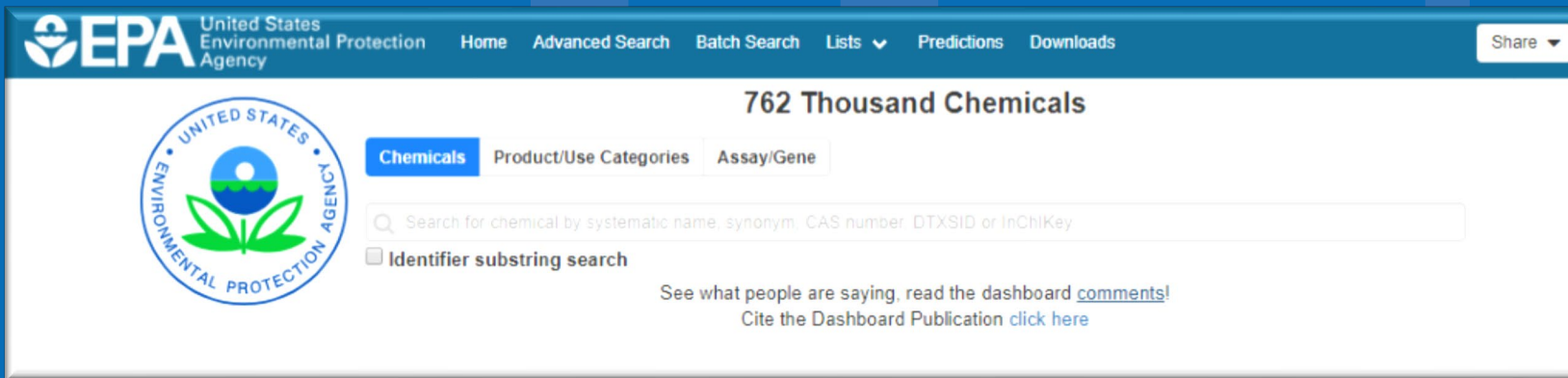


Public Exposure Information on the CompTox Dashboard

*John Wambaugh¹, Kristin Isaacs²,
Katherine Phillips², Antony Williams¹*

1. National Center for Computational Toxicology, Office of Research and Development, U.S. EPA
2. Computational Exposure Division, National Exposure Research Laboratory, Office of Research and Development, U.S. EPA



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

International Society of Exposure Science
*Pre-Conference Course: Application of New
Approach Methodologies for Exposure
Assessment and Prioritization – Tools for
Researchers and Regulators Including use of
Quantitative Structure Use Relationships
(QSUR)*
Ottawa, Canada
August 25, 2018

Chemical Regulation in the United States

- Park *et al.* (2012): At least 3221 chemicals present in pooled human blood samples, many appear to be exogenous albeit at low levels
- 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)



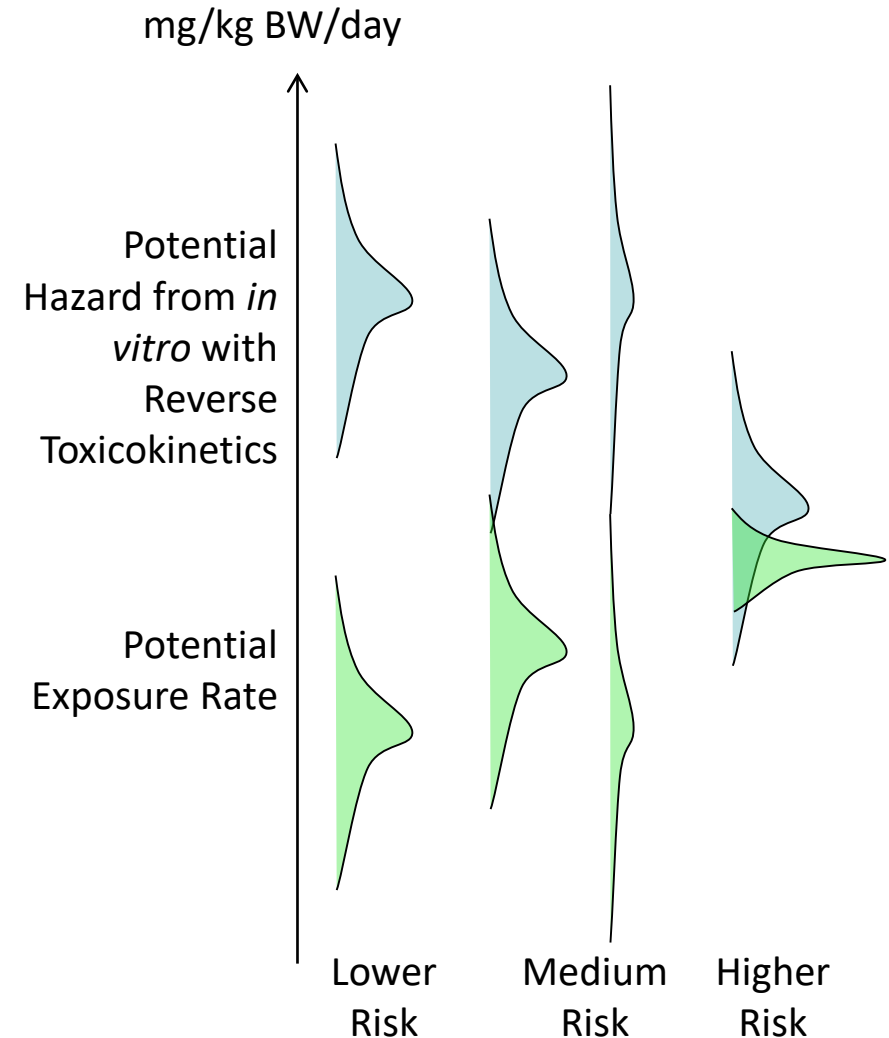
Chemical Regulation in the United States

- 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
 - New alternative methodologies (NAMs) are being developed to prioritize these existing and new chemicals for testing



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 new approach methodologies that can prioritize those chemicals most worthy of additional study
- **High throughput risk prioritization** needs:
 1. high throughput **hazard** characterization (Dix et al., 2007, Collins et al., 2008)
 2. high throughput **exposure** forecasts (Wambaugh et al., 2013, 2014)
 3. high throughput **toxicokinetics** (*i.e.*, dose-response relationship) linking hazard and exposure

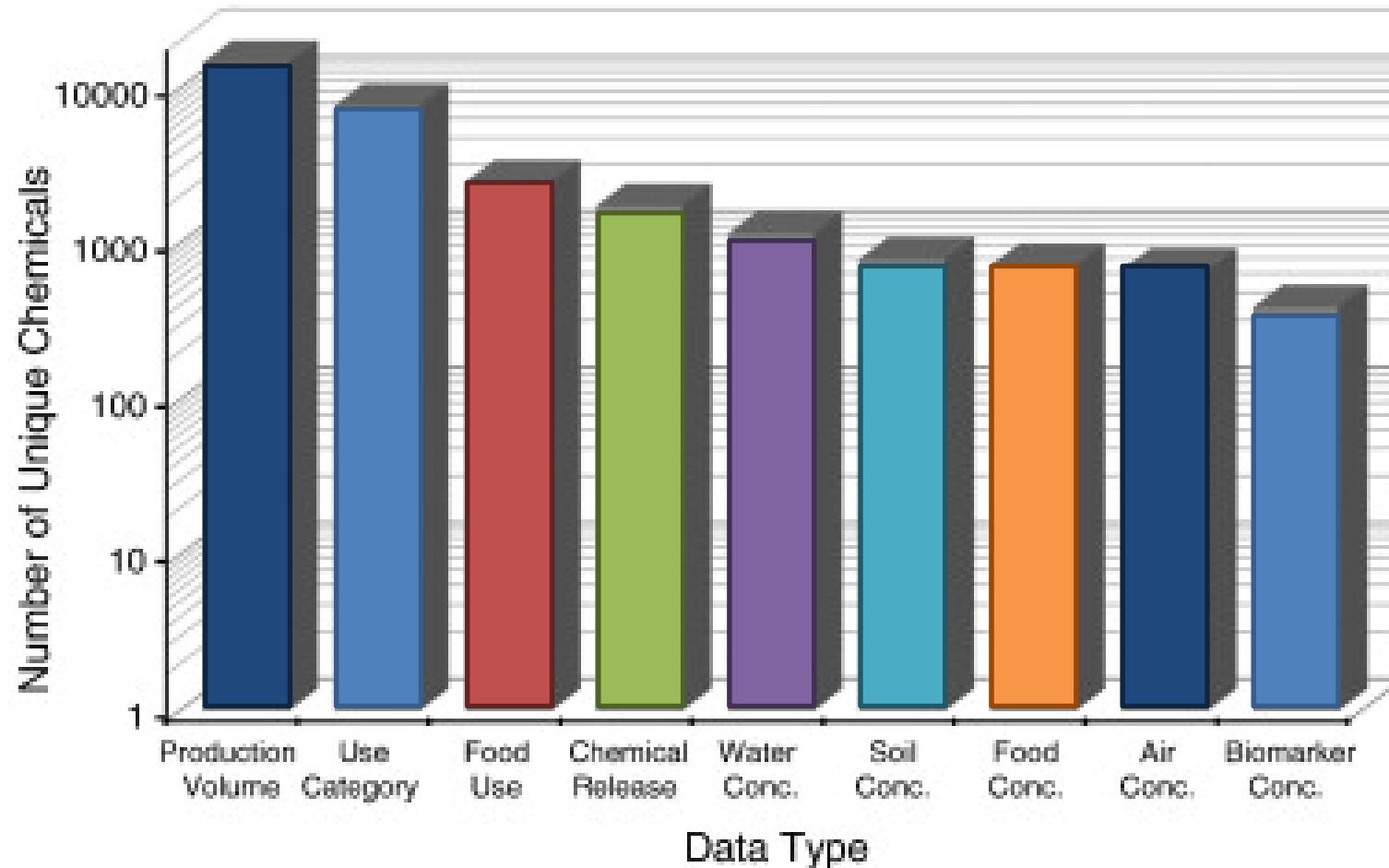


Rotroff et al. (2010)

Wetmore et al. (2012, 2014, 2015)

Limited Available Data for Exposure Estimation

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



Can we develop new tools to generate the exposure information we need?

Understanding Exposure to Chemicals

Rudel et al., 2003



Research

Phthalates, Alkylphenols, Pesticides, Polybrominated Diphenyl Ethers, and Other Endocrine-Disrupting Compounds in Indoor Air and Dust

RUTHANN A. RUDEL,^{a,*} DAVID E. CAMANN,^a JOHN D. SPENGLER,^b LEO R. KORN,^b AND JULIA G. BRODY^c
Silent Spring Institute, 29 Crafts Street, Newton, Massachusetts 02458, Southwest Research Institute, 6220 Culebra Road, P.O. Box 28510, San Antonio, Texas 78228-0510, Environmental Science and Engineering Program, Harvard University School of Public Health, Landmark Center, 401 Park Drive, Boston, Massachusetts 02115, and Division of Biometrics, University of Medicine and Dentistry of New Jersey, School of Public Health, 335 George Street, Liberty Plaza, Suite 2200, New Brunswick, New Jersey 08903-2688

Chemicals identified as endocrine-disrupting compounds (EDCs) have widespread consumer uses, yet little is known about indoor exposure. We sampled indoor air and dust in 120 homes, analyzing for 89 organic chemicals identified as EDCs. Fifty-two compounds were detected in air and 66 were detected in dust. These are the first reported measures in residential environments for over 30 of the compounds, including several detected at the highest concentrations. The number of compounds detected per home ranged from 13 to 28 in air and from 6 to 42 in dust. The most abundant compounds in air included phthalates (plasticizers, emulsifiers), *o*-phenylphenol (disinfectant), 4-nonylphenol (detergent metabolite), and 4-*tert*-butylphenol (adhesive) with typical concentrations in the range of 50–1500 ng/m³. The penta- and tetrabrominated diphenyl ethers (flame retardants) were frequently detected in dust, and 2,3-dibromo-1-propanol, the carcinogenic intermediate of a flame retardant banned in 1977, was detected in air and dust. Twenty-three pesticides were detected in air and 27 were detected in dust, the most abundant being

Introduction

Current widespread interest in a range of health effects potentially associated with endocrine-disrupting compounds (EDCs) has made exposure assessment for these compounds a priority. Studies of potential health effects associated with EDCs have been hampered by lack of information about the major sources of exposure to EDCs. Furthermore, because many EDCs act additively through a common mechanism of action or have antagonistic or other interactive effects by operating at different points in cell signaling systems, consideration of exposure to mixtures is critical in studies of health effects (1–7). These questions are particularly important in relation to indoor environments, which have been identified as an important source of chemical exposures (8–11). People spend a large fraction of their time indoors, and indoor sources of chemicals, coupled with limited ventilation and slow chemical degradation processes, cause increased pollutant concentrations indoors. In fact, indoor air specifically has been described as “one of the most serious environmental risks to human health” (8).

Many high production volume chemicals—including some already identified as EDCs—have consumer uses (e.g., in plastics, detergents, and other household and consumer products) that make them potentially important indoor contaminants. While a number of comprehensive exposure studies have been conducted or are underway to characterize residential exposures to selected contaminants, particularly volatile organic compounds, pesticides, and polycyclic aromatic hydrocarbons (PAHs), these studies have been limited to a small number of compounds and have focused on characterizing exposure pathways and sources (12–18). We were unable to locate exposure data for many of our compounds of interest, including alkylphenols, parabens, polybrominated diphenyl ethers (PBDEs), and many of the estrogenic phenolic compounds such as bisphenol A. We located only one (unpublished) study of substantial size that has characterized phthalate concentrations in indoor air (18).

The primary objective of this study is to provide an assessment of household exposure to a broad suite of organic chemicals that have been identified as EDCs. Indoor air and dust were selected for analysis because many EDCs are used in consumer products and building materials (6, 19), so these chemicals would be expected indoors. Indoor air has been identified as an important source of chemical exposure, while house dust has been demonstrated to be an important exposure pathway in young children (9, 10). Dust also provides

TABLE 4. Most Abundant Chemicals

Ten Chemicals with Highest 90th Percentile Concentrations	
air (ng/m ³) ^a	dust (μg/g) ^a
diethyl phthalate (1,600) 100	bis(2-ethylhexyl) phthalate (854) 100
<i>o</i> -phenylphenol (440) 100	benzyl butyl phthalate (277) 100
di- <i>n</i> -butyl phthalate (430) 100	di- <i>n</i> -butyl phthalate (43.9) 98
4-nonylphenol (230) 100	nonylphenol diethoxylate (18.9) 86
bis(2-ethylhexyl) phthalate (210) 68	bis(2-ethylhexyl) adipate (16.6) 100
diisobutyl phthalate (150) 100	<i>trans</i> -permethrin (16.5) 53
benzyl butyl phthalate (68) 44	piperonyl butoxide (15.1) 66
4- <i>tert</i> -butylphenol (43) 100	diethyl phthalate (10.8) 89
nonylphenol monoethoxylate (41) 95	nonylphenol monoethoxylate (8.55) 86
bis(2-ethylhexyl) adipate (22) 99	<i>cis</i> -permethrin (7.04) 45
10 Pesticides with Highest 90th Percentile Concentrations	
air (ng/m ³) ^a	dust (μg/g) ^a
<i>o</i> -phenylphenol (440) 100	<i>trans</i> -permethrin (16.5) 53
heptachlor ^b (19) 44	piperonyl butoxide (15.1) 66
propoxur (16) 49	<i>cis</i> -permethrin (7.04) 45
<i>γ</i> -chlorodane ^b (12) 53	methoxychlor ^b (3.38) 54
chlorpyrifos (12) 38	4,4'-DDT ^b (3.19) 65
pentachlorophenol ^b (10) 58	pentachlorophenol ^b (2.42) 86
diazinon (9.0) 40	chlorpyrifos ^b (1.87) 18
<i>α</i> -chlorodane ^b (8.8) 51	carbaryl (1.72) 43
chlorothalonil (3.4) 17	propoxur (1.70) 42
3,5,6-trichloro-2-pyridinol (1.1) 13	bendiocarb (1.11) 12

^a Percent detection in italics. ^b Indicates banned or restricted-use pesticide (at time of sample collection).

Rudel et al., 2008

EPIDEMIOLOGY

Welcome Ovid Subscriber!

Your institution's Ovid subscription entitles you to view the full text of articles on the Epidemiology journal website.

Correlations Between Urinary Phthalate Metabolites and Phthalates, Estrogenic Compounds 4-Butyl phenol and *o*-Phenyl phenol, and Some Pesticides in Home Indoor Air and House Dust

Rudel, R A^{*}; Dodson, R E^{*}; Newton, E[†]; Zota, A R^{*}; Brody, J G^{*}

Epidemiology: November 2008 - Volume 19 - Issue 6 - p S332

doi: 10.1097/01.ede.0000340529.83416.do

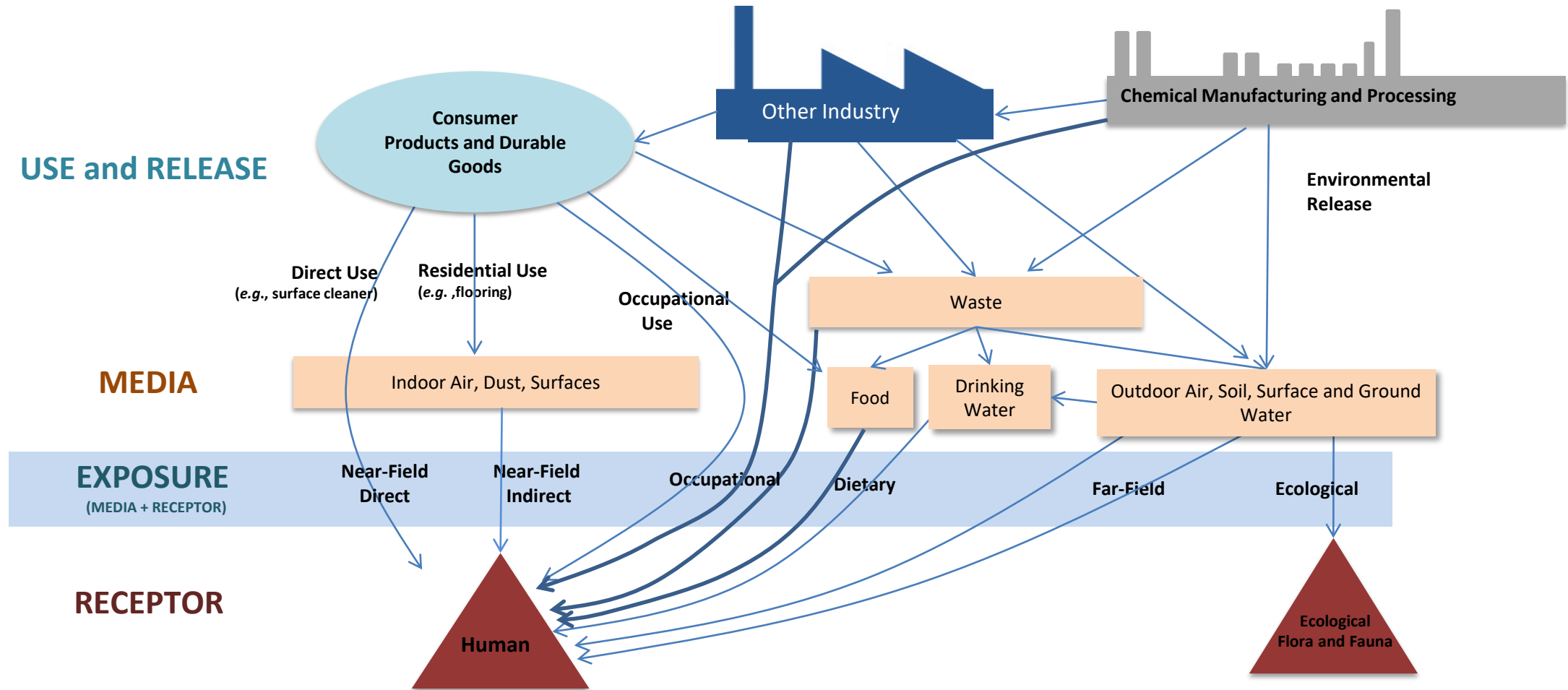
Abstracts: ISEE 20th Annual Conference, Pasadena, California, October 12–16, 2008: Contributed Abstracts

“Overall, these data show that concentrations of many EDCs in biological samples and indoor air and dust co-vary, suggesting that some EDC mixtures may originate from common exposure sources and highlighting potential confounding by other EDCs in health effect studies of phthalates. Future work will utilize factor analysis to **identify source profiles** of EDC mixtures that are associated with urinary phthalate levels.”

2013: ExpoCast Team “Helps” With Correlating Exposure and Biomarkers

- The EPA’s Exposure Forecaster (ExpoCast) team works to develop new methods for exposure
- To correlate metabolites in urine with parent chemical exposures from dust and consumer products we need to know (at least):
 - What chemicals are we talking about? CAS are not unique! Names are certainly not unique.
 - Are there any toxicity data?
 - Which metabolites link to which parent chemicals?
 - What consumer products contain which chemicals?
 - What does a given concentration of a chemical in urine (or plasma) imply about total body burden and exposure?
 - What are the relevant physicochemical properties?
- In 2013 when we set out to do this analysis **we could not answer most of these questions**
- Now you can with the CompTox dashboard: <https://comptox.epa.gov/dashboard>

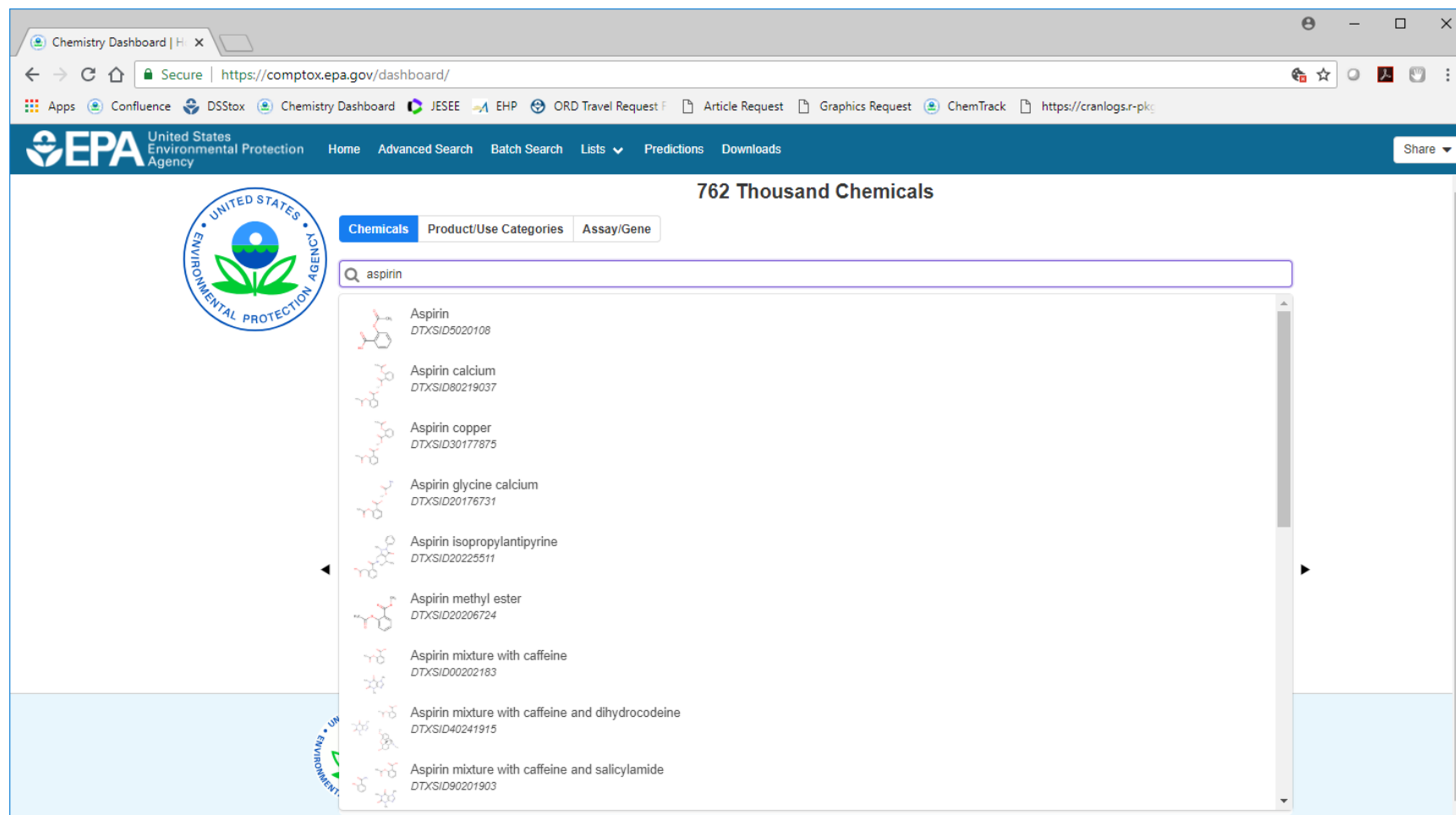
Understanding Exposure is a Systems Problem



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

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6



The screenshot shows the EPA Chemistry Dashboard interface. The browser address bar displays <https://comptox.epa.gov/dashboard/>. The dashboard header includes the EPA logo and navigation links: Home, Advanced Search, Batch Search, Lists, Predictions, and Downloads. A search bar on the left contains the text "aspirin". The main content area, titled "762 Thousand Chemicals", displays a list of search results for "aspirin". Each result includes a chemical structure icon, the name of the chemical, and its DTXSID.

Chemical Name	DTXSID
Aspirin	DTXSID5020108
Aspirin calcium	DTXSID80219037
Aspirin copper	DTXSID30177875
Aspirin glycine calcium	DTXSID20176731
Aspirin isopropylantipyrine	DTXSID20225511
Aspirin methyl ester	DTXSID20206724
Aspirin mixture with caffeine	DTXSID00202183
Aspirin mixture with caffeine and dihydrocodeine	DTXSID40241915
Aspirin mixture with caffeine and salicylamide	DTXSID90201903

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

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6

Chemistry Dashboard x

Secure | <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020108>

Apps Confluence DSSTox Chemistry Dashboard JESEE EHP ORD Travel Request F Article Request Graphics Request ChemTrack <https://cranlogs.r-pk>

EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads Copy Share Submit Comment Search all data

Aspirin

50-78-2 | DTXSID5020108
Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

SIMILAR COMPOUNDS

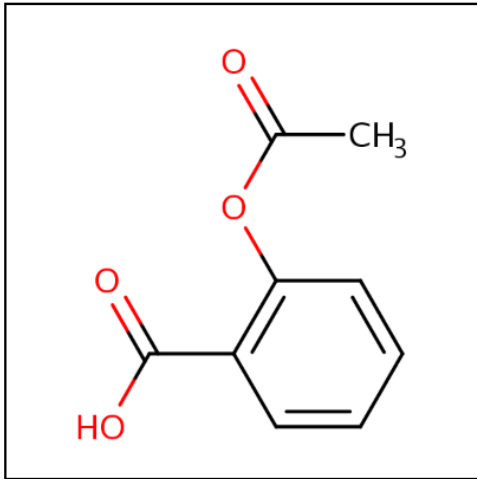
GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

LITERATURE

LINKS



Wikipedia

Aspirin, also known as acetylsalicylic acid (ASA), is a medication used to treat pain, fever, or inflammation. Specific inflammatory conditions in which aspirin is used include Kawasaki disease, pericarditis, and rheumatic fever. Aspirin given shortly after a heart attack decreases the risk of death. Aspirin is also used long-term to help prevent heart attacks, ischaemic strokes, and blood clots in people at high risk. It may also decrease the

[Read more](#)

Intrinsic Properties

Molecular Formula: $C_9H_8O_4$ [Mol File](#) [Find All Chemicals](#)

Average Mass: 180.159 g/mol [Isotope Mass Distribution](#)

Monoisotopic Mass: 180.042259 g/mol

Structural Identifiers

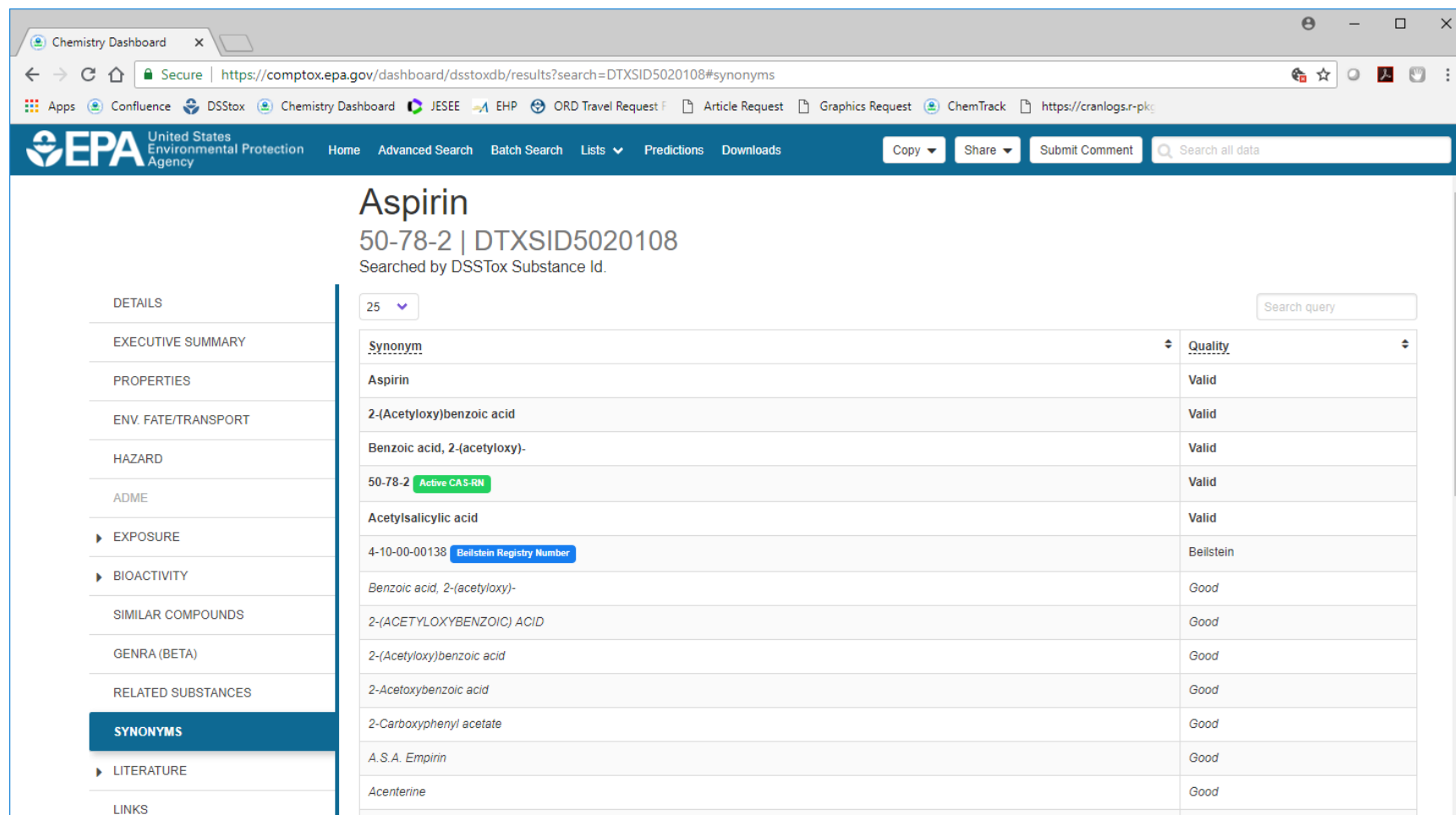
Linked Substances

Presence in Lists

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

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6



Chemistry Dashboard

Secure | <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020108#synonyms>

EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

Aspirin

50-78-2 | DTXSID5020108

Searched by DSSTox Substance Id.

25

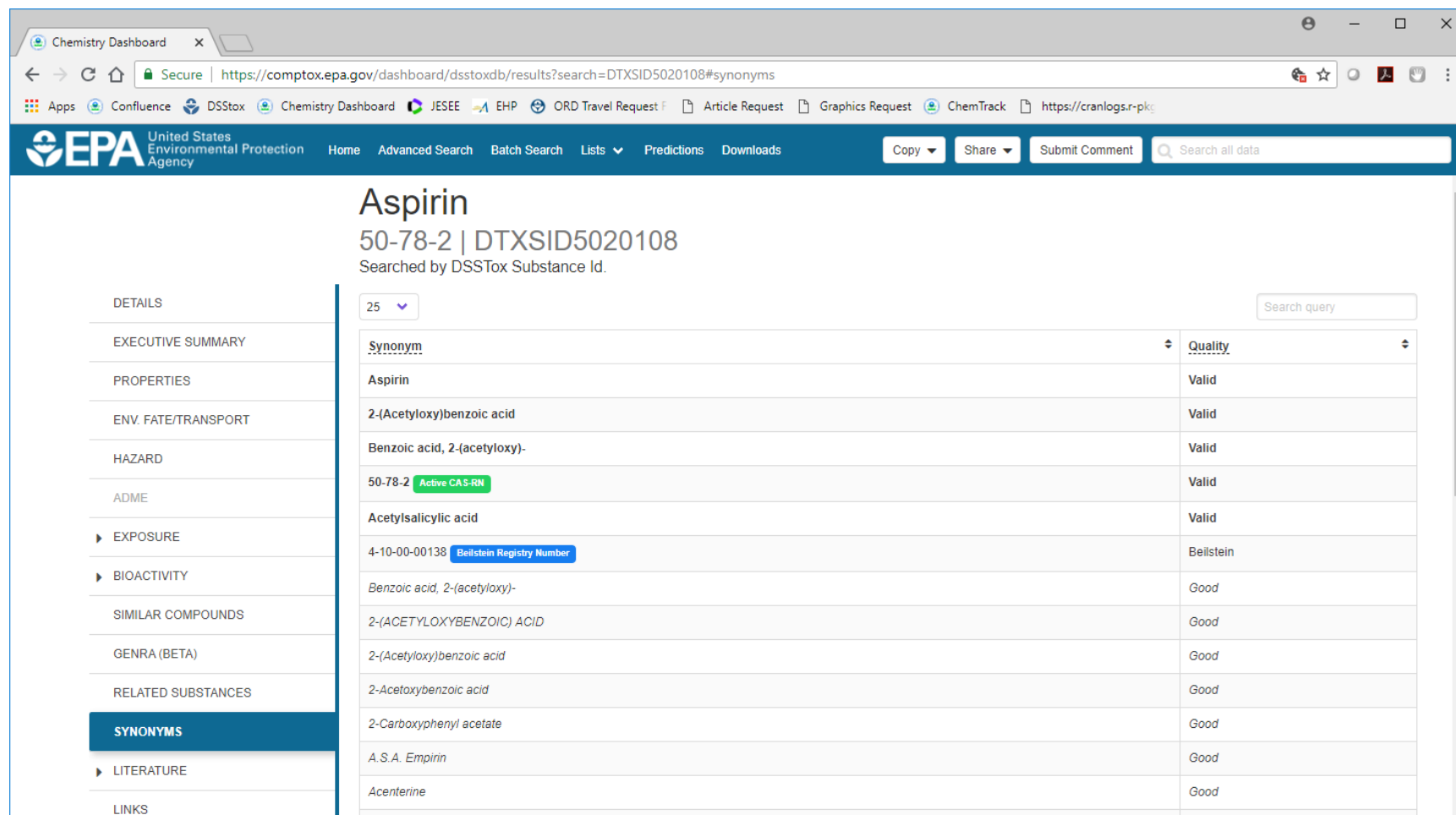
Search query

Synonym	Quality
Aspirin	Valid
2-(Acetyloxy)benzoic acid	Valid
Benzoic acid, 2-(acetyloxy)-	Valid
50-78-2 Active CAS-RN	Valid
Acetylsalicylic acid	Valid
4-10-00-00138 Beilstein Registry Number	Beilstein
Benzoic acid, 2-(acetyloxy)-	Good
2-(ACETYLOXYBENZOIC) ACID	Good
2-(Acetyloxy)benzoic acid	Good
2-Acetoxybenzoic acid	Good
2-Carboxyphenyl acetate	Good
A.S.A. Empirin	Good
Acenterine	Good

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

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6



Chemistry Dashboard

Secure | <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020108#synonyms>

EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

Aspirin

50-78-2 | DTXSID5020108
Searched by DSSTox Substance Id.

25

Search query

Synonym	Quality
Aspirin	Valid
2-(Acetyloxy)benzoic acid	Valid
Benzoic acid, 2-(acetyloxy)-	Valid
50-78-2 Active CAS-RN	Valid
Acetylsalicylic acid	Valid
4-10-00-00138 Beilstein Registry Number	Beilstein
Benzoic acid, 2-(acetyloxy)-	Good
2-(ACETYLOXYBENZOIC) ACID	Good
2-(Acetyloxy)benzoic acid	Good
2-Acetoxybenzoic acid	Good
2-Carboxyphenyl acetate	Good
A.S.A. Empirin	Good
Acenterine	Good

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

50-78-2 Active CAS-RN

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6

Chemistry Dashboard

Secure | <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020108#synonyms>

Apps Confluence DSStox Chemistry Dashboard JESEE EHP ORD Travel Request F Article Request Graphics Request ChemTrack <https://cranlogs.r-pkg>

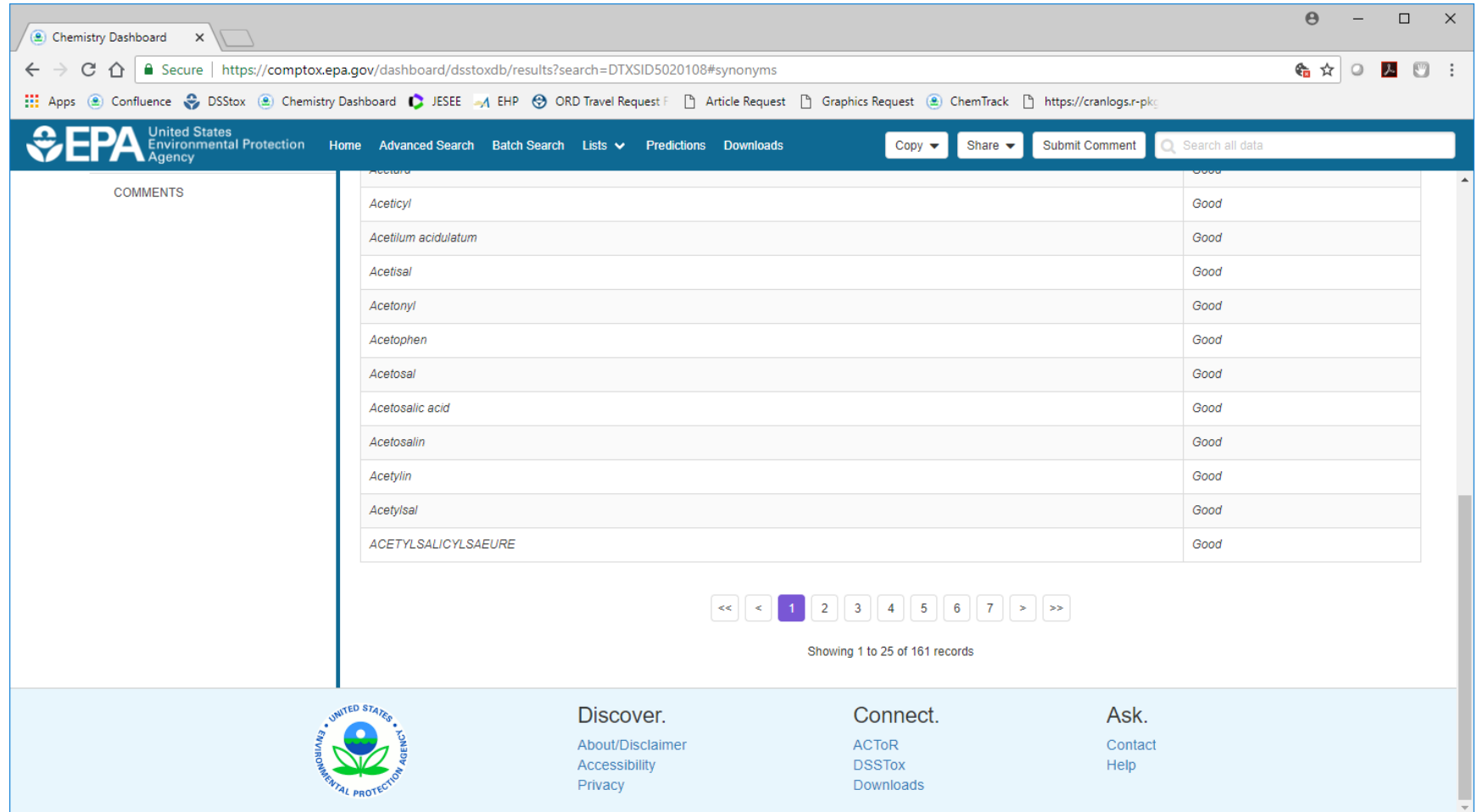
EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads Copy Share Submit Comment Search all data

ADME	50-78-2 Active CAS-RN	Valid
EXPOSURE	Acetylsalicylic acid	Valid
BIOACTIVITY	4-10-00-00138 Beilstein Registry Number	Beilstein
SIMILAR COMPOUNDS	Benzoic acid, 2-(acetyloxy)-	Good
GENRA (BETA)	2-(ACETYLOXYBENZOIC) ACID	Good
RELATED SUBSTANCES	2-(Acetyloxy)benzoic acid	Good
SYNONYMS	2-Acetoxybenzoic acid	Good
LITERATURE	2-Carboxyphenyl acetate	Good
LINKS	A.S.A. Empirin	Good
COMMENTS	Aceterine	Good
	Acetard	Good
	Aceticyl	Good
	Acetium acidulatum	Good
	Acetisal	Good
	Acetonyl	Good
	Acetophen	Good
	Acetosol	Good
	Acetosalic acid	Good
	Acetosalin	Good

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

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6



The screenshot shows the EPA Chemistry Dashboard interface. The browser address bar displays the URL: <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020108#synonyms>. The dashboard header includes the EPA logo, navigation links (Home, Advanced Search, Batch Search, Lists, Predictions, Downloads), and a search bar. The main content area displays a table of search results for the query "DTXSID5020108#synonyms". The table has two columns: "Name" and "Quality". The results are as follows:

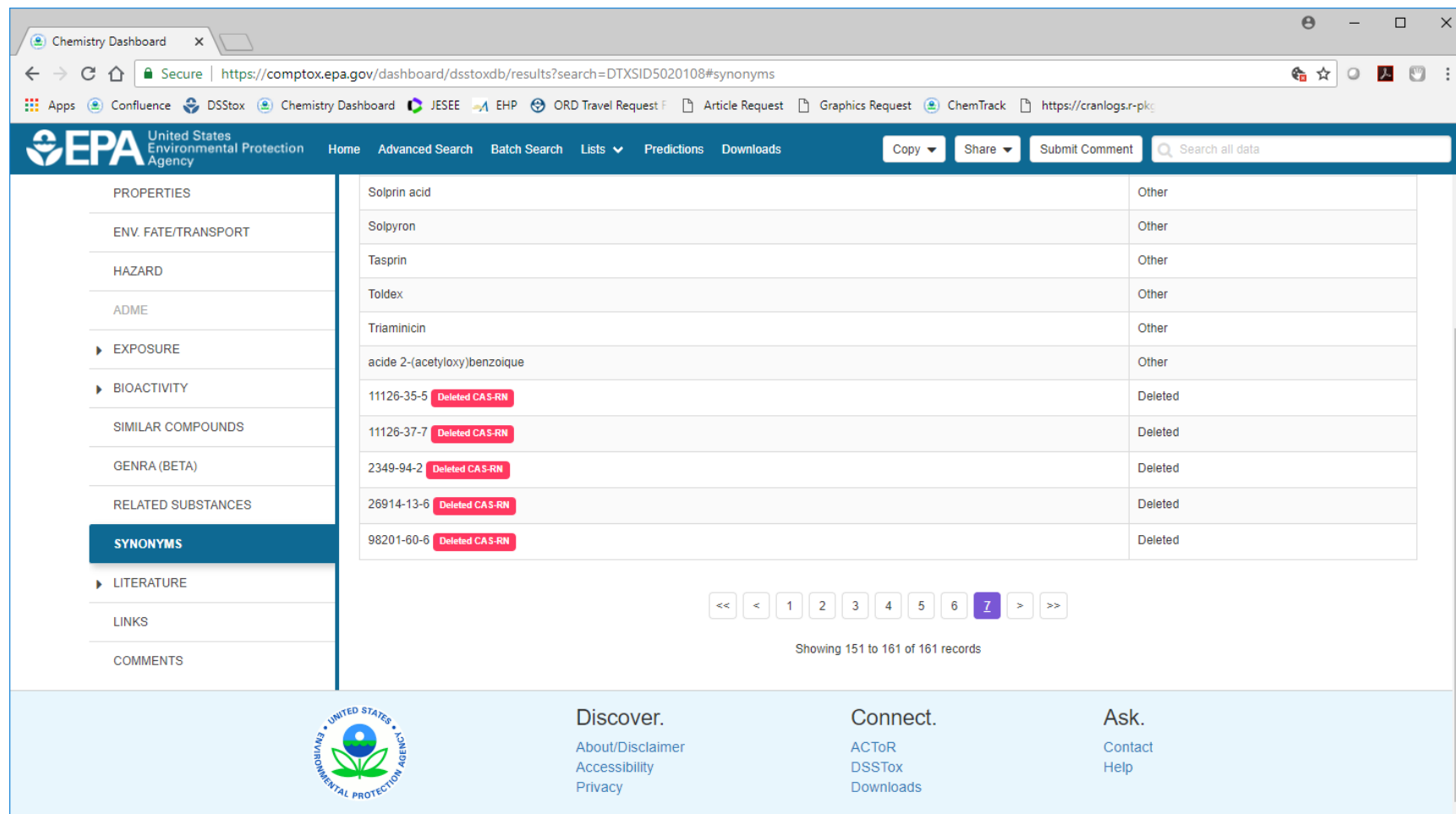
Name	Quality
Acetyl	Good
Acetium acidulatum	Good
Acetisal	Good
Acetonyl	Good
Acetophen	Good
Acetosol	Good
Acetosalic acid	Good
Acetosalin	Good
Acetylin	Good
Acetylsal	Good
ACETYLSALICYLSAEURE	Good

At the bottom of the table, there is a pagination control showing "Showing 1 to 25 of 161 records". The footer of the dashboard includes the EPA logo, navigation links (Discover, Connect, Ask), and a search bar.

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

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6



Chemistry Dashboard

Secure | <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020108#synonyms>

Apps Confluence DSSTox Chemistry Dashboard JESEE EHP ORD Travel Request Article Request Graphics Request ChemTrack <https://cranlogs.r-pk>

EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads Copy Share Submit Comment Search all data

PROPERTIES	Other
Solprin acid	Other
Solpyron	Other
Tasprin	Other
Toldex	Other
Triaminicin	Other
acide 2-(acetyloxy)benzoique	Other
11126-35-5 Deleted CAS-RN	Deleted
11126-37-7 Deleted CAS-RN	Deleted
2349-94-2 Deleted CAS-RN	Deleted
26914-13-6 Deleted CAS-RN	Deleted
98201-60-6 Deleted CAS-RN	Deleted

Showing 151 to 161 of 161 records

Discover.
About/Disclaimer
Accessibility
Privacy

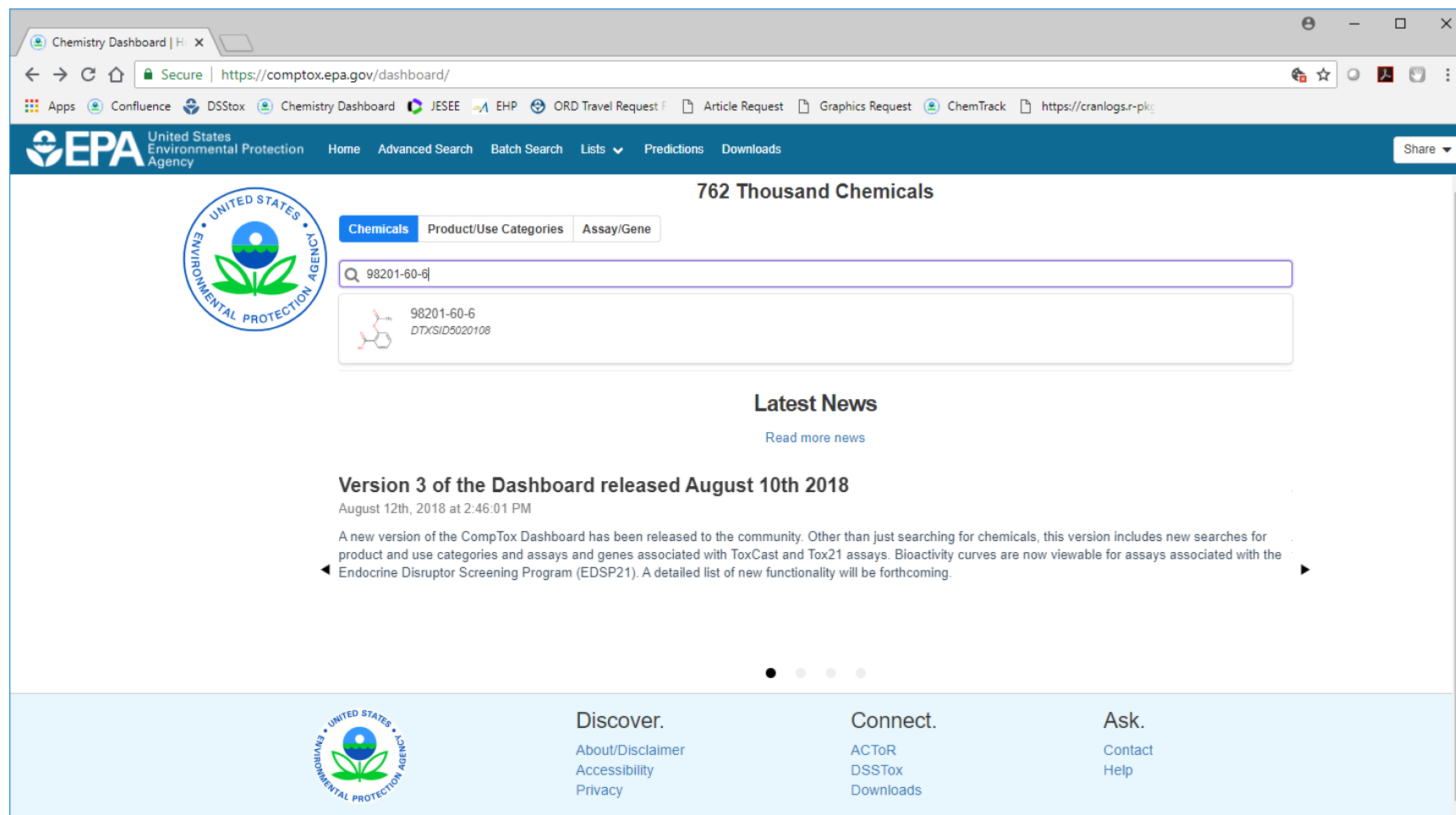
Connect.
ACToR
DSSTox
Downloads

Ask.
Contact
Help

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

Chemical Identity

- What chemicals are we talking about? Chemical Abstracts Service (CAS) Registry Numbers are not unique:
 - Aspirin (CAS #: 50-78-2)
 - Deleted CAS #: 2349-94-2, 11126-35-5, 11126-37-7, 26914-13-6, 98201-60-6

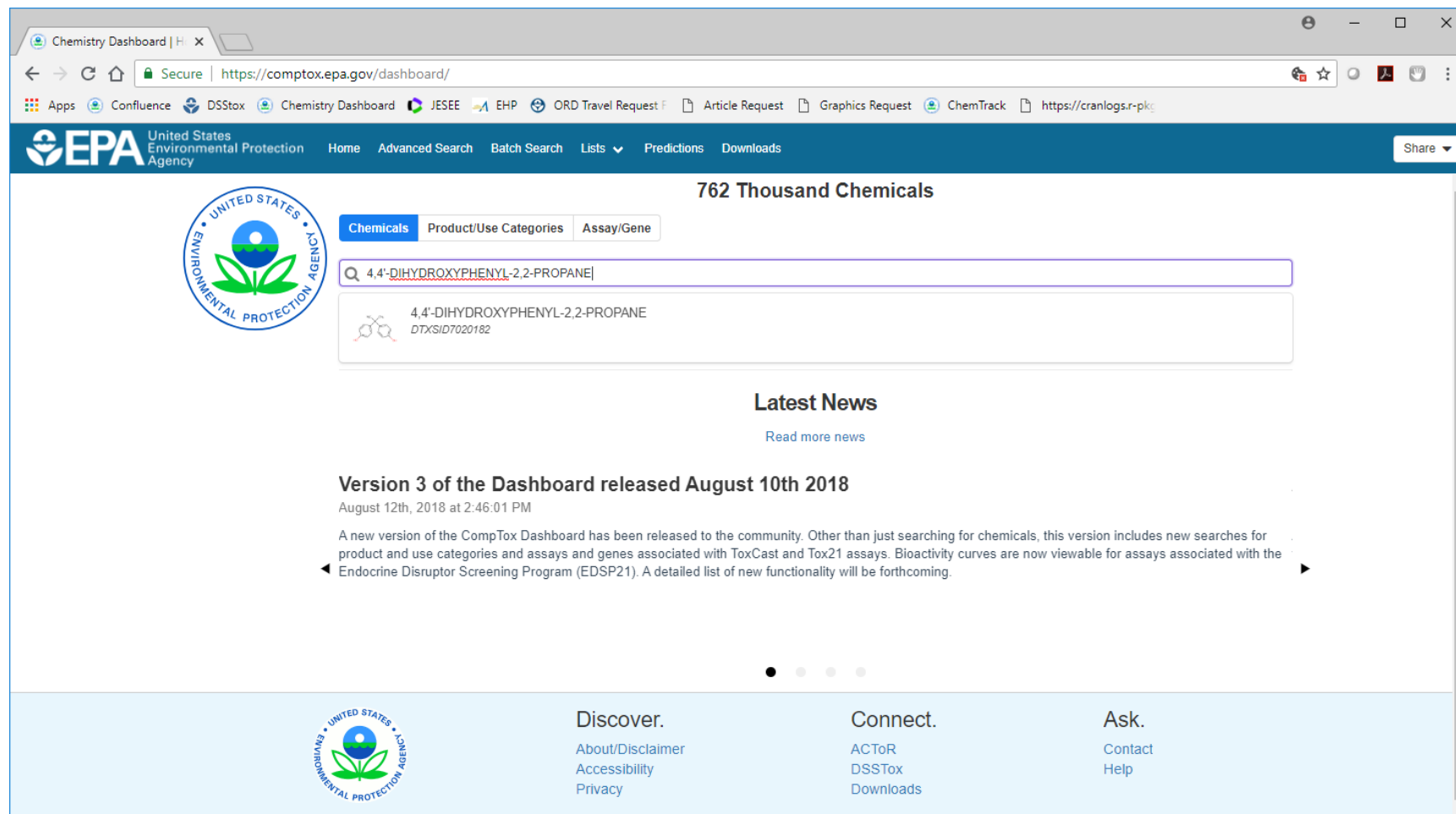


The screenshot shows the EPA Chemistry Dashboard interface. The browser address bar displays <https://comptox.epa.gov/dashboard/>. The page header includes the EPA logo and navigation links: Home, Advanced Search, Batch Search, Lists, Predictions, and Downloads. A search bar contains the text "98201-60-6". Below the search bar, a result is displayed for "98201-60-6" with the identifier "DTXSID5020108". The page also features a "Latest News" section with a headline "Version 3 of the Dashboard released August 10th 2018" and a sub-headline "August 12th, 2018 at 2:46:01 PM". The news text states: "A new version of the CompTox Dashboard has been released to the community. Other than just searching for chemicals, this version includes new searches for product and use categories and assays and genes associated with ToxCast and Tox21 assays. Bioactivity curves are now viewable for assays associated with the Endocrine Disruptor Screening Program (EDSP21). A detailed list of new functionality will be forthcoming." The footer contains the EPA logo and three columns of links: "Discover." (About/Disclaimer, Accessibility, Privacy), "Connect." (ACToR, DSSTox, Downloads), and "Ask." (Contact, Help).

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

Chemical Identity

- What chemicals are we talking about? Names are certainly not unique: *4,4'-DIHYDROXYPHENYL-2,2-PROPANE*



The screenshot shows the EPA Chemistry Dashboard interface. The browser address bar displays <https://comptox.epa.gov/dashboard/>. The dashboard header includes the EPA logo and navigation links: Home, Advanced Search, Batch Search, Lists, Predictions, and Downloads. A search bar on the left contains the text "4,4'-DIHYDROXYPHENYL-2,2-PROPANE". Below the search bar, the results show the chemical name and its DTXSID (DTXSID7020182). The dashboard also features a "Latest News" section with a headline "Version 3 of the Dashboard released August 10th 2018" and a brief description of the new version's features. The footer contains the EPA logo and links for "Discover.", "Connect.", and "Ask.".

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

Chemical Identity

- What chemicals are we talking about? Names are certainly not unique: *4,4'-DIHYDROXYPHENYL-2,2-PROPANE*

Chemistry Dashboard x

Secure | <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID7020182>

Apps Confluence DSStox Chemistry Dashboard JESEE EHP ORD Travel Request F Article Request Graphics Request ChemTrack <https://cranlogs.r-pkg.org>

EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads Copy Share Submit Comment Search all data

Bisphenol A

80-05-7 | DTXSID7020182
Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

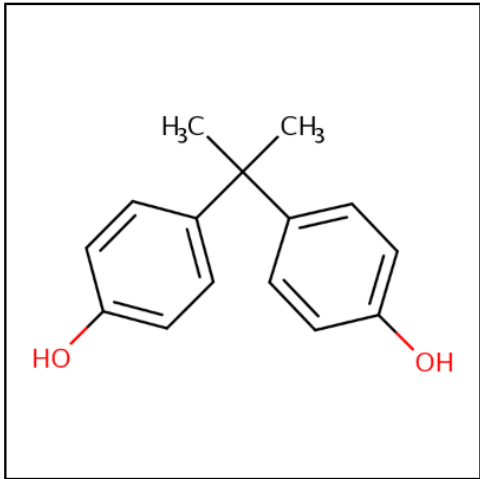
GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS



Wikipedia

Bisphenol A (BPA) is an organic synthetic compound with the chemical formula $(\text{CH}_3)_2\text{C}(\text{C}_6\text{H}_4\text{OH})_2$ belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water. It has been in commercial use since 1957.

BPA is a starting material for the synthesis of plastics, primarily

...
[Read more](#)

Intrinsic Properties

Molecular Formula: $\text{C}_{15}\text{H}_{16}\text{O}_2$ Mol File Find All Chemicals

Average Mass: 228.291 g/mol Isotope Mass Distribution

Monoisotopic Mass: 228.11503 g/mol

Structural Identifiers

Linked Substances

Presence in Lists

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

Chemical Lists

- Can find chemicals by lists

Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Lists of Chemicals

List of Assays

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

The screenshot displays the EPA Chemistry Dashboard interface. The top navigation bar includes links for Advanced Search, Batch Search, Lists, Predictions, and Downloads. The main content area is titled "Bisphenol A" with the identifier "80-05-7 | DTXSID7020182" and a note "Searched by DSSTox Substance Id.". A sidebar on the left provides a "DETAILS" menu with options like EXECUTIVE SUMMARY, PROPERTIES, ENV. FATE/TRANSPORT, HAZARD, ADME, EXPOSURE, BIOACTIVITY, SIMILAR COMPOUNDS, GENRA (BETA), RELATED SUBSTANCES, SYNONYMS, and LITERATURE. The central part of the page features a chemical structure of Bisphenol A, which consists of two phenol rings connected by a central carbon atom bonded to two methyl groups. The right-hand panel contains a "Wikipedia" section describing Bisphenol A (BPA) as an organic synthetic compound with the chemical formula $(\text{CH}_3)_2\text{C}(\text{C}_6\text{H}_4\text{OH})_2$, and an "Intrinsic Properties" section listing molecular formula, average mass, and monoisotopic mass.

Chemical Lists

- Can find chemicals by lists

Chemistry Dashboard | H X

Secure | https://comptox.epa.gov/dashboard/chemical_lists

Apps Confluence DSStox Chemistry Dashboard JESEE EHP ORD Travel Request F Article Request Graphics Request ChemTrack <https://cranlogs.r-pkg>

EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

Select List

Show 10 entries

Search:

Download

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
AEGLVALUES	Acute exposure guideline levels	2018-04-20	174	Acute exposure guideline levels (AEGLs) describe the human health effects from once-in-a-lifetime, or rare, exposure to airborne chemicals.
ALGALTOX	Algal Toxins	2017-11-21	54	A set of algal toxins of interest
ARCHEMICALS	Androgen Receptor Chemicals	2018-05-01	110	The list of chemicals used to identify references with in vitro AR binding . From Kleinstreuer et al http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.6b00347
ATHENSUS	University of Athens Surfactant and Suspect List	2017-07-14	60	ATHENSUS is a compilation of suspects, predicted transformation products and surfactants screened in wastewater by University of Athens, as described in Gago-Ferrero et al 2015, DOI: 10.1021/acs.est.5b03454
ATSDRLST	ATSDR Toxic Substances Portal Chemical List	2017-03-11	200	The Agency for Toxic Substances and Disease Registry (ATSDR) is a federal public health agency of the U.S. Department of Health and Human Services.
ATSDRMRLS	ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances	2018-05-02	756	The ATSDR Minimal Risk Levels (MRLs) were developed as an initial response to the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)
BISPHENOLS	Bisphenol Compounds	2018-01-09	52	This list represents a collection of Bisphenol Compounds
CCL4	CCL 4 Chemical Candidate List	2018-08-14	96	The Contaminant Candidate List (CCL) is a list of contaminants that are currently not subject to any proposed or promulgated national primary drinking water regulations, but are known or anticipated to occur in public water systems.
CERAPP	CERAPP: Collaborative Estrogen Receptor	2018-08-15	32290	CERAPP uses predictive computational models trained on HTS data to evaluate thousands

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

Chemical Lists

- Can find chemicals by lists

Chemistry Dashboard | https://comptox.epa.gov/dashboard/chemical_lists/bisphenols

Apps Confluence DSStox Chemistry Dashboard JESEE EHP ORD Travel Request F Article Request Graphics Request ChemTrack <https://cranlogs.r-pkg>

EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads Share Search all data

Bisphenol Compounds

Search BISPENOLS Chemicals

☐ Substring search

List Details

Description: This list represents a collection of Bisphenol Compounds, specifically Bisphenol A analogues

Number of Chemicals: 52

52 chemicals

Download / Send

Show info: DTXSID CASRN TOXCAST Select all

Sort by: DTXSID

Filter by: Name or CASRN Hide

Phenolphthalein

Dibhenolic acid

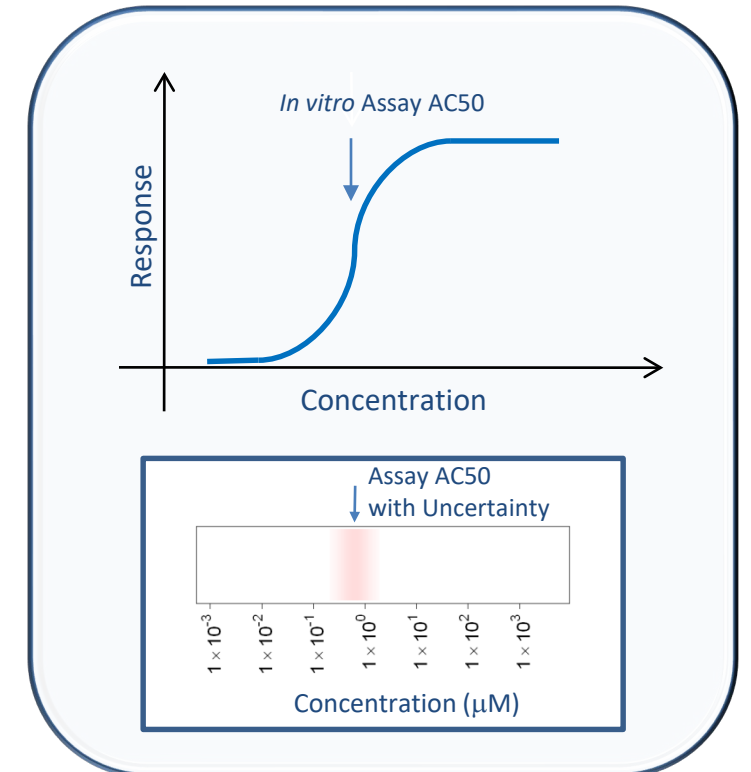
Bisphenol 4,4'-di(4-methyl-1-propenyl-2

4,4'-Sulphonediic(2,6-dibromophenol

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

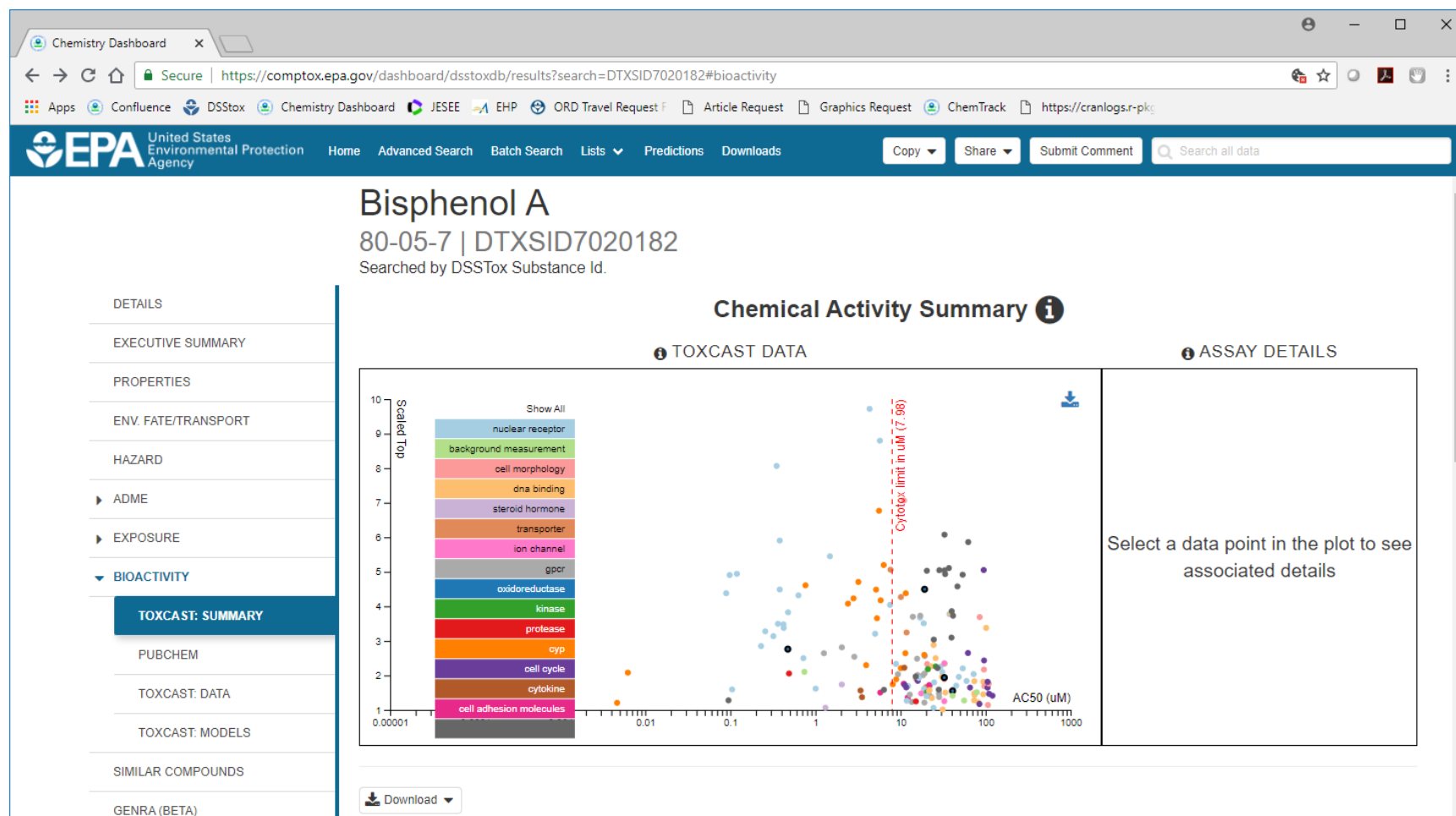
High-Throughput Bioactivity Screening

- The dashboard provides *in vivo* and *in vitro* toxicity data, where available
- **Most chemicals do not have *in vivo* toxicity data available (Judson, 2008)**
 - Bisphenol A vs. Bisphenol S
- **Tox21:** Examining >10,000 chemicals using ~50 assays intended to identify interactions with biological pathways (Schmidt, 2009)
- **EPA Toxicity Forecaster (ToxCast):**
 - For a subset (>3000) of Tox21 chemicals run >1000 additional assay endpoints (Judson et al., 2010)
- Data are being revised, new chemicals tested, new assays added



Chemical Bioactivity Data

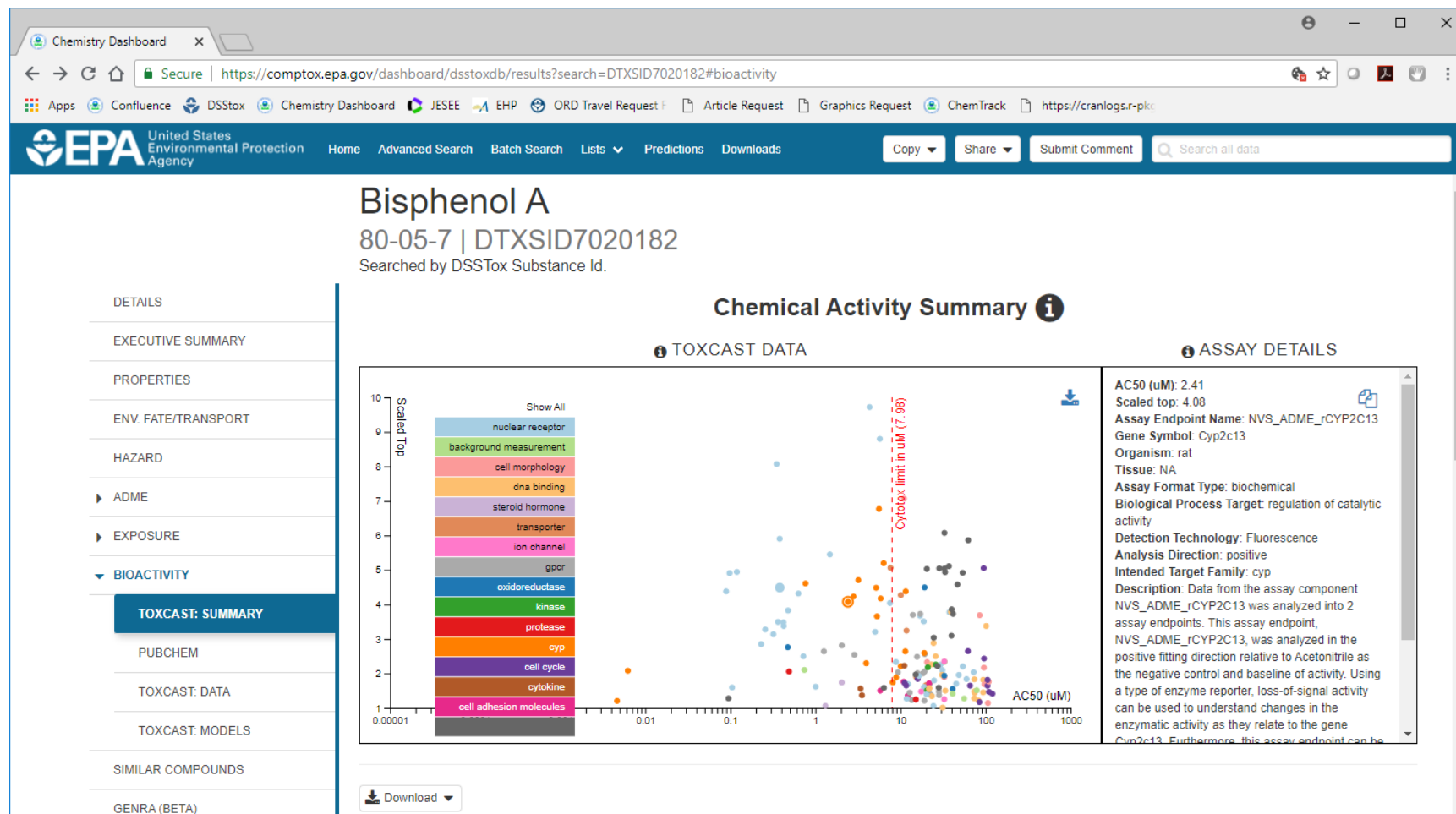
- Data from the ToxCast and Tox21 projects are available through the dashboard



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

Chemical Bioactivity Data

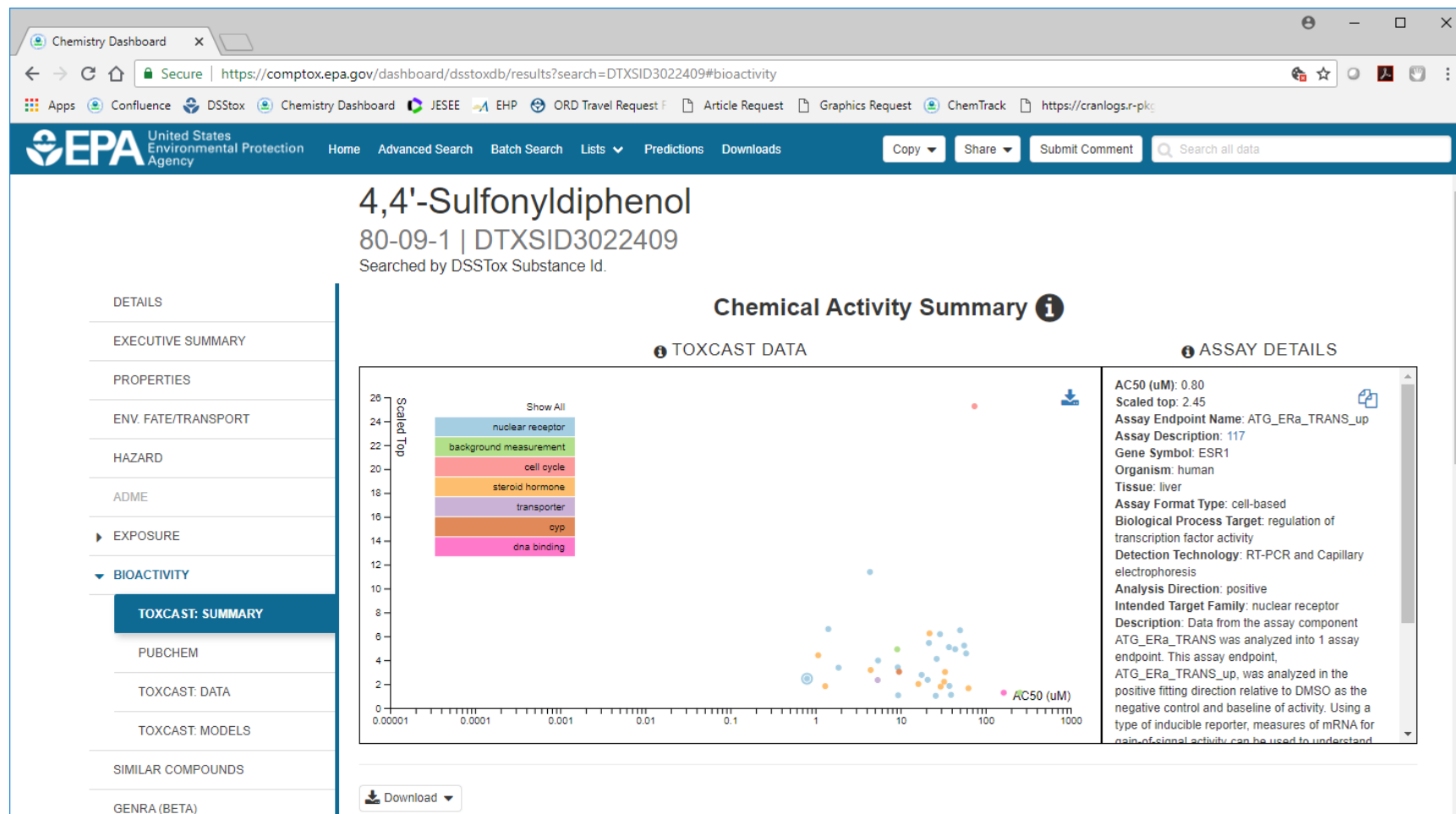
- Data from the ToxCast and Tox21 projects are available through the dashboard



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

Chemical Bioactivity Data

- Data from the ToxCast and Tox21 projects are available through the dashboard



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

Mapping Parent Chemicals to NHANES Analytes

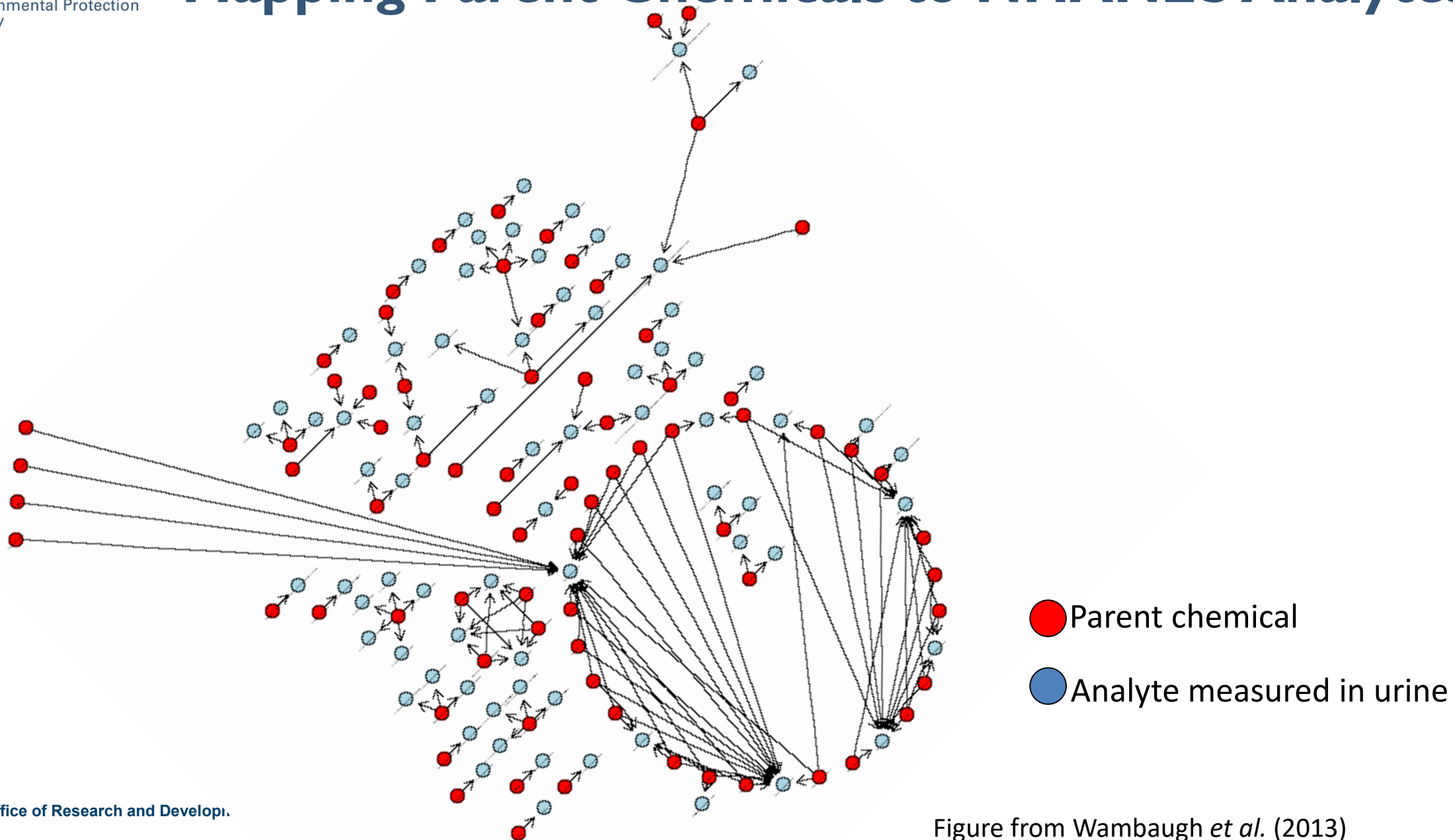


Figure from Wambaugh *et al.* (2013)

Parent-Metabolite Linkage

Chemistry Dashboard x

Secure | <https://comptox.epa.gov/dashboard/dsstoxdb/results?search=%20diazinon#related-substances>

Apps Confluence DSStox Chemistry Dashboard JESEE EHP ORD Travel Request F Article Request Graphics Request ChemTrack <https://cranlogs.r-pkg>

EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads Copy Share Submit Comment Search all data

Diazinon

333-41-5 | DTXSID9020407
Searched by Approved Name.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

LITERATURE

LINKS

5 chemicals

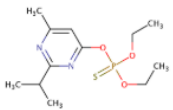
Download / Send

Show info: DTXSID CASRN TOXCAST Select all

Sort by: Relationship

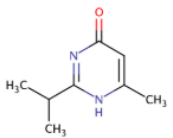
Filter by: Name or CASRN Hide

Searched Chemical



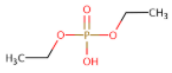
Diazinon
DTXSID: DTXSID9020407
CASRN: 333-41-5
TOXCAST: 52/717

Transformation Product




2-Isopropyl-6-methyl-4-pyrimidone
DTXSID: DTXSID1027502
CASRN: 2814-20-2
TOXCAST: 1/276

Transformation Product




Diethyl hydrogen phosphate
DTXSID: DTXSID1044699
CASRN: 598-02-7
TOXCAST: 0

Transformation Product



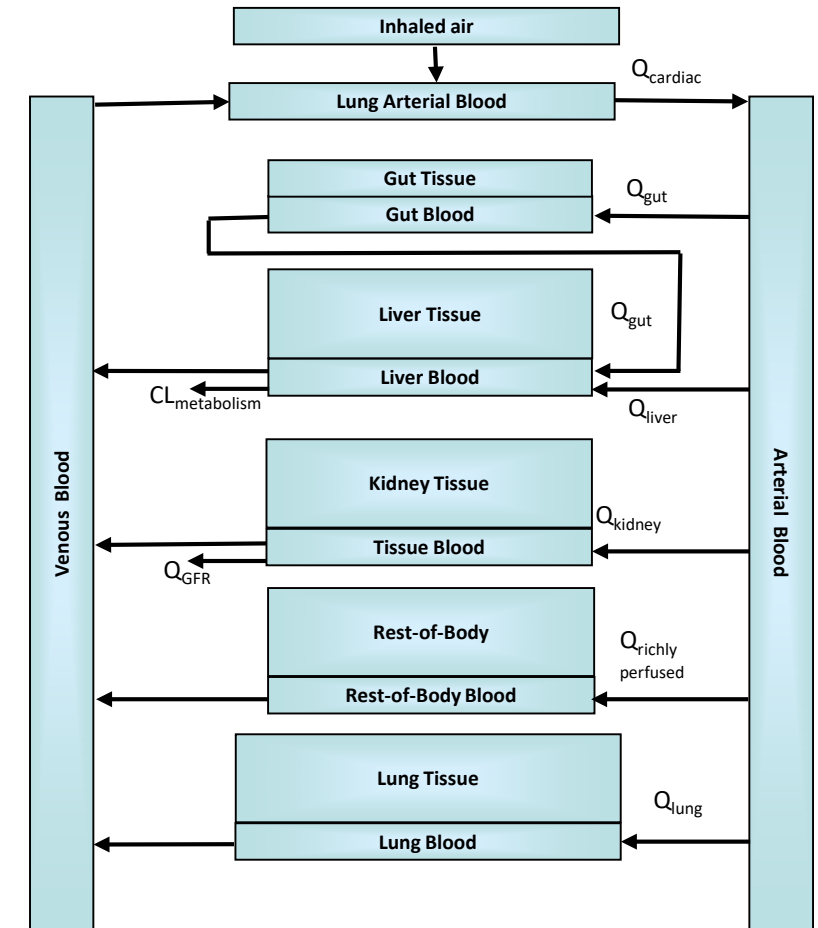
Transformation Product



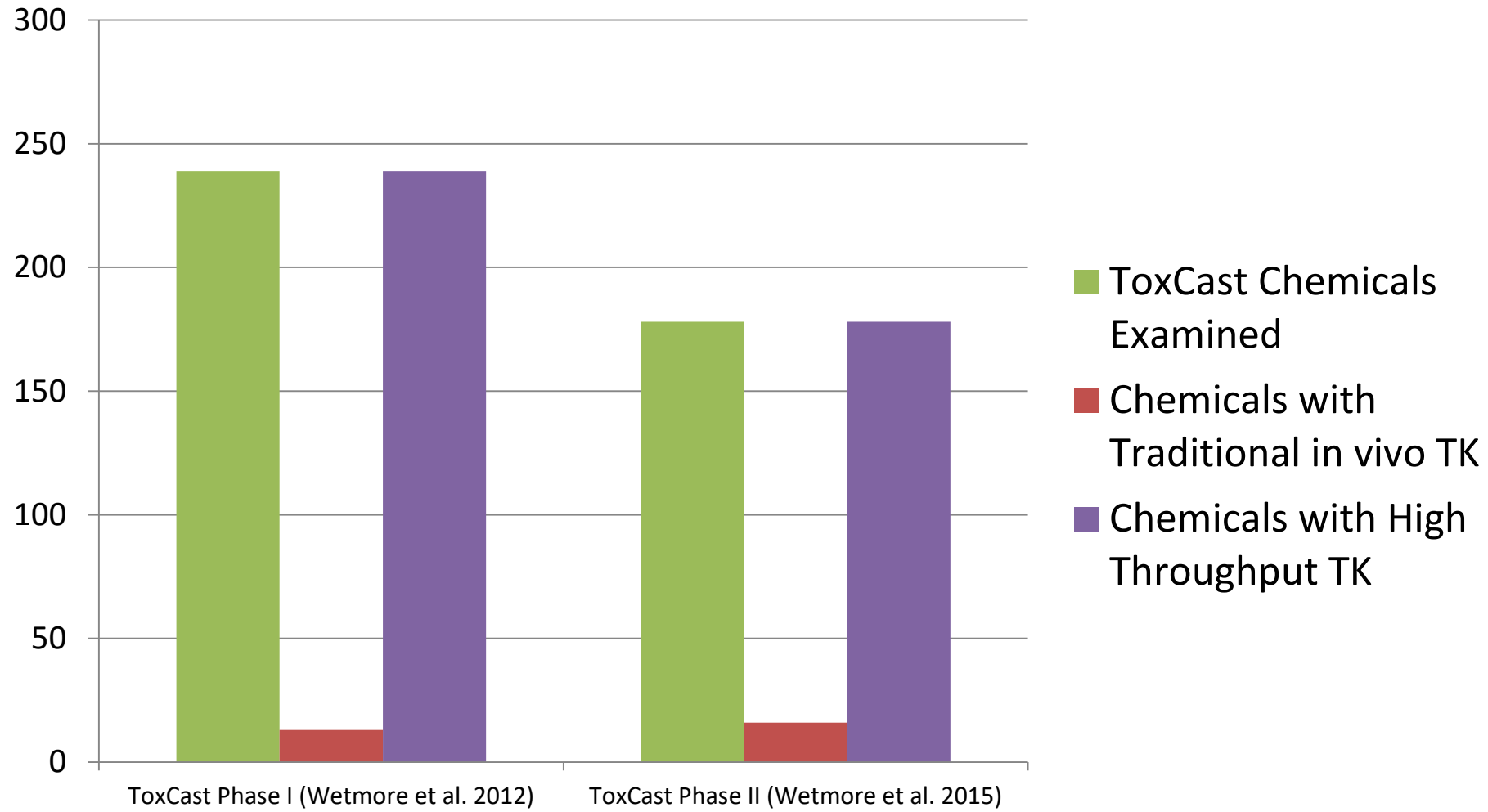
<https://comptox.epa.gov/dashboard/>

Chemical Toxicokinetics

- Toxicokinetics (TK) provides a bridge between toxicity and exposure assessment by predicting tissue concentrations due to exposure
 - However traditional TK methods are resource intensive
- Need to understand what the human body does with a chemical:
 - absorption, distribution, metabolism, excretion (ADME)
- Can relate *in vitro* bioactive concentrations (μM) to steady-state human doses (mg/kg body weight/day) using reverse toxicokinetics (Wetmore et al., 2012 and 2015):
 - You divide by the steady-state plasma concentration, C_{ss} , to convert μM to mg/kg body weight/day
 - Dashboard will give you C_{ss} predicted by HHTK

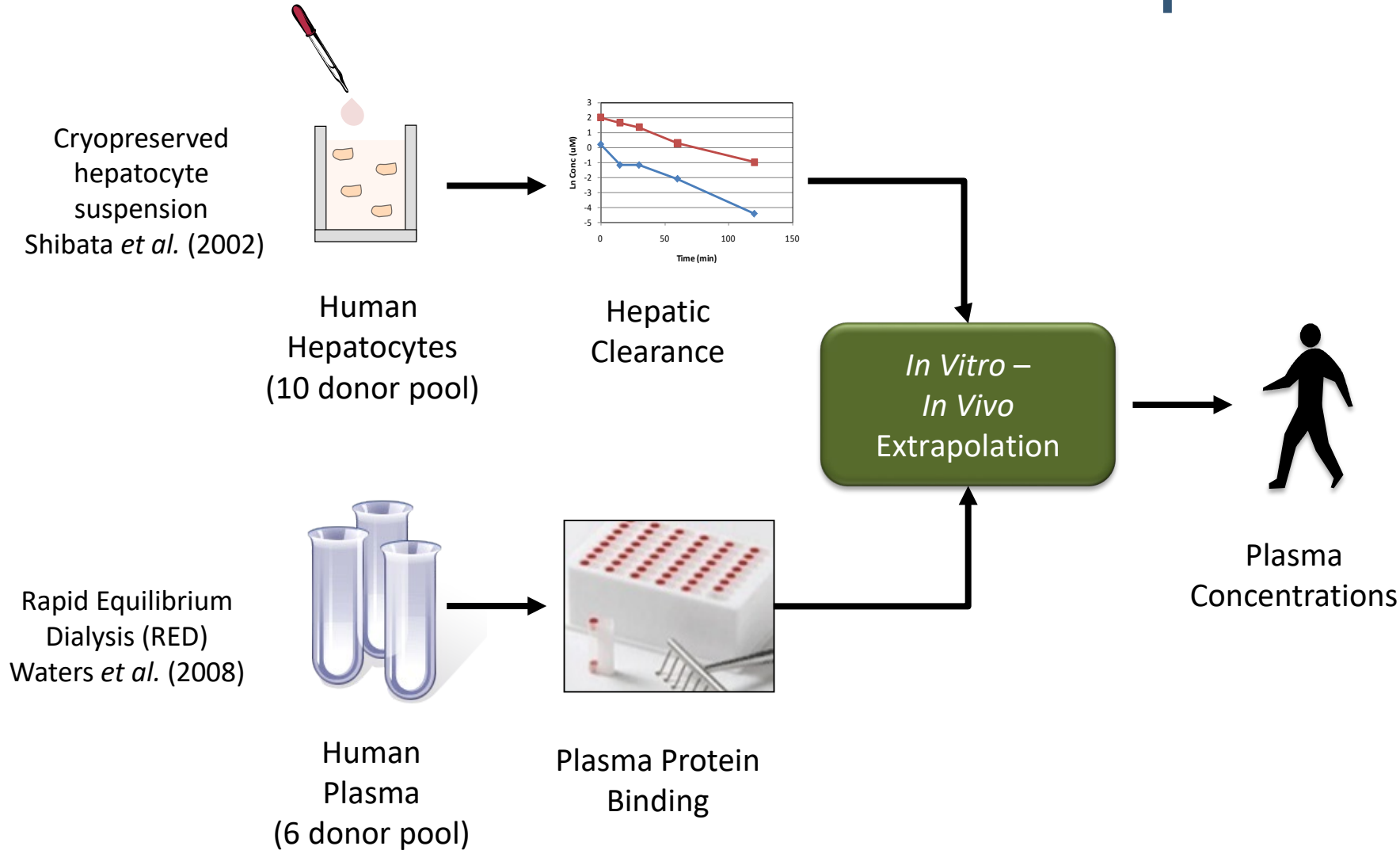


The Need for *In Vitro* Toxicokinetics



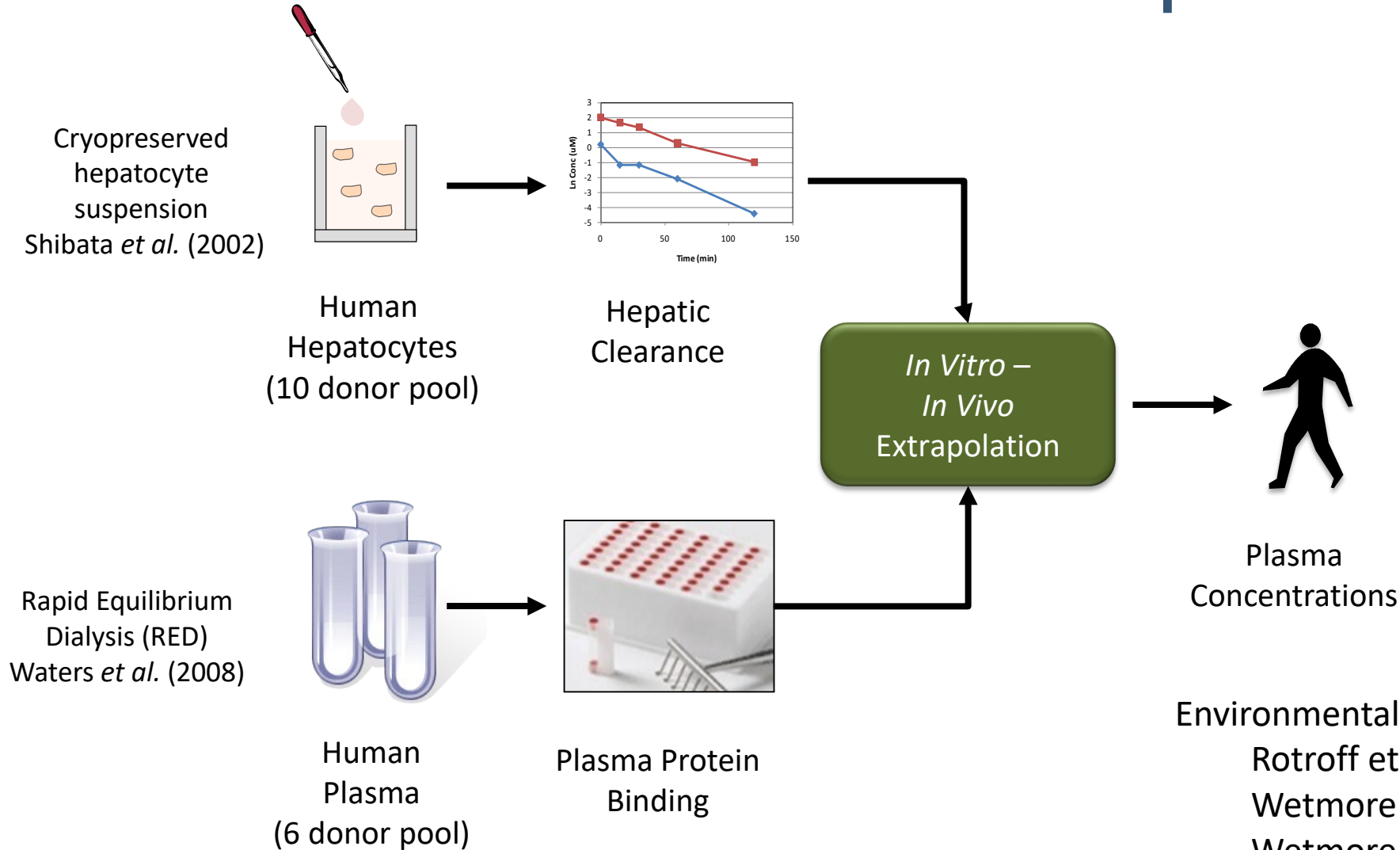
- Studies like Wetmore et al. (2012, 2015), addressed the need for TK data using *in vitro* methods

High-Throughput Toxicokinetics (HTTK) for *In Vitro-In Vivo* Extrapolation (IVIVE)



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

High-Throughput Toxicokinetics (HTTK) for *In Vitro-In Vivo* Extrapolation (IVIVE)



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

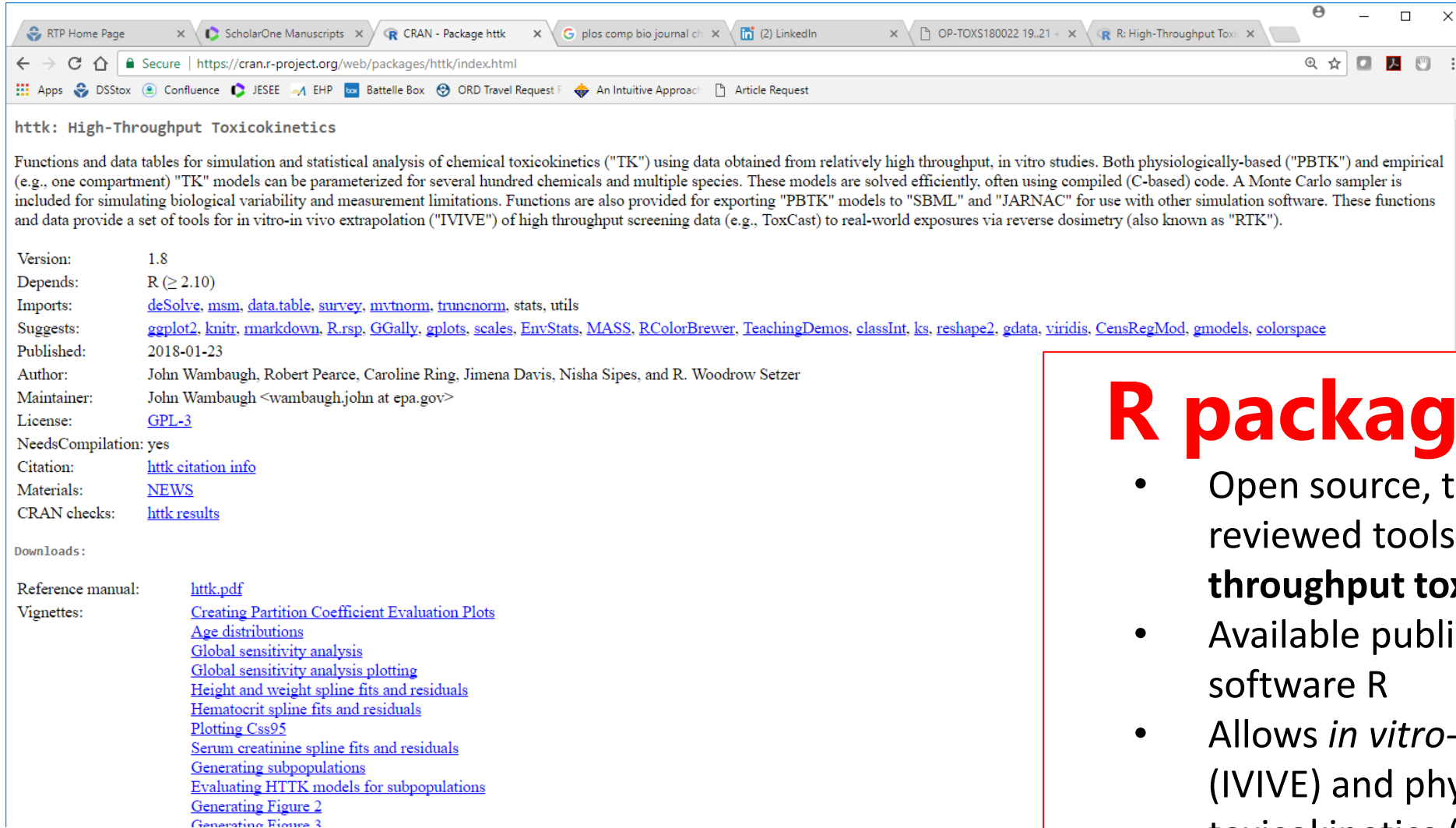
Environmental chemicals:

Rotroff *et al.* (2010) 35 chemicals
Wetmore *et al.* (2012) +204 chemicals
Wetmore *et al.* (2015) +163 chemicals
Wambaugh *et al.* (in prep.) + ~300 chemicals

Figure from Barbara Wetmore

Open Source Tools and Data for HTTK

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



The screenshot shows a web browser window with multiple tabs. The active tab is "CRAN - Package httk". The address bar shows the URL "https://cran.r-project.org/web/packages/httk/index.html". The page content includes the package name "httk: High-Throughput Toxicokinetics", a description of its functions, and a list of metadata.

httk: High-Throughput Toxicokinetics

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

Version: 1.8
Depends: R (≥ 2.10)
Imports: [deSolve](#), [msm](#), [data.table](#), [survey](#), [mvtnorm](#), [truncnorm](#), stats, utils
Suggests: [ggplot2](#), [knitr](#), [rmarkdown](#), [R.rsp](#), [GGally](#), [gplots](#), [scales](#), [EnvStats](#), [MASS](#), [RColorBrewer](#), [TeachingDemos](#), [classInt](#), [ks](#), [reshape2](#), [gdata](#), [viridis](#), [CensRegMod](#), [gmodels](#), [colorspace](#)
Published: 2018-01-23
Author: John Wambaugh, Robert Pearce, Caroline Ring, Jimena Davis, Nisha Sipes, and R. Woodrow Setzer
Maintainer: John Wambaugh <wambaugh.john at epa.gov>
License: [GPL-3](#)
NeedsCompilation: yes
Citation: [httk citation info](#)
Materials: [NEWS](#)
CRAN checks: [httk results](#)

Downloads:

Reference manual: [httk.pdf](#)
Vignettes: [Creating Partition Coefficient Evaluation Plots](#), [Age distributions](#), [Global sensitivity analysis](#), [Global sensitivity analysis plotting](#), [Height and weight spline fits and residuals](#), [Hematocrit spline fits and residuals](#), [Plotting C_{ss}95](#), [Serum creatinine spline fits and residuals](#), [Generating subpopulations](#), [Evaluating HTTK models for subpopulations](#), [Generating Figure 2](#), [Generating Figure 3](#)

R package "httk"

- Open source, transparent, and peer-reviewed tools and data for **high throughput toxicokinetics (httk)**
- Available publicly for free statistical software R
- Allows *in vitro-in vivo* extrapolation (IVIVE) and physiologically-base toxicokinetics (PBTk)

Physicochemical Properties

- Measured and predicted physicochemical properties are available
 - OPEn structure–activity/property Relationship App (OPERA) by Mansouri, et al. (2018)

Chemistry Dashboard

Not secure | comptox-prod.epa.gov/dashboard/dsstoxdb/results?search=triclosan#properties

Apps Confluence DSStox Chemistry Dashboard JESEE EHP ORD Travel Request F Article Request Graphics Request ChemTrack https://cranlogs.r-pkg

EPA United States Environmental Protection Agency Home Advanced Search Batch Search Lists Predictions Downloads Copy Share Submit Comment Search all data

Triclosan

3380-34-5 | DTXSID5032498

Searched by Approved Name.

Property

Summary

Download

Columns

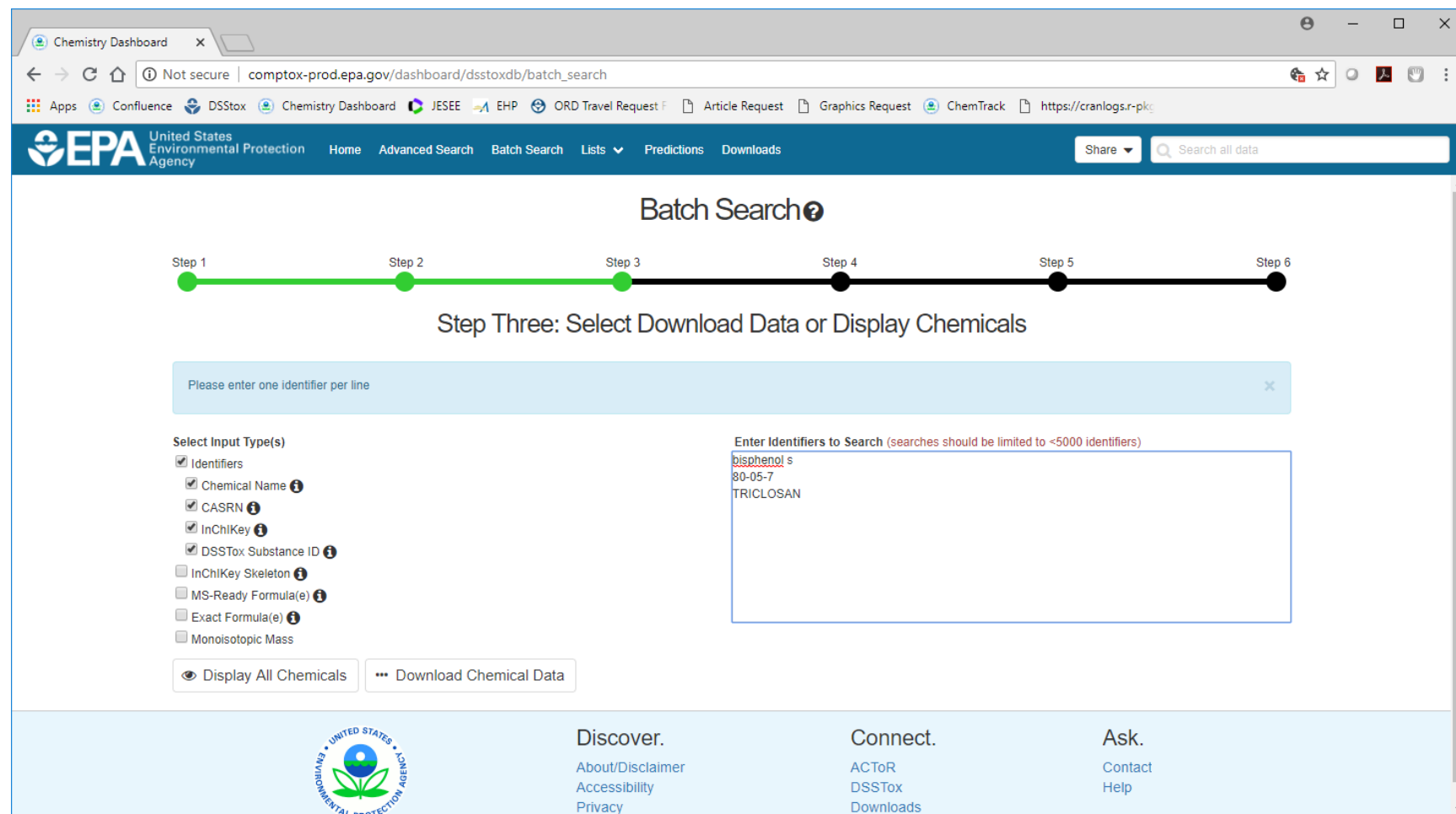
Search query

Property	Experimental average	Predicted average	Experimental median	Predicted median	Experimental range	Predicted range	Unit
LogP: Octanol-Water	4.76 (1)	4.79		5.06	4.76	3.78 to 5.27	
Melting Point	48.0 (11)	107	57.0	108	-40.0 to 57.5	73.4 to 141	°C
Boiling Point	-	352		345	-	327 to 374	°C
Water Solubility	3.45e-5 (1)	2.48e-5		2.15e-5	3.45e-5	4.54e-6 to 5.16e-5	mol/L
Vapor Pressure	-	9.72e-6		2.90e-6	-	4.53e-7 to 3.26e-5	mmHg
Flash Point	-	165		165	-	162 to 169	°C
Surface Tension	-	51.7		-	-	51.7	dyn/cm
Index of Refraction	-	1.63		-	-	1.63	
Molar Refractivity	-	69.3		-	-	69.3	cm³
Polarizability	-	27.5		-	-	27.5	Å³

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

Bulk Download of Data

- Can download databases and spreadsheets of data using Batch Search

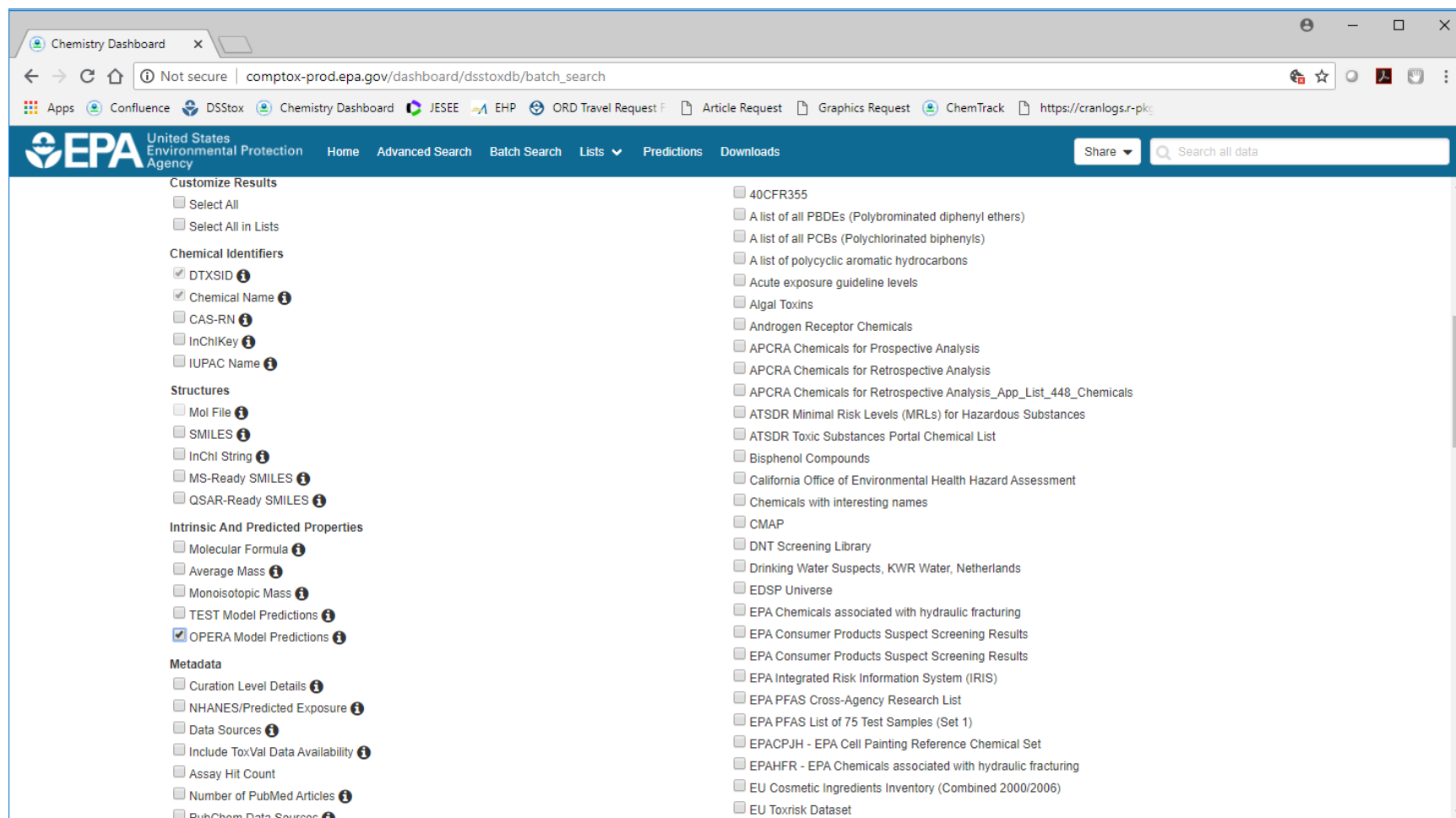


The screenshot shows the EPA Chemistry Dashboard's Batch Search interface. At the top, a progress bar indicates six steps, with Step 3, 'Select Download Data or Display Chemicals', currently active. Below the progress bar, a light blue box prompts the user to 'Please enter one identifier per line'. To the left, under 'Select Input Type(s)', several options are checked: Identifiers, Chemical Name, CASRN, InChIKey, and DSSTox Substance ID. To the right, a text area titled 'Enter Identifiers to Search' contains the text 'bisphenol s', '80-05-7', and 'TRICLOSAN'. At the bottom of the main content area, there are two buttons: 'Display All Chemicals' and 'Download Chemical Data'. The footer of the page includes the EPA logo, the text 'Discover.' with links to 'About/Disclaimer', 'Accessibility', and 'Privacy'; 'Connect.' with links to 'ACToR', 'DSSTox', and 'Downloads'; and 'Ask.' with links to 'Contact' and 'Help'.

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

Bulk Download of Data

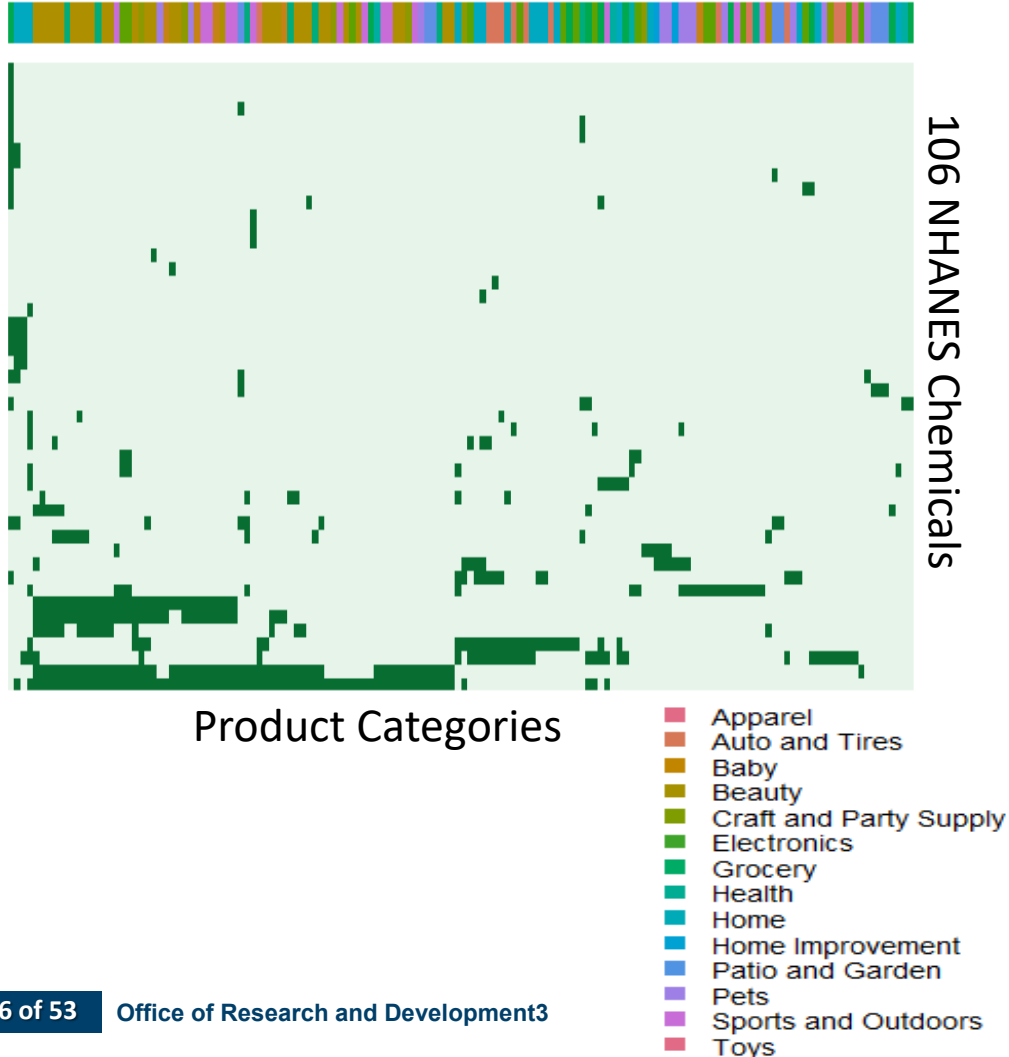
- Can download databases and spreadsheets of data using Batch Search



The screenshot shows the EPA Chemistry Dashboard's Batch Search interface. The browser address bar displays the URL: https://comptox-prod.epa.gov/dashboard/dsstoxdb/batch_search. The page features a navigation bar with links to Home, Advanced Search, Batch Search, Lists, Predictions, and Downloads. A search bar on the right allows users to search all data. The main content area is divided into two columns. The left column contains a 'Customize Results' section with checkboxes for 'Select All' and 'Select All in Lists'. Below this are sections for 'Chemical Identifiers' (including DTXSID, Chemical Name, CAS-RN, InChIKey, and IUPAC Name), 'Structures' (including Mol File, SMILES, InChI String, MS-Ready SMILES, and QSAR-Ready SMILES), 'Intrinsic And Predicted Properties' (including Molecular Formula, Average Mass, Monoisotopic Mass, TEST Model Predictions, and OPERA Model Predictions), and 'Metadata' (including Curation Level Details, NHANES/Predicted Exposure, Data Sources, Include ToxVal Data Availability, Assay Hit Count, Number of PubMed Articles, and PubChem Data Sources). The right column lists various data sources, each with a checkbox: 40CFR355, A list of all PBDEs (Polybrominated diphenyl ethers), A list of all PCBs (Polychlorinated biphenyls), A list of polycyclic aromatic hydrocarbons, Acute exposure guideline levels, Algal Toxins, Androgen Receptor Chemicals, APCRA Chemicals for Prospective Analysis, APCRA Chemicals for Retrospective Analysis, APCRA Chemicals for Retrospective Analysis_App_List_448_Chemicals, ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances, ATSDR Toxic Substances Portal Chemical List, Bisphenol Compounds, California Office of Environmental Health Hazard Assessment, Chemicals with interesting names, CMAP, DNT Screening Library, Drinking Water Suspects, KWR Water, Netherlands, EDSP Universe, EPA Chemicals associated with hydraulic fracturing, EPA Consumer Products Suspect Screening Results, EPA Consumer Products Suspect Screening Results, EPA Integrated Risk Information System (IRIS), EPA PFAS Cross-Agency Research List, EPA PFAS List of 75 Test Samples (Set 1), EPACPJH - EPA Cell Painting Reference Chemical Set, EPAHFR - EPA Chemicals associated with hydraulic fracturing, EU Cosmetic Ingredients Inventory (Combined 2000/2006), and EU Toxrisk Dataset.

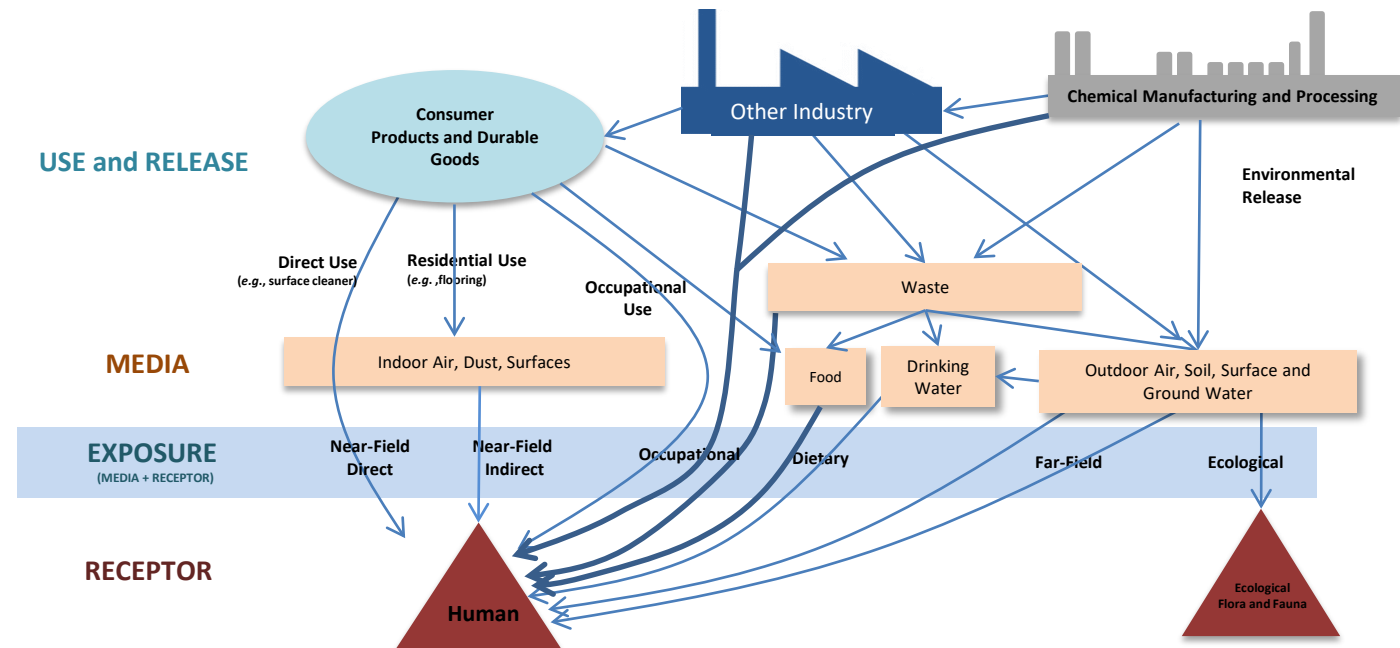
<https://comptox.epa.gov/dashboard/>

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



Chemical Sources

- Near field sources have been known to be important at least since 1987 – see Wallace, *et al.*
- *Hard to know what chemicals are in which materials*
- Dashboard provides this information, and will be addressed in depth in subsequent lectures



Knowledge of Exposure Pathways Limits High Throughput Exposure Models

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

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



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[†]

[†]Department of Public Health Sciences, University of California, Davis, California 95616, United States
[‡]Quantitative Sustainability Assessment Division, Department of Management Engineering, Technical University of Denmark, Kgs. Lyngby 2800, Denmark
[§]Department of Environmental Health Sciences, University of Michigan, Ann Arbor, Michigan 48109, United States
^{||}ARC Arnot Research and Consulting, Toronto, Ontario M4M 1W4, Canada
[⊥]Department of Physical and Environmental Sciences, University of Toronto, Scarborough, Toronto, Ontario M1C 1A4, Canada
[¶]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[∇]The Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina 27709, United States
[○]Harvard School of Public Health and School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States
[◆]Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States
[¶]School of Public Health, University of California, Berkeley, California 94720, United States

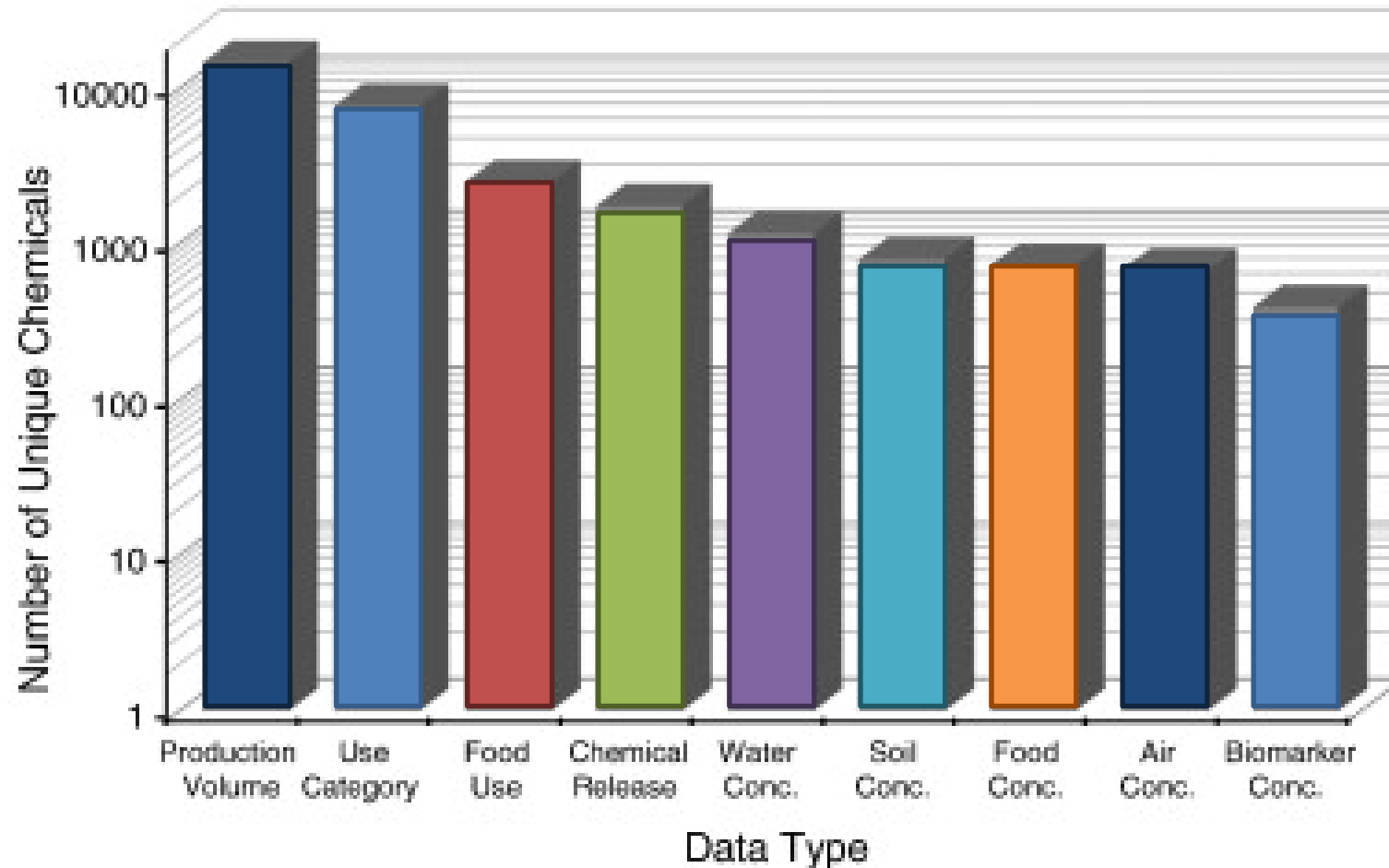
 Supporting Information

ABSTRACT: We present a risk-based high-throughput screening

Potential exposure from exposure Potential hazard from in vitro

Limited Available Data for Exposure Estimation

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



Can we develop new tools to generate the exposure information we need?

Decision Trees

WAMBAUGH ET AL. | 63

- Decision trees are a useful tool for making predictions of how something should be classified
- Unfortunately, they are unstable
Dietterich (2000)

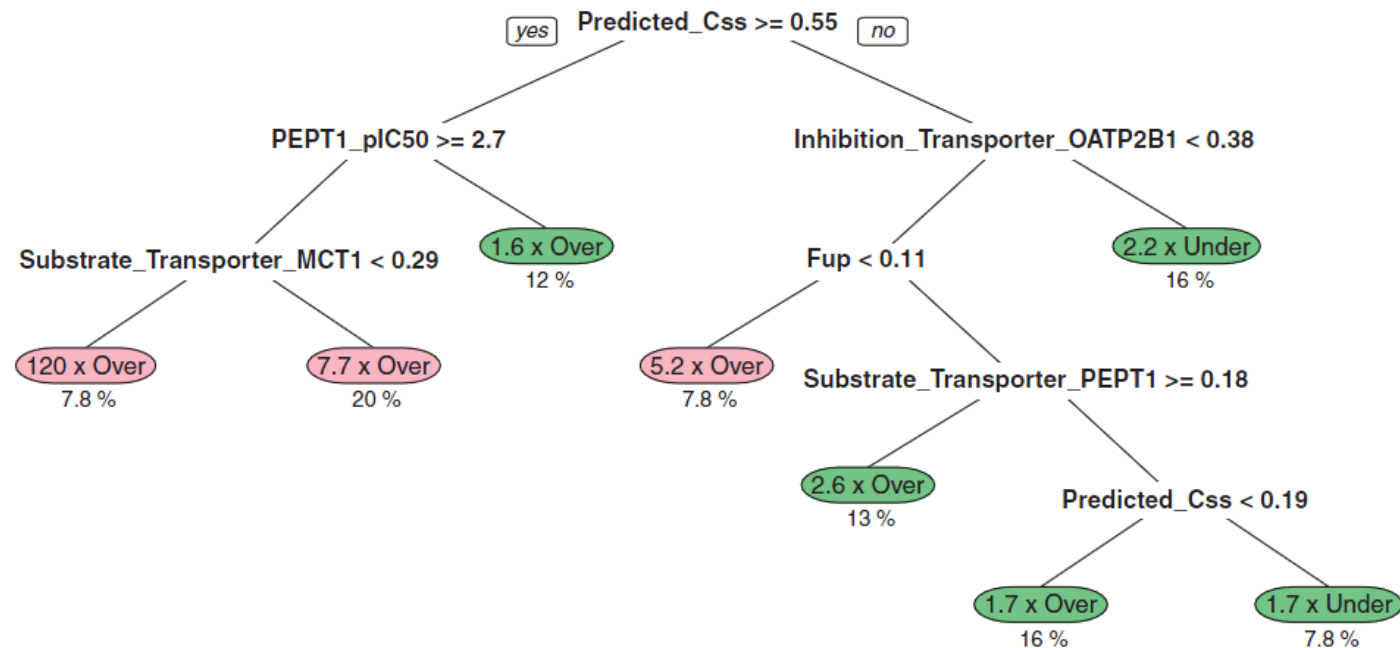
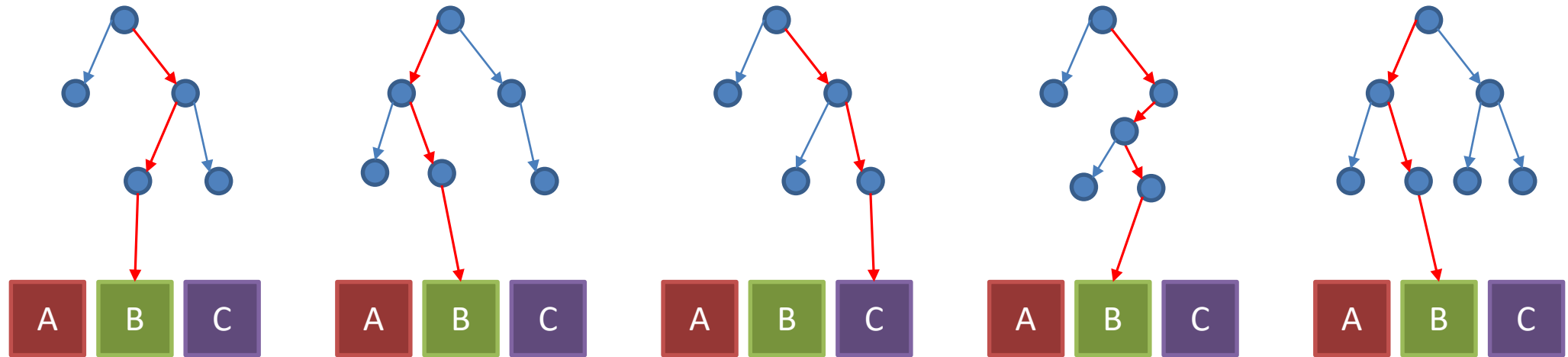


FIG. 5. A recursive partitioning regression tree was used to classify the discrepancy between the C_{ss} predicted from *in vitro* data and the *in vivo* C_{ss} (Obach *et al.*, 2008; Wetmore *et al.*, 2012). Each “leaf” of the tree shows a group of chemicals for which HTTK either overestimates C_{ss} (making conservative predictions) or underestimates C_{ss} . For all but 3 groups, the predictions are on the order of the observed C_{ss} (approximately within a factor of $3.2\times$ greater or lesser). For the other 3 groups, the C_{ss} is $5.2\times$, $7.7\times$, and $120\times$ overestimated. The dashed line indicates the identity (perfect predictor) line.

- In Wambaugh *et al.* (2015) various chemical properties, including Fraction unbound in plasma (F_{up}) and transporter affinities were used to predict whether C_{ss} would be over or underestimated

Ensemble Predictions

- “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)
- Every model gets a “vote” – can think of this probabilistically



Four votes for “B” – 80%

Bootstrap AGGREGatING: Bagging

- How do we get multiple decision trees? We use the method of Random Forests (Brieman, 2001)
- Construct multiple training sets that are subsets of the available data
- The models corresponding to each data subset each get a vote
- Estimate the error of each tree using the data not in the subset
 - Out of bag (OOB) error

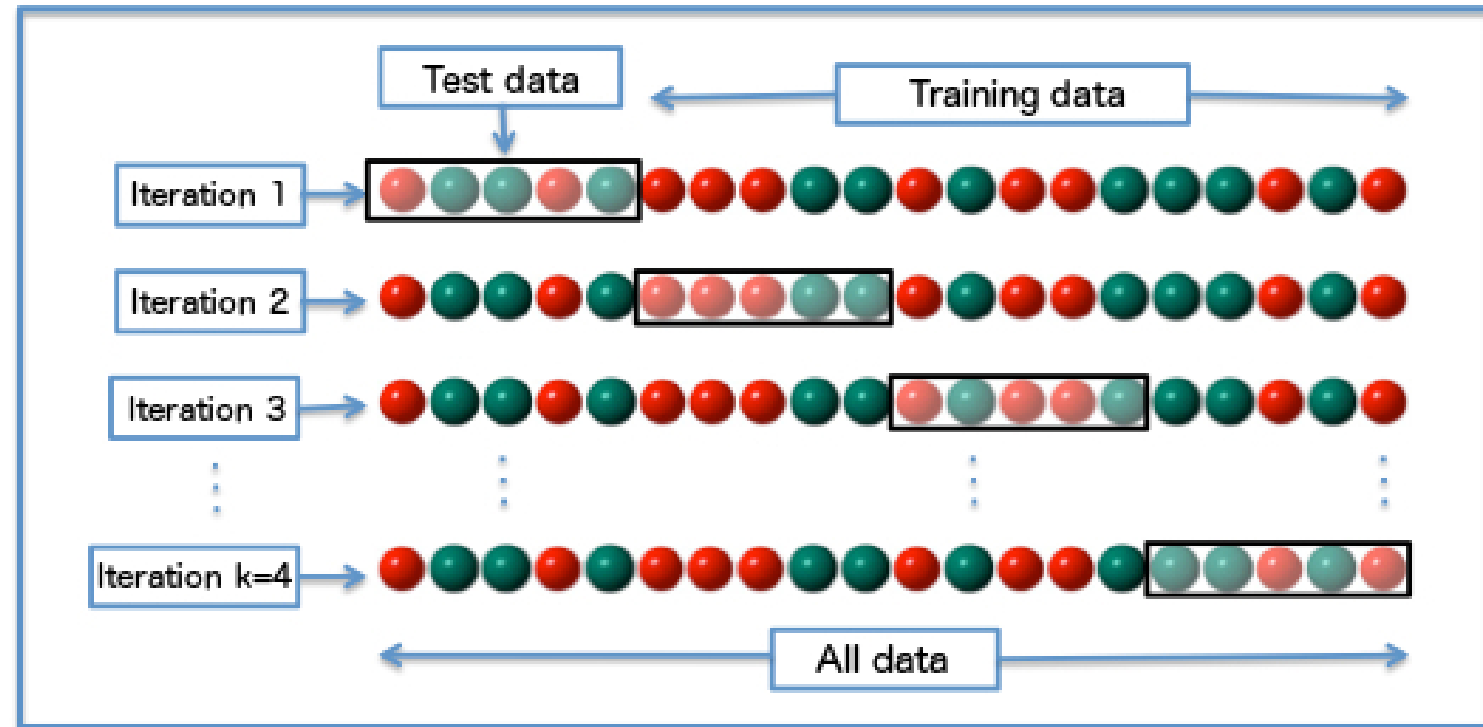


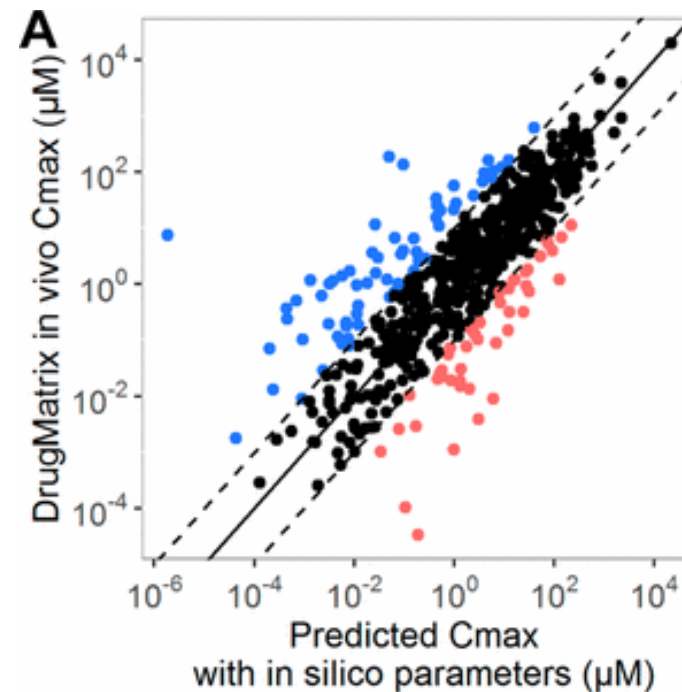
Image from Wikipedia article on “cross-validation”

Descriptor Subsets

- In Random Forests, each decision point only considers a random subset of the available predictors (Ho, 1995)
- Allows for large predictors sets, such as chemical structure chemotypes (thousands of structure features)
- Because only a subset of predictors are evaluated for each branching point in each tree, can evaluate how well do trees that include predictor X perform relative to trees that don't include predictor X
- This is a measure of predictor importance (Archer et al., 2008)
- Can tell us which parameters drive the model
- Need to be careful about correlated variables

Descriptor Subsets

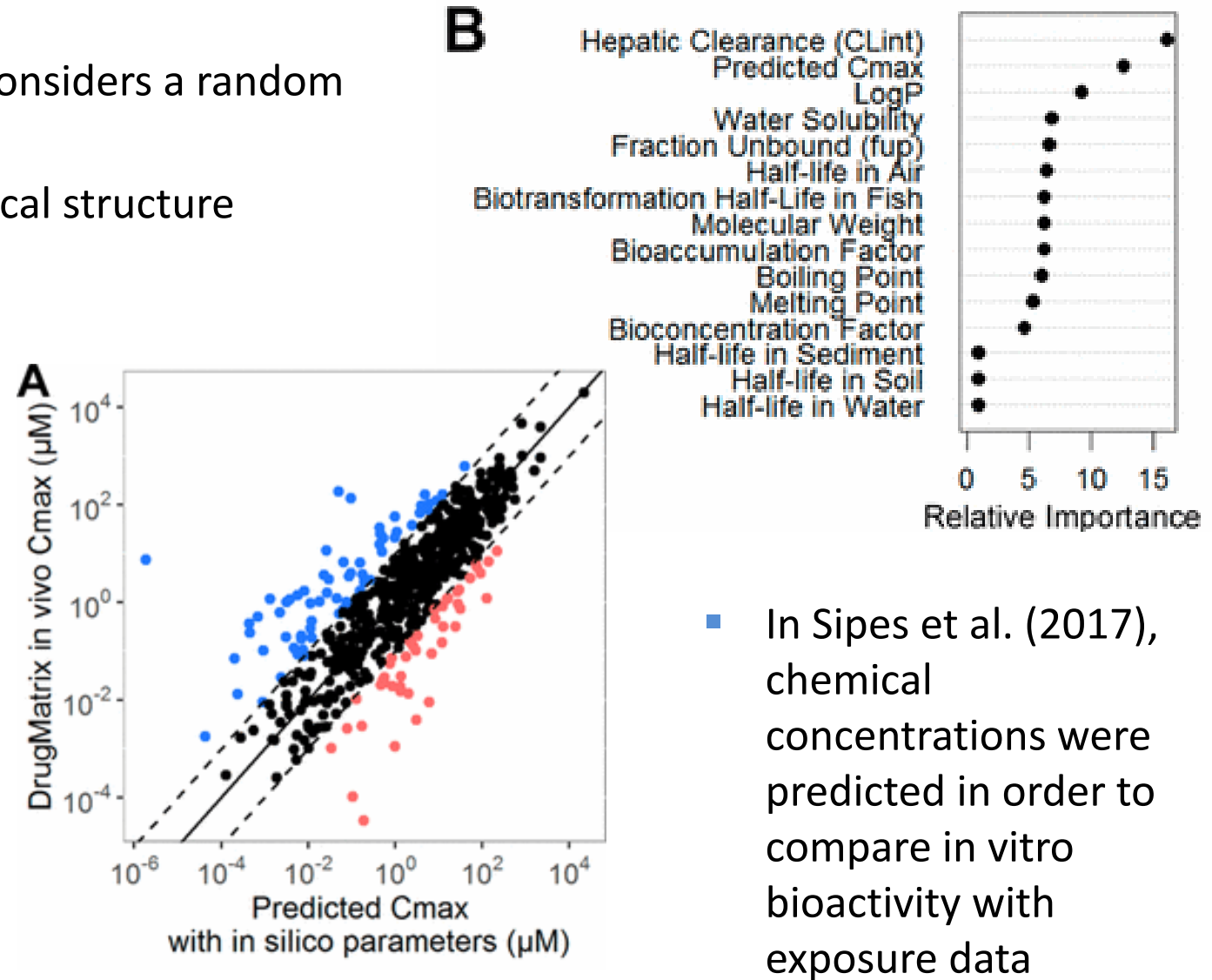
- In Random Forests, each decision point only considers a random subset of the available predictors (Ho, 1995)
- Allows for large predictors sets, such as chemical structure chemotypes (thousands of structure features)
- Because only a subset of predictors are evaluated for each branching point in each tree, can evaluate how well do trees that include predictor X perform relative to trees that don't include predictor X
- This is a measure of predictor importance (Archer et al., 2008)
- Can tell us which parameters drive the model
- Need to be careful about correlated variables



- In Sipes et al. (2017), chemical concentrations were predicted in order to compare in vitro bioactivity with exposure data

Descriptor Subsets

- In Random Forests, each decision point only considers a random subset of the available predictors (Ho, 1995)
- Allows for large predictors sets, such as chemical structure chemotypes (thousands of structure features)
- Because only a subset of predictors are evaluated for each branching point in each tree, can evaluate how well do trees that include predictor X perform relative to trees that don't include predictor X
- This is a measure of predictor importance (Archer et al., 2008)
- Can tell us which parameters drive the model
- Need to be careful about correlated variables



Model Performance

- In addition to OOB (out of bag) error rate, we can look at how well the average prediction of the models does in classifying the data: true positives (**TP**), false positives (**FP**), true negatives (**TN**), false negatives (**FN**)
- $\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$
- $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$
- Want to do both, balanced accuracy = $(\text{Sensitivity} + \text{Specificity}) / 2$
- Can use the option “sampsize” in randomForest R package to make sure that the training sets are balanced

	Actual Positives	Actual Negatives
Predicted Positives	TP	FP
Predicted Negatives	FN	TN

Confusion Matrix

R Code

- Liaw and Wiener's R package "randomForest" ported the original Random Forests Fortran by Leo Breiman and Adele Cutler into R
- Can do both classification and regression (we have not discussed regression much here)

Actual R code:

```
#Built in 1888 Swiss Fertility Data
```

```
help(swiss) # Display data information
```

R Documentation

Swiss Fertility and Socioeconomic Indicators (1888) Data

Description

Standardized fertility measure and socio-economic indicators for each of 47 French-speaking provinces of Switzerland at about 1888.

Usage

```
swiss
```

Format

A data frame with 47 observations on 6 variables, *each* of which is in percent, i.e., in *[0, 100]*.

[,1] Fertility	Ig, 'common standardized fertility measure'
[,2] Agriculture	% of males involved in agriculture as occupation
[,3] Examination	% draftees receiving highest mark on army examination
[,4] Education	% education beyond primary school for draftees.
[,5] Catholic	% 'catholic' (as opposed to 'protestant').
[,6] Infant.Mortality	live births who live less than 1 year.

All variables but 'Fertility' give proportions of the population.

Details

(paraphrasing Mosteller and Tukey):

Switzerland, in 1888, was entering a period known as the *demographic transition*; i.e., its fertility was beginning to fall from the high level typical of underdeveloped countries.

R Code

- Liaw and Wiener's R package "randomForest" ported the original Random Forests Fortran by Leo Breiman and Adele Cutler into R
- Can do both classification and regression (we have not discussed regression much here)

Actual R code:

```
swiss #display the data
```

RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
> swiss
```

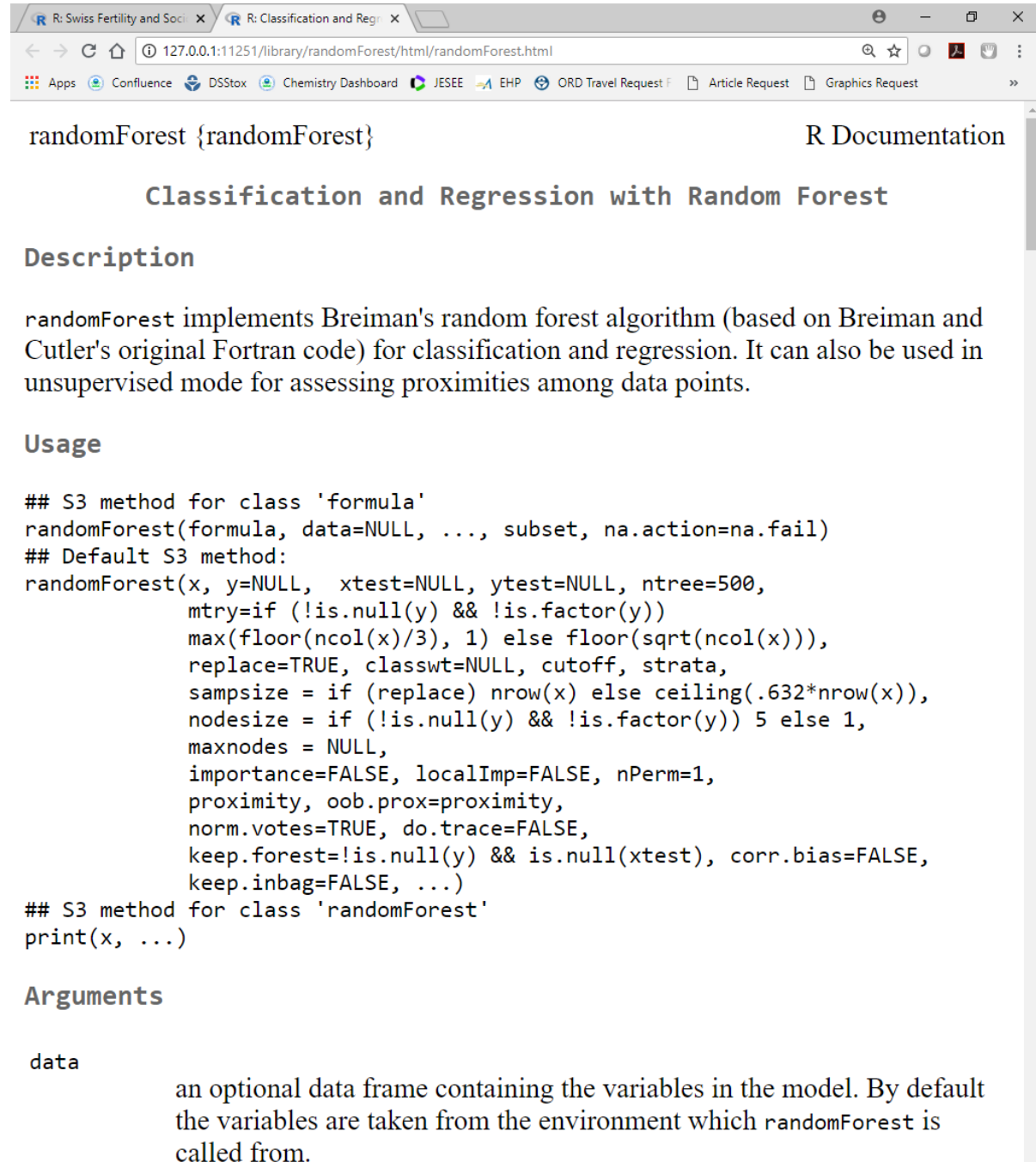
	Fertility	Agriculture	Examination	Education	Catholic	Infant.Mortality
Courtelary	80.2	17.0	15	12	9.96	22.2
Delemont	83.1	45.1	6	9	84.84	22.2
Franches-Mnt	92.5	39.7	5	5	93.40	20.2
Moutier	85.8	36.5	12	7	33.77	20.3
Neuveville	76.9	43.5	17	15	5.16	20.6
Porrentruy	76.1	35.3	9	7	90.57	26.6
Broye	83.8	70.2	16	7	92.85	23.6
Glane	92.4	67.8	14	8	97.16	24.9
Gruyere	82.4	53.3	12	7	97.67	21.0
Sarine	82.9	45.2	16	13	91.38	24.4
Veveyse	87.1	64.5	14	6	98.61	24.5
Aigle	64.1	62.0	21	12	8.52	16.5
Aubonne	66.9	67.5	14	7	2.27	19.1
Avenches	68.9	60.7	19	12	4.43	22.7
Cossonay	61.7	69.3	22	5	2.82	18.7
Echallens	68.3	72.6	18	2	24.20	21.2
Grandson	71.7	34.0	17	8	3.30	20.0
Lausanne	55.7	19.4	26	28	12.11	20.2
La Vallee	54.3	15.2	31	20	2.15	10.8
Lavaux	65.1	73.0	19	9	2.84	20.0
Morges	65.5	59.8	22	10	5.23	18.0
Moudon	65.0	55.1	14	3	4.52	22.4
Nyone	56.6	50.9	22	12	15.14	16.7
Orbe	57.4	54.1	20	6	4.20	15.3
Oron	72.5	71.2	12	1	2.40	21.0
Payerne	74.2	58.1	14	8	5.23	23.8
Paysd'enhaut	72.0	63.5	6	3	2.56	18.0
Rolle	60.5	60.8	16	10	7.72	16.3
Vevey	58.3	26.8	25	19	18.46	20.9
Yverdon	65.4	49.5	15	8	6.10	22.5
Conthey	75.5	85.9	3	2	99.71	15.1
Entremont	69.3	84.9	7	6	99.68	19.8
Herens	77.3	89.7	5	2	100.00	18.3
Martigwy	70.5	78.2	12	6	98.96	19.4
Monthey	79.4	64.9	7	3	98.22	20.2
St Maurice	65.0	75.9	9	9	99.06	17.8
--	--	--	--	--	--	--

R Code

- Liaw and Wiener's R package "randomForest" ported the original Random Forests Fortran by Leo Breiman and Adele Cutler into R
- Can do both classification and regression (we have not discussed regression much here)

Actual R code:

```
library(randomForest) #load Random Forest  
help(randomForest) # get help
```



The screenshot shows a web browser window displaying the documentation for the `randomForest` R package. The browser's address bar shows the URL `127.0.0.1:11251/library/randomForest/html/randomForest.html`. The page title is "randomForest {randomForest}" and it includes a link to "R Documentation". The main heading is "Classification and Regression with Random Forest". Below this, there is a "Description" section stating that `randomForest` implements Breiman's random forest algorithm for classification and regression. The "Usage" section shows the `randomForest` function signature and its default arguments. The "Arguments" section lists the `data` argument, describing it as an optional data frame.

randomForest {randomForest} [R Documentation](#)

Classification and Regression with Random Forest

Description

`randomForest` implements Breiman's random forest algorithm (based on Breiman and Cutler's original Fortran code) for classification and regression. It can also be used in unsupervised mode for assessing proximities among data points.

Usage

```
## S3 method for class 'formula'  
randomForest(formula, data=NULL, ..., subset, na.action=na.fail)  
## Default S3 method:  
randomForest(x, y=NULL, xtest=NULL, ytest=NULL, ntree=500,  
             mtry=if (!is.null(y) && !is.factor(y))  
               max(floor(ncol(x)/3), 1) else floor(sqrt(ncol(x))),  
             replace=TRUE, classwt=NULL, cutoff, strata,  
             sampsize = if (replace) nrow(x) else ceiling(.632*nrow(x)),  
             nodesize = if (!is.null(y) && !is.factor(y)) 5 else 1,  
             maxnodes = NULL,  
             importance=FALSE, localImp=FALSE, nPerm=1,  
             proximity, oob.prox=proximity,  
             norm.votes=TRUE, do.trace=FALSE,  
             keep.forest=!is.null(y) && is.null(xtest), corr.bias=FALSE,  
             keep.inbag=FALSE, ...)  
## S3 method for class 'randomForest'  
print(x, ...)
```

Arguments

data
an optional data frame containing the variables in the model. By default the variables are taken from the environment which `randomForest` is called from.

R Code

- Liaw and Wiener's R package "randomForest" ported the original Random Forests Fortran by Leo Breiman and Adele Cutler into R
- Can do both classification and regression (we have not discussed regression much here)

Actual R code:

```
# Set regions with below median fertility to
false, above median to true:
swiss$Fertility <-
swiss$Fertility>median(swiss$Fertility)
# Turn it into a "factor" which is how R
describes a classification
swiss$Fertility <-
as.factor(swiss$Fertility)
# Build a random forests model:
mdl <- randomForest(Fertility~.,data=swiss)
```

```
Call:
  randomForest(formula = Fertility ~ ., data = swiss)
      Type of random forest: classification
      Number of trees: 500
No. of variables tried at each split: 2

      OOB estimate of  error rate: 25.53%
Confusion matrix:
      FALSE TRUE class.error
FALSE      18      6  0.2500000
TRUE       6     17  0.2608696
> |
```

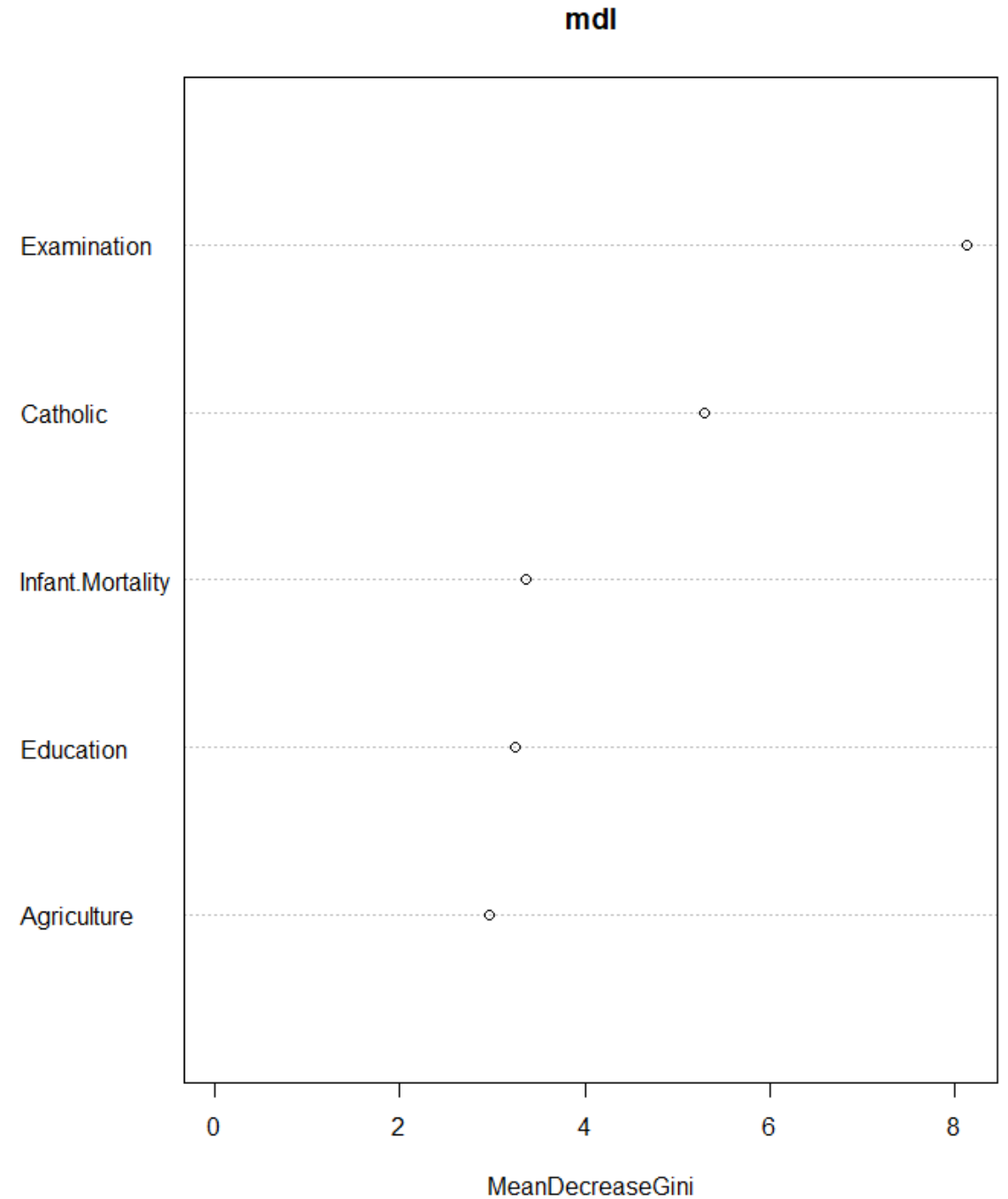

R Code

- Liaw and Wiener's R package "randomForest" ported the original Random Forests Fortran by Leo Breiman and Adele Cutler into R
- Can do both classification and regression (we have not discussed regression much here)

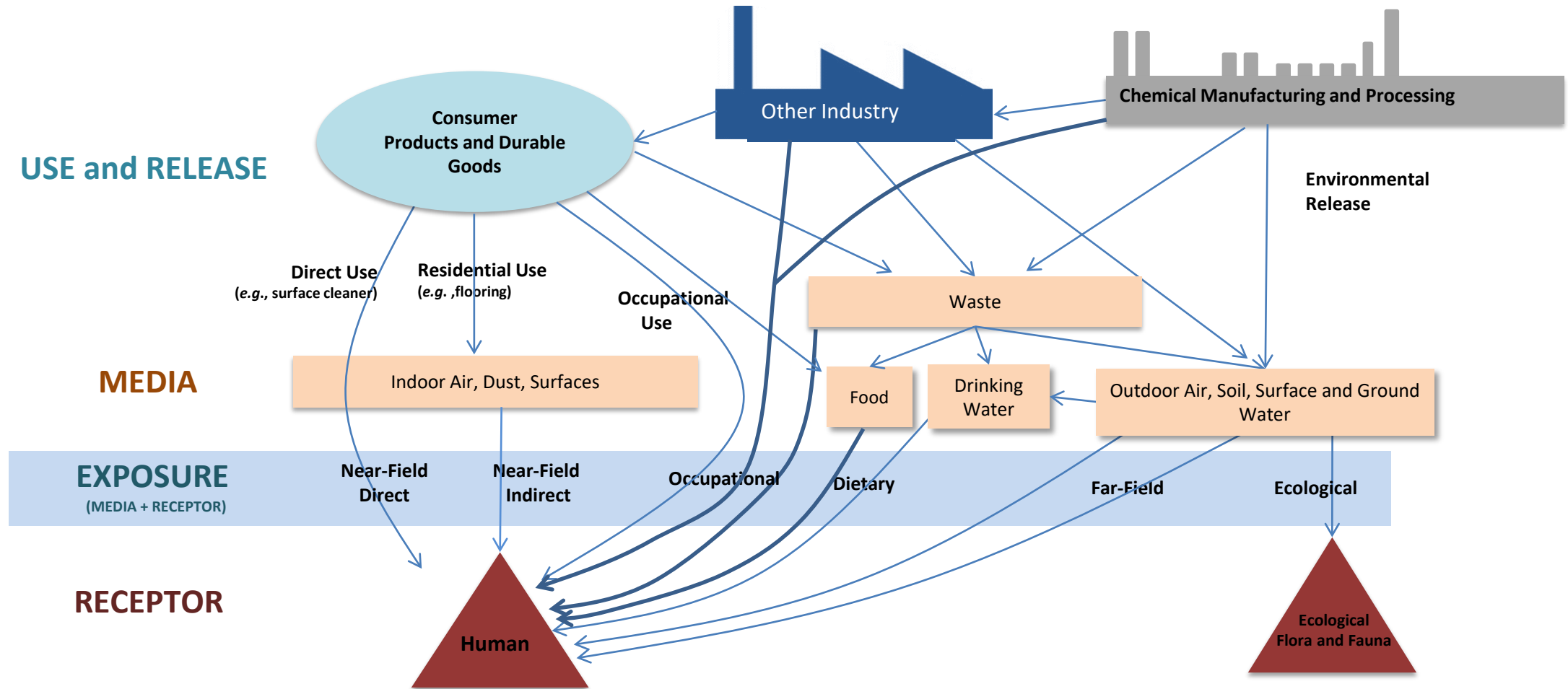
Actual R code:

```
varImpPlot(mdl) # Variable importance
```

Examination: percent of draftees
receiving highest mark on army
examination



Understanding Exposure is a Systems Problem



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

Finding the Right Data

We used Random Forests to relate chemical structure and properties to exposure pathway

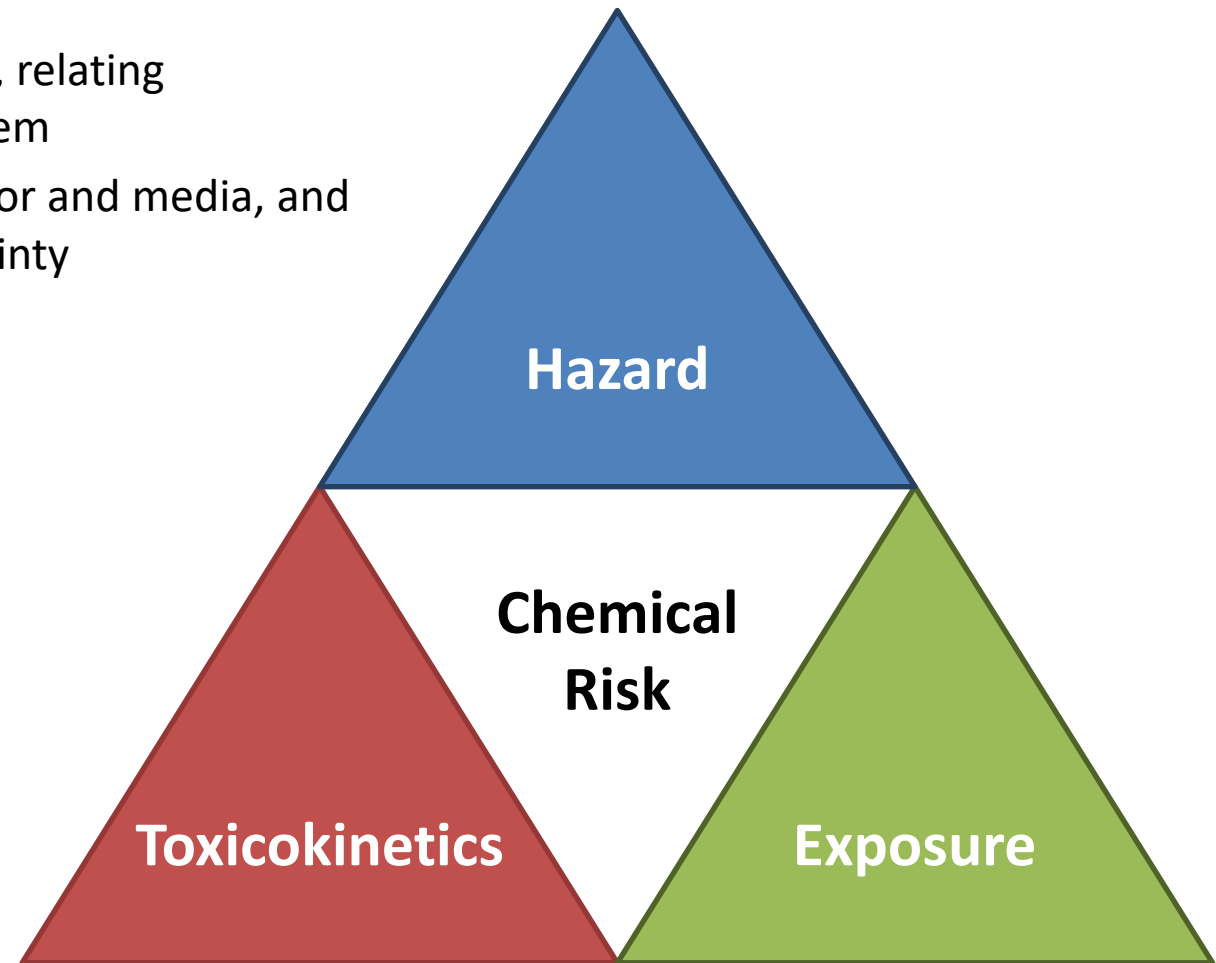
- Data curation is the rate limiting step for application of Random Forests
- Need a set of positive **and** negative examples with descriptors
- Thanks to various software tools and the speed of modern computers, once you have the data building the models is relatively easy

	NHANES Chemicals	Positives	Negatives	OOB Error Rate	Positives Error Rate	Balanced Accuracy	Sources of Positives	Sources of Negatives
Dietary	24	2523	8865	27	32	73	FDA CEDI, ExpoCast, CPDat (Food, Food Additive, Food Contact), NHANES Curation	Pharmapendium, CPDat (non-food), NHANES Curation
Near-Field	49	1622	567	26	24	74	CPDat (consumer_use, building_material), ExpoCast, NHANES Curation	CPDat (Agricultural, Industrial), FDA CEDI, NHANES Curation
Far-Field Pesticide	94	1480	6522	21	36	80	REDs, Swiss Pesticides, Stockholm Convention, CPDat (Pesticide), NHANES Curation	Pharmapendium, Industrial Positives, NHANES Curation
Far Field Industrial	42	5089	2913	19	16	81	CDR HPV, USGS Water Occurrence, NORMAN PFAS, Stockholm Convention, CPDat (Industrial, Industrial_Fluid), NHANES Curation	Pharmapendium, Pesticide Positives, NHANES Curation

- There are low levels of thousands of chemicals in commerce, relating exposures and health effects is an important unsolved problem
- The exposure pathway is the actual interaction of the receptor and media, and this event is often confounded by various sources of uncertainty

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





Chemical Safety for Sustainability (CSS) Research Program

Collaborators

Rapid Exposure and Dosimetry (RED) Project

NCCT

Chris Grulke
Greg Honda*
Richard Judson
Matthew Linakis*
Andrew McEachran*
Ann Richard
Risa Sayre*
Woody Setzer
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

Lead CSS Matrix Interfaces:

John Kenneke (NERL)
John Cowden (NCCT)

***Trainees**

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
Lesla Aylward
Technical University of Denmark
Peter Fantke
Tox Strategies
Caroline Ring
Miyoun Yoon
Unilever
Beate Nicol
Cecilie Rendal
Ian Sorrell
United States Air Force
Heather Pangburn
University of California, Davis
Deborah Bennett
University of Michigan
Olivier Jolliet
University of Texas, Arlington
Hyeong-Moo Shin

References

- Filer, Dayne L., et al. "tcpl: The ToxCast Pipeline for High-Throughput Screening Data." *Bioinformatics* (2016): btw680.
- Howgate, E., et al. "Prediction of in vivo drug clearance from in vitro data. I: impact of inter-individual variability" *Xenobiotica* 2006;36:473-497
- Israili and Dayton "Human Alpha-1-Glycoprotein and Its Interactions with Drugs" *Drug metabolism reviews* 2001;33:161-235
- Jamei, et al. "The Simcyp® population-based ADME simulator." *Expert opinion on drug metabolism & toxicology* 2009b;5:211-223
- Johnson, et al. "Prediction of the clearance of eleven drugs and associated variability in neonates, infants and children." *Clinical pharmacokinetics* (2006)
- Judson, R. S., et al., (2010) "In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environmental Health Perspectives* 118(4), 485-492.
- McNally, et al., "PopGen: a virtual human population generator." *Toxicology* *2014)
- 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*, (2017a)
- Pearce, Robert, et al. "Evaluation and Calibration of High-Throughput Predictions of Chemical Distribution to Tissues," *Journal of Pharmacokinetics and Pharmacodynamics* (2017b)
- Price et al., "Instructions for Use of Software Physiological Parameters for PBPK Modeling Version 1.3 (P3MTM 1.3)." 2003
- Ring, Caroline, et al., "Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability", *Environment International*, *in press*
- Rotroff, Daniel, et al., (2010) "Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening." *Tox. Sciences* 117(2), 348-58
- Routledge, P., "The plasma protein binding of basic drugs. *British journal of clinical pharmacology* 1986;22:499-506
- Sipes, Nisha, et al. "An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library", *Environmental Science and Technology* (2017)
- Wambaugh, John F., et al. "High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals." *Env. science & technology* (2014).
- Wambaugh, John F., et al. "Toxicokinetic triage for environmental chemicals." *Toxicological Sciences* (2015): kfv118.
- 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." *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.
- Yasuda, et al., "The role of ethnicity in variability in response to drugs: focus on clinical pharmacology studies." *Clinical Pharmacology & Therapeutics* 2008;84:417-423
- Yoon, M., et al. (2014). "Evaluation of simple in vitro to in vivo extrapolation approaches for environmental compounds." *Toxicology in Vitro* 28(2), 164-170,