

The Physics of Kevin Bacon: Complex Systems and Environmental Science

John F. Wambaugh

*National Center for Computational Toxicology
Office of Research and Development
United States Environmental Protection Agency
Research Triangle Park, North Carolina 27711*

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>

Wake Technical Community College
February 27, 2019

EPA Office of Research and Development

- The Office of Research and Development (ORD) is the scientific research arm of EPA
 - 626 peer-reviewed journal articles in 2017 and 456 so far in 2018
- Research is conducted by ORD's three national laboratories, four national centers, and two offices organized to address:
 - Hazard, exposure, risk assessment, and risk management
- 13 facilities across the United States
- Research conducted by a combination of Federal scientists; contract researchers; and postdoctoral, graduate student, and post-baccalaureate trainees



ORD Facility in
Research Triangle Park, NC

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 Borassi¹, Pierluigi Crescenzi², Michel Habib³,
Walter Koster⁴, Andrea Marino^{5,*}, and Frank Takes⁴

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

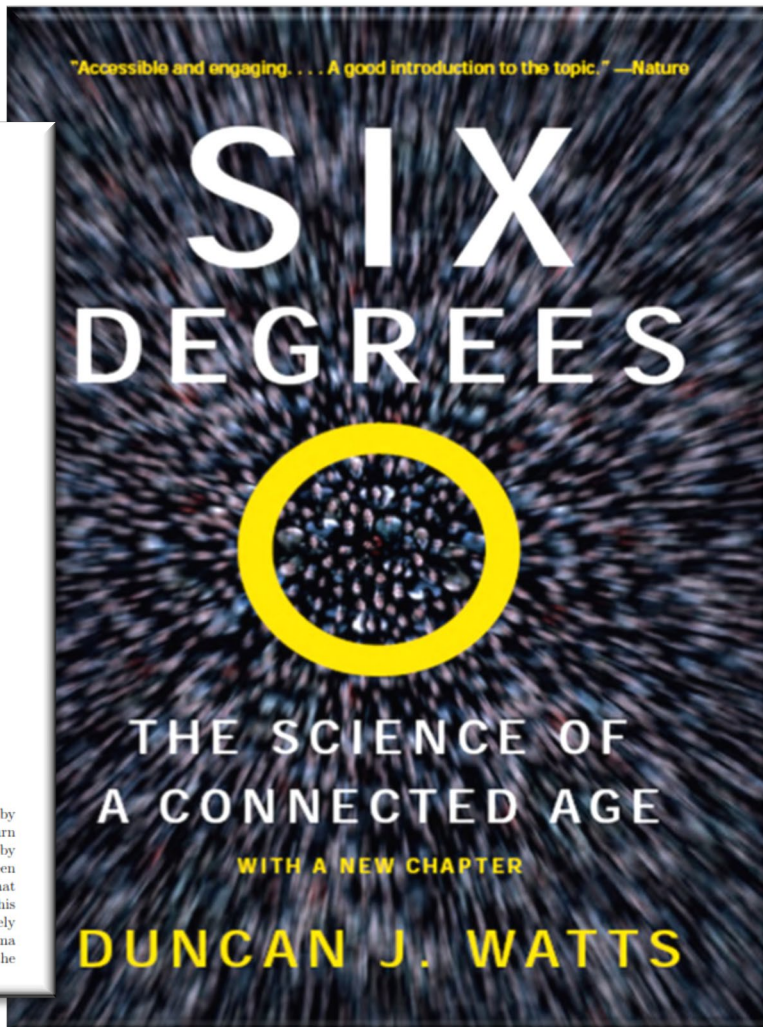
⁵ Dipartimento di Informatica, Università di Milano, Italy

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 $O(|V| \cdot |E|)$, 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 $O(|E|)$. 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 Karinty 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).



kins

Kevin Bacon and Graph Theory

KEVIN BACON AND GRAPH THEORY

Brian Hopkins

ADDRESS: 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.

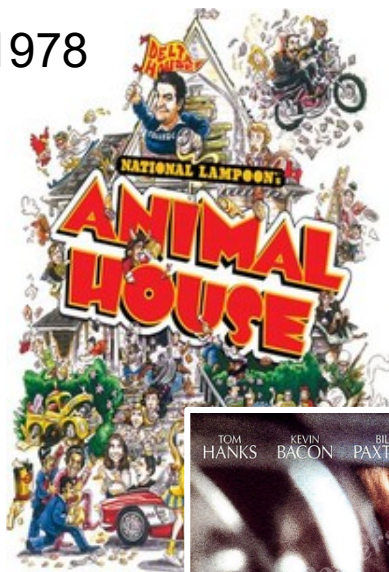
YWORDS: Cinema, finite mathematics, graph theory, popular culture, six degrees of separation, weighted averages.

1 INTRODUCTION

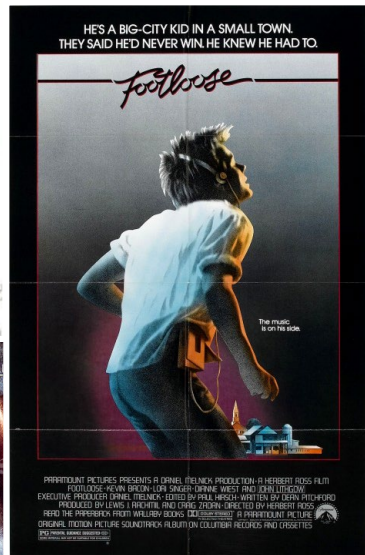
Graph theory is the mathematics of connections. It has wide applications to many interconnected systems: transportation networks, epidemiology, and the Internet, to name just a few. But we teach graph theory with pictures of handfuls of dots and lines. There is one large system that is easy to work with,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

Kevin Bacon

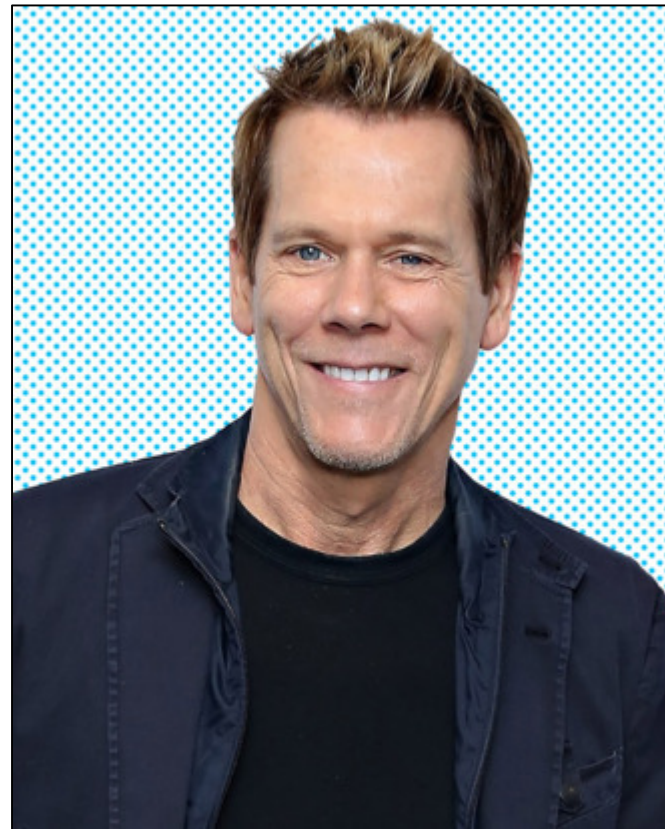
1978



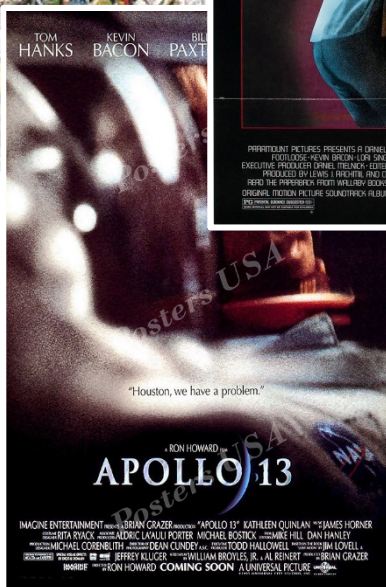
1984



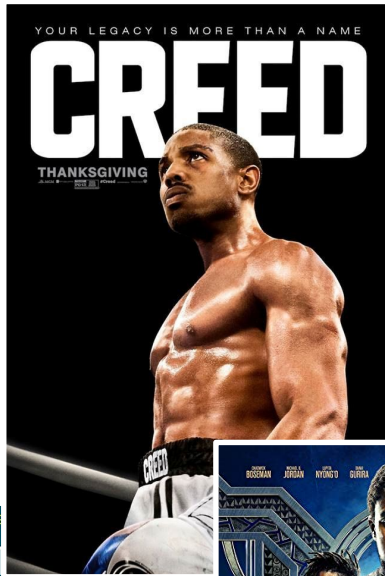
1992



1995



Michael B. Jordan

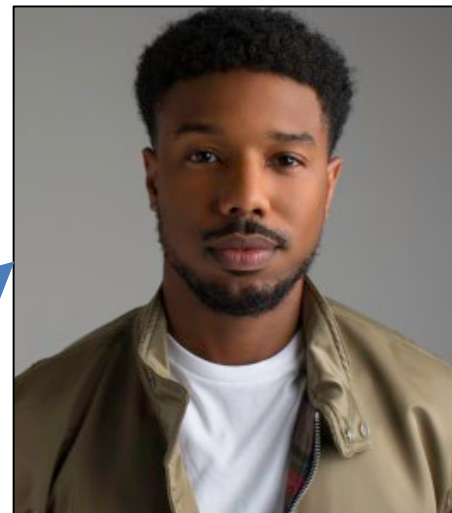


Connectedness to Michael B. Jordan

Hail Caesar
McDormand &
Channing Tatum

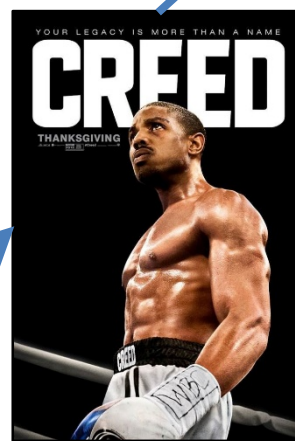


GI Joe: Retaliation
Tatum & Bruce Willis



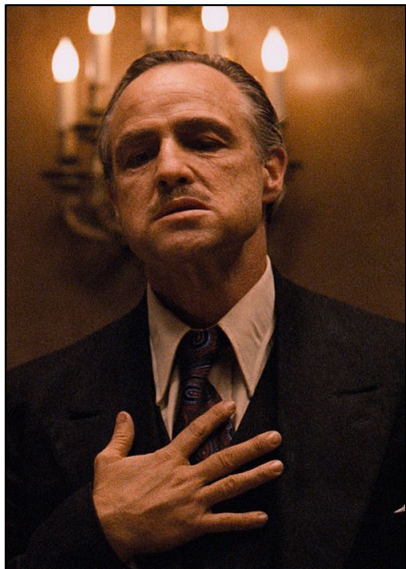
Frances McDormand
Best Actress Winner 2018

Expendables
Willis &
Sylvester Stallone



Creed
Stallone & Jordan

Connectedness to Michael B. Jordan

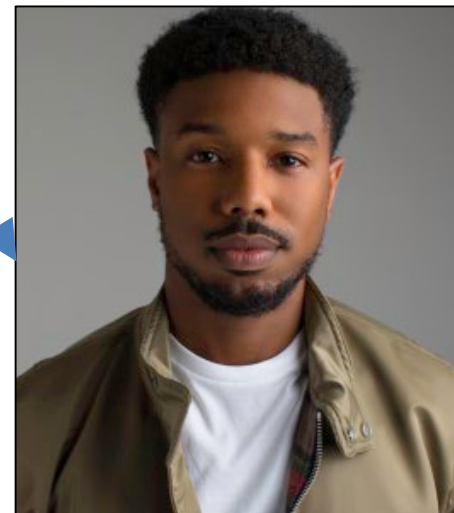


Marlon Brando
Best Actor 1954 and 1972
Died 2004

**Avengers:
Infinity War**
Paltrow &
Chadwick
Boseman



Black Panther
Boseman & Jordan



Superman
with Gene Hackman



The Royal Tenenbaums
Hackman & Gwyneth Paltrow

Small World Networks

Travers and
Milgram (1969):

Collins and Chow (1998)

Watts and Strogatz (1998)

news and views

letters to nature

typical slower than $\sim 1 \text{ km s}^{-1}$) might differ significantly from what is assumed by current modelling efforts². 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 mitigation³ through disruption and deflection, or for resource exploitation⁴. Such predictions would require detailed reconnaissance concerning the composition and internal structure of the targeted object. □

Received 4 February; accepted 18 March 1998.

1. Asphang, E. & Melosh, H. J. The Sticky impact of Phobos: A dynamical model. *Lunar Init.* **100**, 154–164 (1996).
2. Asphang, E. et al. Mechanical and geological effects of impact cratering on Ida. *Lunar Init.* **120**, 158–184 (1996).
3. Stone, M. C., Asphang, E., Melosh, H. J. & Greenberg, R. Impact craters on asteroids: Does strength govern gravity control the size? *Lunar Init.* **120**, 399–371 (1996).
4. Love, S. B. & Ahrens, T. J. Catastrophic impacts on gravity dominated asteroids. *Lunar Init.* **120**, 141–152 (1996).
5. Melosh, H. J. & Ryan, E. V. Asteroids: Shattered but not dispersed. *Lunar Init.* **120**, 263–264 (1997).
6. Housen, K. R., Schmidt, R. M. & Holappke, K. A. Crater crater scaling laws: Fundamental forms based on dimensional analysis. *J. Geophys. Res.* **88**, 2485–2499 (1983).
7. Holappke, K. A. & Schmidt, R. M. Prior state solutions and geologic predictions in cratering mechanics. *J. Geophys. Res.* **94**, 6359–6379 (1989).
8. Housen, K. R. & Holappke, K. A. On the fragmentation of comets and planetary satellites. *Lunar Init.* **84**, 226–233 (1990).
9. Benz, W. & Asphang, E. Simulations of brittle solids using smooth particle hydrodynamics. *Comput. Phys. Commun.* **87**, 233–260 (1993).
10. Asphang, E. et al. Mechanical and geological effects of impact cratering on Ida. *Lunar Init.* **120**, 158–184 (1996).
11. Hudson, R. S. & Oates, S. J. Shape of asteroid 4960 Camilla (1989 PB) from inversion of radar images. *Science* **265**, 948–951 (1994).
12. Oates, S. J. et al. Asteroid radar geometry. *Astroph. J.* **382**, 1480–1502 (1991).
13. Ahrens, T. J. & O'Keefe, D. In: *Impact and Explosion Cratering* (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 649–656 (Tergreen, New York, 1977).
14. Tikhonov, I. H. Metric signatures of state for hypervelocity impact. *General Atomic Report GA-5216*, San Diego, 1982.
15. Nakamura, A. R. Fujiwara, A. Velocity distribution of fragments formed in a simulated collisional disruption. *Lunar Init.* **92**, 132–146 (1991).
16. Benz, W. & Asphang, E. Simulations of brittle solids using smooth particle hydrodynamics. *Comput. Phys. Commun.* **87**, 233–260 (1993).
17. Benoit, W. E., Nolan, M. C., Greenberg, R. & Kolodziej, R. A. Velocity distributions among colliding asteroids. *Lunar Init.* **120**, 255–268 (1996).
18. Rubin, M. J. et al. Galileo encounter with 951 Gaspra—First pictures of an asteroid. *Science* **277**, 1647–1652 (1997).
19. Rubin, M. J. et al. Galileo's encounter with 243 Ida: An overview of the imaging experiment. *Lunar Init.* **120**, 1–31 (1996).
20. Asphang, E. & Melosh, H. J. The Sticky impact of Phobos: A dynamical model. *Lunar Init.* **100**, 154–164 (1996).
21. Asphang, E. et al. Mechanical and geological effects of impact cratering on Ida. *Lunar Init.* **120**, 158–184 (1996).
22. Housen, K. R., Schmidt, R. M. & Holappke, K. A. Crater crater scaling laws: Fundamental forms based on dimensional analysis. *J. Geophys. Res.* **88**, 2485–2499 (1983).
23. Vornicki, L. et al. 3D-ME: Theory of 3D-ME: Impact of C-asteroid. *Science* **278**, 2109–2112 (1997).
24. Asphang, E. et al. Impact evolution of elongated, irregular bodies. *Lunar Planet. Sci. Conf. (Abstr.)* **XXVIII**, 63–64 (1987).
25. Love, S. G., Hiltz, E. & Bowdler, D. E. Target porosity effects in impact cratering and collisional disruption. *Lunar Init.* **120**, 248–255 (1996).
26. Fujiwara, A., Cornell, P., Davis, D. R., Ryan, E. V. & DDMartin, 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 elongated, irregular bodies. *Lunar Init.* **120**, 96–100 (1997).
28. Brown, T. J. & Harris, A. W. Deflection and fragmentation of near-Earth asteroids. *Nature* **380**, 429–430 (1992).
29. *Science of Near Earth Space* (eds Lewis, J. S., Matthews, M. S. & Guertler, M. L.) (Univ. Arizona Press, Tucson, 1995).

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: asphang@rockwell.com).

440

Nature © Macmillan Publishers Ltd 1998

NATURE VOL 393 | JUNE 1998

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 oscillators¹, Josephson junction arrays², excitable media³, neural networks^{4–7}, spatial games⁸, genetic control networks⁹ 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^{10,11} (popularly known as six degrees of separation¹²). The small-world 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 $0 < p < 1$, to 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 connected¹³. In this regime, we find that $L \sim n/2k \gg 1$ and $C \sim 3/4$ as $p \rightarrow 0$, while $L \sim \ln(n)/\ln(k)$ and $C \sim C_{\text{random}} = 4/k$ 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 $C(p) \gg C_{\text{random}}$. 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

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

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 Erdős number, which measures the number of links needed to connect one to the prolific mathematician Paul Erdős through jointly authored papers. For example, individuals have an Erdős number of 1 if they co-authored a paper with Erdős. If one of their co-authors wrote a paper with Erdős, then they have an Erdős number of 2, and so forth. It has been pointed out¹ that Dan Kleiman has a combined Erdős/Bacon number of 3 because he wrote a paper with Erdős 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 Separation², 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 phenomenon³, arises from pioneering empirical work by Milgram⁴ 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 issue⁵, 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

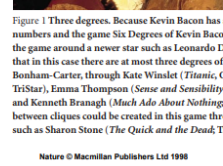


Figure 1 Three degrees. Because Kevin Bacon has appeared in many films, most actors have low Bacon numbers and the game Six Degrees of Kevin Bacon has declined in popularity. It is possible to centre that in this case there are at most three degrees of separation between DiCaprio and Helena Bonham-Carter, 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).

length is short, scaling logarithmically with the size of the network.

What Watts and Strogatz⁵ 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).

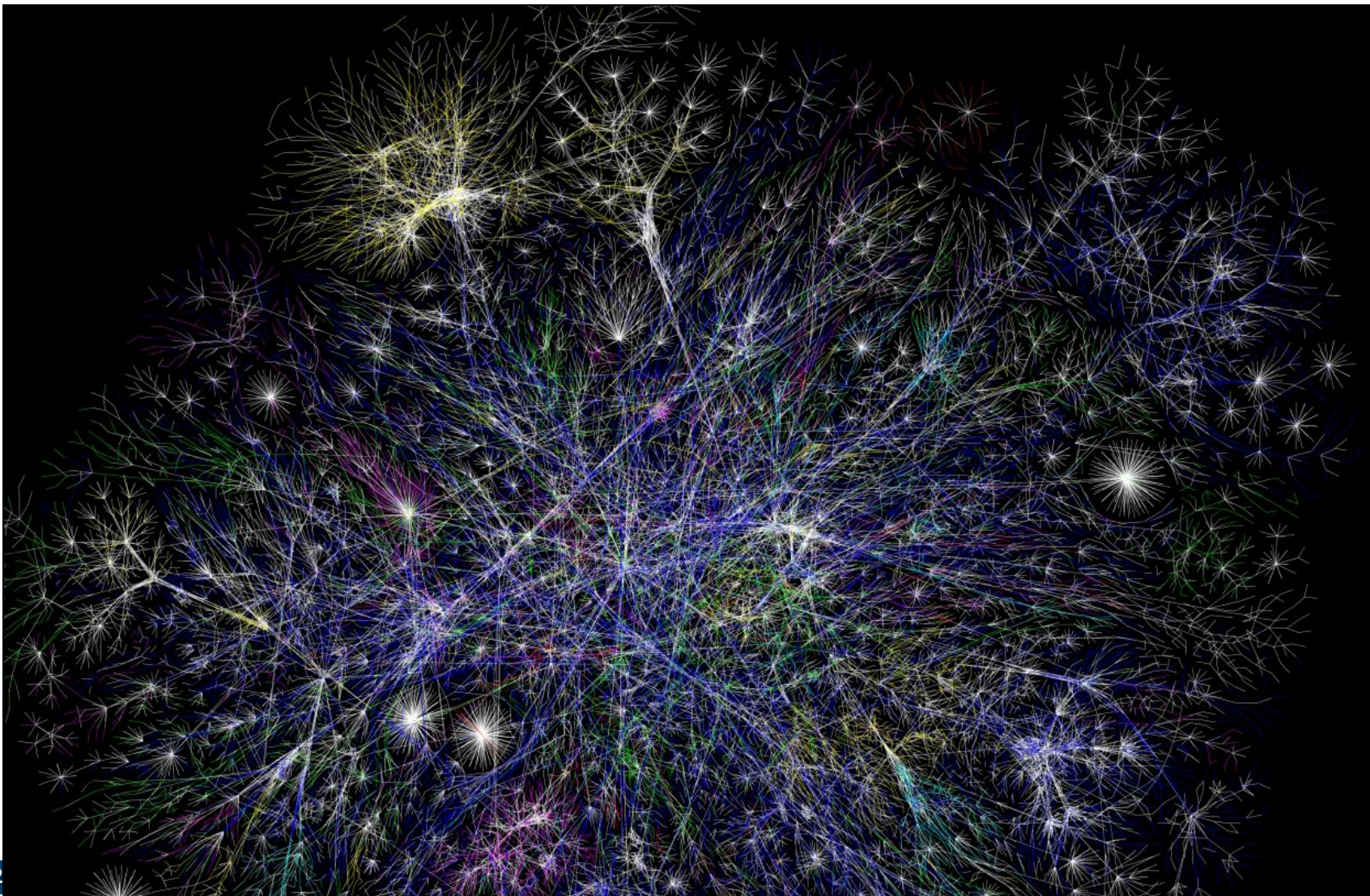
Small-world connectivity has consequences that could be good or bad,

ARTIST REWIND COLLECTION

Nature © Macmillan Publishers Ltd 1998

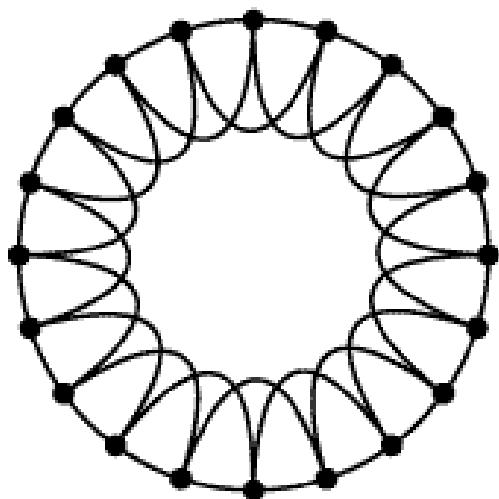
409

The Internet (Circa 2005)

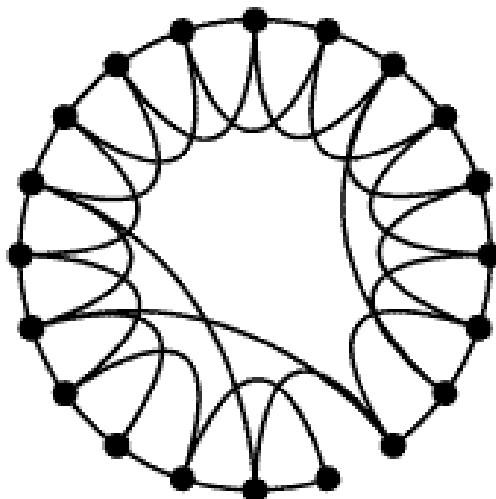


Complex is Not the Same as Random

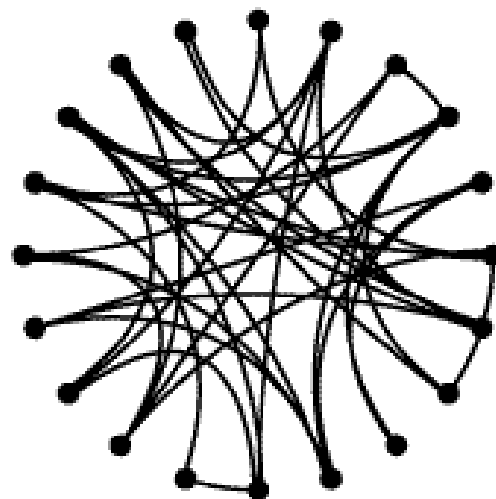
Regular



Small-world



Random



$p = 0$

Increasing randomness

$p = 1$

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)
- Different testing requirements exist for food additives, pharmaceuticals, and pesticide active ingredients (NRC, 2007)



November 29, 2014

Chemical Regulation in the United States

- Most other chemicals, ranging from industrial waste to dyes to packing materials, are covered by the Toxic Substances Control Act (TSCA)
 - Thousands of chemicals on the market were either “grandfathered” in or were allowed without experimental assessment of hazard, toxicokinetics, or exposure
 - Thousands of new chemical use submissions are made to the EPA every year
- TSCA was updated in June, 2016 to allow evaluation of these and other chemicals
 - Methods are being developed to prioritize these existing and new chemicals for testing



November 29, 2014

Chemical Regulation in the United States

“Of the thousands of chemicals listed for commercial use in the United States, EPA has used its authority to limit or ban five chemicals since TSCA was enacted.”

TSCA was updated in June, 2016 to allow more rapid evaluation of chemicals

- Methods are being developed to prioritize these existing and new chemicals for testing

GAO

United States Government Accountability Office

Report to Congressional Requesters

March 2013

TOXIC SUBSTANCES

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



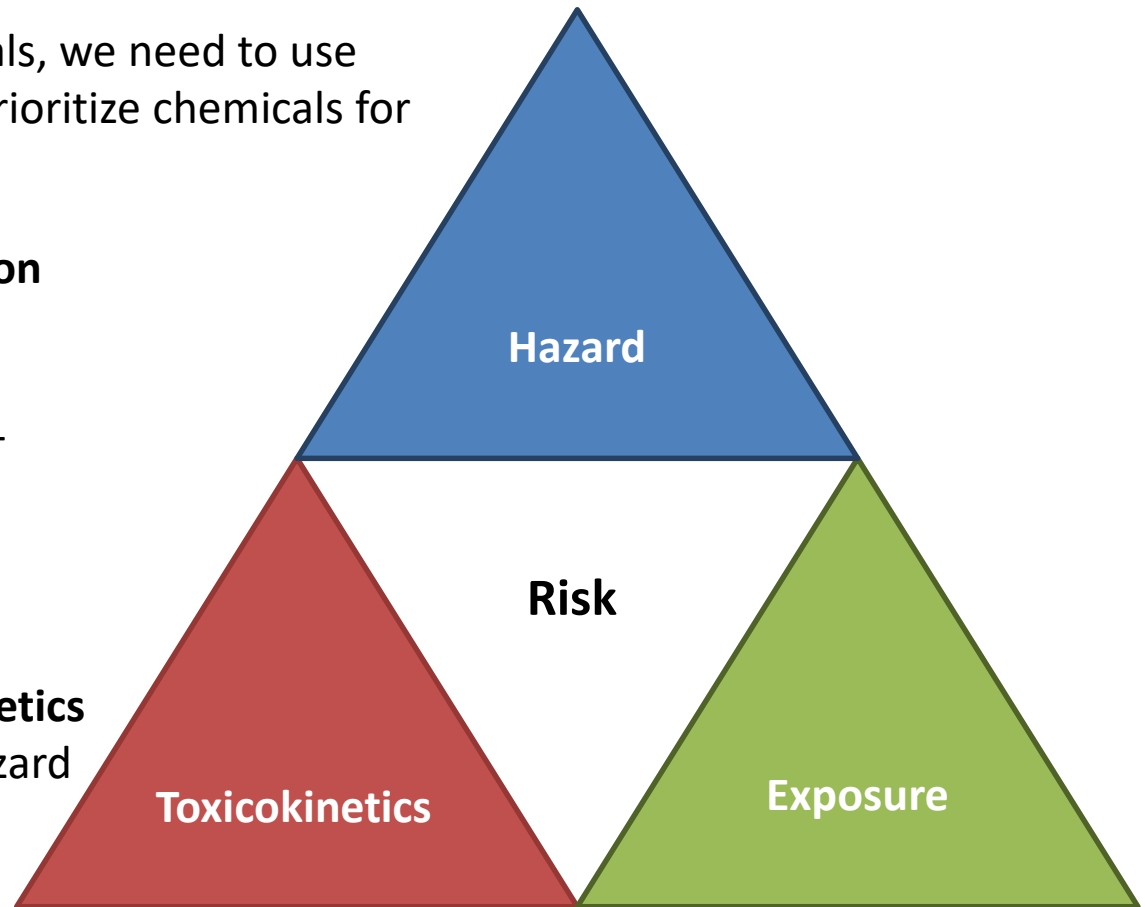
G A O

Accountability • Integrity • Reliability

GAO-13-249

Chemical Risk = Hazard x Exposure

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

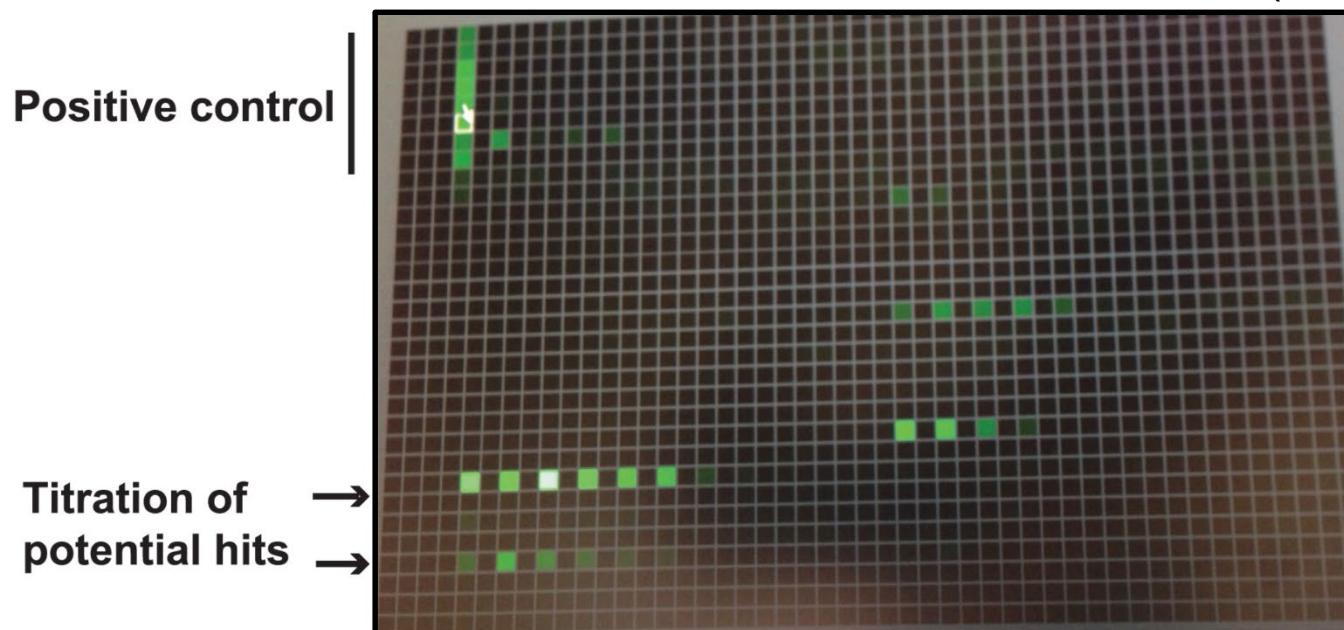


High-throughput Screening for Chemical Hazard

Hertzberg and Pope (2000):

- “New technologies in high-throughput screening have significantly increased throughput and reduced assay volumes”
- “Key advances over the past few years include new fluorescence methods, detection platforms and liquid-handling technologies.”

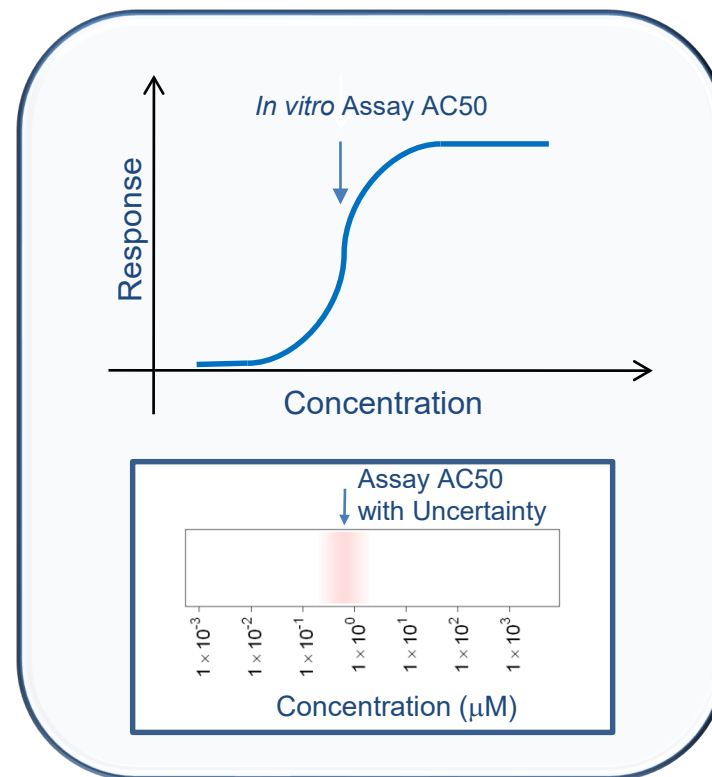
Kaewkhaw et al. (2016)



High-Throughput Testing of Chemicals

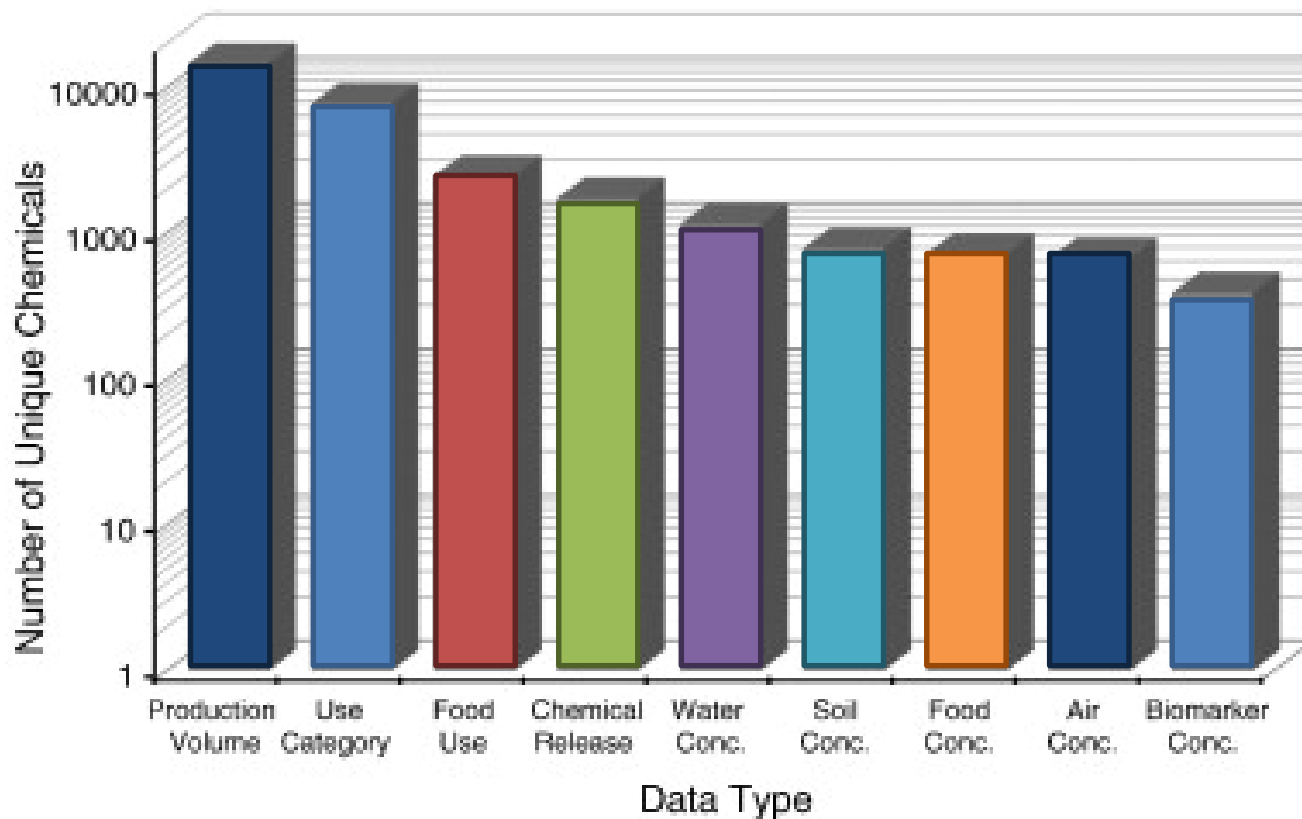


- We might estimate points of departure *in vitro* using high throughput screening (HTS)
- **Tox21:** Examining >8,000 chemicals using ~50 assays intended to identify interactions with biological pathways (Schmidt, 2009)
- **ToxCast:** For a subset (>2000) of Tox21 chemicals ran >1100 additional assays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

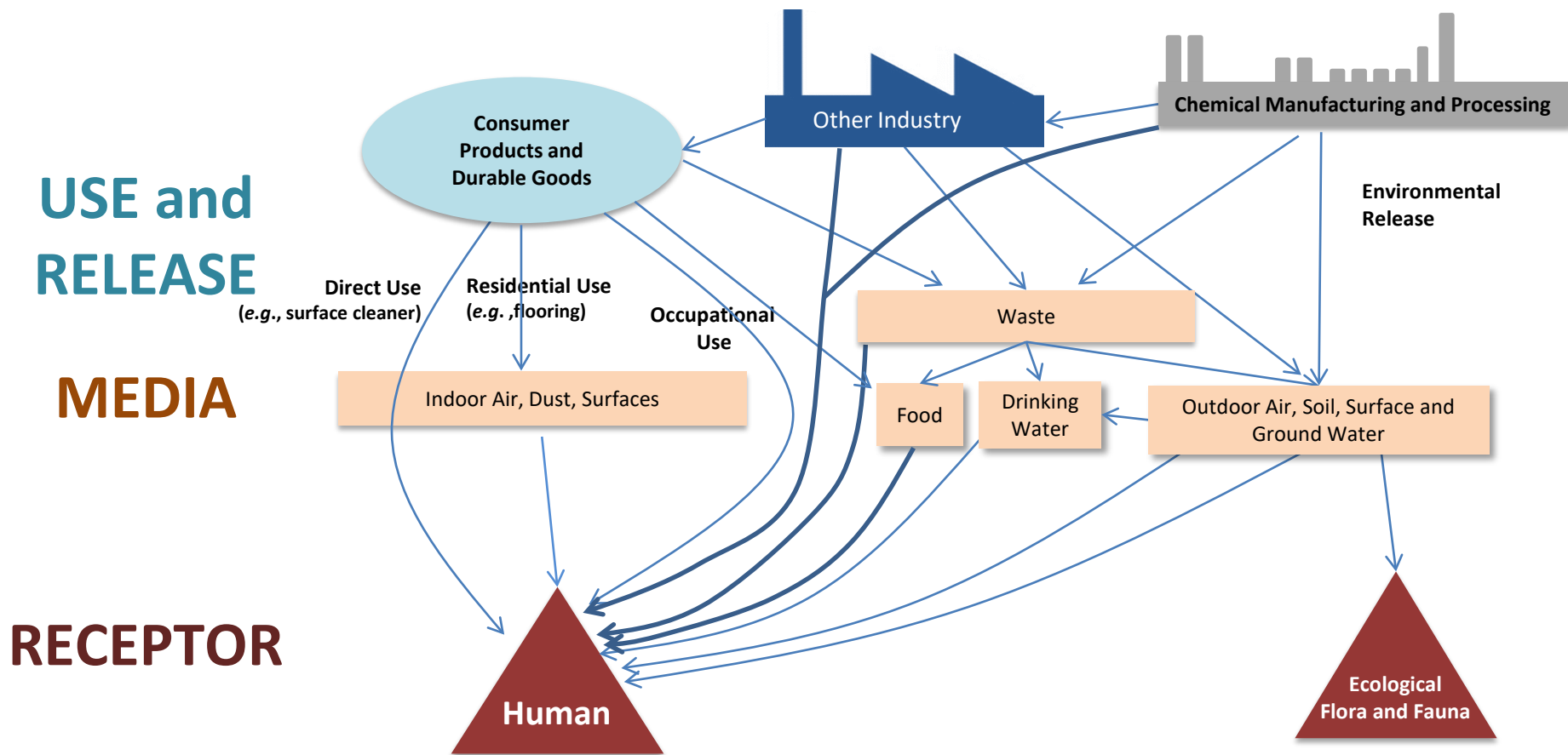


Limited Available Data for Exposure Estimation

Most chemicals lack public exposure-related data beyond production volume (Egeghy et al., 2012)



Exposure is a Complex Problem



Exposure is a Complex Problem

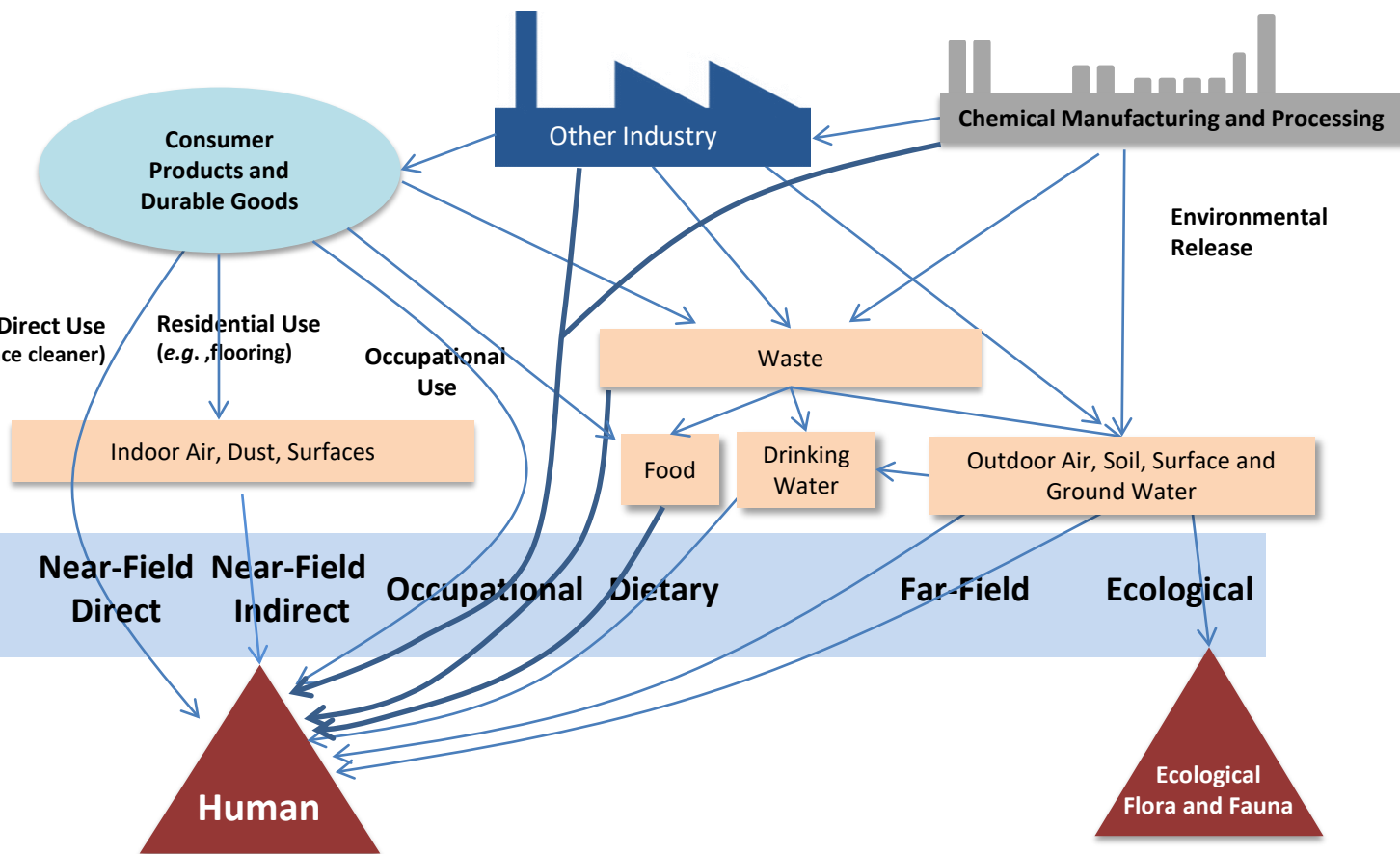
**USE and
RELEASE**

MEDIA

EXPOSURE

(MEDIA + RECEPTOR)

RECEPTOR



The Exposure Event is Often Unobservable



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

What Do We Know About Exposure?

Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey (NHANES) provides an important tool for monitoring public health

Large, ongoing CDC survey of US population: demographic, body measures, medical exam, biomonitoring (health and exposure), ...

Designed to be representative of US population according to census data

Data sets [publicly available](http://www.cdc.gov/nchs/nhanes.htm) (<http://www.cdc.gov/nchs/nhanes.htm>)

Includes measurements of:

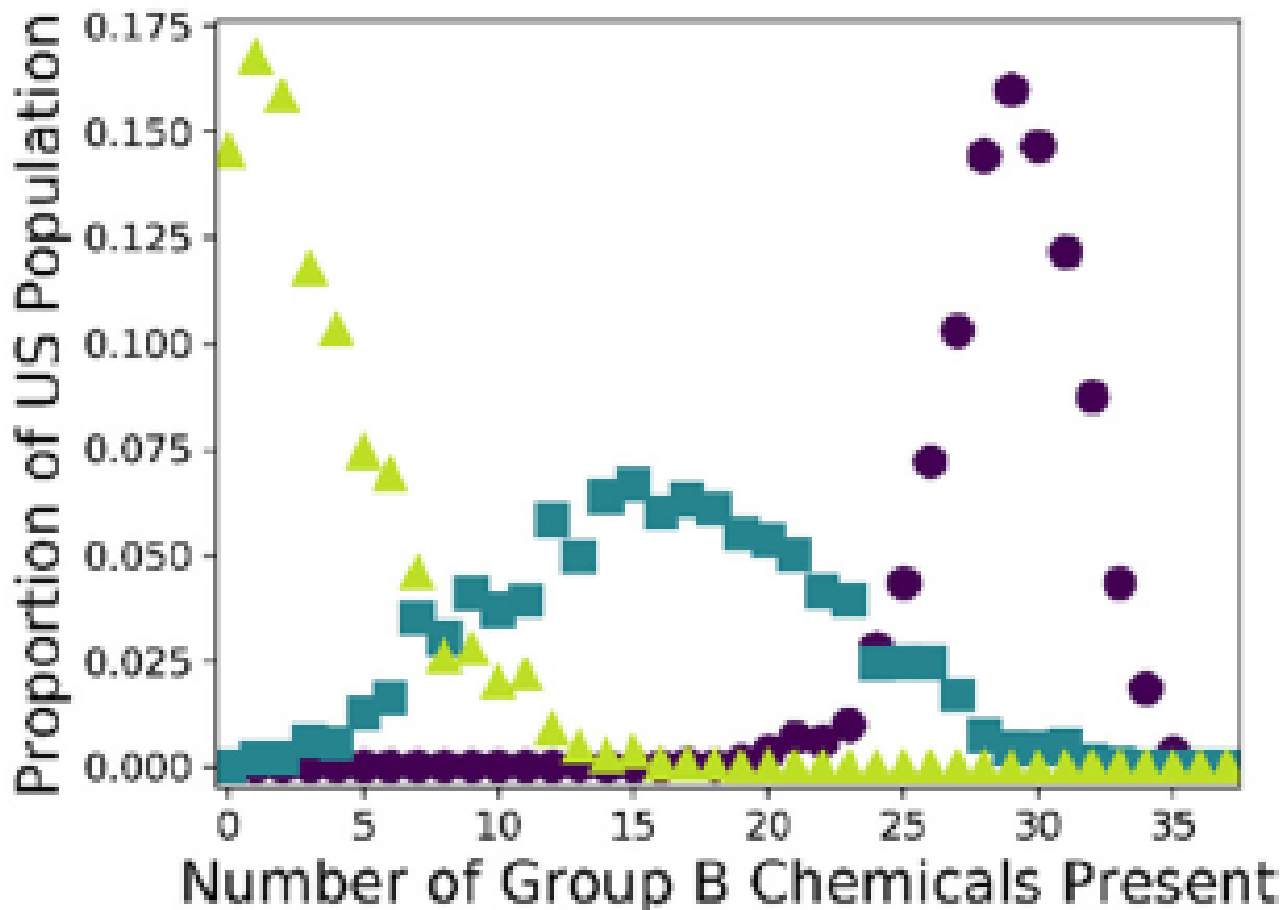
- Body weight
- Height
- **Chemical analysis of blood and urine**



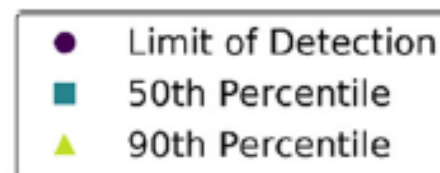
National Health and Nutrition Examination Survey

Co-Occurrence of Chemicals in Individuals

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



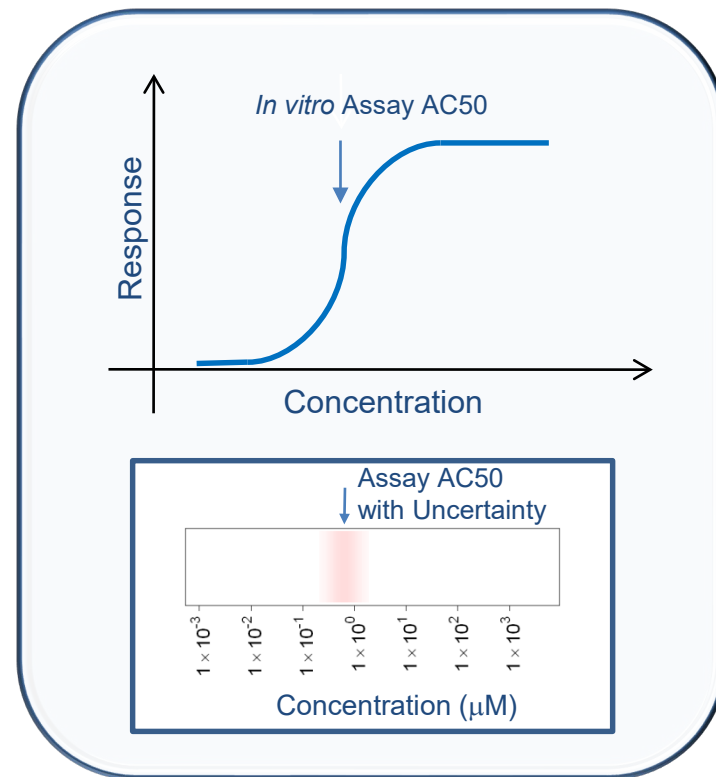
Ideally we would use some sort of chemical toxicity-informed point of departure but don't have that for all chemicals



High-Throughput Testing of Chemicals



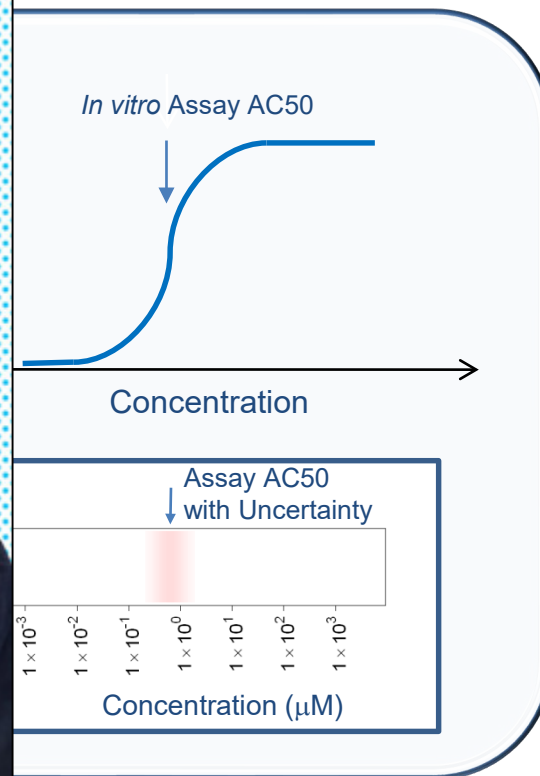
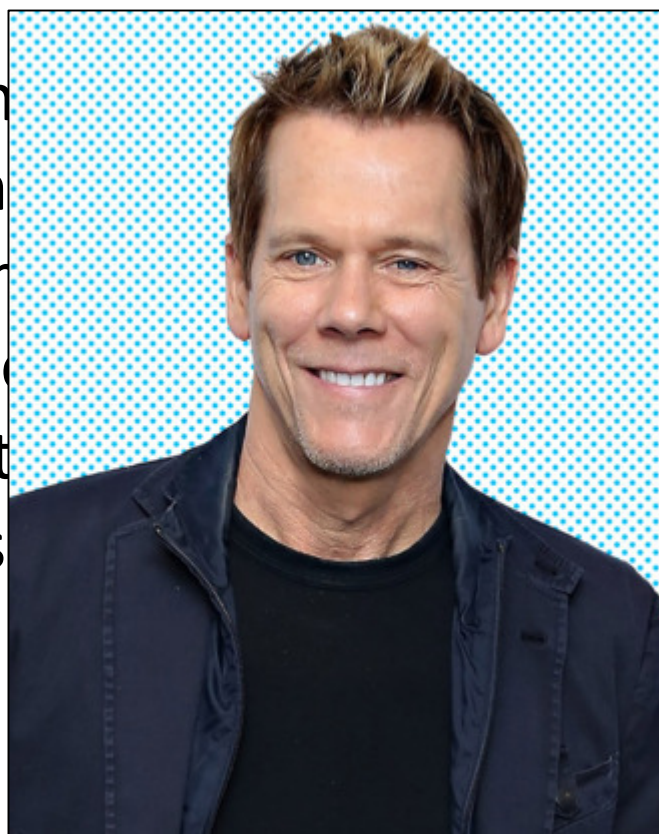
37 chemicals means 2^{37}
combinations that's
134,217,728 combinations –
Tox21 has tested ~8000
chemicals over the last ten
years



High-Throughput Testing of Chemicals



37 chemicals in
combination
134,217,728 combinations
Tox21 has tested
chemicals over the
years

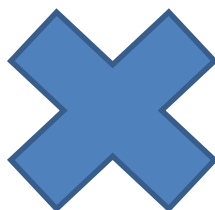


The Structure of Chemical Exposure

finch species

Loxigilla noxis
Melanospiza richardson
Tiara olivacea
Tiara bicolor
Tiara canora
Loxipasser anoxanthus

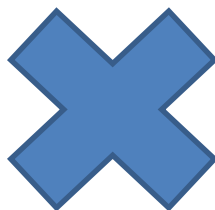
Cuba
Hispanolia
Jamaica
Puerto Rico
Guadeloupe



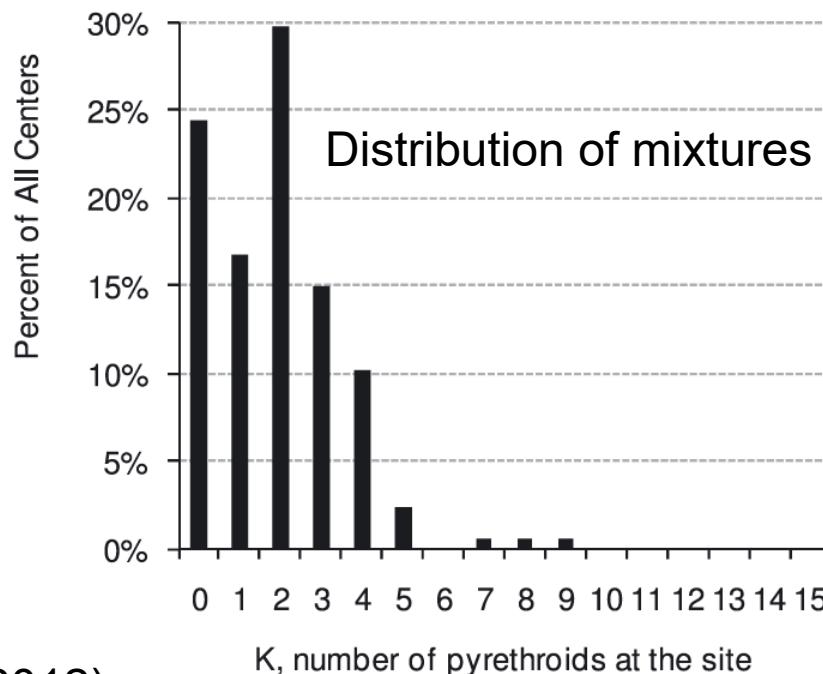
chemical species

chemical 1
chemical 2
chemical 3
chemical 4
chemical 5
chemical 6

site 1
site 2
site 3
site 4
site 5



- For n chemicals 2^n combinations are possible
 - However, not all are observed
- Diamond (1975): Not all finch species present on all islands of Caribbean
- Tornero-Velez et al. (2012): Not all chemical combinations present at all sites



Frequent Itemset Mining (FIM)

- Frequent itemset mining (FIM, Borgelt, 2012) identifies the prevalence (probability from 0-100%) that a set of items that co-occur in a “transaction”



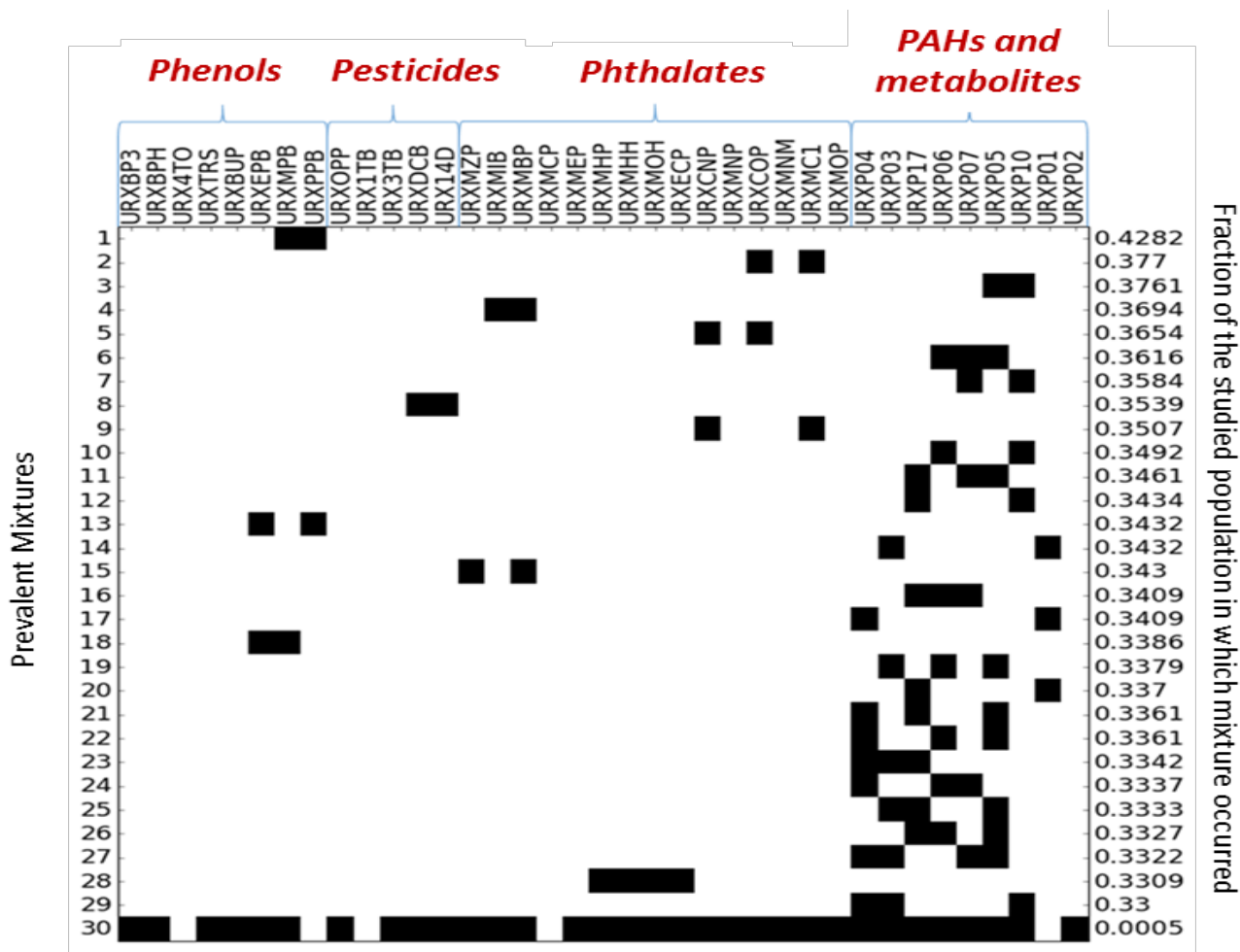
Frequent Itemset Mining (FIM)

- Frequent itemset mining (FIM, Borgelt, 2012) identifies the prevalence (probability from 0-100%) that a set of items that co-occur in a “transaction”
- Between 5pm and 7pm, customers tended to co-purchase beer and diapers

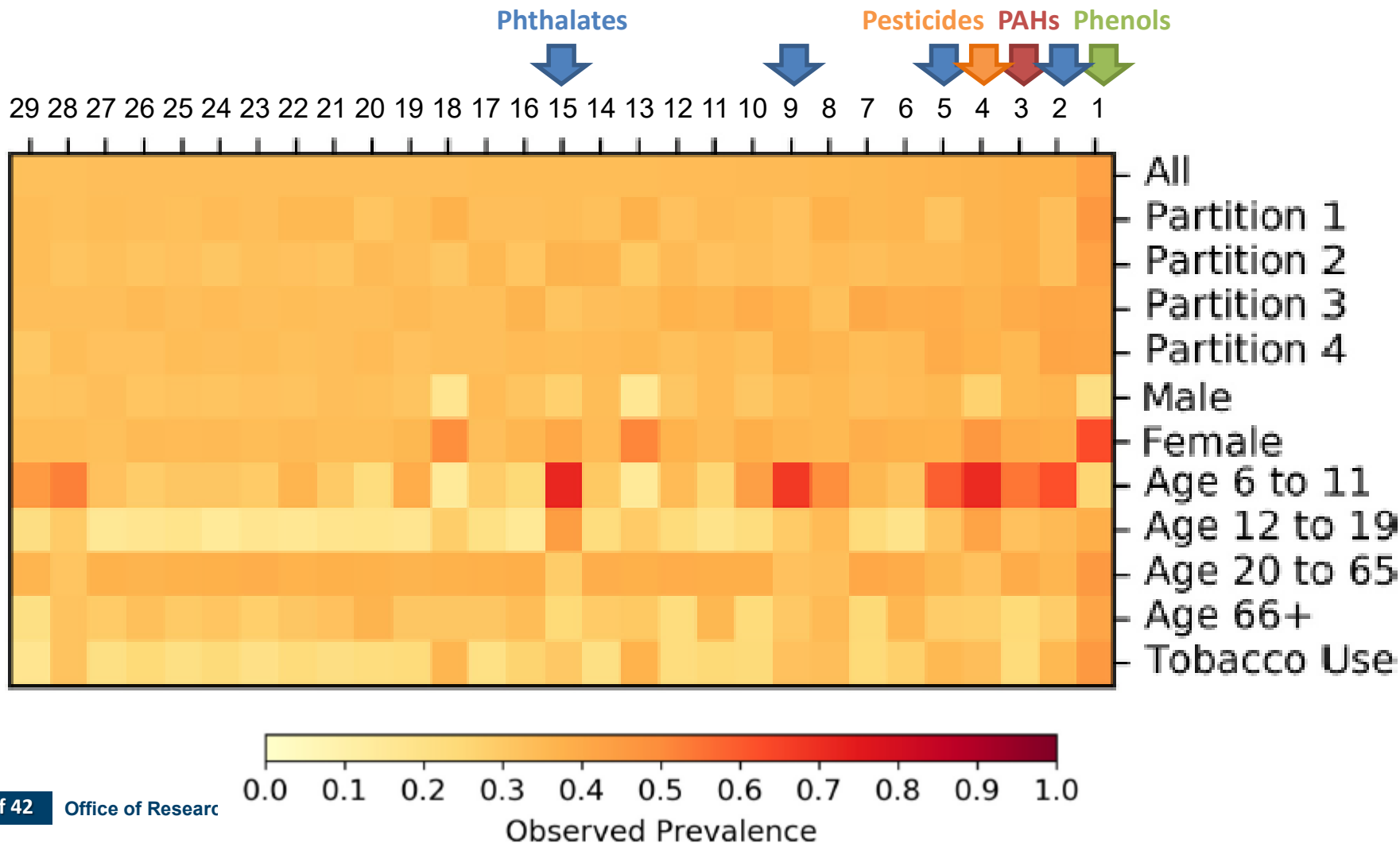


Identifying Prevalent Mixtures

- Kapraun et al. (2017) used FIM to identify combinations of items (chemicals) that co-occur together within CDC NHANES samples from same individual
- Used total population median concentration as threshold for “presence”
- Identified a few dozen mixtures present in >30% of U.S. population

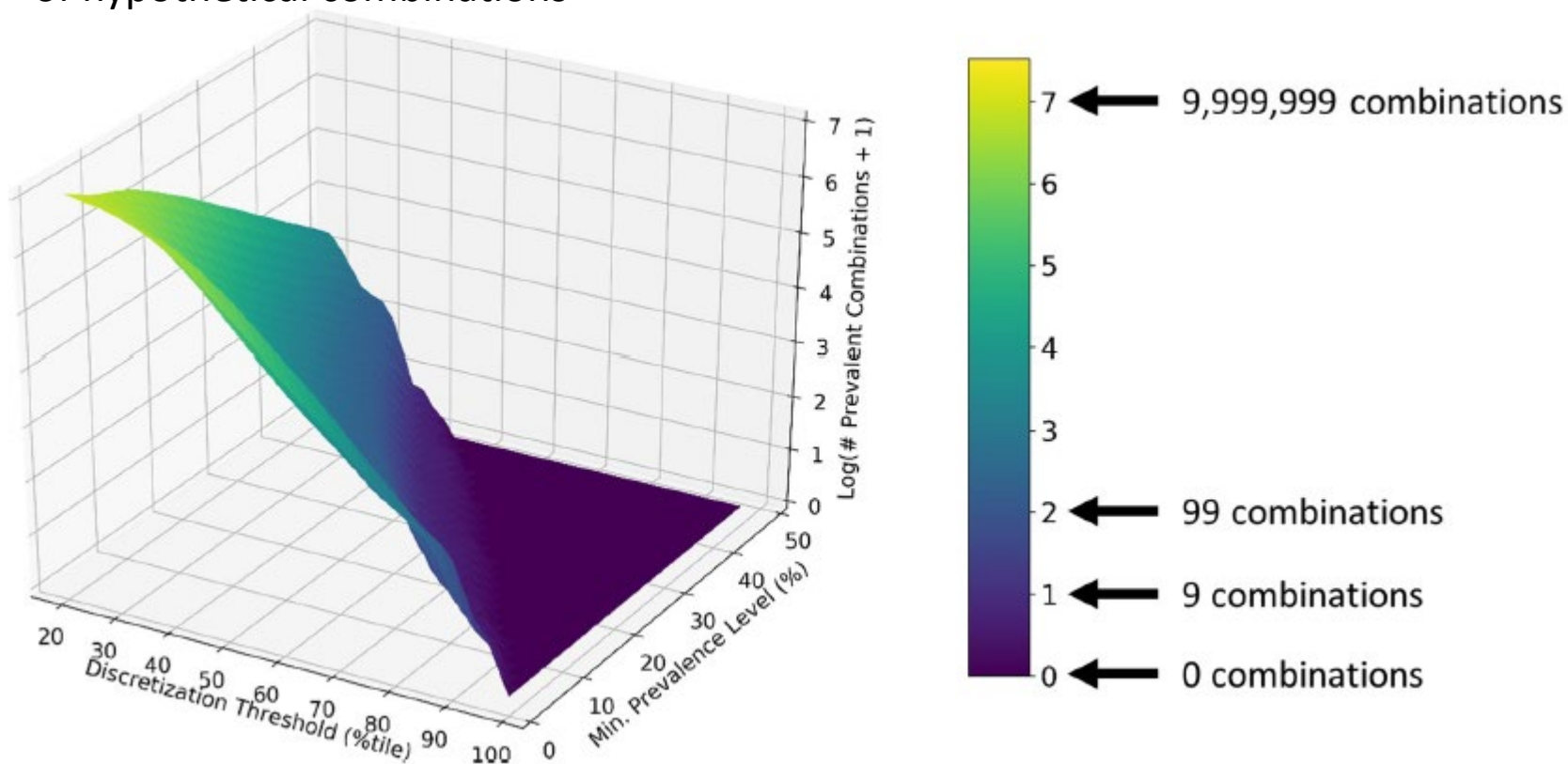


Demographic-Specific Prevalence of Combinations



A Testable Number of Combinations

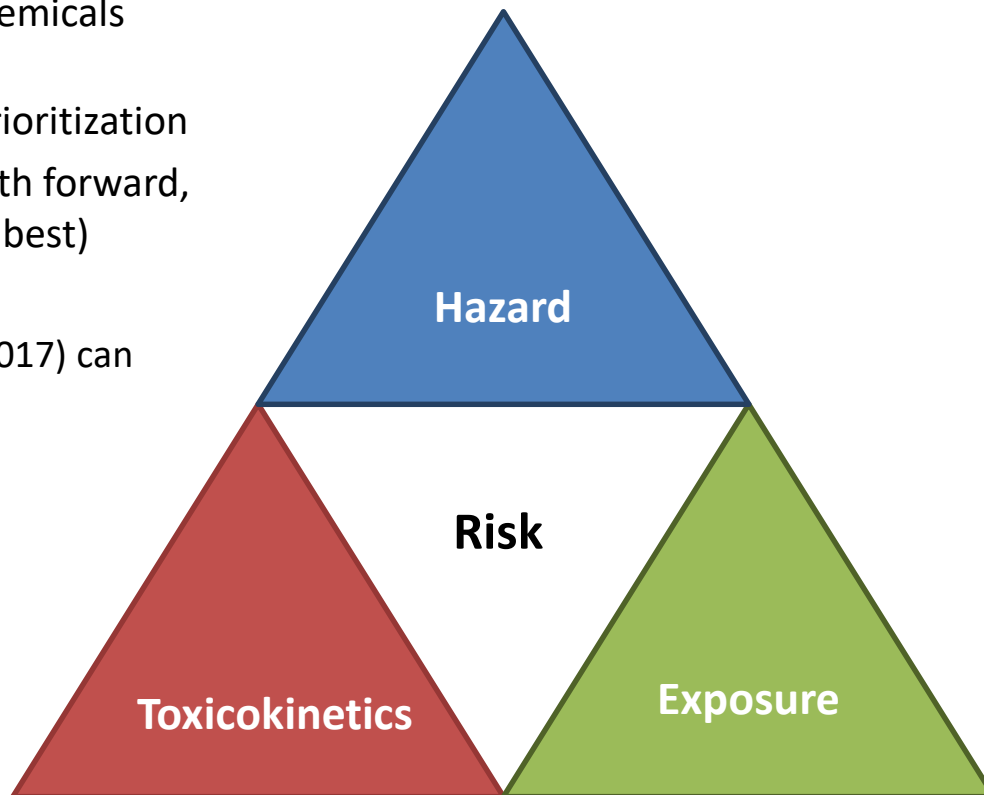
While high throughput screening (HTS) allows thousands of tests, there are millions of hypothetical combinations



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

Conclusions

- We would like to know more about the risk posed by thousands of chemicals in the environment – which ones should we start with?
- Using *in vitro* methods originally developed for pharmaceuticals, we can make useful predictions of hazard and TK for large numbers of chemicals
- Exposure data is also key to risk-based prioritization
 - Consensus modeling provides one path forward, but only as good as available data (at best)
- EPA's CompTox dashboard (Williams et al, 2017) can help you:
 - Identify chemicals
 - Find toxicity data
 - Find lists of chemicals
 - Find metabolites
 - Identify products
 - Find toxicokinetic information
 - Get physicochemical properties
 - Batch download data





The ExpoCast Project (Exposure Forecasting)

NCCT

Chris Grulke
Greg Honda*
Richard Judson
Ann Richard
Risa Sayre*
Mark Sfeir*
Rusty Thomas
John Wambaugh
Antony Williams

NRMRL

Xiaoyu Liu

NHEERL

Linda Adams
Christopher Ecklund
Marina Evans
Mike Hughes
Jane Ellen Simmons

NERL

Cody Addington*
Craig Barber
Namdi Brandon*
Peter Egeghy
Hongtai Huang*
Kristin Isaacs
Ashley Jackson*
Charles Lowe*
Dawn Mills*
Seth Newton

Katherine Phillips
Paul Price
Jeanette Reyes*
Randolph Singh*
Jon Sobus
John Streicher*
Mark Strynar
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
Barbara Wetmore

***Trainees**

Collaborators

Arnot Research and Consulting
Jon Arnot
Johnny Westgate
**Institut National de
l'Environnement et des Risques
(INERIS)**
Frederic Bois
Integrated Laboratory Systems
Kamel Mansouri
National Toxicology Program
Mike Devito
Steve Ferguson
Nisha Sipes
Ramboll
Harvey Clewell
ScitoVation
Chantel Nicolas
Silent Spring Institute
Robin Dodson
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Aylward
Technical University of Denmark
Peter Fantke
Tox Strategies
Caroline Ring
Miyoun Yoon
Unilever
Beate Nicol
Cecilie Rendal
Ian Sorrell
United States Air Force
Matt Linakis
Heather Pangburn
University of California, Davis
Deborah Bennett
University of Michigan
Olivier Jolliet
University of Texas, Arlington
Hyeong-Moo Shin

References

- Arnot, J. A.; Zhang, X.; Kircanski, I.; Hughes, L.; Armitage, J. Develop Sub-Module for Direct Human Exposures to Consumer Products. Technical Report for the U.S. Environmental Protection Agency; ARC Arnot Research & Consulting, Inc.: Toronto, ON, Canada, 2014.
- Bennett, D. H.; Furtaw, E. J., Fugacity-based indoor residential pesticide fate model. *Environmental Science & Technology* 2004, 38, (7), 2142-2152.
- Borgelt, C. (2012). Frequent item set mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(6), 437-456.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Breyer, Stephen. Breaking the vicious circle: Toward effective risk regulation. Harvard University Press, 2009
- Diamond JM. Assembly of species communities. Pp. 342-444 in Cody ML, Diamond JM (eds). *Ecology and Evolution of Communities*. Cambridge, MA: Belknap Press, 1975.
- Egeghy, P. P., et al. (2012). The exposure data landscape for manufactured chemicals. *Science of the Total Environment*, 414, 159-166.
- Ernstoff, A. S.; Fantke, P.; Huang, L.; Jolliet, O., High-throughput migration modelling for estimating exposure to chemicals in food packaging in screening and prioritization tools. *Food and Chemical Toxicology* 2017, 109, 428-438.
- Filer, Dayne L.. "The ToxCast analysis pipeline: An R package for processing and modeling chemical screening data." US Environmental Protection Agency: http://www.epa.gov/ncct/toxcast/files/MySQL%20Database/Pipeline_Overview.pdf (2014)
- Goldsmith, M. R., et al. (2014). Development of a consumer product ingredient database for chemical exposure screening and prioritization. *Food and chemical toxicology*, 65, 269-279.
- Huang, L.; Ernstoff, A.; Fantke, P.; Csiszar, S. A.; Jolliet, O., A review of models for near-field exposure pathways of chemicals in consumer products. *Science of The Total Environment* 2017, 574, 1182-1208.
- Huang, L.; Jolliet, O., A parsimonious model for the release of volatile organic compounds (VOCs) encapsulated in products. *Atmospheric Environment* 2016, 127, 223-235.
- Jolliet, O.; Ernstoff, A. S.; Csiszar, S. A.; Fantke, P., Defining Product Intake Fraction to Quantify and Compare Exposure to Consumer Products. *Environmental Science & Technology* 2015, 49, (15), 8924-8931.
- Judson, R. S., Magpantay, F. M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., ... & Houck, K. A. (2015). Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicological Sciences*, 148(1), 137-154.
- 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.
- Kapraun, Dustin et al., "A Method for Identifying Prevalent Chemical Combinations in the US Population," *Environmental Health Perspectives*, 2017
- Kavlock, Robert, et al. "Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management." *Chemical research in toxicology* 25.7 (2012): 1287-1302.
- MacLeod, Matthew, et al. "The state of multimedia mass-balance modeling in environmental science and decision-making." (2010): 8360-8364
- National Academies of Sciences, Engineering, and Medicine. (2017). *Using 21st century science to improve risk-related evaluations*. National Academies Press.
- National Research Council. (1983). *Risk Assessment in the Federal Government: Managing the Process Working Papers*. National Academies Press.
- National Research Council. (2007) *Toxicity testing in the 21st century: a vision and a strategy*. National Academies Press.
- O'Connell, S. G., Kincl, L. D., & Anderson, K. A. (2014). Silicone wristbands as personal passive samplers. *Environmental science & technology*, 48(6), 3327-3335.
- Park, Youngja, H., et al. "High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring." *Toxicology* 295:47-55 (2012)
- Pearce, Robert, et al. "htk: R Package for High-Throughput Toxicokinetics." *Journal of Statistical Software*, 2017
- Phillips, Katherine A., et al. "High-throughput screening of chemicals as functional substitutes using structure-based classification models." *Green Chemistry* (2017).
- Phillips, Katherine A., et al. "Suspect Screening Analysis of Chemicals in Consumer Products", accepted at *Environmental Science & Technology*.
- Rotroff, Daniel, et al., (2010) "Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening." *Tox. Sciences* 117(2), 348-58
- Schmidt, Charles W. "TOX 21: new dimensions of toxicity testing." *Environmental health perspectives* 117.8 (2009): A348.
- Shin, H.-M.; McKone, T. E.; Bennett, D. H., Intake Fraction for the Indoor Environment: A Tool for Prioritizing Indoor Chemical Sources. *Environmental Science & Technology* 2012, 46, (18), 10063-10072.
- 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.
- Wallace et al., "The TEAM Study: Personal exposures to toxic substances in air, drinking water, and breath of 400 residents of New Jersey, North Carolina, and North Dakota ." *Environmental Research* 43: 209-307 (1987)
- Tornero-Velez et al. (2012) "Biogeographical Analysis of Chemical Co-Occurrence to Identify Priorities for Mixtures Research": *Risk Analysis*, 32(2) 224-236
- 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. "Evaluating In Vitro-In Vivo Extrapolation" accepted at *Toxicological Sciences*,.
- Wang, Y.-H. (2010). "Confidence Assessment of the Simcyp Time-Based Approach and a Static Mathematical Model in Predicting Clinical Drug-Drug Interactions for Mechanism-Based CYP3A Inhibitors." *Drug Metabolism and Disposition* 38(7), 1094-1104
- Wetmore, Barbara A., et al. "Integration of dosimetry, exposure and high-throughput screening data in chemical toxicity assessment." *Toxicological Sciences* (2012): kfr254.
- Wetmore, Barbara A., et al. "Incorporating High-Throughput Exposure Predictions with Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing." *Toxicological Sciences* 148.1 (2015): 121-136.
- Zhang, X.; Arnot, J. A.; Wania, F., Model for screening-level assessment of near-field human exposure to neutral organic chemicals released indoors. *Environmental science & technology* 2014, 48, (20), 12312-12319.