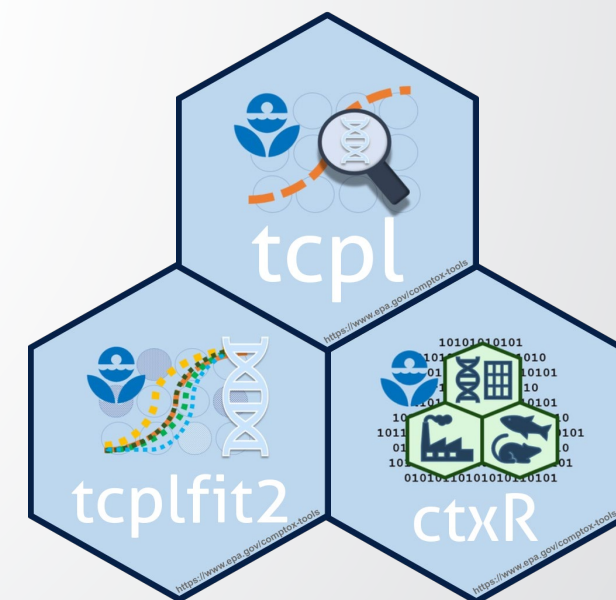




# Navigating Software and Database Updates to ToxCast: Targeted Bioactivity Data for Toxicology

ASCCT 13<sup>th</sup> Annual Meeting

October 29<sup>th</sup> 2024





# Outline & Disclaimer

- ToxCast Tools for Users: Biological Coverage
- Summary of Software & Database Updates: Past & Future of ToxCast's invitrodb
- Activity & Potency Estimates 101
- Exploring ToxCast: Accessing Data & Plotting for Different Needs
- ToxCast Tools for Users: Use Cases
- Q&A

*The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the US Environmental Protection Agency.*







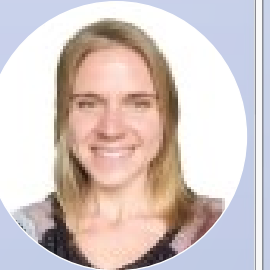


*Company or product names do not constitute endorsement by US EPA.*



# ToxCast Team



We are an interdisciplinary team (software developers, toxicologists, biologists, and statisticians) working alongside several internal/external collaborators (EPA labs, contract vendors, or other partners) to process data for a diverse stakeholder and user group with various research needs.

|  |  |  |  |   |   |   |   |  |
|--|--|--|--|---|---|---|---|--|
|  <p><b>Madison Feshuk</b></p> <ul style="list-style-type: none"><li>• <i>Lead, Pipelining and Curation</i></li></ul> |  <p><b>Jason Brown</b></p> <ul style="list-style-type: none"><li>• <i>Lead, tcpl Development</i></li></ul> |  <p><b>Sarah Davidson-Fritz</b></p> <ul style="list-style-type: none"><li>• <i>Lead, tcplfit2 Development and Statistics</i></li></ul> |  <p><b>Carter Thunes</b></p> <ul style="list-style-type: none"><li>• <i>tcpl Development and Pipelining</i></li><li>• <i>ORAU-SSC</i></li></ul> |  <p><b>Ashley Ko</b></p> <ul style="list-style-type: none"><li>• <i>Pipelining and Curation</i></li><li>• <i>ORAU-SSC</i></li></ul> |  <p><b>Paul Kruse</b></p> <ul style="list-style-type: none"><li>• <i>Lead, ctxR Development</i></li></ul> |  <p><b>Kelly Carstens</b></p> <ul style="list-style-type: none"><li>• <i>SME, DNT</i></li></ul> |  <p><b>Katie Paul Friedman</b></p> <ul style="list-style-type: none"><li>• <i>SME, ToxCast Project Lead</i></li></ul> |  <p><b>Colleen Elonen</b></p> <ul style="list-style-type: none"><li>• <i>SCDCD Project Liaison</i></li></ul> |
|--|--|--|--|---|---|---|---|--|

# Session Learning Objectives

- Examine the breadth of data available in the ToxCast program.
- Learn how to utilize assay annotation information to find and interpret data for particular targets.
- Understand changes in the ToxCast Pipeline including additional models, bidirectional curve-fitting, and continuous hit calling.
- Develop confidence in reviewing and using ToxCast data for different tasks, such as bioactivity data for weight-of-evidence or finding strong positives or negatives in an assay endpoint.
- Review how to access ToxCast data via the [CompTox Chemicals Dashboard](#), [CTX APIs](#), and [data downloads](#).





# ToxCast Tools for Users: Biological Coverage

Katie Paul Friedman



# Key Challenge: Too many chemicals, not enough data or time

## 1984 NAS Report

### Toxicity Testing Strategies to Determine Needs and Priorities

Steering Committee on Identification of Toxic and Potentially Toxic  
Chemicals for Consideration by the National Toxicology Program

Board on Toxicology and Environmental Health Hazards

Commission on Life Sciences

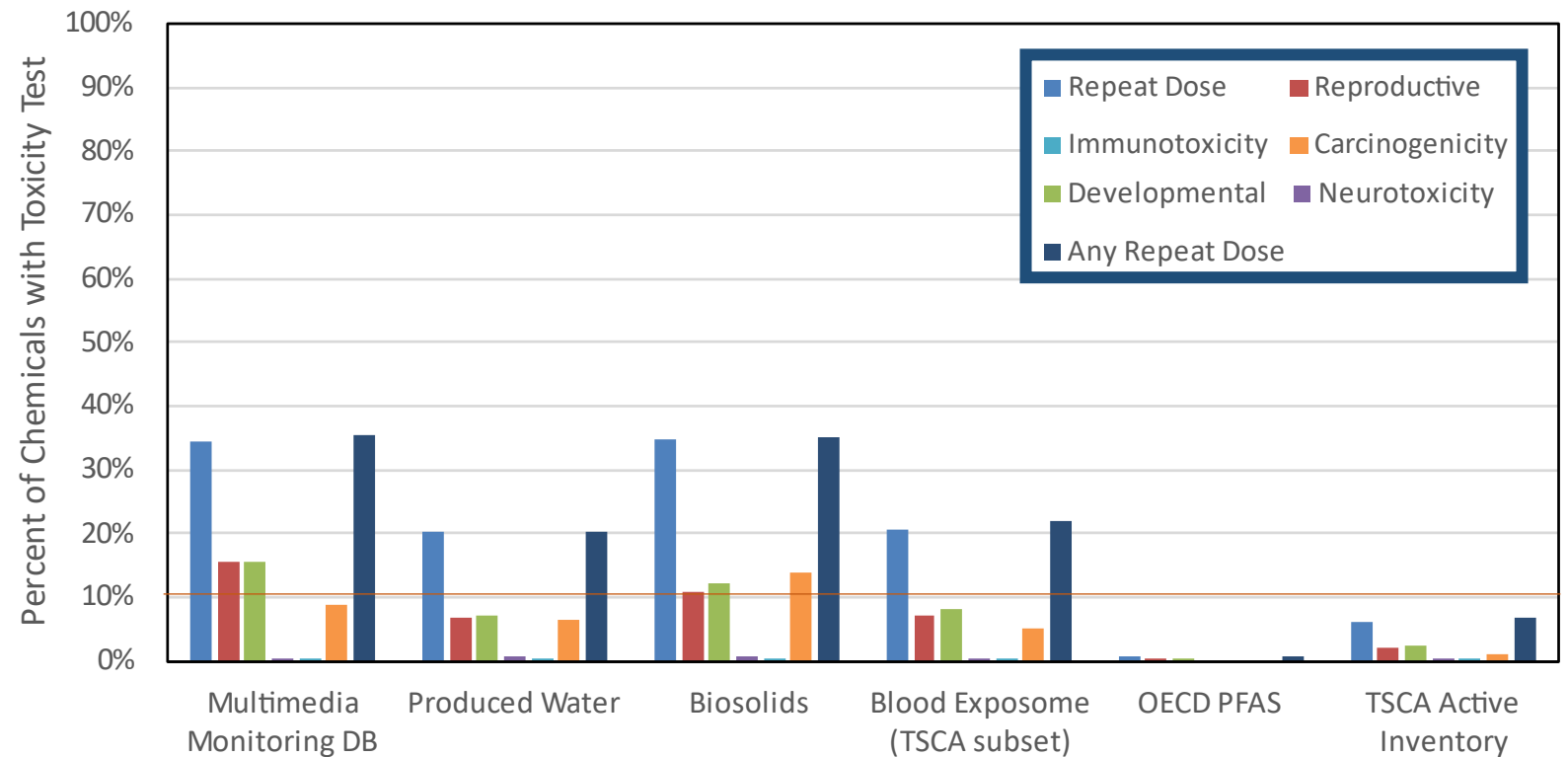
National Research Council

- Major challenge is too many chemicals and not enough data
- Total # chemicals = 65,725
- Chemicals with no toxicity data of any kind = ~46,000

NATIONAL ACADEMY PRESS  
Washington, D.C. 1984

## Toxicity Data and Human Health Assessments: 2022

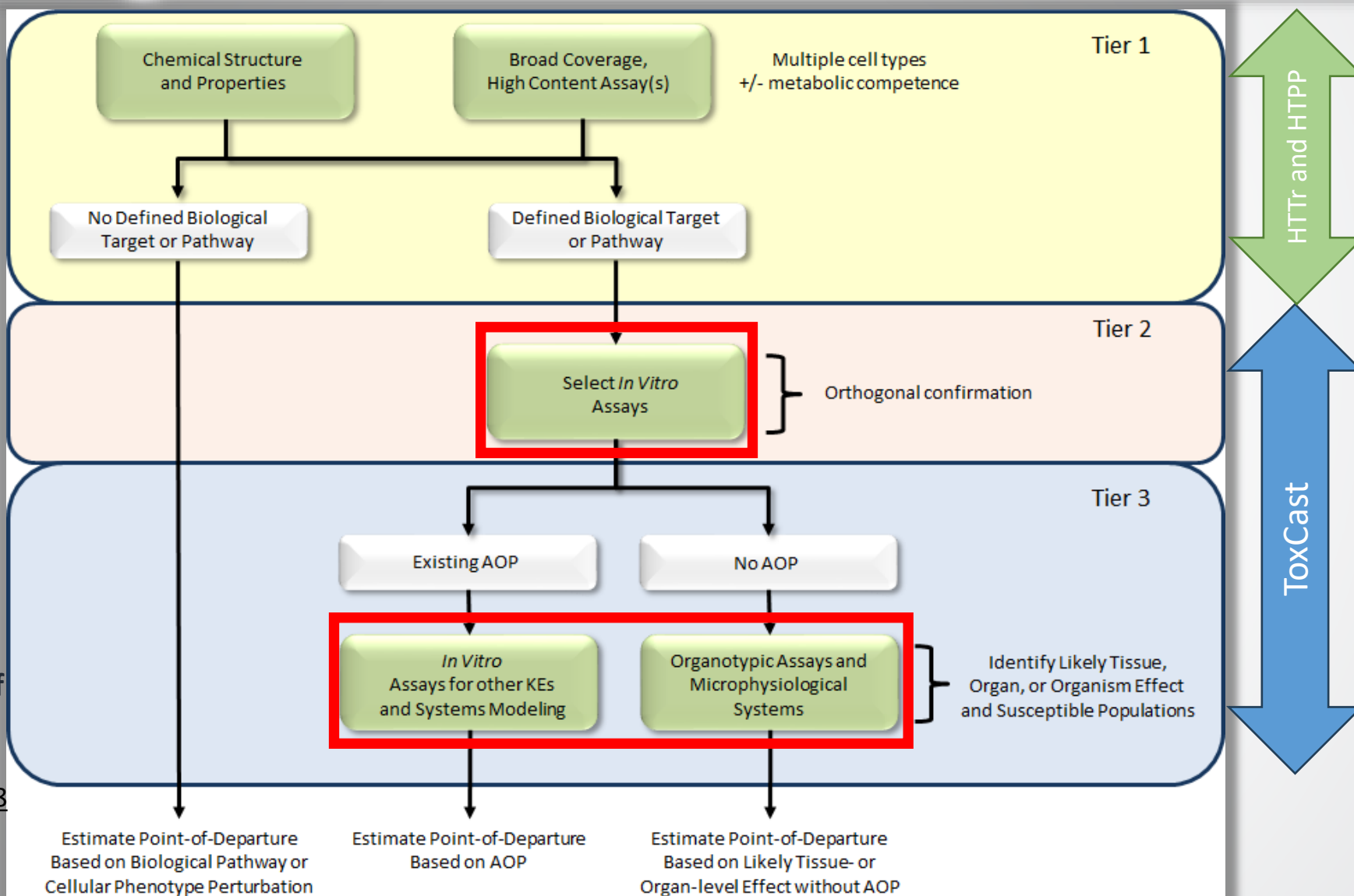
2020 survey of 19 countries and regions: **350,000 chemicals and mixtures of chemicals** are registered in one or more inventories<sup>1</sup>



**There is not enough time or money to generate traditional animal-based hazard data for all of these chemicals and their mixtures**



# Available bioactivity NAMs that are targeted or provide more biological complexity

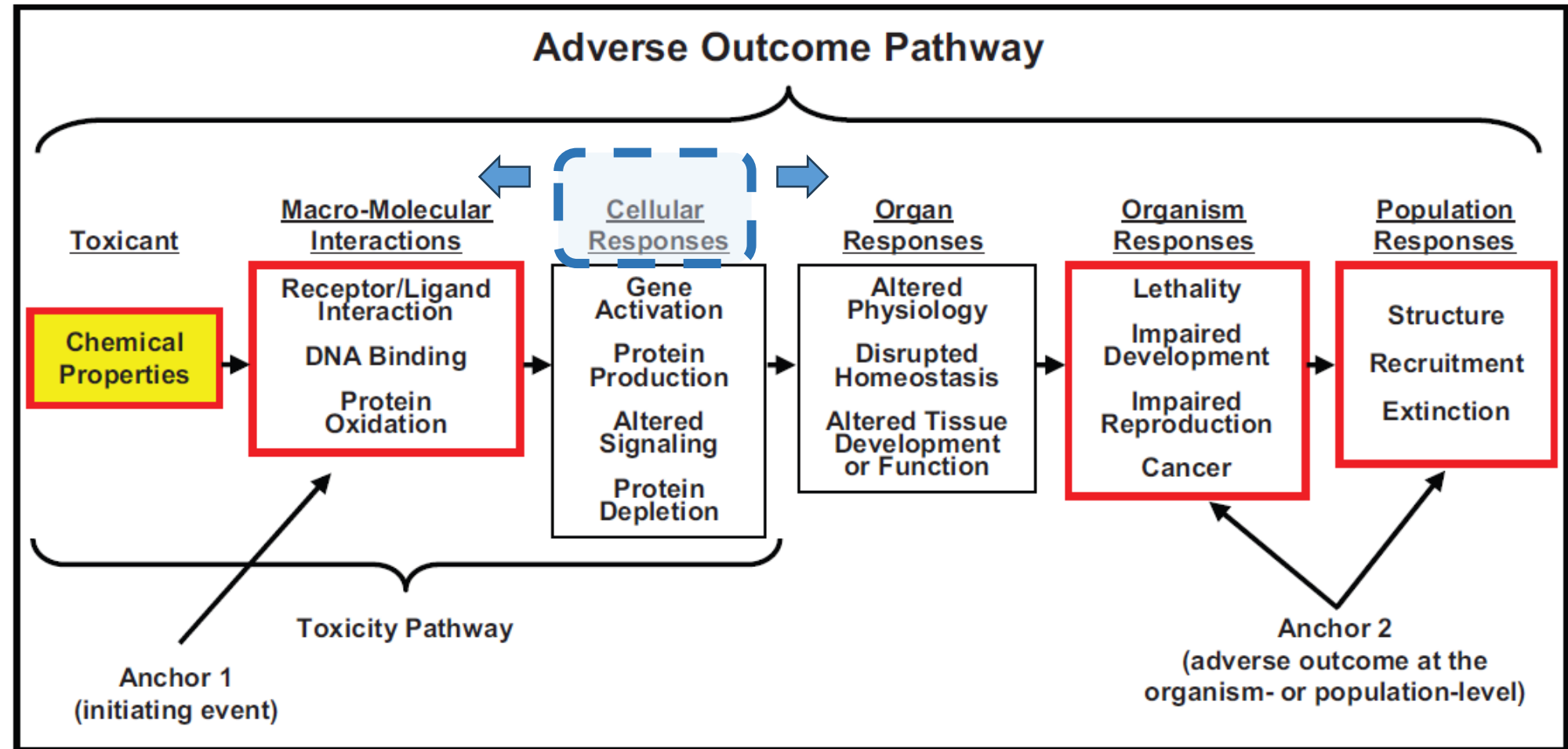


The NexGen Blueprint of  
CompTox at US EPA

Thomas et al.

(2019) [10.1093/toxsci/kfz058](https://doi.org/10.1093/toxsci/kfz058)

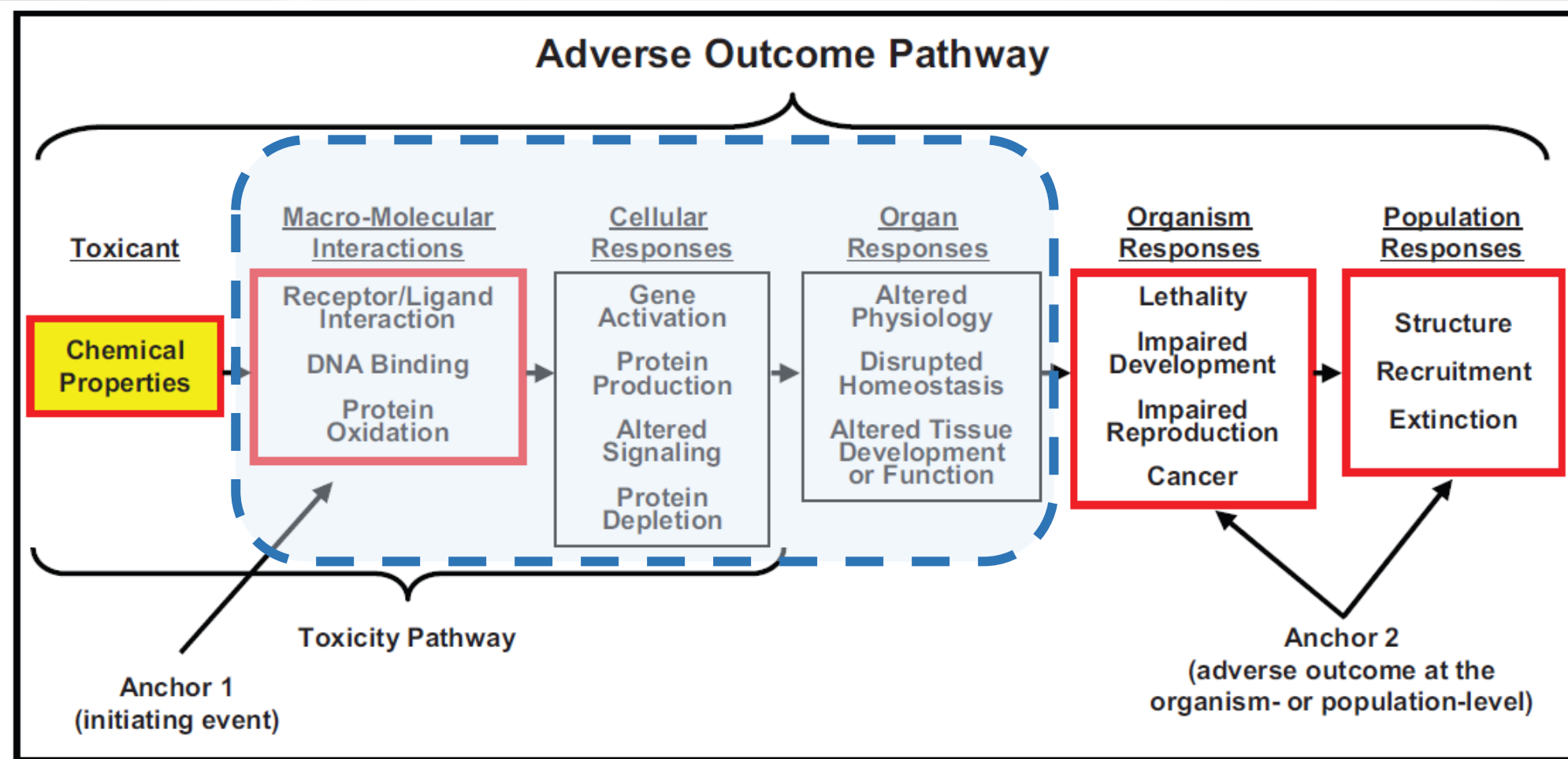
# Broad profiling NAMs tend to probe cellular responses



Ankley et al. (2010)  
[10.1002/etc.34](https://doi.org/10.1002/etc.34)

- Broad profiling NAMs in use interrogate gene expression and cell morphological responses
- These data may be used to infer upstream interactions or downstream organ responses

# Heterogeneous targeted NAMs in ToxCast address a range of event types in the AOP framework





# What is ToxCast?

ToxCast encompasses pipelining software (*tcpl*, along with dependency *tcplfit2*), a database of NAM information (known as invitrodb), and efforts to curate and make these data informative.



# ToxCast Database Coverage

The **Toxicity Forecaster (ToxCast)** program curates and makes publicly available targeted bioactivity screening data. Latest database release (v4.2) includes:

26 Assay Sources

655 Unique Assays

1570 Unique Endpoints

9614 Chemicals

Including a TOX21 assay source for data generated by the TOX21 program



Diverse biology with **over 500 mapped gene targets**, including:



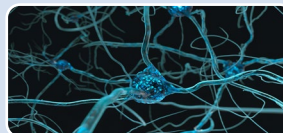
**Endocrine-Related:** Estrogen Receptor, Androgen Receptor, Thyroid, Steroidogenesis



**Cellular Signaling Pathways:** Cytotoxicity, Proliferation, Stress, Mitochondrial Toxicity

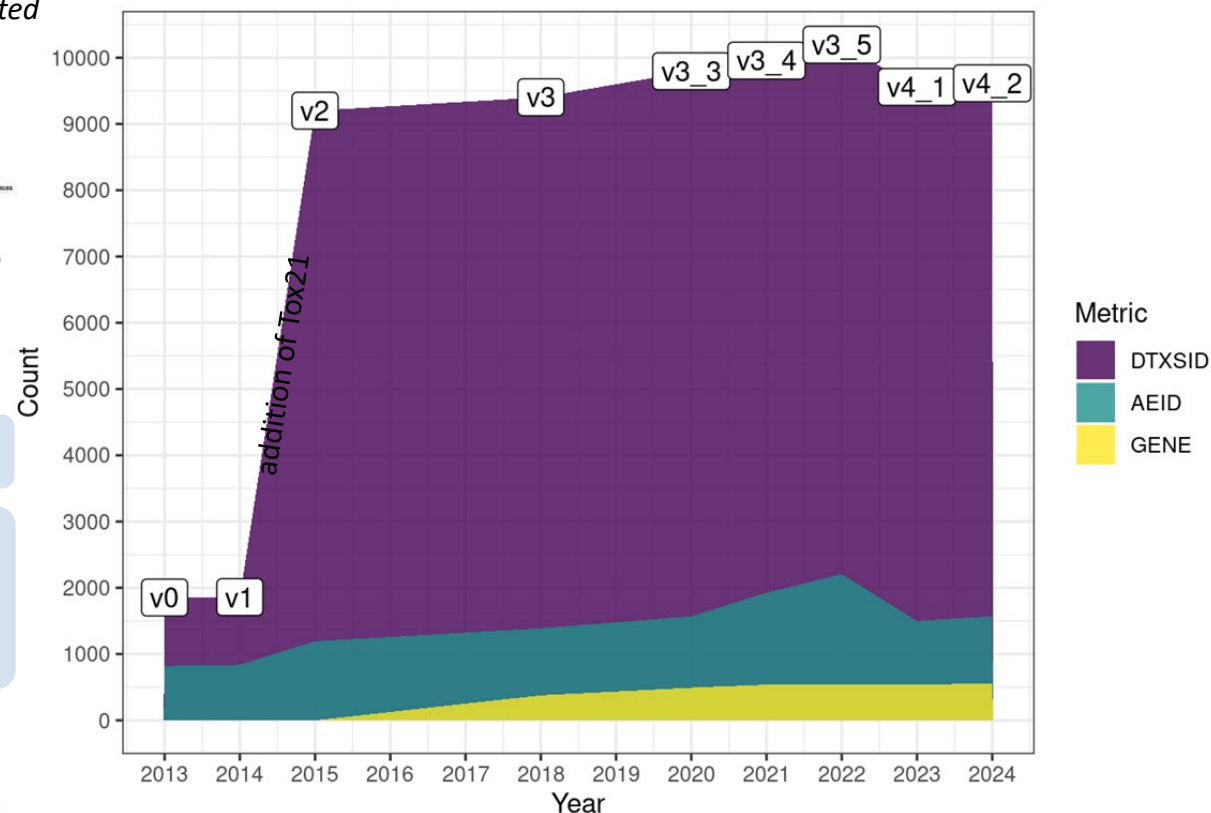


**Protein Interactions:** Receptors, Transporters, Ion Channels, Enzymes



**Complex Responses, e.g.** Immune Response, Development, Neurotoxicity, etc.

ToxCast Data Counts, 2013-2024

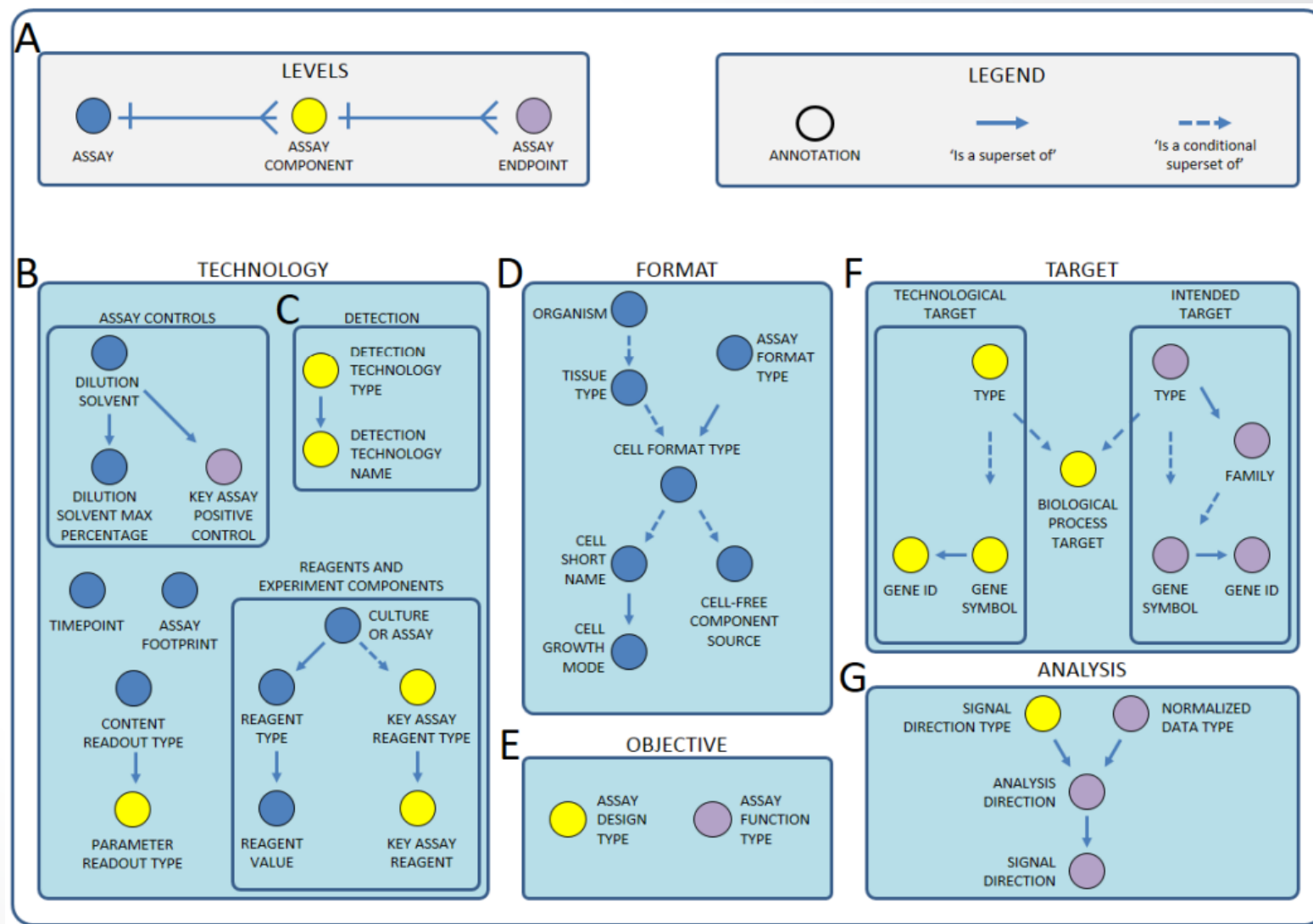


As of October 2024, v4.1 data is available on CCD and CTX Bioactivity API. V4.2 is available to download.



# ToxCast assays have been annotated for many types of details

- Using the Bioassay Ontology (BAO) framework to capture four types of information, each annotation is assigned as a feature to an assay element level:
  - Assay Source:** *Who* conducted the assay
  - Assay:** *What* assay platform was used
  - Assay Component:** “Raw” readout of *what* was measured
  - Assay (component) Endpoint:** *How* the measurement is interpreted (i.e. normalized component data)





# ToxCast assays have many formats

Counts based on invitrodb v4.2

## Cell free and tissue-derived cell-free assays (481 endpoints)

| Assay Source     | Binding | Enzymatic Activity | Protein complementation |
|------------------|---------|--------------------|-------------------------|
| Eurofins (ERF)   | X       | X                  | X                       |
| NovaScreen (NVS) | X       | X                  |                         |
| CCTE             | X       | X                  | X                       |
| Tox21            |         |                    | X                       |

- All NVS and majority of ERF assays are cell-free
- CCTE and Tox21 are diverse sources that have a variety of cell formats

**Whole embryo zebrafish developmental assays (46 endpoints)**  
from CCTE Padilla and the Tanguay laboratory at Oregon State University

## Cell based assays: 1043 endpoints (266 endpoints from primary cells)

| Assay Source                        | Metabolic Capacity                                | Primary Cells |
|-------------------------------------|---|---------------|
| ACEA                                |   |               |
| Apredica (APR)                      | HepG2   |               |
| Attagene (ATG)                      | HepG2 HG19 subclone for elevated Phase I capacity |               |
| BioSeek (BSK)                       |   | X             |
| CEETOX                              |   |               |
| CellzDirect (CLD)                   | Human sandwich cultured hepatocytes               | X             |
| LifeTech Expression Analysis (LTEA) | HepaRG  |               |

*...and many more! Examine assay details*

- Some annotations are hierarchical
  - e.g., general 'intended\_target\_family' and more specific 'intended\_target\_family\_sub'





# Auxiliary Annotations

- Capture additional information, including:
  - Assay list presence
  - Linkages to relevant Adverse Outcome Pathways (AOPs) and Key Events (KEs)
  - Relevant gene identifier(s) from National Center for Biotechnology Information (NCBI)
  - Reagents or experimental conditions
  - Publications describing assay design or results

Navigator

Filter objects

invitrodb

- Tables
  - assay
  - assay\_component
  - assay\_component\_endpoint
  - assay\_component\_map
  - assay\_descriptions
  - assay\_list
  - assay\_list\_aeid
  - assay\_reagent
  - assay\_reagent\_armitage
  - assay\_reference
  - assay\_source
  - chemical
  - chemical\_analytical\_qc
  - chemical\_lists
  - citations
  - class
  - cytotox
  - etl\_metadata
  - flyway\_schema\_history
  - gene
  - intended\_target
  - invitrodb\_dd
  - mc0
  - mc1
  - mc2
  - mc2\_aeid
  - mc2\_methods
  - mc3
  - mc3\_aeid
  - mc3\_methods

| Result Grid  |      |                                |                               |              |                |                                      |                     |                      |             |                              |                  |                      |                          |
|--|------|--------------------------------|-------------------------------|--------------|----------------|--------------------------------------|---------------------|----------------------|-------------|------------------------------|------------------|----------------------|--------------------------|
| Filter Rows: Edit: Export/Import: Wrap Cell Content: |      |                                |                               |              |                |                                      |                     |                      |             |                              |                  |                      |                          |
|  | aeid | acid                           | assay_component_endpoint_name | export_ready | internal_ready | assay_component_endpoint_desc        | assay_function_type | normalized_data_type | burst_assay | key_positive_control         | signal_direction | intended_target_type | intended_target_type_sub |
| 2  | 1    | ACEA_ER_80hr                   |                               | 1            | 1              | Data from the assay component ACE... | signaling           | percent_activity     | 0           | 17b-Estradiol                | gain             | pathway              | pathway-specified        |
| 4  | 2    | APR_HepG2_CellCycleArrest_1hr  |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin      | bidirectional    | pathway              | pathway-specified        |
| 6  | 3    | APR_HepG2_CellLoss_1hr         |                               | 1            | 1              | Data from the assay component APR... | viability           | log2_fold_induction  | 0           | Camptothecin;Anisomycin;P... | bidirectional    | cellular             | cellular                 |
| 8  | 4    | APR_HepG2_MicrotubuleCSK_1hr   |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin;P... | bidirectional    | cellular             | cellular                 |
| 10   | 5    | APR_HepG2_MitoMass_1hr         |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin;P... | bidirectional    | cellular             | mitochondria             |
| 12   | 6    | APR_HepG2_MitoMembPot_1hr      |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin;P... | bidirectional    | cellular             | mitochondria             |
| 14   | 7    | APR_HepG2_MitoticArrest_1hr    |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin;P... | bidirectional    | pathway              | pathway-specified        |
| 16   | 8    | APR_HepG2_NuclearSize_1hr      |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin;P... | bidirectional    | cellular             | nucleus                  |
| 18   | 9    | APR_HepG2_P-H2AX_1hr           |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin      | bidirectional    | pathway              | pathway-specified        |
| 20   | 10   | APR_HepG2_p53Act_1hr           |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin      | bidirectional    | pathway              | pathway-specified        |
| 22   | 11   | APR_HepG2_StressKinase_1hr     |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin      | bidirectional    | pathway              | pathway-specified        |
| 24   | 12   | APR_HepG2_CellCycleArrest_24hr |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin      | bidirectional    | pathway              | pathway-specified        |
| 26   | 13   | APR_HepG2_CellLoss_24hr        |                               | 1            | 1              | Data from the assay component APR... | viability           | log2_fold_induction  | 1           | Camptothecin;Anisomycin;P... | bidirectional    | cellular             | cellular                 |
| 28   | 14   | APR_HepG2_MicrotubuleCSK_24hr  |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Camptothecin;Anisomycin      | bidirectional    | cellular             | cellular                 |
| 30   | 15   | APR_HepG2_MitoMass_24hr        |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Paditaxel;CCCP               | bidirectional    | cellular             | mitochondria             |
| 32   | 16   | APR_HepG2_MitoMembPot_24hr     |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Paditaxel;CCCP               | bidirectional    | cellular             | mitochondria             |
| 34   | 17   | APR_HepG2_MitoticArrest_24hr   |                               | 1            | 1              | Data from the assay component APR... | signaling           | log2_fold_induction  | 0           | Paditaxel;CCCP               | bidirectional    | pathway              | pathway-specified        |

Result Grid

Form Editor

Field Types

Query Stats

Execution Plan

# Reference “cheat sheet” for Assay Sources in *invitrodb*

*Invitrodb version 4.2 (released Sept 2024) contains 26 different assay sources*

| Assay source | Description   | Biological Focus  |
|--------------|---|---|
| ACEA         | ACEA Biosciences: Real-time, label-free, cell growth assay based on a microelectronic impedance readout   | Endocrine (ER-induced proliferation)  |
| APR          | Apredica: CellCiphr High Content Imaging system   | Hepatic cells (HepG2)   |
| ArunA        | ArunA Biomedical: Neuronal migration and neural network formation measured in a human H9-derived embryonic stem cell model  | Cell cycle, neurodevelopment  |
| ATG          | Attagene: Multiplexed pathway profiling platform in metabolically enhanced HepG2 subclone   | Nuclear receptor and stress response profile  |
| BSK          | Bioseek: BioMAP system providing uniquely informative biological activity profiles in complex human primary co-culture systems  | Immune response, inflammation, vascular   |
| CEETOX       | Ceetox/OpAns: HT-H295R assay ( <a href="#">OECD Test No. 456</a> )  | Endocrine (steroidogenesis)   |
| CLD          | CellzDirect: Expression of xenobiotic metabolizing enzyme and transporter genes measured in nuclease protection assays (qNPA)   | Liver (Phase I/Phase II/ Phase III expression)  |
| ERF          | Eurofins: Many cell-free (and some cell-based) assays across a very large range of receptor and protein-based targets. NVS replacement                                      | Many targets: Receptor binding; transporter protein binding; ion channels; enzyme inhibition            |
| IUF          | Leibniz Research Institute: Neuronal based models, including radial glia migration, oligodendrocyte migration, neuronal differentiation, and parallel cytotoxicity measures | <a href="#">Developmental neurotoxicity in vitro battery (DNT-IVB)</a>                                  |
| LTEA         | Life Tech Expression Analysis: Gene expression measured in HepaRG cells following 48-hour chemical exposure   | Liver toxicity model via transcription factor regulated-metabolism and markers of oxidative/cell stress |



# Assay Sources (continued)

| Assay Source | Description   | Biological Focus  |
|--------------|---|---|
| NVS          | Novascreen: Large diverse suite of cell-free binding and biochemical assays   | Many targets: Receptor binding; transporter protein binding; ion channels; enzyme inhibition      |
| OT           | Odyssey Thera: Protein-fragment complementation to understand novel protein:protein interactions  | Endocrine (ER and AR)   |
| STM          | Stemina: Human pluripotent stem cell-based assay measuring changes in cellular metabolism   | Developmental toxicity  |
| Tanguay      | Tanguay Lab at the Oregon State University Sinnhuber Aquatic Research Laboratory under a Material Transfer Agreement (MTA) for the ToxCast chemical library   | Developmental toxicity in zebrafish   |
| TOX21        | Tox21/NCGC: Interagency agreement between the NIH, NTP, FDA and EPA. NIH Chemical Genomics Center (NCGC) is the primary screening facility running ultra high-throughput screening assays across a large interagency-developed chemical library | Many targets and formats<br><a href="https://tripod.nih.gov/tox/">https://tripod.nih.gov/tox/</a> |
| UKN          | University of Konstanz: High content imaging assays to characterize neural migration and neurite outgrowth  | <a href="#">Developmental neurotoxicity in vitro battery (DNT-IVB)</a>                            |
| UPITT        | Johnston Lab at the University of Pittsburgh: Androgen receptor nuclear translocation assay under a Material Transfer Agreement (MTA) for the ToxCast chemical library  | Endocrine (AR related)  |
| VALA         | VALA Sciences: High-content screening assay to assess neurovascular unit (NVU) hazard potential   | Neurodevelopment, wound recovery (angiogenesis)   |

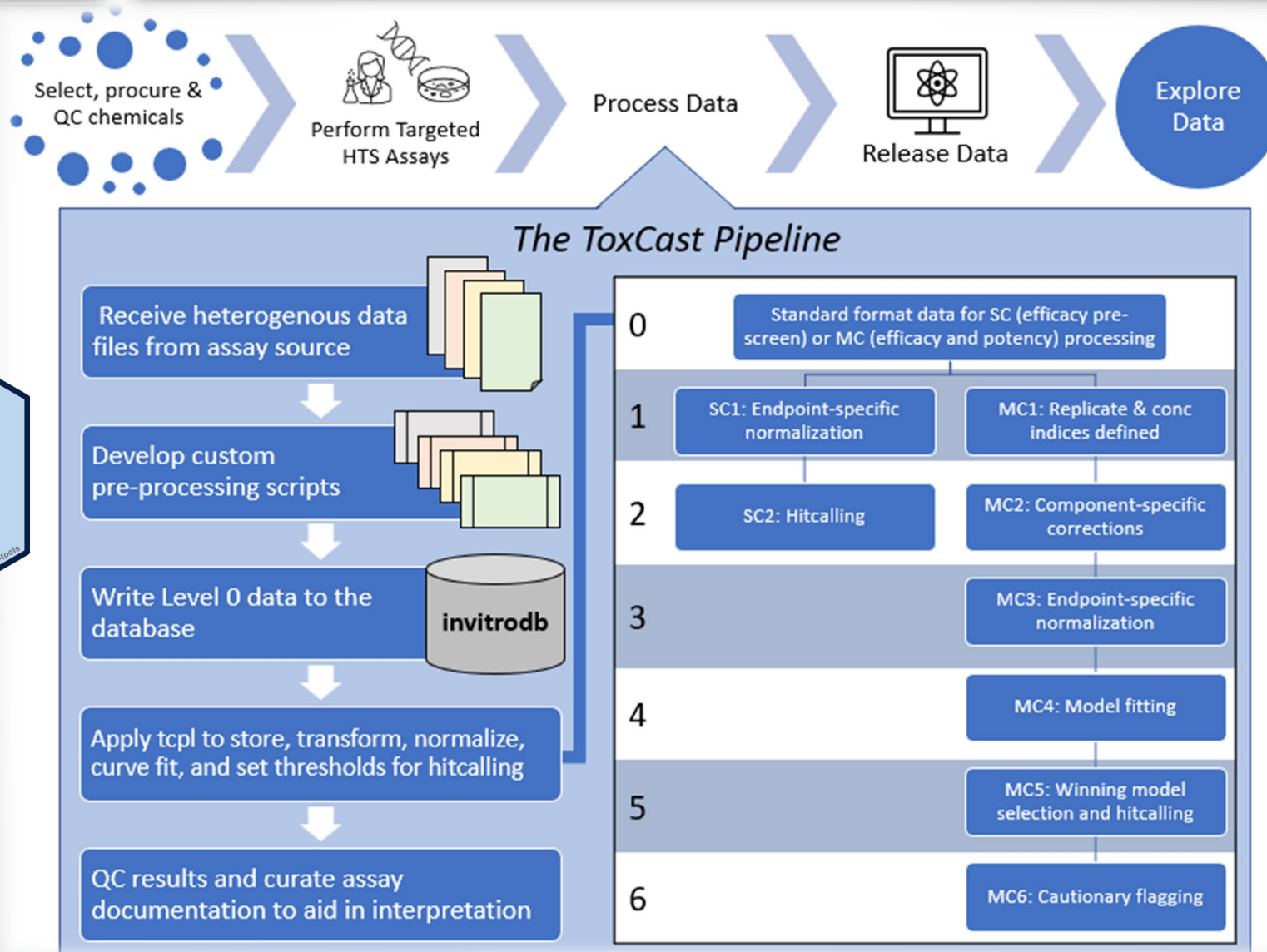
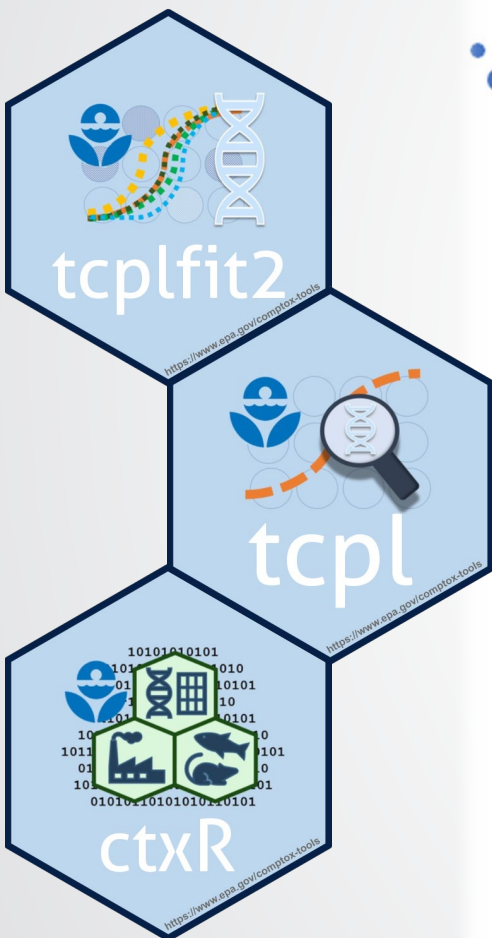


## With each release, more assay endpoints and more chemical x endpoint data are released

*Assay sources developed and generating data “in-house” within the EPA Office of Research and Development (subset shown here). Many of these targeted assays were developed to fill specific data gaps of interest.*

| Assay Source    | Description   | Biological Focus   |
|-----------------|---|--|
| CCTE_Deisenroth | Deisenroth laboratory at the EPA Center for Computational Toxicology and Exposure             | Endocrine disruption and developmental toxicity, as well as assays incorporating xenobiotic metabolism   |
| CCTE_GLTED      | Laboratories at EPA CCTE Great Lakes Toxicology and Ecology Division                          | Ecotoxicology and thyroid bioactivity  |
| CCTE_Padilla    | Padilla laboratory at the EPA Center for Computational Toxicology                             | Developmental toxicity in zebrafish  |
| CCTE_Mundy      | Mundy laboratory at the EPA Center for Computational Toxicology and Exposure                  | High content imaging to characterize neurodevelopment<br><a href="#">Developmental neurotoxicity in vitro battery (DNT-IVB)</a>                                      |
| CCTE_Shafer     | Shafer laboratory at the EPA Center for Computational Toxicology                              | Microelectrode array (MEA) to characterize neuroactivity and neurodevelopment (DNT) hazard<br><a href="#">Developmental neurotoxicity in vitro battery (DNT-IVB)</a> |
| CCTE_Simmons    | Simmons laboratory at the EPA Center for Computational Toxicology                             | Estrogen bioactivity, mitochondrial toxicity, and thyroid hormone disruption   |
| CPHEA_Stoker    | Stoker and Laws laboratories at the EPA Center for Public Health and Environmental Assessment | Thyroid bioactivity - Sodium-iodide cotransporter (NIS)  |

# Process Overview

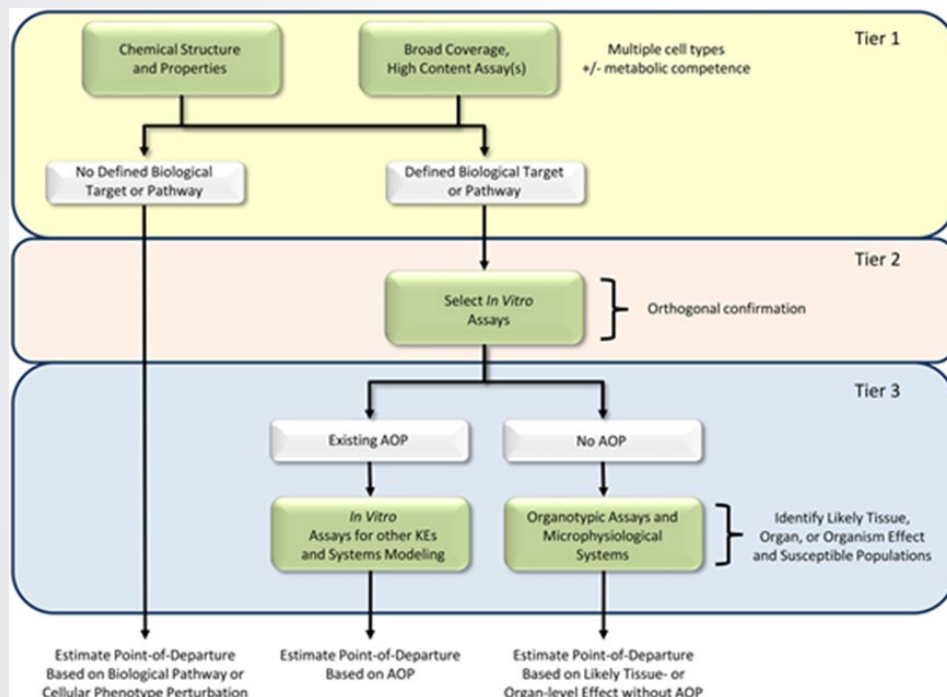




# Summary of Software & Database Updates: Past & Future of ToxCast's invitrodb

Madison Feshuk

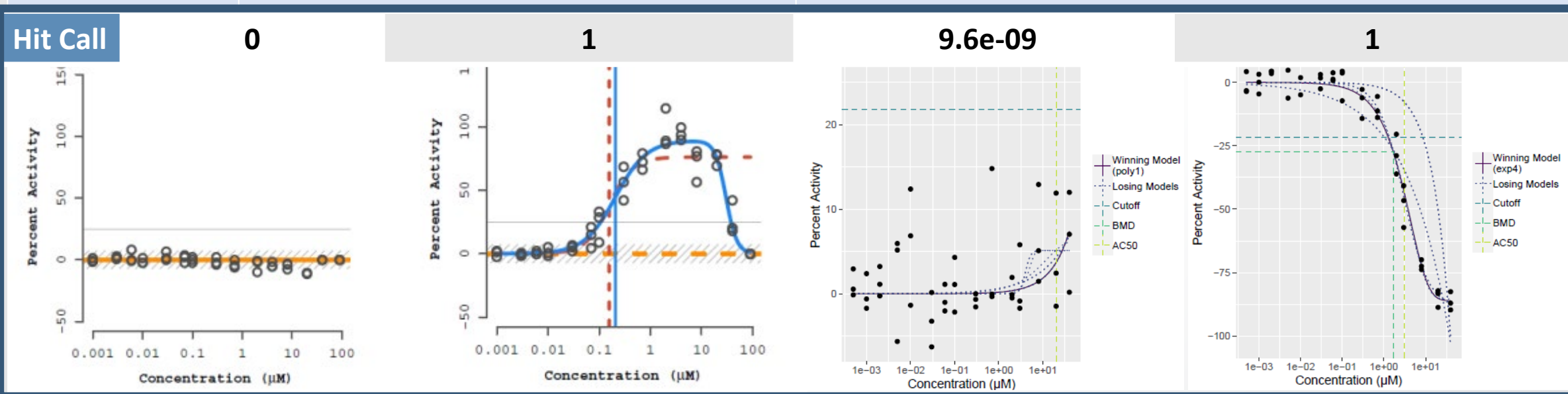
How can tiers of the CompTox Blueprint be connected when the curve-fitting may be different between them?



- Use a single curve-fitting approach across tiers, encoded by *tcplfit2*, to enable comparison across all bioactivity data
- Models in *tcplfit2* are based on models in BMDEpress2
- Flexibility to add more curve-fitting models in the future to better capture the varied response behavior observed in *in vitro* NAMs
- Reduction in data redundancy to simplify interpretation, annotations, and modeling tasks that utilize ToxCast data as input
- Improve interoperability of all bioactivity data

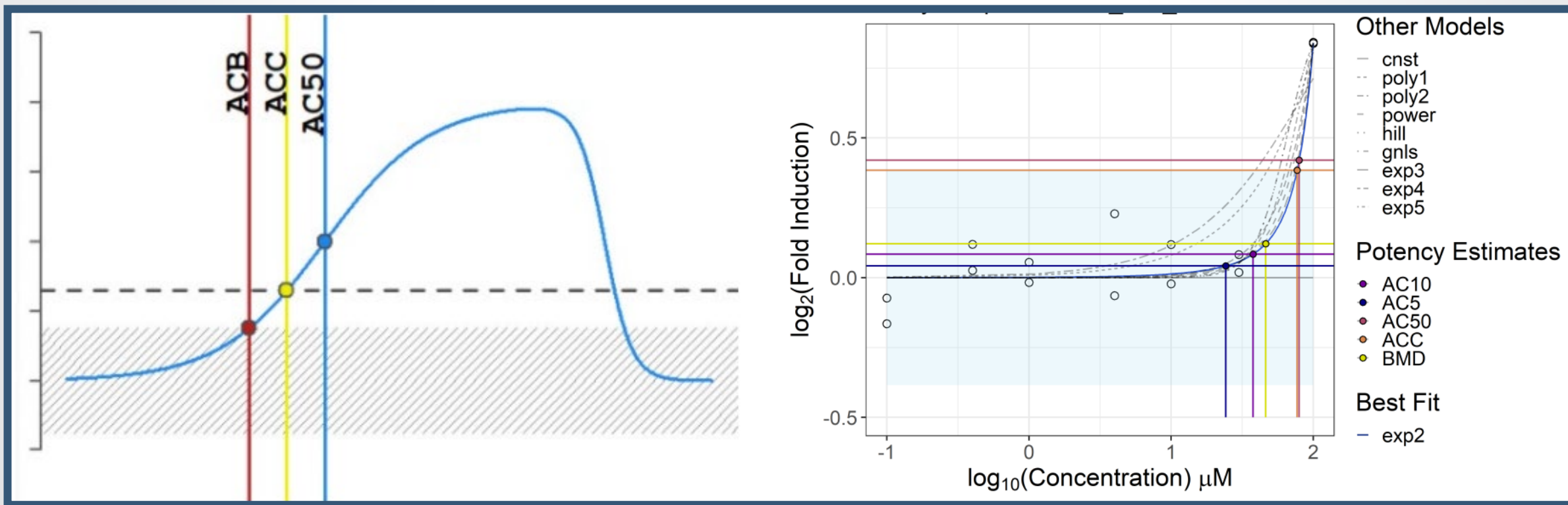
Together, these updates to *tcpl* and *invitrodb* improve the utility of ToxCast data within an integrated NAM strategy and unified open-source software approach.

| Enhancement          | <i>invitrodb</i> v3.5 and <i>tcp</i> v2.0   | <i>invitrodb</i> v4.0 and <i>tcp</i> v3.0   |
|----------------------|---|---|
| Curve-fitting models | Constant, Hill, and Gain-loss   | Additional models based on BMDExpress and encoded by R package dependency <i>tcpIfit2</i> |
| Activity hit calls   | <b>Discrete:</b> hit call (hitc)=1 active, hitc=0 inactive, or hitc=(-1) unable to fit (usually due to fewer than 4 concentrations) | <b>Continuous:</b> hitc between -1 to 1   |
| Plotting             | Several <i>tcp</i> functions  | Streamlined <b>tcpPlot()</b> function   |



Note: the concentration-response data are not identical but merely demonstrative of changes resulting from updates

| Enhancement       | <i>invitrodb v3.5</i> and <i>tcpl v2.0</i>  | <i>invitrodb v4.0</i> and <i>tcpl v3.0</i>   |
|-------------------|---|--|
| Potency estimates | Based on specified levels of response, e.g. <b>ACB</b> (activity concentration at baseline), <b>ACC</b> (cutoff), and <b>AC50</b> (50% of maximal response) | Additional potency and uncertainty estimates, e.g. a benchmark dose ( <b>BMD</b> ) as defined by the Benchmark Response ( <b>BMR</b> ) level |



| Enhancement                       | <i>invitrodb</i> v3.5 and <i>tcpl</i> v2.0  | <i>invitrodb</i> v4.0+ and <i>tcpl</i> v3.0+  |
|-----------------------------------|---|---|
| Stand-alone pipelining            | <i>tcplLite</i> employed flat csv files structured like <i>invitrodb</i> for stand-alone pipelining applications  | <i>tcplLite</i> was deprecated.<br><a href="#">tcplfit2</a> can be used for stand-alone data processing   |
| Endpoint structure and annotation | <i>tcpl</i> only fit in the positive analysis direction therefore dual endpoints were registered to capture gain and loss of signal.                                    | Given bidirectional fitting, a single endpoint is sufficient to capture both gain and loss of signal. Many endpoints were removed and/or renamed, and annotations were updated to reflect this paradigm shift. Continued curation efforts enable better data aggregation. |
| Schema changes                    | Processed data was previously stored in “wide” format with a fixed number of columns in the Level 4 (mc4) and Level 5 (mc5) tables based on three curve-fitting models. | Complete <i>tcplfit2</i> model parameters are captured within the mc4_param and mc5_param tables, allowing for generic fitting and hit calling, with summary-level statistics now only stored in mc4 and mc5.   |
| Fit Categories                    | Based <i>winning model</i> , active or inactive hitc, efficacy, and relationship between the AC50 and the concentration range screened.                                 | A more generic approach based on active or inactive hitc, efficacy, and relationship between the AC50 and the concentration range screened.   |
| Cautionary Flags                  | Programmatically generated to indicate curve characteristics or potential anomalies in data   | Coded logic updated to encompass BMD, bidirectional fitting, and continuous hitc  |

To understand the impacts of *tcpl* updates on ToxCast data, we compared:

- **invitrodb v3.5**, processed using tcpl v2.1.0, and
- **invitrodb v4.0**, which includes the same data as invitrodb v3.5 but reprocessed with tcpl v3.0.1 into an updated database schema to accommodate enhancements

## The ToxCast pipeline: updates to curve-fitting approaches and database structure

M. Feshuk<sup>1</sup>, L. Kolaczowski<sup>1,2</sup>, K. Dunham<sup>1,2</sup>, S. E. Davidson-Fritz<sup>1</sup>,  
K. E. Carstens<sup>1</sup>, J. Brown<sup>1</sup>, R. S. Judson<sup>1</sup> and K. Paul Friedman<sup>1\*</sup>

<sup>1</sup>Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, Durham, NC, United States, <sup>2</sup>National Student

invitrodb v4.1 is the first public release in the new schema. invitrodb v4.0 was a beta release to assess impact of updates.

<https://doi.org/10.3389/ftox.2023.1275980>



# *Invitrodb* v4.2 Highlights

- *Invitrodb* v4.2 (September 2024) includes the following notable updates:
  - Data Updates
    - New data and new endpoints
    - Assay lists
    - Administered Equivalent Doses (AED) provided as *mc7*
    - Chemical Analytical QC and Applicability Domain
  - Curve-fitting
    - Biphasic polynomial 2 (poly2) model
    - Benchmark dose (BMD) bounding
  - Directionality of response
    - Negative “not applicable” hit calls
  - Schema changes
  - Auxiliary Annotations
    - GD211 Aligned Assay Description Documents

## ***Counts Change from v4.1:***

*0 Sources*  
*+32 Assays*  
*+71 Endpoints*  
*+55 Chemicals*  
*+23 Gene Mappings*

***Check out the v4.2 release note for full details***

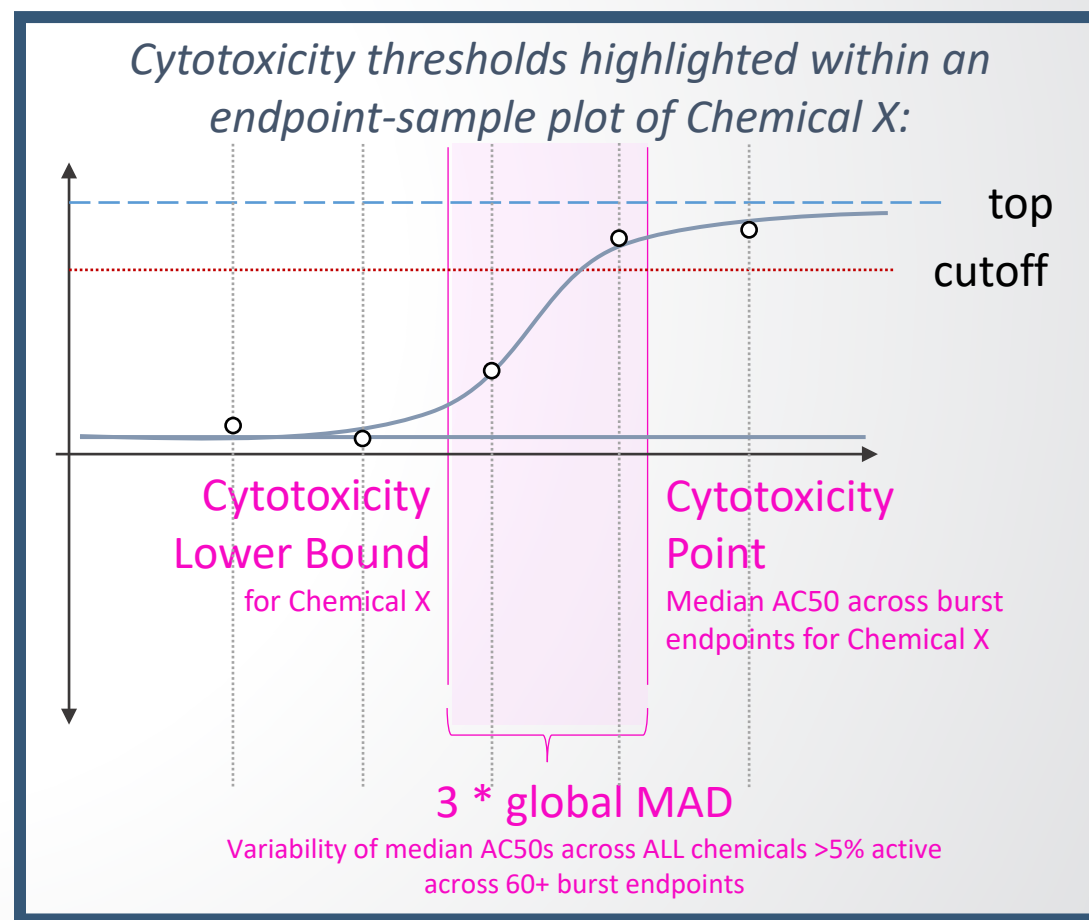


# Cytotoxicity Burst Distribution

- Cytotoxicity Burst phenomenon describes generalized cell stress and/or cytotoxicity observed above some threshold concentration. These thresholds can help contextualize if other bioactivity data may be confounded by assay interference resulting from cytotoxicity, particularly when parallel viability assessments are unavailable.

Considerations:

1. `global_mad`
2. Cytotoxicity Point and Lower Bound
3. Default value





# Cytotoxicity Burst Distribution

- Cytotoxicity Burst phenomenon describes generalized cell stress and/or cytotoxicity observed above some threshold concentration. These thresholds can help contextualize if other bioactivity data may be confounded by assay interference resulting from cytotoxicity, particularly when parallel viability assessments are unavailable.

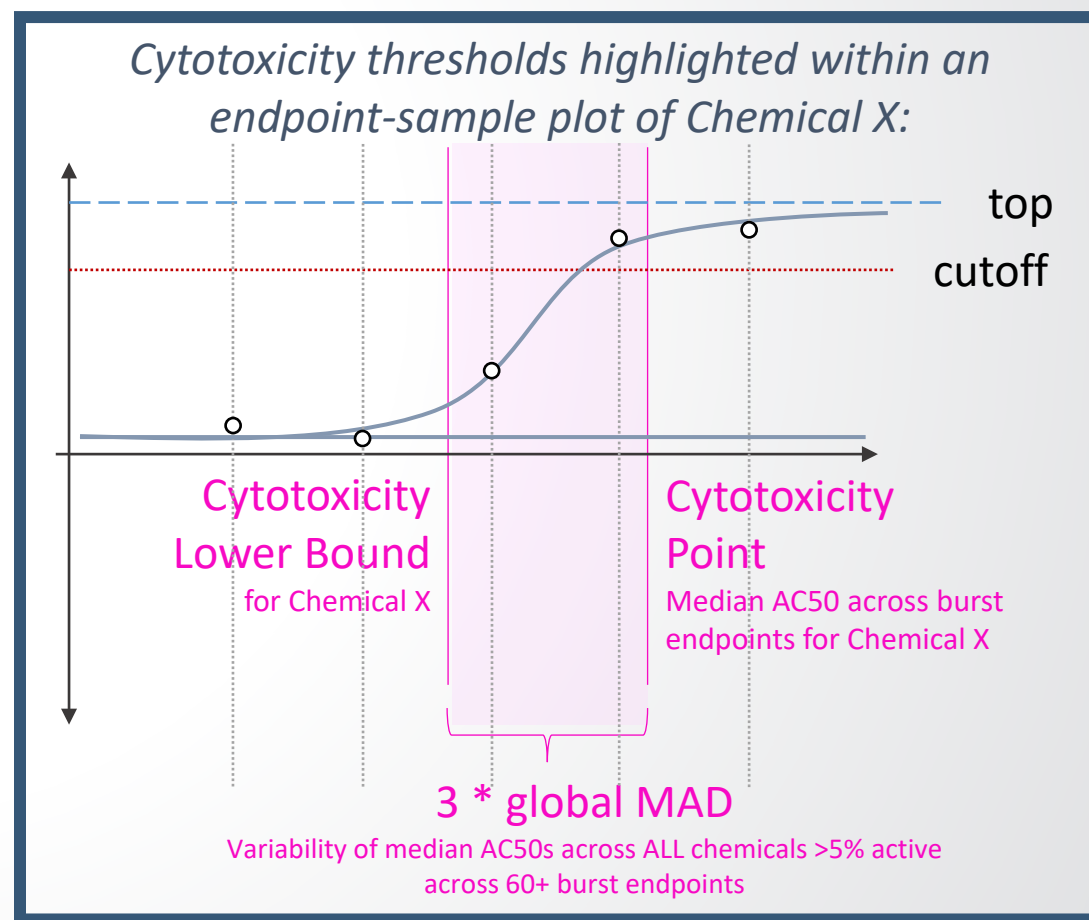
## Considerations:

### 1. What defines whether a chemical is included in the *global\_mad* calculation?

- A list of endpoints across several cell lines and assay technologies are annotated as “burst assays”. The global MAD is an estimate of the variance for all chemicals tested in greater than or equal to 60 burst endpoints (ntst) and with an active hit call in at least 5% of “burst” assay endpoints tested (burstperc)
  - global\_MAD = Median of the median absolute deviation of AC50s of chemicals tested in ntst ≥ 60 and burstperc > 0.05*

### 2. Cytotoxicity Point and Lower Bound

### 3. Default value





# Cytotoxicity Burst Distribution

- Cytotoxicity Burst phenomenon describes generalized cell stress and/or cytotoxicity observed above some threshold concentration. These thresholds can help contextualize if other bioactivity data may be confounded by assay interference resulting from cytotoxicity, particularly when parallel viability assessments are unavailable.

## Considerations:

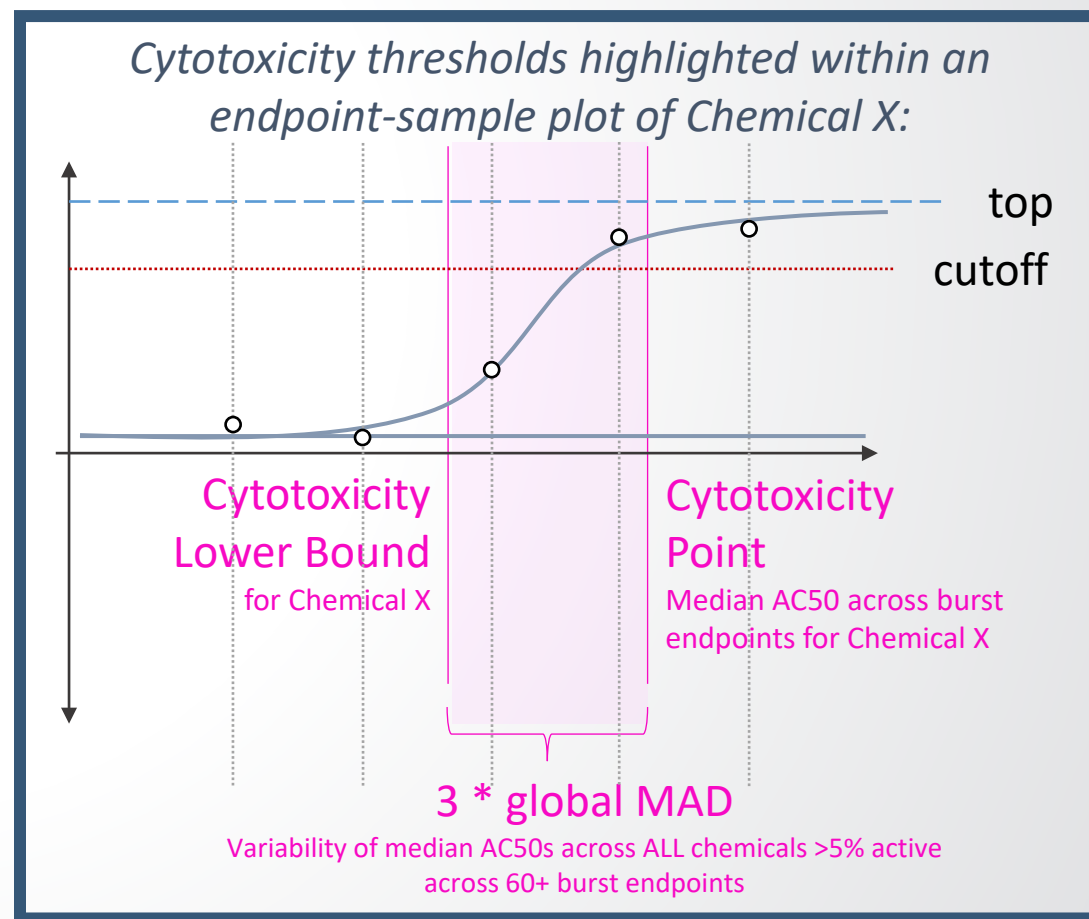
### 1. What defines whether a chemical is included in the *global\_mad* calculation?

- A list of endpoints across several cell lines and assay technologies are annotated as “burst assays”. The global MAD is an estimate of the variance for all chemicals tested in greater than or equal to 60 burst endpoints (ntst) and with an active hit call in at least 5% of “burst” assay endpoints tested (burstperc)
  - global\_MAD* = Median of the median absolute deviation of AC50s of chemicals tested in  $ntst \geq 60$  and  $burstperc > 0.05$

### 2. When is a *cytotoxicity point* reported for a chemical?

- If a chemical has  $\geq 5\%$  hit percent or  $\geq 2$  hits across “burst” endpoints,
  - Cytotox point* = median(AC50 across “burst” endpoints where  $burstperc > 0.05$  or  $n_{hit} \geq 2$ )
  - Cytotox lower bound* = *Cytotox point* –  $3 * global\_MAD$

### 3. Default value





# Cytotoxicity Burst Distribution

- Cytotoxicity Burst phenomenon describes generalized cell stress and/or cytotoxicity observed above some threshold concentration. These thresholds can help contextualize if other bioactivity data may be confounded by assay interference resulting from cytotoxicity, particularly when parallel viability assessments are unavailable.

## Considerations:

### 1. What defines whether a chemical is included in the *global\_mad* calculation?

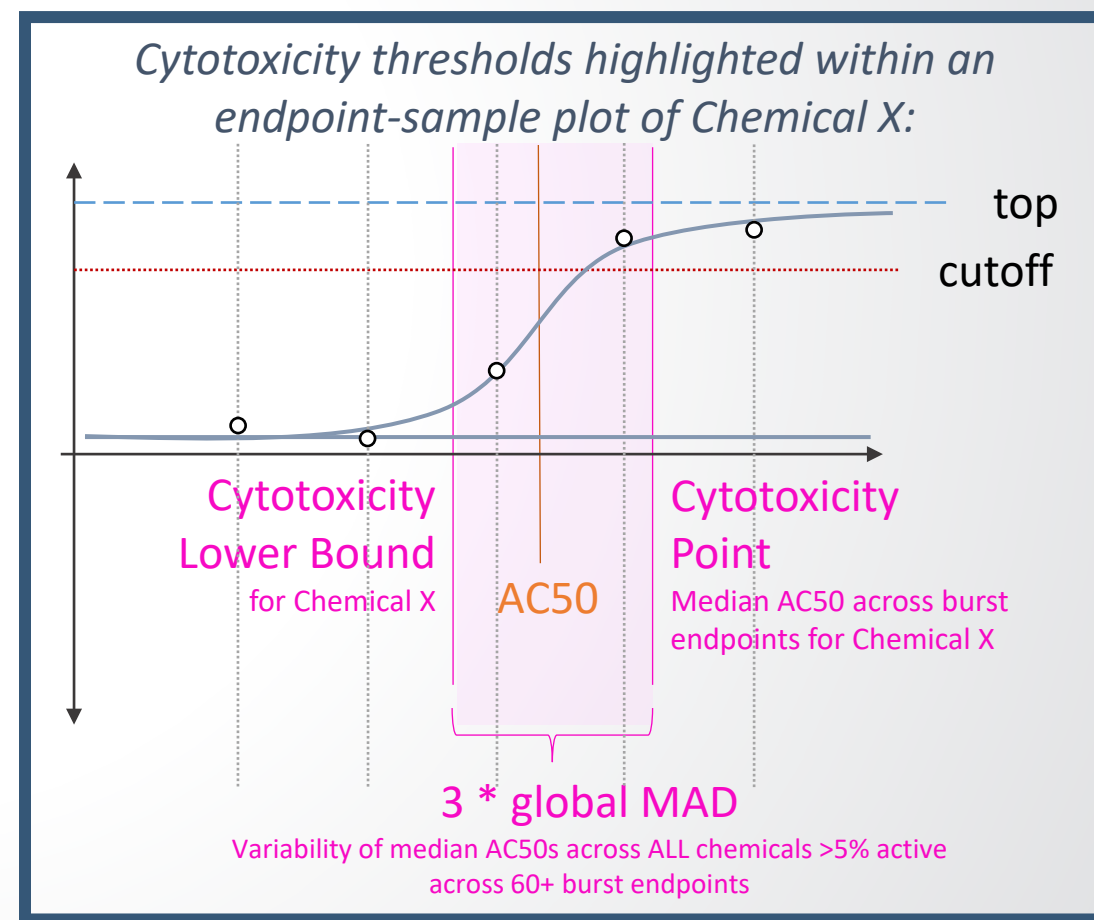
- A list of endpoints across several cell lines and assay technologies are annotated as “burst assays”. The global MAD is an estimate of the variance for all chemicals tested in greater than or equal to 60 burst endpoints (ntst) and with an active hit call in at least 5% of “burst” assay endpoints tested (burstperc)
  - global\_MAD = Median of the median absolute deviation of AC50s of chemicals tested in ntst≥60 and burstperc > 0.05*

### 2. When is a *cytotoxicity point* reported for a chemical?

- If a chemical has  $\geq 5\%$  hit percent or  $\geq 2$  hits across “burst” endpoints,
  - Cytotox point = median(AC50 across “burst” endpoints where burstperc>0.05 or nhit≥2)*
  - Cytotox lower bound = Cytotox point – 3 \* global\_MAD*

### 3. When is a *default value* reported for a chemical?

- If a chemical has *burstperc*<0.05 or 1 nhit, default value assumed because we lack enough data to compute a median or assume an estimate of variance in cell stress/cytotoxicity data. Chemicals not tested in any “burst assays” are also assigned a default value.
  - Default value = 3 on the log10 scale or 1000 micromolar on the arithmetic scale*
    - Arbitrary high concentration given it’s outside the typical testing range for ToxCast data and in principle all compounds are toxic if given in high enough concentration.





# Assay Description Documents

- Given heterogeneity of ToxCast assays, assay documentation aligned with the OECD Guidance Document 211 (GD211) international standardization efforts can make ToxCast data more useful and interpretable for use in regulatory decision-making and research applications.
- Given major ToxCast software and database enhancements plus a desire to automate report generation, a complete overhaul to existing assay documentation process was undertaken.
- A compiled report and assay description documents for **809/1570 endpoints** were released in September 2024, accompanying *invitrodb* v4.2

| Endpoint Report Coverage   |   |   |
|--|---|---|
| Assay Sources  |   | Biology   |
| <ul style="list-style-type: none"><li>• ACEA</li><li>• ARUNA</li><li>• ATG</li><li>• BSK</li><li>• CLD</li><li>• CCTE_Shafer</li><li>• CCTE_Deisenroth</li><li>• CCTE_GLTED</li><li>• CCTE_Mundy</li></ul> | <ul style="list-style-type: none"><li>• CCTE_Padilla</li><li>• IUF</li><li>• LTEA</li><li>• OT</li><li>• STM</li><li>• TOX21</li><li>• Tanguay</li><li>• UKN</li><li>• VALA</li></ul> | <ul style="list-style-type: none"><li>• Androgen Receptor (AR)</li><li>• Estrogen Receptor (ER)</li><li>• Developmental Toxicity</li><li>• Developmental Neurotoxicity</li><li>• Immunotoxicity</li><li>• Steroidogenesis Bioactivity</li><li>• Thyroid Bioactivity</li><li>• Non-mammalian Vertebrate<ul style="list-style-type: none"><li>• Zebrafish</li></ul></li></ul> |



**tcpl: The ToxCast Data Analysis Pipeline**  
*Supporting Accessible Bioactivity Data for Toxicology*

Introduction  
Overview  
ToxCast Publications  
Connection Configuration  
Database Structure  
Assay Registration  
Chemical Registration  
Level 0 Pre-Processing  
Data Processing  
Data Interpretation  
Data Retrieval in invitrodb  
Data Retrieval via API  
Example Integrations with Other Computational Toxicology Tools

## **tcpl: The ToxCast Data Analysis Pipeline** *Supporting Accessible Bioactivity Data for Toxicology*

### Introduction

This vignette provides an overview of the *tcpl* package, including set up, Database Structure, Pre-processing Requirements, Assay and Chemical Registration, Data Processing, Data Interpretation, and Data Retrieval with *invitrodb* and via API.



### Overview

The ToxCast Data Analysis Pipeline (*tcpl*) is an R package that manages, curve-fits, plots, and stores ToxCast data to populate its linked MySQL database, *invitrodb*. The U.S. Environmental Protection Agency (EPA)'s Toxicity Forecaster (ToxCast™) program includes *in vitro* medium- and high-throughput screening (HTS) assays for the prioritization and hazard characterization of thousands of chemicals of interest. Targeted and confirmatory assays (like ToxCast assays) comprise Tiers 2-3 of the Computational Toxicology Blueprint (Thomas et al., 2019), and employ automated chemical screening technologies to evaluate the effects of chemical exposure on living cells and biological macromolecules, such as proteins.

The *tcpl* package is a flexible analysis pipeline is capable of efficiently processing and storing large volumes of data. The diverse data, received in heterogeneous formats from numerous vendors, are transformed to a standard computable format via *Level 0 Preprocessing* then loaded into the database by vendor-specific R

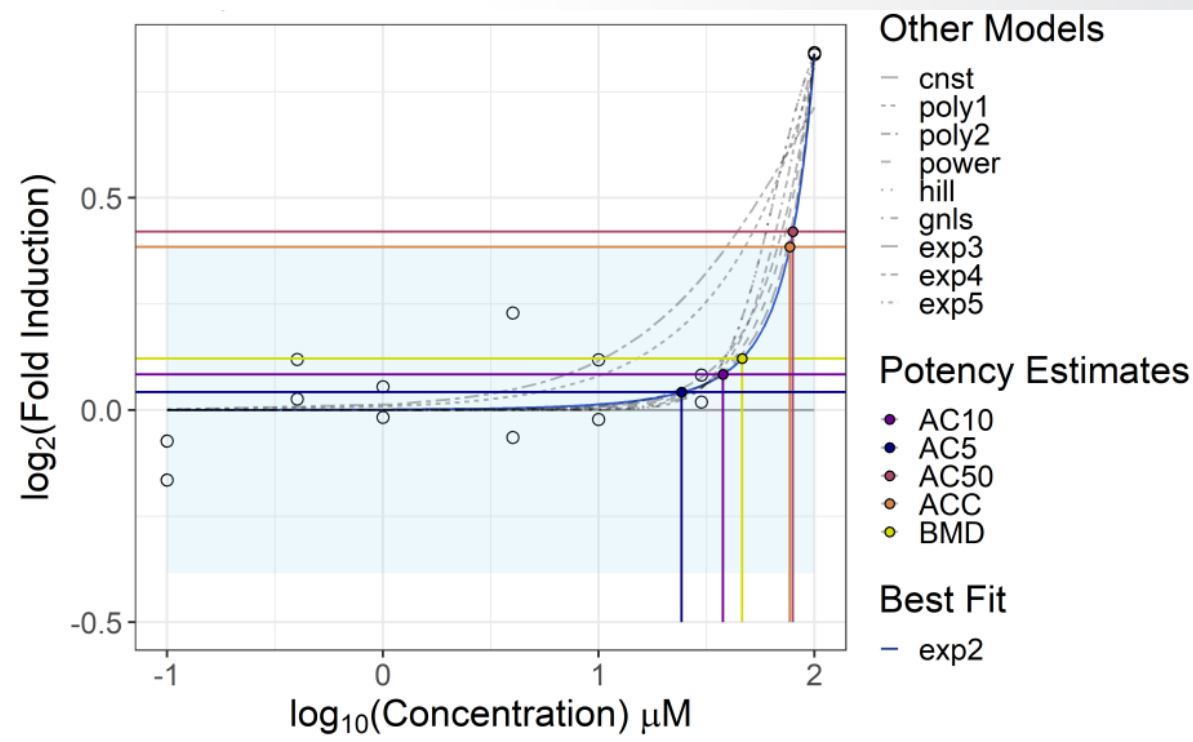
- New consolidated vignette
- Plotting updates
- API integration



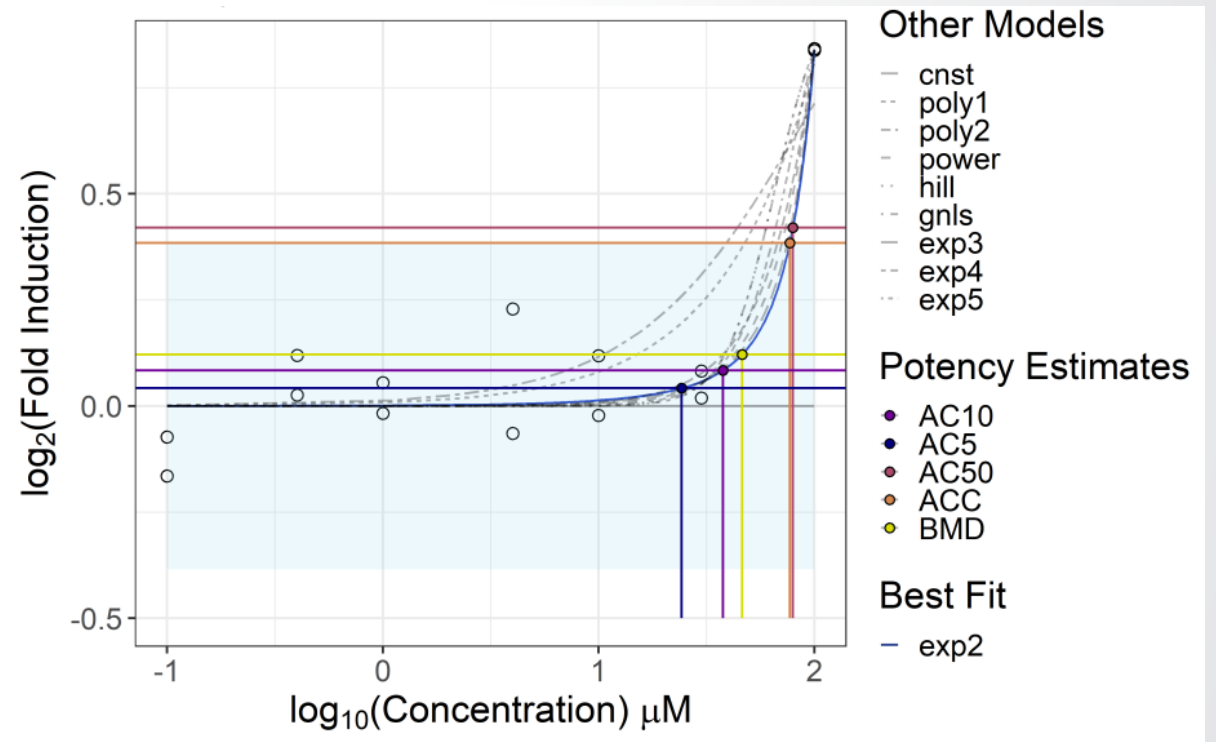
# Activity & Potency Estimates 101

Sarah Davidson-Fritz

- Multiple concentration experiments allow for evaluating a compound's impact on biological response with increasing concentrations
- Concentration-response modeling predicts the underlying relationship to assess **bioactivity** and **potency**



- Advantages:
  - Evaluates the underlying shape of the concentration-response relationship
  - Deriving a point-of-departure (POD) is **not dependent** upon experimental concentrations



*tcplfit2* includes the constant, Hill, and gain-loss models and recently incorporated other models present in BMDEExpress

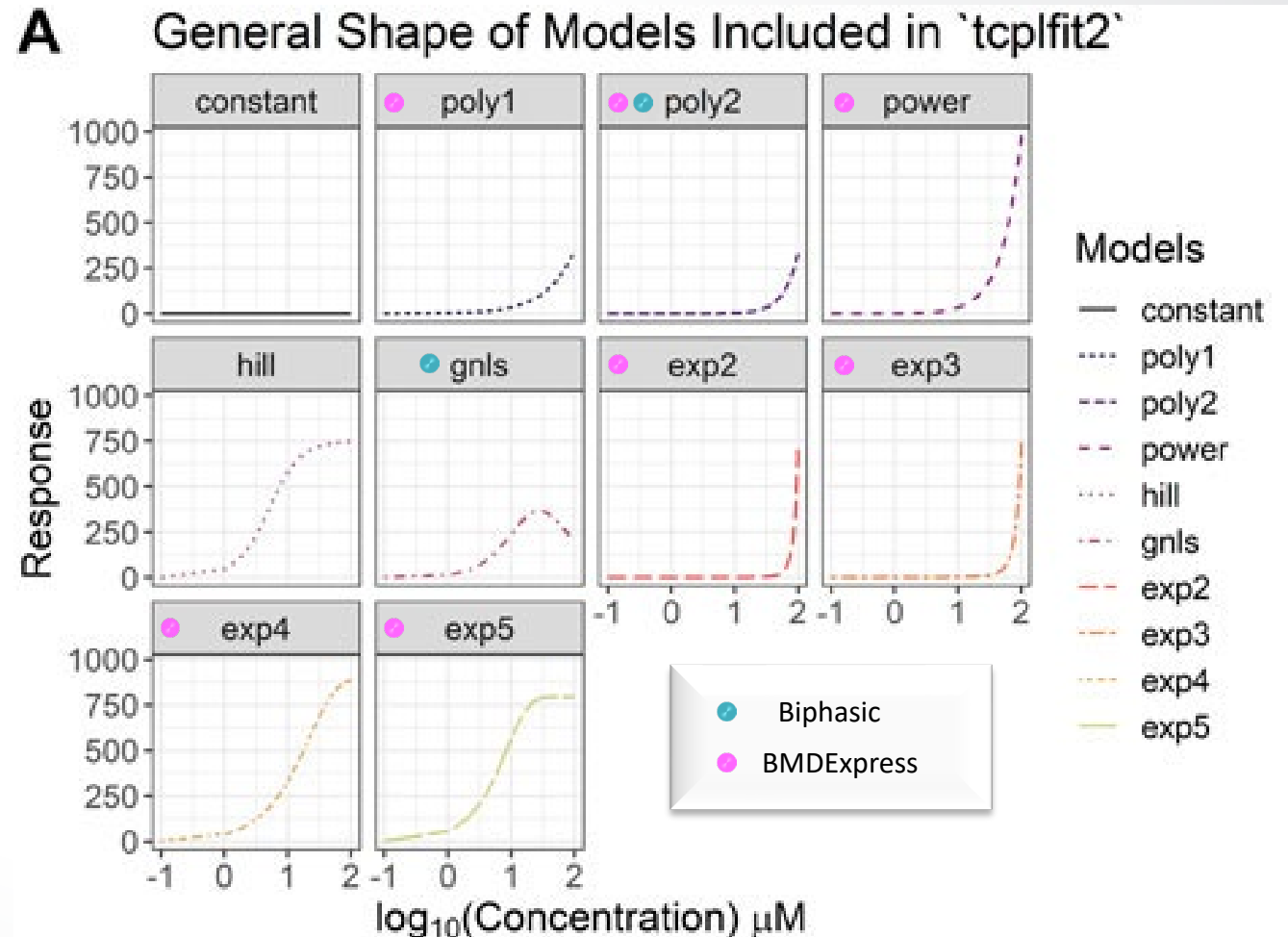
Models included are meant to capture the general range of possible curve shapes a biological responses may take

All model in *tcplfit2* assume the biological responses are continuous

## Winning Model:

The model with the *lowest Akaike Information Criterion (AIC)* is selected as the best model fit to the data

$$AIC = 2k - 2\ln(\hat{L})$$

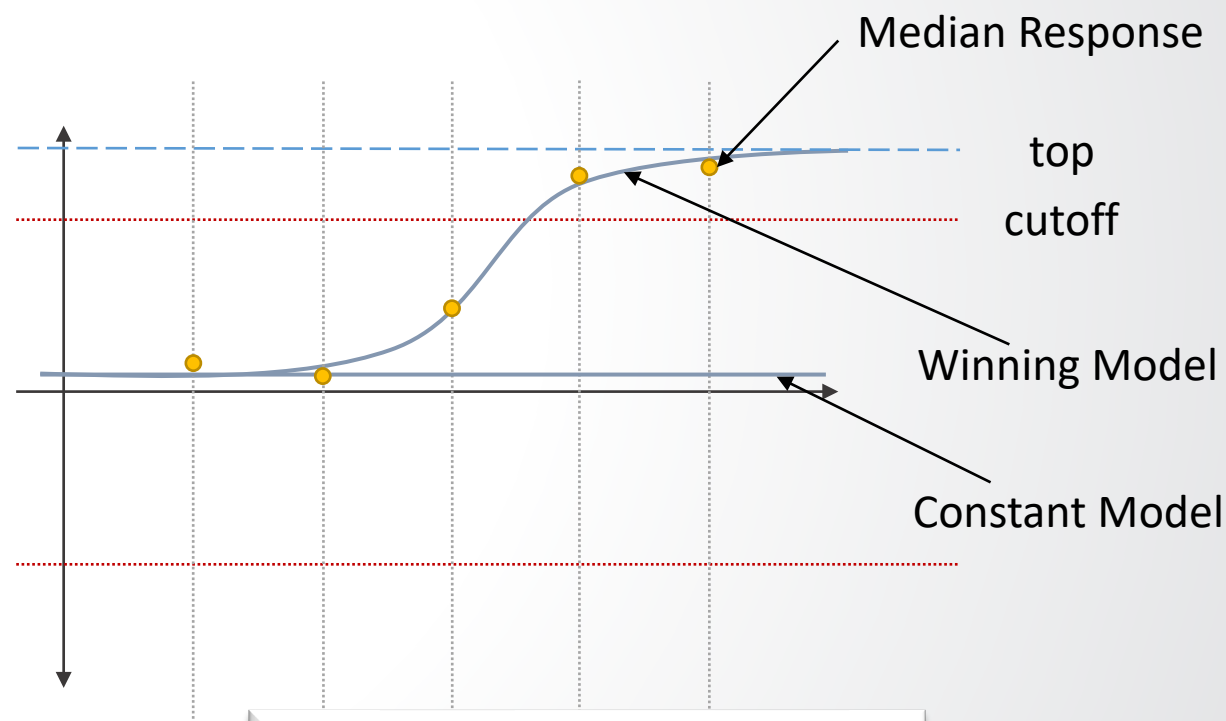


Is the concentration-response curve we are fitting indicative of meaningful biological activity?

**Metric: Continuous Hit-call**

Activity of concentration-response curves in tcpl v3.1 are indicated by the estimated continuous hit-call (hitc), which is the product of the three proportional weights:

- **$p_1$ : “the winning AIC value is less than that of the constant model”**
  - Determine whether the constant model – if allowed to win – is a better fit than the winning model – i.e. is the winning model essentially flat or not.
- **$p_2$ : “at least one median response is greater than the cutoff”**
  - At least one dose group has a central tendency of the response values “outside” the cutoff band (consider bi-directional).
  - Response is greater than cutoff in “+” direction and less than cutoff in “-” direction.
- **$p_3$ : “the top of the fitted curve is above the cutoff”**
  - Determine whether the predicted maximal response exceeds the cutoff, i.e. the response corresponding to the effect size of interest.



$$hitc \in [0,1]$$

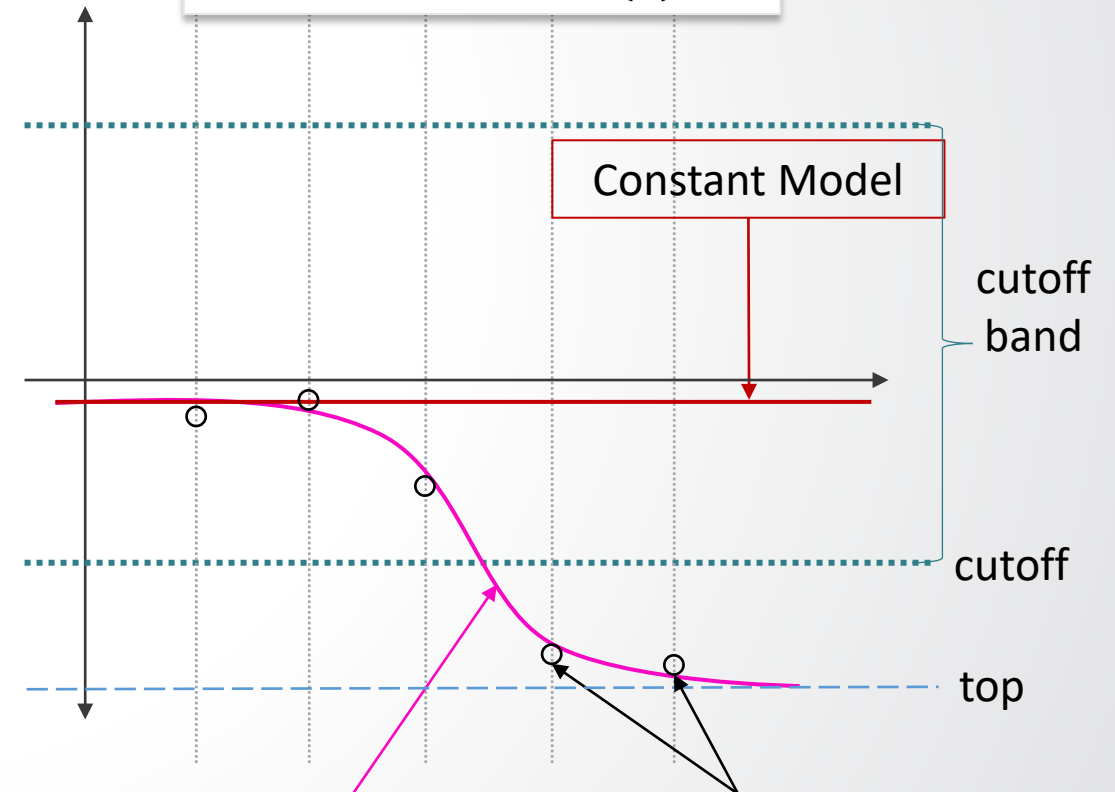
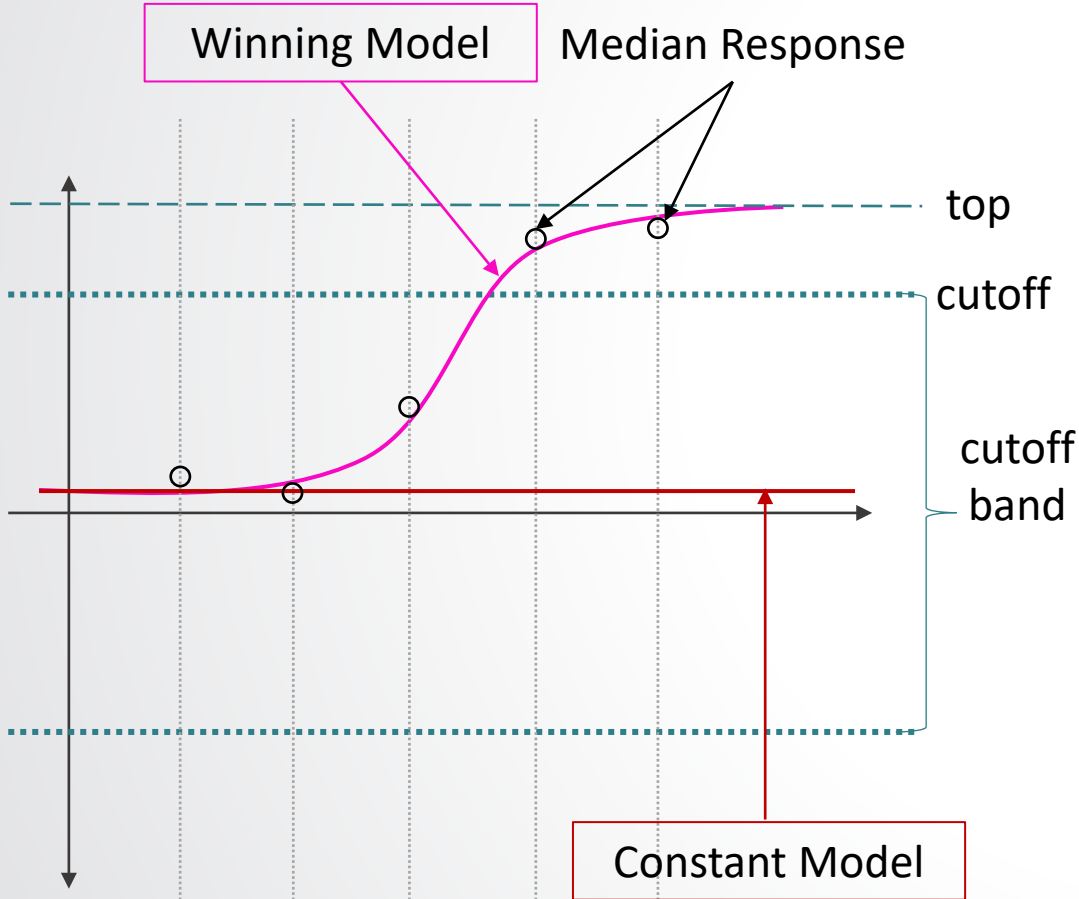
Active Response:  $hitc \geq 0.9$



# Proportional Weight: $p_1$

$p_1$  - address whether the winning model is a more appropriate fit compared to a flat curve fit

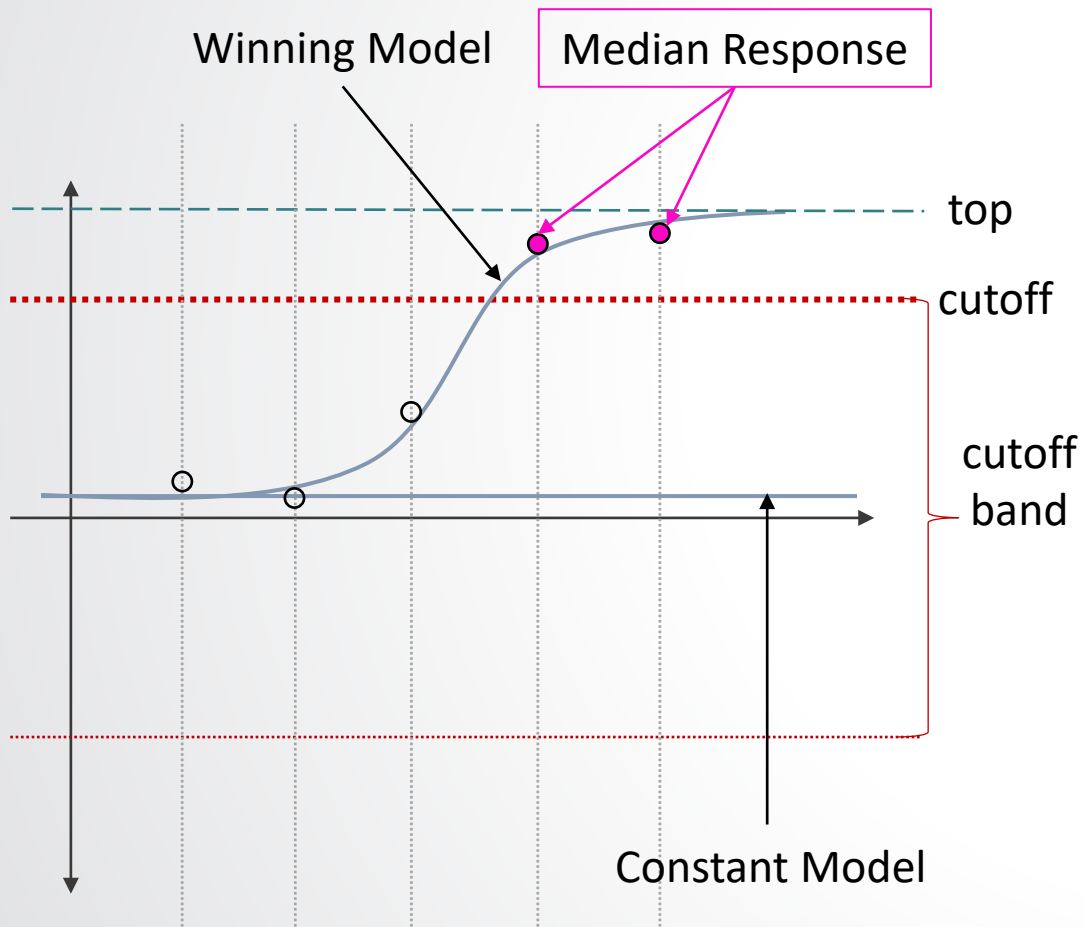
$$AIC = 2k - 2\ln(\hat{L})$$



\* Both cases here are examples of what should produce a 'high'  $p_1$  value. Winning model and data indicate response different from baseline.

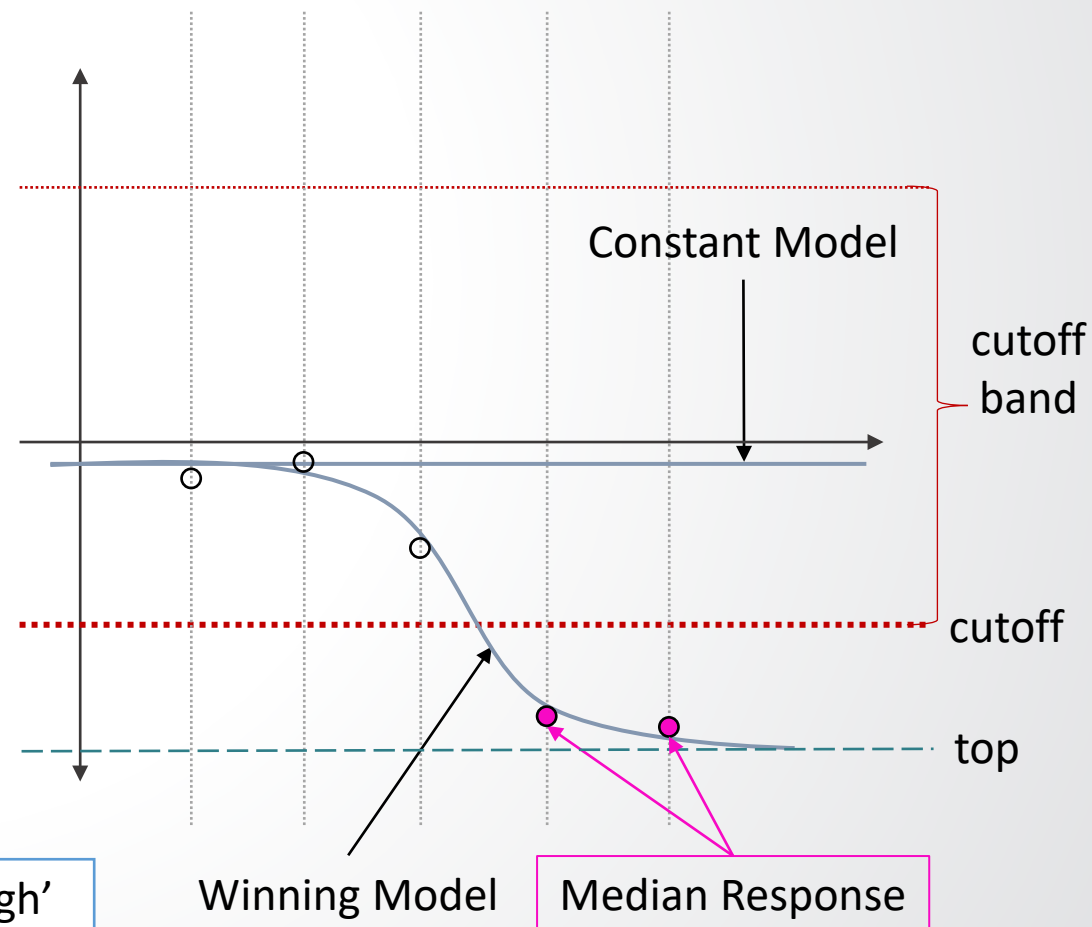


# Proportional Weight: $p_2$



\* Both cases here are examples of what should produce a 'high'  $p_2$  value. Two median responses outside cutoff band.

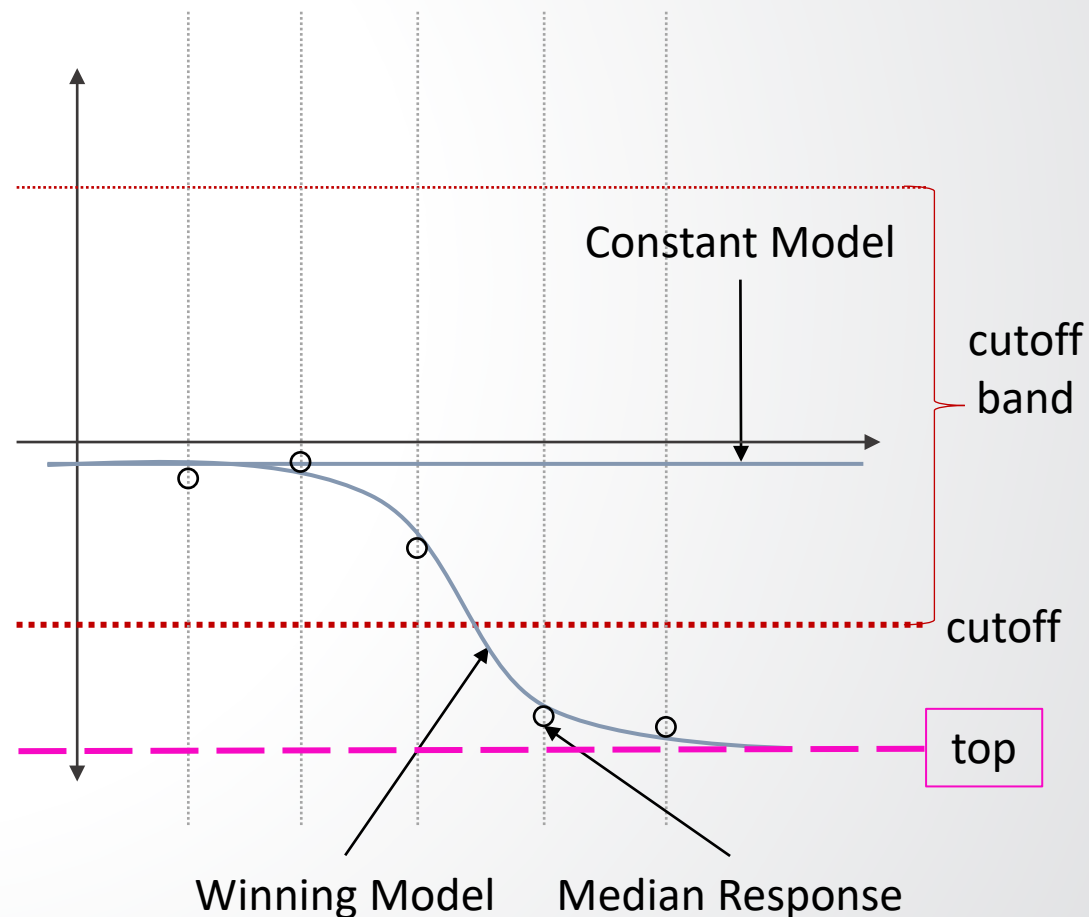
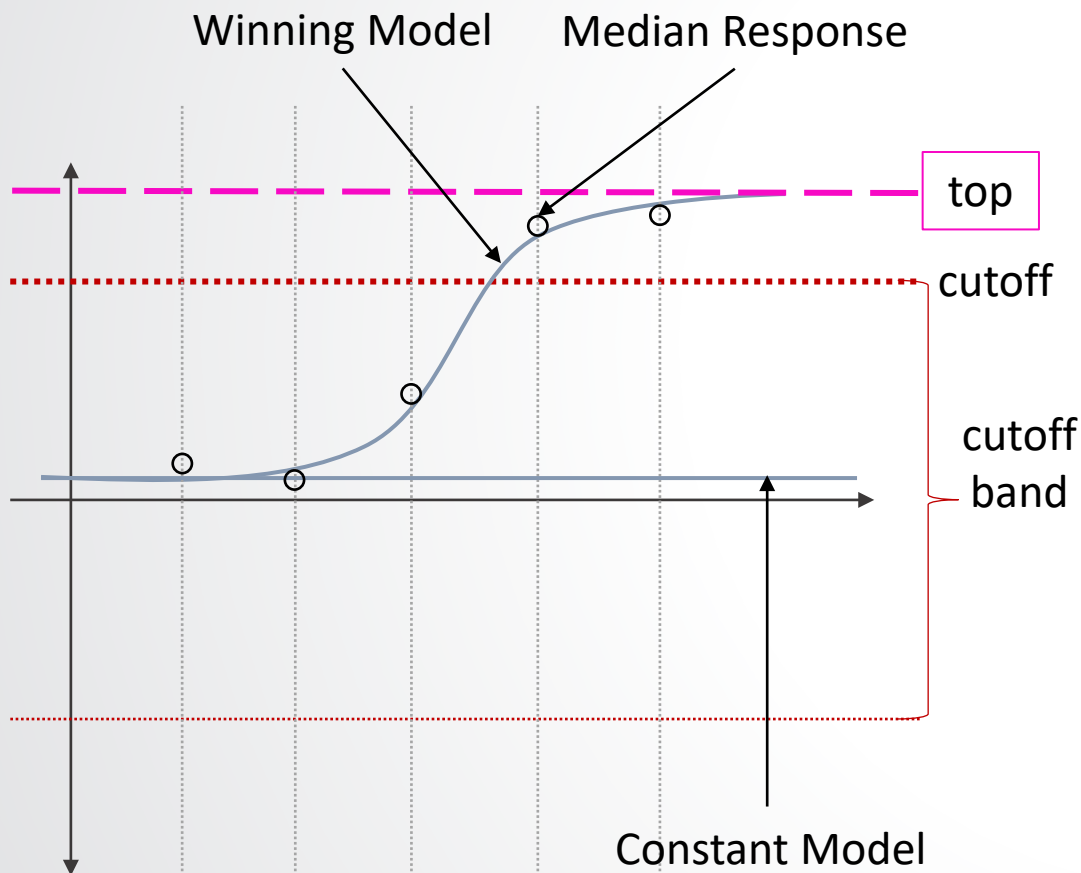
$p_2$ - address whether there are any median response values "outside" the cutoff





# Proportional Weight: $p_3$

$p_3$ - address relationship of how close the maximum predicted response is to the cutoff



\* Both cases here are examples of what should produce a 'high'  $p_3$  value. 'top' is outside the cutoff band.

# Updates in Hit Calling & Impacts

- **All data is fit bidirectionally** - *i.e.*, responses are not assumed to be in the positive direction
- Overwrite methods in *tcpl* multiply the hitc by -1 if the 'top' is not in the anticipated direction
  - This occurs after the initial hitc calculation
  - A negative hitc value allow us to indicate when a response occurs in biologically irrelevant direction
- **Minimal changes in aggregate hitc**

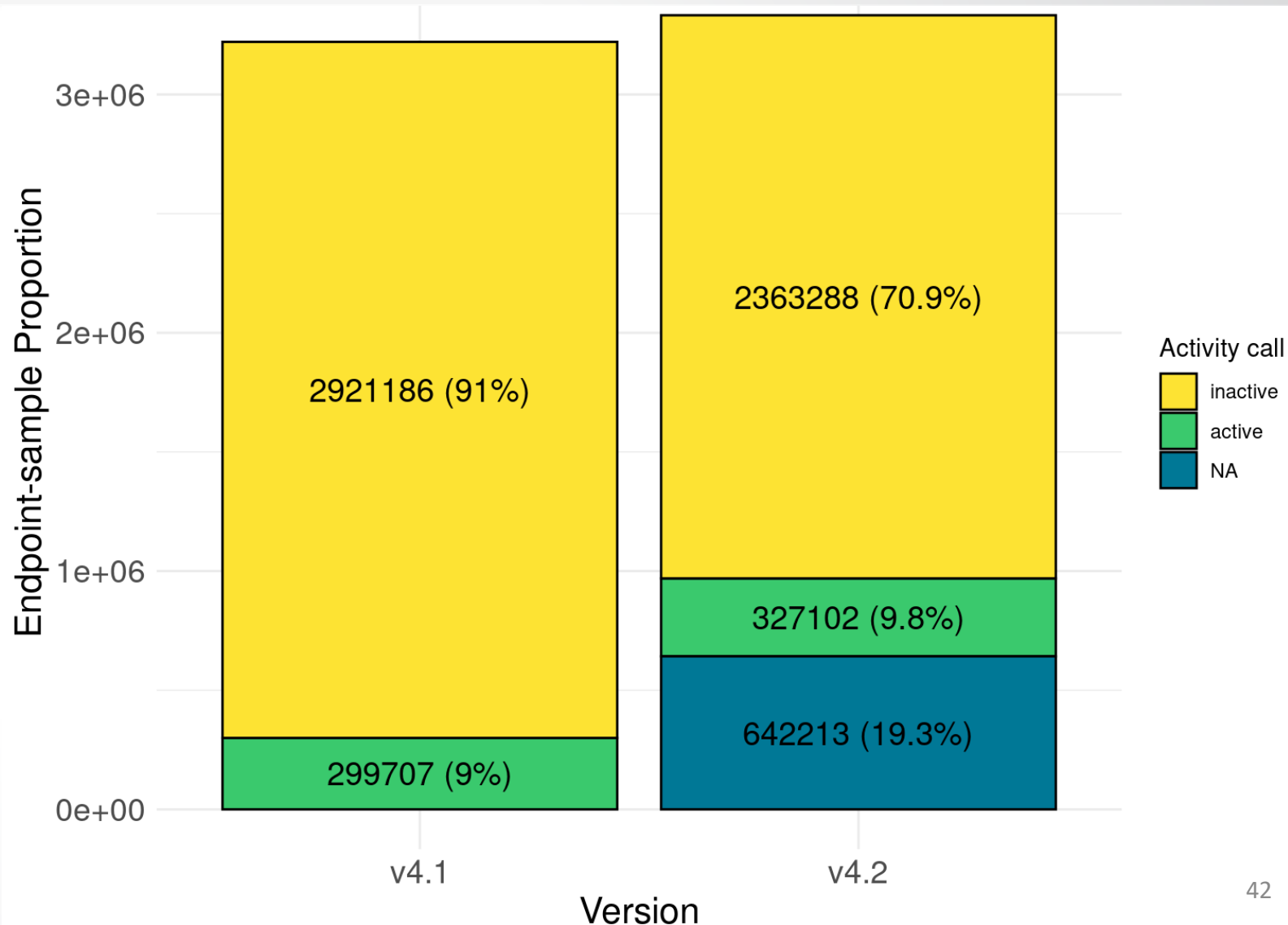
**Inactive Response:**

$0 \leq \text{hitc} < 0.9$

**Active Response:  $\text{hitc} \geq 0.9$**

**Overwrite Hit-call Methods:**

$\text{hitc} < 0 \Rightarrow \text{hitc} = \text{NA}$



What is the minimum concentration necessary to induce a meaningful biological activity?

**Metric(s): Activity Concentration (AC), Benchmark Dose (BMD)**



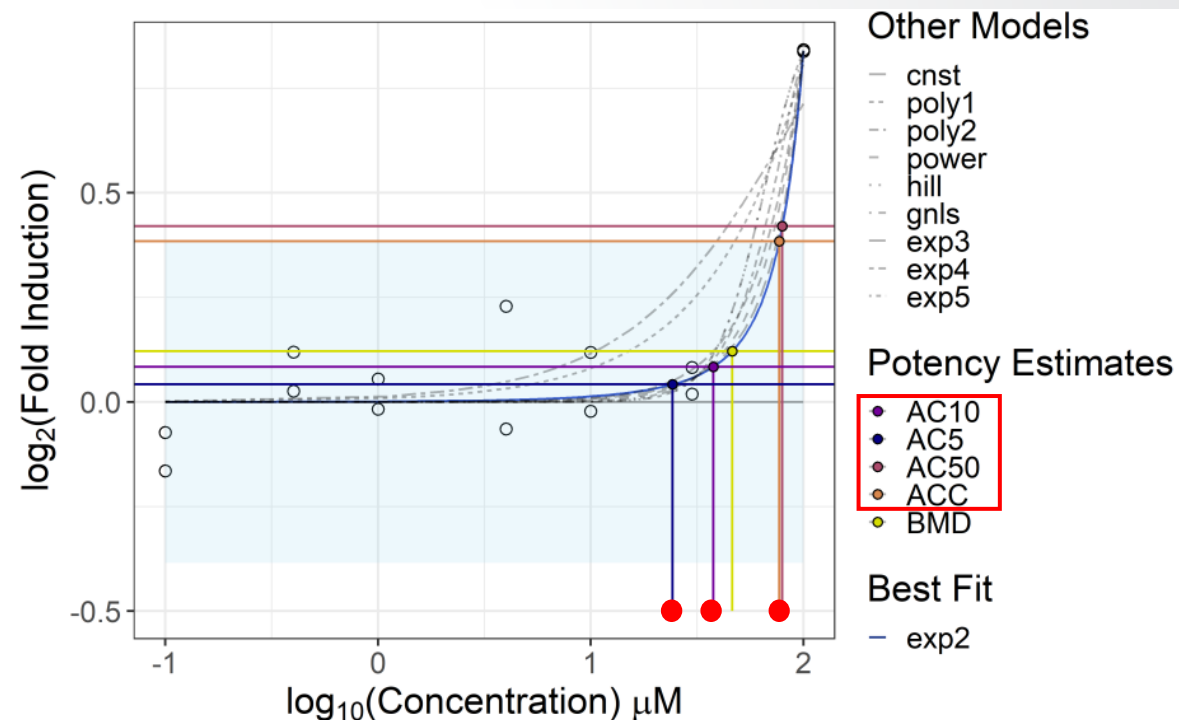
# Activity Concentrations

An activity concentration is the estimated concentration inducing a specified level of response (activity).

*tcpI* estimates and tracks several different activity concentrations.

| Activity Concentration (uM) | Specified Level of Response                     |
|-----------------------------|---|
| AC5                         | 5% of the maximal response                      |
| AC10                        | 10% of the maximal response                     |
| AC50                        | 50% of the maximal response                     |
| ACC                         | Response at the user-defined cutoff (threshold) |

*Additional potency metrics (not shown) are also computed and stored at level 5.*





# Benchmark Dose (BMD)

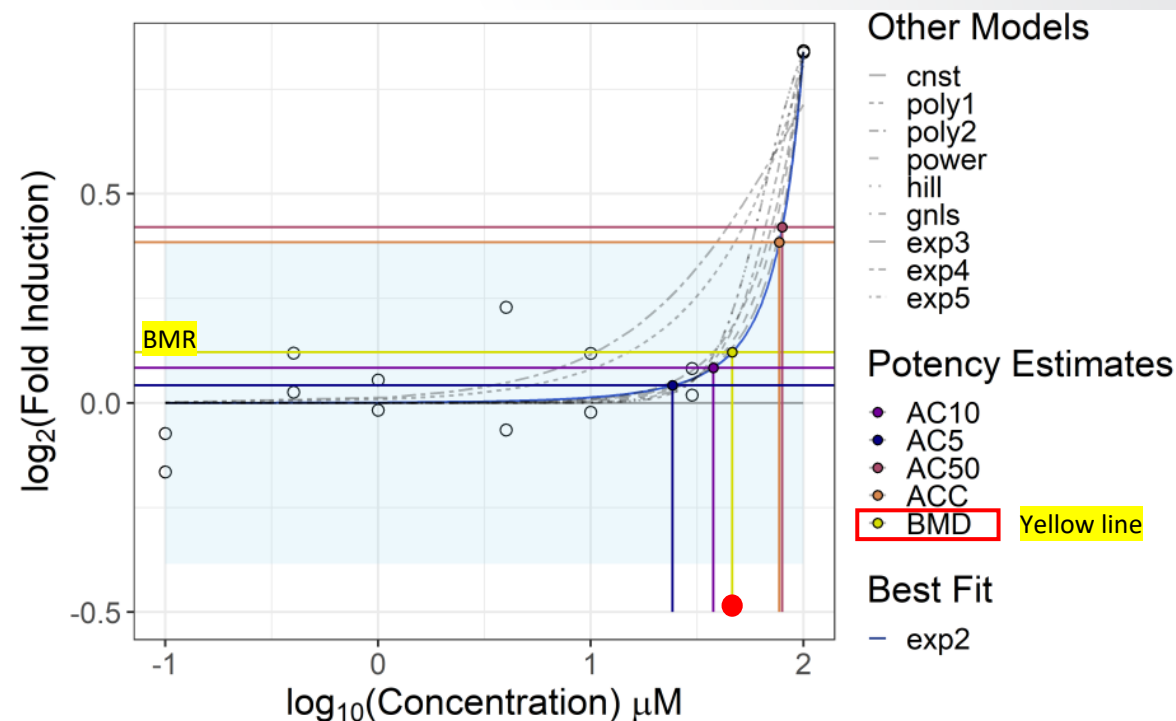
The benchmark dose (BMD) is the concentration inducing a specified benchmark response (BMR).

*tcpl* uses the following definitions and assumptions for setting the BMR:

- BMR is a change from the mean response at baseline ( $\mu(b)$ ) by some multiple ( $c$ ) of the standard deviation of the baseline ( $sd(b)$ ).

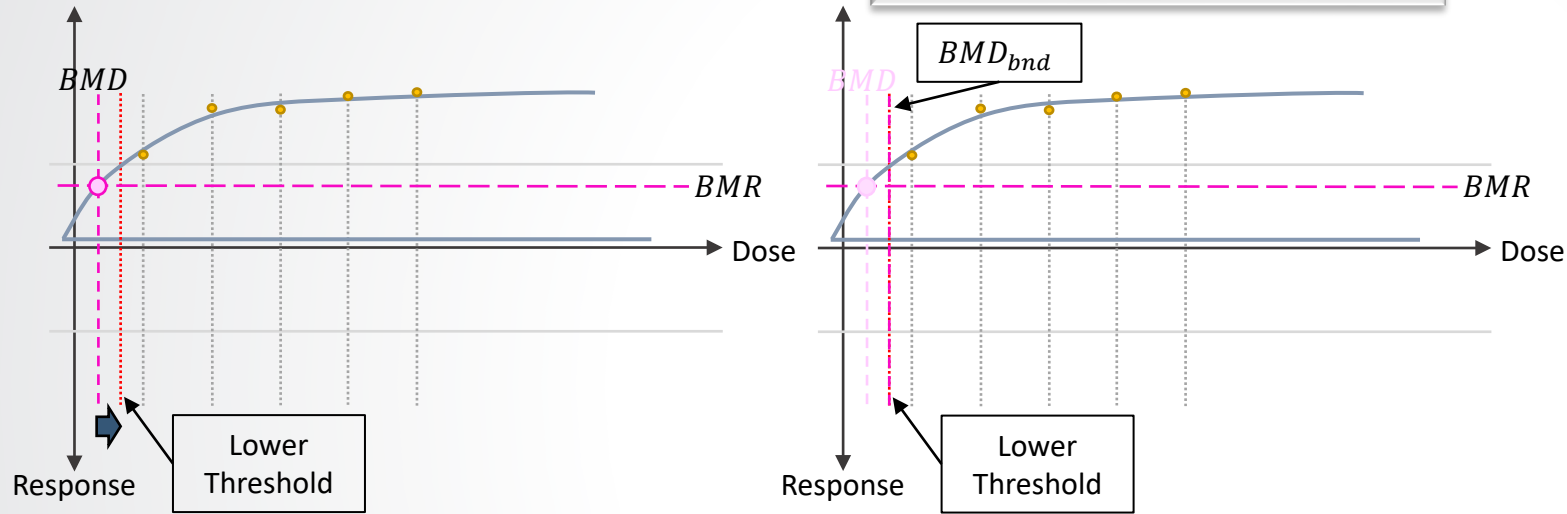
$$\mu(b) + c * sd(b) = BMR = \mu(BMD)$$

- Here, the baseline ( $b$ ) is defined as samples from the two lowest concentrations across chemicals within an assay endpoint and the  $c = 1.349^a$ .

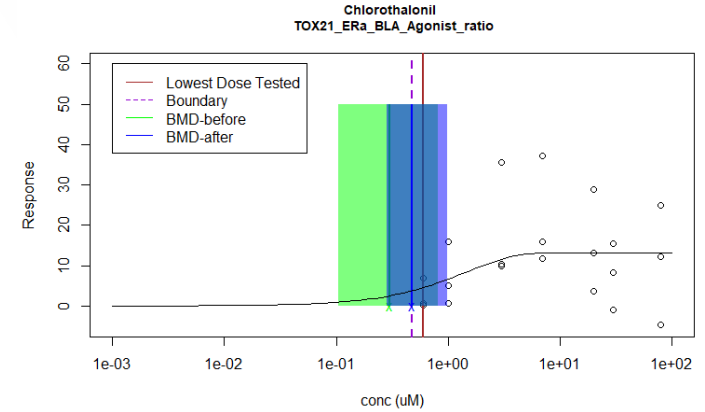
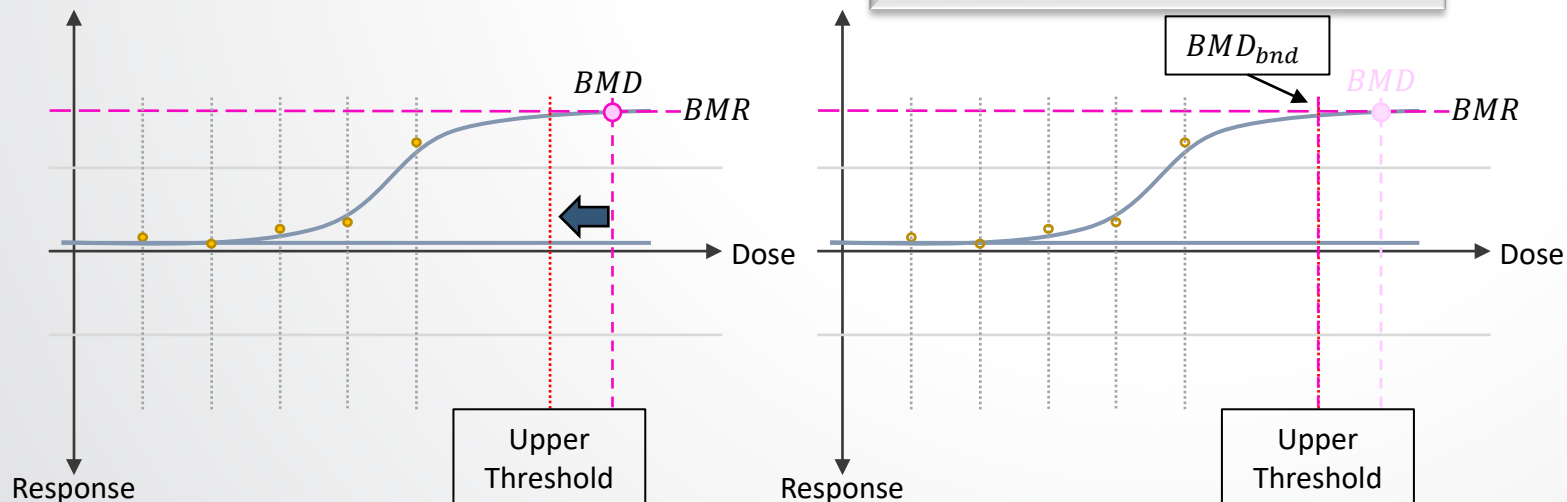


# BMD Bounding

## Lower Threshold Bounding

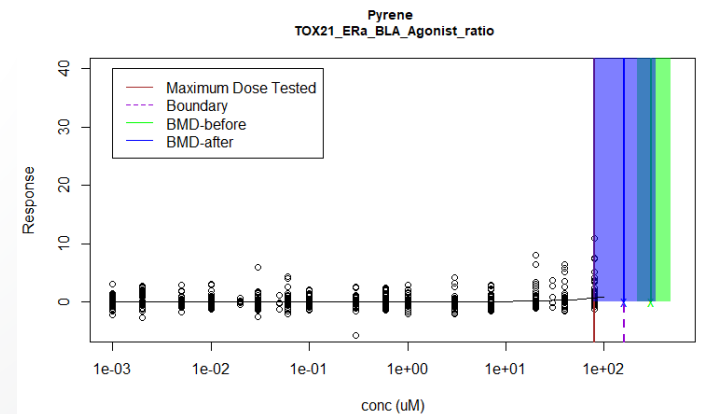


## Upper Threshold Bounding

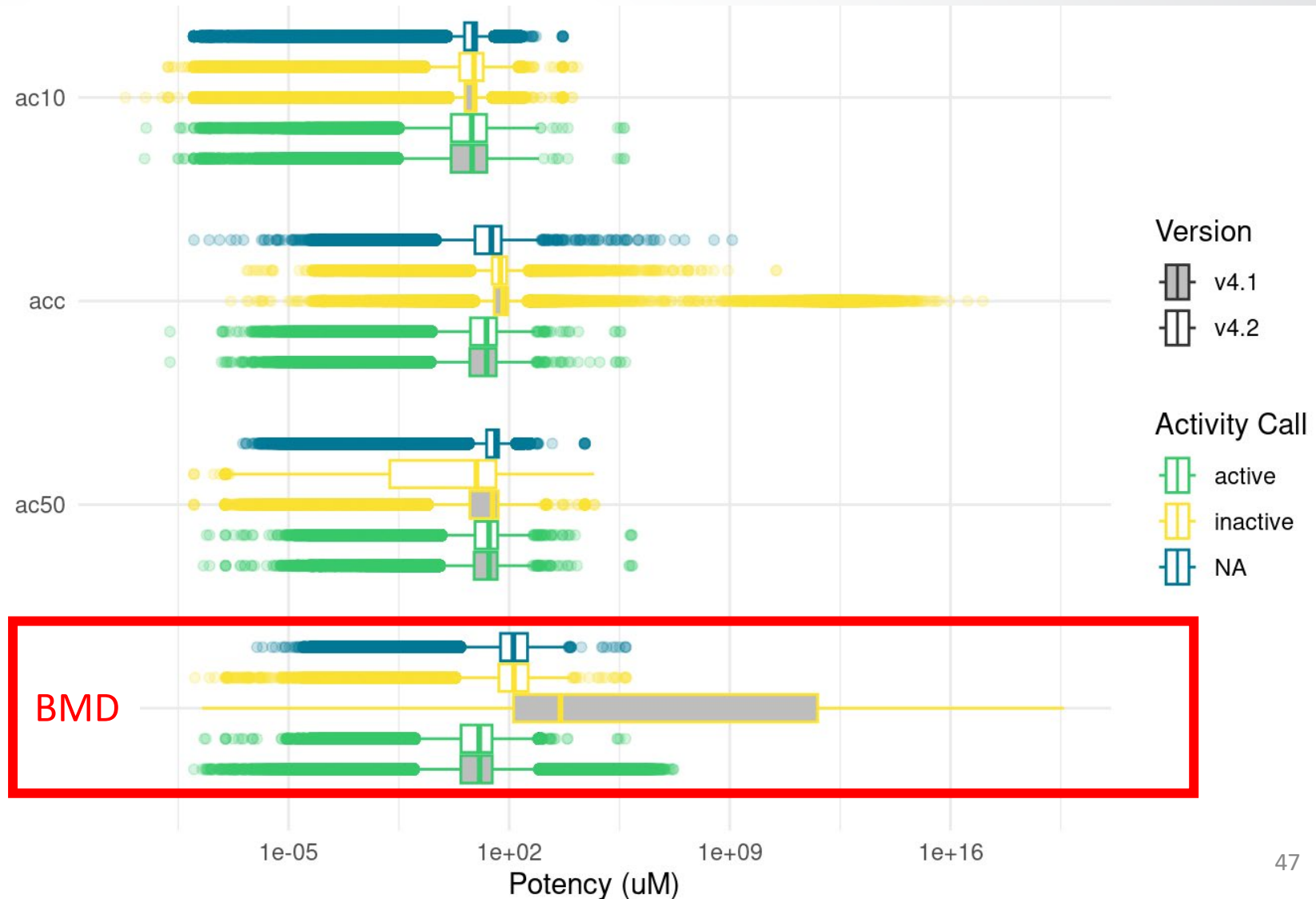


[tcplfit2 vignette: BMD bounding](#)

Fig. 6 & 8



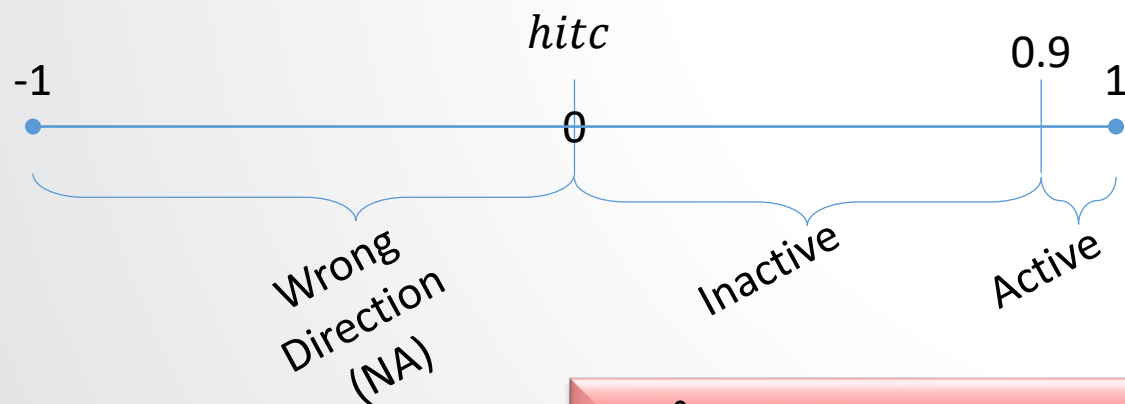
- Uncertainty in the BMD estimate is measured with a 90% confidence interval (CI):
  - BMDL – lower CI limit
  - BMDU – upper CI limit
- BMD thresholds were implemented to bound extreme BMD estimates outside the tested concentration range
  - Lower BMD threshold:*
    - $0.1 * \text{lowest conc}$
  - Upper BMD threshold:*
    - $10 * \text{highest conc}$
- invitrodb v4.2 – implementation of BMD bounding resulted in fewer extreme BMDs



# Activity & Potency Data Interpretation

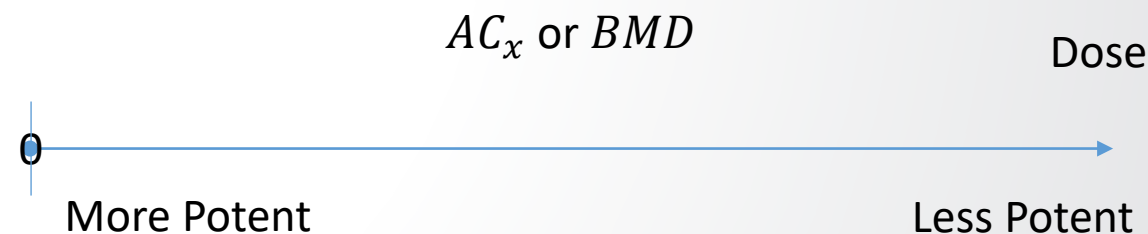
## Activity

Is the concentration-response curve we are fitting indicative of meaningful biological activity?



## Potency

What is the minimum concentration necessary to induce a meaningful biological activity?



Wider CI's indicate greater uncertainty in the BMD estimate



Chemical-endpoint pairs with  
 $hitc \geq 0.9$  &  $AC_x$  and/or  $BMD$  close to 0





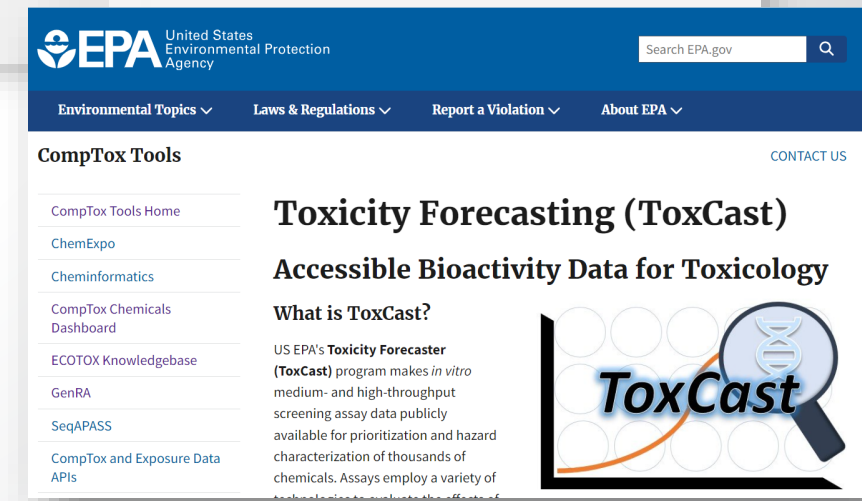
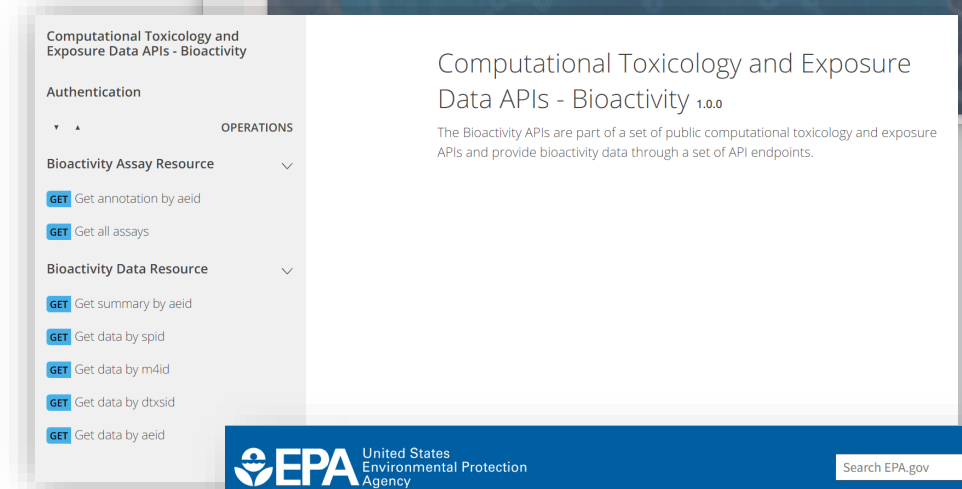
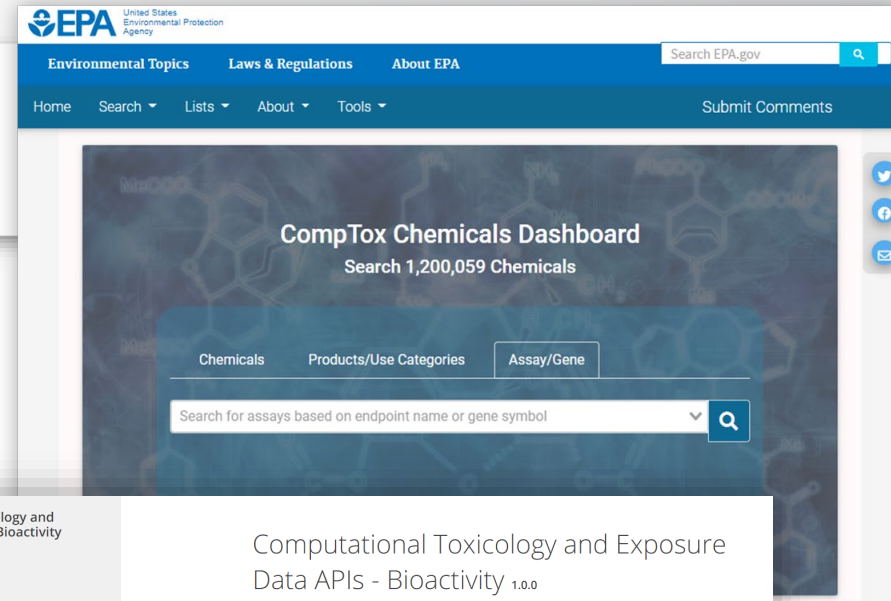
# Exploring ToxCast: Accessing Data & Plotting for Different Needs

Jason Brown and Carter Thunes



# Exploring ToxCast

- ToxCast data is accessible via:
  - [CompTox Chemicals Dashboard](#)
  - [Computational Toxicology and Exposure \(CTX\) Bioactivity API](#)
  - [Downloadable Data](#)

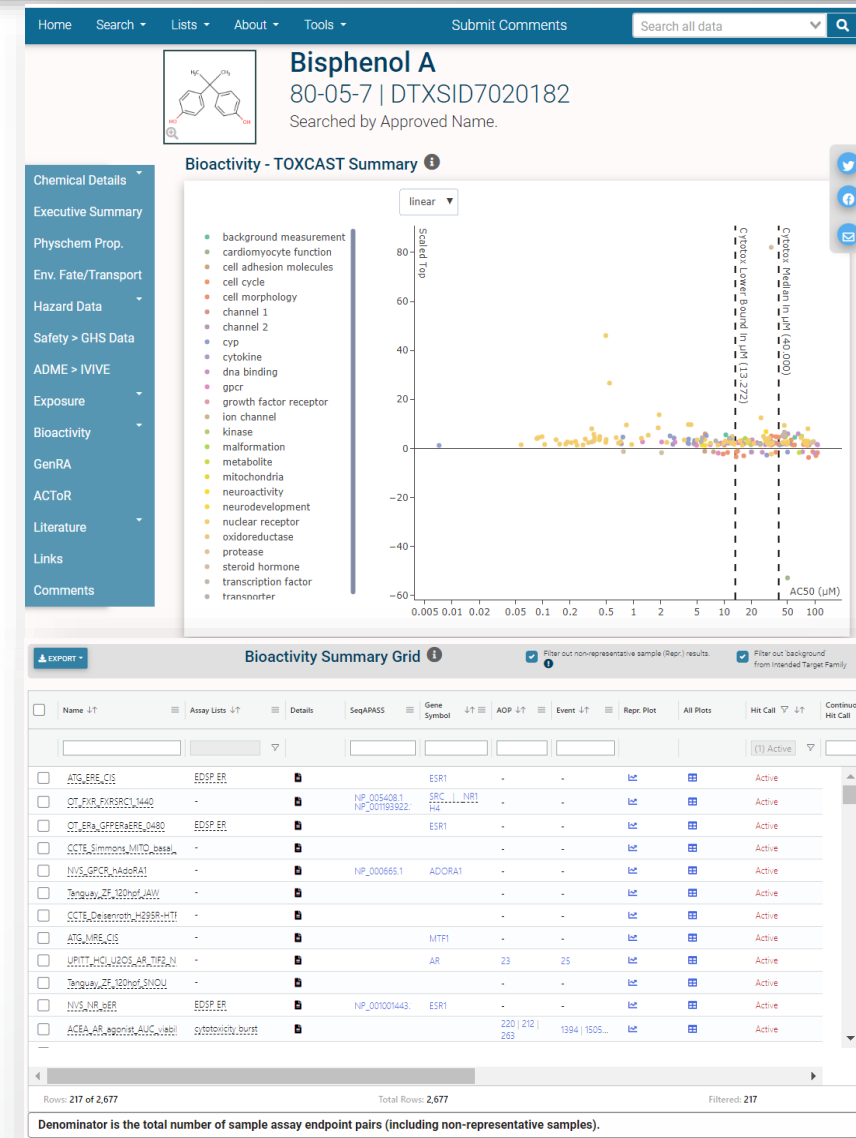




# CompTox Chemicals Dashboard(CCD)

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

- CCD's ToxCast bioactivity module presents a view of potency and relative efficacy metrics across ToxCast endpoints for chemicals of interest
- Users can easily sort, filter, and export ToxCast results and assay descriptions
- As of October 2024, ToxCast data reflects invitrodb v4.1



Example on right: Bisphenol A

<https://comptox.epa.gov/dashboard/chemical/invitrodb/DTXSID7020182>



# ToxCast Data Downloads

<https://www.epa.gov/comptox-tools/exploring-toxcast-data>

- Data downloads allow users to set up their own personal instance of the invitrodb MySQL database and interact with the data directly via the tcpl R package
- This is a preferred option for more customized or programmatic ToxCast data needs, or if users want to do their own data processing

## tcpl: ToxCast Data Analysis Pipeline

A set of tools for processing and modeling high-throughput and high-content chemical screening data. The package was developed for the the chemical screening data generated by the US EPA ToxCast program, but can be used for diverse chemical screening efforts.

Version: 3.1.0  
Depends: R (≥ 3.5.0)  
Imports: [data.table](#) (≥ 1.9.4), [DBI](#), [RMariaDB](#), [numDeriv](#), [RColorBrewer](#), [utils](#), [stats](#), [methods](#), [graphics](#), [grDevices](#), [sqldf](#), [dplyr](#), [tidyr](#), [plotly](#), [tcplfit2](#), [ggplot2](#), [gridExtra](#), [stringr](#)  
Suggests: [roxygen2](#), [knitr](#), [prettydoc](#), [rmarkdown](#), [htmlTable](#), [testthat](#) (≥ 3.0.0), [reshape2](#), [viridis](#), [kableExtra](#), [colorspace](#), [magrittr](#), [vdiff](#)  
Published: 2023-10-06  
Author: Richard S Judson [ctb, ths], Dayne L Filer [aut], Jason Brown [cre], Sarah E Davidson-Fritz [ctb], Madison Feshuk [ctb], Lori Kolaczowski [ctb], Kurt Dunham [ctb], Carter Thunes [ctb], Ashley Ko [ctb], Todd Zurlinden [ctb], Parth Kothiya [ctb], Woodrow R Setzer [ctb], Matthew T Martin [ctb, ths], Katie Paul Friedman [ctb]  
Maintainer: Jason Brown <brown.jason@epa.gov>  
License: [MIT](#) + file [LICENSE](#)  
URL: <https://github.com/USEPA/CompTox-ToxCast-tcpl>  
NeedsCompilation: no  
Materials: [NEWS](#)  
CRAN checks: [tcpl results](#)

## CompTox Tools

[CompTox Tools Home](#)

[ChemExpo](#)

[Cheminformatics](#)

[CompTox Chemicals Dashboard](#)

[ECOTOX Knowledgebase](#)

[GenRA](#)

[SeqAPASS](#)

[CompTox and Exposure APIs](#)

[Downloadable Computational Toxicology Data](#)

[Toxicity Forecasting \(ToxCast\)](#)

[Exploring ToxCast Data](#)

[Generating ToxCast Data](#)

[Chemical Coverage](#)

[Chemical Procurement Workflow](#)

[CONTACT US](#)

## Exploring ToxCast Data

On this page:

[Download ToxCast Data](#) | [Using ToxCast Data](#) | [Citations](#)

The US EPA Toxicity Forecaster (ToxCast) program is a world leader in providing an accessible data resource with bioactivity data for toxicology. ToxCast assays come from many sources, including external contracts, cooperative agreements, and internal research laboratories. These assays evaluate chemical effects on many biological processes, such as mitochondrial, developmental, neurological, and cell cycle function, as well as targets such as nuclear receptors, enzymes, transcriptional expression, etc. The ToxCast program takes heterogeneous assay data and computationally processes them into a singular resource that supports chemical evaluations, user interfaces such as the [CompTox Chemicals Dashboard](#), and research applications inside and outside of EPA.

This page provides links to all relevant ToxCast data. The most recent ToxCast data is available in the [invitrodb v4.2 database](#). The database was released in September 2024. Data files from previously published ToxCast data releases are still [available for download](#).

## Download ToxCast Data

### • Most Recent ToxCast Database Release

**(*invitrodb* v4.2) and Associated Software:** The ToxCast Data Analysis Pipeline R package (tcpl) processes, models, and visualizes ToxCast concentration-response data and populates a linked MySQL Database, *invitrodb*. The tcpl R package enables flexible data processing and analysis. Additionally, *tcpl* employs EPA-maintained R dependency packages: *tcplFit2* as its curve fitting utility and *ctxR* for API integration. The database release includes a MySQL database, release notes, summary files, assay information and concentration response plots. *tcpl* and *invitrodb* provide a standard for consistent and reproducible curve-fitting and data management for diverse, targeted in vitro assay data with readily available documentation, thus enabling sharing and use of these data in myriad toxicology applications.

• [Download Database Package](#)

• [Download the tcpl R Package](#)

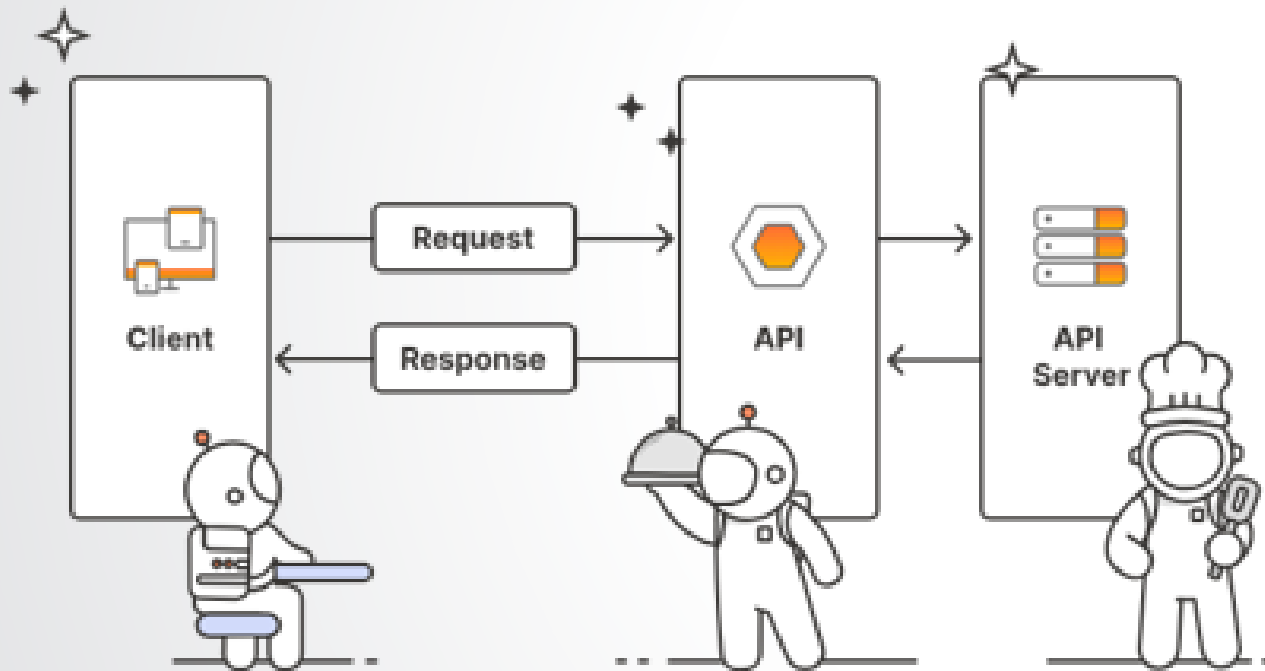
### Additional Resources

- [About ToxCast](#)
- [ToxCast Publications](#)



# Data Downloads Includes:

- Release note
  - Document describing changes in the database and software between versions
- Summary files
  - Zip file containing chemical by assay variable matrix output, *assay annotations*, *cytotoxicity burst output*, and *assay-gene mappings* in Excel and Rdata formats
    - *Excel files provided outside of zip file in addition for convenience*
  - Descriptions of all summary files included in README
- Assay Description Documents
  - Documents by Endpoint aligned to the OECD GD211 specifications
- Plots
  - Zip file of plots by assay source for all single and multi-concentration data
- MySQL data
  - Database dump file (.sql) with installation instructions available in README

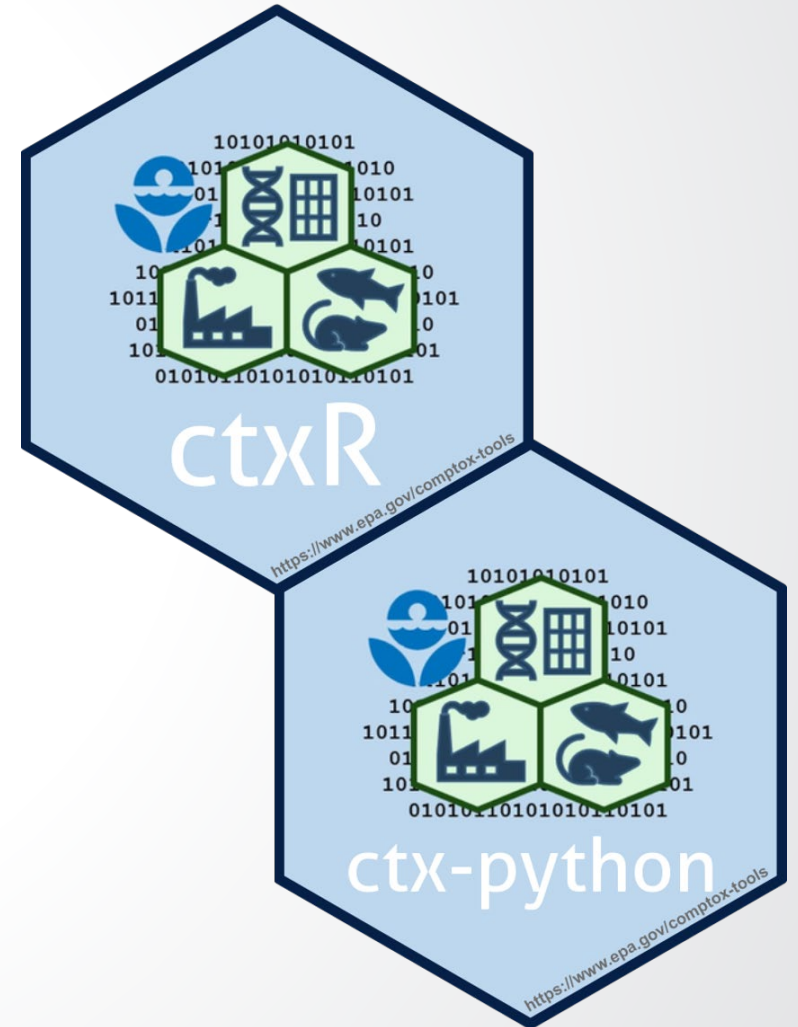


- An **Application Programming Interface (API)** allows software applications to communicate and exchange data through HTTP requests/responses
  - “Middleman” between a client and server
- Packages in different programming languages may wrap APIs
- May require user authentication



# CTX APIs and Clients

- **Computational Toxicology and Exposure APIs (CTX APIs)** make EPA's computational landscape easily available
  - Data domains include Bioactivity, Chemical, Exposure, and Hazard
  - Visit the [CTX APIs webpage](https://www.epa.gov/comptox-tools) for more information
- APIs are particularly useful for those with programmatic data needs, opposed to those where an interactive application like CCD is more useful
- Client packages (*ctxR* and *ctx-python*) facilitate API access





# Bioactivity API

- Sourced from the ToxCast's *invitrodb v4.1*
- Seven endpoints structured across the following resources:
  - Bioactivity Assay Resource
    - Annotations by aeid
    - Annotations for all assays
  - Bioactivity Data Resource
    - Summary data by aeid
    - Endpoint-sample data by aeid, spid (sample), m4id (endpoint-sample), and dtxsid (chemical)

## CompTox Tools

[CONTACT US](#)

[CompTox Tools Home](#)

[ChemExpo](#)

[Cheminformatics](#)

[CompTox Chemicals  
Dashboard](#)

[ECOTOX Knowledgebase](#)

[GenRA](#)

[SeqAPASS](#)

[CompTox and Exposure APIs](#)

[Downloadable  
Computational Toxicology  
Data](#)

## Computational Toxicology and Exposure APIs

Application Programming Interfaces (APIs) allow programmatic access to the EPA's computational toxicology and exposure data resources. APIs provided by EPA enable users to extract specific data from various databases and integrate them into their applications. APIs can effectively automate the process of accessing and downloading the data, such as data used to populate the [CompTox Chemicals Dashboard](#).

As part of EPA's commitment to provide "open data", these Computational Toxicology and Exposure APIs (CTX APIs) are publicly available data resources for anyone to access and use. These APIs are hosted on cloud.gov, a secure cloud environment managed by the General Services Administration specifically for U.S. federal government applications. These data are free of all copyright restrictions, and are fully and freely available for both non-commercial and commercial use. All data are also available for download on the [Downloadable Computational Toxicology Data page](#).

### Limited Access APIs

Users need an individual API key to access. Please [send an email](#) (ccte\_api@epa.gov) to request a FREE API key.

#### About CTX APIs

Learn about getting started with CTX APIs.

[About](#)

#### Data Domains

Learn more about available CTX API data domains.

[Domains](#)

#### Client Packages

Client API packages have been developed to support access to the CTX APIs.

[Access Clients](#)

### Computational Toxicology and Exposure Data APIs - Bioactivity

#### Authentication

#### OPERATIONS

#### Bioactivity Assay Resource ▾

[GET](#) Get annotation by aeid

[GET](#) Get all assays

#### Bioactivity Data Resource ▾

[GET](#) Get summary by aeid

[GET](#) Get data by spid

[GET](#) Get data by m4id

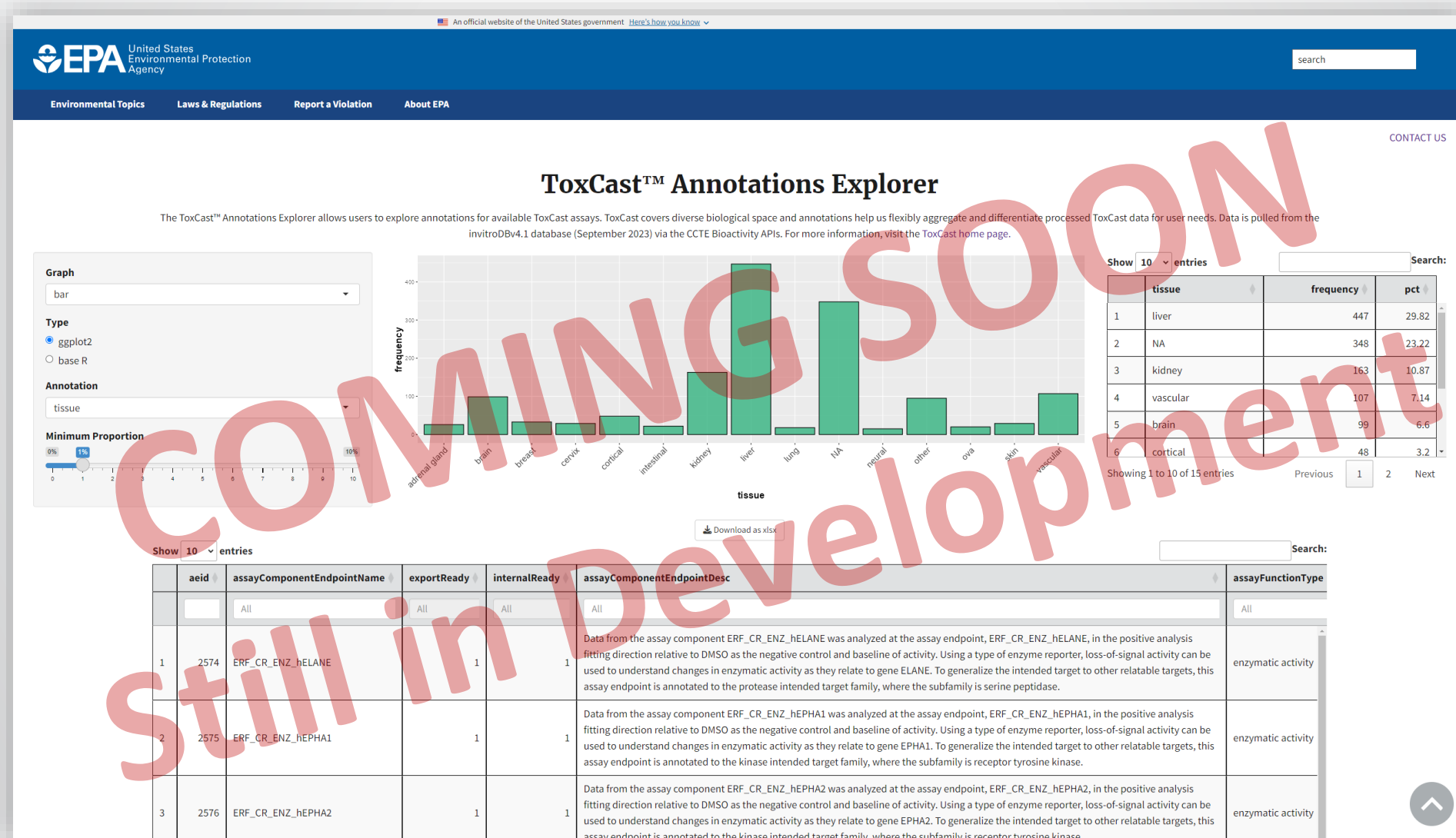
[GET](#) Get data by dtxsid

[GET](#) Get data by aeid



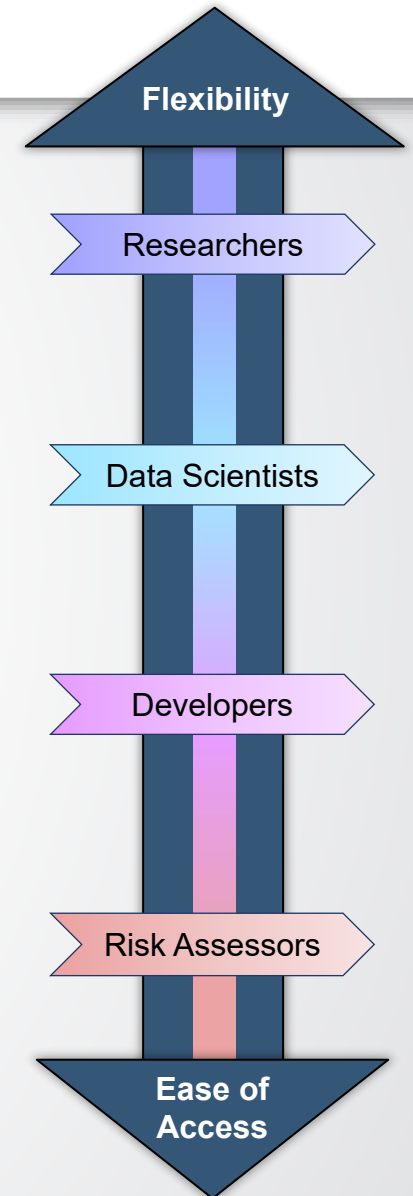
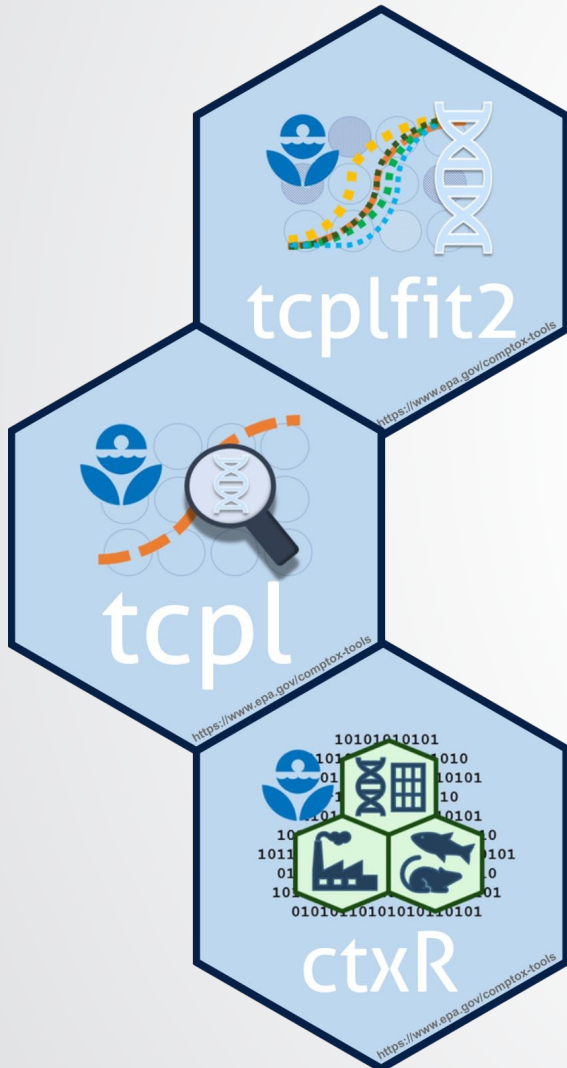
# ToxCast Annotations Explorer

Example RShiny application programmatically pulling annotations from the CTX Bioactivity API



# *tcpl-ctxR* Integration

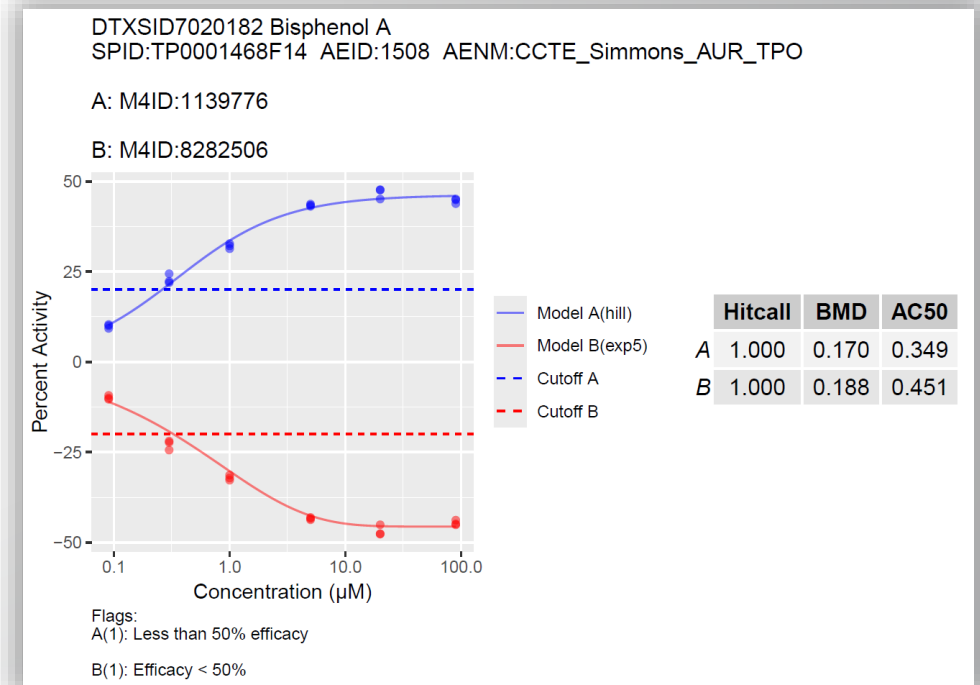
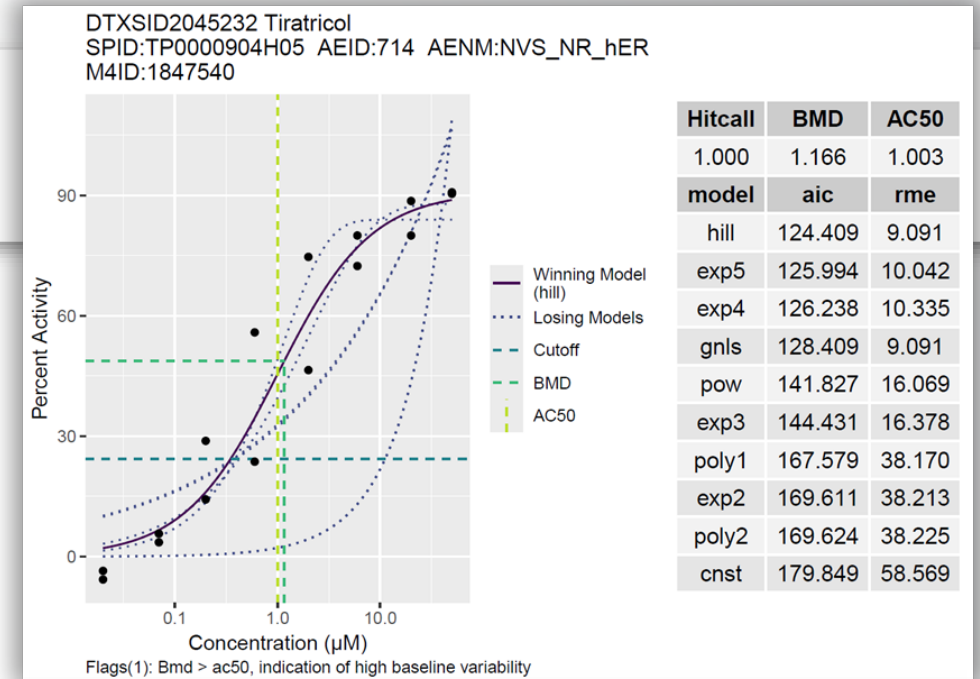
- ToxCast has a spectrum of users which require access to the bioactivity data in different environments
  - We have processing alternatives, but there was a need for a new standalone *tcpl* environment for non-database connections
    - *tcplLite*, the former standalone environment, which used static csv files as a source, is deprecated and *tcplFit2* package encouraged for standalone curvefitting applications
- Regular ToxCast users may find it easier to use *tcpl*, which has integrated *ctxR*'s bioactivity functions to make the API data retrievable in a familiar format.
  - However, several *ctxR* functions can be used to access the CTX Bioactivity API data if desired, but *tcpl* integration removes extra formatting steps





# Plotting Updates

- tcplPlot updates in *tcpl* v3.2
  - API connection supported
  - Simplified parameters
  - Options for cautionary flags to display on plots as well as consistent Y-axis (response) scaling
  - Single-concentration plotting
  - Comparison plotting
    - Advanced comparison plotting across data connections using tcplPlotLoadData





# ToxCast Tools for Users: Use Cases

Katie Paul Friedman



# Using ToxCast: from putative hazard to metrics of estimated risk

Hypotheses about **mode-of-action** or **adverse outcome pathways** (single or many chemicals)

*Explore a single chemical or build machine learning models for bioactivity*

Contributions to “**weight of evidence**” for mechanistic inference

*Many different bioactivities to examine plausible mode of action (e.g., effects on mitochondria, metabolism, development)*

Endocrine assessments using models and assay data

*A particularly common use case*

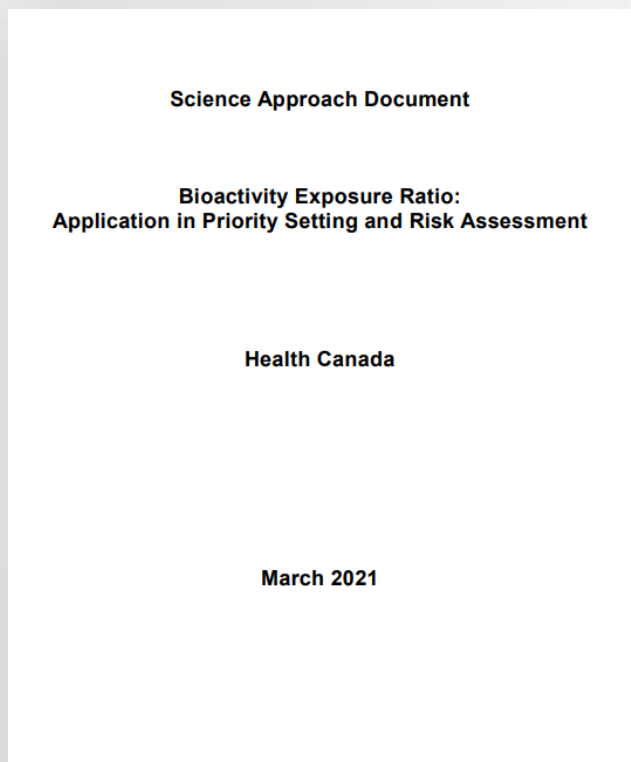
Calculation of a bioactivity-based **point of departure** (single or many chemicals)

Quantitative estimation of a risk metric: **bioactivity:exposure ratio (BER)**



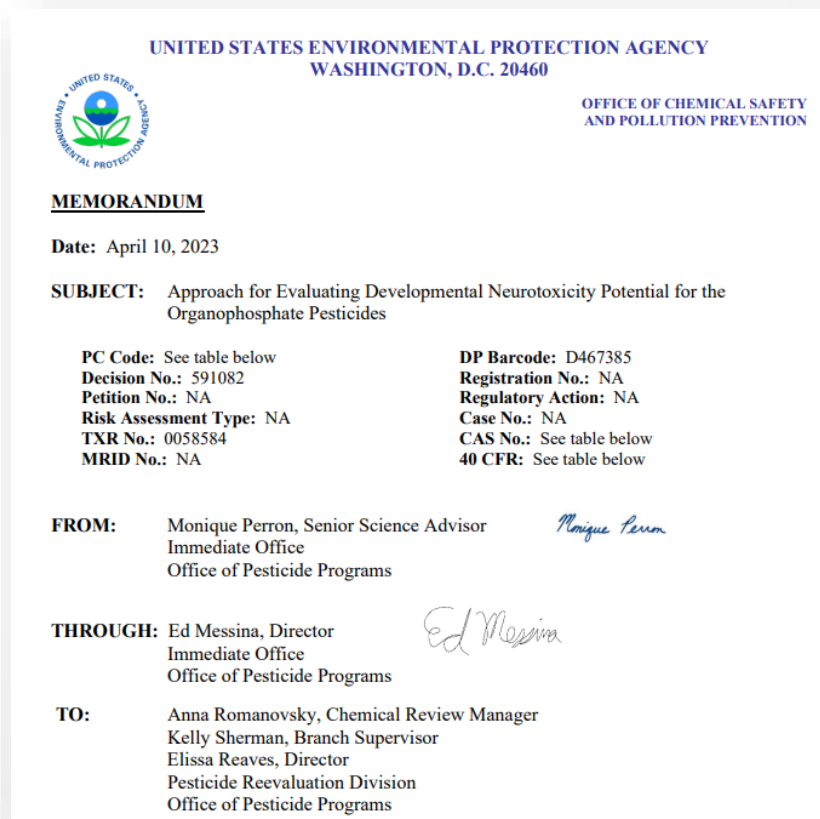
# ToxCast data inform research and regulatory applications

In managing risk evaluations of large inventories of chemicals, such as Canada's Domestic Substances List



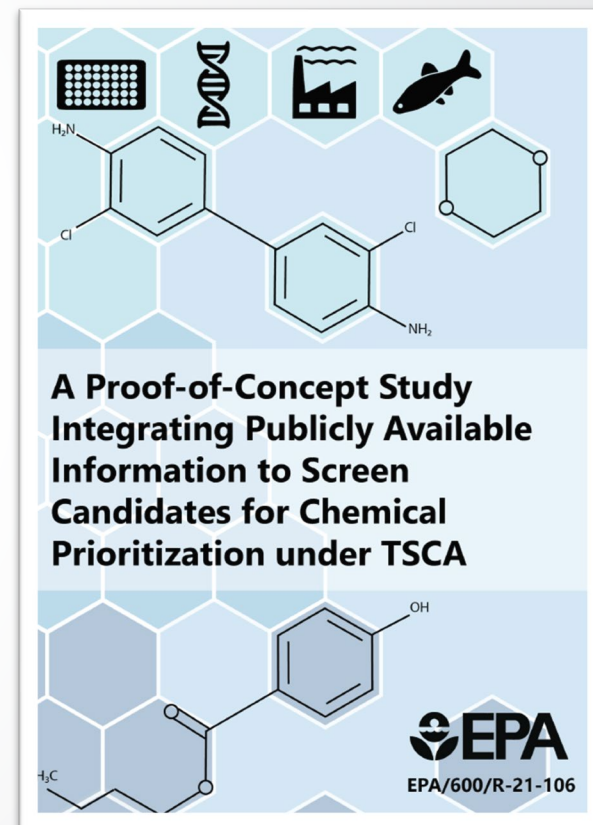
<https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/science-approach-document-bioactivity-exposure-ratio-application-priority-setting-risk-assessment.html>

In addressing data gaps within a weight-of-evidence for risk of developmental neurotoxicity of organophosphate insecticides



<https://www.regulations.gov/document/EPA-HQ-OPP-2008-0915-0056>

In proof-of-concept work to identify existing chemicals for further evaluation under the Toxic Substances Control Act



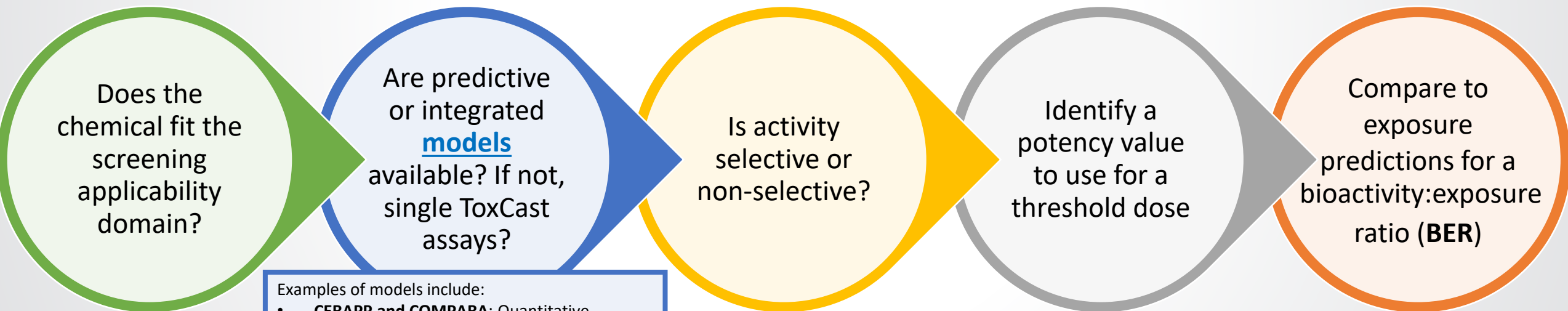
<https://www.epa.gov/sciencematters/proof-concept-case-study-integrating-publicly-available-information-screen>



# Examining a single substance for endocrine bioactivity

A generic workflow is illustrated which suggests examining:

- Amenability of the substance for screening and sample quality;
- Models or ToxCast assays available; and,
- Whether the activity is likely to be selective or not.



Examples of models include:

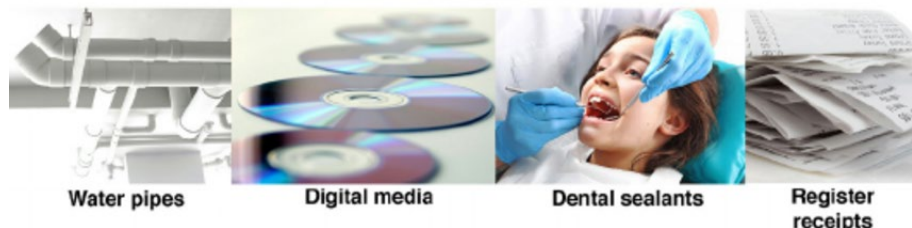
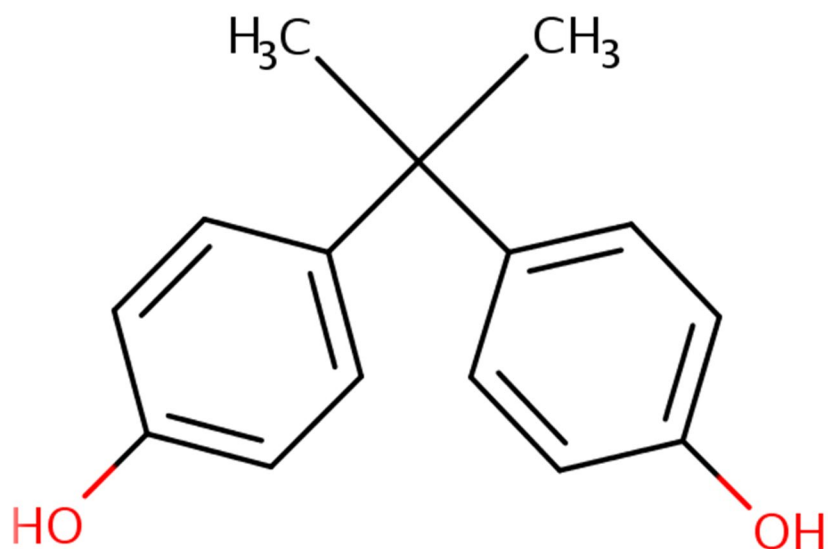
- **CERAPP and COMPARA:** Quantitative structure-activity relationship (QSAR)
- **ToxCast ER and AR Pathway Model:** Computational model of endocrine bioactivity data
- **ToxCast HT-H295R Model:** Statistical model of steroidogenesis activity



# Vignette Introduction

Madison Feshuk

# Chemical of Interest: Bisphenol A (BPA)



- Primarily industrial chemical used in production of many products
  - Polycarbonate plastics, epoxy resins, and vinyl ester resins
  - Food container liners, eyewear, pipes, sealants, dental composite, printed circuit boards
  - Minor uses as flame retardants used in plastic, antioxidant in brake fluid, thermal paper
  - Through recycling, BPA can reach other products
- Humans appear to be exposed primarily through food packaging manufactured using BPA, although those products account for less than 5 percent of the BPA used in US
  - Note that rapid metabolism to an inactive form in humans greatly lowers internal exposure of BPA
  - Food and Drug Administration (FDA) has purview over food packaging
- BPA remains prevalent in biomonitoring data (from sources such as NHANES) suggesting continuous exposure (largely food and water) due to lack of bioaccumulation
- There are concerns about its potential risk, particularly on children's health and the environment



# Vignette Links

- Toxicity Forecasting (ToxCast) home page <https://www.epa.gov/comptox-tools/toxicity-forecasting-toxcast>
  - Exploring ToxCast Data → Download Database Package
- Tcpl CRAN: <https://cran.r-project.org/web/packages/tcpl/index.html>
- Tcpl GitHub: <https://github.com/USEPA/CompTox-ToxCast-tcpl>
- CCD: <https://comptox.epa.gov/dashboard/>
  - Single Chemical Search “BPA” >Navigate to ToxCast tab> ToxCast Summary plot (AC50 vs Scaled Top (max modeled response/cutoff), cytotoxicity burst median and lower bounds)
  - Bioactivity grid (Adding additional fields like Annotations, Inspecting plots)
  - Search by gene “estrogen”
  - Search by assay “ACEA\_ER\_80hr”
  - Lists of Assay vs Chemicals > Send to Batch
  - Batch Search Export of ToxCast AC50 values
- CTX APIs <https://www.epa.gov/comptox-tools/computational-toxicology-and-exposure-apis>
- Bioactivity APIs <https://api-ccte.epa.gov/docs/bioactivity.html>
  - Overview of different request types
- • ctxR for accessing APIs <https://cran.r-project.org/web/packages/ctxR/index.html>



# Thank you for attending!



We love to hear from our users.  
Please reach out if you have questions or feedback.



**Madison Feshuk**

- *Lead, Pipelining and Curation*



**Jason Brown**

- *Lead, tcpl Development*



**Sarah Davidson-Fritz**

- *Lead, tcplfit2 Development and Statistics*



**Carter Thunes**

- *tcpl Development and Pipelining*
- *ORAU-SSC*



**Ashley Ko**

- *Pipelining and Curation*
- *ORAU-SSC*



**Paul Kruse**

- *Lead, ctxR Development*



**Kelly Carstens**

- *SME, DNT*



**Katie Paul Friedman**

- *SME, ToxCast Project Lead*



**Colleen Elonen**

- *SCDCD Project Liaison*



# Assay Lists

- Assay Lists organize ToxCast endpoints which have been studied in CCTE publications and/or other projects.
- In addition to updating existing assay lists in *invitrodb*, new assay lists were added

| List Name                          | List Description   | Endpoint Count |
|------------------------------------|--|----------------|
| Cytotoxicity Burst                 | Assays used to define the cytotoxicity burst region  | 76             |
| ToxCast ER Pathway Model           | Estrogen receptor assays used in ToxCast ER Pathway model  | 21             |
| ToxCast AR Pathway Model           | Androgen receptor assays used in ToxCast AR Pathway model. See 10.1016/j.yrtph.2020.104764 and 10.1021/acs.chemrestox.6b00347  | 17             |
| Thyroid Bioactivity                | Assays related to the thyroid adverse outcome pathway network  | 60             |
| <i>Steroidogenesis Bioactivity</i> | Assays related to steroidogenesis  | 29             |
| <i>Developmental Toxicity</i>      | Assays associated with developmental toxicity  | 75             |
| <i>DNT-IVB</i>                     | Assays included in the developmental neurotoxicity in vitro battery, see <a href="https://www.regulations.gov/document/EPA-HQ-OPP-2020-0263-0006">https://www.regulations.gov/document/EPA-HQ-OPP-2020-0263-0006</a> | 68             |
| <i>Immunosuppression</i>           | Assays associated with markers of immunosuppression, see 10.14573/altex.2203041  | 6              |
| <i>Cardiotoxicity</i>              | Assays associated with cardiotoxicity  | 5              |
| <i>Non-mammalian Vertebrate</i>    | Assays associated with non-mammalian vertebrate species  | 98             |

Note that endpoints may be associated with multiple assay lists



## Level 7 - Administered Equivalent Doses

- “mc7” previously could be used to understand uncertainty by bootstrapping ToxCast data with “toxboot”. This was deprecated by BMD bounds in tcpl v3.
- As the next extension to the ToxCast pipeline to assist with interpretation, the new “mc7” provides users with pre-calculated estimates of the *in vivo* human administered equivalent dose (AEDs, in mg/kg/day) based on the *in vitro* bioactive concentrations using several potency metrics (AC50, ACC, and BMD) and a subset of models from the [High-throughput Toxicokinetics R package httk](#).

| m7id | m4id     | aeid | potency_val_type | aed_type        | aed_val | aed_val_unit | interindividual_var_perc | httk_model     | invitrodb_version | httk_version | created_date     |
|------|----------|------|------------------|-----------------|---------|--------------|--------------------------|----------------|-------------------|--------------|------------------|
| 1    | 12485543 | 64   | ac50             | aed.ac50.3css50 | 38.05   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 2    | 12499190 | 67   | ac50             | aed.ac50.3css50 | 38.82   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 3    | 12503739 | 68   | ac50             | aed.ac50.3css50 | 46.35   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 4    | 12508288 | 69   | ac50             | aed.ac50.3css50 | 31.11   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 5    | 12512837 | 70   | ac50             | aed.ac50.3css50 | 23.17   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 6    | 12517386 | 71   | ac50             | aed.ac50.3css50 | 43.69   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 7    | 12521935 | 72   | ac50             | aed.ac50.3css50 | 27.8    | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 8    | 12531033 | 74   | ac50             | aed.ac50.3css50 | 47.98   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |
| 9    | 12558327 | 80   | ac50             | aed.ac50.3css50 | 65.21   | mg/kg/day    | 50                       | 3compartmentss | invitrodb_v_4_2   | httk_v_2_3_1 | 2024-06-21 20:00 |



# Analytical QC and Applicability Domain

- The “chemical\_analytical\_qc” table of invitrodb is a resource of applicability domain information at the substance and sample level curated from a retrospective analysis of the analytical QC data for the ToxCast/Tox21 chemical library.
  - Consult full table for additional details, including cautionary flags and annotations, as well as [tcpl’s vignette](#), under Data Interpretation>Analytical QC & Applicability Domain, for more information.
- v4.2 include **7878 substance-level** and **24159 sample-level calls**. 6231 (79.09%) substance-level calls pass, whereas 19246 (79.66%) of sample-level calls pass.

| dsstox_substance_id | chnm                      | spid         | qc_level  | pass_or_caution | t0 | t4 | call | annotation   | flags      | average_mass  | log10_vapor_pressure_OP | logKow_octanol | created_date |
|---------------------|---------------------------|--------------|-----------|-----------------|----|----|------|--------------|------------|---------------|-------------------------|----------------|--------------|
| DTXSID0045412       | Benserazide hydrochloride | NA           | substance | pass            | A  | Fc | T    | NULL         | Room te... | 293.0778483   | -10.416897620621151     | -1.96662       | 2024-08-05   |
| DTXSID0045412       | Benserazide hydrochloride | NULL         | sample    | pass            | A  | Fc | T    | NULL         | Room te... | 293.0778483   | -10.416897620621151     | -1.96662       | 2024-08-05   |
| DTXSID7024164       | Bensulfuron-methyl        | NA           | substance | caution         | C  | D  | T    | Evidence ... | Room te... | 410.08962011  | -13.671914103390696     | 2.28924        | 2024-08-05   |
| DTXSID7024164       | Bensulfuron-methyl        | NULL         | sample    | caution         | C  | D  | T    | Evidence ... | Room te... | 410.08962011  | -13.671914103390696     | 2.28924        | 2024-08-05   |
| DTXSID9032329       | Bensulide                 | NA           | substance | pass            | A  | A  | S    | NULL         |            | 397.060508777 | -6.095936656340898      | 4.20853        | 2024-08-05   |
| DTXSID9032329       | Bensulide                 | Tox21_300733 | sample    | pass            | A  | A  | S    | NULL         |            | 397.060508777 | -6.095936656340898      | 4.20853        | 2024-08-05   |
| DTXSID9032329       | Bensulide                 | TP0000235H12 | sample    | pass            | A  | A  | S    | NULL         |            | 397.060508777 | -6.095936656340898      | 4.20853        | 2024-08-05   |
| DTXSID9032329       | Bensulide                 | TP0000248H12 | sample    | pass            | A  | A  | S    | NULL         |            | 397.060508777 | -6.095936656340898      | 4.20853        | 2024-08-05   |
| DTXSID9032329       | Bensulide                 | TP0000821H06 | sample    | pass            | A  | A  | S    | NULL         |            | 397.060508777 | -6.095936656340898      | 4.20853        | 2024-08-05   |
| DTXSID9032329       | Bensulide                 | TP0000971D01 | sample    | pass            | A  | A  | S    | NULL         |            | 397.060508777 | -6.095936656340898      | 4.20853        | 2024-08-05   |
| DTXSID9032329       | Bensulide                 | TP0001194L16 | sample    | pass            | A  | A  | S    | NULL         |            | 397.060508777 | -6.095936656340898      | 4.20853        | 2024-08-05   |



# Chemical Lists

- These *chemical\_lists* table contains lists curated by the US EPA in the Distributed Structure-Searchable Toxicity (DSSTox) database.
  - This is a convenience table that is refreshed by stored procedure along with *sample* table from ChemTrack and *chemical* table from DSSTox. See CTX Chemical API or data downloads for versioned data
- The **tcplLoadChemList** function allows the user to subdivide the chemical IDs based on presence in different chemical lists.
- This replaces deprecated *chemical\_library* table and **tcplLoadCLib** function

| chemical_lists_id | chid  | dsstox_substance_id | list_acronym | list_name                                      | list_desc                                      |
|-------------------|-------|---------------------|--------------|--|--|
| 1                 | 20005 | DTXSID7020005       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 2                 | 20006 | DTXSID2020006       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 3                 | 20007 | DTXSID7020007       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 4                 | 20009 | DTXSID7020009       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 5                 | 20014 | DTXSID6020014       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 6                 | 20015 | DTXSID1020015       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 7                 | 20020 | DTXSID0020020       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 8                 | 20021 | DTXSID5020021       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |
| 9                 | 20022 | DTXSID0020022       | CPDBAS       | Carcinogenic Potency Database Summary Table... | Carcinogenic Potency Database Summary Table... |



# New Endpoints and Data

- **67 new endpoints** released with multiple-concentration (MC) data, including:
  - DevTox Germ Layer Reporter assay from [CCTE Deisenroth Lab](#)
  - Zebrafish developmental assays from [CCTE Padilla](#) and [Oregon State University Tanguay Labs](#)
  - Assays from Toxicology in the 21st Century (TOX21) federal consortium including the following targets: Acetylcholinesterase (AChE), hERG potassium channel, and various cytochromes P450 (CYPs)
  - New Eurofins (ERF) nuclear receptor (NR) assays, including estrogen receptors, peroxisome proliferator-activated receptors (PPAR), Retinoid X receptors (RXRs)
- **8 new endpoints** released with single-concentration (SC) data
- **Plus, new data to existing endpoints** (ATG, DNT)

**Total: *111,710 MC + 42,569 SC endpoint-samples added***



# Flag Table Decoding

| Level 6 Flag Name        | Level 6 Flag Description   |
|--------------------------|--|
| modl.directionality.fail | Model directionality is questionable as data points are split in positive and negative axis. tcplFit2 models assume data is zero-centered and the absolute response is increasing  |
| low.nrep                 | Average number of replicates per conc is less than 2   |
| low.nconc                | Number of concentrations tested is less than 4   |
| bmd.high                 | Bmd > ac50, indication of high baseline variability  |
| singlept.hit.high        | Only highest conc above baseline, active   |
| singlept.hit.mid         | Only one conc above baseline, active   |
| multipoint.neg           | Multiple points above baseline, inactive   |
| gnls.lowconc             | Complete gain-loss curve not within concentration range tested, as the "Gain" AC50 less than lowest concentration tested or the "Loss" AC50 greater than mean concentration tested |
| noise                    | Noisy data (rme>coff)  |
| border                   | Borderline activity with top $\leq 1.2 \cdot \text{coff}$ or top $\geq 0.8 \cdot \text{coff}$  |
| efficacy.50              | Less than 50% efficacy   |
| ac50.lowconc             | AC50 less than lowest concentration tested   |
| viability.gnls           | Cell viability assay fit with gain-loss winning model  |



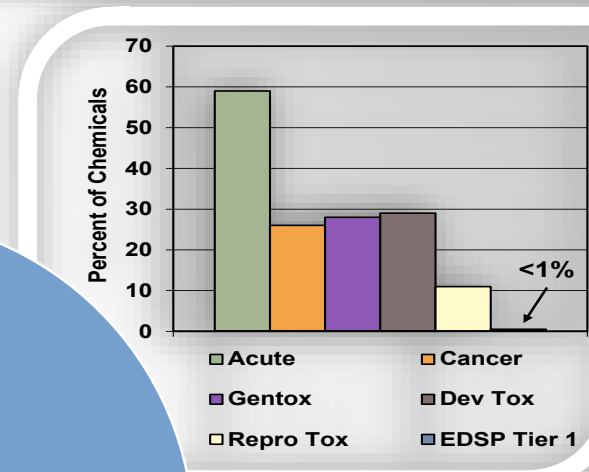
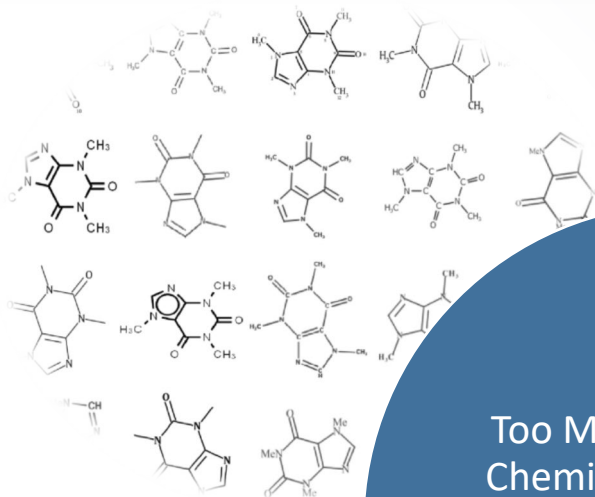
## v4.1 to v4.2 Schema Changes

- Additional summary statistics are now stored in the *mc4* table to appropriately consider the directionality of the response
- Processing now avoids having to switch between logged and unlogged concentration values between levels (“logc” → “conc”)
- Deprecated tables have been removed:
  - *chemical\_library*, *pfas*, and *chemical\_assay\_count*
- Existing tables have been updated:
  - *mc7*, *assay\_descriptions*, and *sample*
- New tables added:
  - *chemical\_lists* and *chemical\_analytical\_qc*

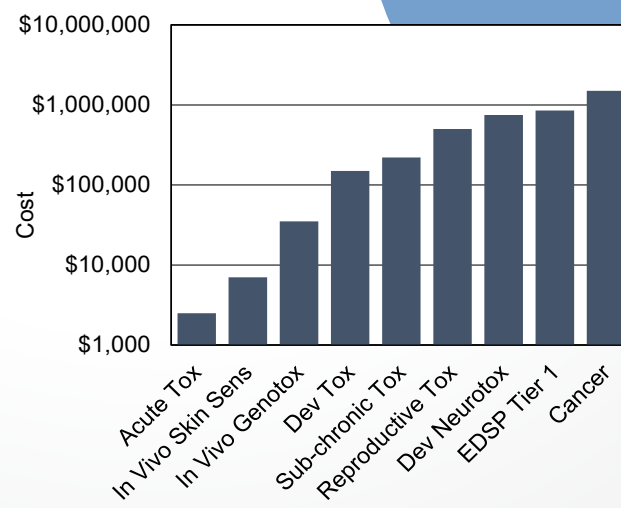


# Extra Slides

- There are a number of limitations to traditional toxicology testing
- EPA needs rapid and efficient methods to prioritize, evaluate and regulate thousands of chemicals in commerce
- CompTox Blueprint outlines a tiered testing strategy for hazard characterization
  - Tier 1: Broad profiling, high content assays
  - Tier 2: Targeted *in vitro* assays (e.g. **ToxCast**)
  - Tier 3: Confirmation using assays of greater biological complexity (e.g. **ToxCast**)



Modified from Judson *et al.*, EHP 2010





# ToxCast Database Coverage

The **Toxicity Forecaster (ToxCast)** program curates and makes publicly available targeted bioactivity screening data. Latest database release (v4.2) includes:

## *Change from v4.1:*

*0 Sources*  
*+32 Assays*  
*+71 Endpoints*  
*+55 Chemicals*  
*+23 Gene Mappings*

26 Assay Sources

655 Unique Assays

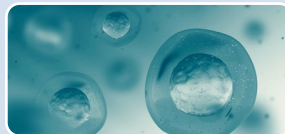
1570 Unique Endpoints

9614 Chemicals

*Including a TOX21 assay source for data generated by the TOX21 program*



Diverse biology with **over 500 mapped gene targets**, including:



**Endocrine-Related:** Estrogen Receptor, Androgen Receptor, Thyroid, Steroidogenesis



**Cellular Signaling Pathways:** Cytotoxicity, Proliferation, Stress, Mitochondrial Toxicity



**Protein Interactions:** Receptors, Transporters, Ion Channels, Enzymes



**Complex Responses, e.g.** Immune Response, Development, Neurotoxicity, etc.

## Activity

Is the concentration-response curve we are fitting indicative of meaningful biological activity?

- Continuous hitcalls close to 0 indicate little to no biological activity in the tested concentration range
- Continuous hitcall close to 1 indicates a plausible biological activity within the tested concentration range

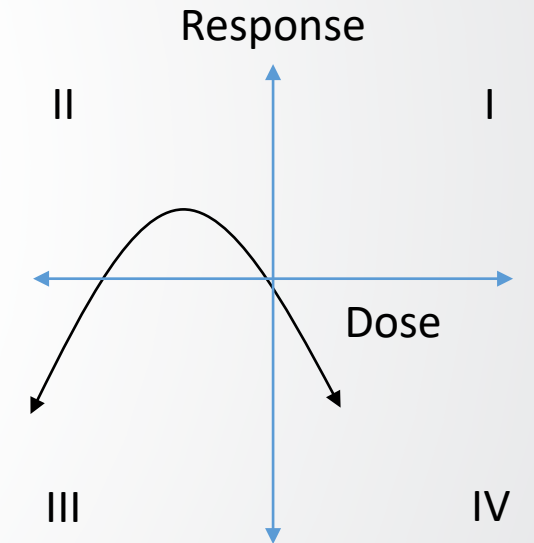
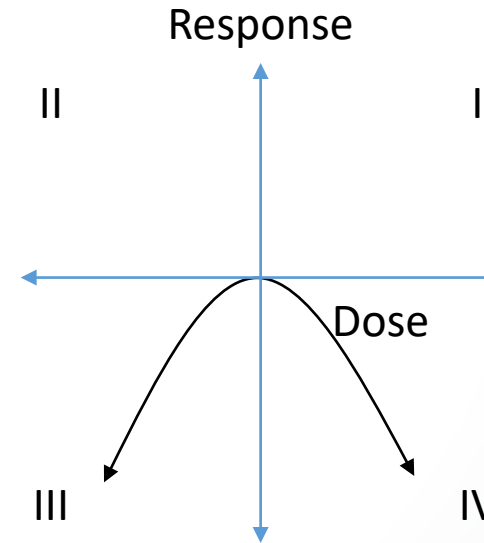
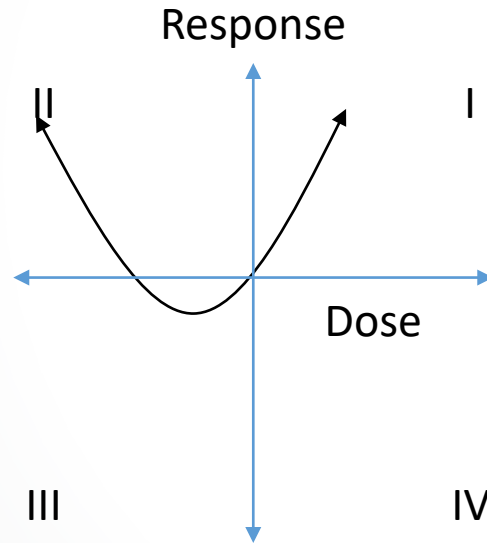
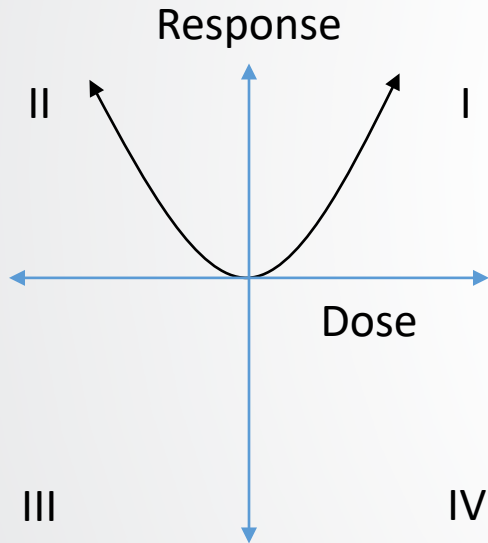
## Potency

What is the minimum concentration necessary to induce a meaningful biological activity?

- Lower BMD/AC estimates indicate more potency – i.e. smaller amounts the chemical induce a meaningful biological response
- Wider 90% CI's indicate more uncertainty in the BMD estimate

# *tcplfit2* Polynomial 2 Implementation

$$f(x) = \frac{\alpha}{\beta}x + \frac{\alpha}{\beta^2}x^2 + \epsilon$$



Vertex

$$x_v = -\frac{b}{2}$$

$$b < 0 \Rightarrow x_v > 0$$

$$b = 0 \Rightarrow x_v = 0$$

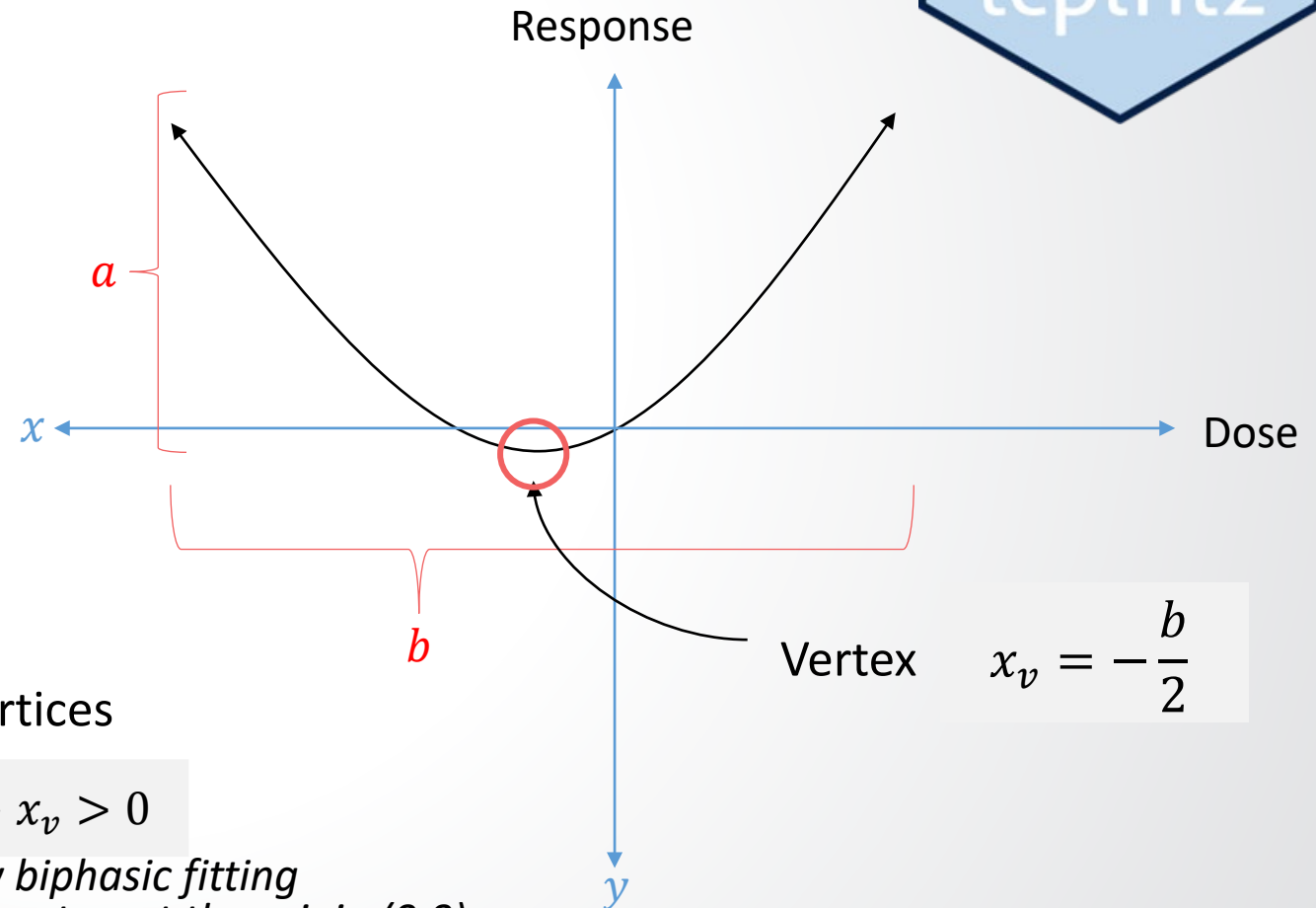
$$b > 0 \Rightarrow x_v < 0$$



- v4.2 includes the *tcplfit2* update to the polynomial 2 (poly2), i.e. quadratic, model parameterization. Model is now in line with BMDEpress

$$f(x) = a \left( \frac{x}{b} + \left( \frac{x}{b} \right)^2 \right) + \epsilon$$

| Parameter  | Interpretation            |
|------------|---------------------------|
| $a$        | Y-scalar (scale response) |
| $b$        | X-scalar (scale dose)     |
| $\epsilon$ | Error term                |



- Constraints  $b$  parameter limited the possible vertices

$$b > 0 \Rightarrow x_v < 0$$

$$b = 0 \Rightarrow x_v = 0 \quad b < 0 \Rightarrow x_v > 0$$

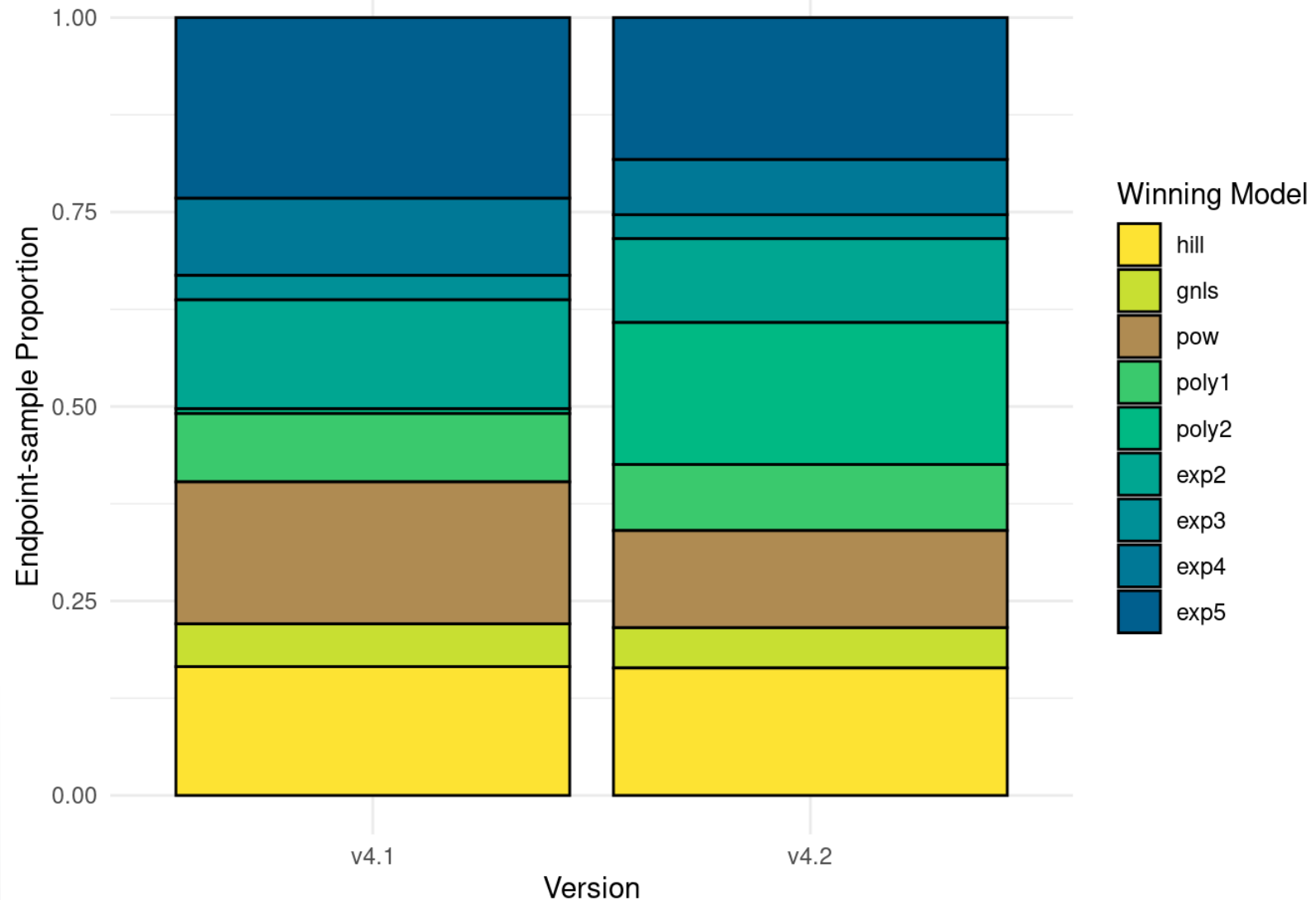
*Original  
(plot shown)*

*Additional options now allow biphasic fitting  
i.e. vertex can be any quadrant or at the origin (0,0)*



# Impacts of Poly2 Update: Winning Model Selection, Actives Only

- With biphasic poly2 fitting, the number of actives with poly2 winning model selected increased from 0.6% in v4.1 to 18.2358% in v4.2





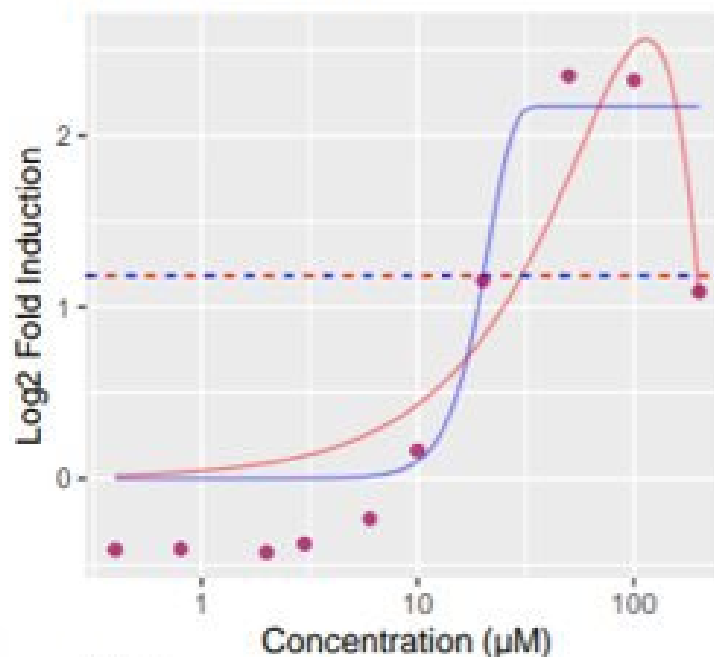
# Impacts of Poly2 Update

- Poly2 overfitting may be occurring when fewer data points, but in most cases, poly2 fits looks better than past fits
- However, additional reparameterization of poly2 model as used in hitcalling is needed in future
  - As shown, the top likelihood calculation assumes top response occurs at max conc, which is *not true for biphasic curves (top occurs at vertex)*
  - Problematic for borderline cases

DTXSID2047357 Surinabant  
SPID:TP0000379F02 AEID:60 AENM:APR\_HepG2\_p53Act\_72hr

A: M4ID:624270

B: M4ID:11881861



A: v4.1

B: v4.2 → Better fit, but flipped inactive

|   | Hitcall | BMD    | AC50   |
|---|---------|--------|--------|
| A | 0.998   | 16.203 | 19.529 |
| B | 0.780   | 14.448 | 32.659 |

Flags:

A(1): Average number of replicates per conc is less than 2

B(2): Inactive with multiple concs above baseline (3\*bmad);  
Average number of replicates per conc < 2



# Bisphenol A: in domain of aqueous cell-based screening?

Chemical domain of applicability

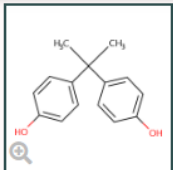
Many successfully screened chemicals in aqueous-based assays have been (but not limited to):

logP -1 to ~6.5 range;

MW 140-480;

log<sub>10</sub> Vapor Pressure < 2.

CompTox Chemicals Dashboard v2.3.0 Home Search Lists About Tools Submit Comments Search all data



## Bisphenol A

80-05-7 | DTXSID7020182

Searched by Approved Name.

Chemical Details

Executive Summary

Physchem Prop.

Env. Fate/Transport

Hazard Data

Safety > GHS Data

### Properties: Summary

Summary



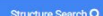
EXPORT











| Property        | Experimental average | Predicted average | Experimental median |
|-----------------|----------------------|-------------------|---------------------|
| Surface tension | -                    | 46.0 (1)          | -                   |

| Property                     | Value  |
|------------------------------|--|
| LogKow                       | 3.32 (lipophilic, likely crosses cell membrane without active transport) |
| Vapor Pressure               | 1.07 e-6 mmHg (not volatile)   |
| Average Mass                 | 228.291 g/mol  |
| Analytical QC of DMSO sample | pass   |

NIH National Center for Advancing Translational Sciences TOX21 SAMPLES Log In

Home QC T0 QC T4 QC Methods Tox21 Program Applications Resources Contact Us

Bisphenol A [SAMPLE\_NAME]   

| Structure   | TOX21 ID & Name  | QC Grade T0  | QC Grade T4  |
|---|--|--|--|
|  | Tox21_202992<br>Bisphenol A<br>  |  MW Confirmed, Purity > 90% |  MW Confirmed, Purity > 90% |
|  | Tox21_400088<br>Bisphenol A<br>  |  MW Confirmed, Purity > 90% |  MW Confirmed, Purity > 90% |

[tripod.nih.gov/tox/samples](https://tripod.nih.gov/tox/samples)



# Overview of endocrine models available

## Consensus QSARs (*in silico*)

CERAPP: Collaborative Estrogen Receptor Activity Prediction Project for agonist, antagonist, and binding prediction [Mansouri *et al.*, 2016, <http://dx.doi.org/10.1289/ehp.1510267>]

COMPARA: Collaborative Modeling Project for Androgen Receptor Activity for agonist, antagonist, and binding prediction [Mansouri *et al.*, 2020, <https://doi.org/10.1289/EHP5580>]

## ToxCast ER and AR pathway models (based on *in vitro* data for multiple assays)

Original models using 18 and 12 assays, respectively, have results on the CompTox Chemicals Dashboard

Confidence score and examination of selectivity can be important for interpreting the overall results

Work is currently in progress to create a set of assays to inform a prospective model with fewer assays

## ToxCast HT-H295R statistical model (based on *in vitro* data for multiple hormones in H295R cells)

Similar to OECD Test Guideline 456

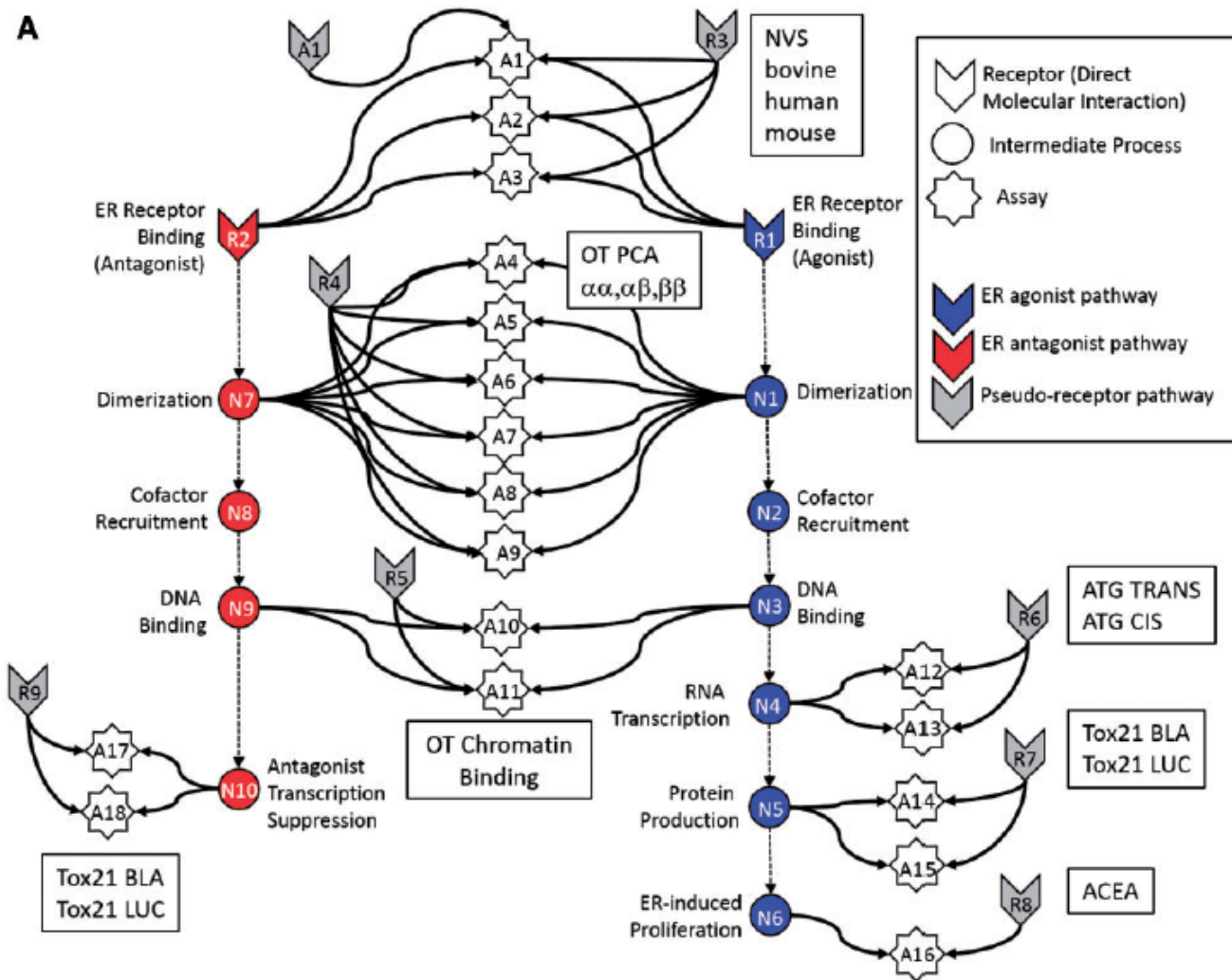
Maximum mean Mahalanobis distance, which compresses the 11-hormone responses into a single value to determine if steroidogenesis has been perturbed in the H295R system

Log2-fold changes by hormone are available in the publication(s)



## ER and AR are unique targets in their assay redundancy, facilitating this approach

- Developed multiple high-throughput screening assays
- Use multiple assays per pathway
  - Different technologies
  - Different points in pathway
- No assay is perfect
  - Assay Interference
  - Noise
- Use a systems biology model to integrate assays
  - Model creates a composite dose-response curve for each chemical to summarize results from all assays
  - Includes penalization for cytotoxicity



- The current model in the CompTox Chemicals Dashboard (v2.3.0) is an update of the 2015 published model but still includes all 18 assays for agonist mode.
- This model has been accepted as an alternative for the ER binding, ER-TA, and uterotrophic assays in the EDSP Tier 1 (<https://www.federalregister.gov/documents/2015/06/19/2015-15182/use-of-high-throughput-assays-and-computational-tools-endocrine-disruptor-screening-program-notice>).
- Only 4 assays that cover key “receptors” or events in the activation of ER can achieve similar performance as the full model ([10.1016/j.yrtph.2017.09.022](https://doi.org/10.1016/j.yrtph.2017.09.022)).



# Regulatory use of endocrine bioactivity models

2015 US Federal Register notice:  
**ToxCast ER pathway model** is a suitable alternative to 3 ER assays in EDSP Tier 1

Federal Register / Vol. 80, No. 118 / Friday, June 19, 2015 / Notices

response  
sclose  
ed by a claim  
the extent  
ordance with,  
section 14 and  
annual public  
ing burden for  
tion is  
hours per  
ed in 5 CFR

lable in the  
elated  
iled  
ion activities  
that is only

Entities:  
ed by this ICR  
fature,  
al substances,  
r of potential

On occasion,  
number of  
ident: 1,  
burden hours:

costs: \$2,388,  
d burden cost  
d cost of \$0 for  
intenance and

the Estimates

16 hours in the

and approval pursuant to 5 CFR 1320.12. EPA will issue another Federal Register document pursuant to 5 CFR 1320.5(a)(1)(iv) to announce the submission of the ICR to OMB and the opportunity to submit additional comments to OMB. If you have any questions about this ICR or the approval process, please contact the technical person listed under **FOR FURTHER INFORMATION CONTACT**.

Authority: 44 U.S.C. 3501 et seq.

Dated: June 10, 2015.

James Jones,  
Assistant Administrator, Office of Chemical Safety and Pollution Prevention.

[FR Doc. 2015-14946 Filed 6-18-15; 8:45 am]

BILLING CODE 6560-50-P

## ENVIRONMENTAL PROTECTION AGENCY

[EPA-HQ-OPPT-2015-0305; FRL-9928-69]

**Use of High Throughput Assays and Computational Tools; Endocrine Disruptor Screening Program; Notice of Availability and Opportunity for Comment**

AGENCY: Environmental Protection Agency (EPA).

ACTION: Notice.

**SUMMARY:** This document describes how EPA is planning to incorporate an alternative scientific approach to screen chemicals for their ability to interact with the endocrine system. This will

efficient screening using alternative test methods to some assays in the Tier 1 battery to protect human environment.

**DATES:** Comments must be submitted on or before August 18, 2015. **ADDRESSES:** Submit your comments to the docket identified by docket id/number EPA-HQ-OPP-2015-0305, one of the following methods:

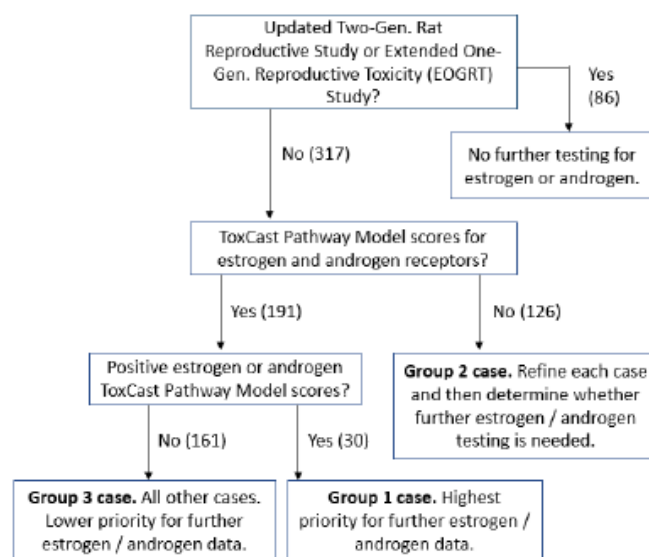
- **Federal eRulemaking:** [www.regulations.gov](http://www.regulations.gov). Follow the instructions for submitting comments.
- **Mail:** Document C (7407M), Office of Pollution Prevention and Toxics (OPPT), Environmental Protection Agency, 1200 Avenue of the Americas, Washington, DC 20460.
- **Hand Delivery:** To arrangements for hand delivery of boxed information, follow the instructions at [www.epa.gov/dockets](http://www.epa.gov/dockets).

Additional instructions for submitting comments or visiting along with more information are available at [www.epa.gov/dockets](http://www.epa.gov/dockets).

**FOR FURTHER INFORMATION:** Technical information contact: Mr. Robbins, Office of Science and Policy (OSCP), Office of Chemical Safety and Pollution Prevention, Environmental Protection Agency, Washington, DC 20460.

2023 EPA-OCSP applied strategy for using **ToxCast ER and AR pathway models** to fill data gaps

Figure 1. Framework for prioritizing the 403 conventional pesticide cases currently in registration review for which an FFDCA section 406(p)(6) determination is needed.



2018 ECHA/EFSA guidance on using ToxCast ER pathway model and other bioactivity data

**ECHA** GUIDANCE **efsa** European Food Safety Authority

ADOPTED (ECHA): 5 June 2018  
ADOPTED (EFSA): 5 June 2018  
doi: 10.2903/j.efsa.2018.5311

**Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1831/2003**

European Chemical Agency (ECHA) and European Food Safety Authority (EFSA) technical support

Niklas Andersson, Maria Arena, Dorothea Kienzl, Peter Lepper, Alfonso Lopez, Francesca Pellizzato, Jose Tarazona

ADOPTED: 22 March 2023  
doi: 10.2903/j.efsa.2023.7968

**Abstract**  
This Guidance describes how to perform the identification of endocrine disruptors following the scientific criteria which are outlined in the Commission Regulation (EU) 2018/606 respectively.

© 2018 European Chemicals Agency and © 2023 European Food Safety Authority

**Keywords:** biocidal product, plant protection product, endocrine disruptor, identification

**Requestor:** European Commission

**Question numbers:** EFSA-Q-2016-00825, ECHA-18-G-01-EN

**Correspondence:** For biological products: [biocides@echa.europa.eu](mailto:biocides@echa.europa.eu)  
For plant protection products: [pesticides.peerreview@efsa.europa.eu](mailto:pesticides.peerreview@efsa.europa.eu)

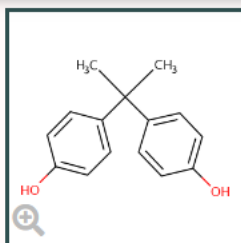
2023 Examples of use by EFSA



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10116401/pdf/EFSA-21-e07968.pdf>

**Statement concerning the testing strategy and timelines proposed by the applicant for the assessment of the endocrine disruption properties of acibenzolar-S-methyl in the context of the review of the approval of the active substance**

European Food Safety Authority (EFSA)



## Bisphenol A

80-05-7 | DTXSID7020182

Searched by Approved Name.

- 2 kinds of models are represented here: *in silico* consensus (Q)SARs and bioactivity-based ToxCast models
- For ToxCast models, >0.1 is positive; 0.001-0.1 is equivocal

### Bioactivity - ToxCast: Models

EXPORT

#### ToxCast Model Predictions

| Model ↓↑   | Receptor ↓↑ | Agonist ↓↑ | Antagonist ↓↑ | Binding ↓↑ |
|--|-------------|------------|---------------|------------|
| <a href="#">COMPARA (Consensus)</a>                    | Androgen    | 0.00       | 1.00          | 1          |
| <a href="#">ToxCast Pathway Model (AUC)</a>            | Estrogen    | 0.450      | 0.00          | -          |
| <a href="#">CERAPP Potency Level (From Literature)</a> | Estrogen    | Weak       | Strong        | Weak       |
| <a href="#">CERAPP Potency Level (Consensus)</a>       | Estrogen    | 1.00       | 1.00          | 1          |
| <a href="#">ToxCast Pathway Model (AUC)</a>            | Androgen    | 0.00       | 0.345         | -          |

CERAPP (literature and model), and ToxCast ER pathway model, suggest estrogen receptor agonism  
COMPARA and ToxCast AR pathway model suggest androgen receptor antagonism

Chemical Details

Executive Summary

Physchem Prop.

Env. Fate/Transport

Hazard Data

Safety > GHS Data

ADME > IVIVE

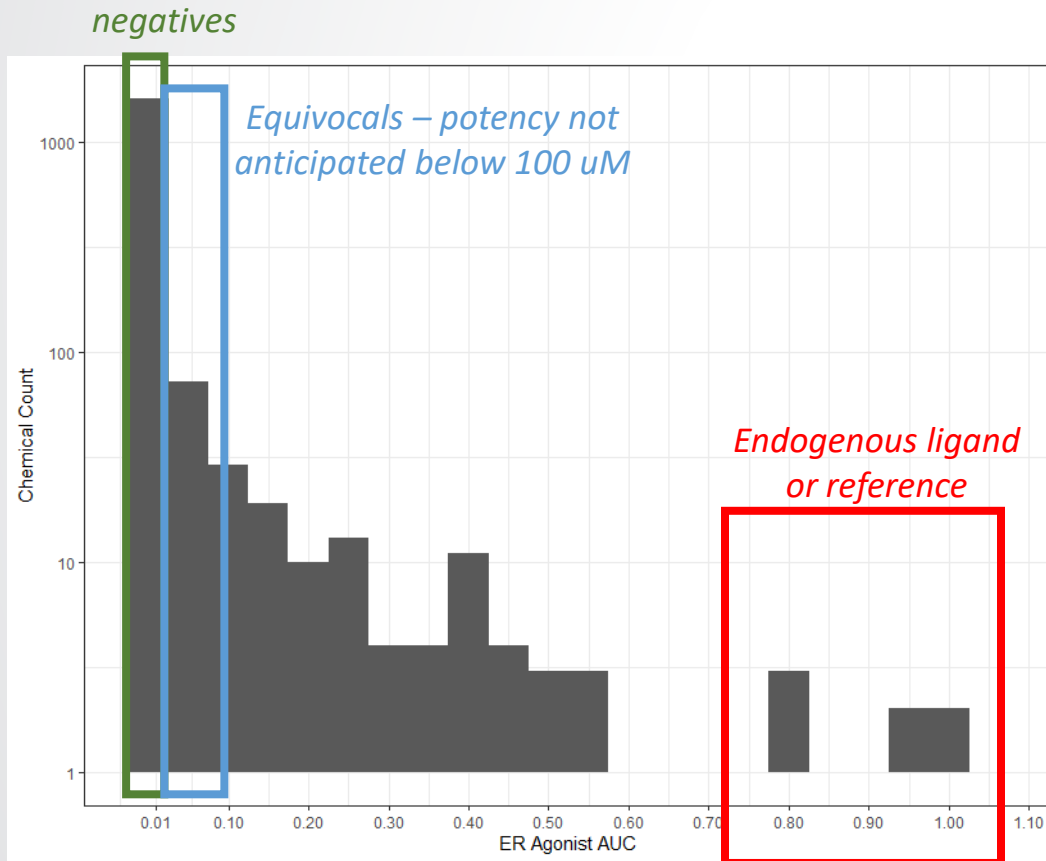
Exposure

Bioactivity

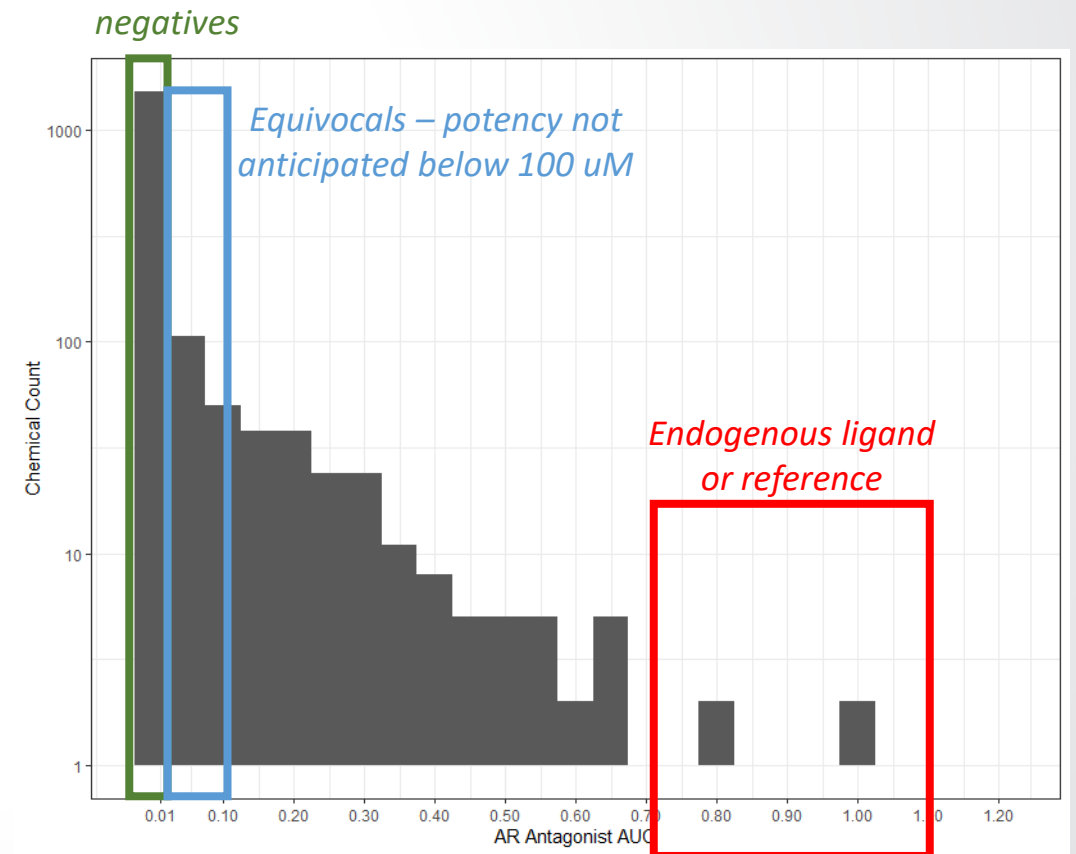


# Interpreting and using ToxCast pathway model scores: relative activity

Distribution of ToxCast ER Pathway Agonist Scores

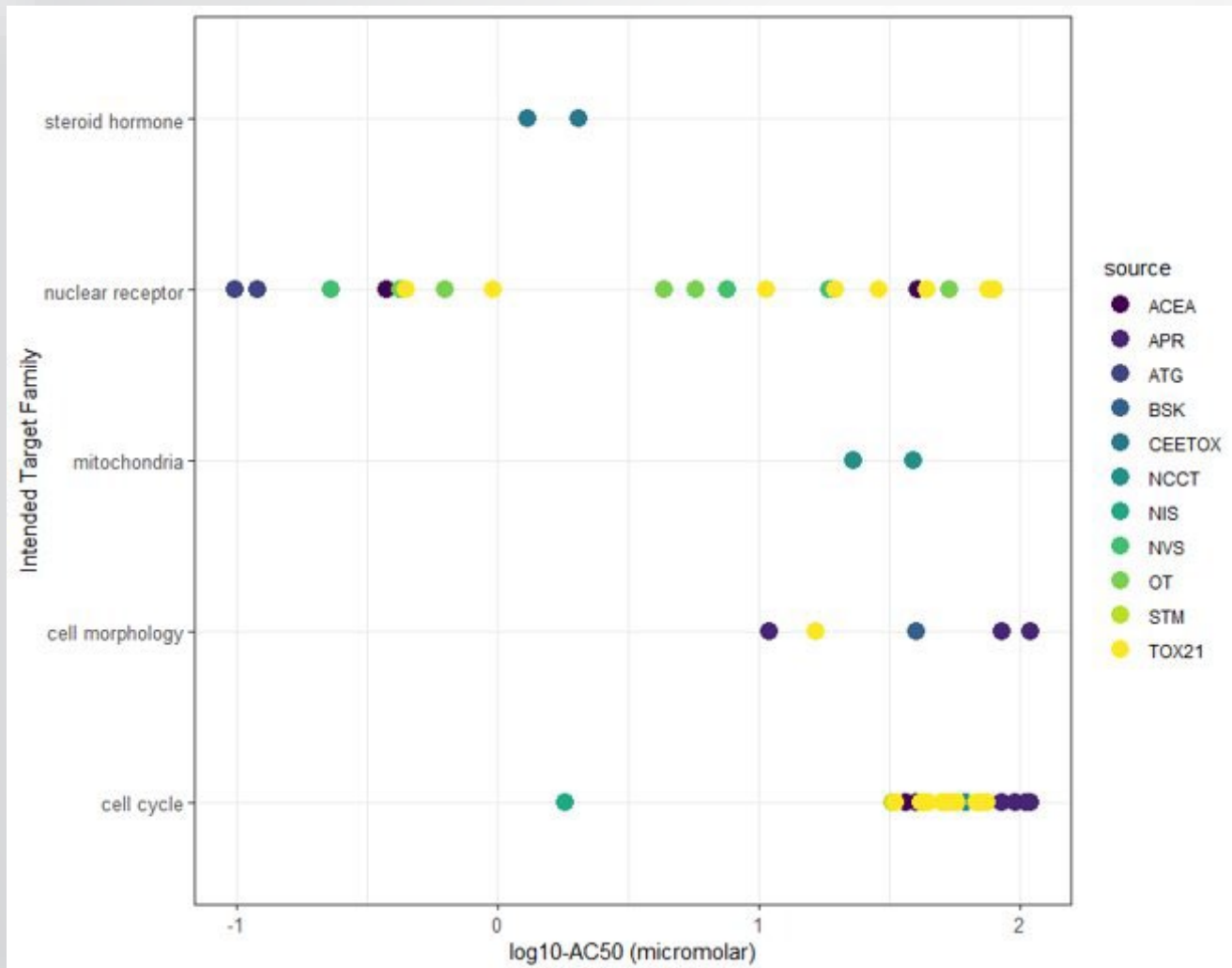


Distribution of ToxCast AR Pathway Antagonist Scores





## Some of Bisphenol A's most potent *in vitro* activity is for the estrogen receptor



Exporting data from CCD, and then plotting by intended target family can be a helpful way of understanding the overall activity of the chemical



# Examining ToxCast data for ESR1 using an export from CCD

| NAME                           | ASSAY_COMPONENT_NAME           | ASSAY_LISTS | GENE_SYMBOL | HIT_CALL | AC50  | ACC  |
|--------------------------------|--------------------------------|-------------|-------------|----------|-------|------|
| ATG_ERa_TRANS                  | ATG_ERa_TRANS                  | EDSP ER     | ESR1        | Active   | 0.09  | 0.03 |
| ATG_ERE_CIS                    | ATG_ERE_CIS                    | EDSP ER     | ESR1        | Active   | 0.1   | 0.05 |
| NVS_NR_mERa                    | NVS_NR_mERa                    | EDSP ER     | Esr1        | Active   | 0.14  | 0.02 |
| NVS_NR_hER                     | NVS_NR_hER                     | EDSP ER     | ESR1        | Active   | 0.23  | 0.16 |
| OT_ER_ERbERb_1440              | OT_ER_ERbERb_1440              | EDSP ER     | ESR2        | Active   | 0.35  | 0.12 |
| OT_ERa_GFPERaERE_0120          | OT_ERa_GFPERaERE_0120          | EDSP ER     | ESR1        | Active   | 0.37  | 0.3  |
| OT_ER_ERbERb_0480              | OT_ER_ERbERb_0480              | EDSP ER     | ESR2        | Active   | 0.37  | 0.1  |
| ACEA_ER_80hr                   | ACEA_ER_80hr                   | EDSP ER     | ESR1        | Active   | 0.37  | 0.2  |
| NVS_NR_bER                     | NVS_NR_bER                     | EDSP ER     | ESR1        | Active   | 0.42  | 0.19 |
| TOX21_ERa_LUC_VM7_Agonist      | TOX21_ERa_LUC_VM7_Agonist      | EDSP ER     | ESR1        | Active   | 0.43  | 0.12 |
| OT_ER_ERaERb_0480              | OT_ER_ERaERb_0480              | EDSP ER     | ESR1   ESR2 | Active   | 0.5   | 0.28 |
| OT_ERa_GFPERaERE_0480          | OT_ERa_GFPERaERE_0480          | EDSP ER     | ESR1        | Active   | 0.65  | 0.38 |
| TOX21_ERa_BLA_Agonist_ratio    | TOX21_ERa_BLA_Agonist_ratio    | EDSP ER     | ESR1        | Active   | 0.96  | 1.37 |
| OT_ER_ERaERb_1440              | OT_ER_ERaERb_1440              | EDSP ER     | ESR1   ESR2 | Active   | 1.92  | 0.09 |
| OT_ER_ERaERa_0480              | OT_ER_ERaERa_0480              | EDSP ER     | ESR1        | Active   | 4.03  | 0.68 |
| OT_ER_ERaERa_1440              | OT_ER_ERaERa_1440              | EDSP ER     | ESR1        | Active   | 4.31  | 1.05 |
| TOX21_ERa_BLA_Antagonist_ratio | TOX21_ERa_BLA_Antagonist_ratio | EDSP ER     | ESR1        | Active   | 31.34 | 9.16 |

Depending on your use case, 0.1 uM appears to be a threshold AC50 value for BPA in estrogen receptor (ESR1) related assays

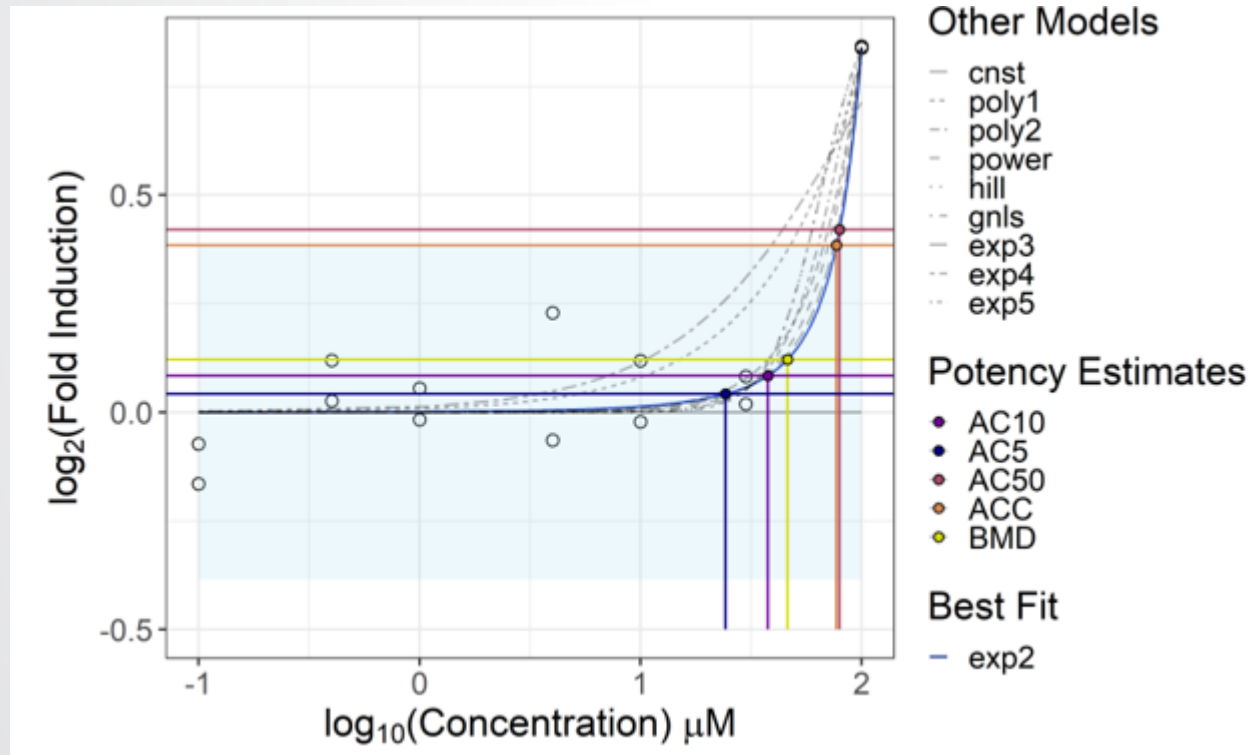
It is important to cite the data and explain the derivation of your point-of-departure

# Calculating a bioactivity-based point-of-departure (POD) and bioactivity:exposure ratio (BER)

First, we need to convert the bioactive concentrations from micromolar to external dose estimates, termed “administered equivalent doses” in mg/kg/day (at least, for oral exposure)

# Defining POD and BER

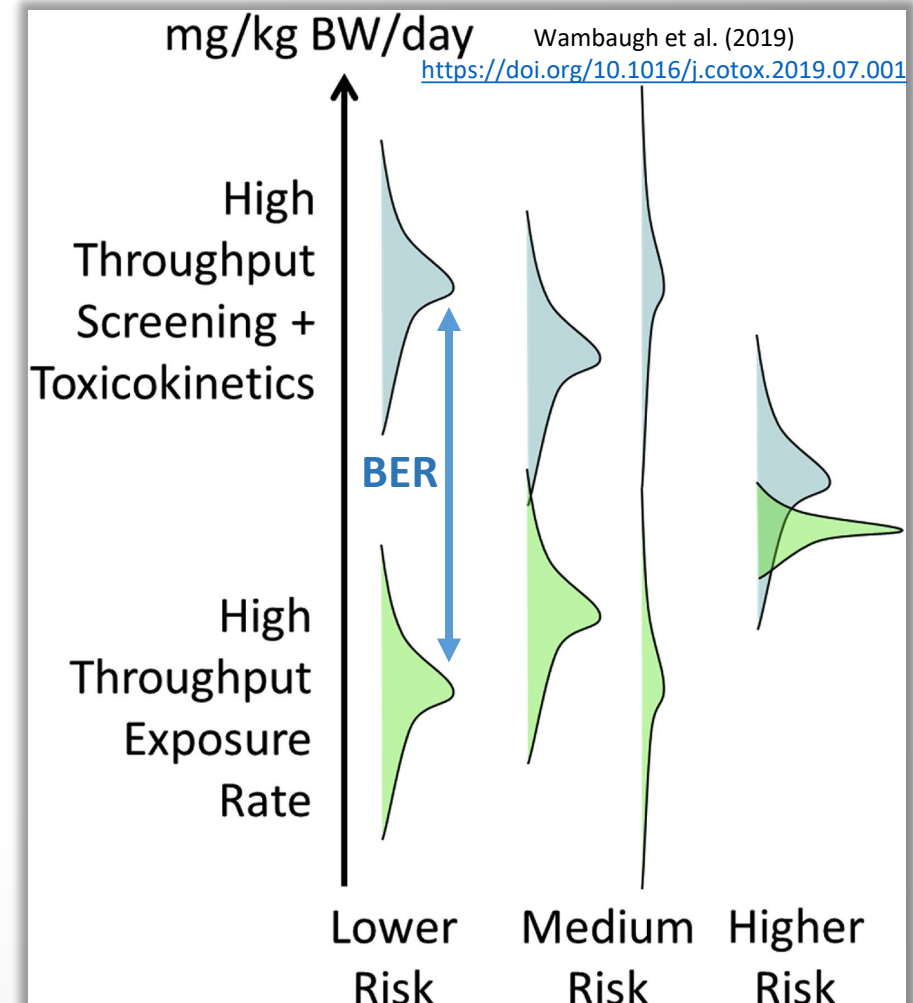
A point-of-departure describes a point on a concentration (or dose) response curve where the activity moves away from the background and can be a first basis for setting health-protective limits



Feshuk et al (2023)

<https://doi.org/10.3389/ftox.2023.1275980>

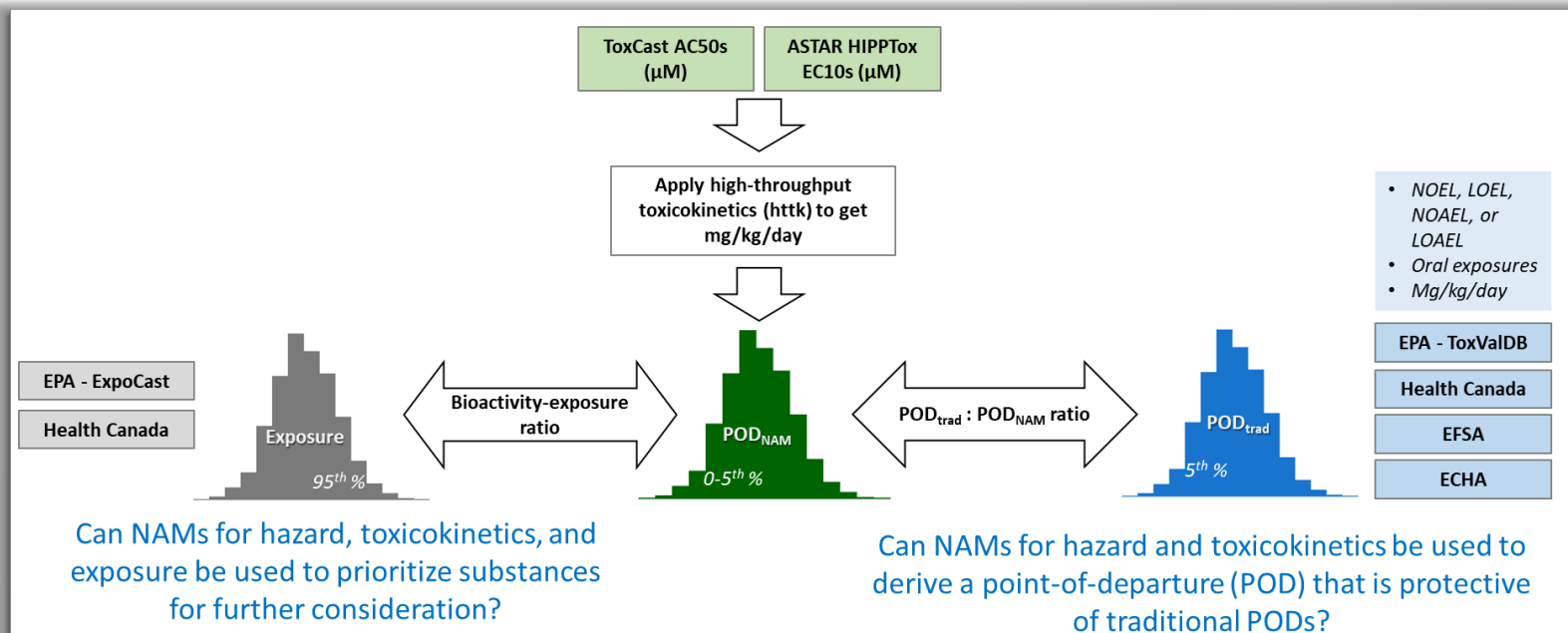
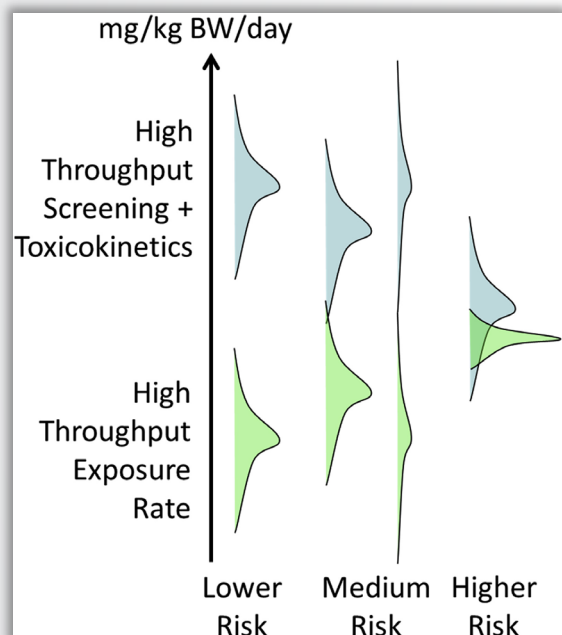
Bioactivity:exposure ratio: quantitative difference between bioactive dose and possible exposure dose





# Bioactivity:exposure ratio is similar to a margin of exposure

*In vitro* bioactive concentrations would be useful to compare to predicted exposures in humans, i.e. for derivation of a bioactivity:exposure ratio (BER).



Paul Friedman et al. (2020)  
<https://doi.org/10.1093/toxsci/kfz201>



# In vitro to in vivo extrapolation (IVIVE) of dose can be a simple approximation

$$[\text{Conc}]_{\text{ss}} = \frac{\text{Dose Rate} * \text{Body Weight}}{\text{CL}_{\text{WholeBody}}}$$

Steady State Blood  
Concentration

CL: clearance

Rowland *et al.*, 1973

Wilkinson and Shand, 1975

Gillette, 1980

Wilkinson, 1987



# In vitro to in vivo extrapolation (IVIVE) of dose can be a simple approximation

## Assumptions:

100% absorption

Linear kinetics

No extrahepatic metabolism

***Whole Body Clearance (CL) = Considering renal and hepatic clearance ( $CL_R$ ,  $CL_H$ ) – adjusted for blood binding ( $F_{ub}$ )***

$$\begin{aligned} \text{Steady State Blood Concentration } [Conc]_{ss} &= \frac{\text{Dose Rate} * \text{Body Weight}}{CL_{\text{WholeBody}}} \\ &= \frac{\text{Dose Rate} * \text{Body Weight}}{CL_R + CL_H} \\ CL_R &= F_{ub} * GFR \\ CL_H &= \frac{F_{ub} * Q_L * CL_{int}}{Q_L + F_{ub} * CL_{int}} \end{aligned}$$

GFR = glomerular filtration rate

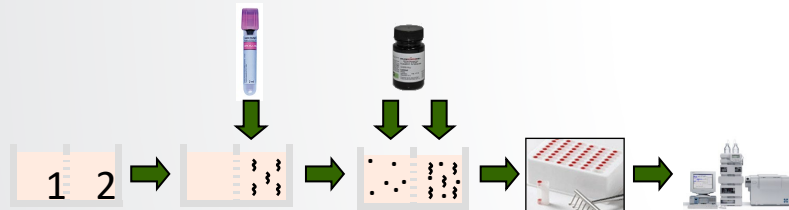
$F_{ub}$  = fraction unbound in blood

$Q_L$  = hepatic blood flow

$CL_{int}$  = intrinsic clearance

## *in vitro* toxicokinetic data

Hepatic clearance from suspended hepatocytes

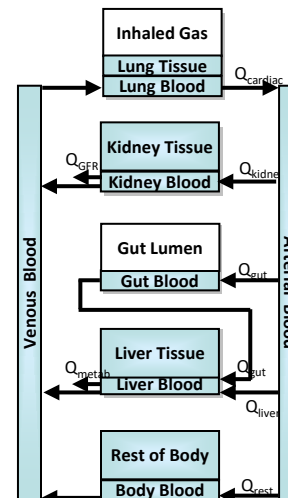


Plasma protein binding



= *httk*

## Generic toxicokinetic models

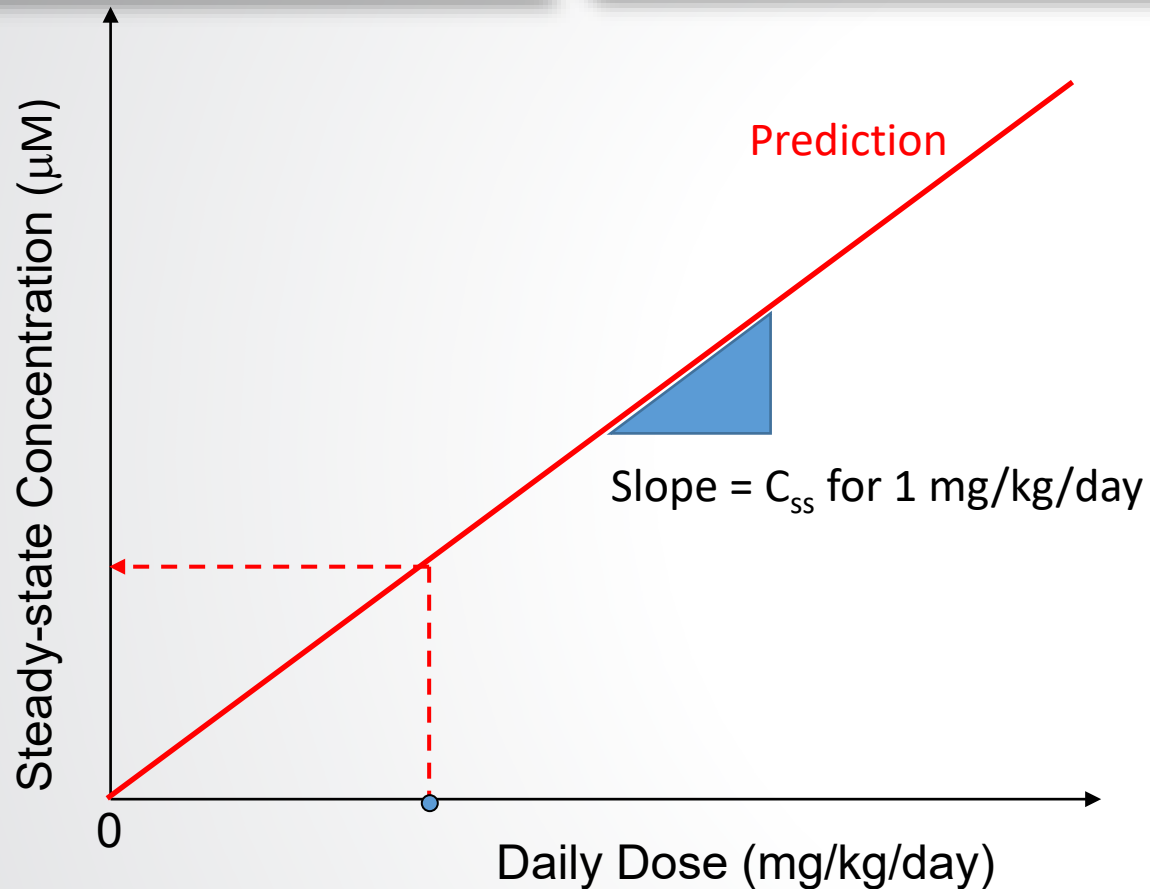


## Some high-level assumptions commonly employed:

- (1) bioactive nominal *in vitro* assay concentration  $\sim$  *in vivo* plasma concentration that would correspond to a similar effect;
- (2) external exposures (in mg/kg/day units) that may have resulted in that plasma concentration can be constructed using estimates of species-specific physiology and Phase I and Phase II enzyme-driven hepatic clearance; and,
- (3) Often, we expect that plasma concentration can be approximated by steady-state kinetics (unless we have enough information to use other dose metrics).

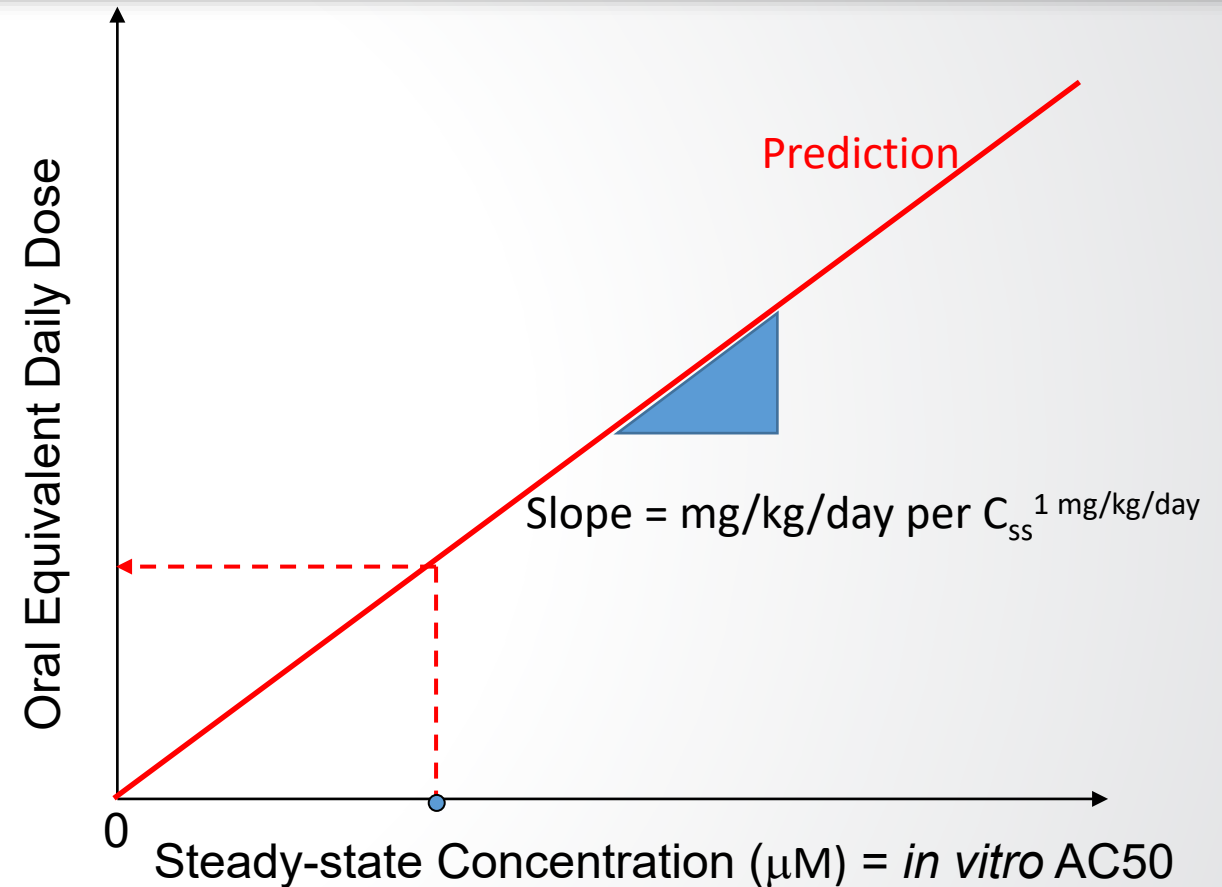


## Steady state in vitro-in vivo extrapolation assumption: blood-to-tissue partitioning $\approx$ cells-to-medium partitioning



$$C_{ss} = \frac{\text{oral dose rate}}{(GFR * F_{ub}) + \left( Q_l * F_{ub} * \frac{Cl_{int}}{Q_l + F_{ub} * Cl_{int}} \right)}$$

Wetmore *et al.* (2012)



- Swap the axes (this is the "reverse" part of reverse dosimetry)
- Can divide bioactive concentration by  $C_{ss}$  for for a 1  $\text{mg/kg/day}$  dose to get oral equivalent dose

Slide from John Wambaugh



## Simple calculation of AED based on 3 compartment steady state model

AED values in mg/kg/day units can be calculated from bioactivity using the following equation:

$$Eq: AED_{50} \left( \frac{\frac{mg}{kg}}{day} \right) = AC_{50}(\mu M) * \frac{\frac{1 \frac{mg}{kg}}{day}}{C_{SS50}}$$

Where the C<sub>ss</sub> (steady-state concentration) values for the median individual based on Monte Carlo simulation of species-specific physiological parameters (C<sub>ss50</sub>) (Pearce et al. 2017) were generated using the 3-compartment steady state model



# A simple approach for using the CompTox Chemicals Dashboard to estimate a POD

- $AC50 \text{ or LEC (micromolar)} * (1 \text{ mg/kg/day/Css (micromolar)}) = \text{AED prediction}$
- httk package optionally implements multiple models that can have increasing complexity based on data available (e.g., using pbtk model or including interindividual toxicokinetic variability).

$$\begin{array}{|c|c|c|c|} \hline 2.35 \text{ mg} & \text{g} & \text{mol} & 1\text{e6 } \mu\text{mol} \\ \hline \text{L} & 1000 \text{ mg} & 228.291 \text{ g} & \text{mol} \\ \hline \end{array} = 10.2938793 \text{ } \mu\text{mol/L} = \text{ } \mu\text{M}$$

Bioactive concentration

$$\begin{array}{|c|c|} \hline 0.1 \text{ } \mu\text{M} & 1 \text{ mg/kg/day} \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline & 10.2938793 \text{ } \mu\text{M} \\ \hline \end{array} = \mathbf{0.0097 \text{ mg/kg/day} = \text{AED}_{95}}$$

**ADME - IVIVE** ⓘ

Search ADME IVIVE

EXPORT

**IVIVE**

| Label ↓↑                          | Species ↓↑ | Measured ↓↑ | Predicted ↓↑ | Units ↓↑ | Model ↓↑         | Percentile ↓↑ | Reference ↓↑  | Data Source Species ↓↑ |
|-----------------------------------|------------|-------------|--------------|----------|------------------|---------------|---------------|------------------------|
|                                   | (1) Human  |             |              |          | (3) 1compartment | (2) NA,95%    |               |                        |
| Fraction Unbound in Plasma        | Human      | 0.04        | NA           |          | NA               | NA            | Wambaugh 2019 | Human                  |
| Volume of Distribution            | Human      | NA          | 6.337        | L/kg     | 1compartment     | NA            | NA            | Human                  |
| PK Half Life                      | Human      | NA          | 28.28        | hours    | 1compartment     | NA            | NA            | Human                  |
| Steady-State Plasma Concentration | Human      | NA          | 2.35         | mg/L     | 3compartmentss   | 95%           | NA            | Human                  |



# Calculate the bioactivity:exposure ratio (BER)

$$\text{BER} = \text{bioactive dose/exposure} = 0.0097/0.0204 = 0.476$$

$$\text{or } \log_{10}(\text{AED}) - \log_{10}(\text{exposure}) = -2.01 - -1.69 = -0.322$$

where both bioactive dose and exposure are in the same units

Chemical Details

Executive Summary

Physchem Prop.

Env. Fate/Transport

Hazard Data

Safety > GHS Data

ADME > IVIVE

Exposure

Bioactivity

GenRA

ACToR

Literature

Links

Exposure - Exposure Predictions (mg/kg-bw/day)

Search Demographics Predictions Data

EXPORT

Demographics Predictions Data

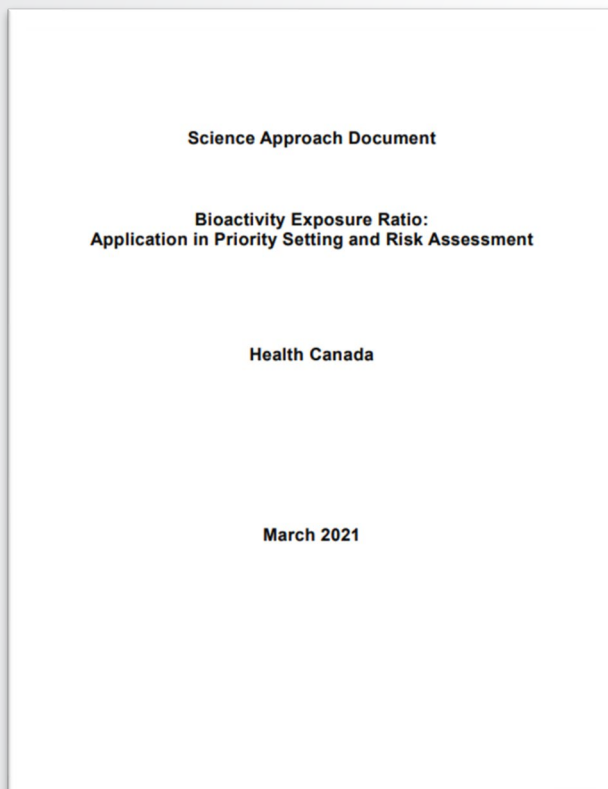
| Demographic        | Predictor                            | Median  | Upper 95%ile | Units     |
|--------------------|--------------------------------------|---------|--------------|-----------|
|                    | (2) SEEM2 Heuristic, SEEM3 Consensus |         |              |           |
| Age 20-65          | SEEM2 Heuristic                      | 5.68e-5 | 1.15e-2      | mg/kg/day |
| Age 66+            | SEEM2 Heuristic                      | 6.61e-5 | 1.95e-2      | mg/kg/day |
| BMI <= 30          | SEEM2 Heuristic                      | 6.25e-5 | 1.36e-2      | mg/kg/day |
| BMI > 30           | SEEM2 Heuristic                      | 7.07e-5 | 1.86e-2      | mg/kg/day |
| Females            | SEEM2 Heuristic                      | 1.24e-5 | 2.90e-3      | mg/kg/day |
| Males              | SEEM2 Heuristic                      | 3.87e-5 | 6.31e-3      | mg/kg/day |
| Repro. Age Females | SEEM2 Heuristic                      | 1.36e-5 | 4.18e-3      | mg/kg/day |
| Total              | SEEM3 Consensus                      | 5.50e-5 | 2.04e-2      | mg/kg/day |

EXPORT



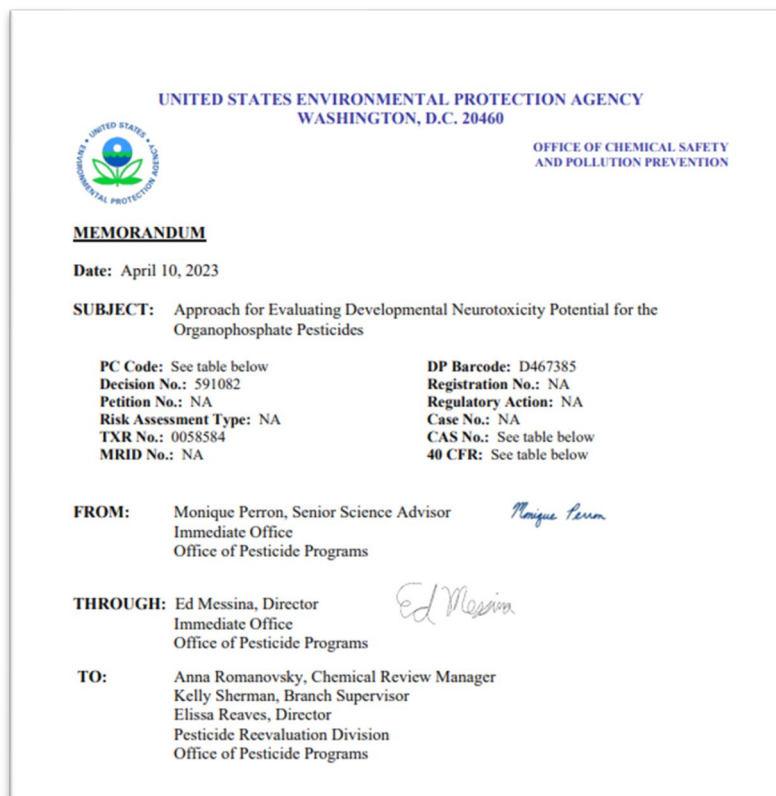
# Examples of BER in the regulatory toxicology

In managing risk evaluations of large inventories of chemicals, such as Canada's Domestic Substances List



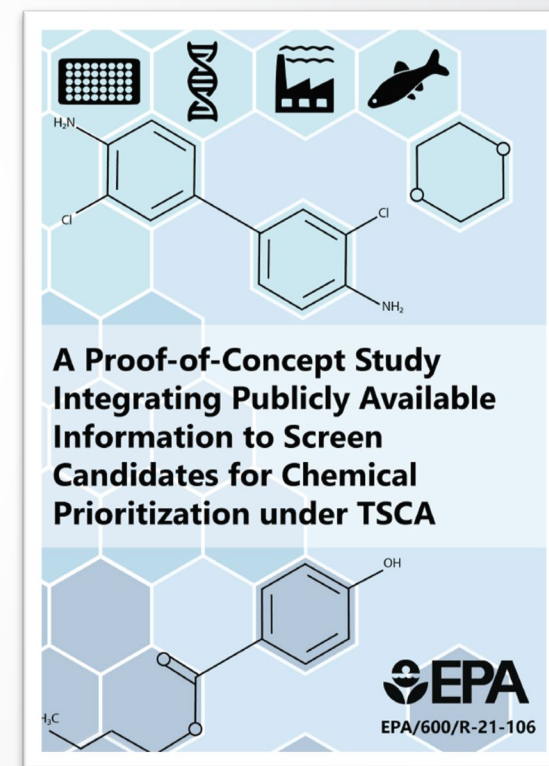
<https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/science-approach-document-bioactivity-exposure-ratio-application-priority-setting-risk-assessment.html>

In addressing data gaps within a weight-of-evidence for risk of developmental neurotoxicity of organophosphate insecticides



<https://www.regulations.gov/document/EPA-HQ-OPP-2008-0915-0056>

In proof-of-concept work to identify existing chemicals for further evaluation under the Toxic Substances Control Act



<https://www.epa.gov/sciencematters/proof-of-concept-case-study-integrating-publicly-available-information-screen>



# Health Canada Scientific Approach Document used BER for prioritization with familiar methods

## Science Approach Document

Bioactivity Exposure Ratio:  
Application in Priority Setting and Risk Assessment

Health Canada

March 2021

