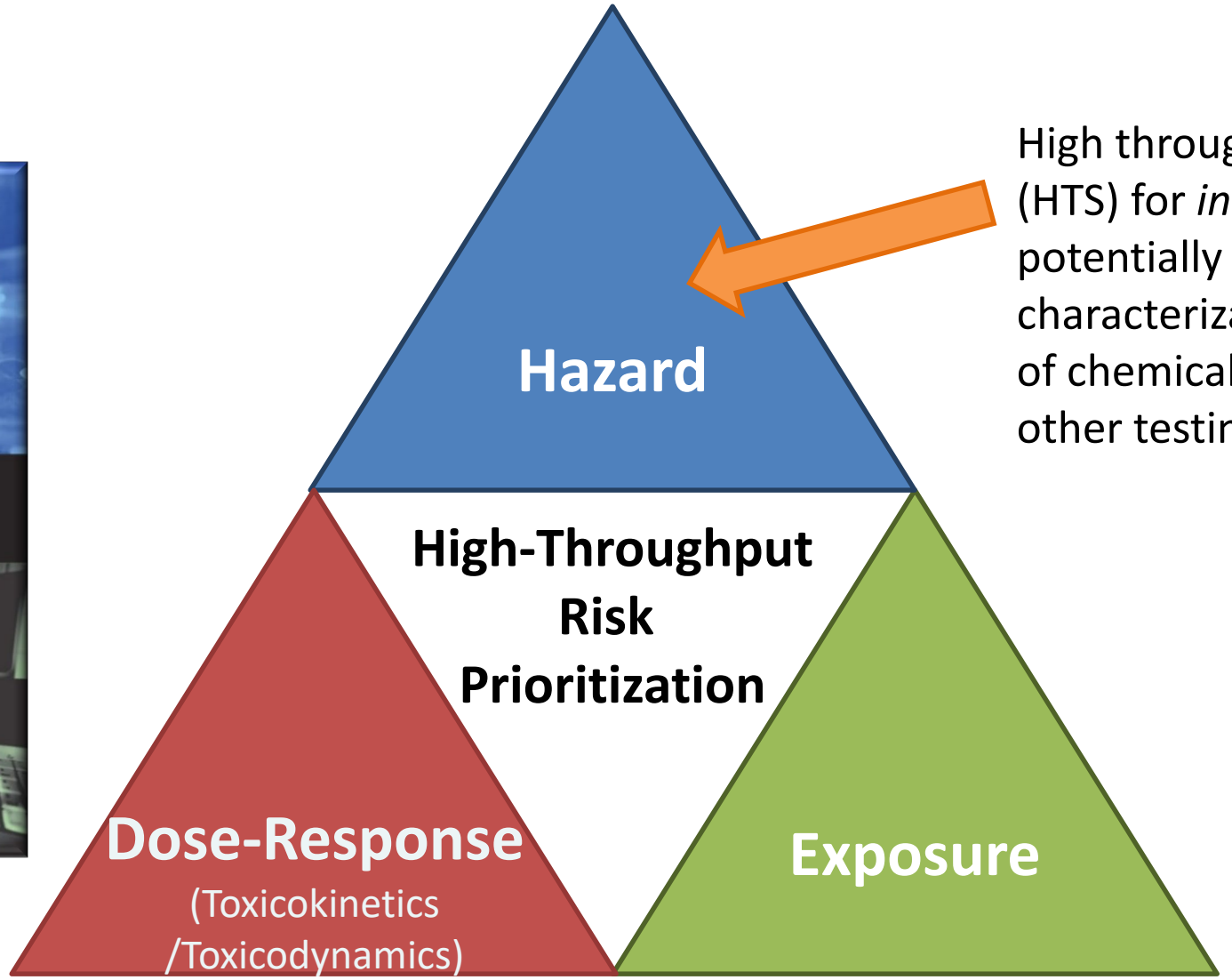
- Therefore, high throughput risk prioritization needs:
  1. High throughput hazard characterization (Dix et al., 2007, Collins et al., 2008)
  2. High throughput exposure forecasts (Wambaugh et al., 2013, 2014)
  3. High throughput toxicokinetics (i.e., dose-response relationship) linking hazard and exposure (Wetmore et al., 2012, 2015)



High-Throughput Risk Prioritization



NRC (2007)



High throughput screening (HTS) for *in vitro* bioactivity potentially allows characterization of thousands of chemicals for which no other testing has occurred

To perform high throughput risk prioritization, we need all three components

High-throughput Screening

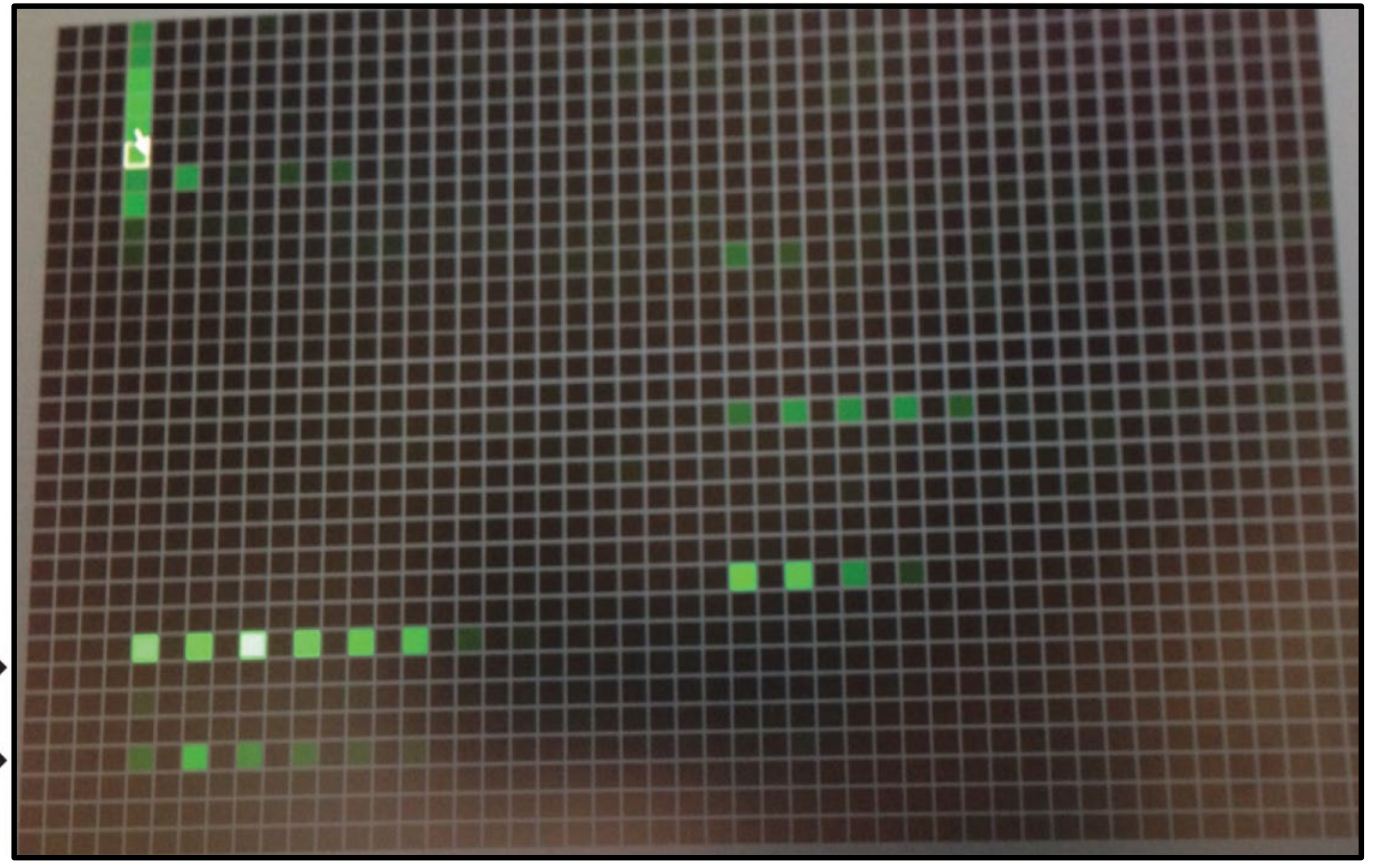
Hertzberg and Pope (2000):

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

Kaewkhaw et al. (2016)

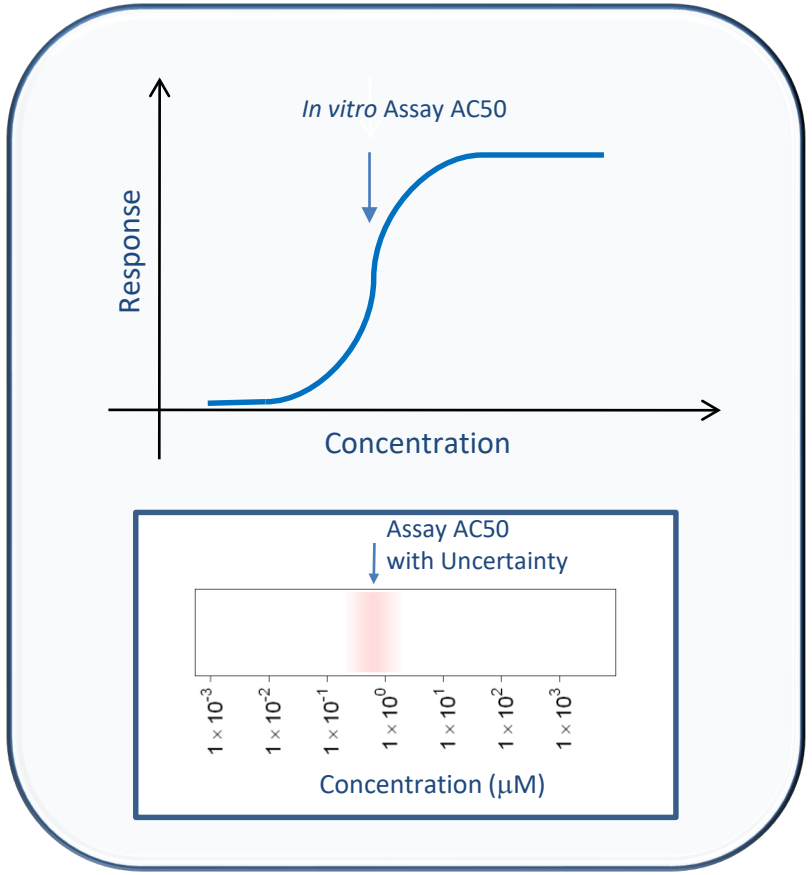
- "…new fluorescence methods, detection platforms and liquid-handling technologies."

- Typically assess many chemicals with a signal readout (e.g., green fluorescent protein).

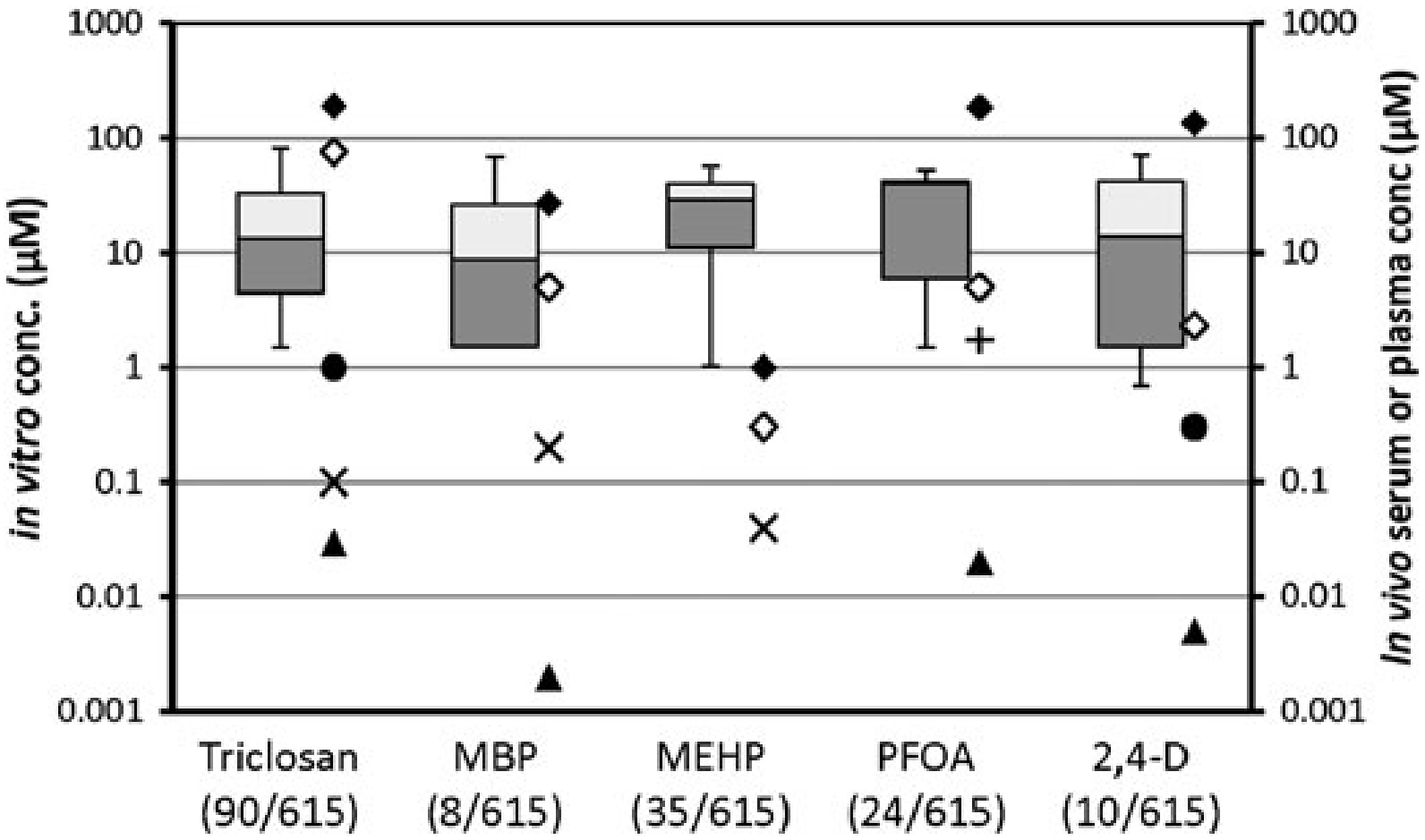


High-Throughput Bioactivity Screening Projects

- We attempt to 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 – AC_{50} – and efficacy if data described by a Hill function, Filer *et al.*, 2016)
- All data are public: http://comptox.epa.gov/dashboard/



The Margin Between Exposure and Hazard



Range of bioactive concentrations across ToxCast assays

Estimated or measured average concentrations associated with the LOAEL in animal studies

NOAEL in animal studies

Humans with chronic exposure reference values (solid circles)

X Volunteers using products containing the chemical

\+ Bio-monitored occupational populations

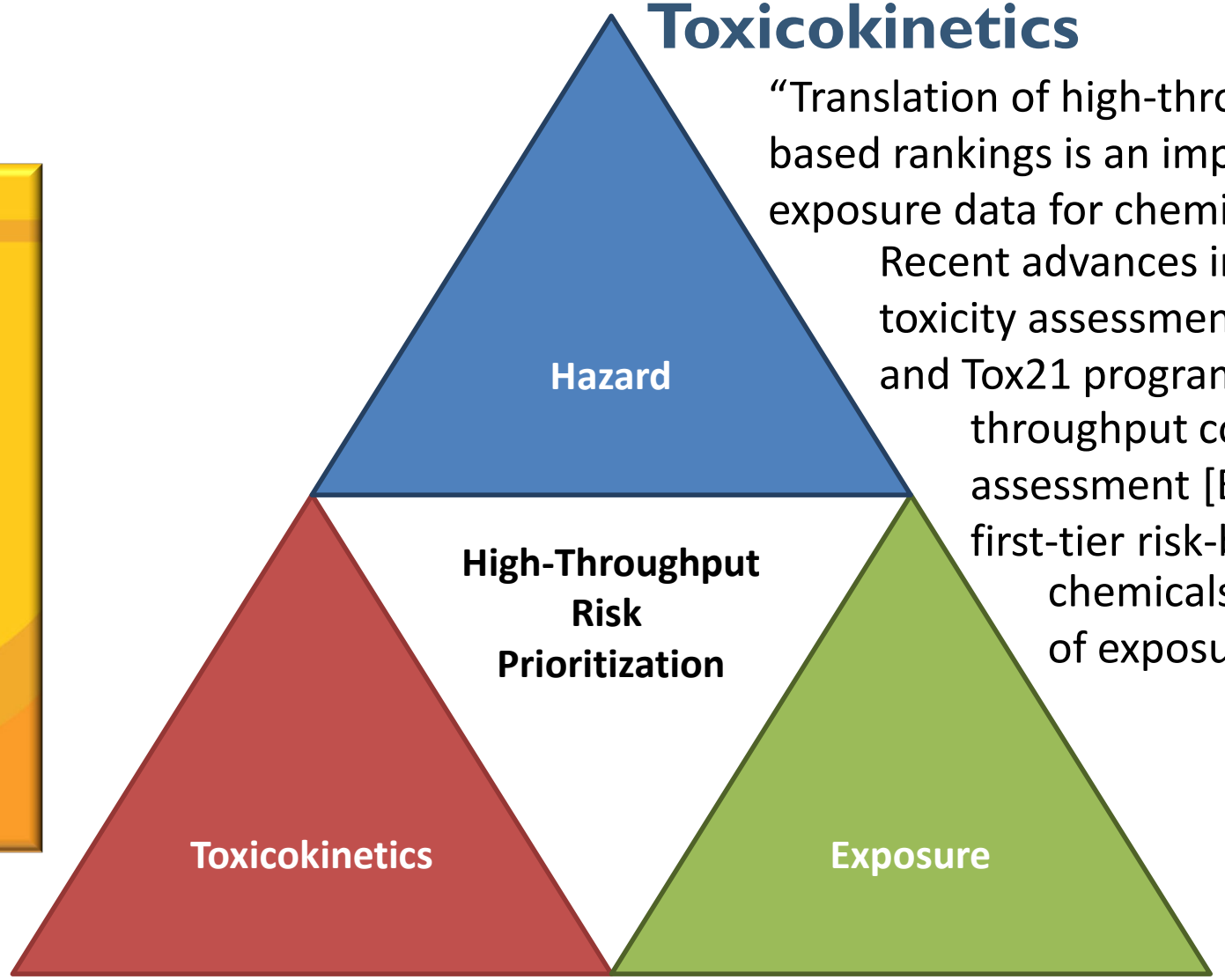
General populations

The five chemicals (as of 2011) with plasma biomonitoring AND ToxCast data… what do we do about the other 1000's?

Aylward and Hays (2011)

Most Chemicals Lack Data on Exposure and Toxicokinetics

NASEM (2017)

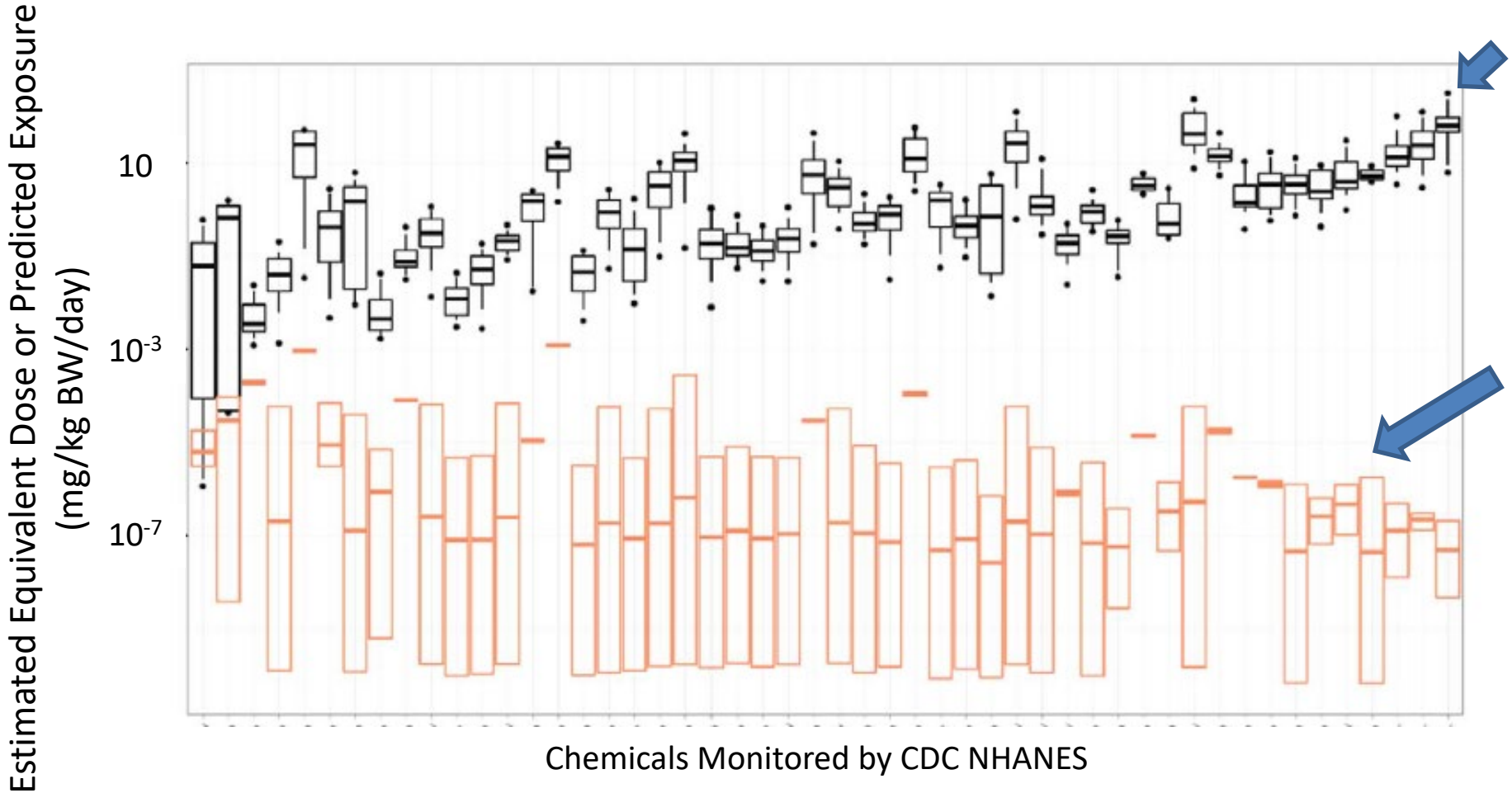


"Translation of high-throughput data into risk-based rankings is an important application of exposure data for chemical priority-setting.

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" - National Academies of Sciences, Engineering, and Medicine (NASEM)

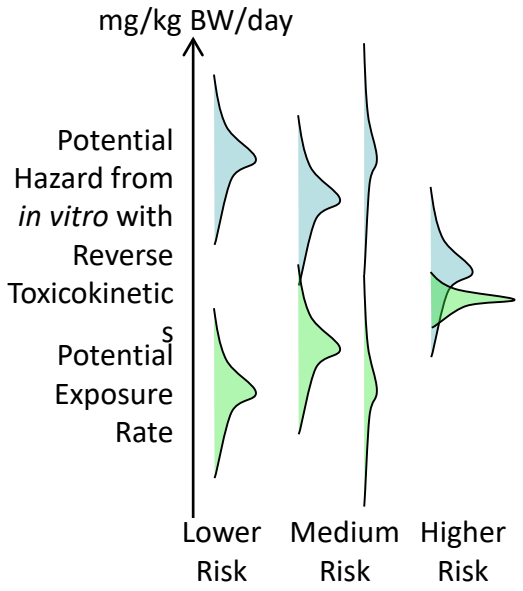
In order to perform risk-based ranking we need data on hazard, toxicokinetics, and exposure…

Chemical Prioritization NAMs



High throughput *in vitro* screening can estimate doses needed to cause bioactivity (e.g., Wetmore et al., 2015)

Exposure intake rates can be inferred from biomarkers (e.g., Ring et al., 2018)



Ring *et al*. (2017)

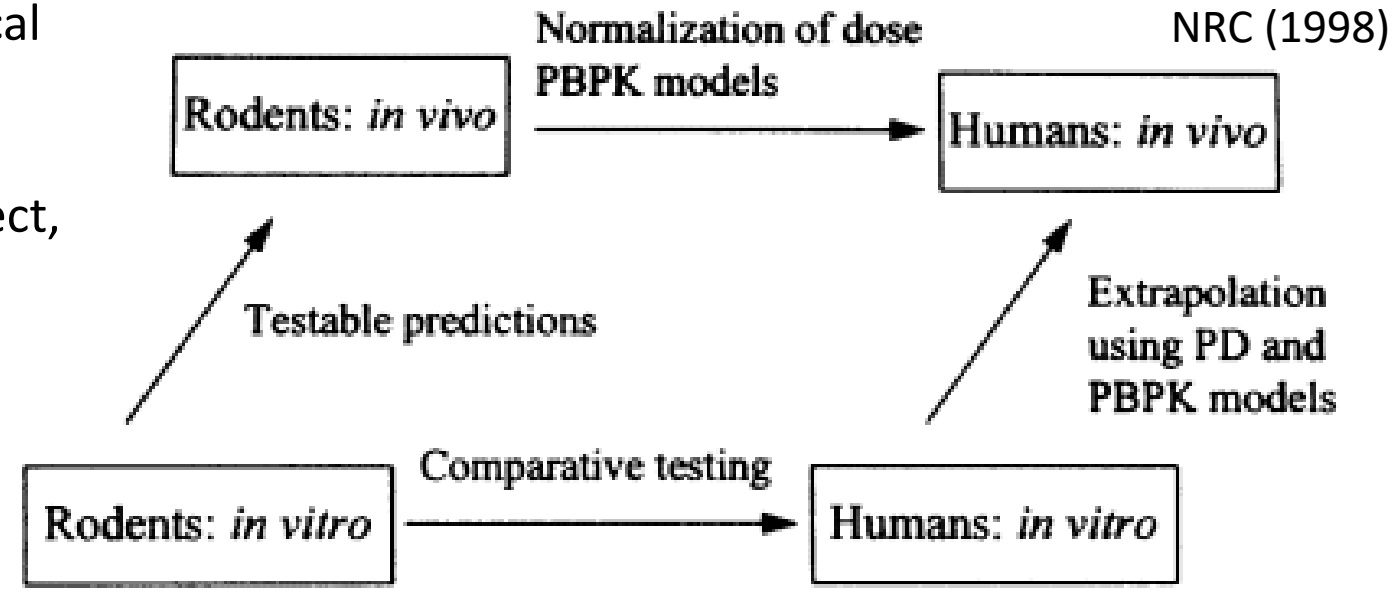
In Vitro - In Vivo Extrapolation (IVIVE)

IVIVE is the use of *in vitro* experimental data to predict phenomena *in vivo*

- IVIVE-PK/TK (**Pharmacokinetics/Toxicokinetics**):
  - Fate of molecules/chemicals in body
  - Considers absorption, distribution, metabolism, excretion (ADME)
  - Uses empirical PK and physiologically-based (PBPK) modeling

- IVIVE-PD/TD (**Pharmacodynamics/Toxicodynamics**):
  - Effect of molecules/chemicals at biological target *in vivo*
  - Assay design/selection important
  - Perturbation as adverse/therapeutic effect, reversible/ irreversible effeccts

- Both contribute to *in vivo* effect prediction

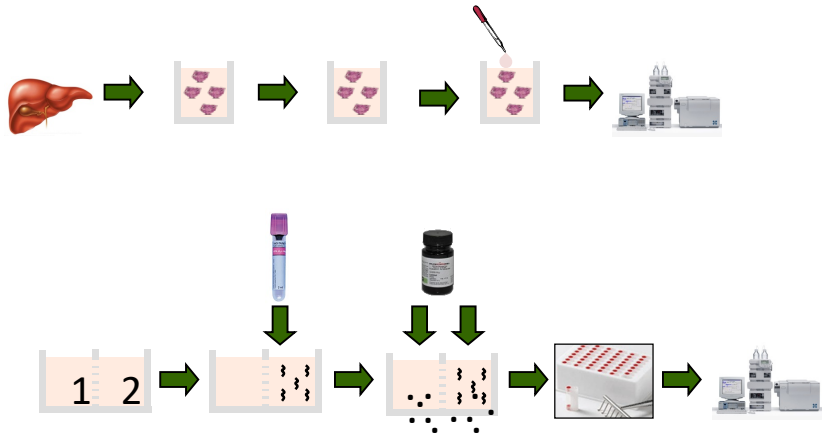


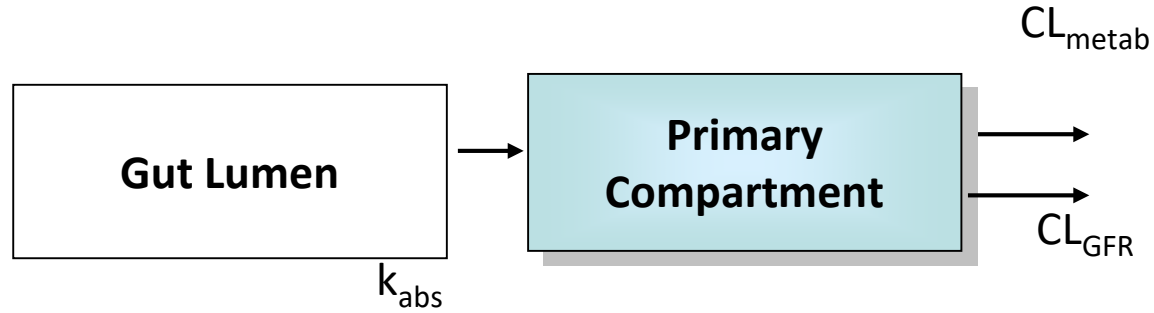
NRC (1998)

High Throughput Toxicokinetics (HTTK)

In vitro **toxicokinetic data + generic toxicokinetic model**

= high(er) throughput toxicokinetics

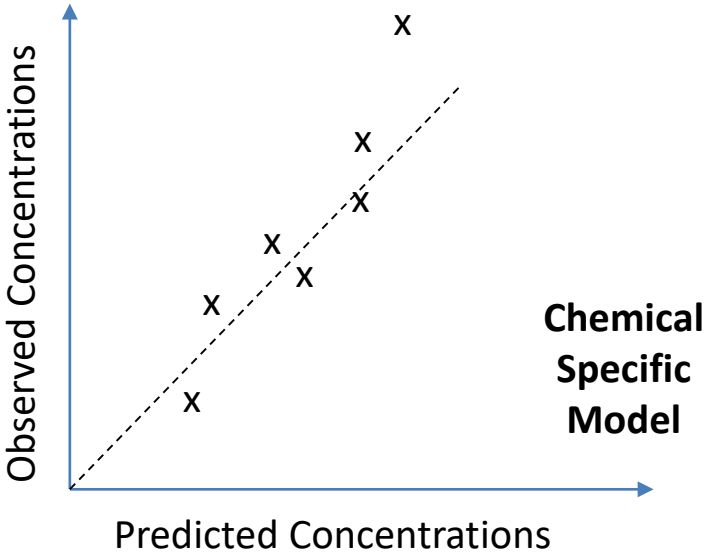




= *httk*

Building Confidence in TK Models

- To evaluate a **chemical-specific TK model** for "chemical x" you can compare the predictions to *in vivo* measured data
  - Can estimate bias
  - Can estimate uncertainty
  - Can consider using model to extrapolate to other situations (dose, route, physiology) where you don't have data

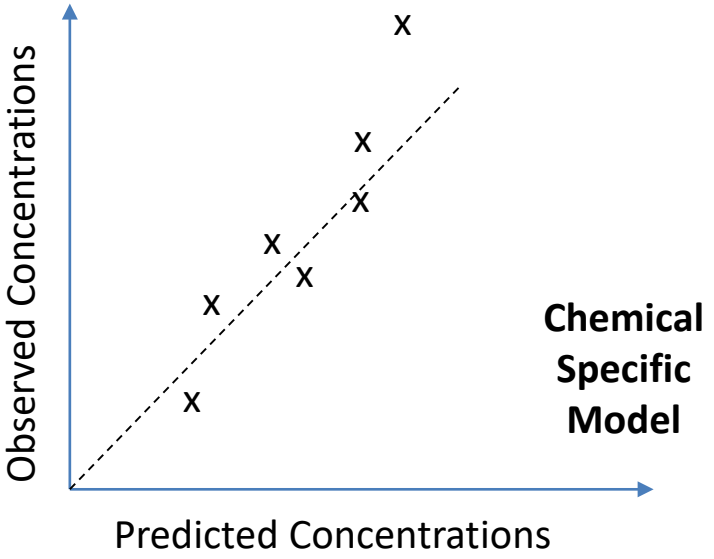


Cohen Hubal et al. (2018)

Building Confidence in TK Models

- To evaluate a **chemical-specific TK model** for "chemical x" you can compare the predictions to *in vivo* measured data
  - Can estimate bias
  - Can estimate uncertainty
  - Can consider using model to extrapolate to other situations (dose, route, physiology) where you don't have data

- However, we do not typically have TK data



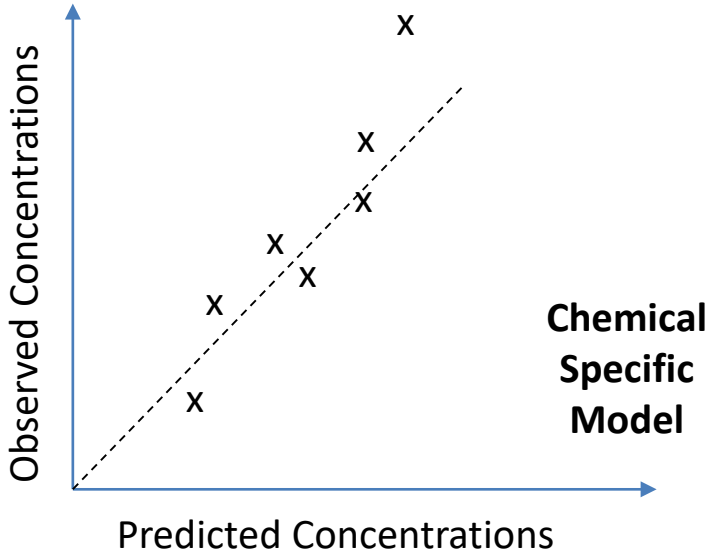
Cohen Hubal et al. (2018)

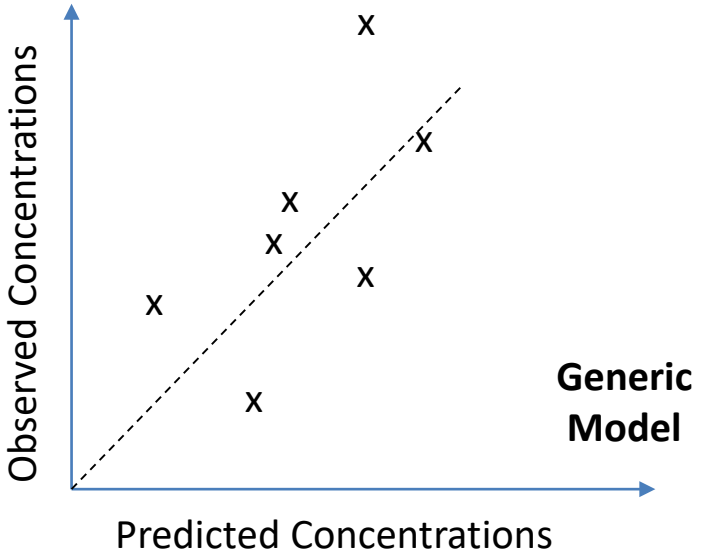
Building Confidence in TK Models

- To evaluate a **chemical-specific TK model** for "chemical x" you can compare the predictions to *in vivo* measured data
  - Can estimate bias
  - Can estimate uncertainty
  - Can consider using model to extrapolate to other situations (dose, route, physiology) where you don't have data

- However, we do not typically have TK data

- We can parameterize a **generic TK model**, and evaluate that model for as many chemicals as we do have data
  - We do expect larger uncertainty, but also greater confidence in model implementation
  - Estimate bias and uncertainty, and try to correlate with chemical-specific properties





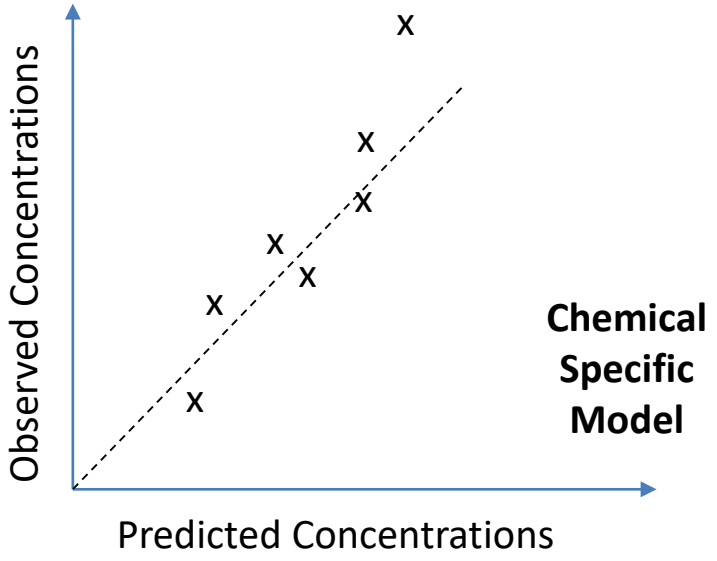
Cohen Hubal et al. (2018)

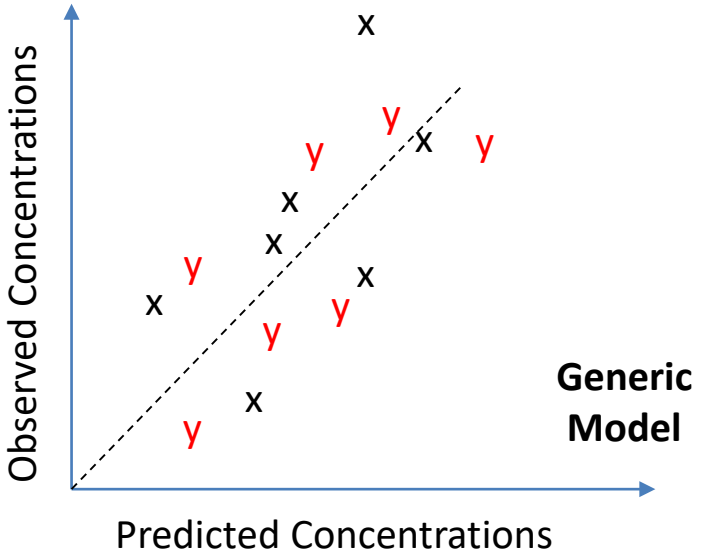
Building Confidence in TK Models

- To evaluate a **chemical-specific TK model** for "chemical x" you can compare the predictions to *in vivo* measured data
  - Can estimate bias
  - Can estimate uncertainty
  - Can consider using model to extrapolate to other situations (dose, route, physiology) where you don't have data

- However, we do not typically have TK data

- We can parameterize a **generic TK model**, and evaluate that model for as many chemicals as we do have data
  - We do expect larger uncertainty, but also greater confidence in model implementation
  - Estimate bias and uncertainty, and try to correlate with chemical-specific properties
  - Can consider using model to extrapolate to other situations (chemicals without *in vivo* data)





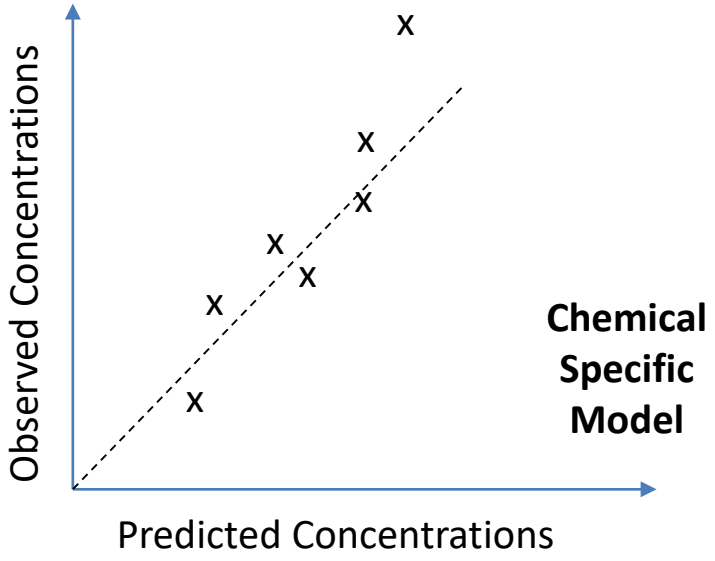
Cohen Hubal et al. (2018)

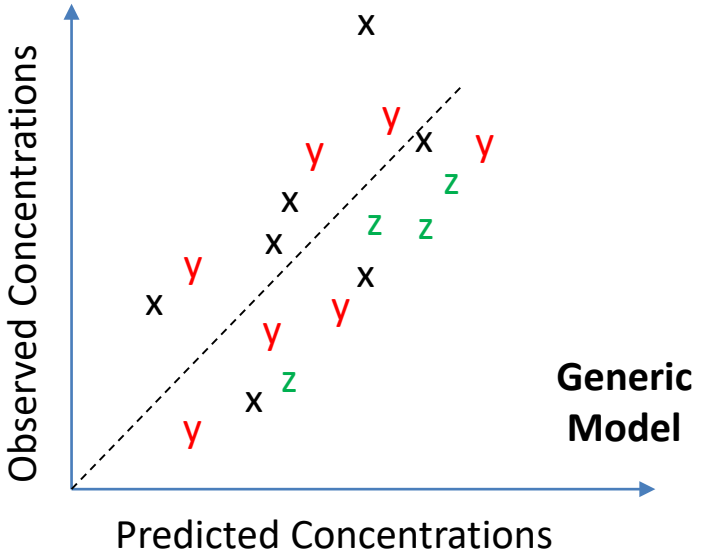
Building Confidence in TK Models

- To evaluate a **chemical-specific TK model** for "chemical x" you can compare the predictions to *in vivo* measured data
  - Can estimate bias
  - Can estimate uncertainty
  - Can consider using model to extrapolate to other situations (dose, route, physiology) where you don't have data

- However, we do not typically have TK data

- We can parameterize a **generic TK model**, and evaluate that model for as many chemicals as we do have data
  - We do expect larger uncertainty, but also greater confidence in model implementation
  - Estimate bias and uncertainty, and try to correlate with chemical-specific properties
  - Can consider using model to extrapolate to other situations (chemicals without *in vivo* data)





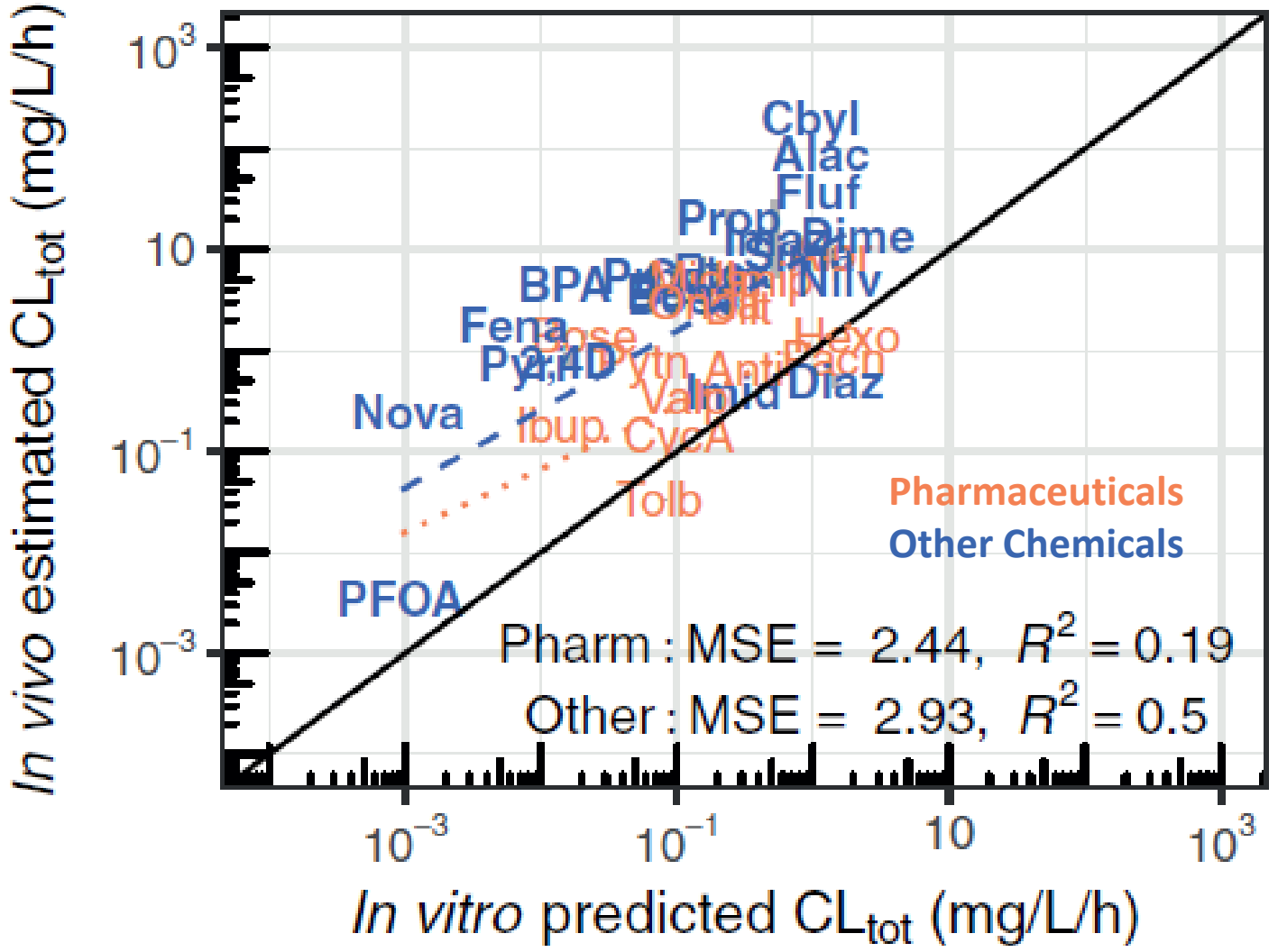
Cohen Hubal et al. (2018)

Evaluation Example

- The HTTK model estimates chemical clearance from the body by two processes:
  - hepatic metabolism (liver)
  - passive glomerular filtration (kidney)

- This appears to work better for pharmaceuticals than other chemicals:
  - ToxCast chemicals are overestimated

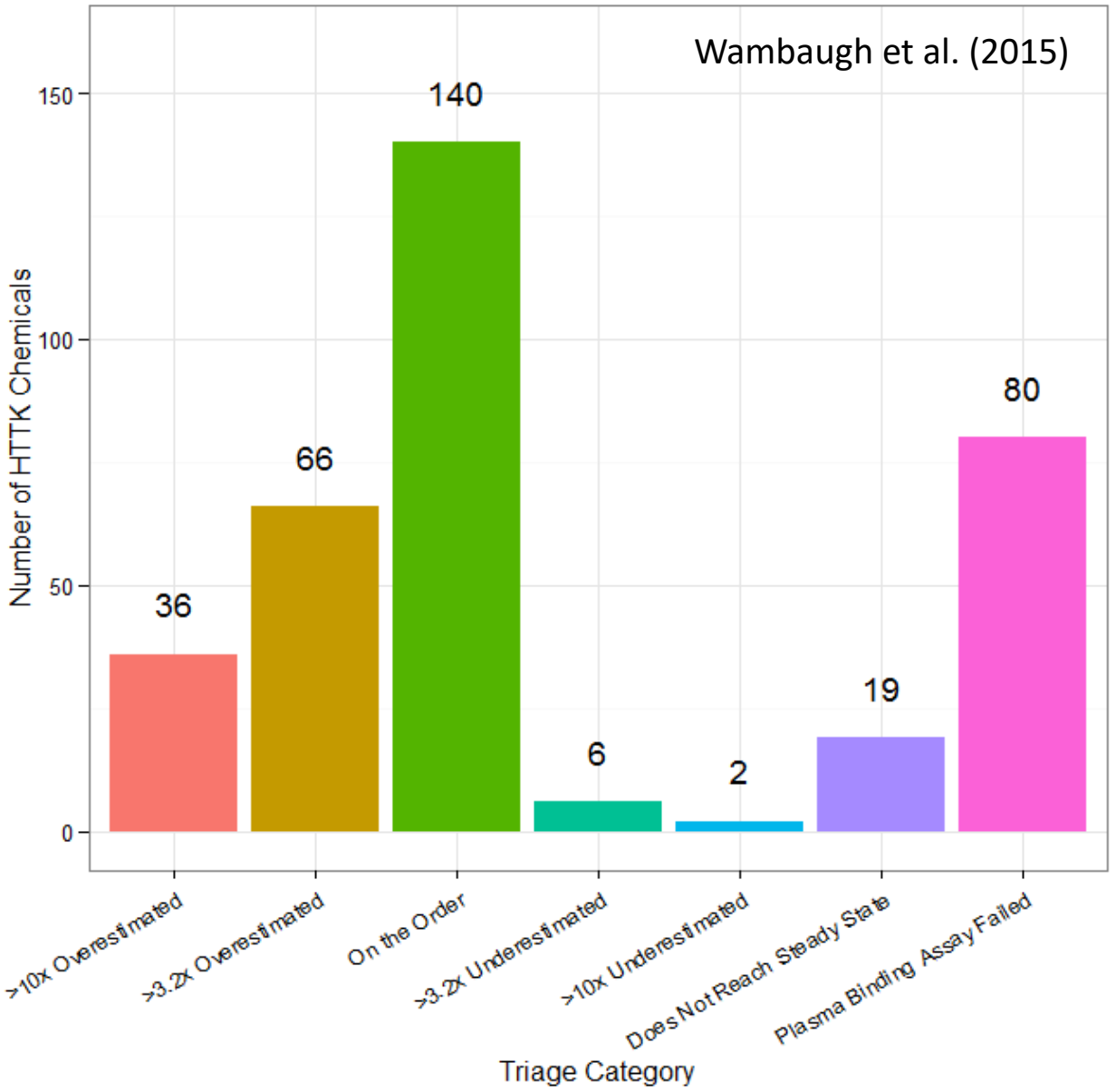
- Non-pharmaceuticals may be subject to extrahepatic metabolism and/or active transport



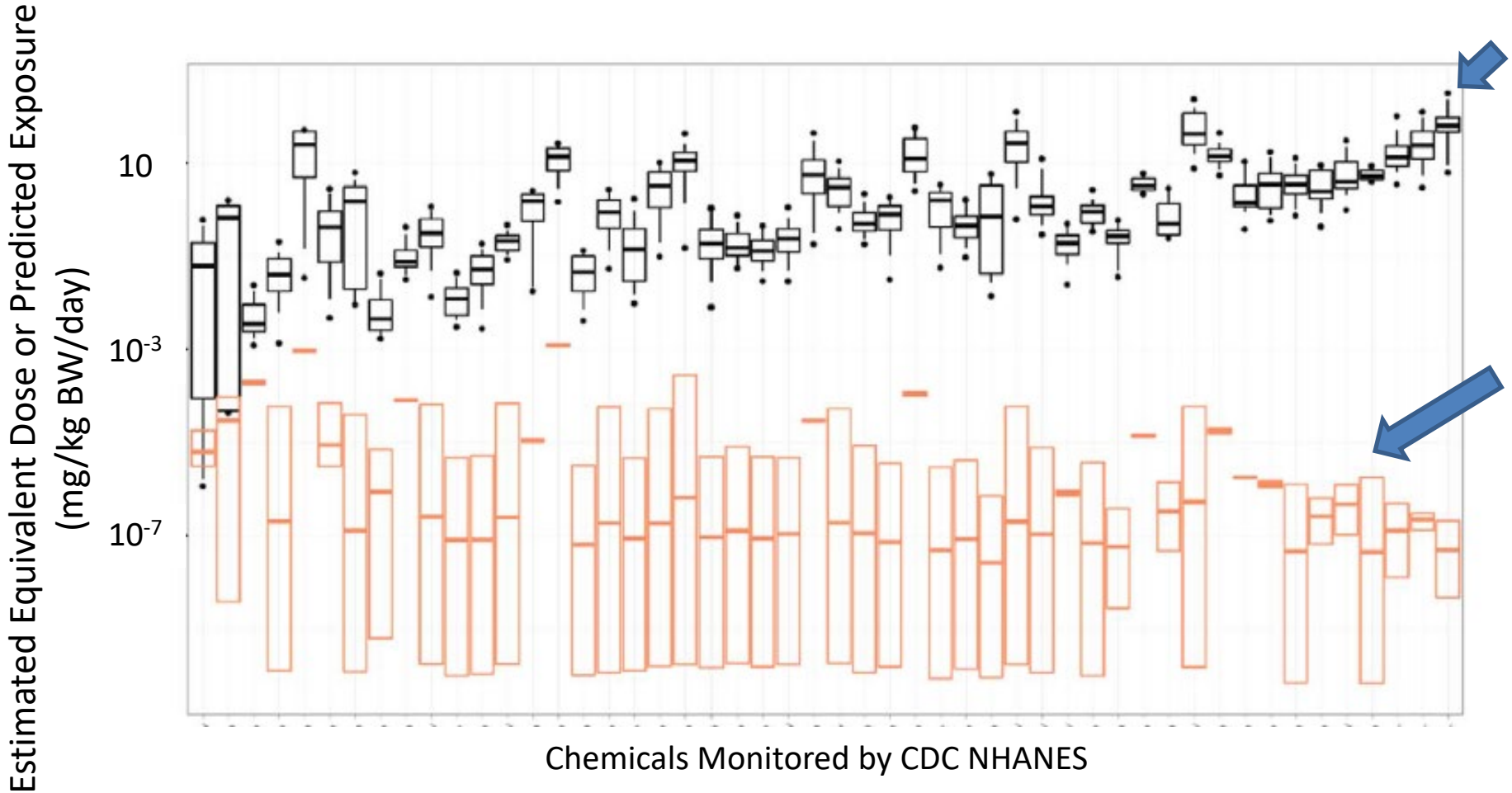
Wambaugh et al. (2018)

Toxicokinetic Triage: When Does TK IVIVE

- Through comparison to *in vivo* data, a cross-validated (random forest) predictor of success or failure of HTTK has been constructed
- All chemicals can be placed into one of seven confidence categories
  - Added categories for chemicals that do not reach steady-state or for which plasma binding assay fails
- Plurality of chemicals end up in the "on the order" bin (within a factor of 3.2x) which is consistent with Wang (2010)

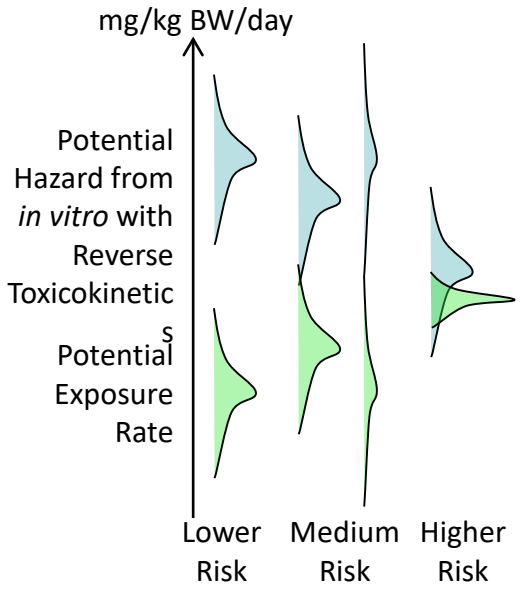


Chemical Prioritization NAMs



High throughput *in vitro* screening can estimate doses needed to cause bioactivity (e.g., Wetmore et al., 2015)

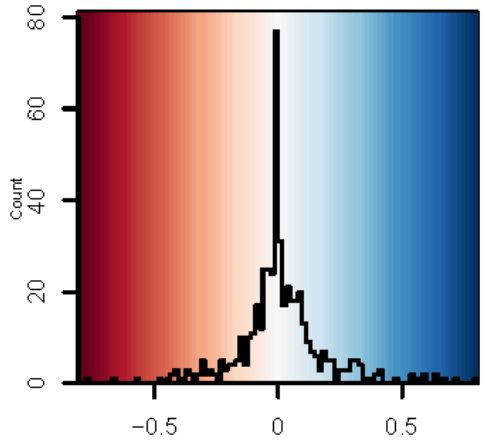
Exposure intake rates can be inferred from biomarkers (e.g., Ring et al., 2018)



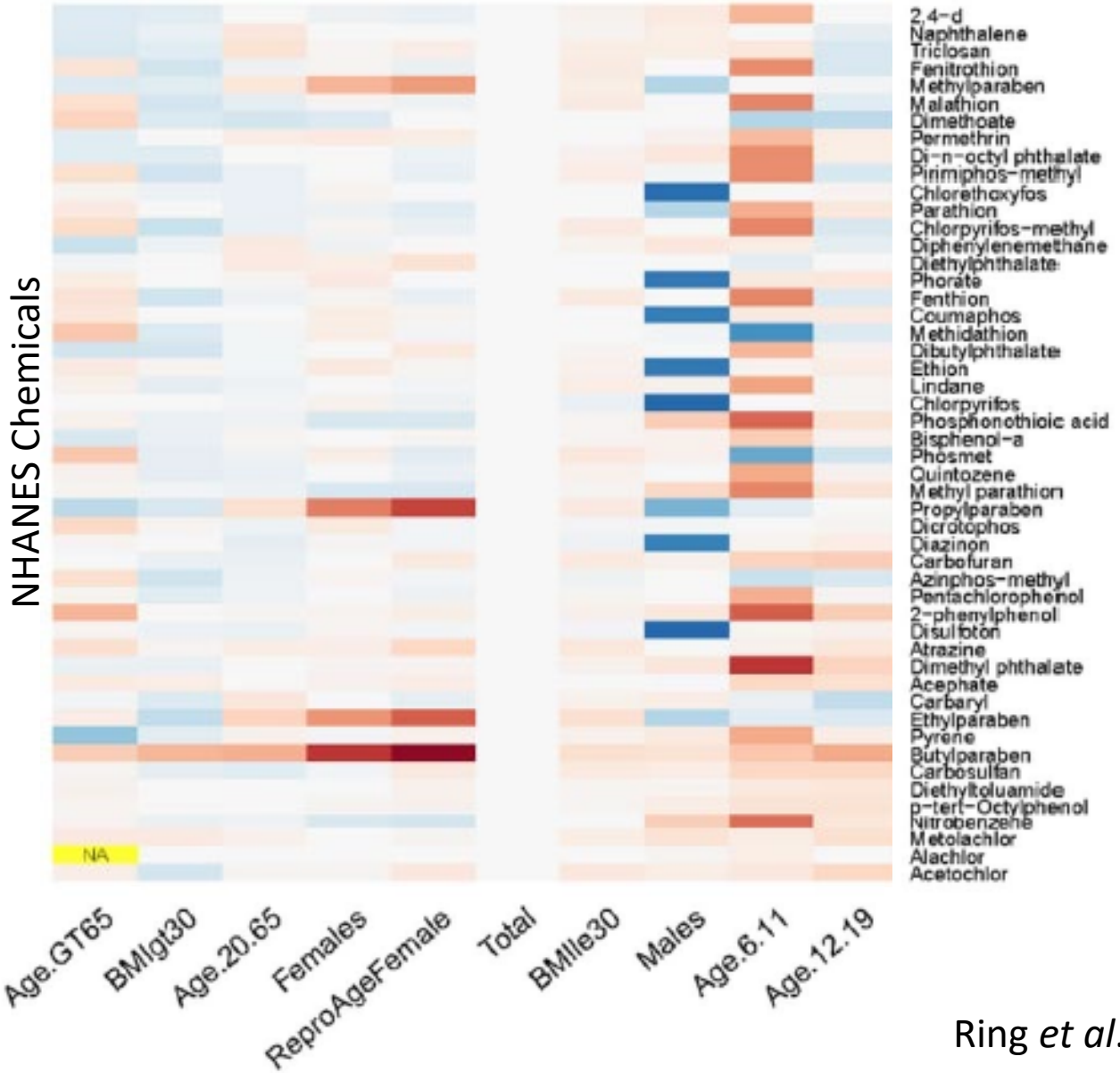
Ring *et al*. (2017)

Life-stage and Demographic Variation in Exposure

- Wambaugh *et al.* (2014) made steady-state inferences of exposure rate (mg/kg/day) from NHANES data for various demographic groups



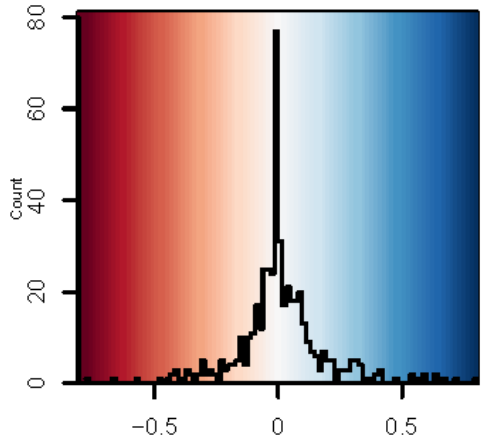
Change in Exposure Relative to Total Population



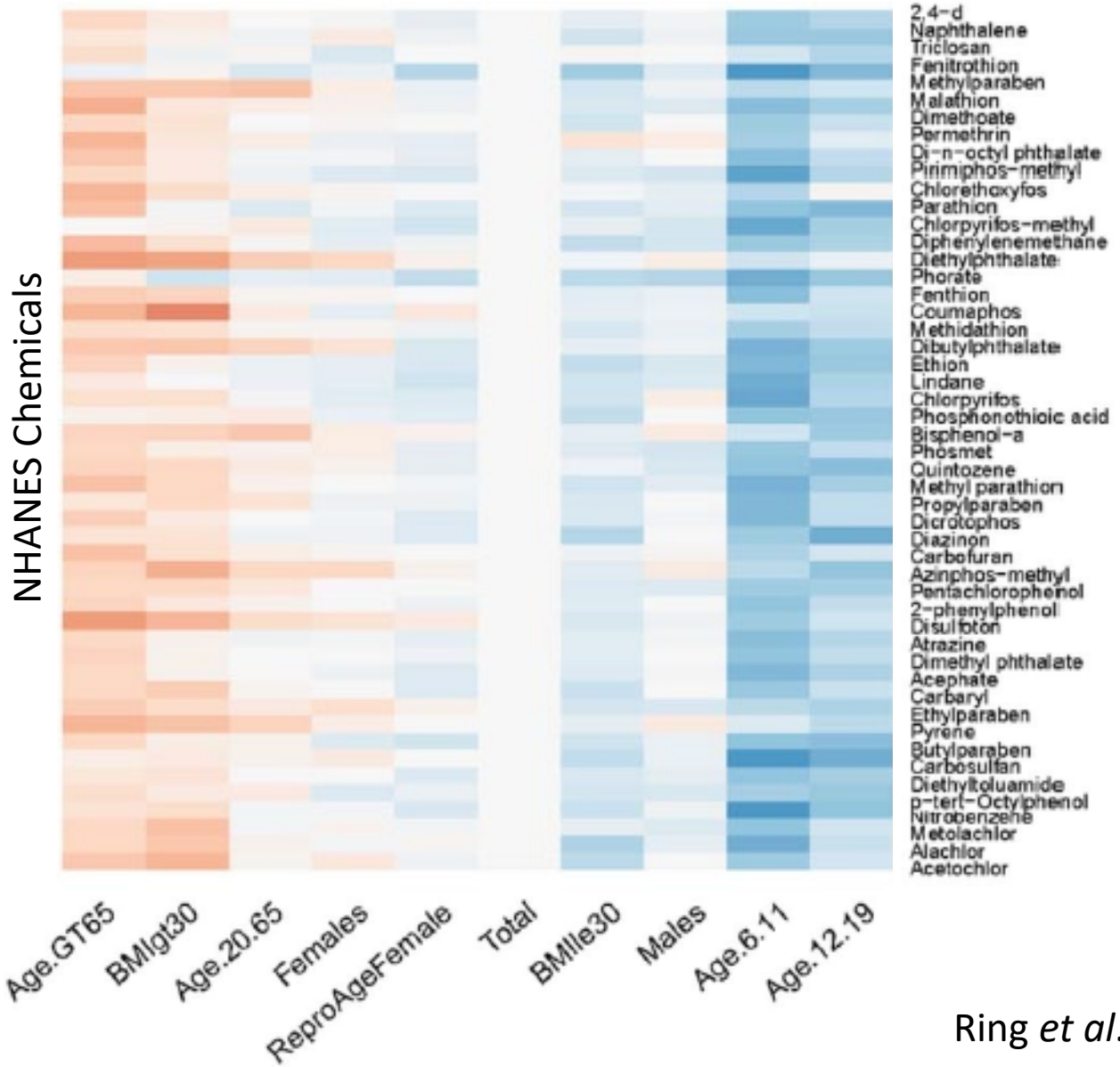
Ring *et al.* (2017)

Life-stage and Demographic Variation in TK

- Ring *et al.* (2017) made demographic-specific predictions of change in plasma concentrations for a 1 mg/kg bw/day exposure



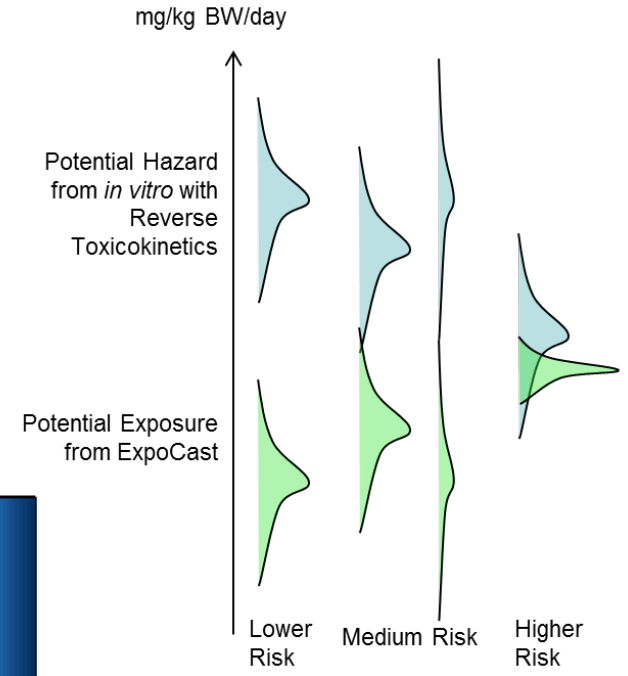
Change in Toxicokinetics Relative to Total Population

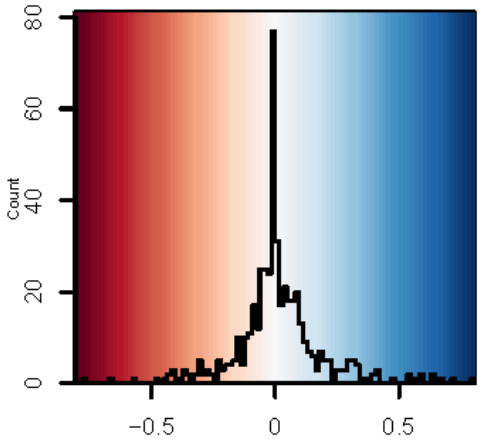


Ring *et al.* (2017)

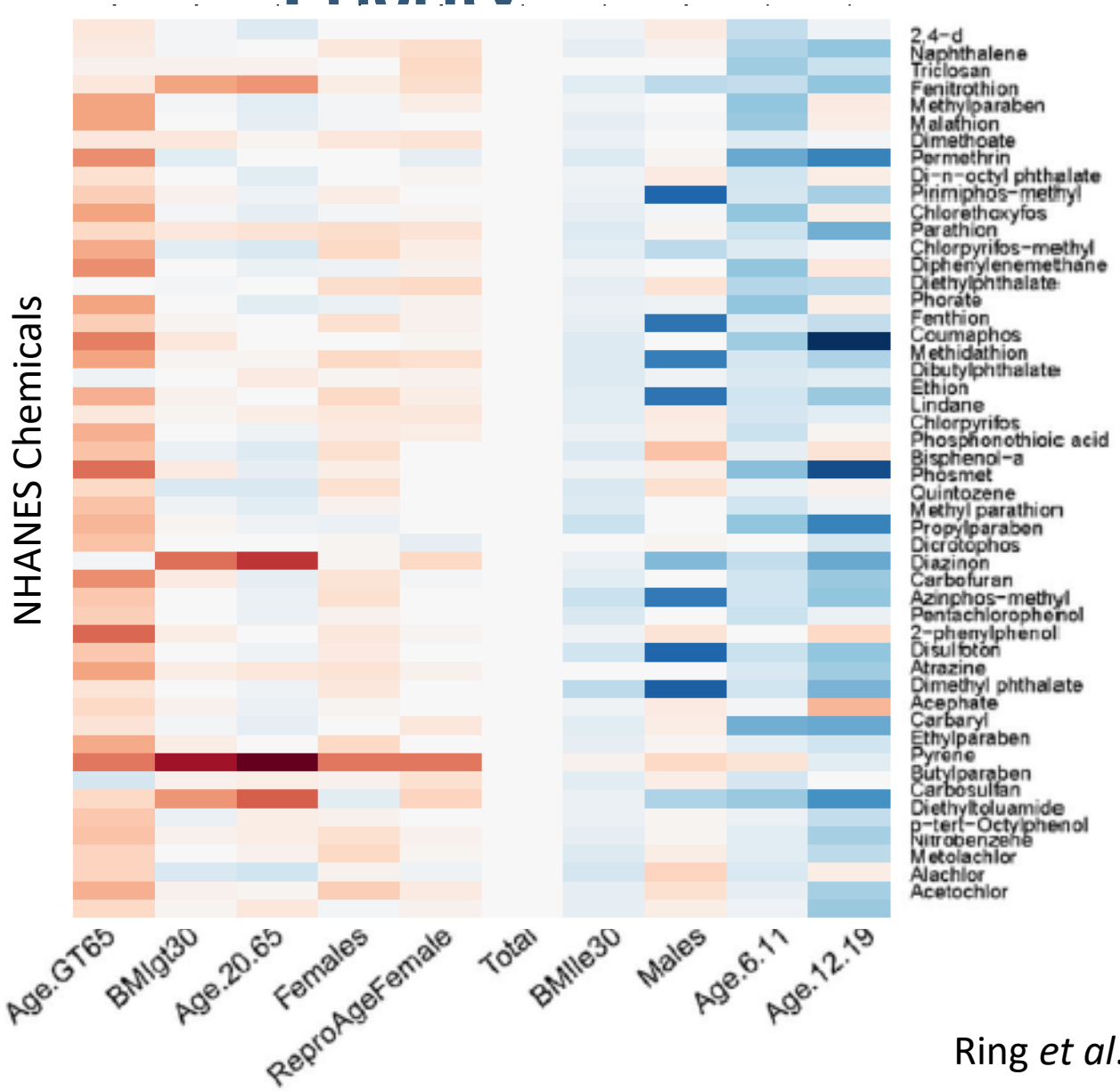
Life-stage and Demographic Variation in Risk Priority

- Can calculate margin between bioactivity and exposure for specific populations





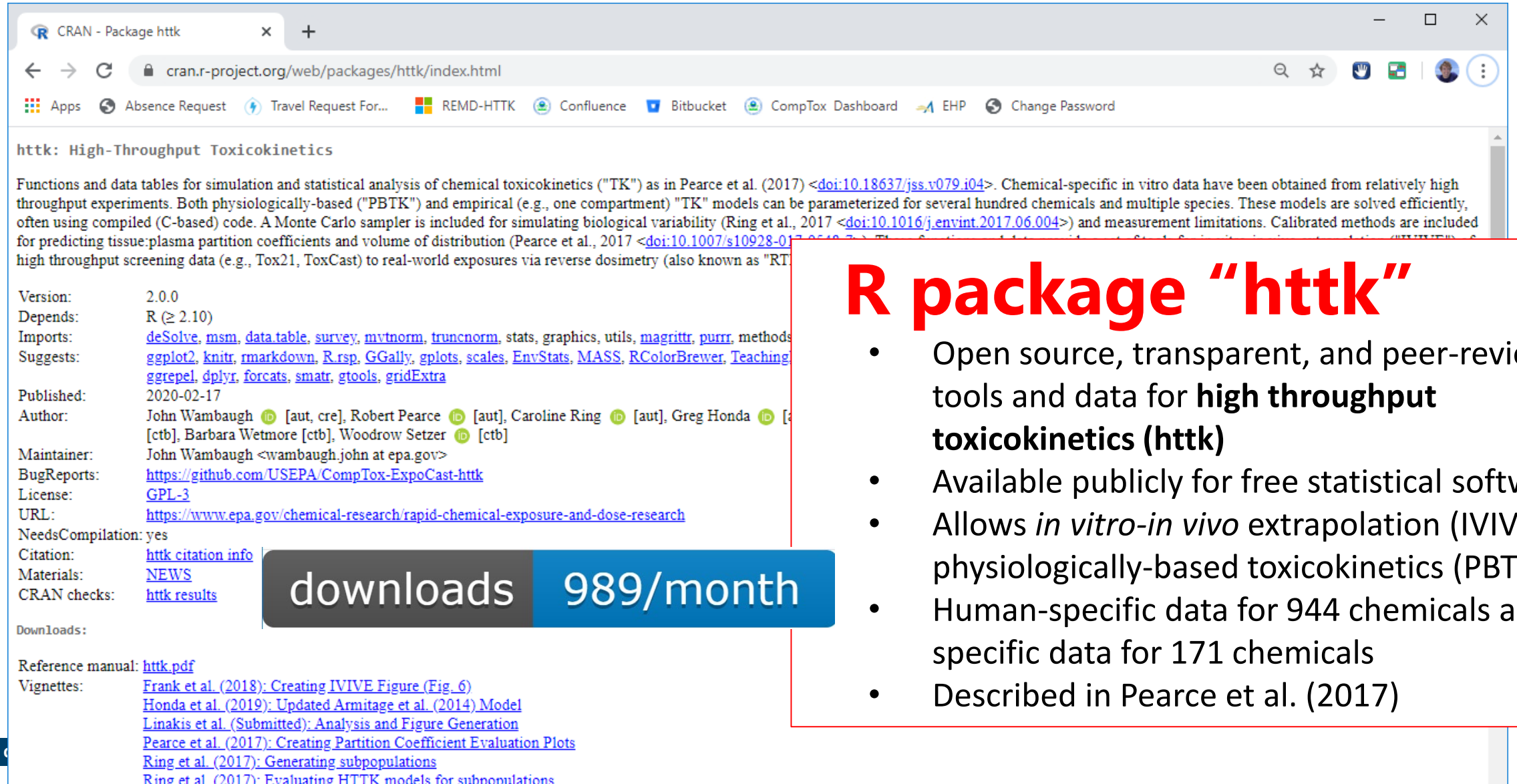
Change in Risk Relative to Total Population



Ring *et al.* (2017)

Open Source Tools and Data for HTTK

https://CRAN.R-project.org/package=httk

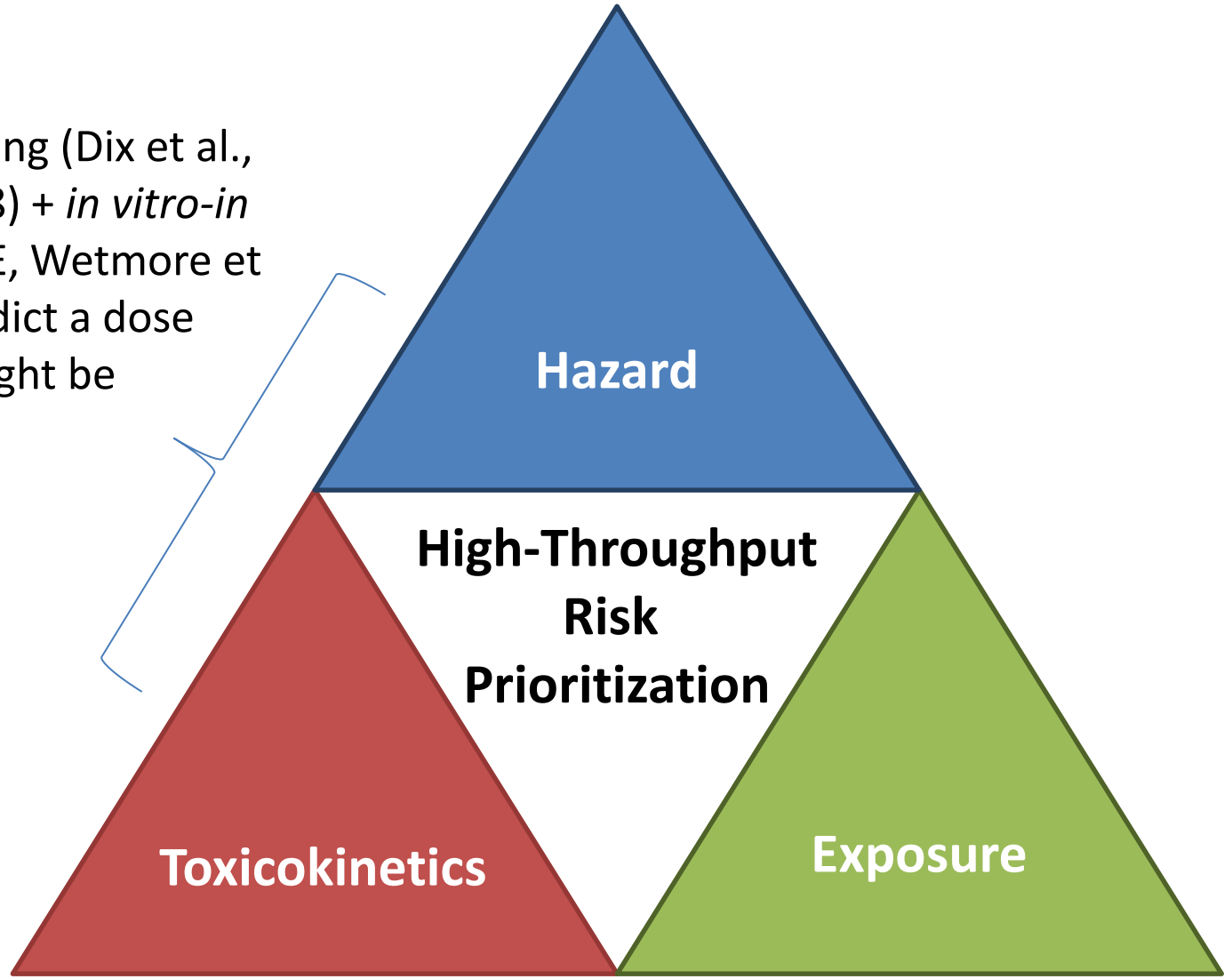


R package "httk"

- Open source, transparent, and peer-reviewed tools and data for **high throughput toxicokinetics (httk)**
- Available publicly for free statistical software R
- Allows *in vitro-in vivo* extrapolation (IVIVE) and physiologically-based toxicokinetics (PBTK)
- Human-specific data for 944 chemicals and rat-specific data for 171 chemicals
- Described in Pearce et al. (2017)

Risk = Hazard x Exposure

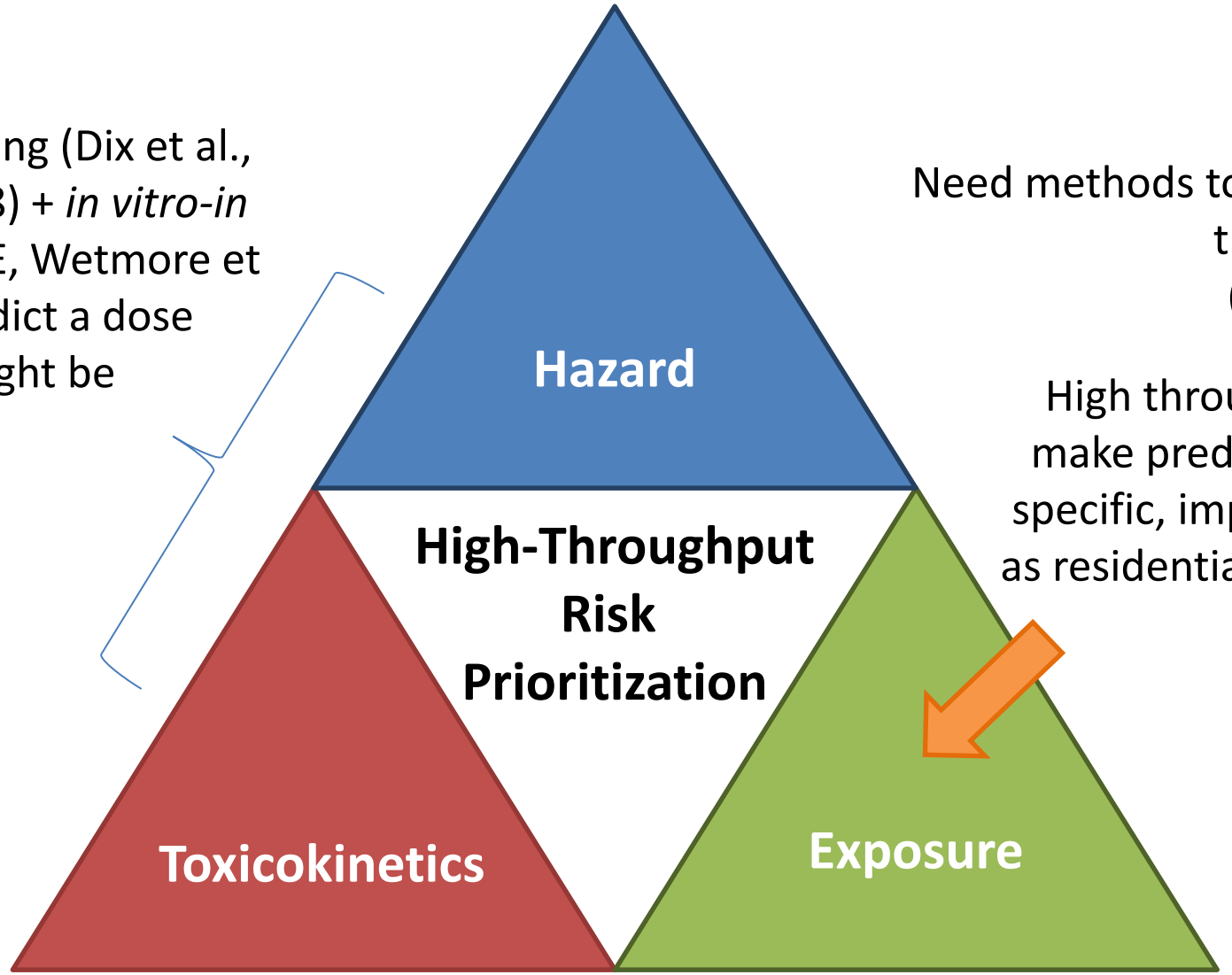
High throughput screening (Dix et al., 2006, Collins et al., 2008) + *in vitro-in vivo* extrapolation (IVIVE, Wetmore et al., 2012, 2015) can predict a dose (mg/kg bw/day) that might be adverse



NRC (1983)

Risk = Hazard x Exposure

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



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

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

NRC (1983)

Understanding Exposure is a Systems Problem

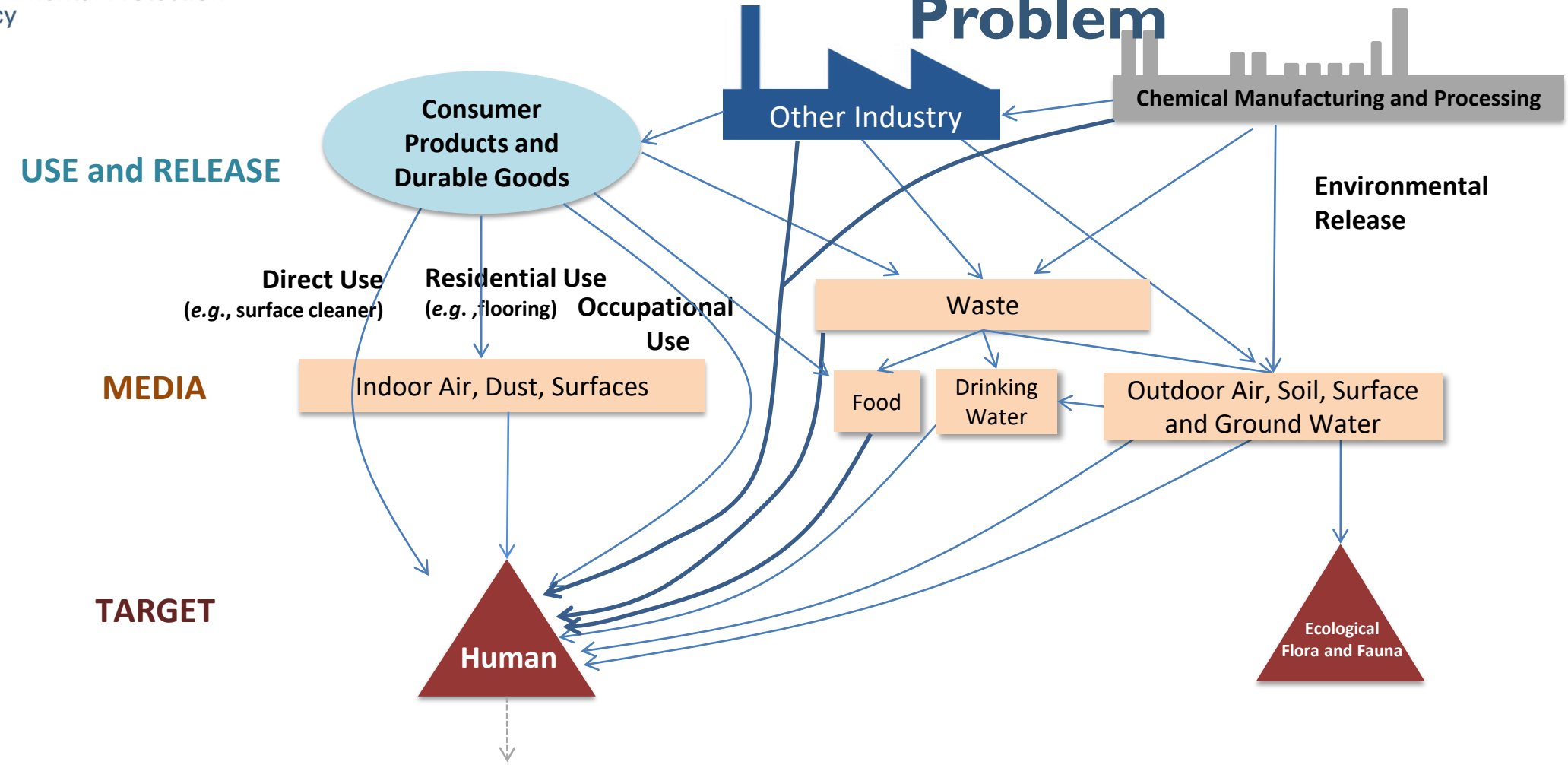
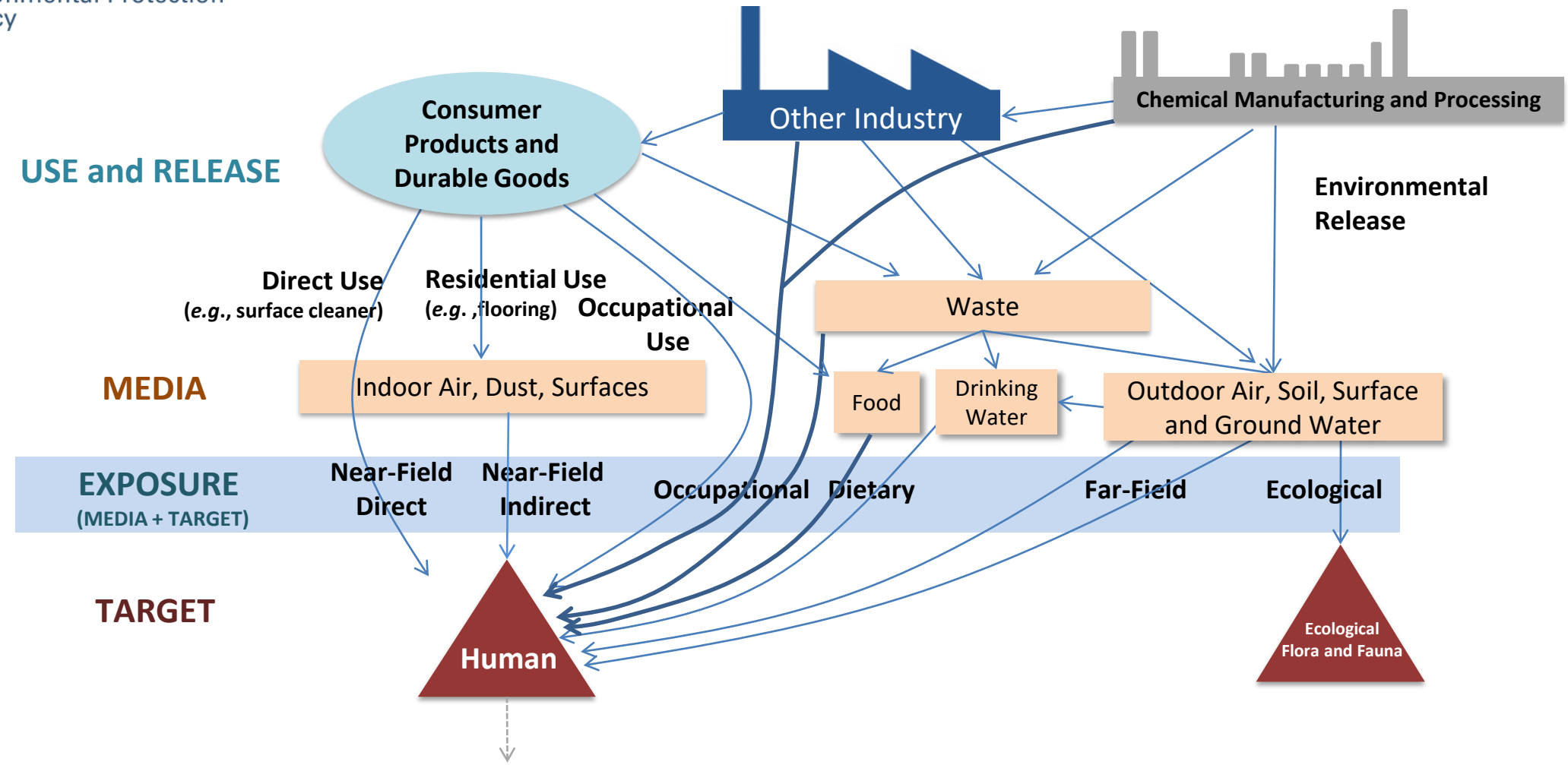


Figure from Kristin Isaacs

Exposure event is often unobservable

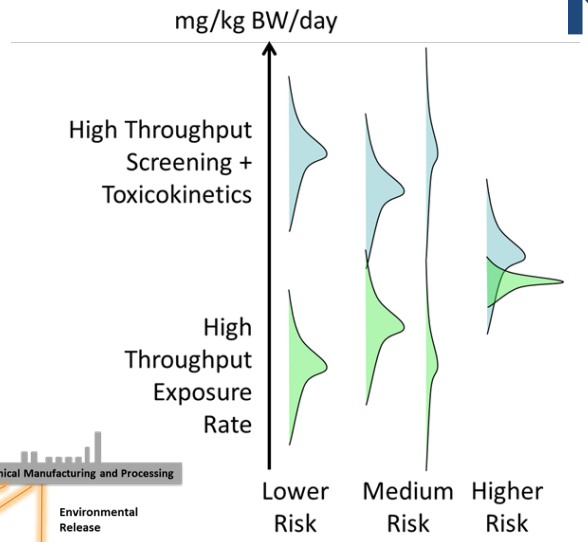


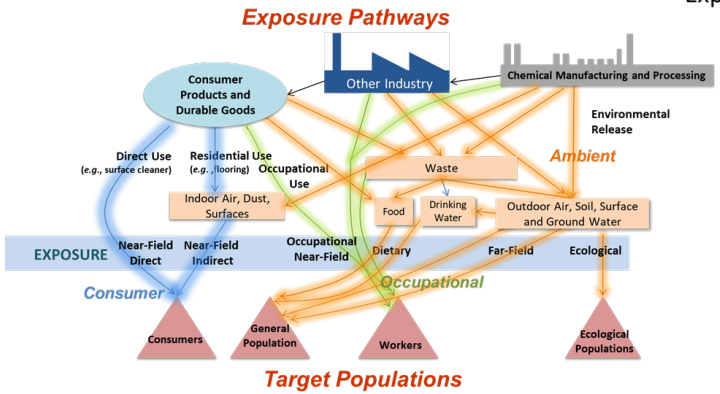
- Can try to predict exposure by characterizing pathway
- Some pathways have much higher average exposures: In home "Near field" sources significant (Wallace, *et al.,* 1987)

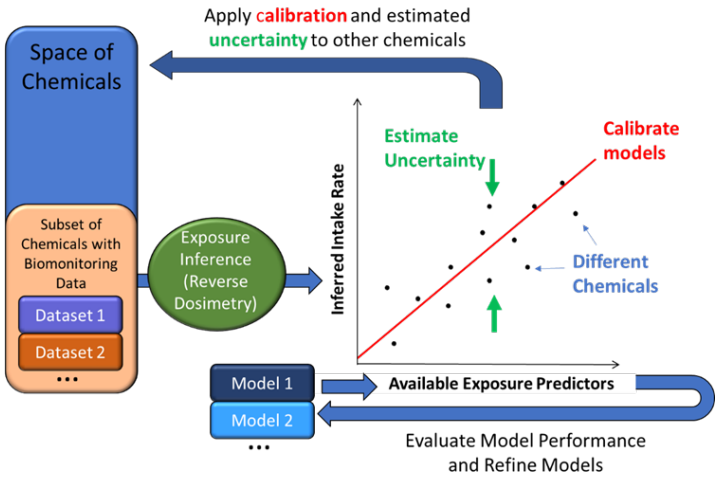
Figure from Kristin Isaacs

NAMs for Exposure Science

United States Environmental Protection Agency







Author's Personal Copy
Available online at www.sciencedirect.com
ScienceDirect

Current Opinion in Toxicology

ELSEVIER

New approach methodologies for exposure science

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

Abstract

Chemical risk assessment relies on knowledge of hazard, the dose–response relationship, and exposure to characterize potential risks to public health and the environment. A chemical with minimal toxicity might pose a risk if exposures are extensive, repeated, and/or occurring during critical windows across the human life span. Exposure assessment involves understanding human activity, and this activity is confounded by interindividual variability that is both biological and behavioral. Exposures further vary between the general population and susceptible or occupationally exposed populations. Recent computational exposure efforts have tackled these problems through the creation of new tools and predictive models. These tools include machine learning to draw inferences from existing data and computer-enhanced screening analyses to generate new data. Mathematical models provide frameworks describing

[9] Department of Earth and Environmental Sciences, University of Texas, Arlington, TX 76019, USA
[10] National Health and Environmental Effects Research Laboratory, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Corresponding author: Wambaugh, John F. (Wambaugh.john@epa.gov)

Current Opinion in Toxicology 2019, 15:76–92

This review comes from a themed issue on **Risk Assessment in Toxicology**

Edited by **Anne Marie Vinggaard** and **Richard Judson**

Available online 31 July 2019

For a complete overview see the Issue and the Editorial

https://doi.org/10.1016/j.cotox.2019.07.001

New Approach Methodologies for Exposure Science

Exposure NAM Class	Description	Traditional Approach	Makes Use of: Measurement	Toxicokinetics	Models	Descriptors	Evaluation	Machine Learning
Measurements	New techniques including screening analyses capable of detecting hundreds of chemicals present in a sample	Targeted (chemical-specific) analyses	-	●	●	●		●
Toxicokinetics	High throughput methods using in vitro data to generate chemical-specific models	Analyses based on in vivo animal studies	●	-		●		●
HTE Models	Models capable of making predictions for thousands of chemicals	Models requiring detailed, chemical- and scenario-specific information	●	●	-	●		
Chemical Descriptors	Informatic approaches for organizing chemical information in a machine-readable format	Tools targeted at single chemical analyses by humans				-		●
Evaluation	Statistical approaches that use the data from many chemicals to estimate the uncertainty in a prediction for a new chemical	Comparison of model predictions to data on a per chemical basis	●	●	●	●	-	●
Machine Learning	Computer algorithms to identify patterns	Manual Inspection of the Data	●	●		●		-
Prioritization	Integration of exposure and other NAMs to identify chemicals for follow-up study	Expert decision making	●	●	●	●	●	●

Wambaugh et al. (2019)

What Do We Know About Exposure? Biomonitoring Data

- Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey (NHANES) provides an important tool for monitoring public health
- Large, ongoing CDC survey of US population: demographic, body measures, medical exam, biomonitoring (health and exposure), …
- Designed to be representative of US population according to census data
- Data sets publicly available (http://www.cdc.gov/nchs/nhanes.htm)
- Includes measurements of:
    - Body weight
    - Height
    - **Chemical analysis of blood and urine**

What Do We Know About Exposure? Exposure Models

- Human chemical exposures can be coarsely grouped into "**near field**" sources that are close to the exposed individual (consumer or occupational exposures) '**far-field**' scenarios wherein individuals are exposed to chemicals that were released or used far away (ambient exposure) (Arnot *et al.*, 2006).

- A model captures knowledge and a hypothesis of how the world works (MacLeod *et al.*, 2010)

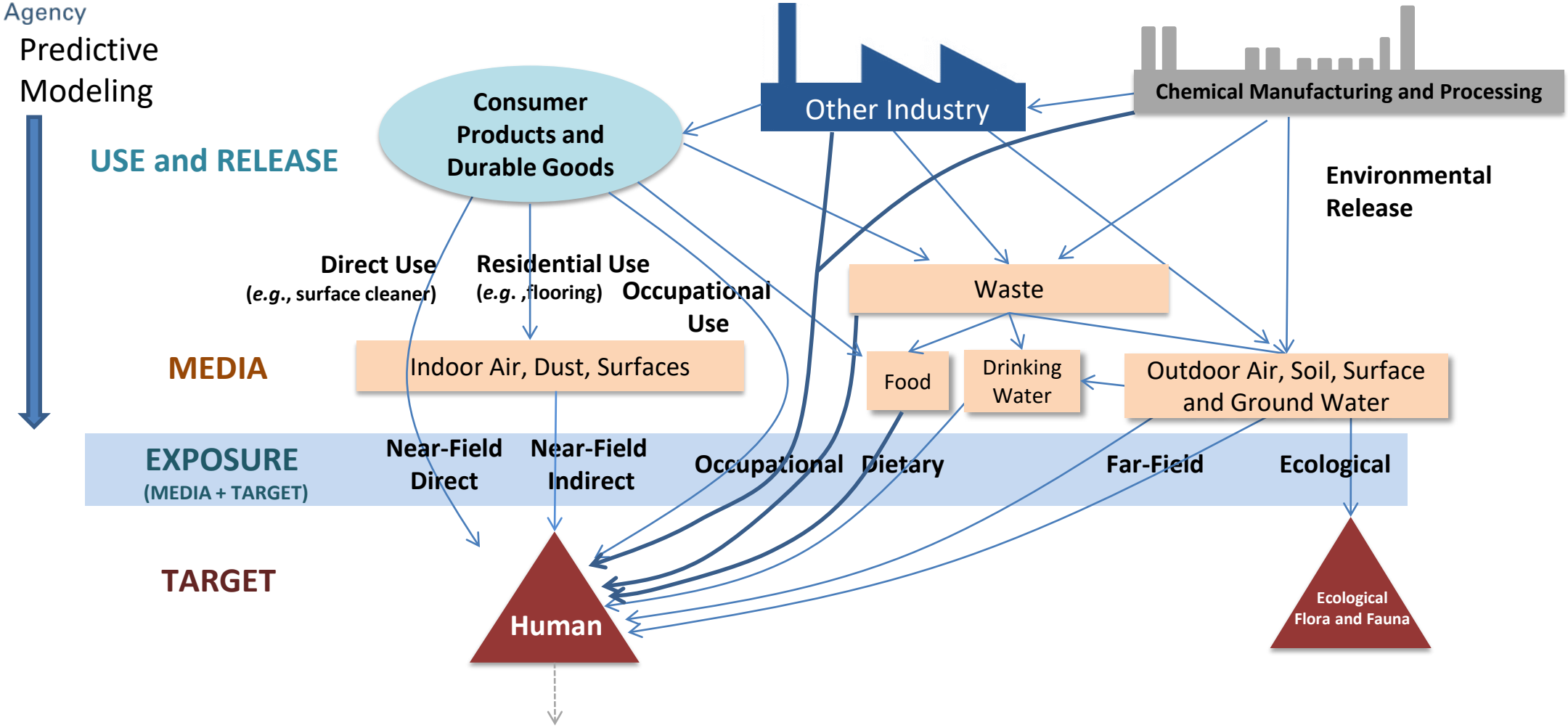
- EPA's EXPOsure toolBOX (EPA ExpoBox) is a toolbox created to assist individuals from within government, industry, academia, and the general public with assessing exposure
  - Includes many, many models

  https://www.epa.gov/expobox

"Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations… The only question of interest is 'Is the model illuminating and useful?'" George Box

Models to Predict Exposure

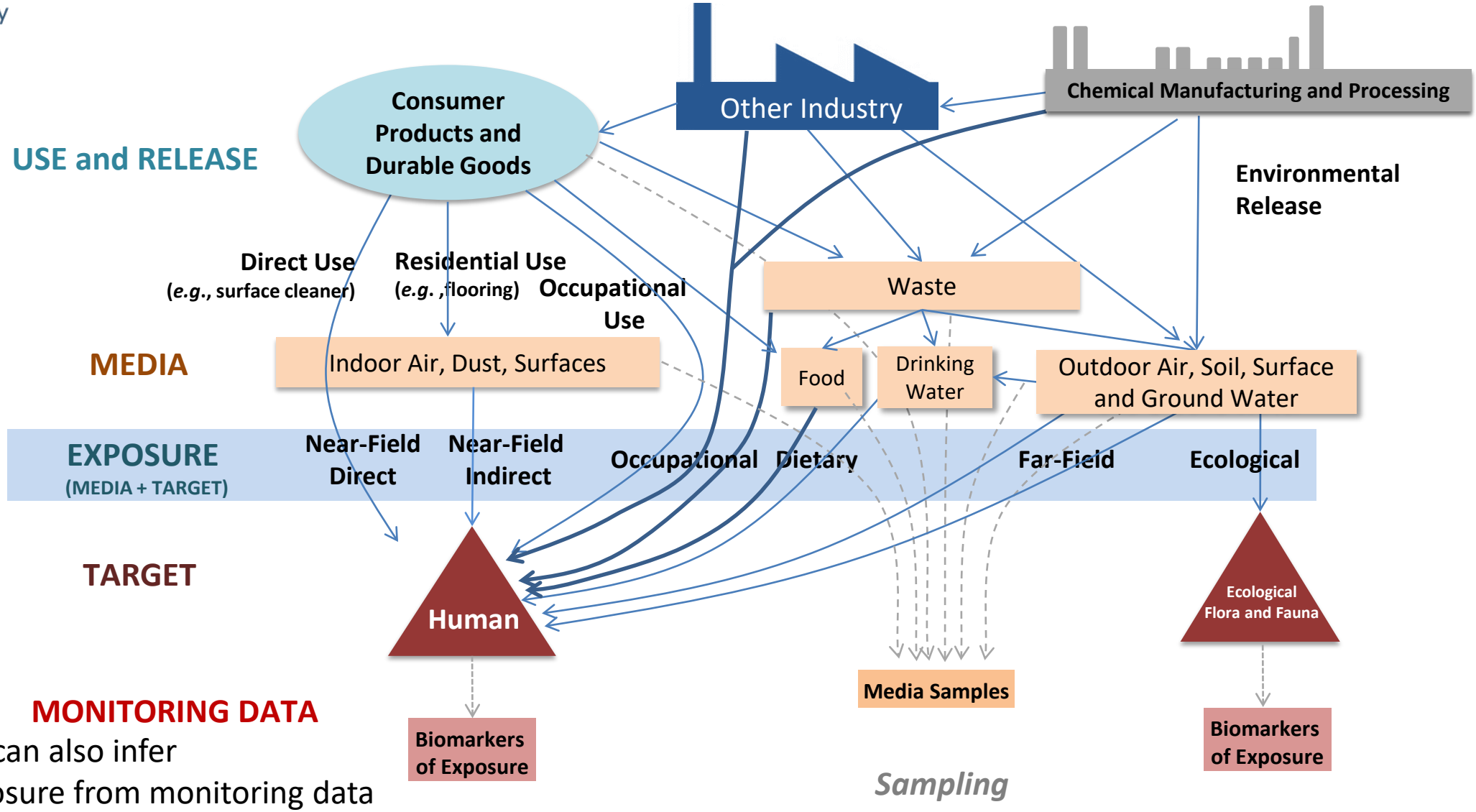
Predictive Modeling



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

Figure from Kristin Isaacs

Monitoring Data



MONITORING DATA

We can also infer exposure from monitoring data

Figure from Kristin Isaacs

Models to Infer Exposure

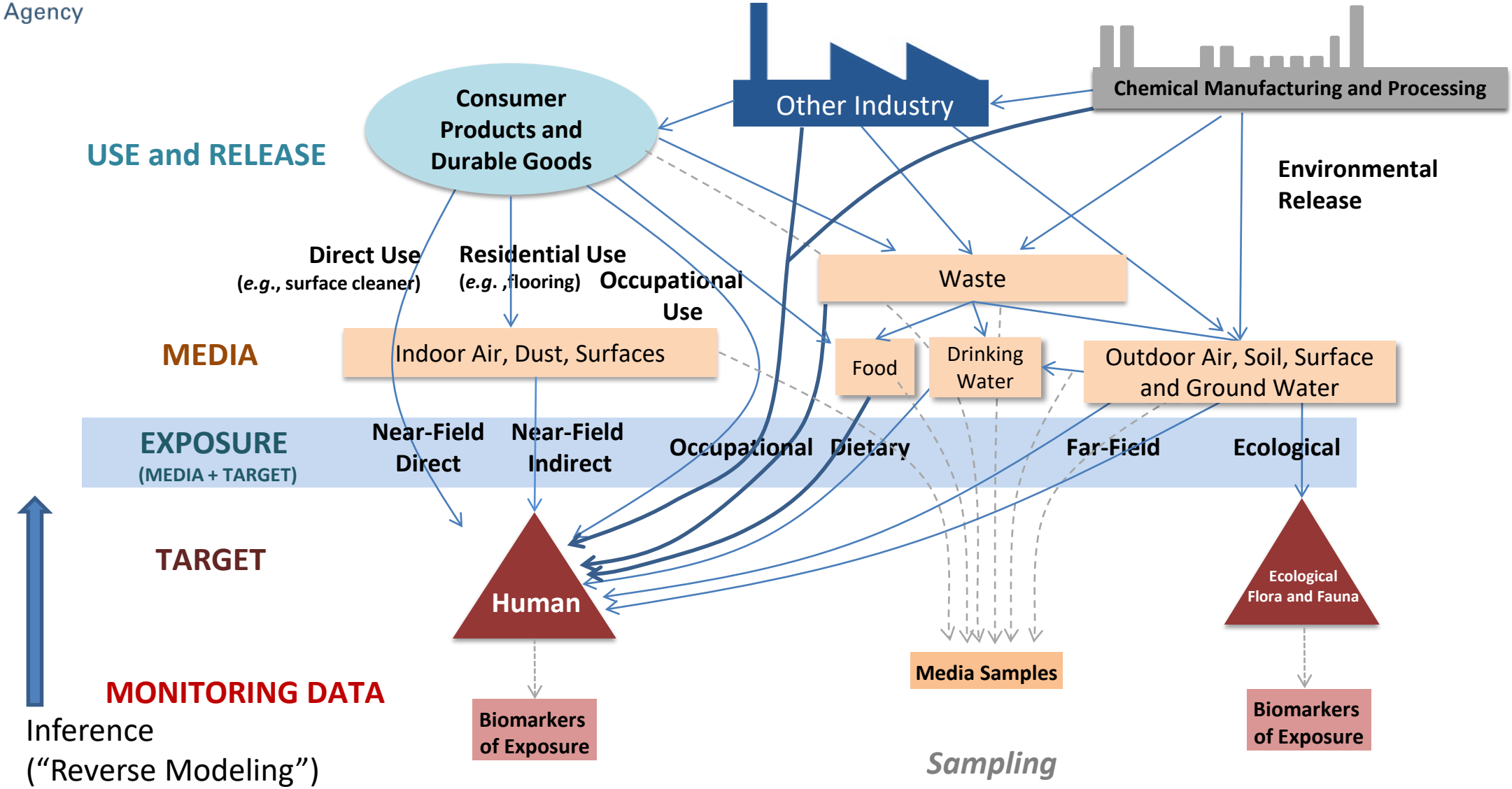


Figure from Kristin Isaacs

Evaluating Models with Monitoring Data

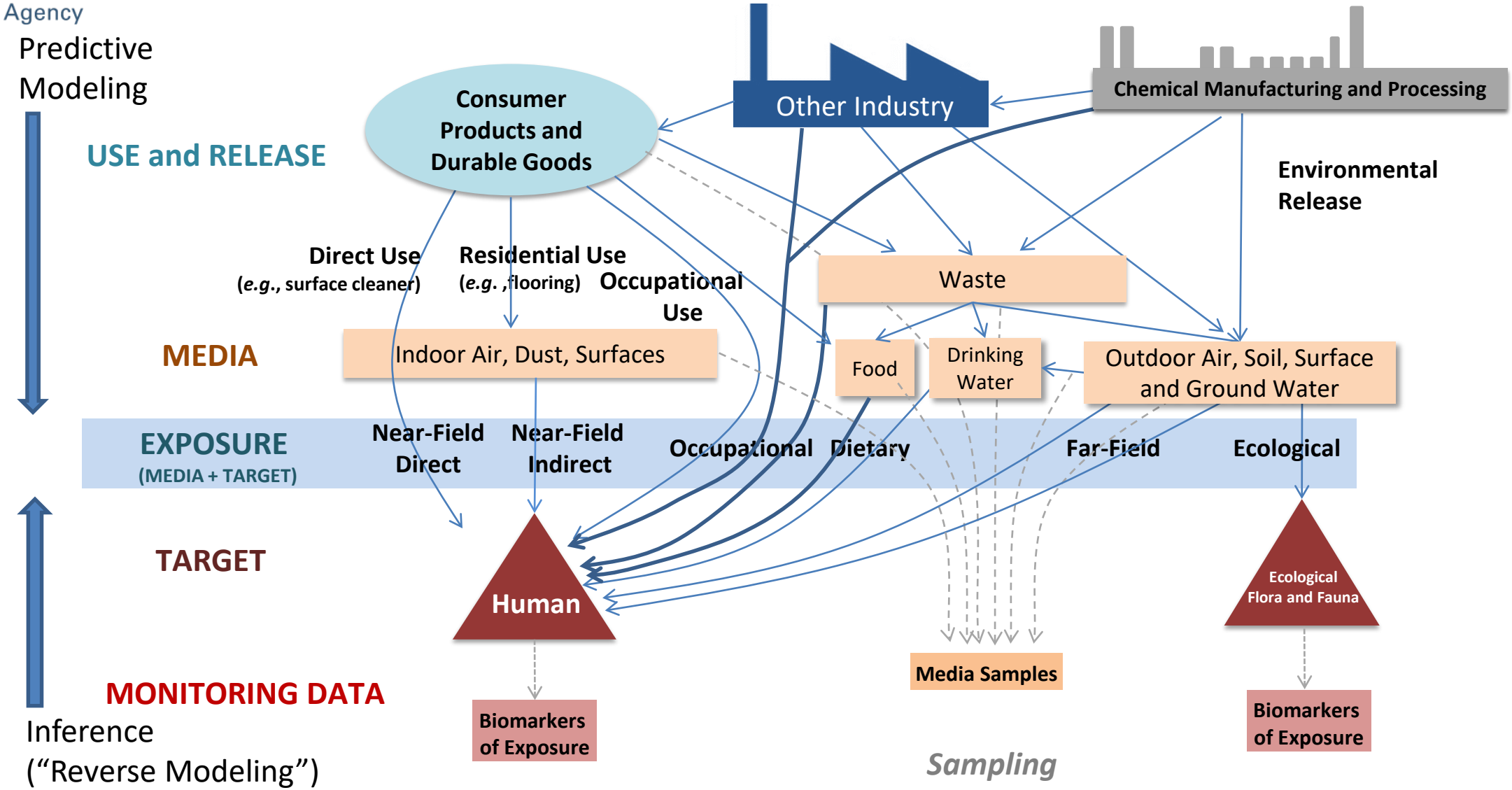
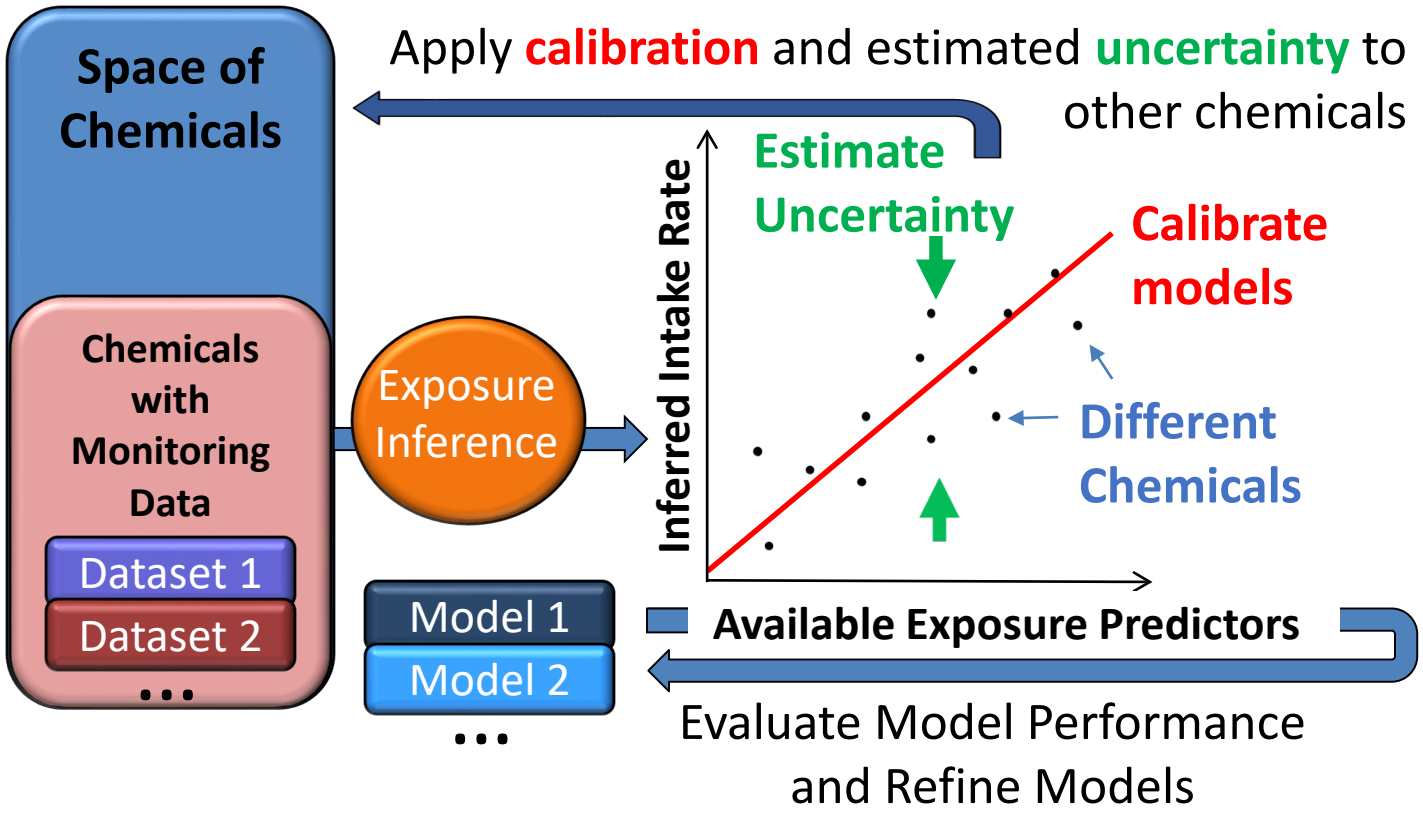
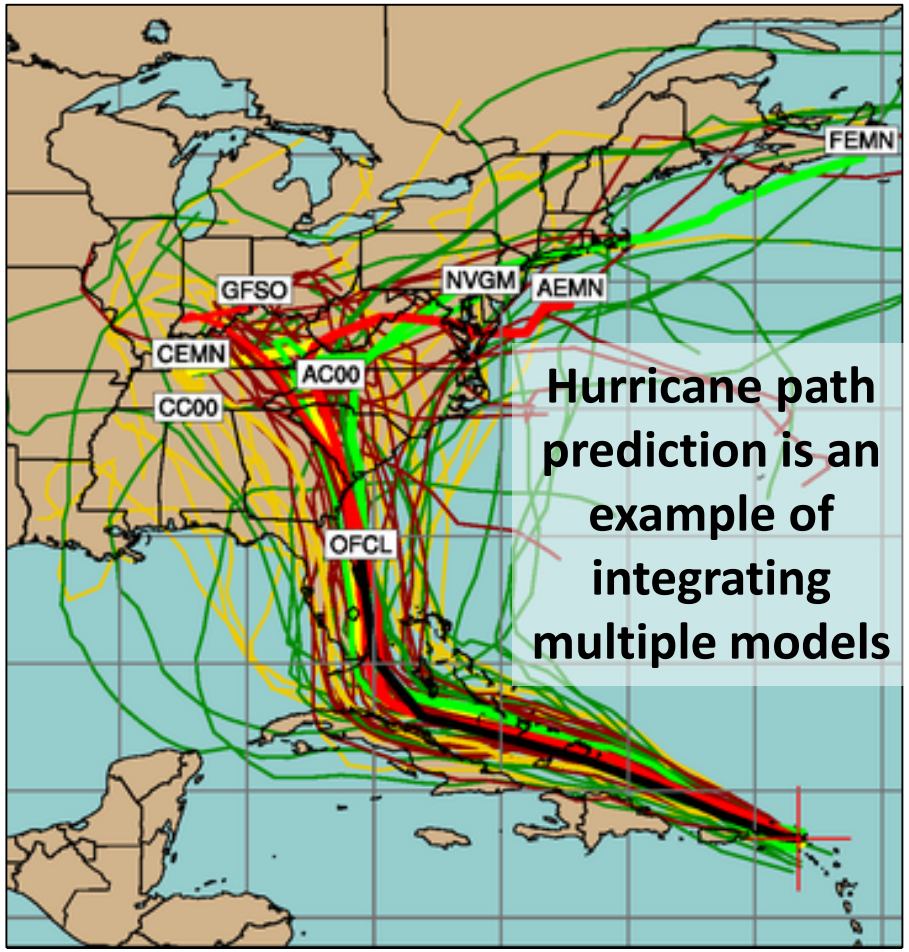


Figure from Kristin Isaacs

Evaluation NAMs: The SEEM Framework

- We use Bayesian methods to incorporate multiple models into consensus predictions for 1000s of chemicals within the **Systematic Empirical Evaluation of Models (SEEM)** (Wambaugh et al., 2013, 2014; Ring et al., 2018)

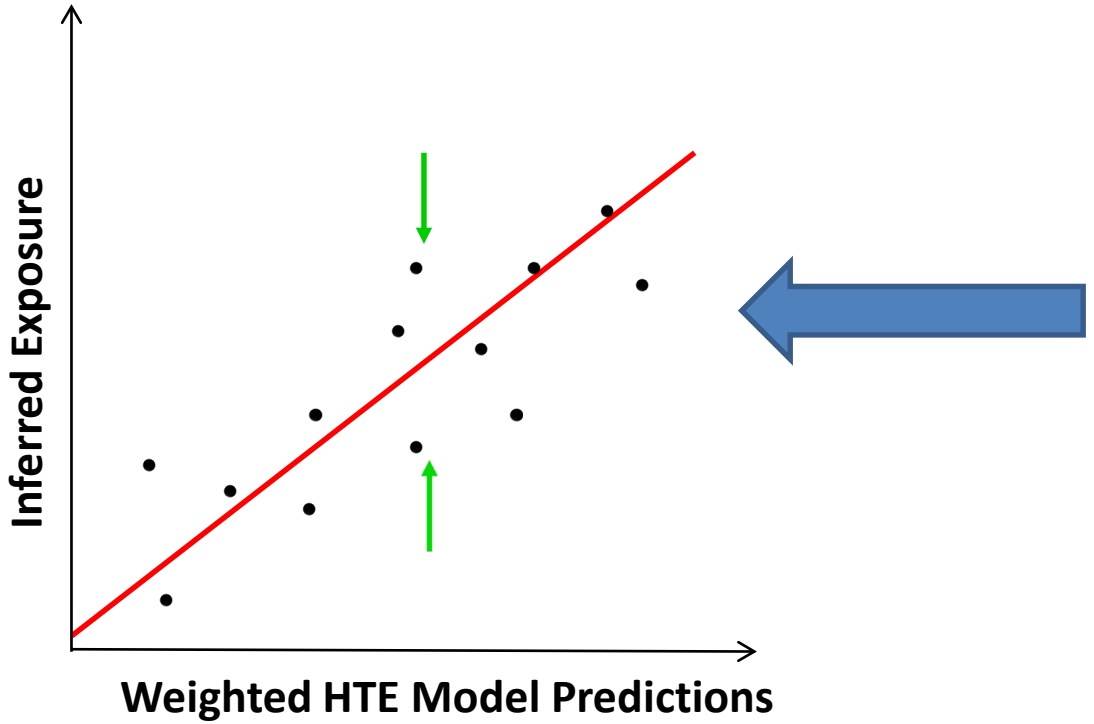




SEEM is a Linear Regression

Multiple regression models:

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

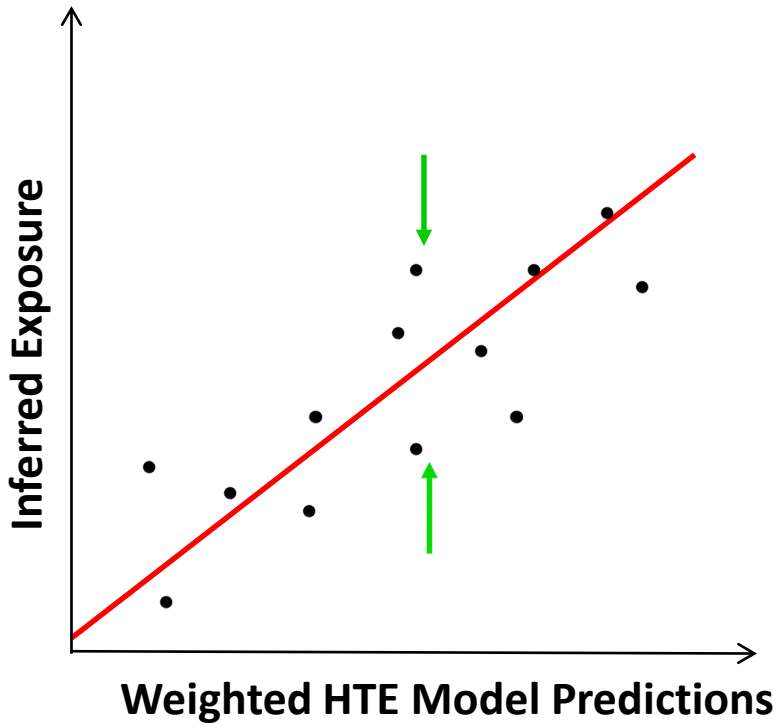


$\varepsilon \sim N(0, \sigma^2)$
Residual error, unexplained by the regression model

SEEM is a Linear Regression

Multiple regression models:

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



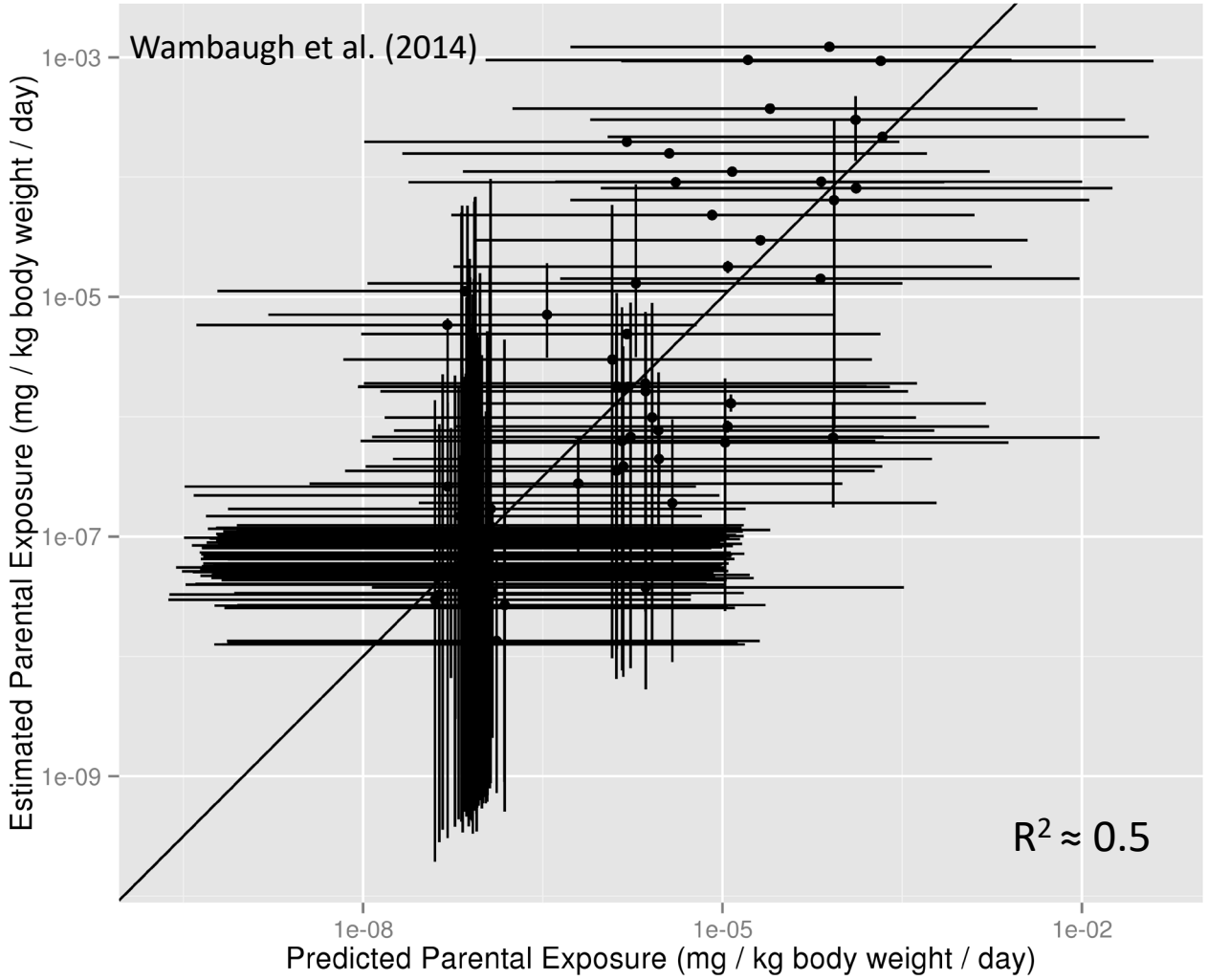
Not all models have predictions for all chemicals

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

What do we do for the rest?

- Assign the average value?
- Zero?

SEEM Analysis of NHANES Data

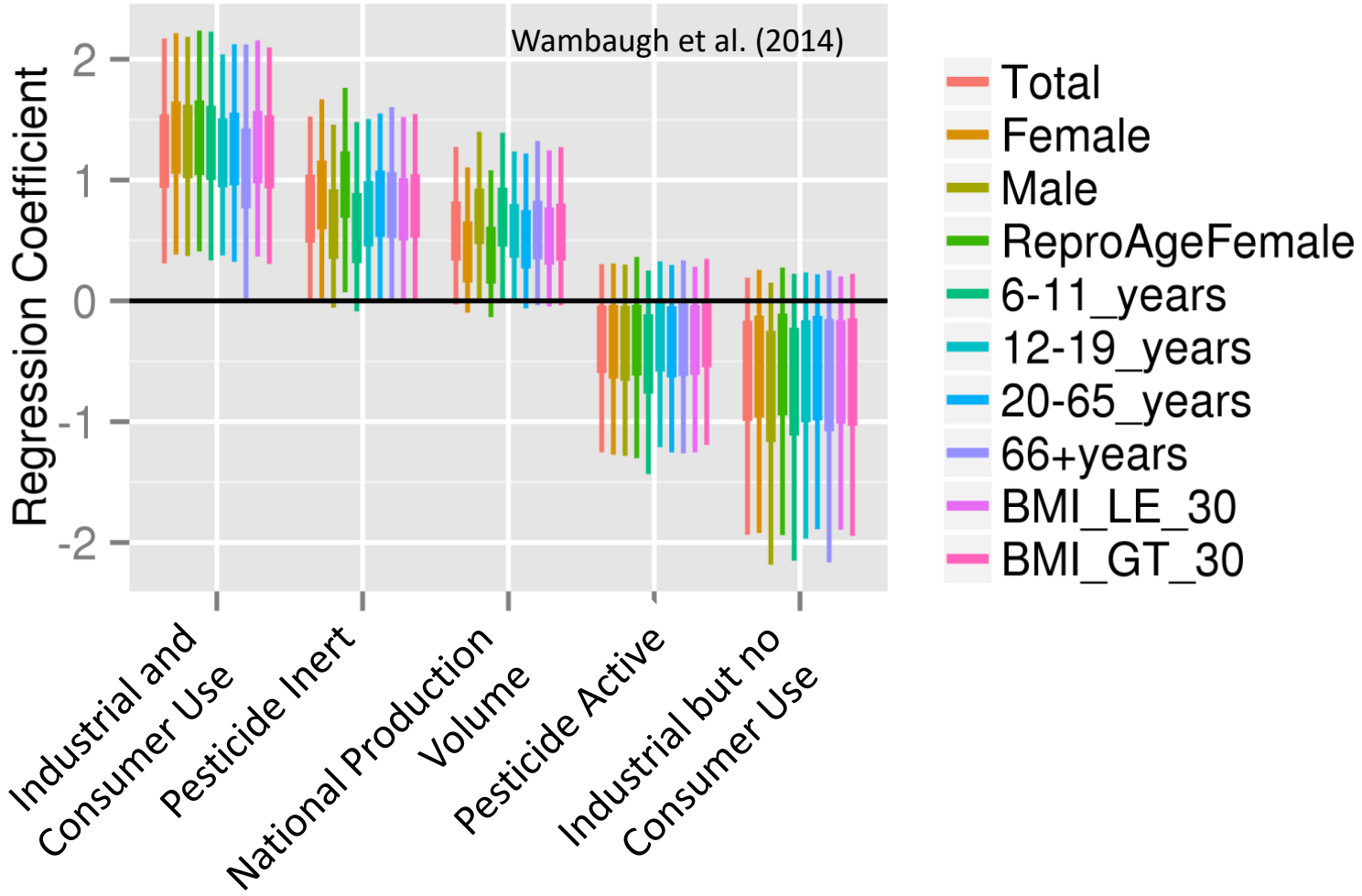


$R^2 \approx 0.5$ indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates

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

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

Heuristics of Exposure



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

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

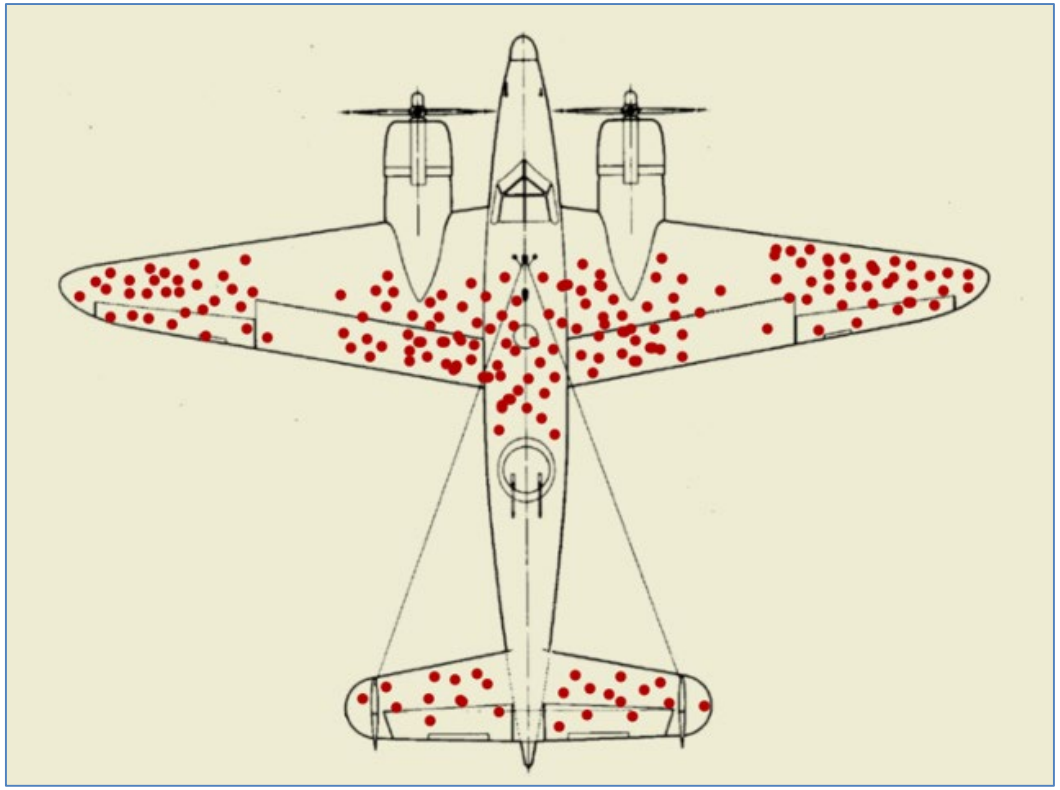
- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume

Correlation is Not Causation

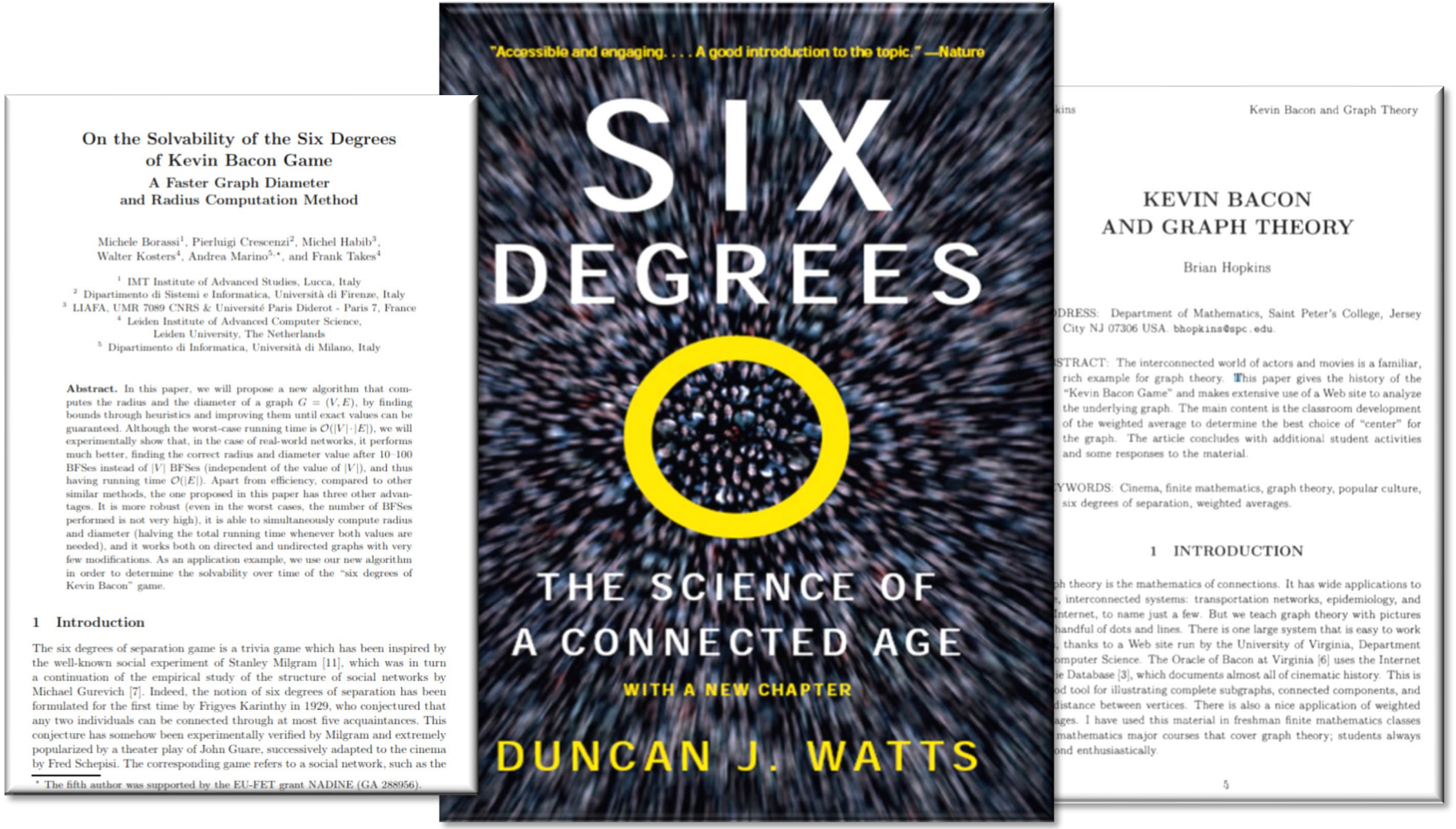
- Wambaugh et al. (2014) found that “pesticide inerts” had higher than average levels in biomonitoring data, while “pesticide actives” had lower than average

- In World War II, there Royal Air Force (UK) wanted to armor planes against anti-aircraft fire
  - Initial proposal was to place armor wherever bullet holes were most common
  - Mathematician Abraham Wald pointed out that they were looking at the planes that had returned
  - *See Drum, Kevin (2010) “The Counterintuitive World”*

- Pesticide inerts have many other uses, but there are more stringent reporting requirements for pesticides
  - **Exposure is occuring by other pathways**



The Six Degrees of Kevin Bacon



On the Solvability of the Six Degrees of Kevin Bacon Game

A Faster Graph Diameter and Radius Computation Method

Michele Borassi[1], Pierluigi Crescenzi[2], Michel Habib[3], Walter Kosters[4], Andrea Marino[5,*], and Frank Takes[4]

[1] IMT Institute of Advanced Studies, Lucca, Italy
[2] Dipartimento di Sistemi e Informatica, Università di Firenze, Italy
[3] LIAFA, UMR 7089 CNRS & Université Paris Diderot - Paris 7, France
[4] Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
[5] Dipartimento di Informatica, Università di Milano, Italy

Abstract. In this paper, we will propose a new algorithm that computes the radius and the diameter of a graph $G = (V, E)$, by finding bounds through heuristics and improving them until exact values can be guaranteed. Although the worst-case running time is $\mathcal{O}(|V| \cdot |E|)$, we will experimentally show that, in the case of real-world networks, it performs much better, finding the correct radius and diameter value after 10–100 BFSes instead of $|V|$ BFSes (independent of the value of $|V|$), and thus having running time $\mathcal{O}(|E|)$. Apart from efficiency, compared to other similar methods, the one proposed in this paper has three other advantages. It is more robust (even in the worst cases, the number of BFSes performed is not very high), it is able to simultaneously compute radius and diameter (halving the total running time whenever both values are needed), and it works both on directed and undirected graphs with very few modifications. As an application example, we use our new algorithm in order to determine the solvability over time of the "six degrees of Kevin Bacon" game.

1 Introduction

The six degrees of separation game is a trivia game which has been inspired by the well-known social experiment of Stanley Milgram [11], which was in turn a continuation of the empirical study of the structure of social networks by Michael Gurevich [7]. Indeed, the notion of six degrees of separation has been formulated for the first time by Frigyes Karinthy in 1929, who conjectured that any two individuals can be connected through at most five acquaintances. This conjecture has somehow been experimentally verified by Milgram and extremely popularized by a theater play of John Guare, successively adapted to the cinema by Fred Schepisi. The corresponding game refers to a social network, such as the

* The fifth author was supported by the EU-FET grant NADINE (GA 288956).

"Accessible and engaging. . . . A good introduction to the topic." —Nature

SIX DEGREES

THE SCIENCE OF A CONNECTED AGE

WITH A NEW CHAPTER

DUNCAN J. WATTS

KEVIN BACON AND GRAPH THEORY

Brian Hopkins

DDRESS: Department of Mathematics, Saint Peter's College, Jersey City NJ 07306 USA. bhopkins@spc.edu.

STRACT: The interconnected world of actors and movies is a familiar, rich example for graph theory. This paper gives the history of the "Kevin Bacon Game" and makes extensive use of a Web site to analyze the underlying graph. The main content is the classroom development of the weighted average to determine the best choice of "center" for the graph. The article concludes with additional student activities and some responses to the material.

YWORDS: Cinema, finite mathematics, graph theory, popular culture, six degrees of separation, weighted averages.

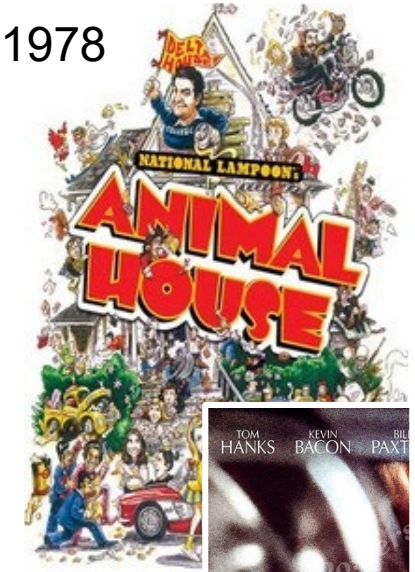
1 INTRODUCTION

h theory is the mathematics of connections. It has wide applications to , interconnected systems: transportation networks, epidemiology, and nternet, to name just a few. But we teach graph theory with pictures handful of dots and lines. There is one large system that is easy to work , thanks to a Web site run by the University of Virginia, Department omputer Science. The Oracle of Bacon at Virginia [6] uses the Internet e Database [3], which documents almost all of cinematic history. This is d tool for illustrating complete subgraphs, connected components, and distance between vertices. There is also a nice application of weighted ages. I have used this material in freshman finite mathematics classes mathematics major courses that cover graph theory; students always ond enthusiastically.

5

Kevin Bacon

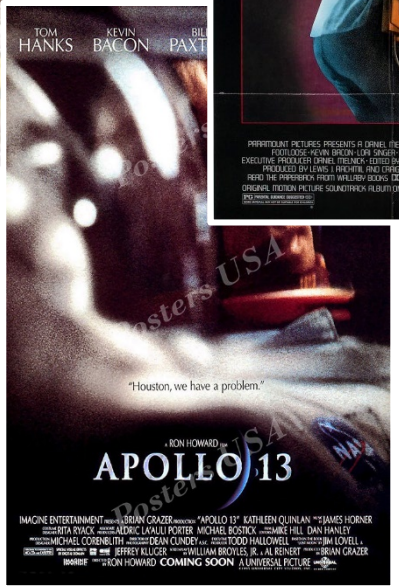
1978

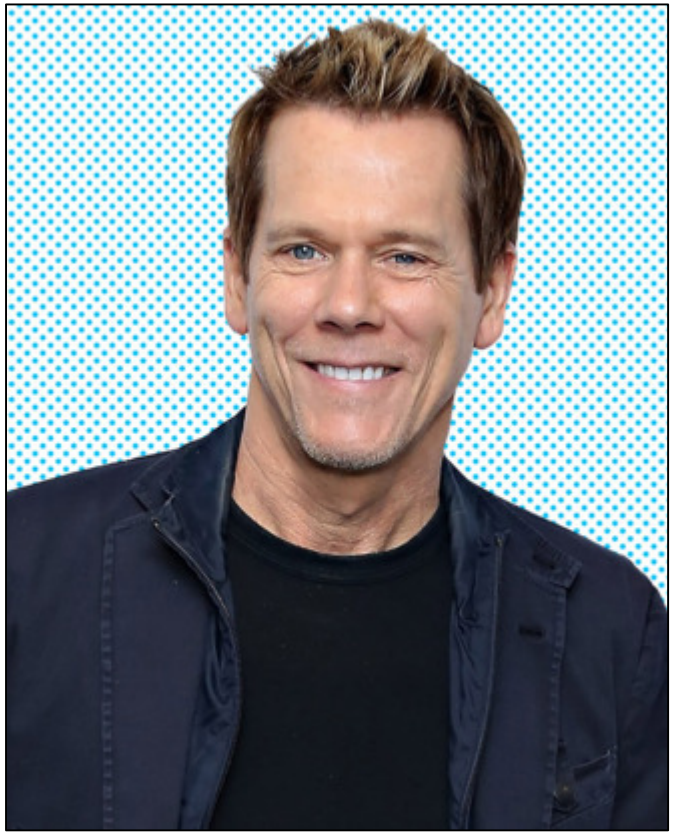


1984

1992

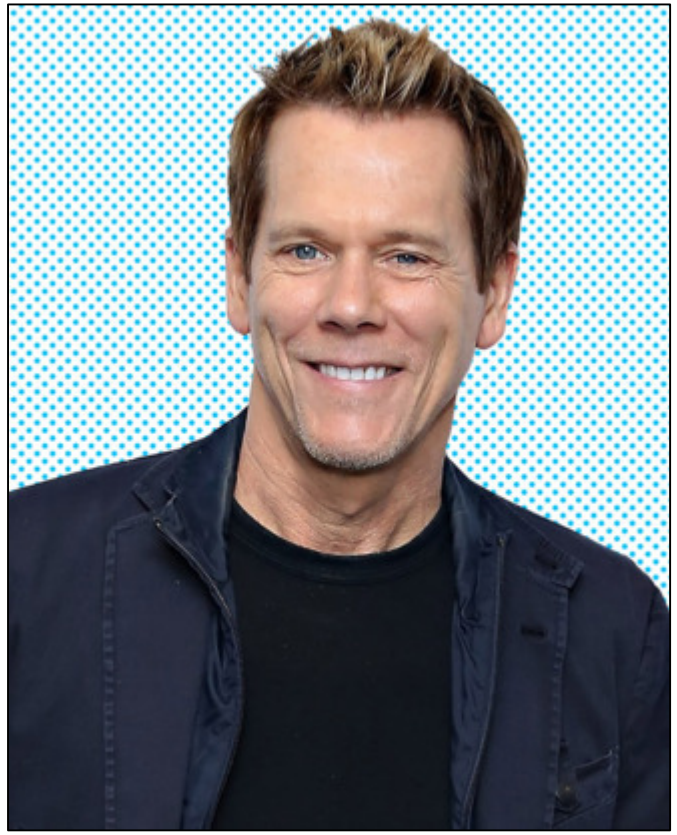
1995





Kevin Bacon

1990



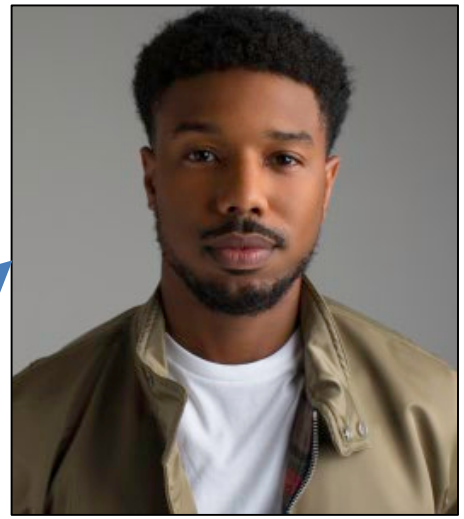
Michael B. Jordan

Connectedness to Michael B. Jordan

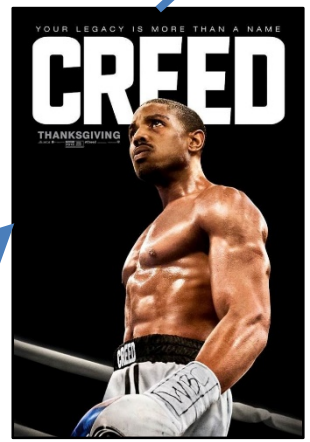
Frances McDormand
Best Actress Winner 2018

Hail Caesar
McDormand &
Channing Tatum

GI Joe: Retaliation
Tatum & Bruce Willis

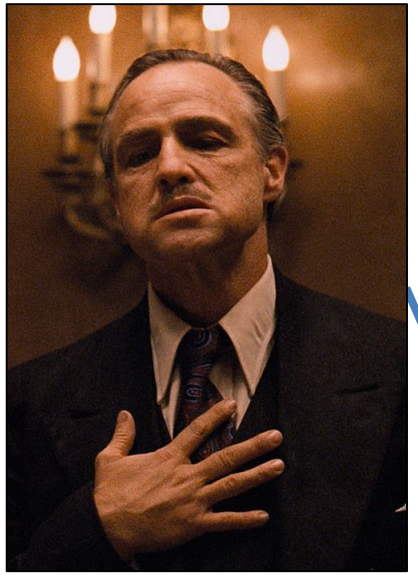


Expendables
Willis &
Sylvester Stallone



Creed
Stallone & Jordan

Connectedness to Michael B. Jordan

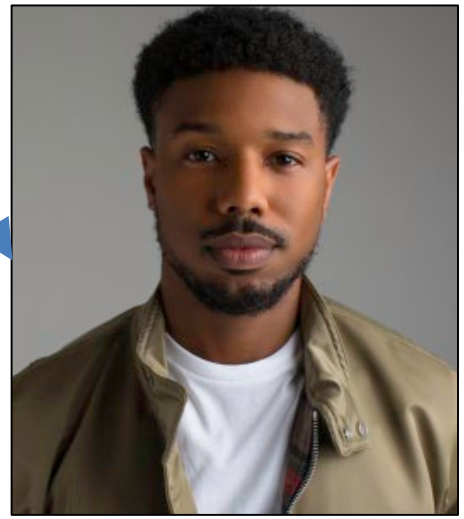


Marlon Brando
Best Actor 1954 and 1972
Died 2004

Avengers: Infinity War
Paltrow & Chadwick Boseman

Superman
with Gene Hackman

Black Panther
Boseman & Jordan



The Royal Tenenbaums
Hackman & Gwyneth Paltrow

Small World Networks

Watts and Strogatz (1998)

letters to nature

typically slower than ~1 km s^{-1}) might differ significantly from what is assumed by current modelling efforts[27]. The expected equation-of-state differences among small bodies (ice versus rock, for instance) presents another dimension of study; having recently adapted our code for massively parallel architectures (K. M. Olson and E.A, manuscript in preparation), we are now ready to perform a more comprehensive analysis.

The exploratory simulations presented here suggest that when a young, non-porous asteroid (if such exist) suffers extensive impact damage, the resulting fracture pattern largely defines the asteroid's response to future impacts. The stochastic nature of collisions implies that small asteroid interiors may be as diverse as their shapes and spin states. Detailed numerical simulations of impacts, using accurate shape models and rheologies, could shed light on how asteroid collisional response depends on internal configuration and shape, and hence on how planetesimals evolve. Detailed simulations are also required before one can predict the quantitative effects of nuclear explosions on Earth-crossing comets and asteroids, either for hazard mitigation[28] through disruption and deflection, or for resource exploitation[29]. Such predictions would require detailed reconnaissance concerning the composition and internal structure of the targeted object. □

Received 4 February; accepted 18 March 1998.

1. Asphaug, E. & Melosh, H. J. The Stickney impact of Phobos: A dynamical model. *Icarus* **101**, 144–164 (1993).
2. Asphaug, E. *et al.* Mechanical and geological effects of impact cratering on Ida. *Icarus* **120**, 158–184 (1996).
5. Nolan, M. C., Asphaug, E., Melosh, H. J. & Greenberg, R. Impact craters on asteroids: Does strength or gravity control their size? *Icarus* **124**, 359–371 (1996).
4. Love, S. J. & Ahrens, T. J. Catastrophic impacts on gravity dominated asteroids. *Icarus* **124**, 141–155 (1996).
5. Melosh, H. J. & Ryan, E. V. Asteroids: Shattered but not dispersed. *Icarus* **129**, 562–564 (1997).
6. Housen, K. R., Schmidt, R. M. & Holsapple, K. A. Crater ejecta scaling laws: Fundamental forms based on dimensional analysis. *J. Geophys. Res.* **88**, 2485–2499 (1983).
7. Holsapple, K. A. & Schmidt, R. M. Point source solutions and coupling parameters in cratering mechanics. *J. Geophys. Res.* **92**, 6350–6376 (1987).
8. Housen, K. R. & Holsapple, K. A. On the fragmentation of asteroids and planetary satellites. *Icarus* **84**, 226–253 (1990).
9. Benz, W. & Asphaug, E. Simulations of brittle solids using smooth particle hydrodynamics. *Comput. Phys. Commun.* **87**, 253–265 (1995).
10. Asphaug, E. *et al.* Mechanical and geological effects of impact cratering on Ida. *Icarus* **120**, 158–184 (1996).
11. Hudson, R. S. & Ostro, S. J. Shape of asteroid 4769 Castalia (1989 PB) from inversion of radar images. *Science* **263**, 940–943 (1994).
12. Ostro, S. J. *et al.* Asteroid radar astrometry. *Astron. J.* **102**, 1490–1502 (1991).
13. Ahrens, T. J. & O'Keefe, J. D. In *Impact and Explosion Cratering* (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 639–656 (Pergamon, New York, 1977).
14. Tillotson, J. H. Metallic equations of state for hypervelocity impact. (General Atomic Report GA-3216, San Diego, 1962).
15. Nakamura, A. & Fujiwara, A. Velocity distribution of fragments formed in a simulated collisional disruption. *Icarus* **92**, 132–146 (1991).
16. Benz, W. & Asphaug, E. Simulations of brittle solids using smooth particle hydrodynamics. *Comput. Phys. Commun.* **87**, 253–265 (1995).
17. Bottke, W. F., Nolan, M. C., Greenberg, R. & Kolvoord, R. A. Velocity distributions among colliding asteroids. *Icarus* **107**, 255–268 (1994).
18. Belton, M. J. S. *et al.* Galileo encounter with 951 Gaspra—First pictures of an asteroid. *Science* **257**, 1647–1652 (1992).
19. Belton, M. J. S. *et al.* Galileo's encounter with 243 Ida: An overview of the imaging experiment. *Icarus* **120**, 1–19 (1996).
20. Asphaug, E. & Melosh, H. J. The Stickney impact of Phobos: A dynamical model. *Icarus* **101**, 144–164 (1993).
21. Asphaug, E. *et al.* Mechanical and geological effects of impact cratering on Ida. *Icarus* **120**, 158–184 (1996).
22. Housen, K. R., Schmidt, R. M. & Holsapple, K. A. Crater ejecta scaling laws: Fundamental forms based on dimensional analysis. *J. Geophys. Res.* **88**, 2485–2499 (1983).
23. Veverka, J. *et al.* NEAR's flyby of 253 Mathilde: Images of a C asteroid. *Science* **278**, 2109–2112 (1997).
24. Asphaug, E. *et al.* Impact evolution of icy regoliths. *Lunar Planet. Sci. Conf. (Abstr.)* **XXVIII**, 63–64 (1997).
25. Love, S. G., Hörz, F. & Brownlee, D. E. Target porosity effects in impact cratering and collisional disruption. *Icarus* **105**, 216–224 (1993).
26. Fujiwara, A., Cerroni, P., Davis, D. R., Ryan, E. V. & DiMartino, M. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, A. S.) 240–265 (Univ. Arizona Press, Tucson, 1989).
27. Davis, D. R. & Farinella, P. Collisional evolution of Edgeworth-Kuiper Belt objects. *Icarus* **125**, 50–60 (1997).
28. Ahrens, T. J. & Harris, A. W. Deflection and fragmentation of near-Earth asteroids. *Nature* **360**, 429–433 (1992).
29. *Resources of Near-Earth Space* (eds Lewis, J. S., Matthews, M. S. & Guerrieri, M. L.) (Univ. Arizona Press, Tucson, 1993).

Acknowledgements. This work was supported by NASA's Planetary Geology and Geophysics Program.

Correspondence and requests for materials should be addressed to E.A. (e-mail: asphaug@earthsci.ucsc.edu).

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall, Cornell University, Ithaca, New York 14853, USA

Networks of coupled dynamical systems have been used to model biological oscillators[1–4], Josephson junction arrays[5,6], excitable media[7], neural networks[8–10], spatial games[11], genetic control networks[12] and many other self-organizing systems. Ordinarily, the connection topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes. Here we explore simple models of networks that can be tuned through this middle ground: regular networks 'rewired' to introduce increasing amounts of disorder. We find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. We call them 'small-world' networks, by analogy with the small-world phenomenon[13,14] (popularly known as six degrees of separation[15]). The neural network of the worm *Caenorhabditis elegans*, the power grid of the western United States, and the collaboration graph of film actors are shown to be small-world networks. Models of dynamical systems with small-world coupling display enhanced signal-propagation speed, computational power, and synchronizability. In particular, infectious diseases spread more easily in small-world networks than in regular lattices.

To interpolate between regular and random networks, we consider the following random rewiring procedure (Fig. 1). Starting from a ring lattice with n vertices and k edges per vertex, we rewire each edge at random with probability p. This construction allows us to 'tune' the graph between regularity ($p = 0$) and disorder ($p = 1$), and thereby to probe the intermediate region $0 < p < 1$, about which little is known.

We quantify the structural properties of these graphs by their characteristic path length $L(p)$ and clustering coefficient $C(p)$, as defined in Fig. 2 legend. Here $L(p)$ measures the typical separation between two vertices in the graph (a global property), whereas $C(p)$ measures the cliquishness of a typical neighbourhood (a local property). The networks of interest to us have many vertices with sparse connections, but not so sparse that the graph is in danger of becoming disconnected. Specifically, we require $n \gg k \gg \ln(n) \gg 1$, where $k \gg \ln(n)$ guarantees that a random graph will be connected[16]. In this regime, we find that $L \sim n/2k \gg 1$ and $C \sim 3/4$ as $p \to 0$, while $L \approx L_{\text{random}} \sim \ln(n)/\ln(k)$ and $C \approx C_{\text{random}} \sim k/n \ll 1$ as $p \to 1$. Thus the regular lattice at $p = 0$ is a highly clustered, large world where L grows linearly with n, whereas the random network at $p = 1$ is a poorly clustered, small world where L grows only logarithmically with n. These limiting cases might lead one to suspect that large C is always associated with large L, and small C with small L.

On the contrary, Fig. 2 reveals that there is a broad interval of p over which $L(p)$ is almost as small as L_{random} yet $C(p) \gg C_{\text{random}}$. These small-world networks result from the immediate drop in $L(p)$ caused by the introduction of a few long-range edges. Such 'short cuts' connect vertices that would otherwise be much farther apart than L_{random}. For small p, each short cut has a highly nonlinear effect on L, contracting the distance not just between the pair of vertices that it connects, but between their immediate neighbourhoods, neighbourhoods of neighbourhoods and so on. By contrast, an edge

*Present address: Paul F. Lazarsfeld Center for the Social Sciences, Columbia University, 812 SIPA Building, 420 W118 St, New York, New York 10027, USA.

Nature © Macmillan Publishers Ltd 1998

Travers and Milgram (1977):

296 arbitrary individuals in Nebraska and Boston were asked to give a letter to an acquaintance most likely to help it reach a target person in Massachusetts. 64 reached the target person, average number of intermediaries was 5.2

Collins and Chow (1998)

news and views

It's a small world

James J. Collins and Carson C. Chow

The concept of Six Degrees of Separation has been formalized in so-called 'small-world networks'. The principles involved could be of use in settings as diverse as improving networks of cellular phones and understanding the spread of infections.

A few years ago, on American campuses, it was popular to play Six Degrees of Kevin Bacon. In this game, participants attempt to link the actor Kevin Bacon to any other actor through as few common films and co-stars as possible. Links are formed directly between Bacon and another actor if they appeared in the same film or indirectly through a chain of co-stars in different films (Fig. 1).

In the world of mathematics, a similar amusement involves assessing one's Erdös number, which measures the number of links needed to connect one to the prolific mathematician Paul Erdös through jointly authored papers. For example, individuals have an Erdös number of 1 if they co-authored a paper with Erdös. If one of their co-authors wrote a paper with Erdös, then they have an Erdös number of 2, and so forth. It has been pointed out[1] that Dan Kleitman has a combined Erdös/Bacon number of 3 because he wrote a paper with Erdös and appeared in *Good Will Hunting* with Minnie Driver, who appeared with Bacon in *Sleepers*.

These games are related to the popular concept of Six Degrees of Separation[2], which is based on the notion that everyone in the world is connected to everyone else through a chain of at most six mutual acquaintances. If two people have one mutual acquaintance, then they have one degree of separation. The estimate of six degrees of separation, which is related to the small-world phenomenon[3,4], arises from pioneering empirical work by Milgram[5] and can be understood heuristically from a somewhat unrealistic assumption of random connectivity. That is, if each person knows about one hundred individuals, and given that there are about a billion people on the Earth, then seven connections or six degrees of separation are enough to link everyone together.

On page 440 of this issue[6], Watts and Strogatz formalize this idea in what they call small-world networks. They demonstrate through numerical simulations that a network need not be very random to get this small-world effect. They consider a connected network with nodes and links. In the friendship analogy, each node represents a person and each link represents a single connection to an acquaintance. They then define two measures. The first is a characteristic path length. This is the smallest number of links it takes to connect one node to another, averaged over all pairs of nodes in the network. The second measure is the clustering coefficient. This measures the amount of cliquishness of the network, that is, the fraction of neighbouring nodes that are also connected to one another. For example, in an all-to-all connected network, the clustering coefficient is one.

An example of a large-world network is one that is regularly and locally connected like a crystalline lattice. Such a network is highly clustered and the characteristic path length is large, scaling with the typical linear dimension of the network. On the other hand, a completely random network is poorly clustered and the characteristic path length is short, scaling logarithmically with the size of the network.

What Watts and Strogatz[6] do is to shift gradually from a regular network to a random network by increasing the probability of making random connections from 0 to 1 (see Fig. 1, page 441). They then measure the characteristic path length and the amount of clustering of the network as a function of the amount of randomness. They find that path length and clustering depend differently on the amount of randomness in the network. The characteristic path length drops quickly, whereas the amount of clustering drops rather slowly. This leads to a small-world network in which the amount of clustering is high and the characteristic path length is short. So a small world can exist even when the cliquishness is imperceptibly different from that of a large world.

The explanation for this effect is that it only takes a few short cuts between cliques to turn a large world into a small world. In the friendship analogy, it only takes a small number of well-connected people to make a world small. The interesting and surprising thing is that it is impossible to determine whether or not you live in a small world or a large world from local information alone. The average person (node) is not directly associated with the key people (the clique-linkers).

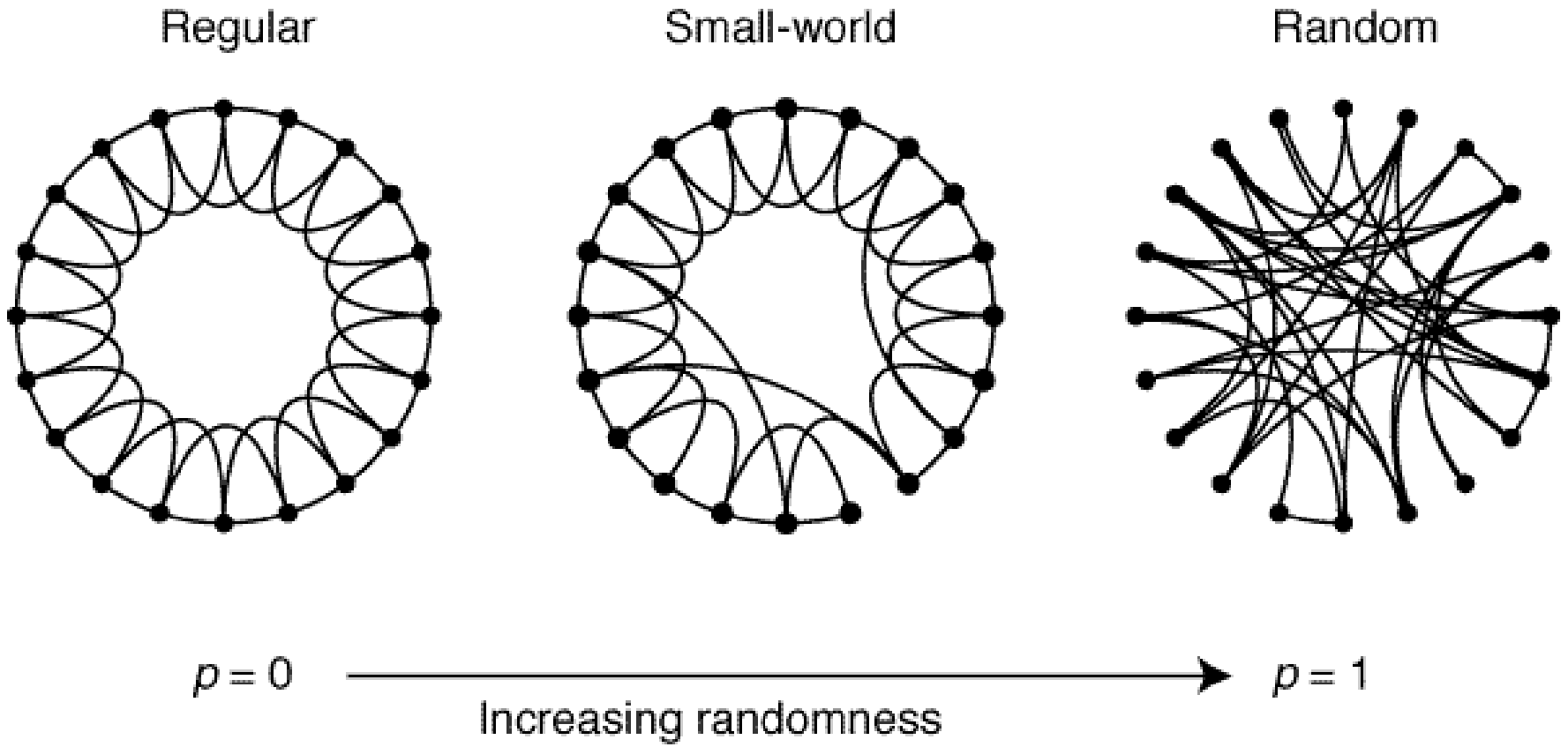
Small-world connectivity has consequences that could be good or bad,

APTHE ROYAL COLLECTION

Figure 1 **Three degrees. Because Kevin Bacon has appeared in many films, most actors have low Bacon numbers and the game Six Degrees of Kevin Bacon has declined in popularity. It is possible to centre the game around a newer star such as Leonardo DiCaprio. These film stills, running clockwise, show that in this case there are at most three degrees of separation between DiCaprio and Helena Bonham-Carter, through Kate Winslet (*Titanic*, Columbia TriStar; *Sense and Sensibility*, Columbia TriStar), Emma Thompson (*Sense and Sensibility*; *Much Ado About Nothing*, Entertainment Films) and Kenneth Branagh (*Much Ado About Nothing*; *Frankenstein*; Columbia TriStar). Short cuts between cliques could be created in this game through some of DiCaprio's well-connected co-stars such as Sharon Stone (*The Quick and the Dead*; TriStar; not shown).**

Nature © Macmillan Publishers Ltd 1998

Complex is Not the Same as Random



Watts and Strogatz (1998)

Knowledge of Exposure Pathways Limits High Throughput Exposure Models

"In particular, the assumption that 100% of [quantity emitted, applied, or ingested] is being applied to each individual use scenario is a very conservative assumption for many compound / use scenario pairs."

This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

Environmental Science & Technology

Article

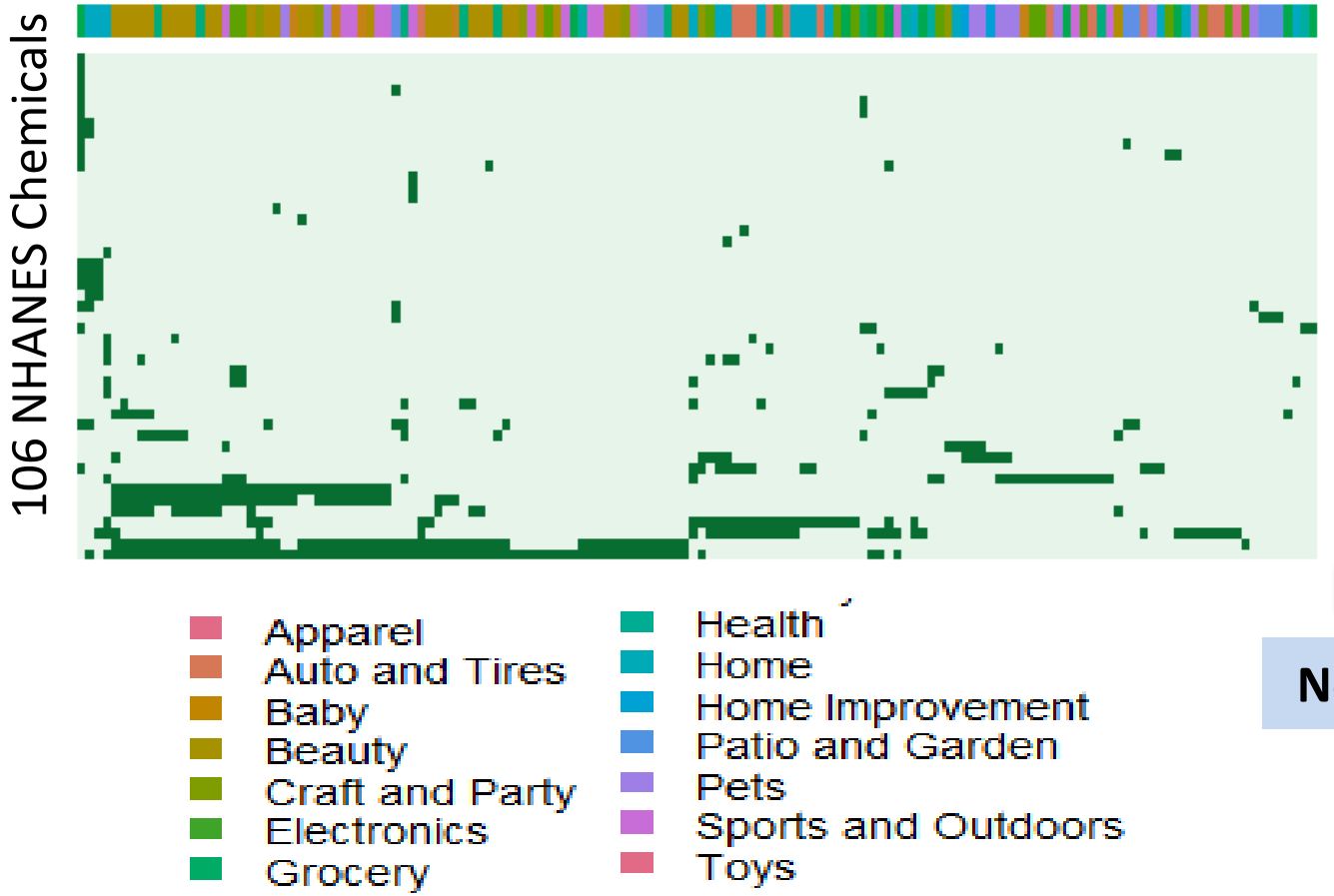
pubs.acs.org/est

Risk-Based High-Throughput Chemical Screening and Prioritization using Exposure Models and in Vitro Bioactivity Assays

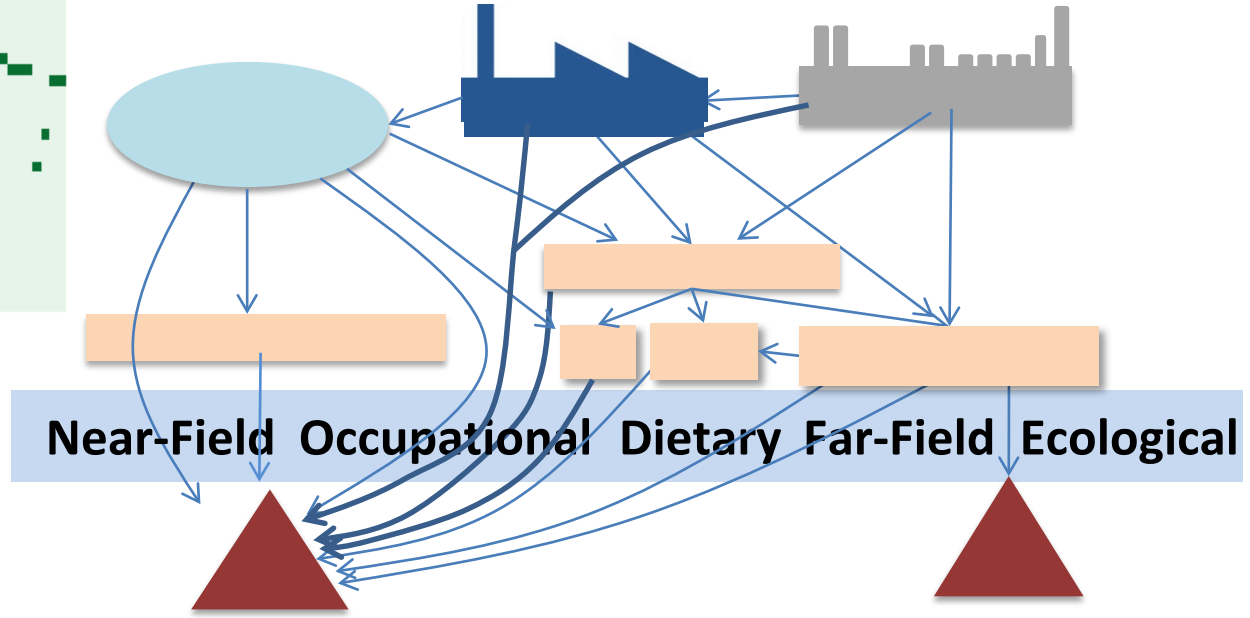
Hyeong-Moo Shin,[*,†] Alexi Ernstoff,[‡,§] Jon A. Arnot,[∥,⊥,#] Barbara A. Wetmore,[▽] Susan A. Csiszar,[§] Peter Fantke,[‡] Xianming Zhang,[○] Thomas E. McKone,[◆,¶] Olivier Jolliet,[§] and Deborah H. Bennett[†]

Chemical Use Identifies Relevant Pathways

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



Some pathways have much higher average exposures!



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

How Can we Know Chemical Use? Chemical Property NAMs

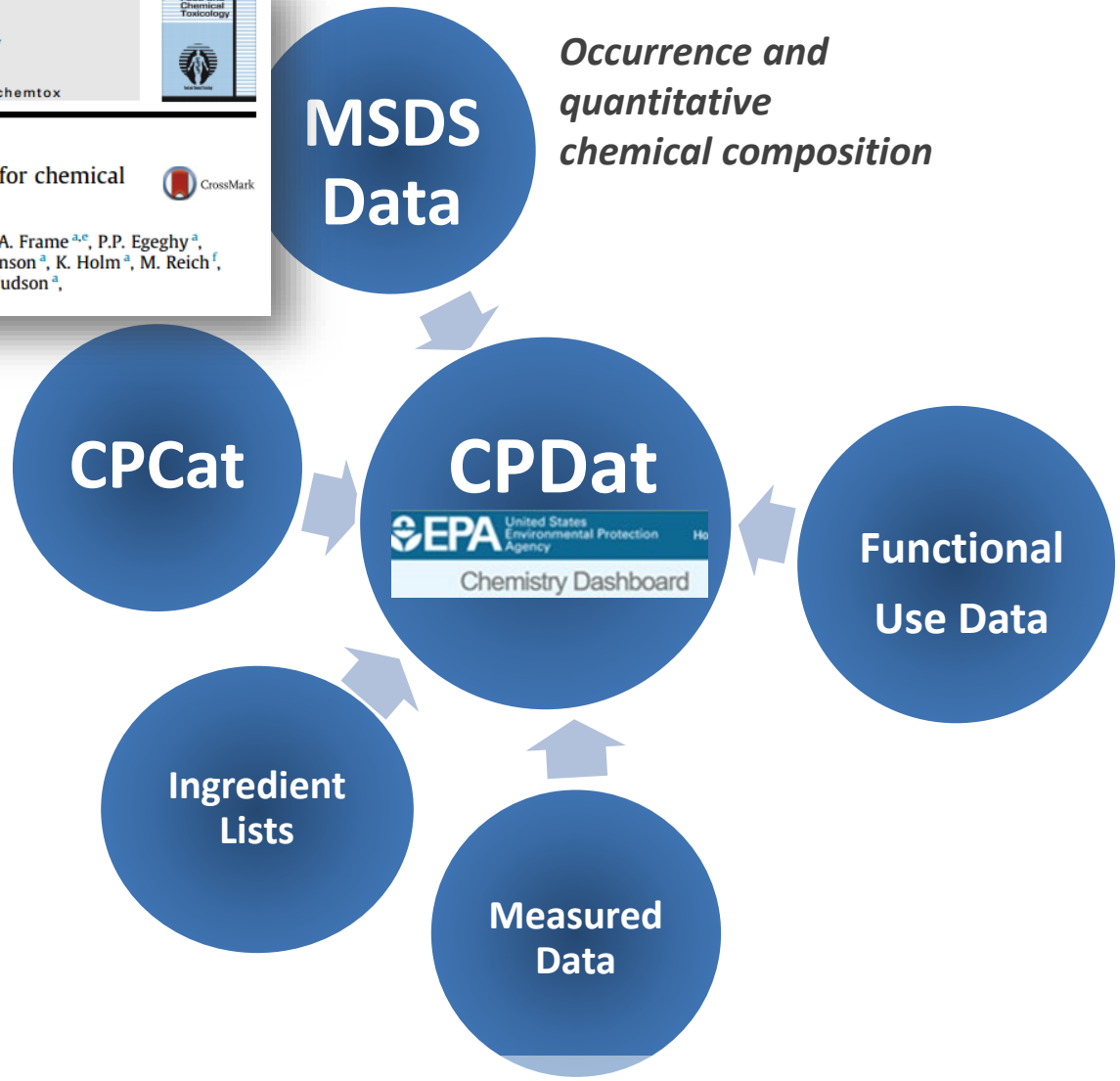
Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

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

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



Slide from Kristin Isaacs

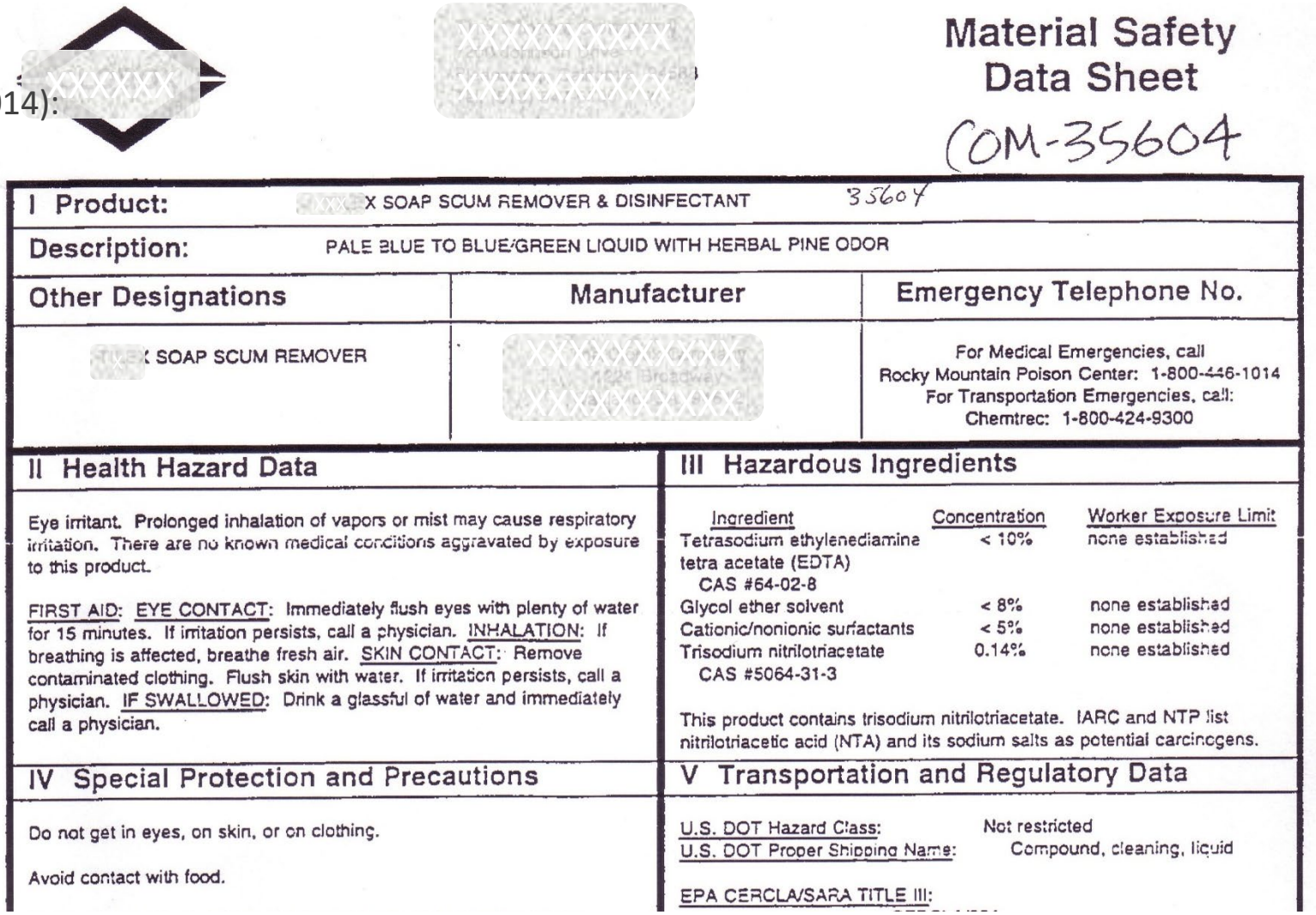
https://comptox.epa.gov/dashboard

CPCPdb: Material Safety Data Sheets

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.
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:

How Can we Know Chemical Use? Chemical Property NAMs

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

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

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

Broad "index" of chemical uses

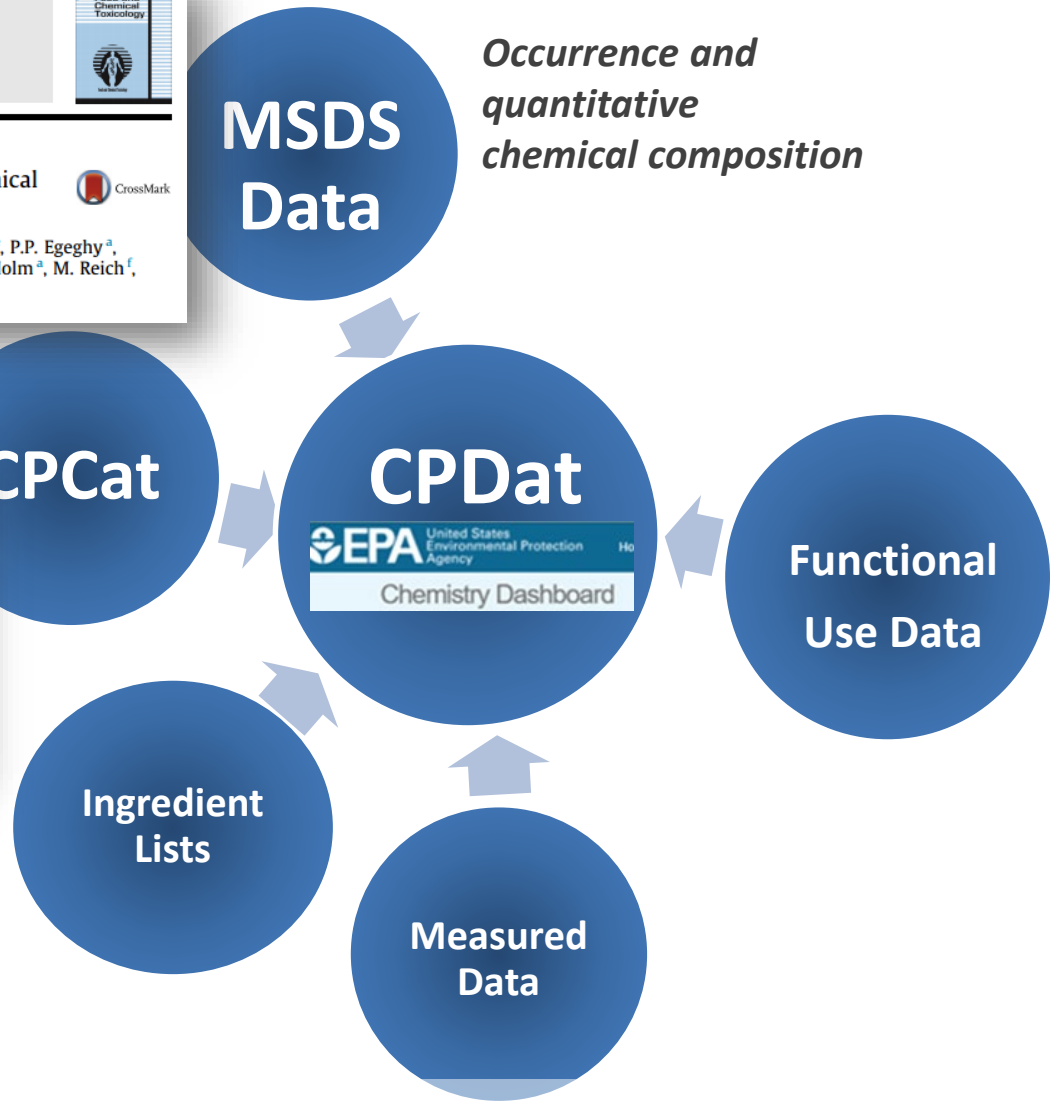
Contents lists available at ScienceDirect

Toxicology Reports

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

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

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



https://comptox.epa.gov/dashboard

Slide from Kristin Isaacs

How Can we Know Chemical Use? Chemical Property NAMs

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

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

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

MSDS Data

Occurrence and quantitative chemical composition

Broad "index" of chemical uses

Contents lists available at ScienceDirect

Toxicology Reports

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

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

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

CPCat

CPDat

EPA United States Environmental Protection Agency

Chemistry Dashboard

Functional Use Data

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

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

www.nature.com/jes

ORIGINAL ARTICLE

Consumer product chemical weight fractions from ingredient lists

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

Ingredient Lists

Measured Data

Occurrence data

https://comptox.epa.gov/dashboard

Slide from Kristin Isaacs

How Can we Know Chemical Use? Chemical Property NAMs

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

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

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

MSDS Data

Occurrence and quantitative chemical composition

Broad "index" of chemical uses

Contents lists available at ScienceDirect

Toxicology Reports

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

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

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

CPCat

CPDat

Chemistry Dashboard

Functional Use Data

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

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

www.nature.com/jes

ORIGINAL ARTICLE

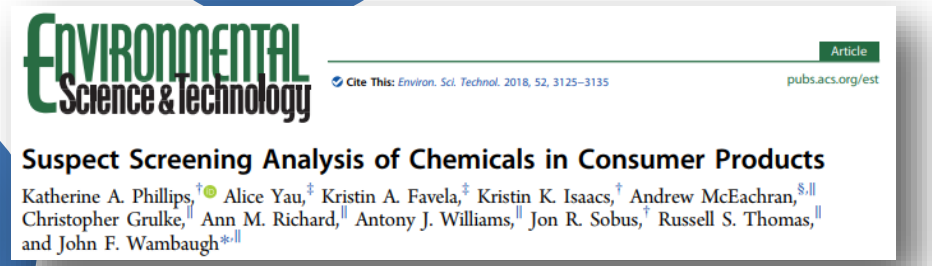
Consumer product chemical weight fractions from ingredient lists

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

Ingredient Lists

Occurrence data

Measured Data


Environmental Science & Technology

Article

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

pubs.acs.org/est

Suspect Screening Analysis of Chemicals in Consumer Products

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

Measurement of chemicals in consumer products

https://comptox.epa.gov/dashboard

Slide from Kristin Isaacs

How Can we Know Chemical Use? Chemical Property NAMs

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

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

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

MSDS Data

Occurrence and quantitative chemical composition

Broad "index" of chemical uses

Contents lists available at ScienceDirect

Toxicology Reports

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

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

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

CPCat

CPDat

EPA United States Environmental Protection Agency

Chemistry Dashboard

Green Chemistry

PAPER

View Article Online

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

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

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

Functional Use Data

The roles that chemicals serve in products

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

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

www.nature.com/jes

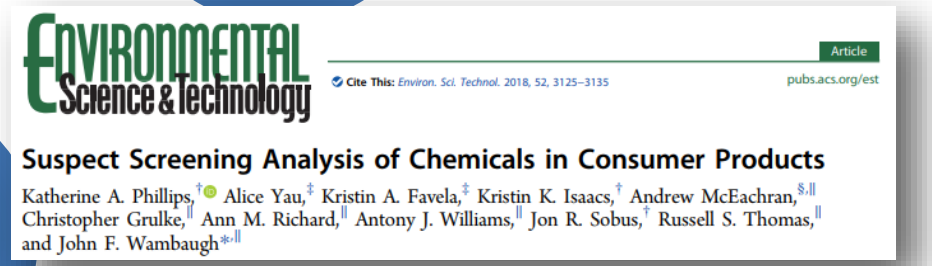
ORIGINAL ARTICLE

Consumer product chemical weight fractions from ingredient lists

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

Ingredient Lists

Occurrence data


Environmental Science & Technology

Article

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

pubs.acs.org/est

Suspect Screening Analysis of Chemicals in Consumer Products

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

Measured Data

Measurement of chemicals in consumer products

https://comptox.epa.gov/dashboard

Slide from Kristin Isaacs

How Can we Know C… Chemical Proper…

Contents lists available at ScienceDirect

Food and Chemical Toxicology

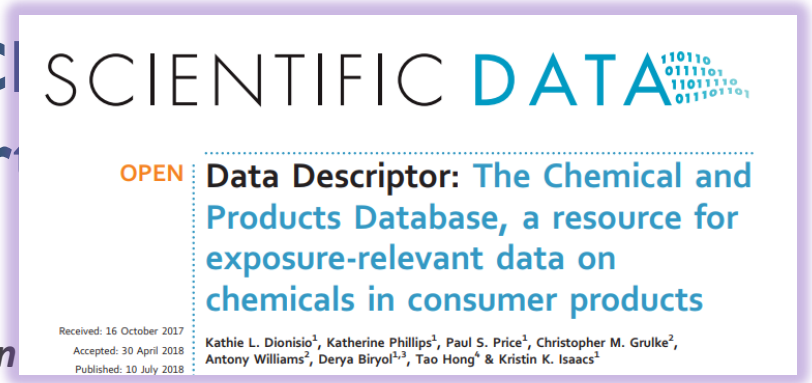
journal homepage: www.elsevier.com/locate/foodchemtox

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

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

MSDS Data

Occurrence and quantitative chemical composition


SCIENTIFIC DATA

OPEN

Data Descriptor: The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products

Received: 16 October 2017
Accepted: 30 April 2018
Published: 10 July 2018

Kathie L. Dionisio[1], Katherine Phillips[1], Paul S. Price[1], Christopher M. Grulke[2], Antony Williams[2], Derya Biryol[1,3], Tao Hong[4] & Kristin K. Isaacs[1]

Broad "index" of chemical uses

Contents lists available at ScienceDirect

Toxicology Reports

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

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

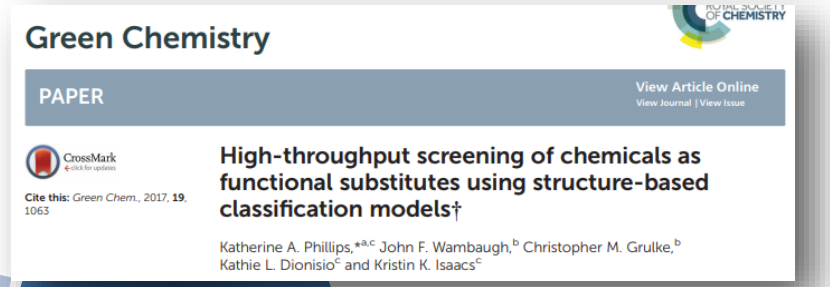
Kathie L. Dionisio[a], Alicia M. Frame[b,1], Michael-Rock Goldsmith[a,2], John F. Wambaugh[b], Alan Liddell[c,3], Tommy Cathey[d], Doris Smith[b], James Vail[b], Alexi S. Ernstoff[e], Peter Fantke[e], Olivier Jolliet[f],

CPCat

CPDat

EPA United States Environmental Protection Agency

Chemistry Dashboard


Green Chemistry

PAPER

View Article Online

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

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

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

Functional Use Data

The roles that chemicals serve in products

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

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

www.nature.com/jes

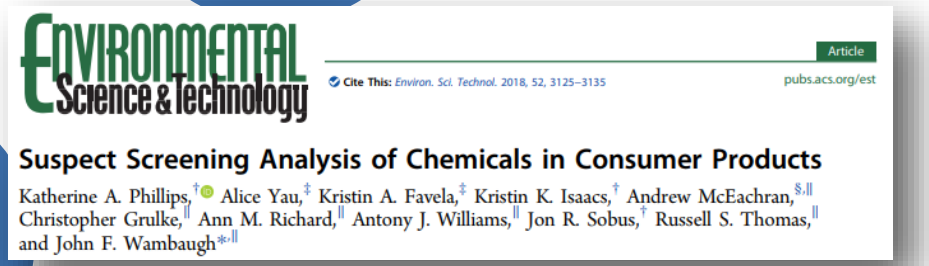
ORIGINAL ARTICLE

Consumer product chemical weight fractions from ingredient lists

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

Ingredient Lists

Occurrence data


Environmental Science & Technology

Article

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

pubs.acs.org/est

Suspect Screening Analysis of Chemicals in Consumer Products

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

Measured Data

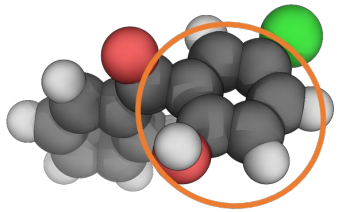
Measurement of chemicals in consumer products

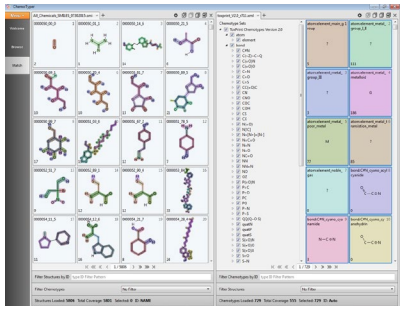
https://comptox.epa.gov/dashboard

Slide from Kristin Isaacs

Exposure NAM: Machine Learning to Fill Data Gaps
EXAMPLE: Predicting Function Based on Structure

Use Database (FUSE)





Chemical Structure and Property Descriptors

Prediction of Of Potential Alternatives from Chemical Libraries

Machine Learning Based Classification Models
(Random Forest, Breiman, 2001)

Phillips et al. (2017)

What is "High Throughput"?

- Tox21: Testing one assay across 10,000 chemicals takes 1-2 days, but only 50 assays have been developed so far that can run that fast

- ToxCast: ~1100 off-the-shelf (pharma) assay-endpoints tested for up to 4,000 chemicals over the past decade, now developing new assays as well

HTS tox assays often use single readout, such as fluorescence, across many chemicals, measuring concentration for toxicokinetics or exposure requires chemical-specific methods…

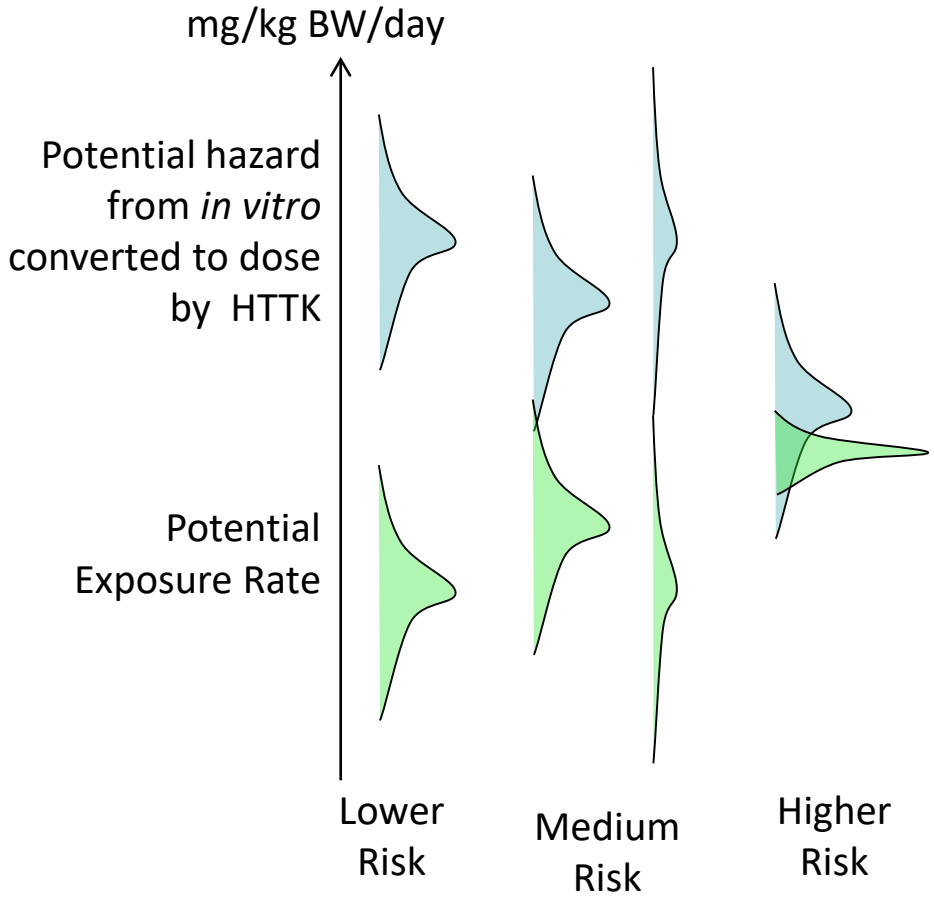
- ExpoCast: Ring et al. made *in silico* predictions for ~480,000 chemicals from structure, but based on NHANES monitoring for ~120 chemicals
  - Quantitative non-targeted analysis (NTA) may eventually provide greater evaluation data to reduce uncertainty

- HTTK: *In vitro* data on 944 chemicals collected for humans, starting with Rotroff et al. (2010)
  - Work continues to develop *in silico* tools, e.g. Sipes et al. (2016)

Our work is not done…

Summary

- A tapestry of laws covers the chemicals people are exposed to in the United States (Breyer, 2009)
- Many chemicals, ranging from industrial waste to dyes to packing materials, are covered by the recently updated Toxic Substances Control Act (TSCA) and administered by the EPA
- New approach methodologies (NAMs) are being developed to prioritize these existing and new chemicals for testing
- All data are being made public:
  - The CompTox Chemicals Dashboard (A search engine for chemicals) http://comptox.epa.gov/
  - R package "httk": https://CRAN.R-project.org/package=httk



The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA

ExpoCast Project (Exposure Forecasting)

Center for Computational Toxicology and Exposure

Linda Adams
Miyuki Breen*
Alex Chao*
Daniel Dawson*
Mike Devito
Kathie Dionisio
Christopher Eklund
Peter Egeghy
Marina Evans
Chris Grulke
Hongtai Huang*
Mike Hughes
Kristin Isaacs
Ashley Jackson*
Richard Judson
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Katherine Phillips
Ann Richard
Risa Sayre*
Mark Sfeir*
Jane Ellen Simmons
Marci Smeltz*
Jon Sobus
Mike Tornero-Velez
Rusty Thomas
Elin Ulrich
Dan Vallero
Barbara Wetmore
John Wambaugh
Antony Williams

Center for Environmental Measurement and Modeling

Hongwan Li
Xiaoyu Liu
Seth Newton
John Streicher*
Mark Strynar

SwRI Southwest Research Institute

cyprotex An Evotec Company

*Trainees

Collaborators

Arnot Research and Consulting
Jon Arnot
Johnny Westgate
Integrated Laboratory Systems
Kamel Mansouri
Xiaoqing Chang
National Toxicology Program
Steve Ferguson
Nisha Sipes
Ramboll
Harvey Clewell
Silent Spring Institute
Robin Dodson
Simulations Plus
Michael Lawless
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Aylward
Technical University of Denmark
Peter Fantke
ToxStrategies
Caroline Ring
Unilever
Beate Nicol
Cecilie Rendal
Ian Sorrell
United States Air Force
Heather Pangburn
Matt Linakis
University of California, Davis
Deborah Bennett
University of Michigan
Olivier Jolliet
University of Texas, Arlington
Hyeong-Moo Shin

United States Environmental Protection Agency

References

Arnot, Jon A., et al. "Screening level risk assessment model for chemical fate and effects in the environment." Environmental science & technology 40.7 (2006): 2316-2323.

Aylward, Lesa L., and Sean M. Hays. "Consideration of dosimetry in evaluation of ToxCast™ data." Journal of Applied Toxicology 31.8 (2011): 741-751.

Breyer, Stephen. Breaking the vicious circle: Toward effective risk regulation. Harvard University Press, 2009

Cohen Hubal, EA, et al. "Advancing internal exposure and physiologically-based toxicokinetic modeling for 21st-century risk assessments." Journal of exposure science & environmental epidemiology (2018).

Collins FS, Gray GM, Bucher JR. Transforming environmental health protection. Science. 2008;319:906–907.

Dionisio, Kathie L., et al. "Exploring consumer exposure pathways and patterns of use for chemicals in the environment." Toxicology reports 2 (2015): 228-237.

Dionisio, Kathie L., et al. "The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products." Scientific data 5 (2018): 180125.

Dix David, et al. "The ToxCast program for prioritizing toxicity testing of environmental chemicals." Toxicol Sci. 2007;95:5–12

Egeghy, P. P., et al. (2012). The exposure data landscape for manufactured chemicals. Science of the Total Environment, 414, 159-166.

Filer, Dayne L., et al. "tcpl: the ToxCast pipeline for high-throughput screening data." Bioinformatics 33.4 (2016): 618-620.

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.

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.

Isaacs, Kristin K., et al. "Consumer product chemical weight fractions from ingredient lists." Journal of Exposure Science and Environmental Epidemiology 28.3 (2018): 216.

Jamei, et al. "The Simcyp® population-based ADME simulator." Expert opinion on drug metabolism & toxicology 2009b;5:211-223

Judson, Richard, et al. "The toxicity data landscape for environmental chemicals." Environmental health perspectives 117.5 (2008): 685-695.

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), ORSFI1-ORSFI11.

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.

Kavlock, R. J., et al. (2018). Accelerating the pace of chemical risk assessment. Chemical research in toxicology, 31(5), 287-290

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

Mansouri, Kamel, et al. "OPERA models for predicting physicochemical properties and environmental fate endpoints." Journal of cheminformatics 10.1 (2018): 10.

McNally, et al., "PopGen: a virtual human population generator." Toxicology 2014

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.

National Research Council. Exposure Science in the 21st Century: a Vision and a Strategy. National Academies Press, 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)

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 19.4 (2017): 1063-1074.

Phillips, Katherine A., et al. "Suspect screening analysis of chemicals in consumer products." Environmental science & technology 52.5 (2018): 3125-3135.

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

Ring, Caroline L., et al. "Consensus modeling of median chemical intake for the US population based on predictions of exposure pathways." Environmental science & technology 53.2 (2018): 719-732.

Rotroff, Daniel M., et al. "Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening." Toxicological Sciences 117.2 (2010): 348-358

Schmidt, Charles W. "TOX 21: new dimensions of toxicity testing." Environmental health perspectives 117.8 (2009): A348.

Shibata, Yoshihiro, et al. "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." Drug Metabolism and disposition 30.8 (2002): 892-896.

Shin, Hyeong-Moo, et al. "Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays." Environmental science & technology 49.11 (2015): 6760-6771.

Sipes, Nisha S., et al. "An intuitive approach for predicting potential human health risk with the Tox21 10k library." Environmental science & technology 51.18 (2017): 10786-10796.

US Congress. "Frank R. Lautenberg Chemical Safety for the 21st Century Act." (2016).

U.S. E.P.A. (2018) "A Working Approach for Identifying Potential Candidate Chemicals for Prioritization." https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/identifying-existing-chemicals-prioritization-under-tsca

U.S. G.A.O.. "Toxic substances: EPA has increased efforts to assess and control chemicals but could strengthen its approach." (2013).

Wallace, Lance A., 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 43.2 (1987): 290-307.

Wambaugh, John F., et al. "High-throughput models for exposure-based chemical prioritization in the ExpoCast project." Environmental science & technology 47.15 (2013): 8479-848.

Wambaugh, John F., et al. "High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals." Environmental science & technology (2014).

Wambaugh, John F., et al. "Toxicokinetic triage for environmental chemicals." Toxicological Sciences 147.1 (2015): 55-67.

Wambaugh, John F., et al. "Evaluating in vitro-in vivo extrapolation of toxicokinetics." Toxicological Sciences 163.1 (2018): 152-169.

Wambaugh, John F., et al. "Assessing Toxicokinetic Uncertainty and Variability in Risk Prioritization" Toxicological Sciences (2019), *in press*

Wambaugh, John F., et al. "New Approach Methodologies for Exposure Science." Current Opinion in Toxicology (2019).

Wang, Ying-Hong. "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 (2010): 1094-1104.

Waters, Nigel J., et al. "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." Journal of pharmaceutical sciences 97.10 (2008): 4586-4595.

Wetmore, Barbara A., et al. "Integration of dosimetry, exposure and high-throughput screening data in chemical toxicity assessment." *Tox. Sciences* (2012)

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.