March 8, 2021

United States
Environmental Protection
Agency

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

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

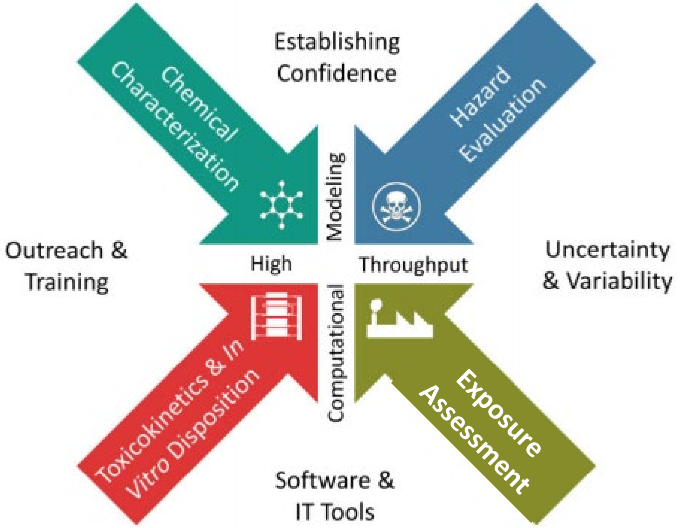
US EPA's ExpoCast Project:

New Approach Methodologies for **Exposure Forecasting**

*"**Investment in 21st century exposure science is now required** to fully realize the potential of the NRC vision for toxicity testing."*
Cohen Hubal (2009)

Lovell and Hegstad (2009): "Obama's FY10 Budget Includes Increased Toxicology":

- Funding allows for complementary exposure predictions from ExpoCast, which is slated to be **launched in FY10**
- Predict the impact of chemicals on the human body using data from ToxCast



Thomas et al. (2019)

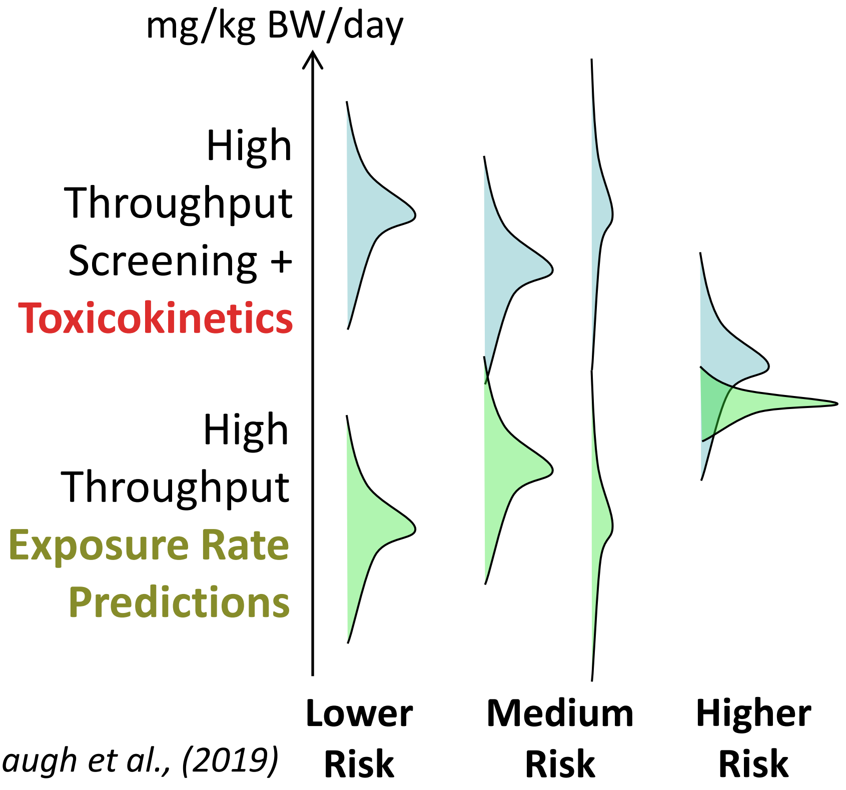
ExpoCast is

Models | Machine Learning | Measurements | Databases | Applied Statistics

Consumer | Ambient | Occupational | Ecological

Since 2010:

- 45 peer-reviewed publications
- 5 STAR grants awarded
- 3 Federal research contracts (SWRI and Battelle)

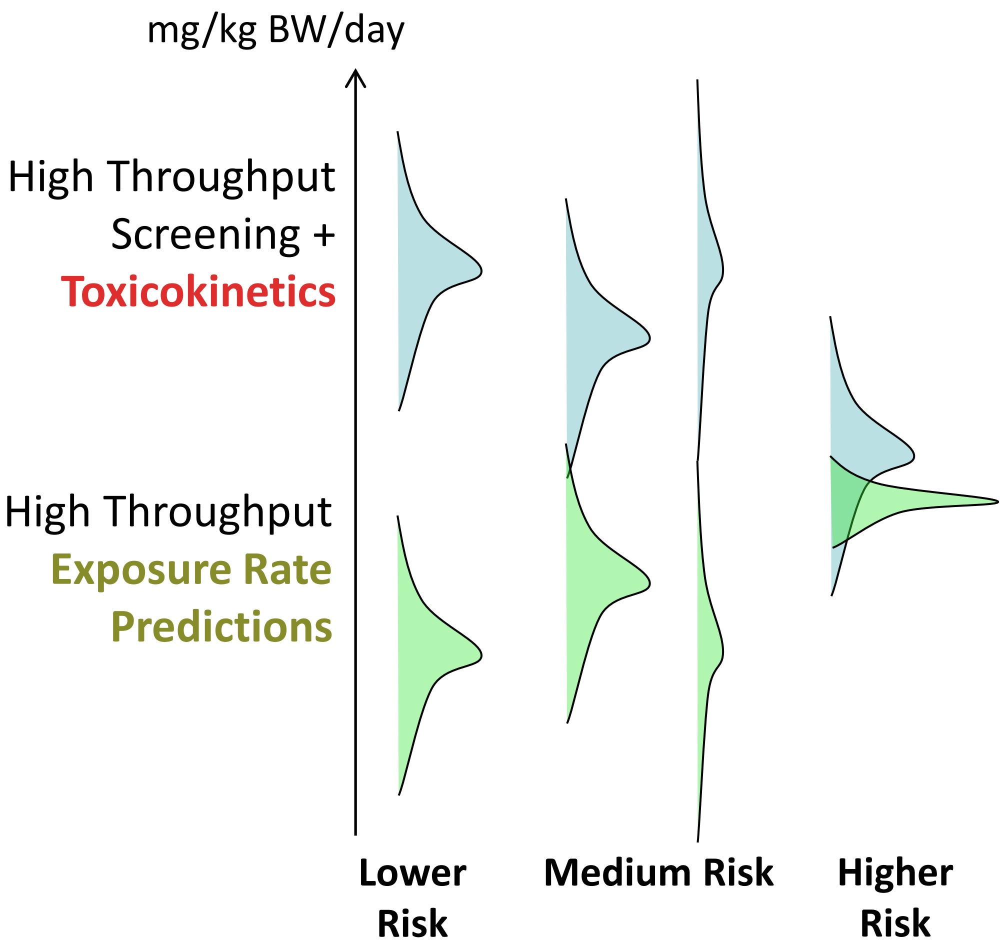


Wambaugh et al., (2019)

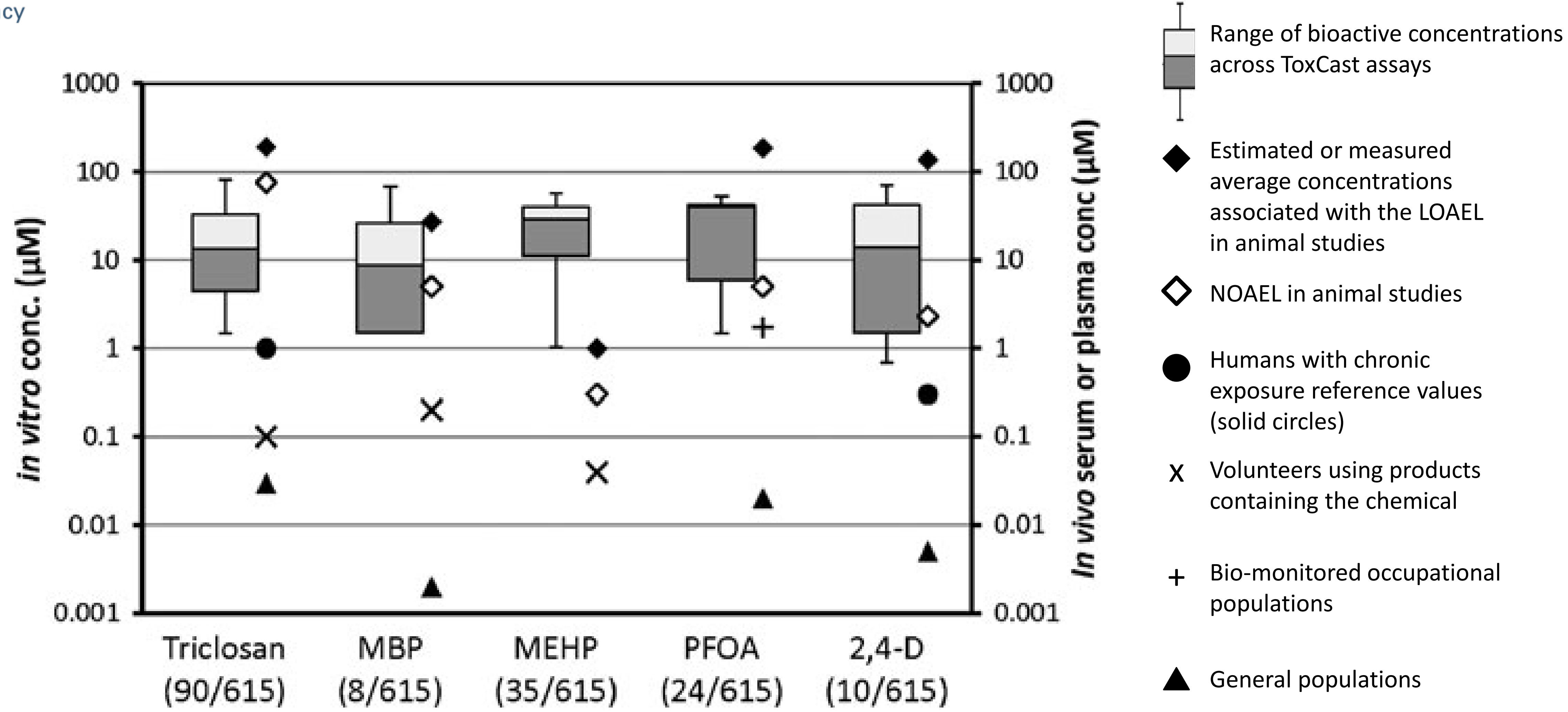
Chemical Risk = Hazard x Exposure

- 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 (that is, dose-response relationship) linking hazard and exposure (Wetmore et al., 2012, 2015)



The Margin Between Exposure and Hazard



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)

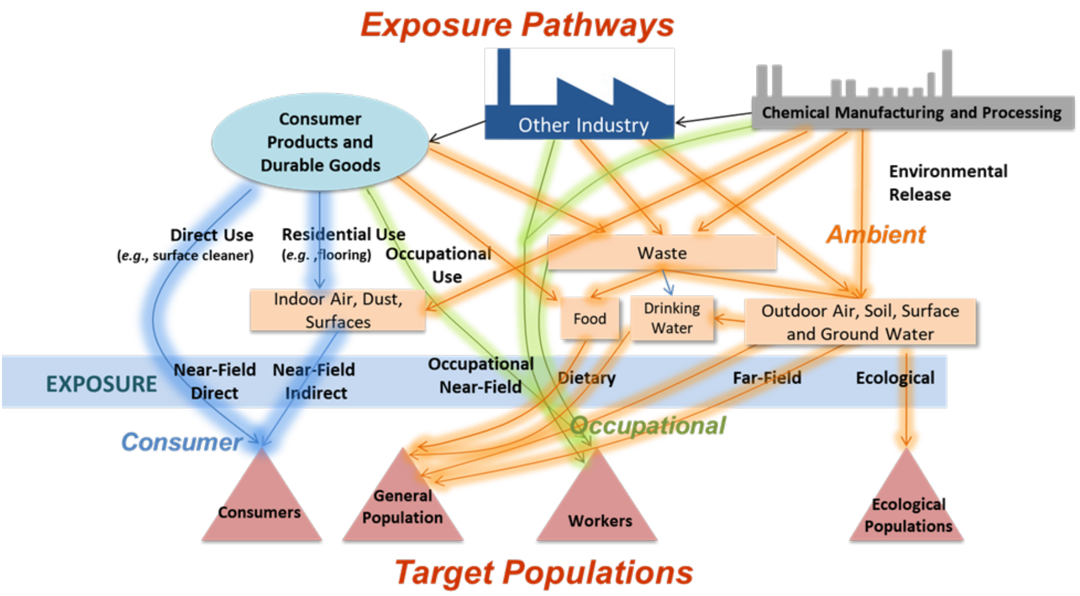
United States
Environmental Protection
Agency

NAMs for Exposure Science

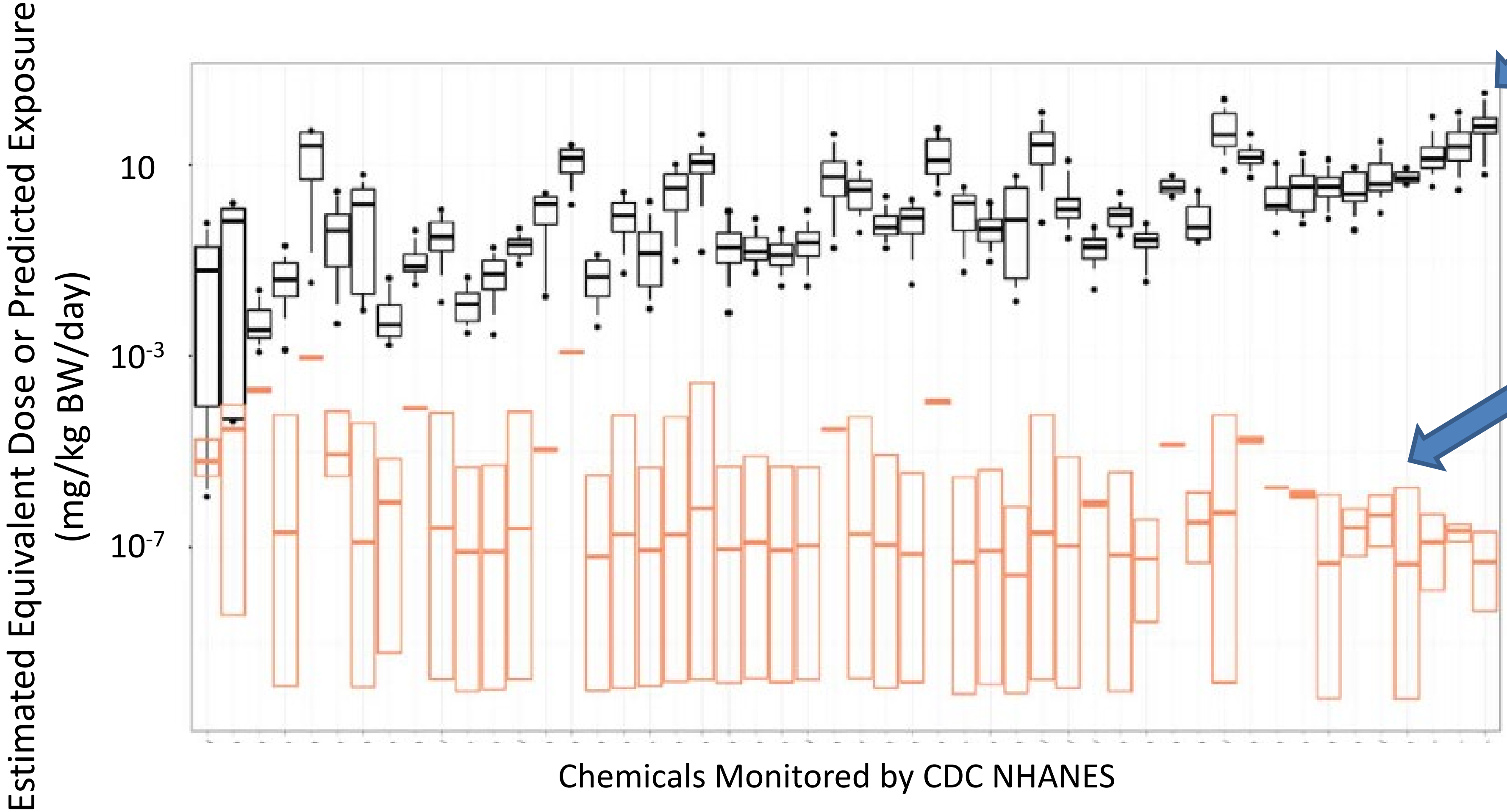
- There are at least 10,000 chemicals produced, used in commerce, and potentially present in the environment
    - Traditional methods are too resource-intensive to address all of these
    - New Approach Methodologies (NAMs) have the potential to address these gaps

- The tools to characterize both toxicity and exposure have evolved significantly in the past decade

- NAMs for exposure science are being developed to enable risk assessors to more rapidly address public health challenges and chemical regulation

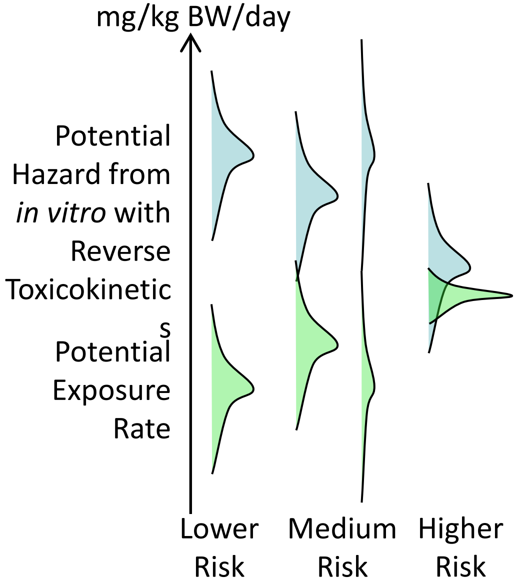


Chemical Prioritization NAMs



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

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



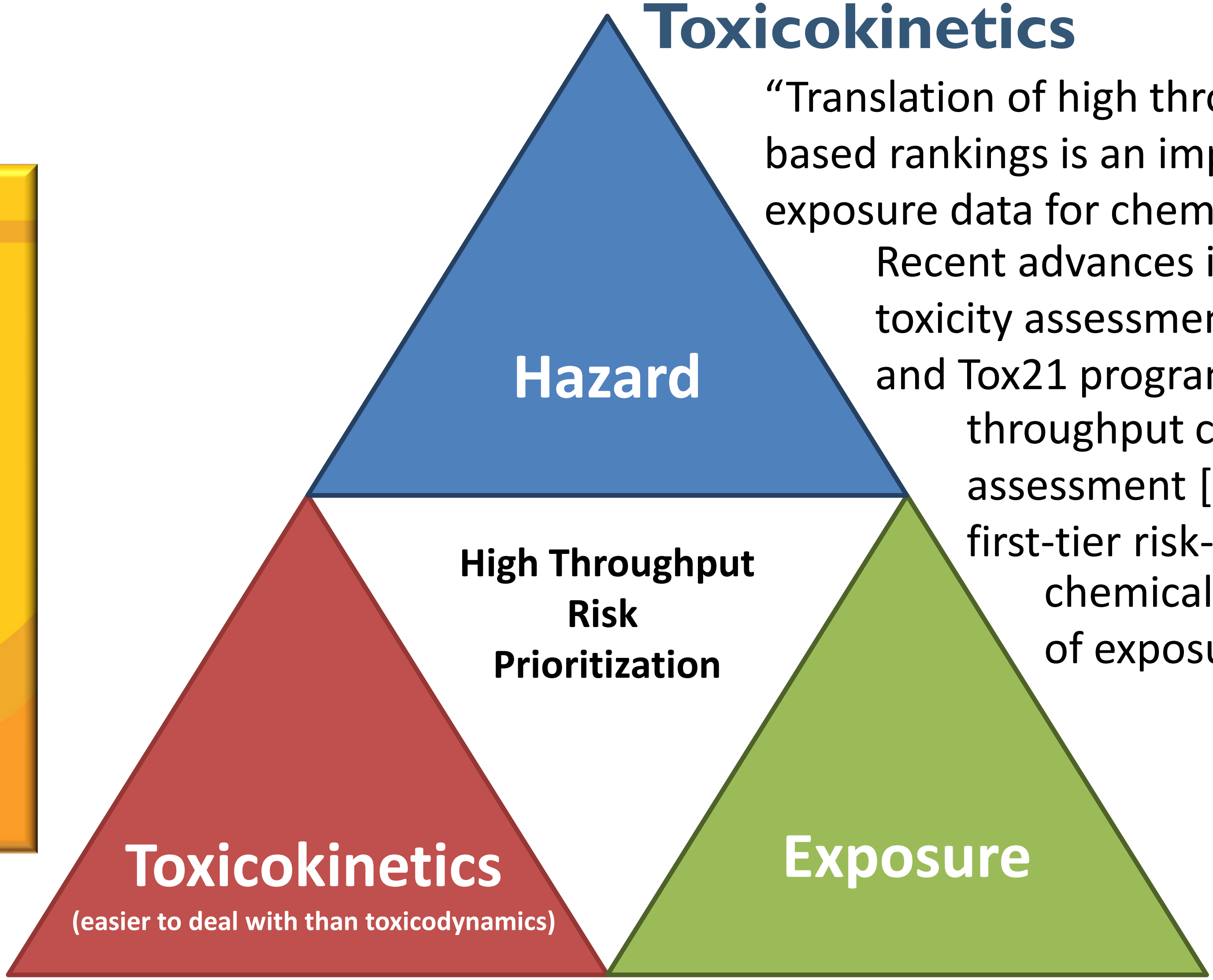
Ring *et al*. (2017)

Most Chemicals Lack Data on Exposure and Toxicokinetics

The National Academies of
SCIENCES · ENGINEERING · MEDICINE
REPORT

USING
21ST CENTURY
SCIENCE
TO IMPROVE
RISK-RELATED
EVALUATIONS

NASEM (2017)



Hazard

High Throughput Risk Prioritization

Toxicokinetics
(easier to deal with than toxicodynamics)

Exposure

"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...

NAMs for Exposure Science

			Makes Use of					
Exposure NAM Class	**Description**	**Traditional Approach**	**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)

NAMs for Exposure Science

			Makes Use of					
Exposure NAM Class	**Description**	**Traditional Approach**	**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)

HTTK: A NAM for Exposure

- To provide toxicokinetic data for larger numbers of chemicals collect *in vitro*, high throughput toxicokinetic (HTTK) data (for example, Rotroff et al., 2010, Wetmore et al., 2012, 2015)
- HTTK methods have been used by the pharmaceutical industry to determine range of efficacious doses and to prospectively evaluate success of planned clinical trials (Jamei, *et al.*, 2009; Wang, 2010)
- The **primary goal** of HTTK is to provide a human dose context for bioactive *in vitro* concentrations from HTS (that is*, in vitro-in vivo* extrapolation, or **IVIVE**) (for example, Wetmore et al., 2015)
- A **secondary goal** is to provide **open-source data and models** for evaluation and use by the broader scientific community (Pearce et al, 2017a)

United States
Environmental Protection
Agency

Lex Parsimoniae
"Law of Parsimony"

"Among competing hypotheses, the one with the fewest assumptions should be selected." William of Occam

"While Occam's razor is a useful tool in the physical sciences, it can be a very dangerous implement in biology. It is thus very rash to use simplicity and elegance as a guide in biological research. "
Francis Crick

"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."
John von Neumann

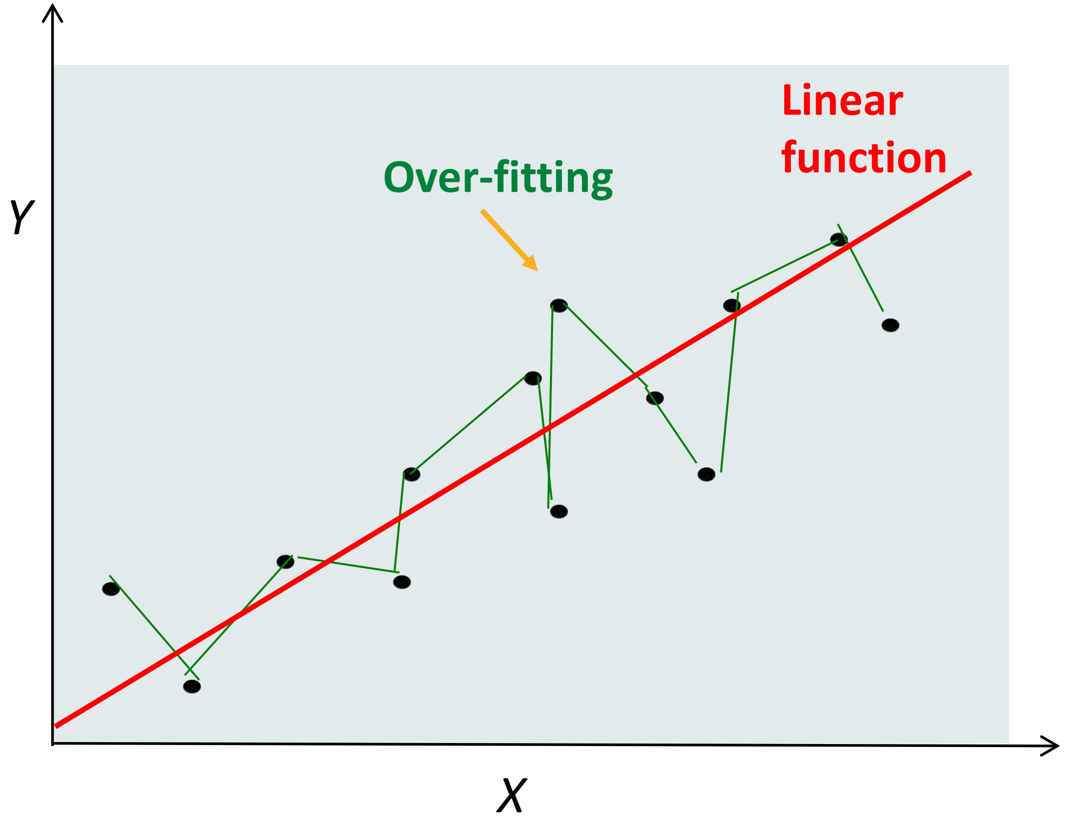
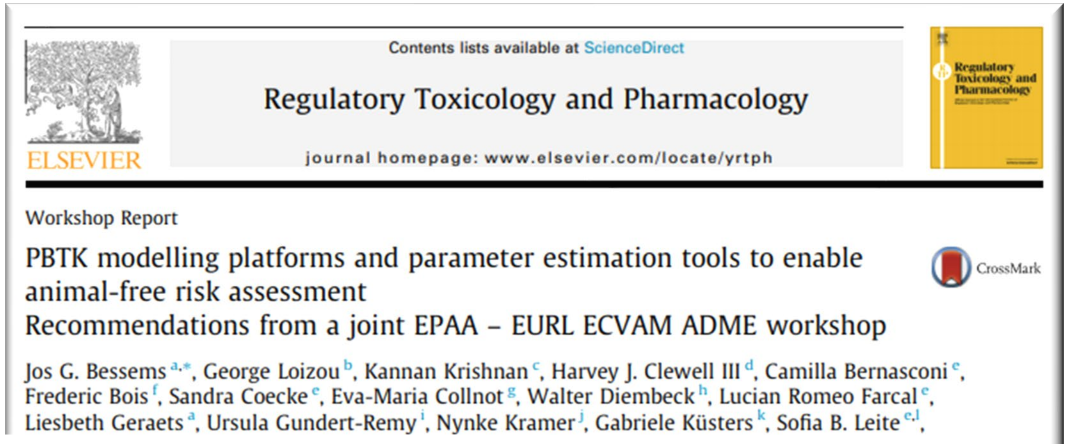


Figure from Anran Wang

Fit for Purpose IVIVE

- We choose to make the complexity of the model and the number of physiological processes appropriate to decision context
- Bessems et al. (2014): We need “a first, relatively quick (‘Tier 1’), estimate” of concentration vs. time in blood, plasma, or cell
- They suggested that we neglect active metabolism – thanks to *in vitro* measurements we can now do better
- We do neglect transport and other protein-specific phenomena



Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

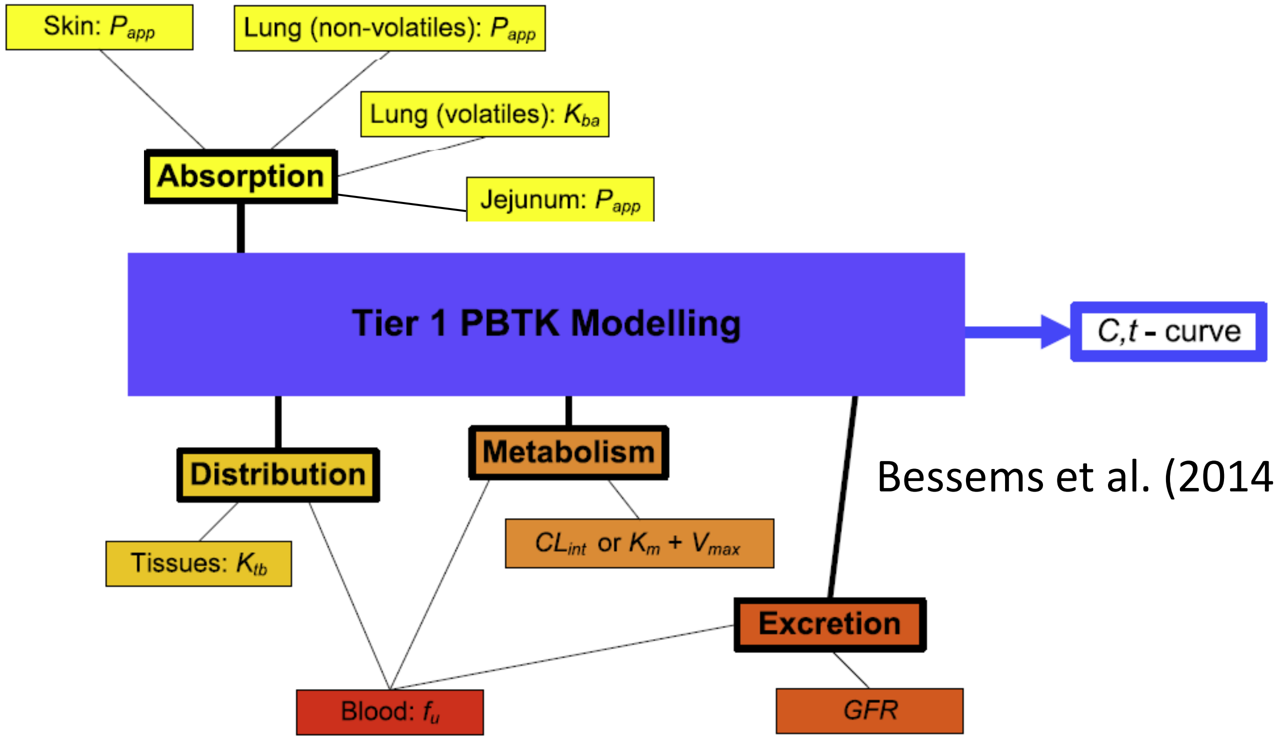
ELSEVIER

Workshop Report

PBTK modelling platforms and parameter estimation tools to enable animal-free risk assessment

Recommendations from a joint EPAA – EURL ECVAM ADME workshop

Jos G. Bessems [a,*], George Loizou [b], Kannan Krishnan [c], Harvey J. Clewell III [d], Camilla Bernasconi [e], Frederic Bois [f], Sandra Coecke [e], Eva-Maria Collnot [g], Walter Diembeck [h], Lucian Romeo Farcal [e], Liesbeth Geraets [a], Ursula Gundert-Remy [i], Nynke Kramer [j], Gabriele Küsters [k], Sofia B. Leite [e,l],



Skin: P_{app}

Lung (non-volatiles): P_{app}

Lung (volatiles): K_{ba}

Absorption

Jejunum: P_{app}

Tier 1 PBTK Modelling

C,t - curve

Distribution

Metabolism

Tissues: K_{tb}

CL_{int} or $K_m + V_{max}$

Excretion

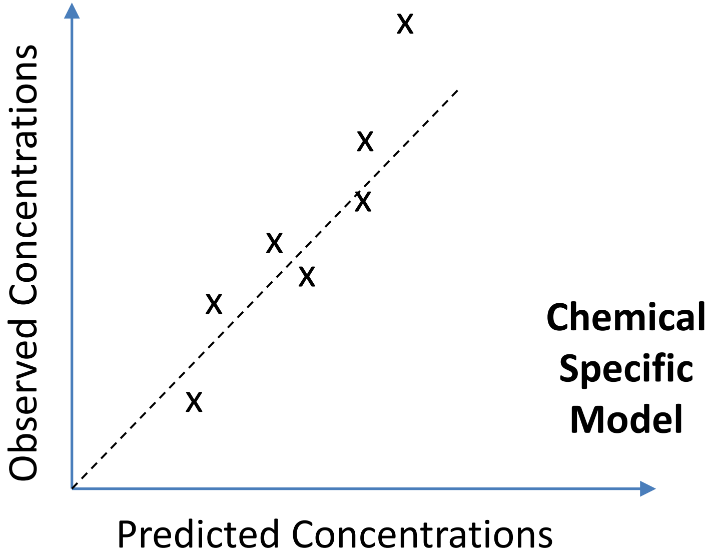
Blood: f_u

GFR

Bessems et al. (2014)

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 have no data



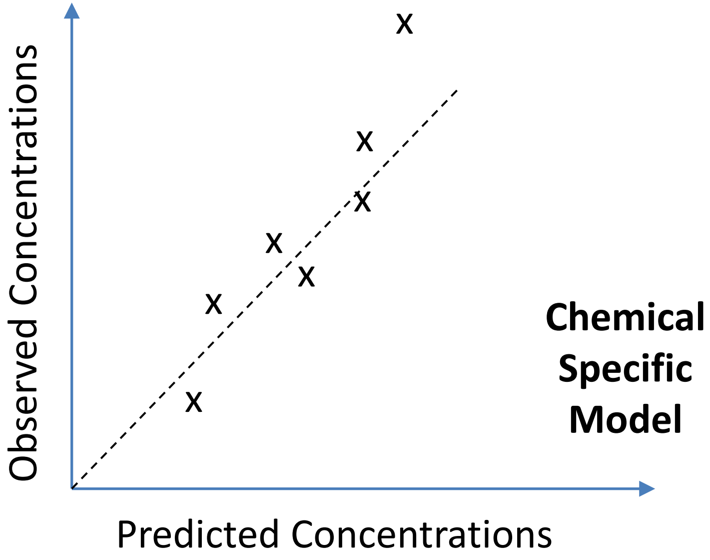
Chemical Specific Model

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 have no data

- However, we do not typically have TK data



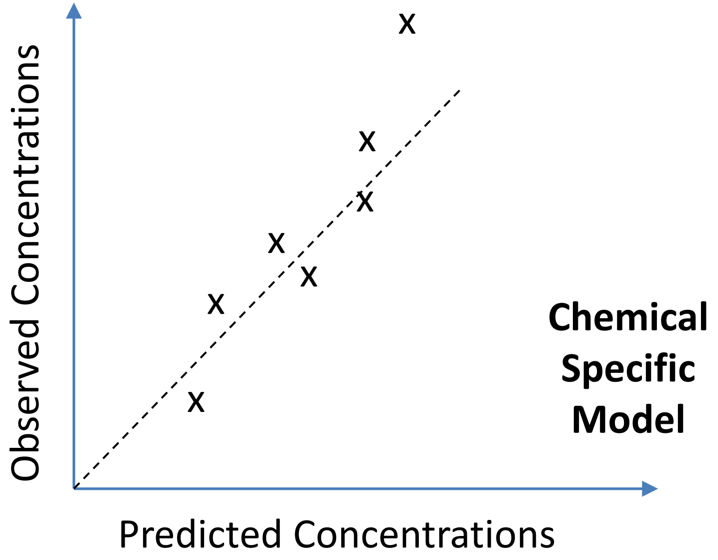
Cohen Hubal et al. (2018)

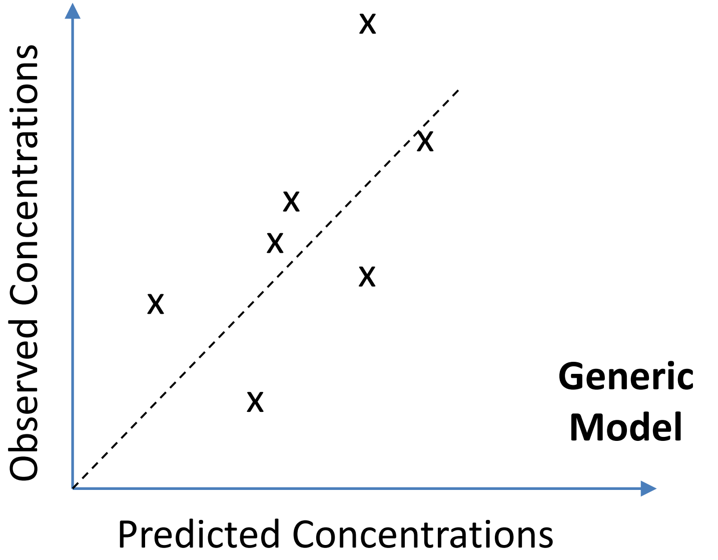
Building Confidence in TK Models

- 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 have no 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





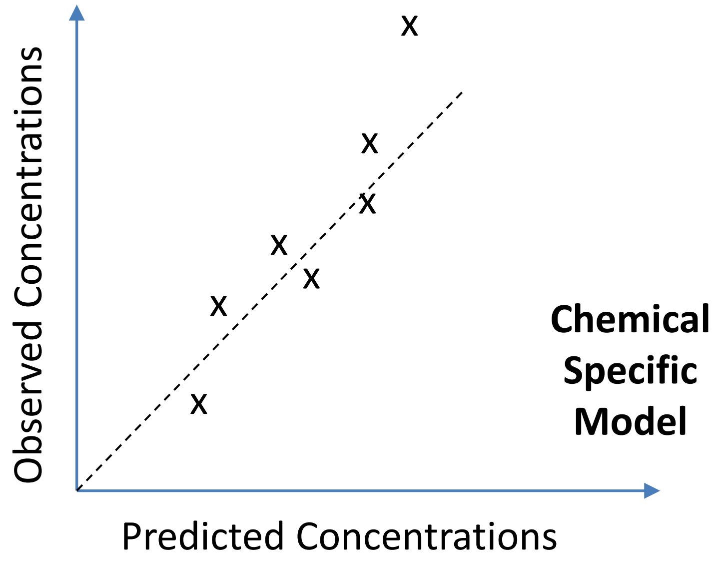
Cohen Hubal et al. (2018)

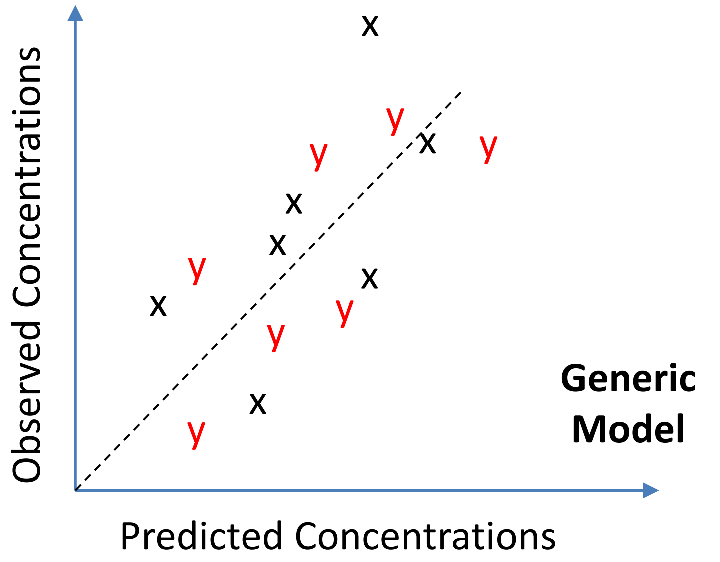
Building Confidence in TK Models

- 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 have no 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)





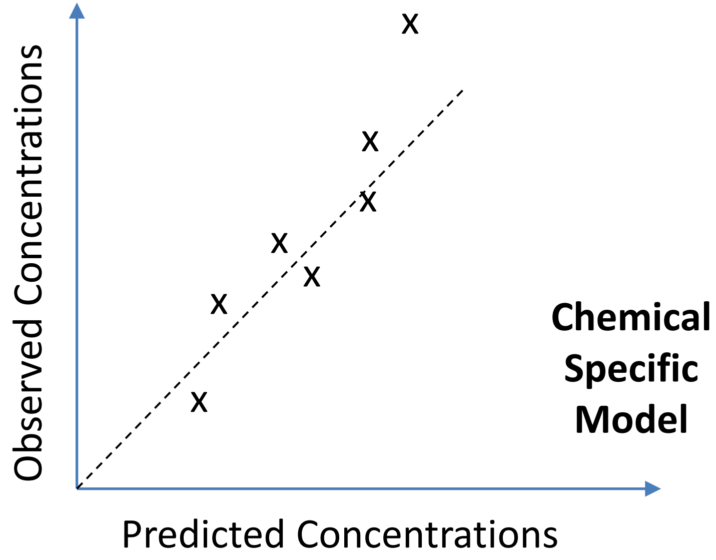
Cohen Hubal et al. (2018)

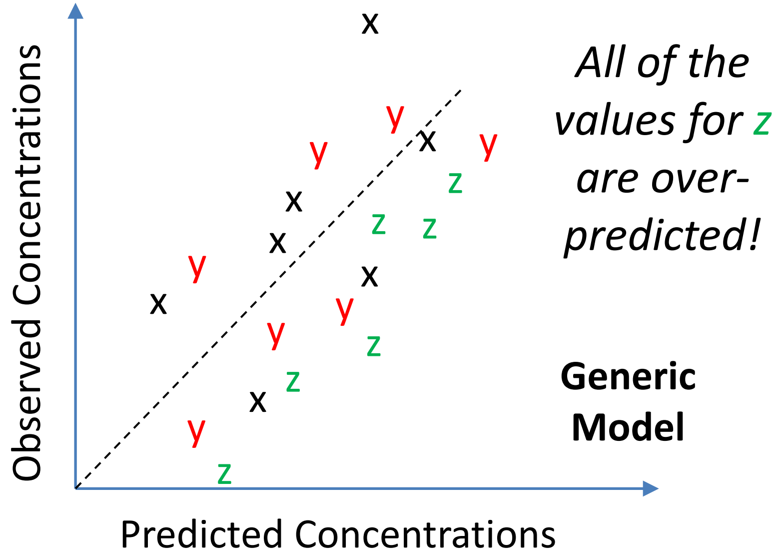
Building Confidence in TK Models

- 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 have no 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)





All of the values for z are over-predicted!

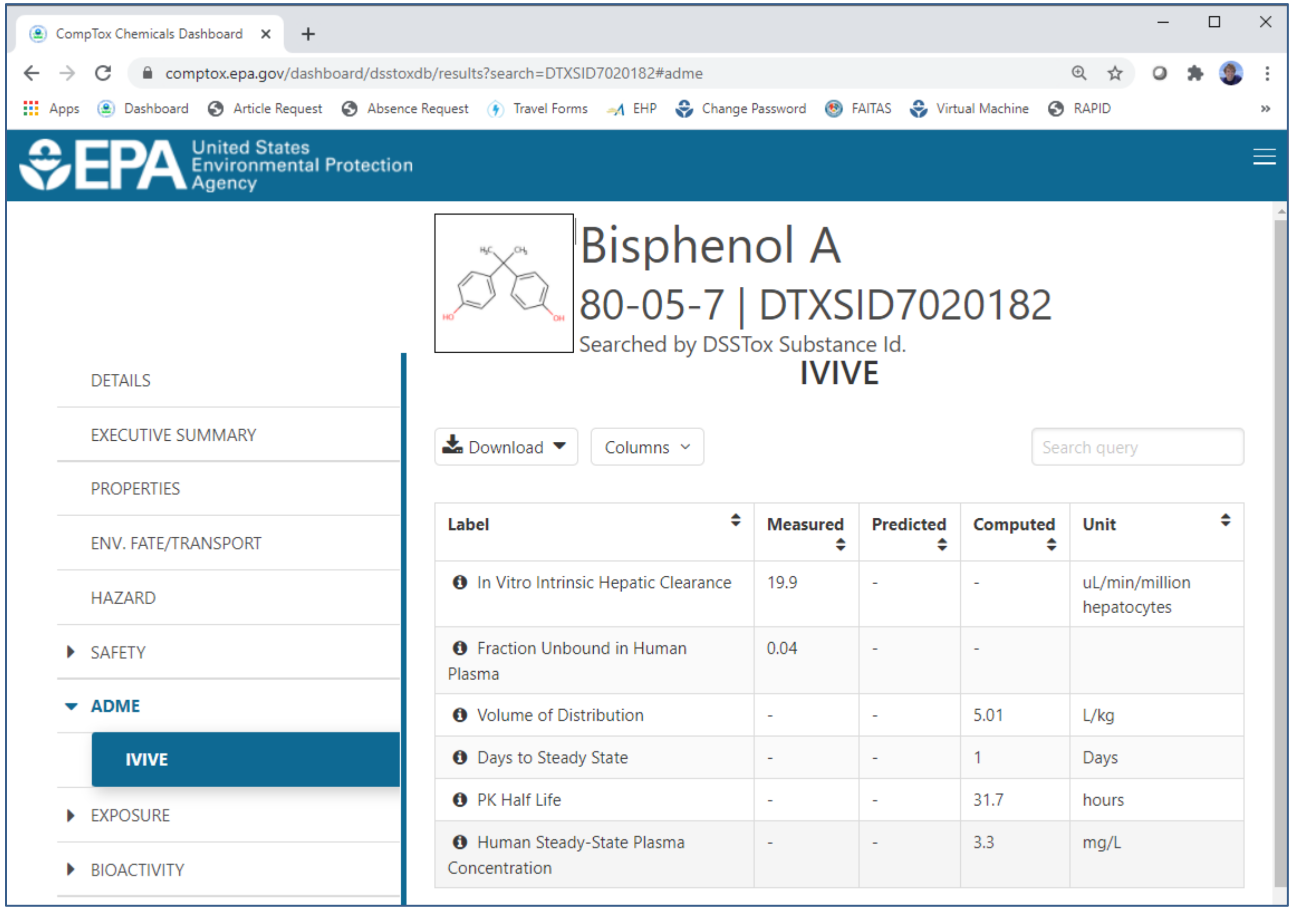
Cohen Hubal et al. (2018)

HTTK on the CompTox Chemicals Dashboard

- The CompTox Chemicals Dashboard provides $C_{ss,95}$ values for >1000 chemicals

https://comptox.epa.gov/dashboard/

- We use EPA's R package "httk" to provide IVIVE predictions
- The value reported is calculated assuming a 1 mg/kg/day dose rate
- We give the upper 95th percentile of the calculated values based on a Monte Carlo simulation of human variability and uncertainty



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

IVIVE

Label	Measured	Predicted	Computed	Unit
In Vitro Intrinsic Hepatic Clearance	19.9	-	-	uL/min/million hepatocytes
Fraction Unbound in Human Plasma	0.04	-	-	
Volume of Distribution	-	-	5.01	L/kg
Days to Steady State	-	-	1	Days
PK Half Life	-	-	31.7	hours
Human Steady-State Plasma Concentration	-	-	3.3	mg/L

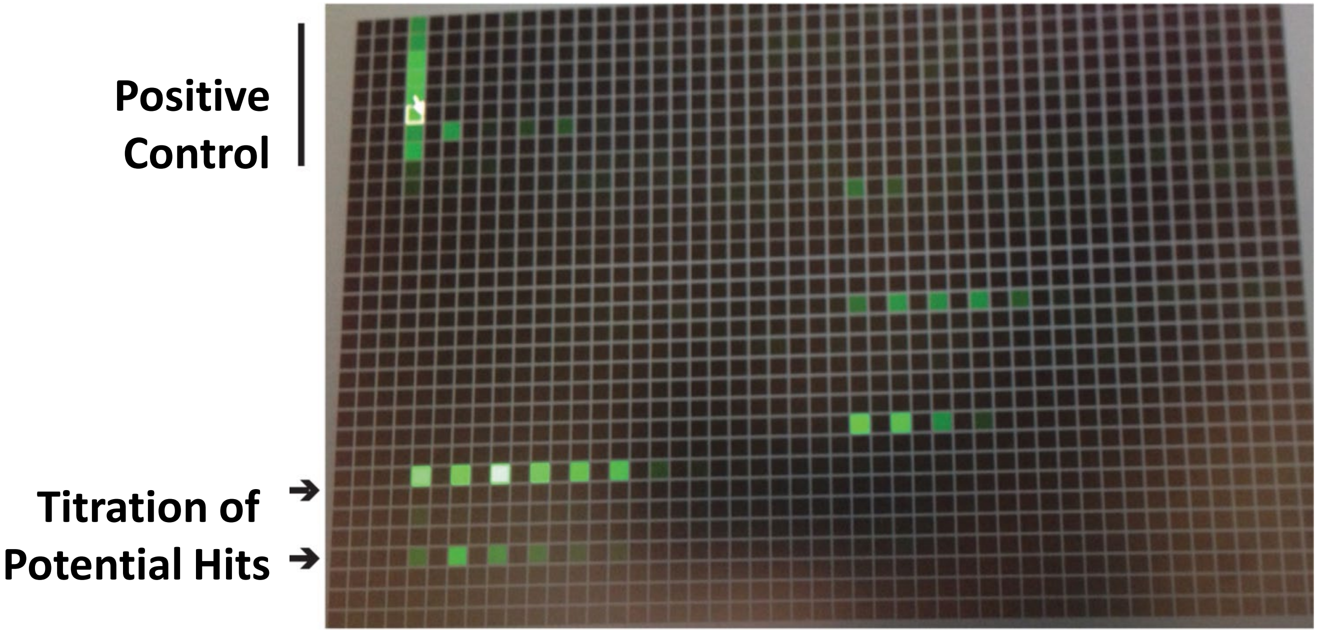
What is “High Throughput”?

- 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…

Kaewkhaw et al. (2016)



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…

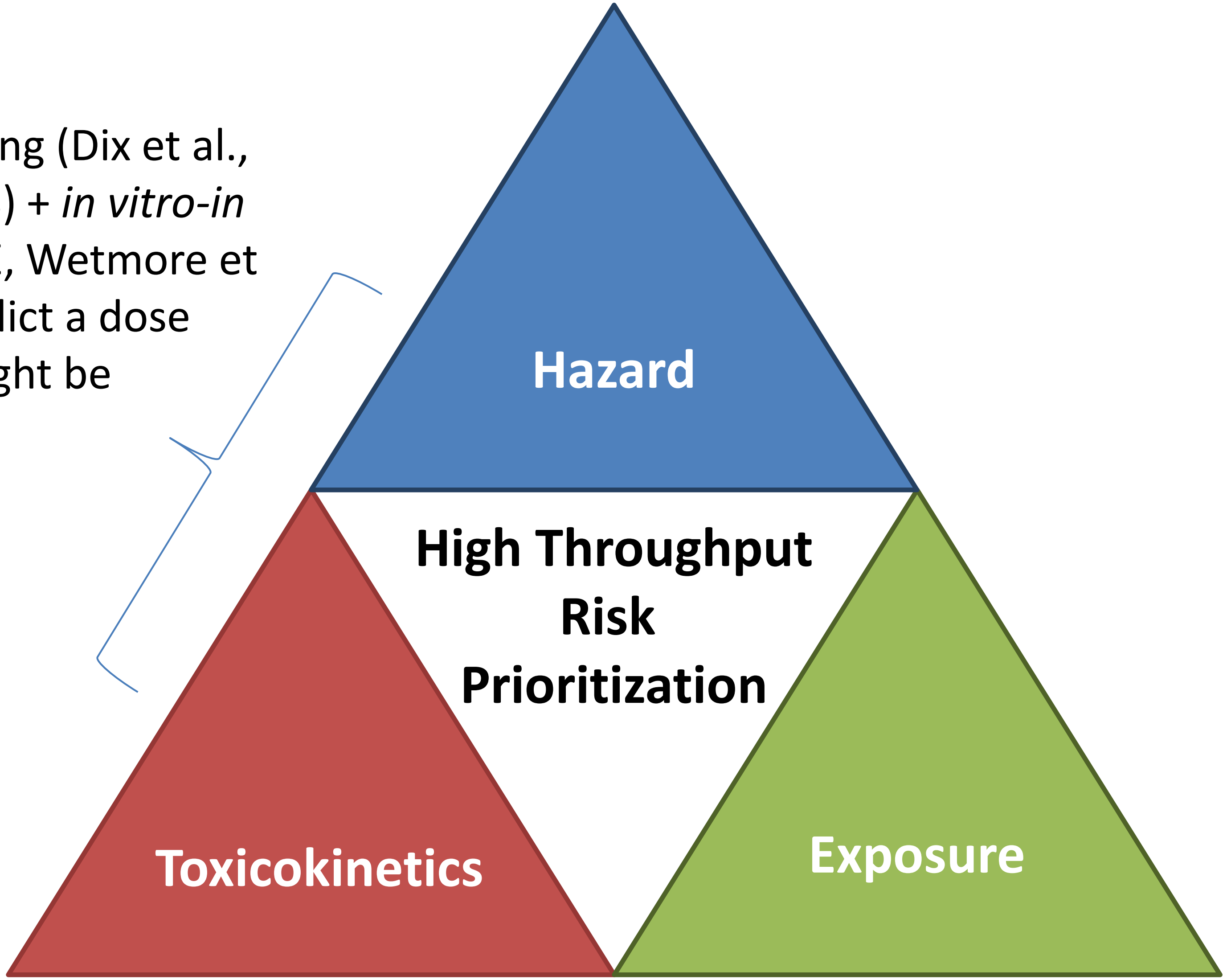
- ExpoCast: Ring et al. made *in silico* predictions for ~480,000 chemicals from structure, but based on NHANES monitoring for ~120 chemicals
    - Quantitative non-targeted analysis (NTA) may eventually provide greater evaluation data to reduce uncertainty

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

Our work is not done…

Risk = Hazard x Exposure

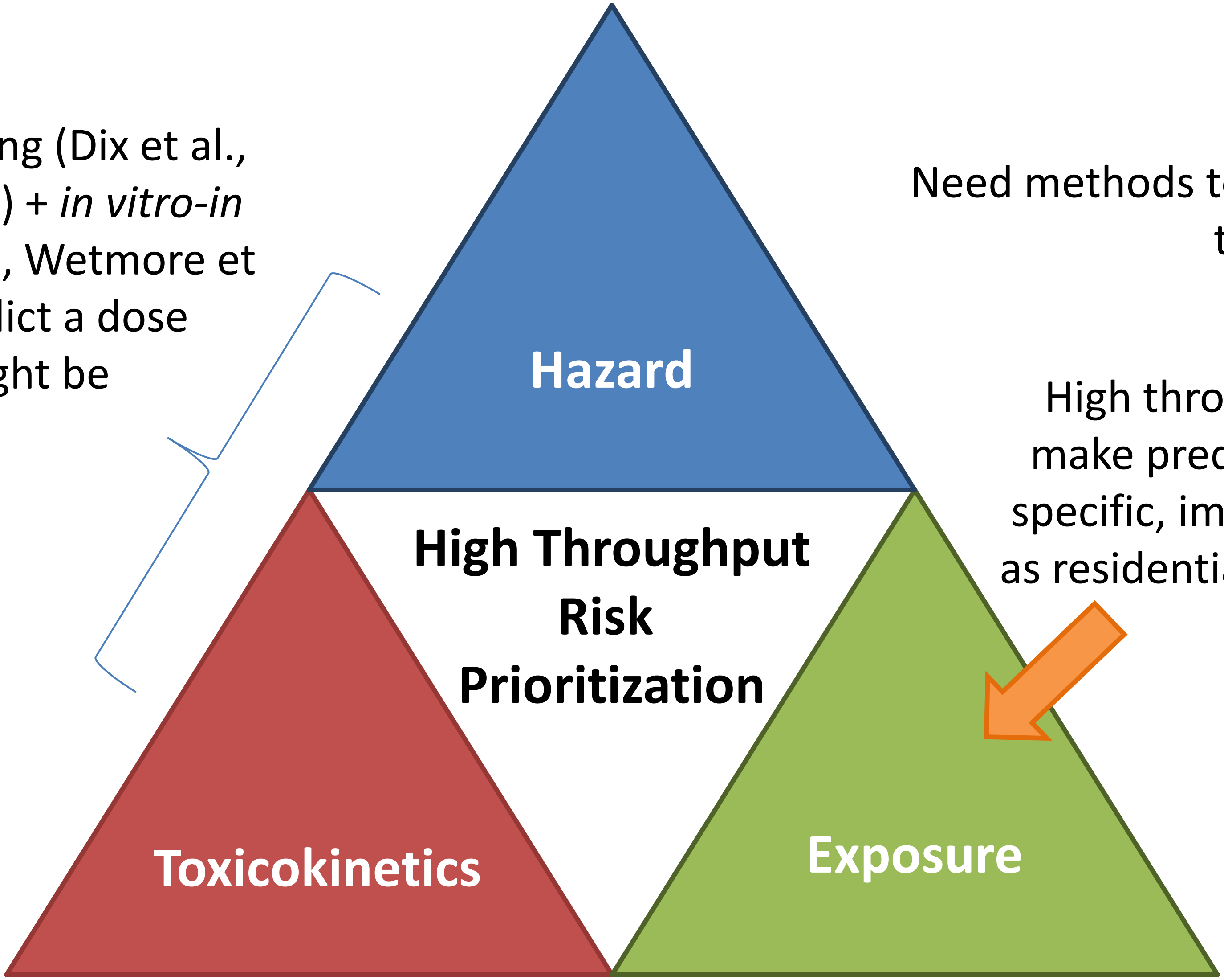
High throughput screening (Dix et al., 2006, Collins et al., 2008) + *in vitro-in vivo* extrapolation (IVIVE, Wetmore et al., 2012, 2015) can predict a dose (mg/kg bw/day) that might be adverse



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)

United States
Environmental Protection
Agency

Exposure is a Complex System

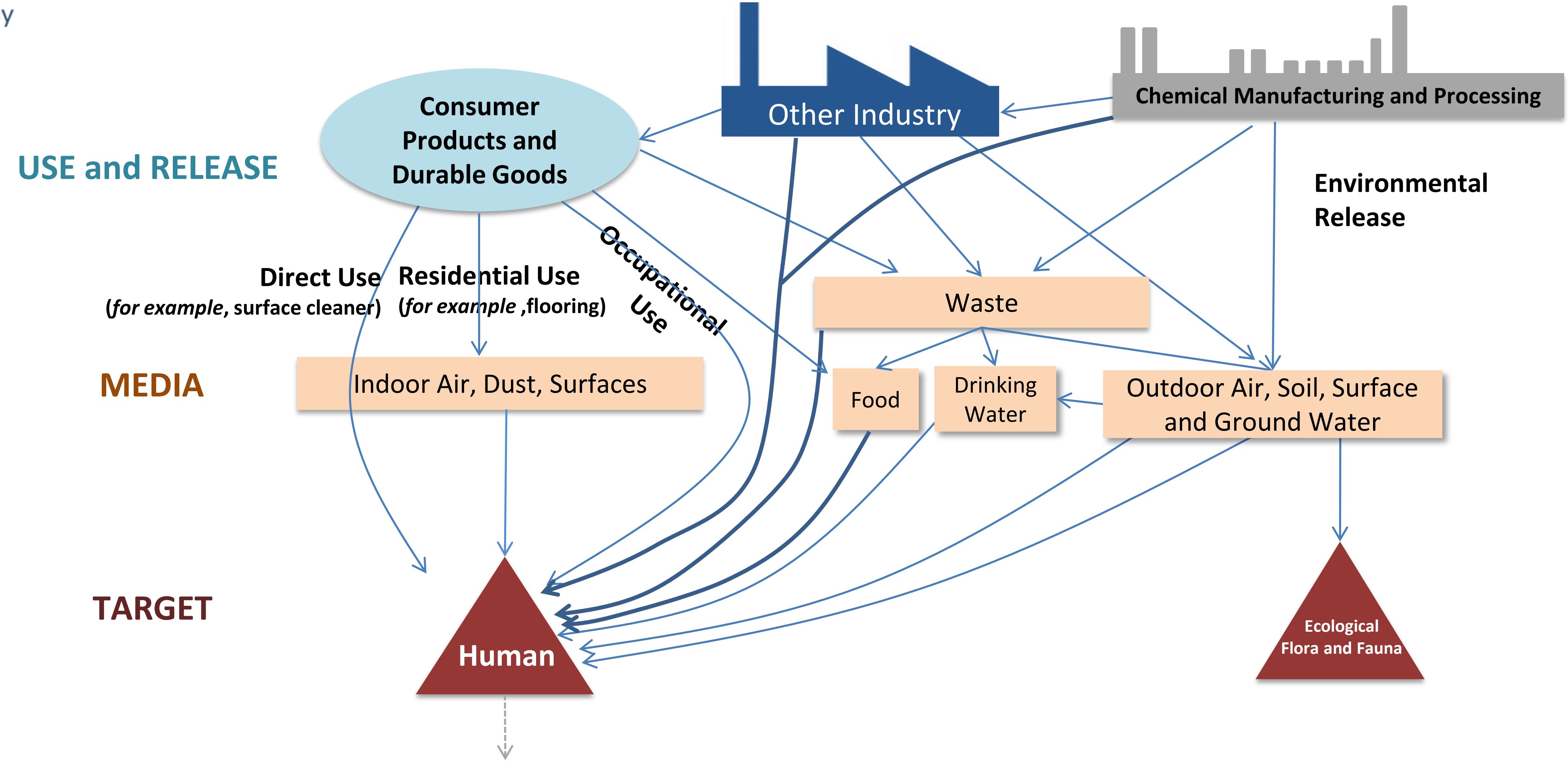
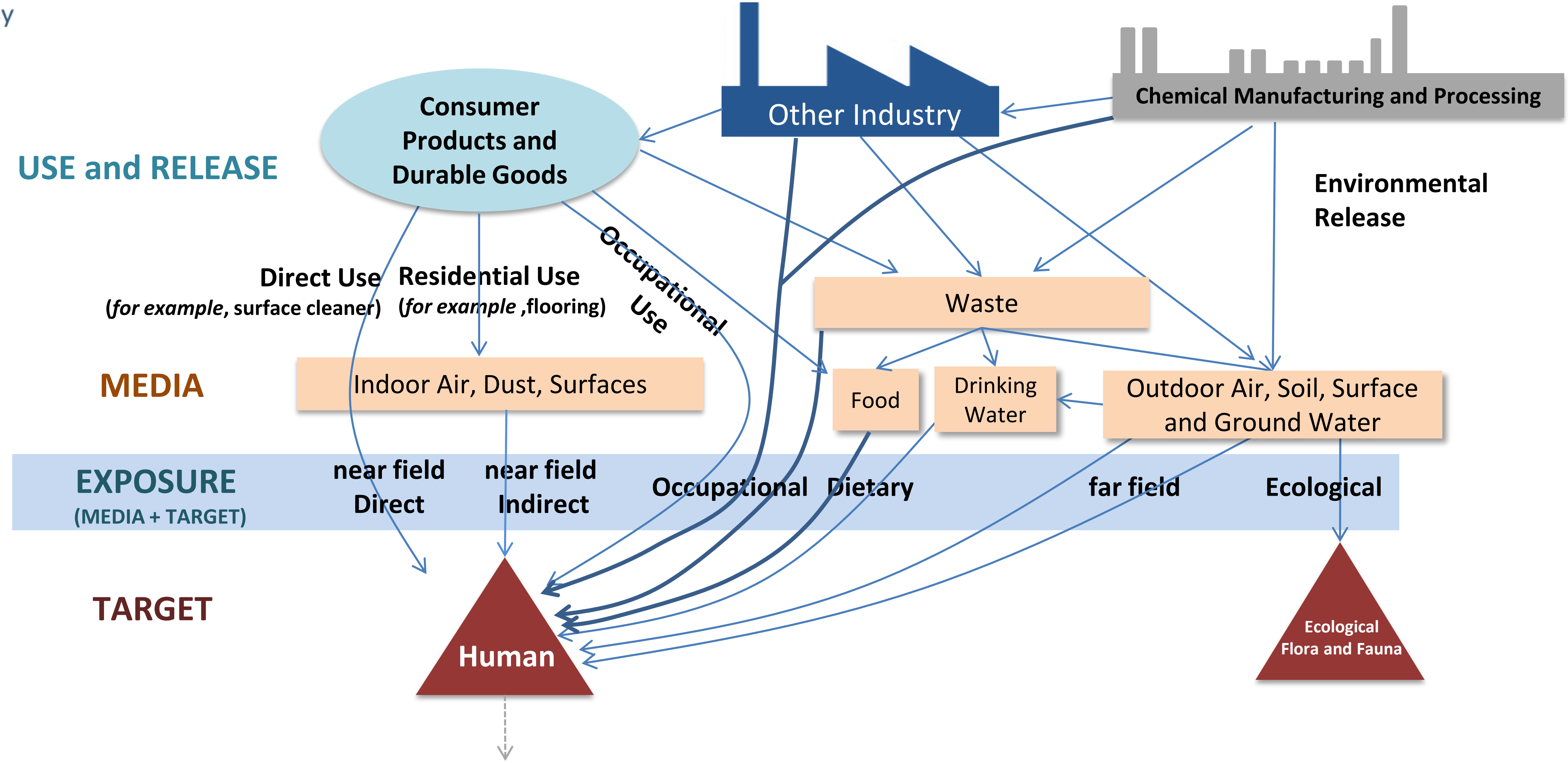


Figure from Kristin Isaacs

United States
Environmental Protection
Agency

The Exposure Event is Often Unobservable



USE and RELEASE

Other Industry

Chemical Manufacturing and Processing

Consumer Products and Durable Goods

Environmental Release

Direct Use
(*for example*, surface cleaner)

Residential Use
(*for example* ,flooring)

Occupational Use

Waste

MEDIA

Indoor Air, Dust, Surfaces

Food

Drinking Water

Outdoor Air, Soil, Surface and Ground Water

EXPOSURE
(MEDIA + TARGET)

near field
Direct

near field
Indirect

Occupational

Dietary

far field

Ecological

TARGET

Human

Ecological Flora and Fauna

- 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

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**

nhanes
National Health and Nutrition Examination Survey

What Do We Know About Exposure?
Exposure Models

- Any model, including those for exposure, capture knowledge and a hypothesis of how the world works

- 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)

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

Everyone Uses Models

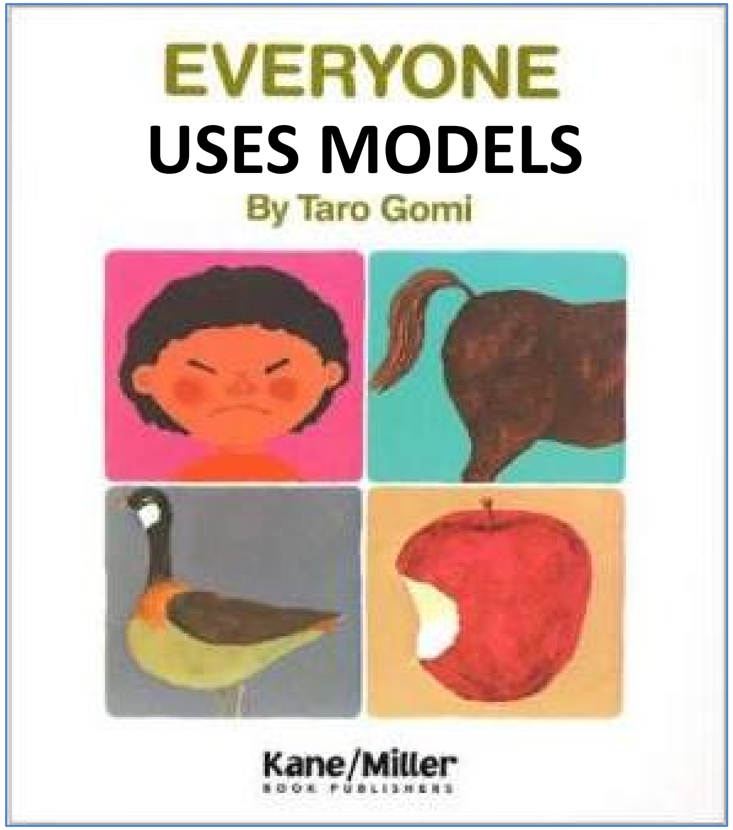
- Toxicology has long relied upon model animal species

- People rely on mental models every day
  - For example, repetitive activities like driving home from work

- Mathematical models offer some significant advantages:
  - Reproducible
  - Can (and should) be transparent

- …with some disadvantages:
  - Sometimes reality is complex
  - Sometimes the model doesn’t always work well
  - How do we know we can extrapolate?

- …that can be turned into advantages:
  - If we have evaluated confidence/uncertainty and know the “domain of applicability” we can make better use of mathematical models



Fit for Purpose Models

- A “fit for purpose” model is an abstraction of a complicated problem that allows us to reach a decision.

  “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

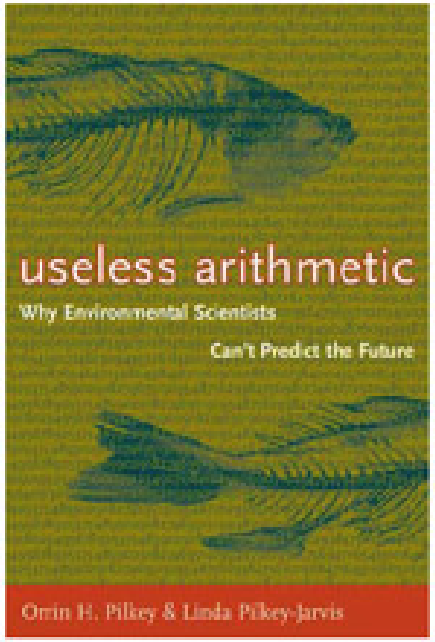
- A fit for purpose model is defined as much by what is omitted as what is included in the model.

- We must accept that there will always be areas in need of better data and models – our knowledge will always be incomplete, and thus we wish to extrapolate.

  - How do I drive to a place I’ve never been before?

How to Make Good Forecasts

Adapted from Nate Silver



useless arithmetic

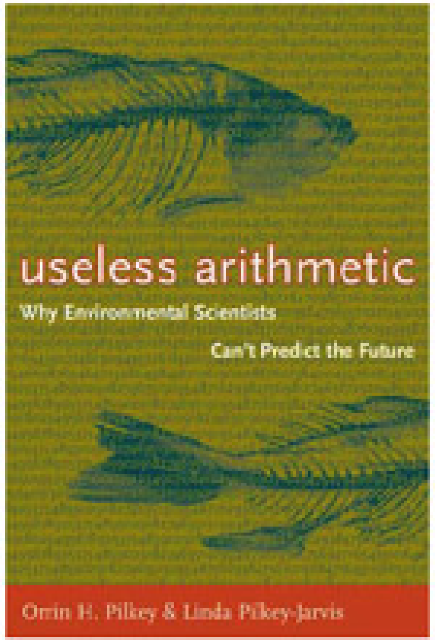
Why Environmental Scientists Can't Predict the Future

Orrin H. Pilkey & Linda Pilkey-Jarvis

Orrin Pilkey &
Olinda Pilkey-Jarvis (2007)

How to Make Good Forecasts

Adapted from Nate Silver

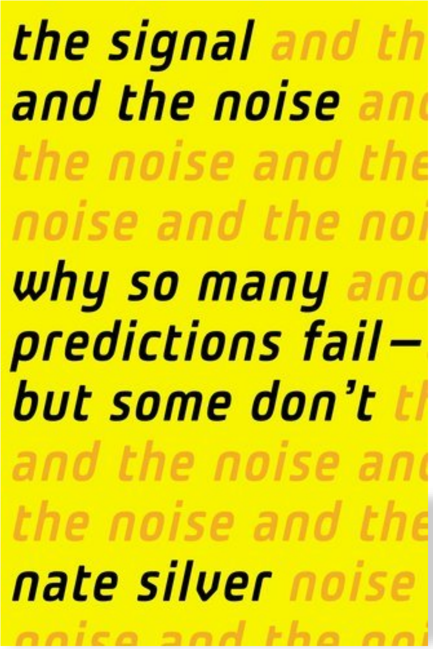


Orrin Pilkey &
Olinda Pilkey-Jarvis (2007)

1) Think probabilistically (especially, Bayesian): We use an approach that evaluates model performance systematically across as many chemicals (and chemistries) as possible

2) Forecasts change: Today's forecast reflects the best available data today but we must accept that new data and new models will cause predictions to be revised

3) Look for consensus: We evaluate as many models and predictors/ predictions as possible

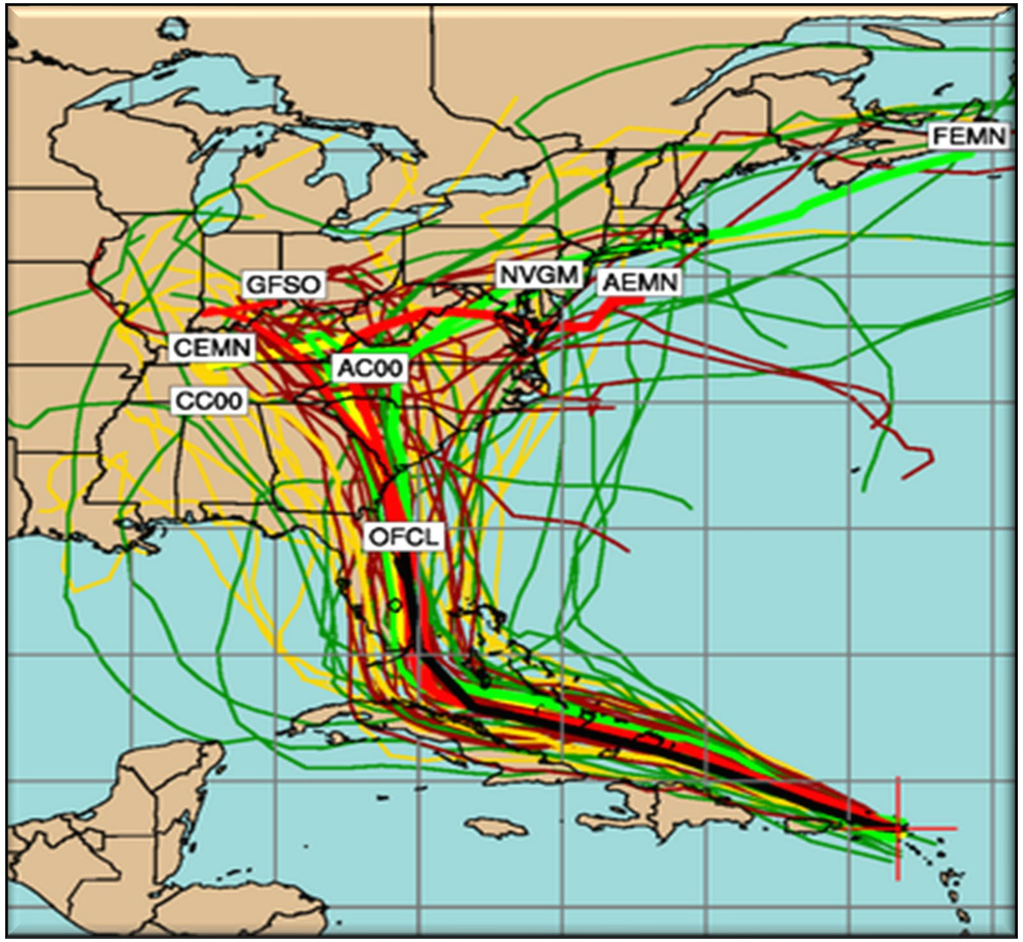


Nate Silver (2012)

In Nate Silver's terminology:
a ***prediction*** is a specific statement
a ***forecast*** is a probabilistic statement

Ensemble Predictions

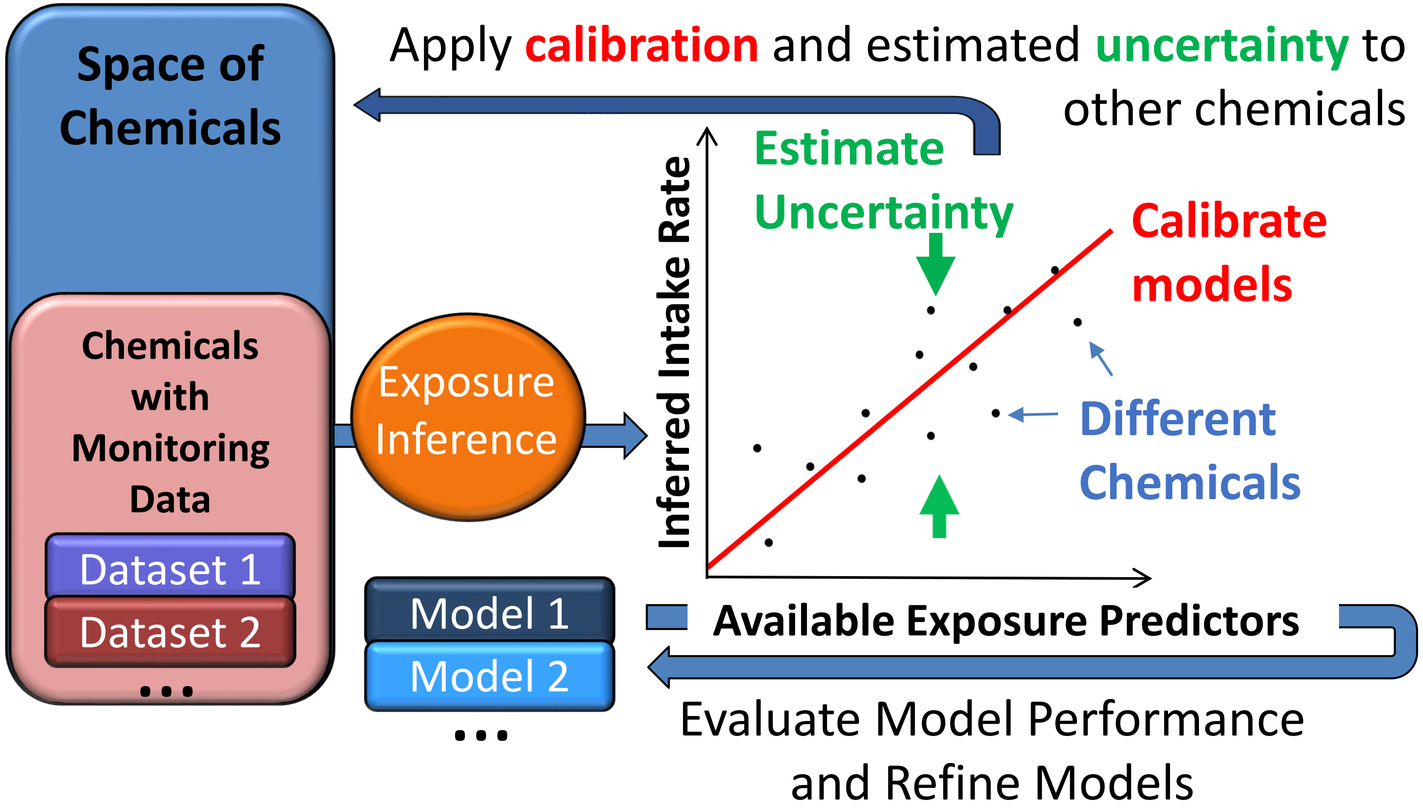
- We can use ensemble methods to make more stable models and characterize uncertainty
- "Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a (weighted) vote of their predictions." Dietterich (2000)
- Ensemble systems have proven themselves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



Hurricane Path Prediction is an Example of Integrating Multiple Models

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)



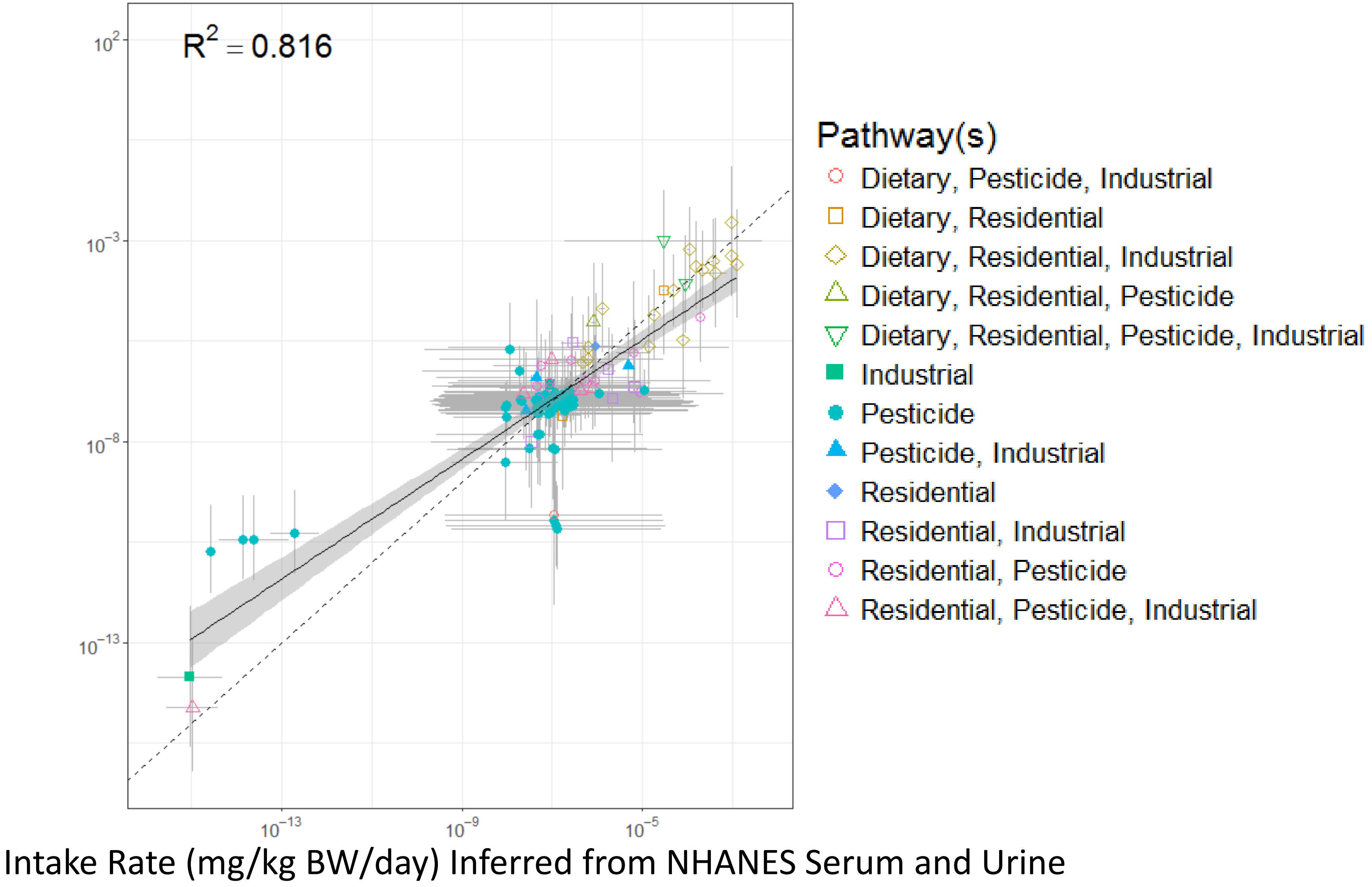
SEEM3 Collaboration

Jon Arnot, Deborah H. Bennett, Peter P. Egeghy, Peter Fantke, Lei Huang, Kristin K. Isaacs, Olivier Jolliet, Hyeong-Moo Shin, Katherine A. Phillips, Caroline Ring, R. Woodrow Setzer, John F. Wambaugh, Johnny Westgate

Predictor	Reference(s)	Chemicals Predicted	Pathway(s)
EPA Inventory Update Reporting and Chemical Data Reporting (CDR) (2015)	US EPA (2018)	7856	All
Stockholm Convention of Banned Persistent Organic Pollutants (2017)	Lallas (2001)	248	far field Industrial and Pesticide
EPA Pesticide Reregistration Eligibility Documents (REDs) Exposure Assessments (Through 2015)	Wetmore et al. (2012, 2015)	239	far field Pesticide
United Nations Environment Program and Society for Environmental Toxicology and Chemistry toxicity model (USEtox) Industrial Scenario (2.0)	Rosenbaum et al. (2008)	8167	far field Industrial
USEtox Pesticide Scenario (2.0)	Fantke et al. (2011, 2012, 2016)	940	far field Pesticide
Risk Assessment IDentification And Ranking (RAIDAR) far field (2.02)	Arnot et al. (2008)	8167	far field Pesticide
EPA Stochastic Human Exposure Dose Simulator High Throughput (SHEDS-HT) near field Direct (2017)	Isaacs (2017)	7511	far field Industrial and Pesticide
SHEDS-HT near field Indirect (2017)	Isaacs (2017)	1119	Residential
Fugacity-based INdoor Exposure (FINE) (2017)	Bennett et al. (2004), Shin et al. (2012)	645	Residential
RAIDAR-ICE near field (0.803)	Arnot et al., (2014), Zhang et al. (2014)	1221	Residential
USEtox Residential Scenario (2.0)	Jolliet et al. (2015), Huang et al. (2016,2017)	615	Residential
USEtox Dietary Scenario (2.0)	Jolliet et al. (2015), Huang et al. (2016), Ernstoff et al. (2017)	8167	Dietary

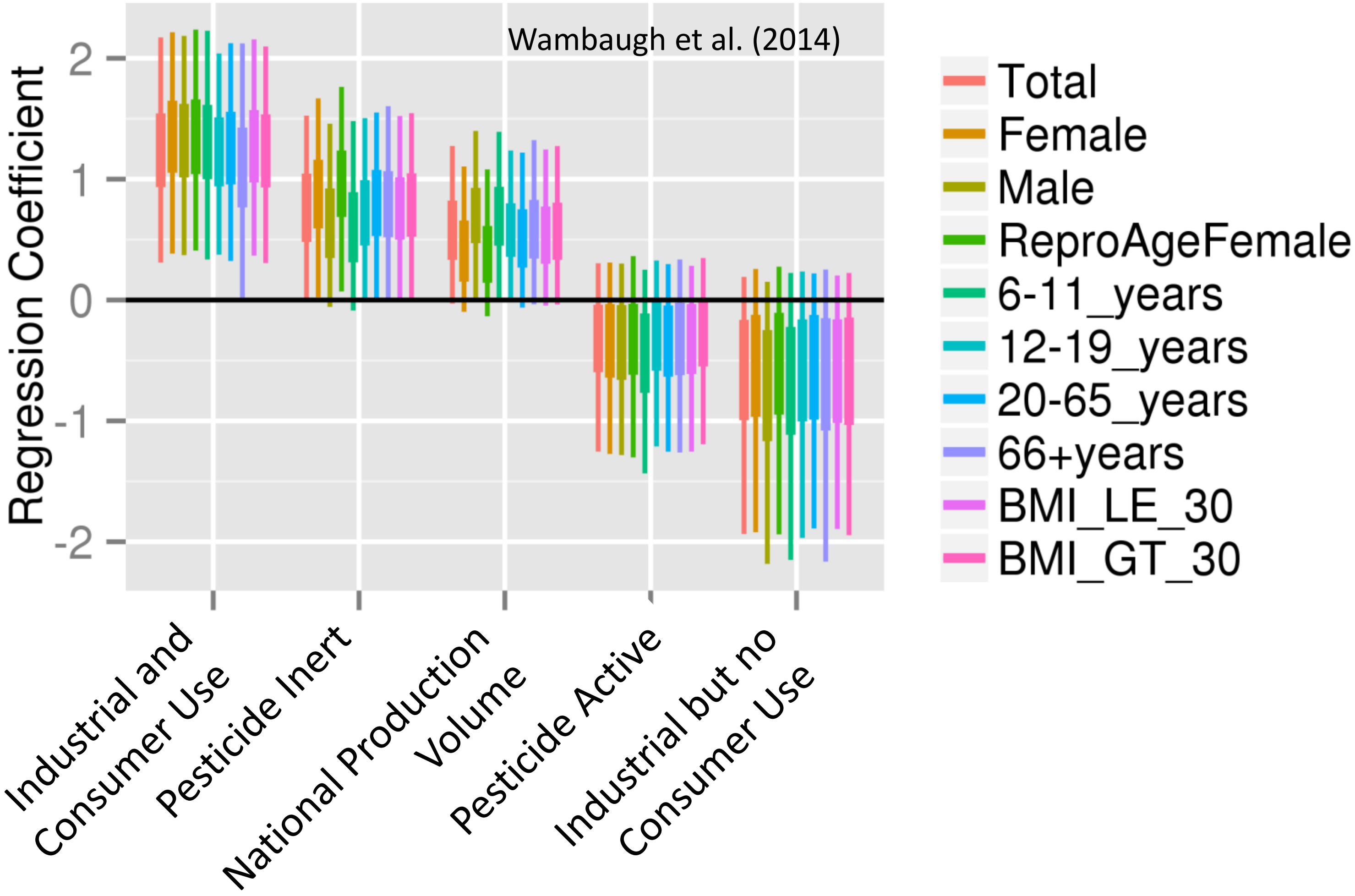
Ring et al. (2018)

SEEM3: Pathway-Based Consensus Modeling



Ring et al., 2018

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

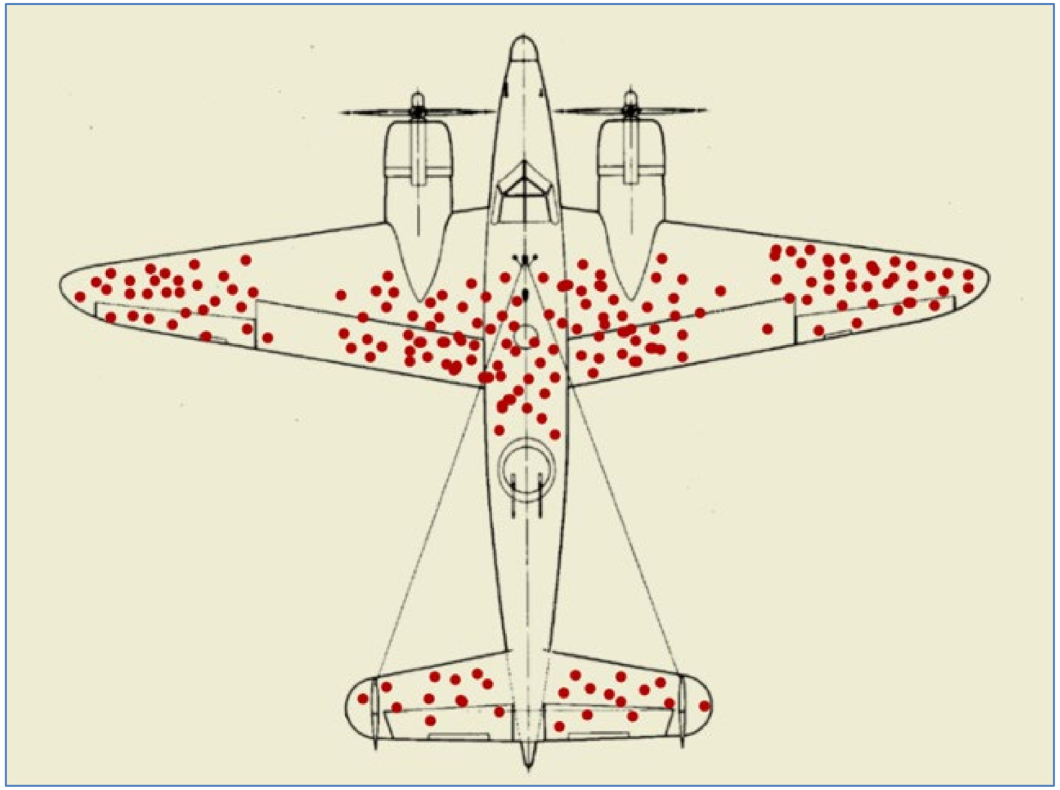
Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

- 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, the 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

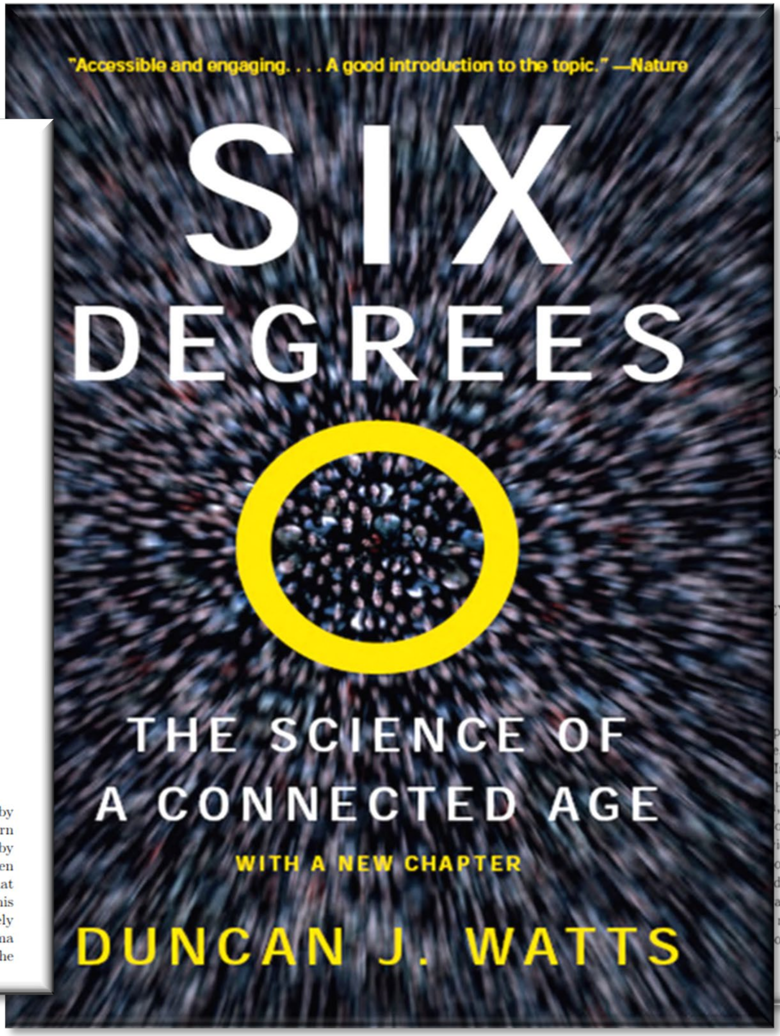
Abstract. In this paper, we will propose a new algorithm that computes the radius and the diameter of a graph $G = (V, E)$, by finding bounds through heuristics and improving them until exact values can be guaranteed. Although the worst-case running time is $\mathcal{O}(|V| \cdot |E|)$, we will experimentally 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).



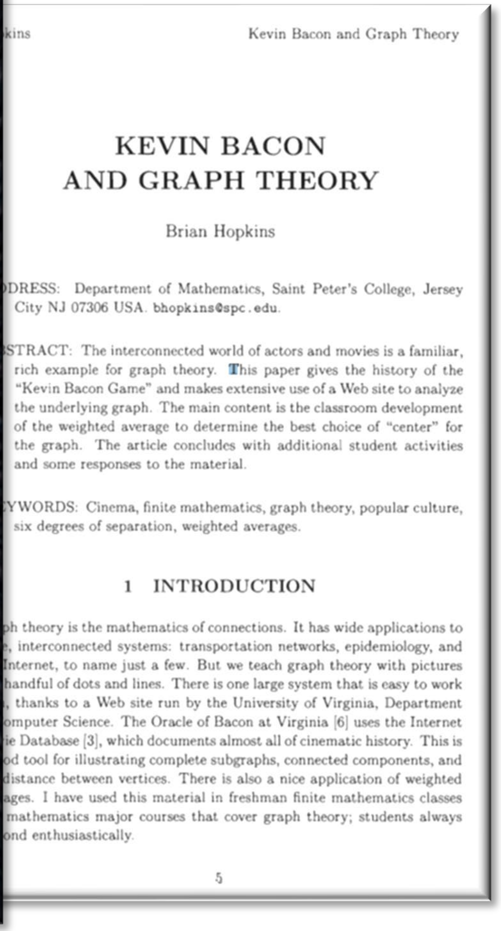
"Accessible and engaging. . . . A good introduction to the topic." —Nature

SIX DEGREES

THE SCIENCE OF A CONNECTED AGE

WITH A NEW CHAPTER

DUNCAN J. WATTS



KEVIN BACON AND GRAPH THEORY

Brian Hopkins

DRESS: 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

ph theory is the mathematics of connections. It has wide applications to , interconnected systems: transportation networks, epidemiology, and Internet, 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 ie Database [3], which documents almost all of cinematic history. This is od 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.

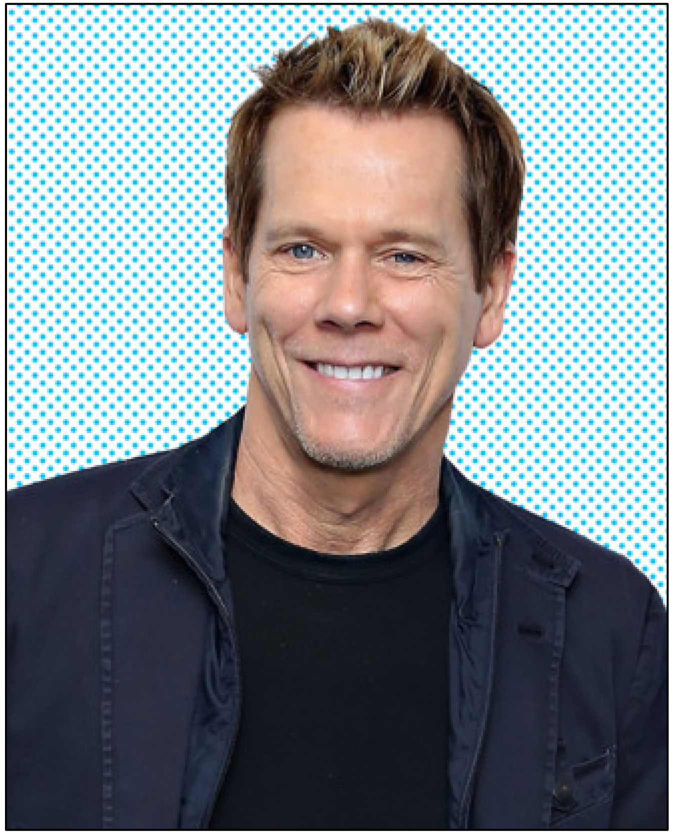
Kevin Bacon

1978

1984

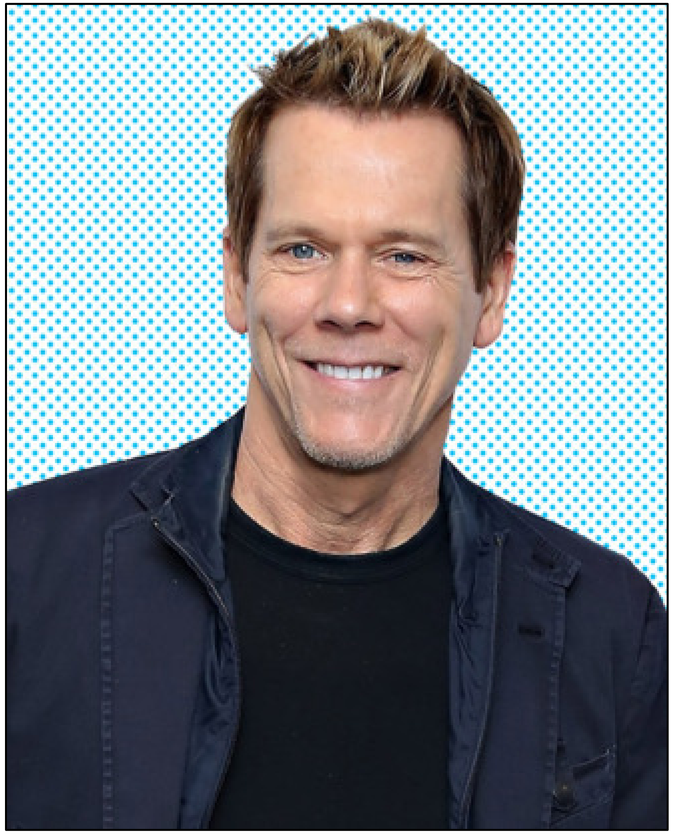
1992

1995



Kevin Bacon

1990



Michael B. Jordan

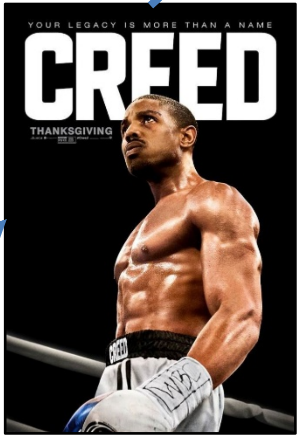
Connectedness to Michael B. Jordan

Frances McDormand
Best Actress Winner 2018

Hail Caesar
McDormand &
Channing Tatum

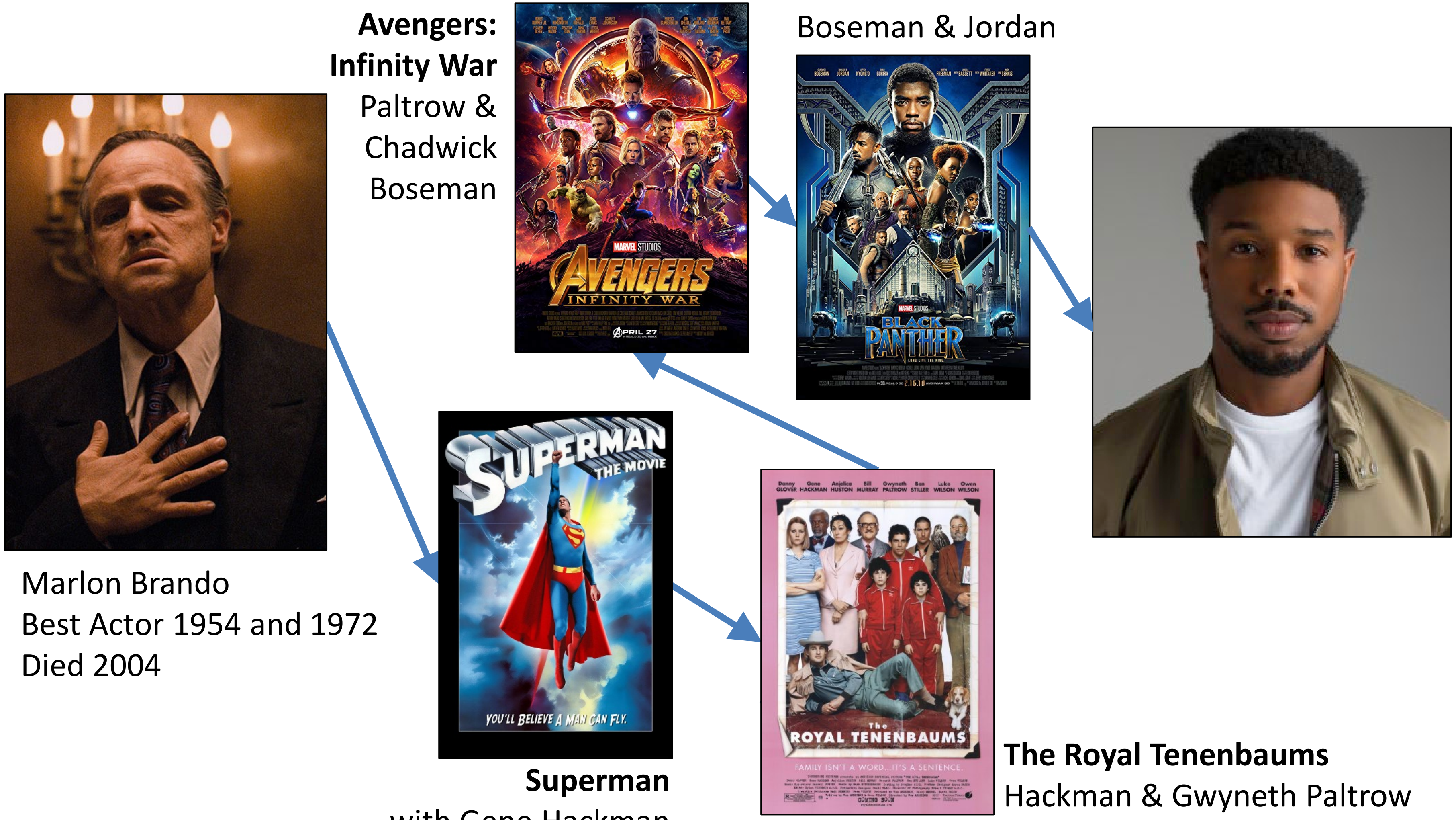
GI Joe: Retaliation
Tatum & Bruce Willis

Expendables
Willis &
Sylvester Stallone



Creed
Stallone & Jordan

Connectedness to Michael B. Jordan



Marlon Brando
Best Actor 1954 and 1972
Died 2004

Superman
with Gene Hackman

The Royal Tenenbaums
Hackman & Gwyneth Paltrow

Avengers: Infinity War
Paltrow & Chadwick Boseman

Black Panther
Boseman & Jordan

United States
Environmental Protection
Agency

Small World Networks

Watts and Strogatz (1998)

letters to nature

typically slower than $\sim 1\,\mathrm{km\,s^{-1}}$) 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).
3. 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 $p \rightarrow 0$, while $L \approx L_{\mathrm{random}} \sim \ln(n)/\ln(k)$ and $C \approx C_{\mathrm{random}} \sim k/n \ll 1$ as $p \rightarrow 1$. 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 $C(p) \gg C_{\mathrm{random}}$. 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.

440

NATURE | VOL 393 | 4 JUNE 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[3] 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[5], 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[5] 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).

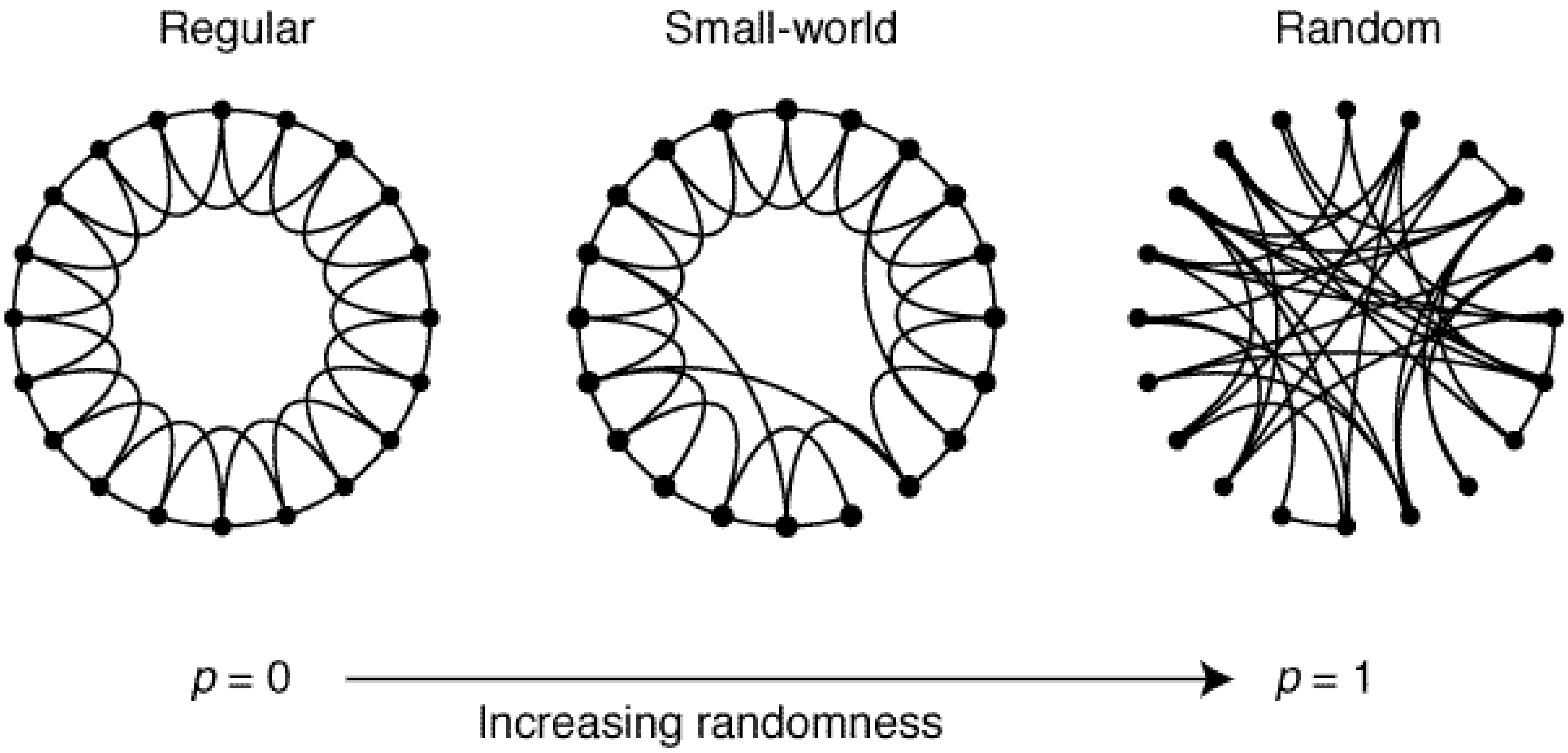
Small-world connectivity has consequences that could be good or bad,

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 | VOL 393 | 4 JUNE 1998

409

Complex is Not the Same as Random



Watts and Strogatz (1998)

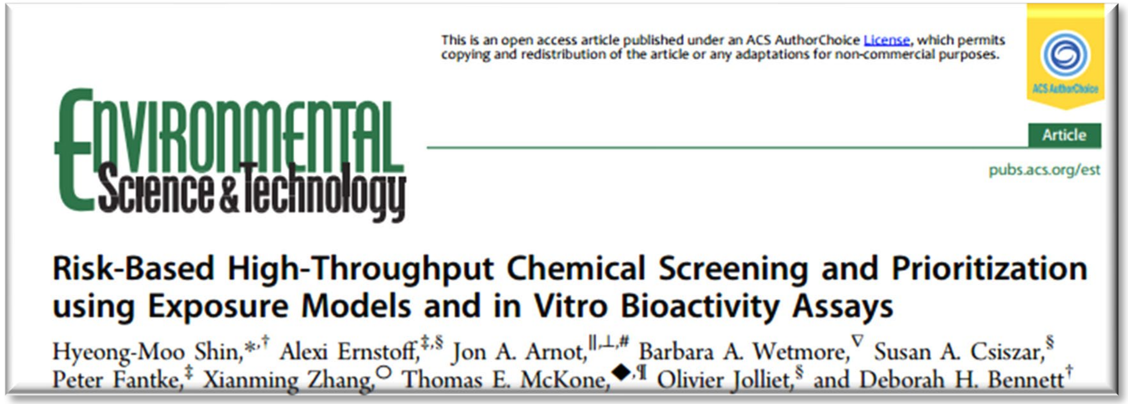
Knowledge of Exposure Pathways Limits High Throughput Exposure Models

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

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

- But we don’t always know how chemicals are used:

> “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

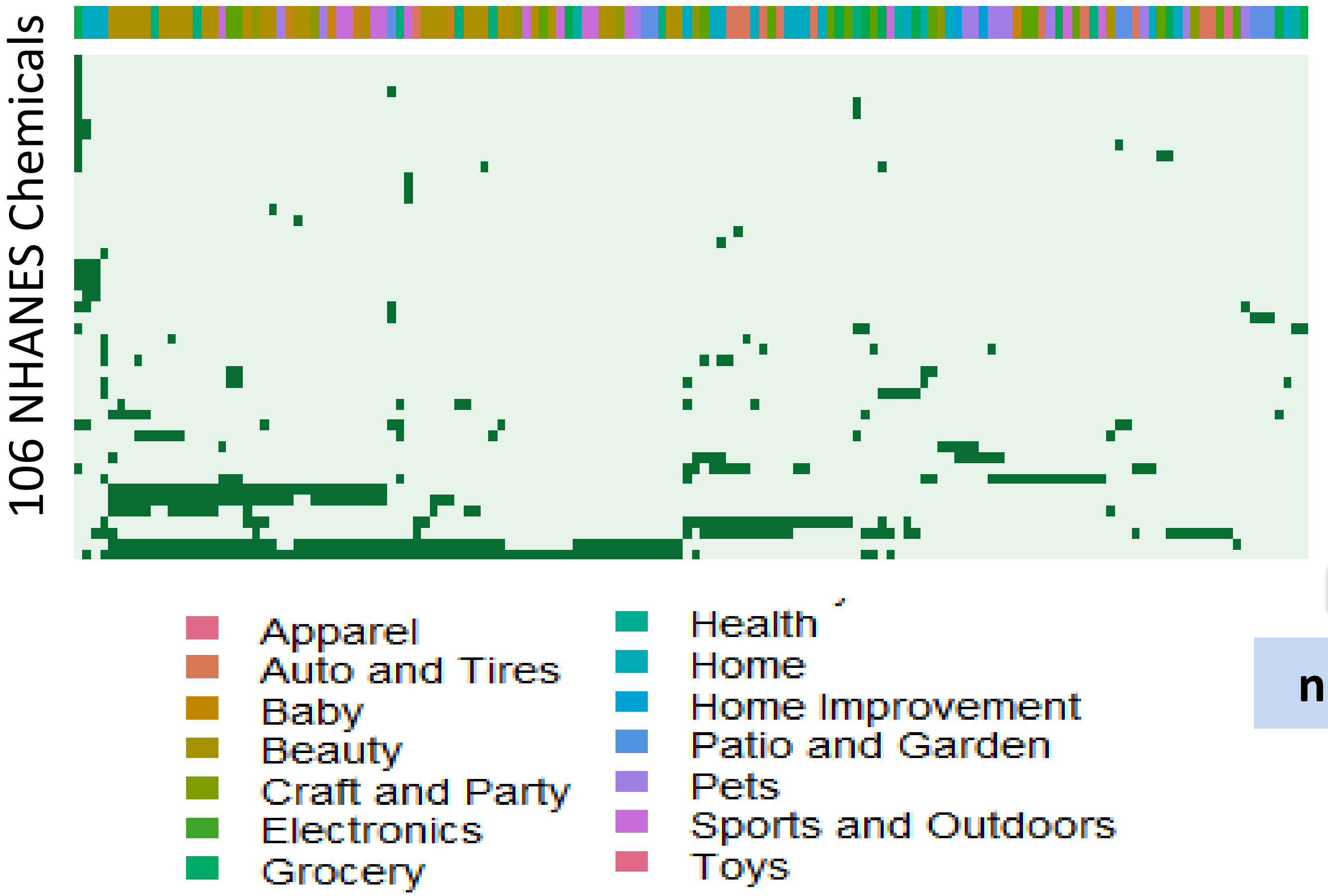
pubs.acs.org/est

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

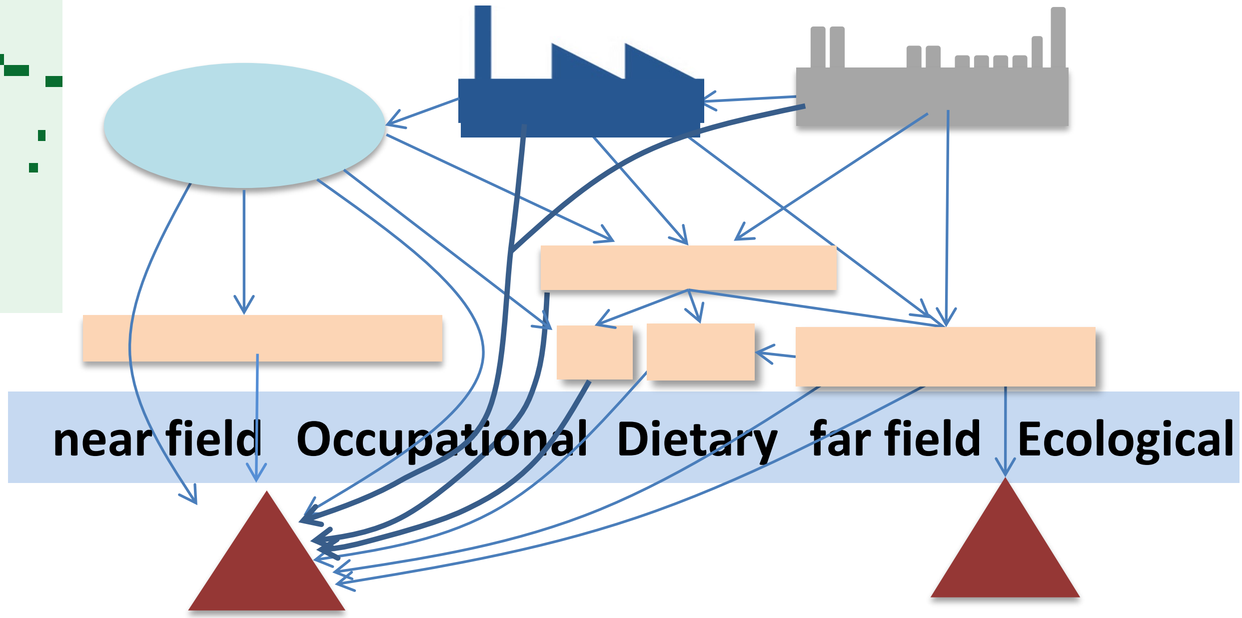
Hyeong-Moo Shin,[*,†] Alexi Ernstoff,[‡,§] Jon A. Arnot,[∥,⊥,#] Barbara A. Wetmore,[∇] Susan A. Csiszar,[§] Peter Fantke,[‡] Xianming Zhang,[○] Thomas E. McKone,[◆,¶] Olivier Jolliet,[§] and Deborah H. Bennett[†]

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

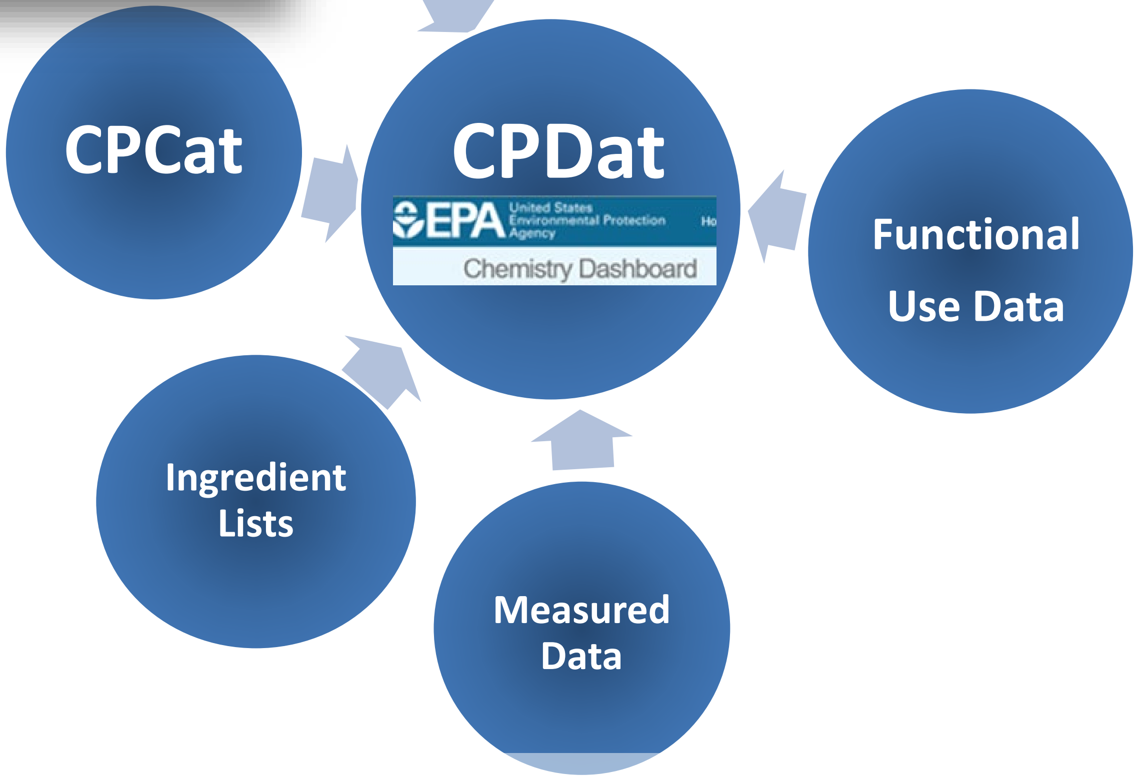
ELSEVIER

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



CPCat

CPDat

EPA United States Environmental Protection Agency

Chemistry Dashboard

Functional Use Data

Ingredient Lists

Measured Data

Slide from Kristin Isaacs

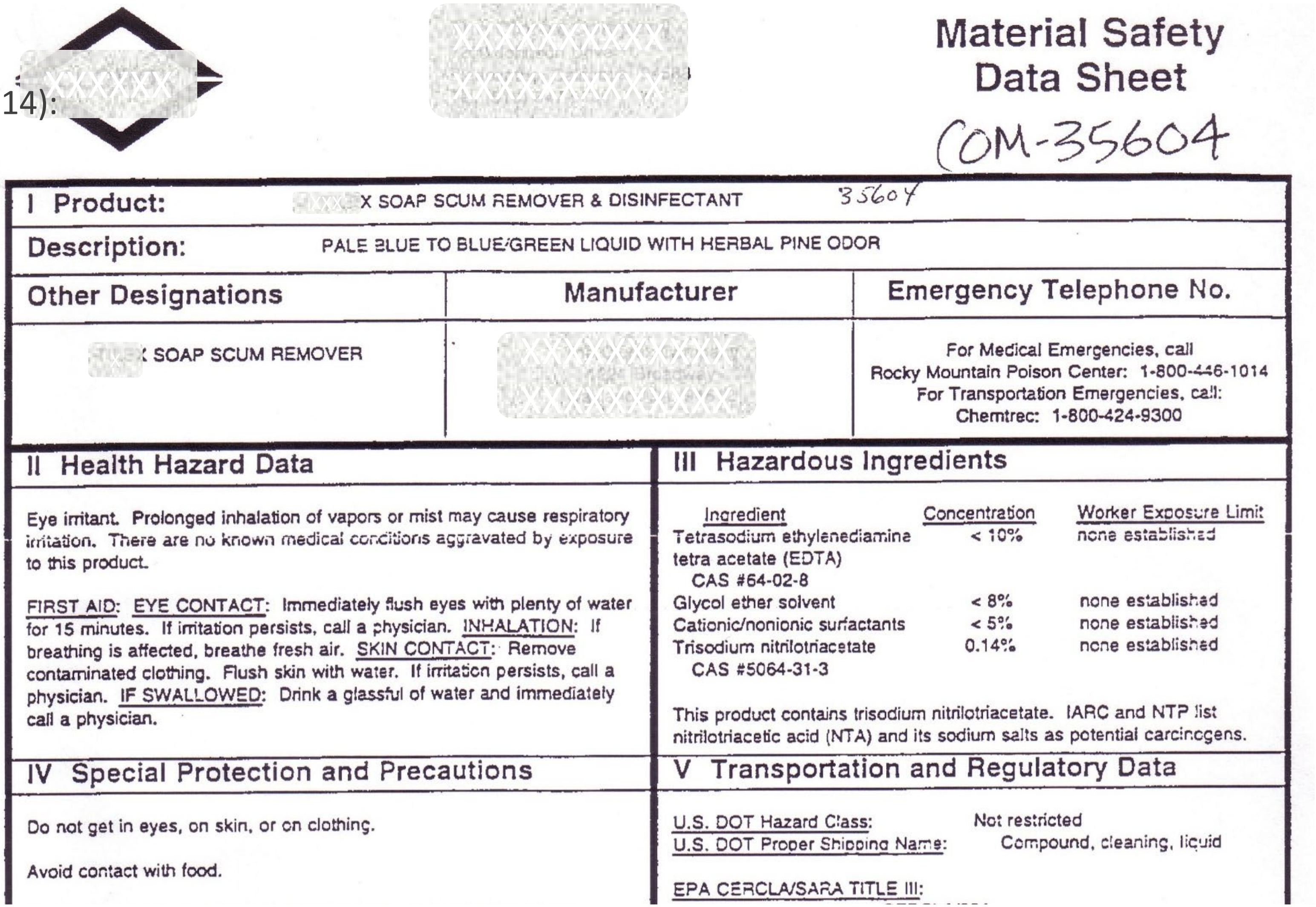
https://comptox.epa.gov/dashboard

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 (for example, home goods, bath soaps, baby) to find each product



[redacted]

[redacted]

Material Safety Data Sheet

COM-35604

I Product:	[redacted]X SOAP SCUM REMOVER & DISINFECTANT 35604
Description:	PALE BLUE TO BLUE/GREEN LIQUID WITH HERBAL PINE ODOR

Other Designations	**Manufacturer**	**Emergency Telephone No.**
[redacted]X SOAP SCUM REMOVER	[redacted]	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:

United States
Environmental Protection
Agency

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

ELSEVIER

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]

Occurrence and quantitative chemical composition

Broad "index" of chemical uses

Contents lists available at ScienceDirect

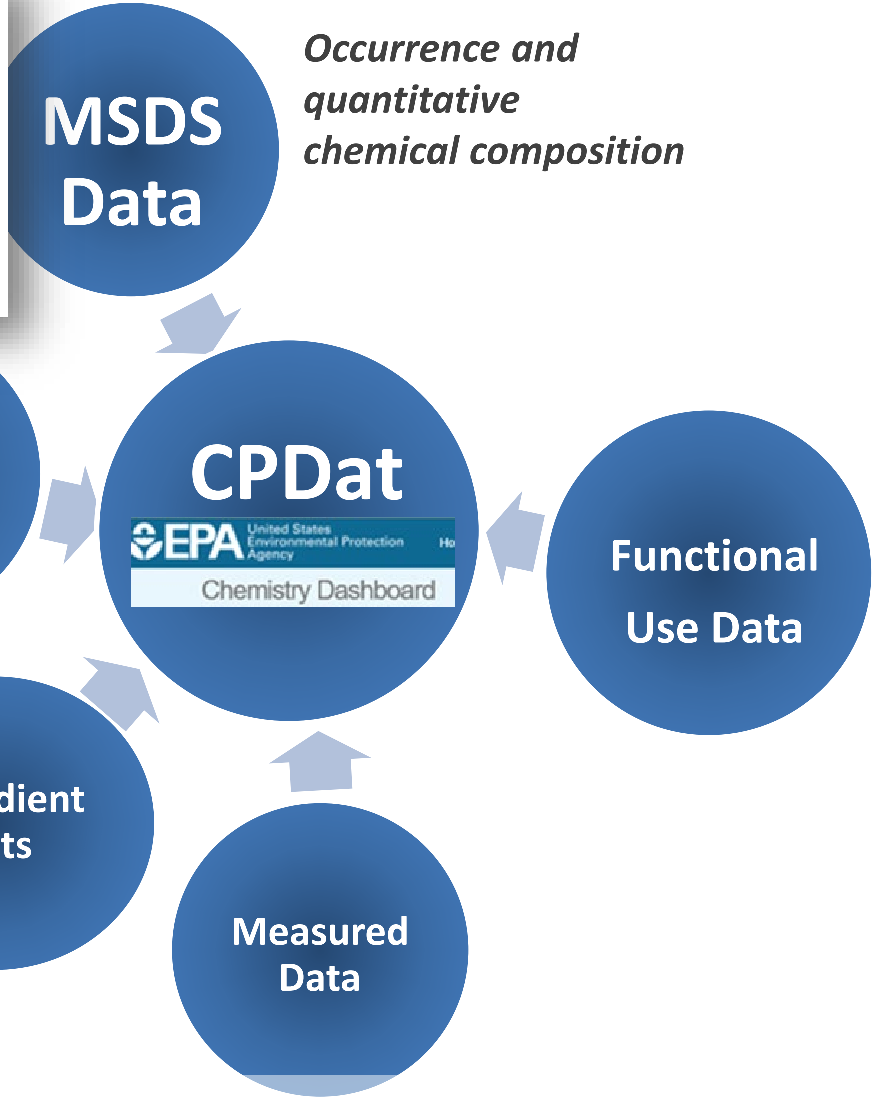
Toxicology Reports

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

ELSEVIER

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],



Slide from Kristin Isaacs

https://comptox.epa.gov/dashboard

How Can we Know Chemical Use?
Chemical Property NAMs


Contents lists available at ScienceDirect

Food and Chemical Toxicology

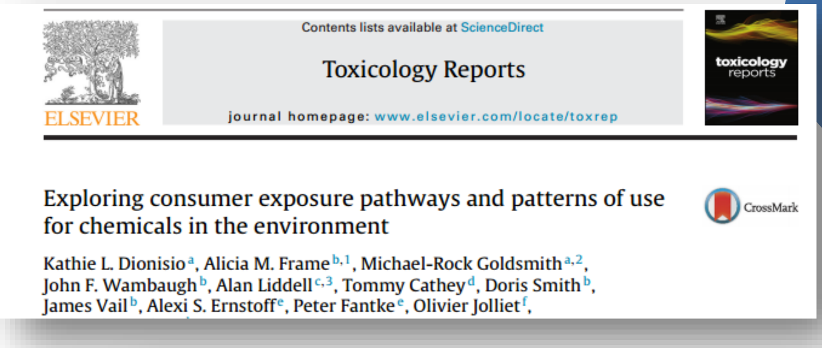
journal homepage: www.elsevier.com/locate/foodchemtox

ELSEVIER

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

ELSEVIER

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

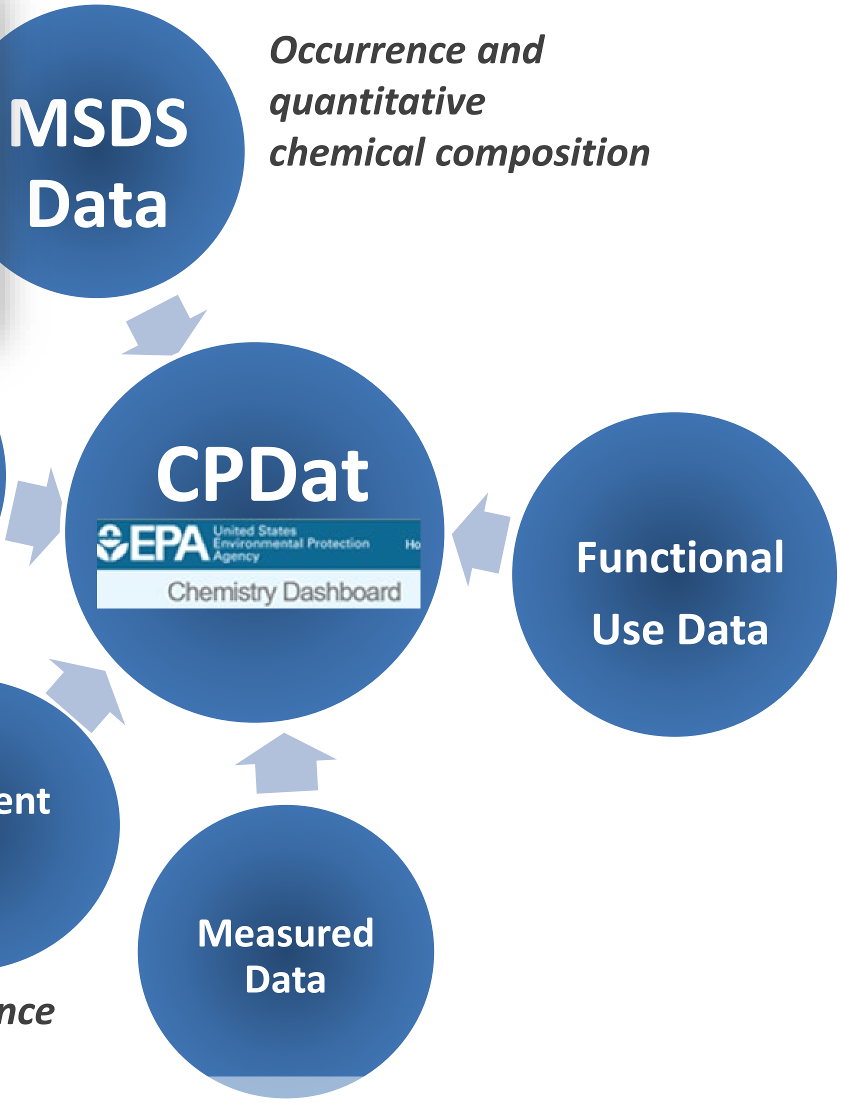
Kathie L. Dionisio[a], Alicia M. Frame[b,1], Michael-Rock Goldsmith[a,2], John F. Wambaugh[b], Alan Liddell[c,3], Tommy Cathey[d], Doris Smith[b], James Vail[b], Alexi S. Ernstoff[e], Peter Fantke[e], Olivier Jolliet[f],


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]



Occurrence and quantitative chemical composition

Occurrence data

Slide from Kristin Isaacs

https://comptox.epa.gov/dashboard

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

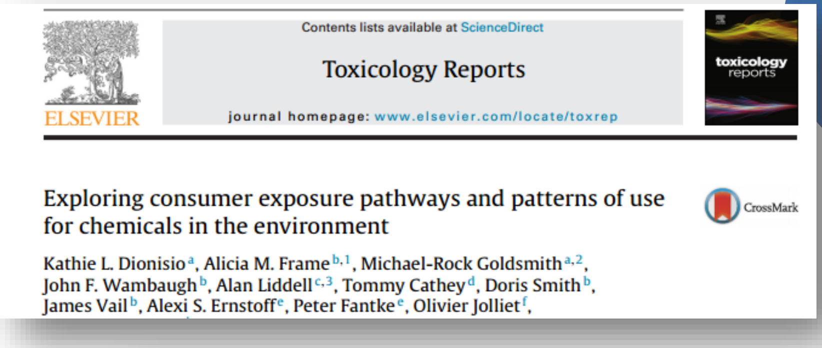
Development of a consumer product ingredient database for chemical exposure screening and prioritization

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

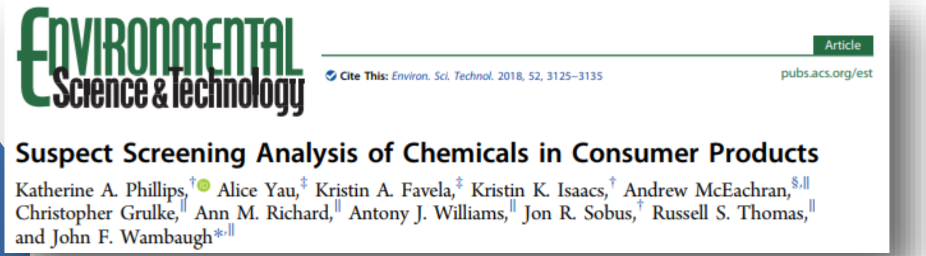
Consumer product chemical weight fractions from ingredient lists

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

Ingredient Lists

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

Slide from Kristin Isaacs

https://comptox.epa.gov/dashboard

United States
Environmental Protection
Agency

How Can we Know Chemical Use?
Chemical Property NAMs

MSDS Data

Occurrence and quantitative chemical composition

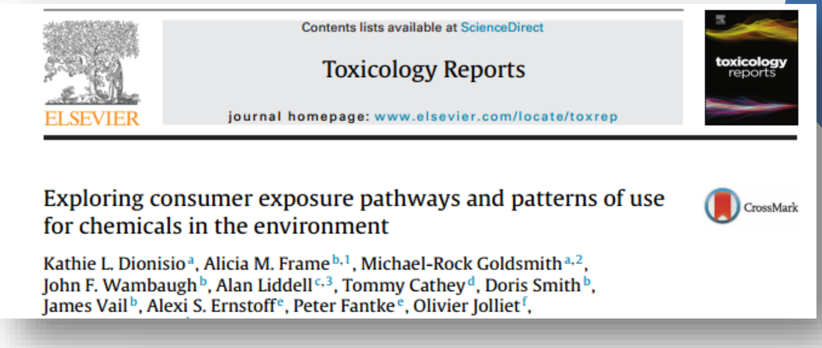

Contents lists available at ScienceDirect
Food and Chemical Toxicology
journal homepage: www.elsevier.com/locate/foodchemtox

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

CPCat


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

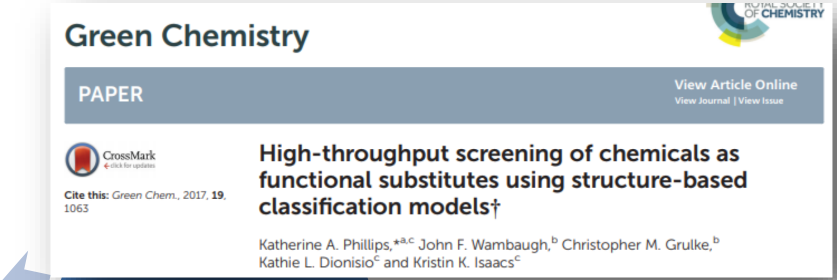
Kathie L. Dionisio[a], Alicia M. Frame[b,1], Michael-Rock Goldsmith[a,2], John F. Wambaugh[b], Alan Liddell[c,3], Tommy Cathey[d], Doris Smith[b], James Vail[b], Alexi S. Ernstoff[e], Peter Fantke[e], Olivier Jolliet[f],

CPDat

Chemistry Dashboard

Functional Use Data

The roles that chemicals serve in products


Green Chemistry

PAPER

View Article Online
View Journal | View Issue

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

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

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

Ingredient Lists

Occurrence data

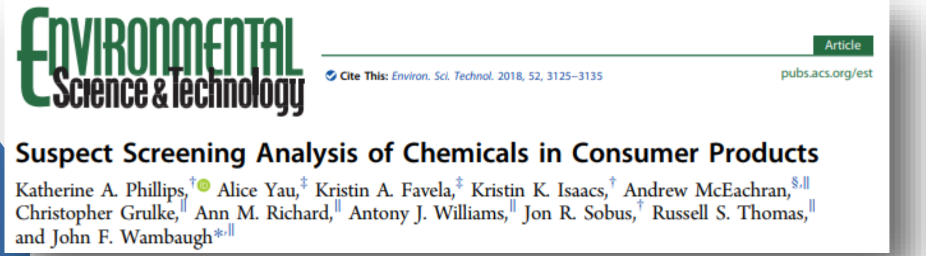

Journal of Exposure Science and Environmental Epidemiology (2018) 28, 216–222
© 2018 Nature America, Inc., part of Springer Nature. All rights reserved 1559-0631/18
www.nature.com/jes

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]

Measured Data

Measurement of chemicals in consumer products


Environmental Science & Technology

Article

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

pubs.acs.org/est

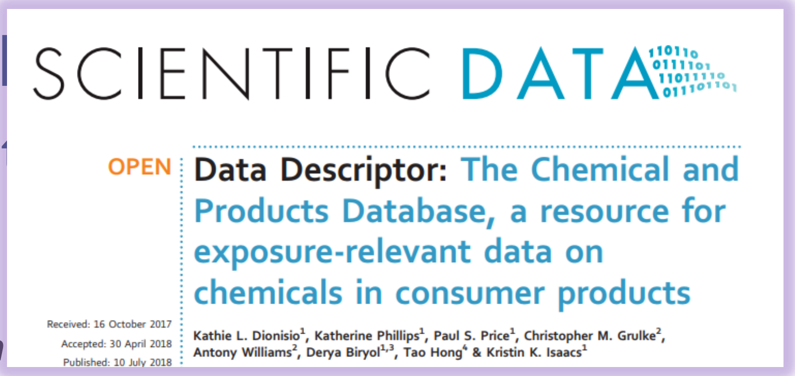
Suspect Screening Analysis of Chemicals in Consumer Products

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[*,∥]

Slide from Kristin Isaacs

https://comptox.epa.gov/dashboard

How Can we Know Cl
Chemical Proper



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]

Contents lists available at ScienceDirect
Food and Chemical Toxicology
journal homepage: www.elsevier.com/locate/foodchemtox
ELSEVIER

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

CrossMark

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

CPCat

Contents lists available at ScienceDirect
Toxicology Reports
journal homepage: www.elsevier.com/locate/toxrep
ELSEVIER

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

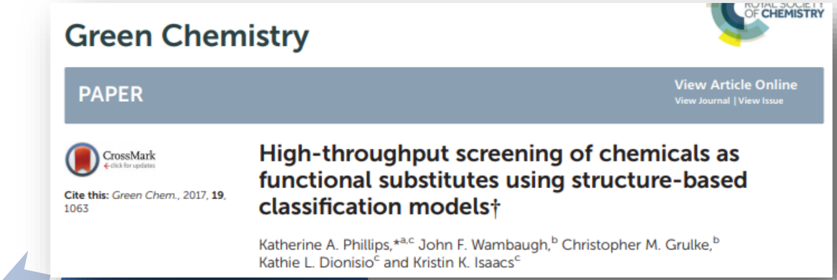
CrossMark

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],

CPDat

EPA United States Environmental Protection Agency

Chemistry Dashboard



Green Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

CrossMark

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

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

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

Ingredient Lists

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

www.nature.com/jes

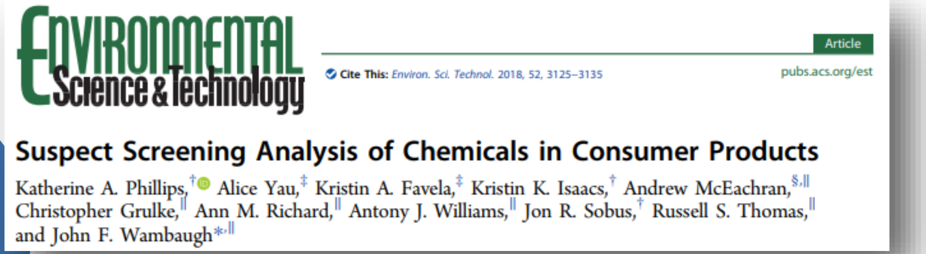
ORIGINAL ARTICLE

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]

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

Slide from Kristin Isaacs

https://comptox.epa.gov/dashboard

United States
Environmental Protection
Agency

How Can we Know Chemical Use?
Chemical Property NAMs

Contents lists available at ScienceDirect

Food and Chemical

ELSEVIER

Development of a
exposure screening

M.-R. Goldsmith[a,*], C.M
R. Edwards[d], D.T. Chang
J. Mitchell[g], D.A. Vallero
T.J. Buckley[a], C.C. Dary

Broad "ind

Contents

Tox

journal homepag

ELSEVIER

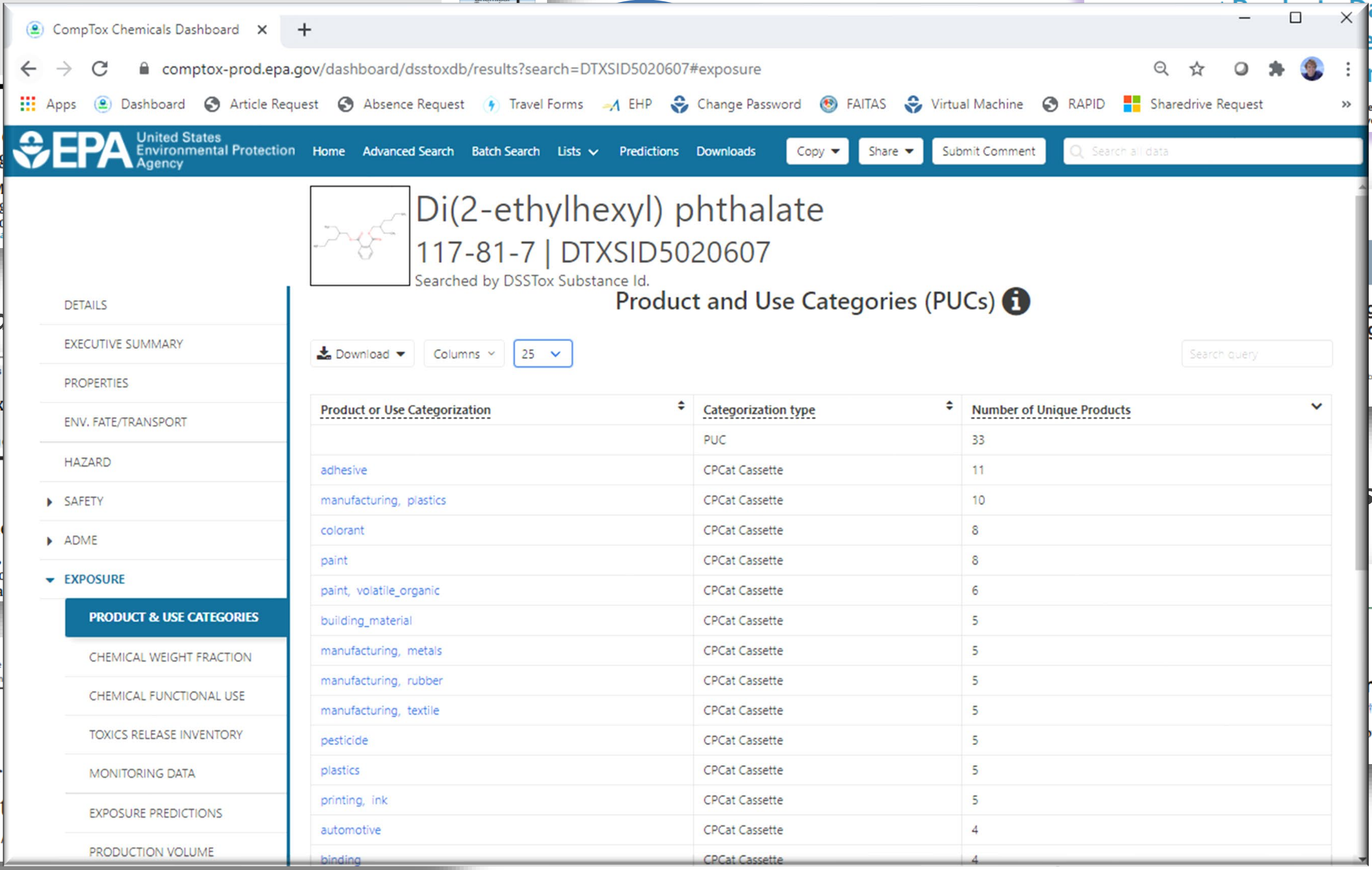
Exploring consumer exposure
for chemicals in the environm

Kathie L. Dionisio[a], Alicia M. Frame[b,1],
John F. Wambaugh[b], Alan Liddell[c,3], To
James Vail[b], Alexi S. Ernstoff[e], Peter Fa

Journal of Exposure Science
© 2018 Nature America, In

ORIGINAL ARTICLE
Consumer pr
ingredient list

Kristin K. Isaacs[1], Katherine



Di(2-ethylhexyl) phthalate

117-81-7 | DTXSID5020607

Searched by DSSTox Substance Id.

Product and Use Categories (PUCs)

Product or Use Categorization	Categorization type	Number of Unique Products
	PUC	33
adhesive	CPCat Cassette	11
manufacturing, plastics	CPCat Cassette	10
colorant	CPCat Cassette	8
paint	CPCat Cassette	8
paint, volatile_organic	CPCat Cassette	6
building_material	CPCat Cassette	5
manufacturing, metals	CPCat Cassette	5
manufacturing, rubber	CPCat Cassette	5
manufacturing, textile	CPCat Cassette	5
pesticide	CPCat Cassette	5
plastics	CPCat Cassette	5
printing, ink	CPCat Cassette	5
automotive	CPCat Cassette	4
binding	CPCat Cassette	4

DATA

The Chemical and
atabase, a resource for
elevant data on
n consumer products

e Phillips[1], Paul S. Price[1], Christopher M. Grulke[2],
ol[1,3], Tao Hong[4] & Kristin K. Isaacs[1]

ROYAL SOCIETY OF CHEMISTRY

View Article Online
View Journal | View Issue

g of chemicals as
g structure-based

Christopher M. Grulke,[b]

that
serve in

Article

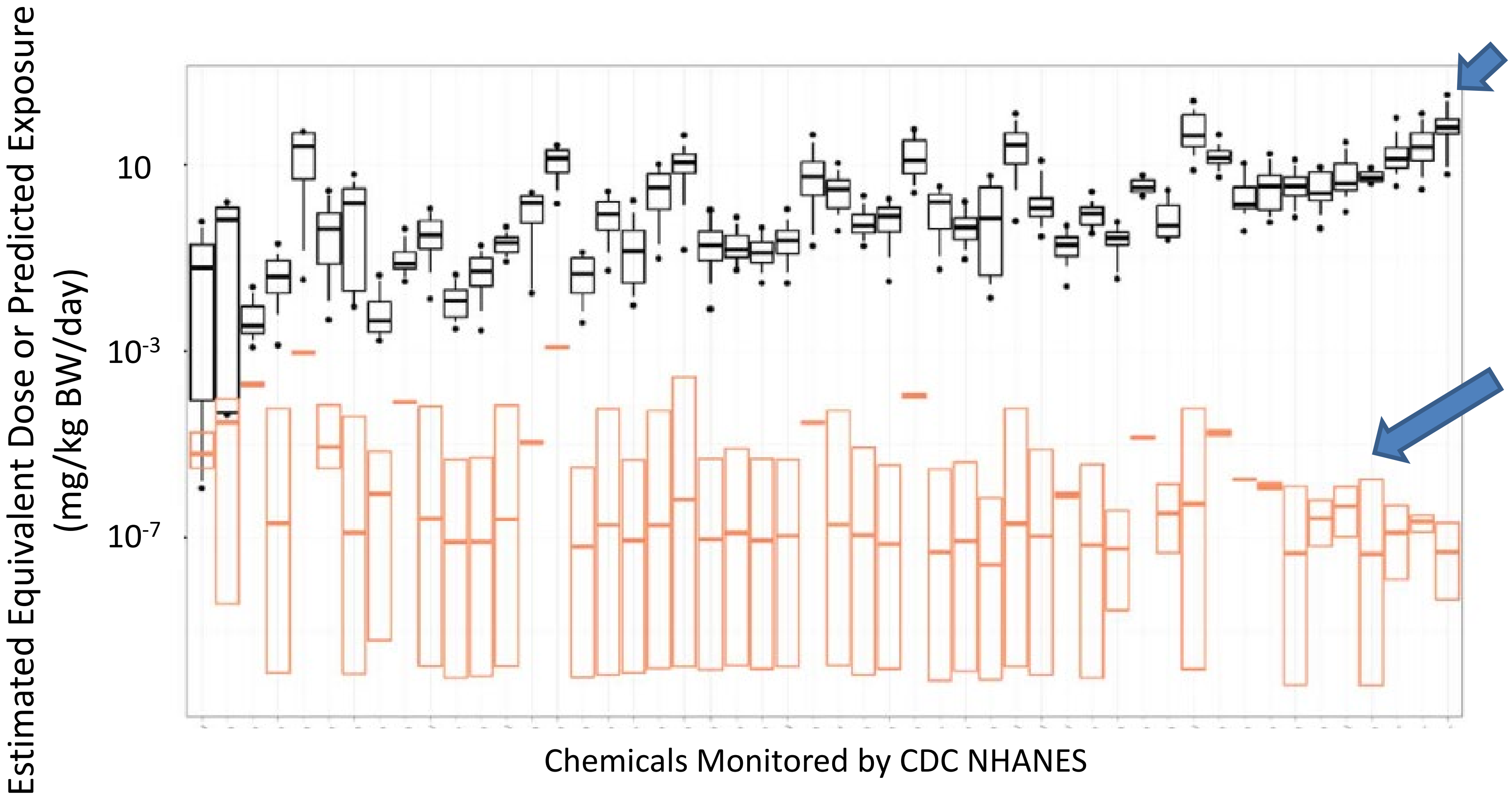
pubs.acs.org/est

nsumer Products
Andrew McEachran,[§,‖]
us,[†] Russell S. Thomas,[‖]

Slide from Kristin Isaacs

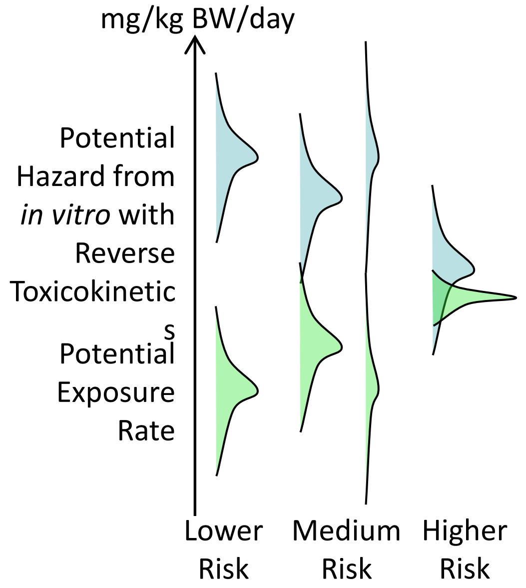
https://comptox.epa.gov/dashboard

Chemical Prioritization NAMs



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

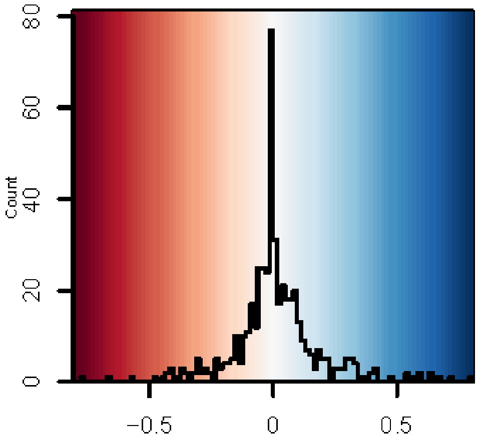
Exposure intake rates can be inferred from biomarkers (for example, Ring et al., 2018)



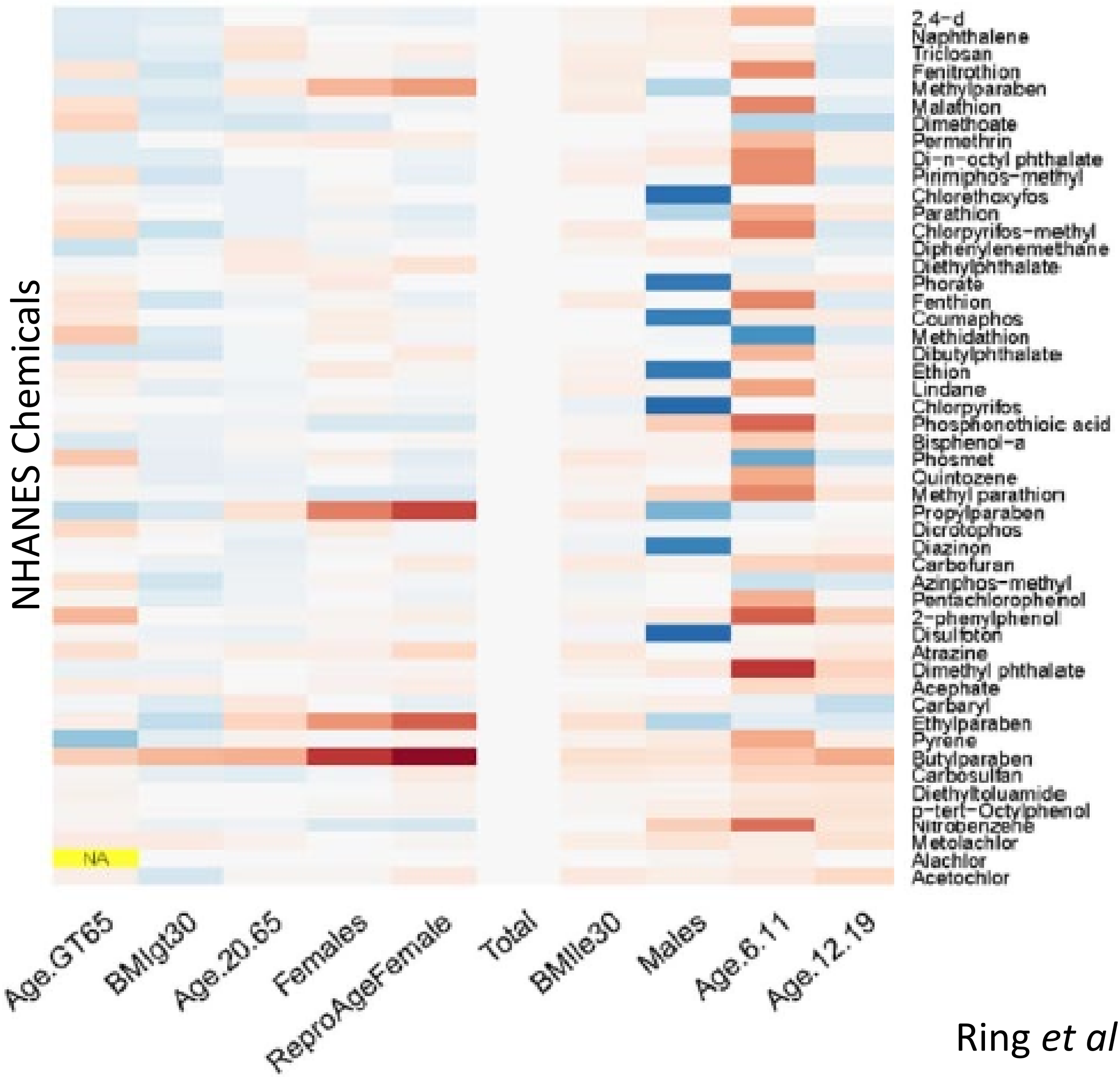
Ring *et al*. (2017)

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



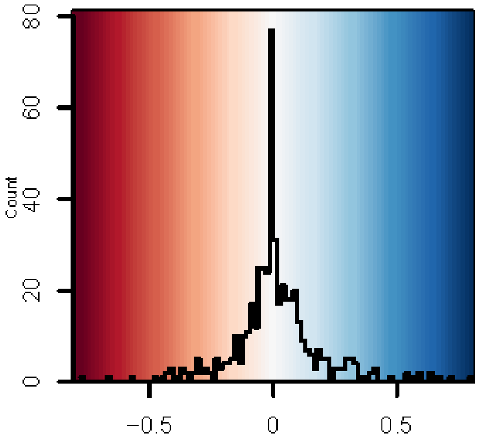
Change in Exposure Relative to Total Population



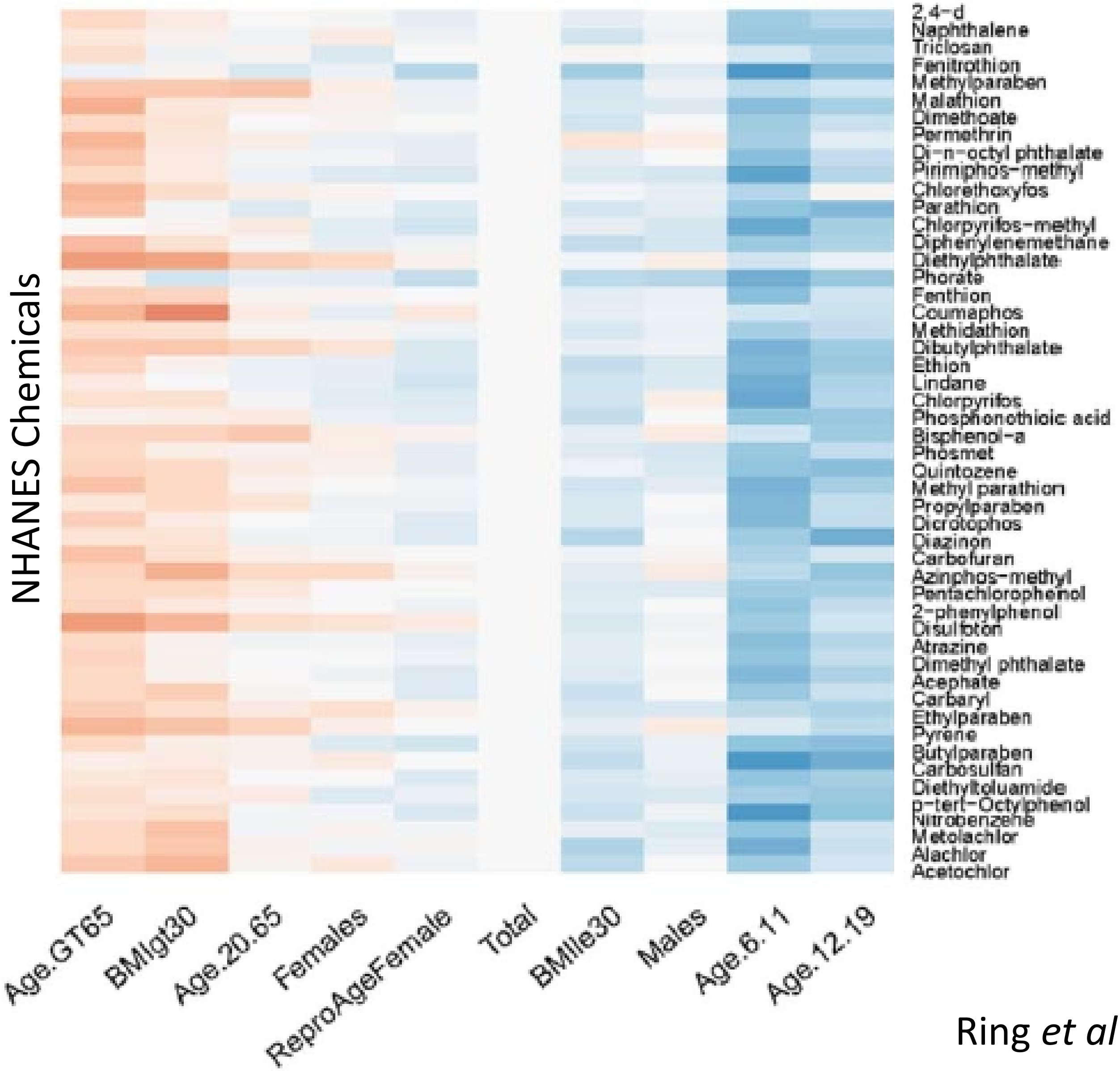
Ring *et al.* (2017)

Life-stage and Demographic Variation in TK

- Ring *et al.* (2017) predicted change in plasma concentrations for a 1 mg/kg bw/day exposure for various demographic groups



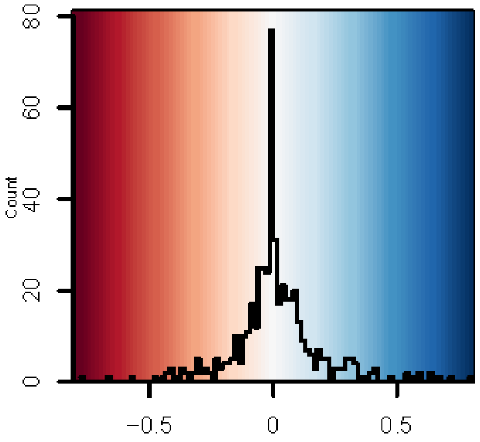
Change in Toxicokinetics Relative to Total Population



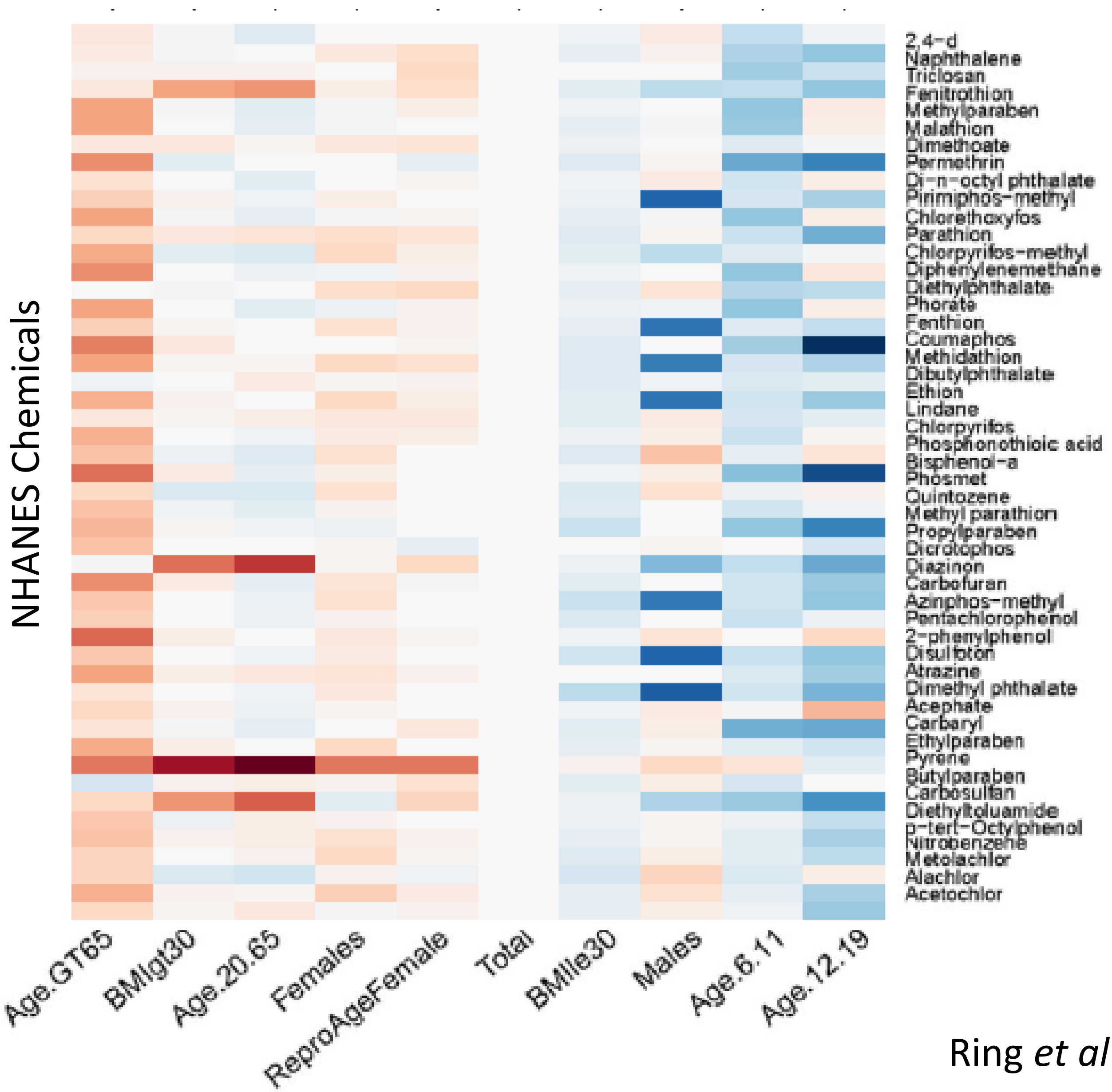
Ring *et al*. (2017)

Life-stage and Demographic Variation in Risk Priority

- Using demographic-specific toxicokinetics and exposure, we can calculate margin between bioactivity and exposure for various demographic groups



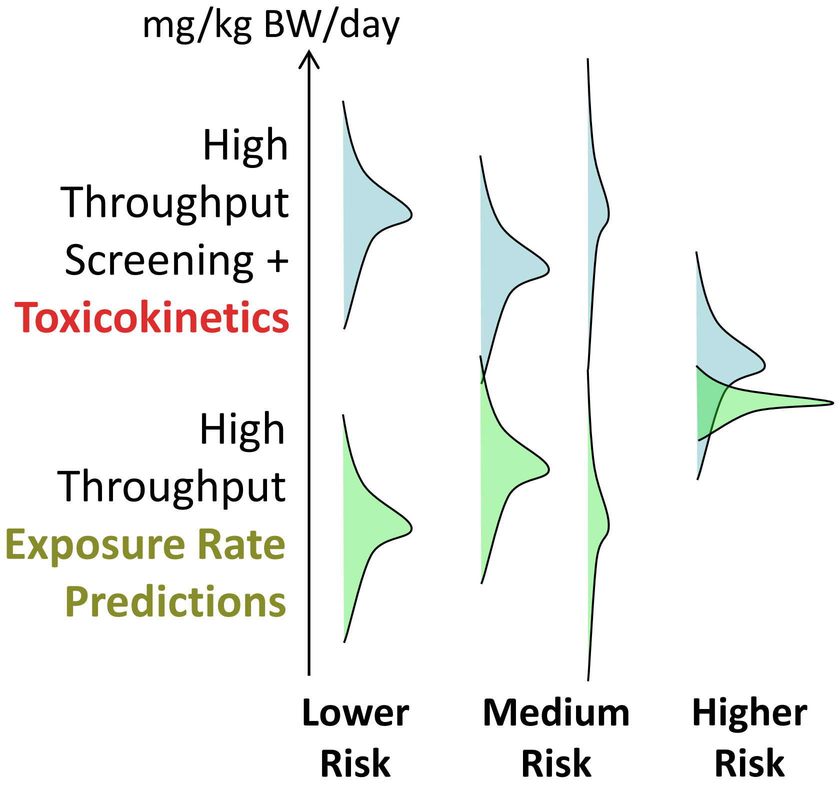
Change in Risk Relative to Total Population



Ring *et al*. (2017)

Summary

- We need to know chemical hazard, exposure, and toxicokinetics to assess risk posed to the public health
- There are tens of thousands of chemicals in commerce in the environment that lack some of these data
- 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/dashboard
  - R package "httk": https://CRAN.R-project.org/package=httk



ExpoCast Project (Exposure Forecasting)

Center for Computational Toxicology and Exposure

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

Louis Groff*
Chris Grulke
Colin Guider*
Mike Hughes
Victoria Hull*
Kristin Isaacs
Richard Judson
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Seth Newton
Katherine Phillips
Paul Price
Tom Purucker

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

CEMM

Hongwan Li*
Xiaoyu Liu
Zachary Robbins*
Mark Strynar

Collaborators

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

*Trainees

United States
Environmental Protection
Agency

US EPA Office of Research and Development

- The Office of Research and Development (ORD) is the scientific research arm of EPA
  - 543 peer-reviewed journal articles in 2019
- 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

1798


Credit: the Research Triangle Foundati

ORD Facility in
Research Triangle Park, NC

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), ORSFl1-ORSFl11.

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.