Quantitative Prediction of Continuous Toxicity Values using Chemical Structure Information

Jessica Wignall

ICF International – Health Sciences Practice

Fairfax, VA

Jessica.Wignall@icfi.com

March 29, 2016

Conflict of Interest Statement

Acknowledgments

- UNC Chapel Hill: Ivan Rusyn (now at Texas A&M), Alexander Tropsha, Eugene Muratov, Denis Fourches, and Jessica Wignall (now at ICF)

- EPA/NCEA: Kate Guyton (now at IARC), Weihsueh Chiu (now at Texas A&M), Vincent Cogliano

- EPA/NCCT: Matthew Martin

- California EPA: Lauren Zeise

- UC San Francisco: Tracey Woodruff

- NC State: David Reif

Overview

- While TTC assigns chemical structures to classes, there is a need for quantitative predictions of toxicity values in certain decision-making contexts

- Quantitative structure activity relationship (QSAR) methods can be used to generate quantitative predictions of doses at which a chemical is likely to cause harm to humans

  - Uncertainty inherent in process, but that uncertainty can be quantified

- More and better *in vivo* dose-response data is being extracted and can be used to improve predictions

Outline

- Collect and Curate Datasets
- Build QSAR Models
- Apply QSAR Models
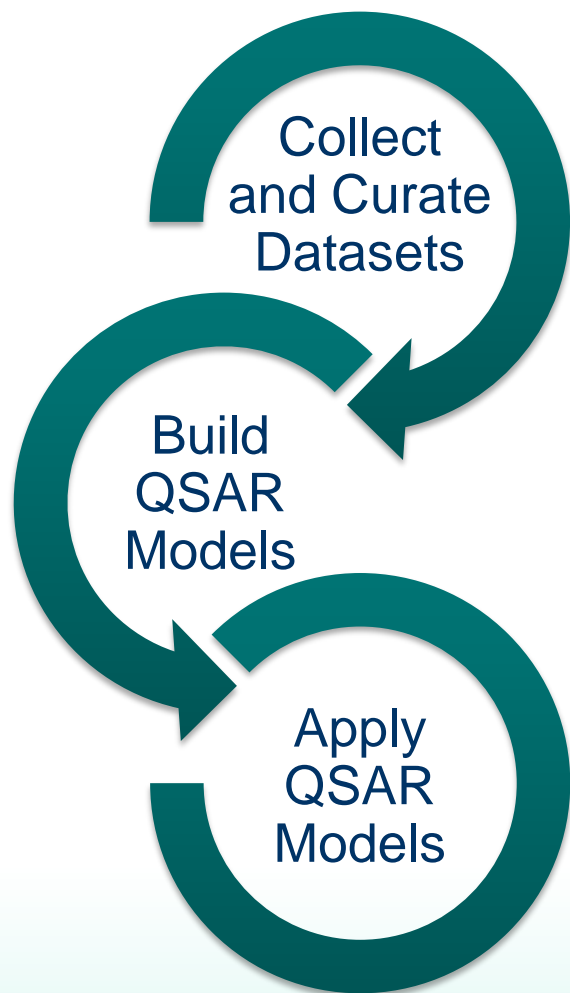
- Used in Decision Making
  - Points of Departure
    - Benchmark Dose Values (Standardized)
  - Other Toxicity Values

- Consider Chemical Space
- Evaluate Performance
- Quantify Uncertainty

- Addressing "no data–no hazard"
- ToxValue.org

Points of Departures Can be Used in Decision Making

Quantitative Dose-Response Assessment

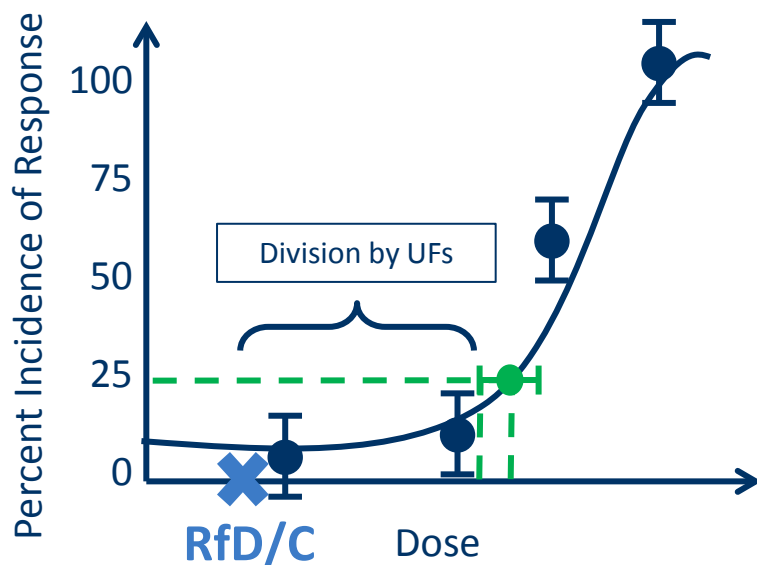Percent Incidence of Response

BMD
BMDL
LOAEL
NOAEL
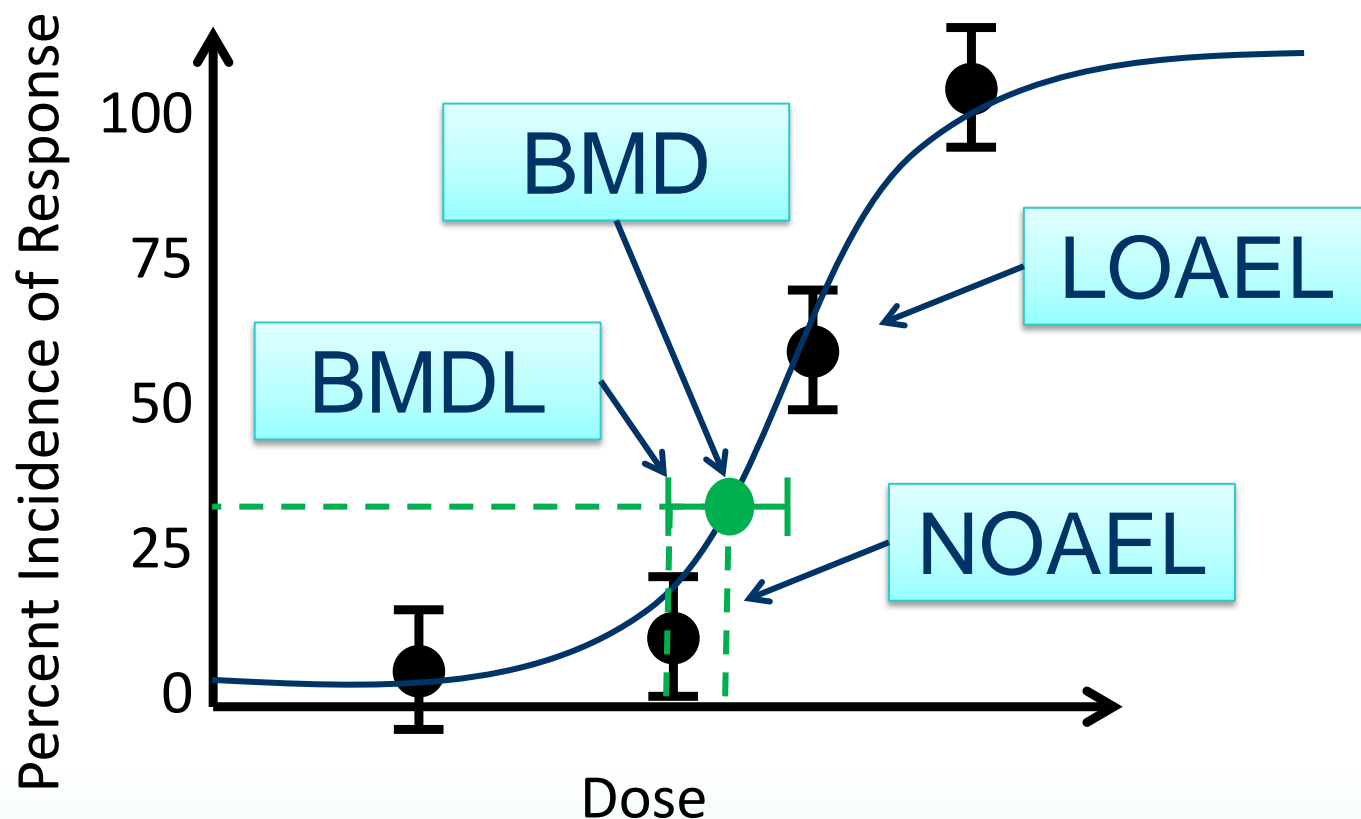
Dose

NOAEL = No observed adverse effect level
LOAEL = Lowest observed adverse effect level
BMD = Benchmark dose
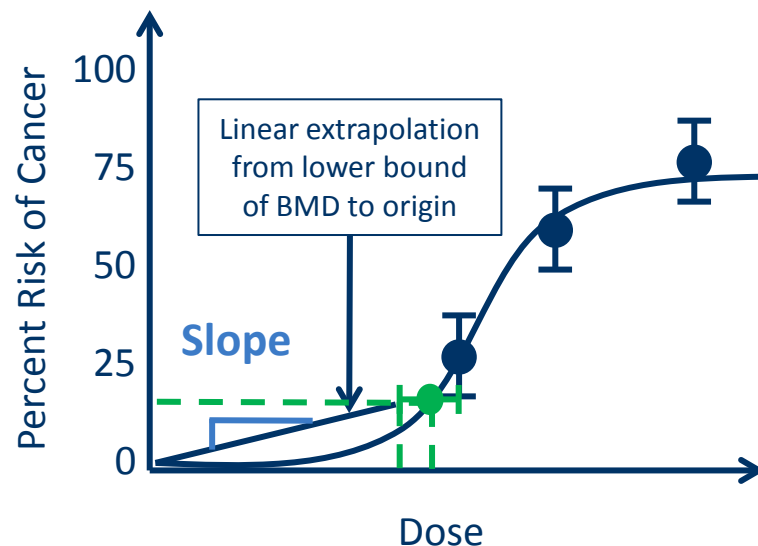BMDL = Benchmark dose lower confidence limit
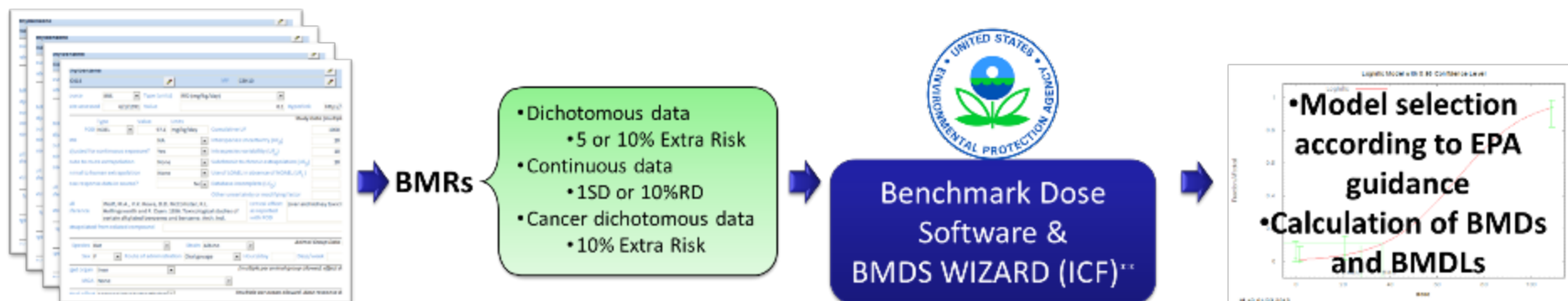
Benchmark Dose: A Data-Driven POD

But there are limitations to BMDs:
- **Time-intensive**
- **Complex**
- **Not all data amenable to modeling**

Standardized Calculation of BMDs and BMDLs for a Large Number of Chemicals

Collected[*] **880** dose-response datasets for

352 unique chemicals with Toxicity Value(s) (*e.g.*, RfD, OSF)

- **~75% of collected datasets <u>can be modeled</u> with BMDS**
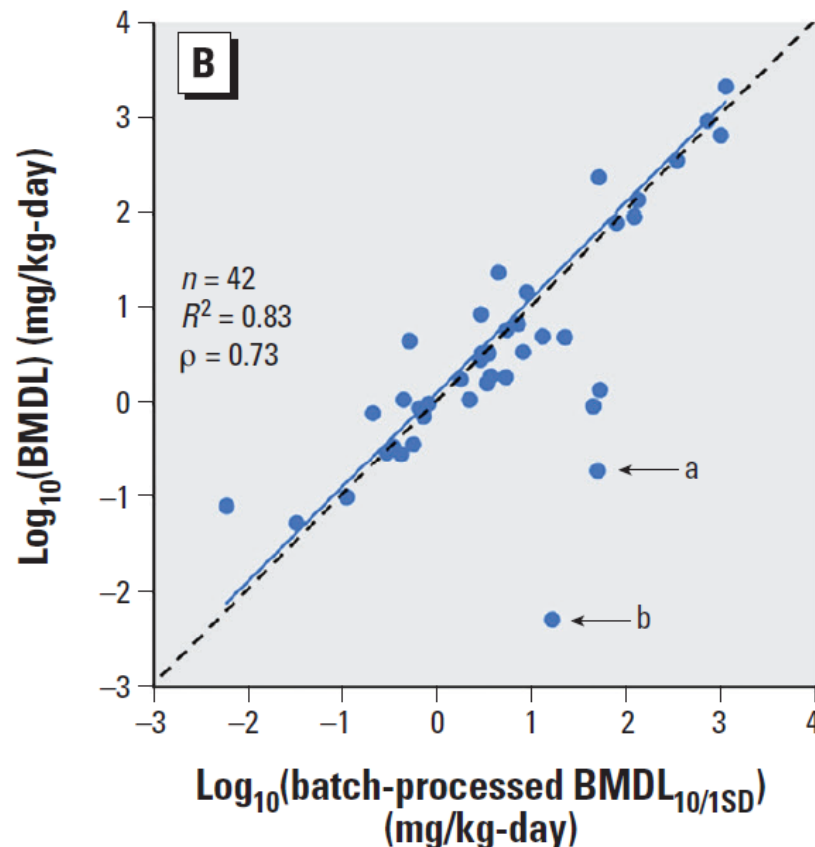- **Batch-calculated BMD/Ls available for over 300 chemicals**

*Under contract with SRC, Inc.
**Available for download from: http://www.icfi.com/insights/products-and-tools/bmds-wizard

See Wignall et al., 2014

Batch-Calculated BMDs and BMDLs are Correlated with Assessment BMDLs
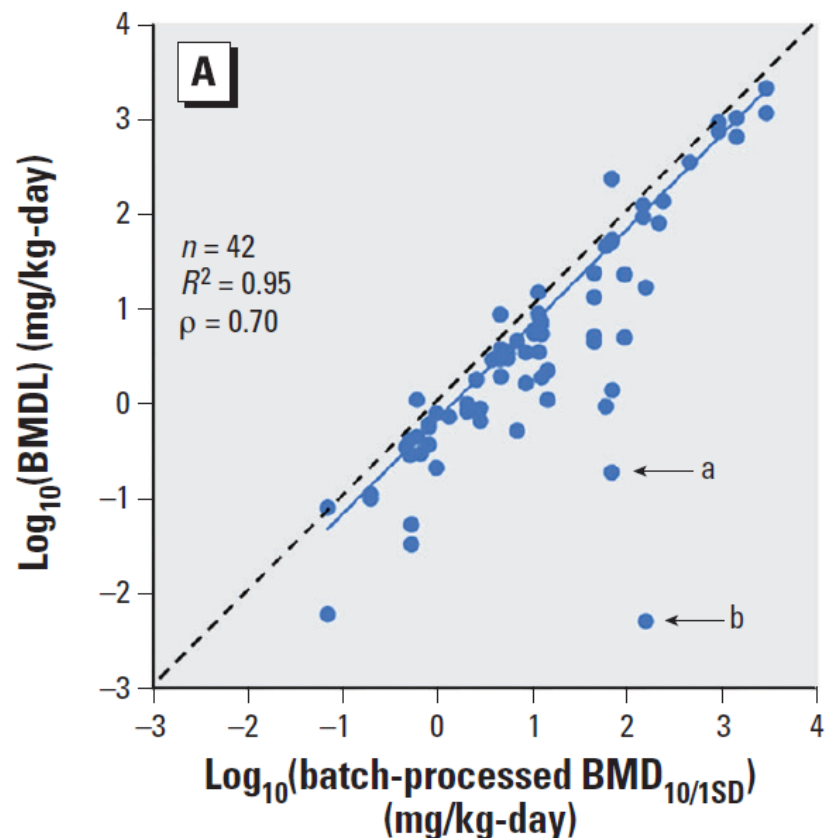
"a" denotes dichloromethane values; "b" denotes trichloroethylene values; both assessments adjusted for human toxicokinetic variability prior to BMD calculation

Source: Wignall et al., 2014

Batch-Calculated BMDLs Can Facilitate Evaluation of Data on a Specific Chemical

Useful for comparisons among multiple health effects and/or multiple studies, and for identification of outliers

Nitroguanidine
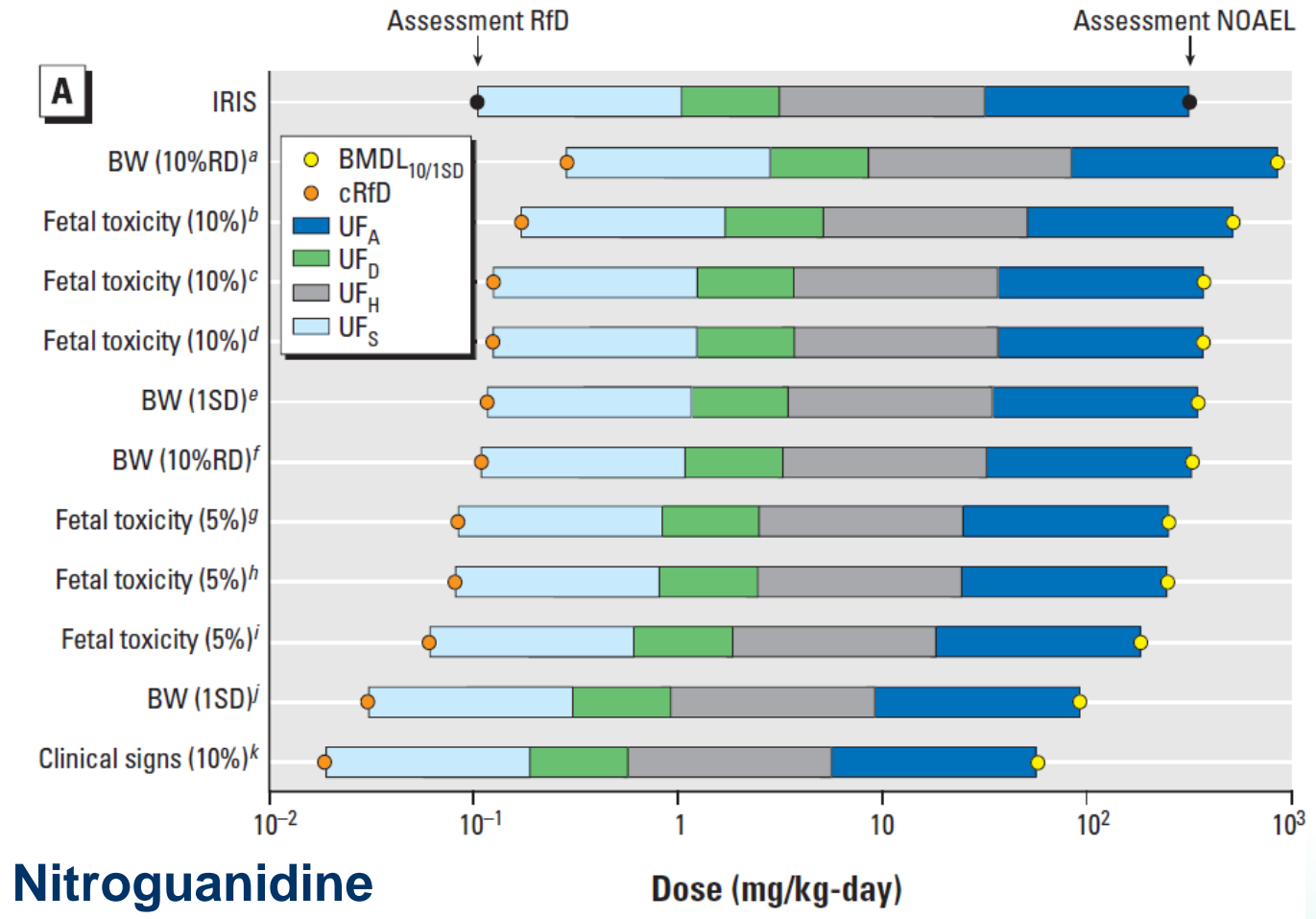
Source: Wignall et al., 2014

Lessons Learned

- BMD/Ls are useful as points of departure
- BMD/Ls can be calculated in a standardized way
- These batch-calculated BMD/Ls can be used for many purposes, including:
  - Evaluating weight of evidence for a chemical, such as across multiple studies or multiple effects
  - Interpreting or using high throughput data, including screening assays or transcriptomics
  - To serve as datasets for QSAR modeling
    - Along with…

Data Exist for Many Types of Toxicity Values

| Toxicity value type	Toxicity value name	Number of compounds with a toxicity value
Oral exposure non-cancer	Reference Dose (RfD)	668
	No Observed Adverse Effect Level (NOAEL)	487
	Benchmark Dose (BMD)	**136**
	Benchmark Dose Lower Level (BMDL)	**136**
Oral exposure cancer	Oral Slope Factor (OSF)	300
	Cancer Potency Value (CPV)	223
Inhalation exposure (non-cancer and cancer)	Reference Concentration (RfC)	149
	Inhalation Unit Risk (IUR)	148

Sources: Integrated Risk Information System; Office of Pesticide Programs; Provisional Peer-Reviewed Toxicity Values; Agency for Toxic Substances and Disease Registry; California EPA; Health Effects Assessment Summary Tables (EPA)

Build Models Based on Chemical Structure Features

Why? To predict values of interest to decision makers

- Chemical structures can be represented by numbers
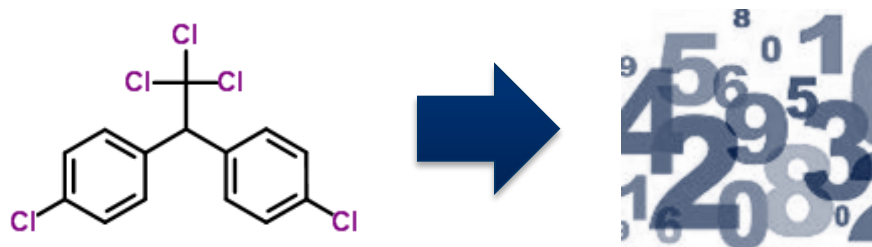
- Statistical algorithms can be used to model the relationship between those features and an outcome of interest
    - f(x) = y → f(chemical descriptors) = outcome
- The algorithm is used to predict the outcome for new chemicals, based on their descriptors
    - Application of algorithm based on structures of original dataset

Chemicals with Toxicity Values Cover a Diverse Chemical Space

- Multi-dimensional chemical descriptor information can be reduced to 2D using PCA

- Plot shows overlap of chemicals with toxicity values and those included in EPA's CERAPP (~32k chemicals)

Principal Component Analysis
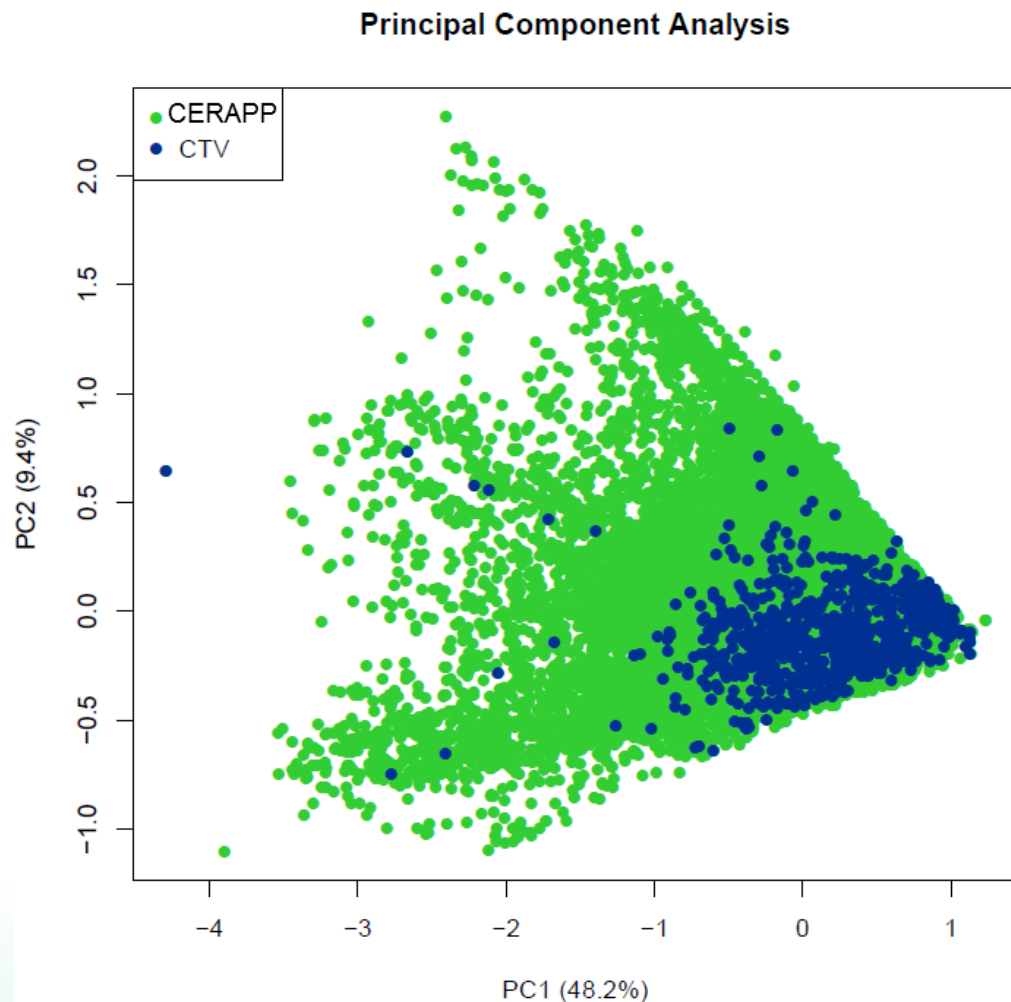
CERAPP = Collaborative Estrogen Receptor Activity Prediction Project

Considerations when Evaluating QSAR Performance

- Model performance should be calculated based on external datasets as much as possible (Tropsha et al., 2003)
- Model performance is limited by how "good" the experimental data is (Lo Piparo et al., 2014)
  - "Prediction errors cannot be better than experimental variability"
- Model performance is improved by using both larger datasets and closely related datasets (McLellan et al., 2011)
- These considerations have implications for predicting *in vivo* outcomes for environmental chemicals, where data is limited and variable.
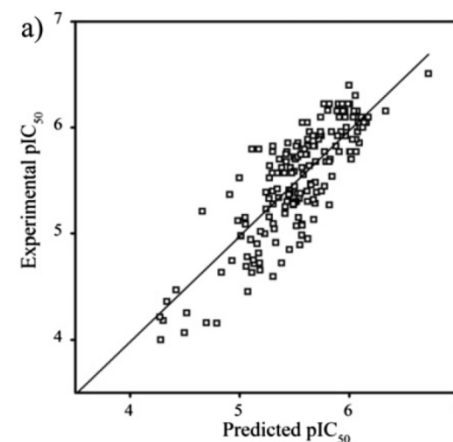
Predictive QSARs Can Be Used to Make Decisions

● Examples of QSAR models and their performance

CAESAR mutagenicity model (Ferrari and Gini 2010)

Source: http://www.caesar-project.eu/index.php?page=results&section=endpoint&ne=3
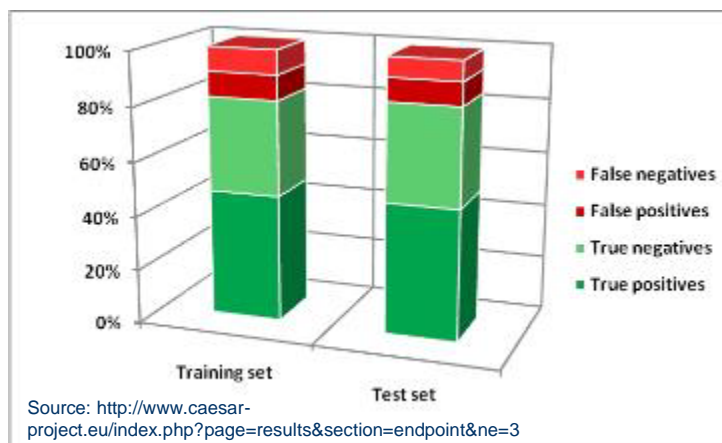
Inhibition of THR binding (Politi et al., 2014)

Prediction of LOAEL (TOPKAT, as reported in Venkatapathy et al., 2009)

Chemical subclass	# of chemicals[a]	# of variables	Adj.-r^2	SD	SE	% chemicals predicted within a factor of				95% chemicals predicted within a factor of
						2	3	4	5	
Acyclics	73	17	0.87	0.85	0.31	73	92	97	100	4
Alicyclics	39	12	0.98	1.49	0.22	94	100			3
Heteroaromatics	68	17	0.85	0.80	0.30	78	92	98	100	4
Multiple Benzenes	83	14	0.78	0.71	0.34	70	92	96	97	4
Single Benzenes	130	23	0.79	0.75	0.34	66	88	94	98	5

[a] Total = 393 chemicals.

Objectives to Build Useful and Predictive Models

1. Predict <u>continuous</u> outcomes that are of use to decision makers, including PODs.

   – Used RfD; NOAEL; BMD; BMDL; OSF; CPV; RfC; and IUR data

2. Facilitate transparency and communication by using publicly available chemical descriptors, easy to understand algorithms, and external validation

   – Descriptor types: cdk + ISIDA → Consensus model

   – Algorithm: Random Forest in Python

   – Validation: 5-fold external cross-validation

3. Provide data through accessible online portals

   – Used online portal ChemBench* to build models

   – Models and predictions available through ToxValue.org

*Carolina Cheminformatics Workbench, developed by the Carolina Exploratory Center for Cheminformatics Research (CECCR); https://chembench.mml.unc.edu/home

Model Performance Varies Across Toxicity Value Type

Toxicity value (# of compounds)	Consensus model Q^2
RfD (668)	0.48
NOAEL (487)	0.51
BMD Non-Cancer (136)	0.34
BMDL Non-Cancer (136)	0.26
OSF (300)	0.43
CPV (223)	0.38
RfC (149)	0.55
IUR (148)	0.38

Distribution of Observed Values
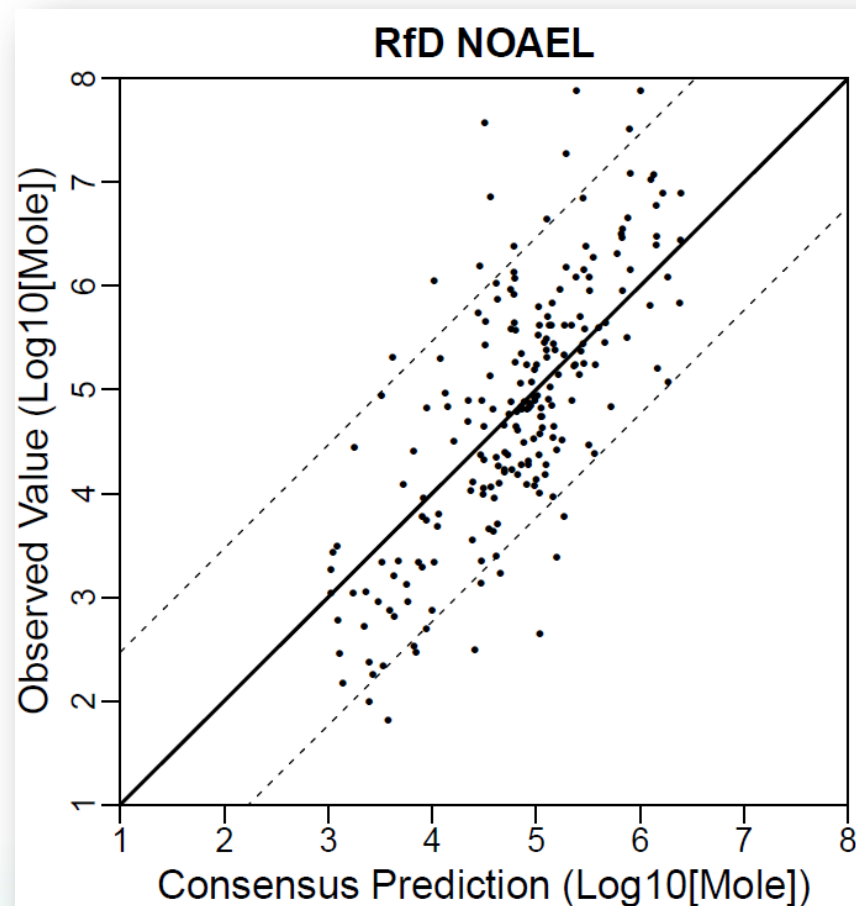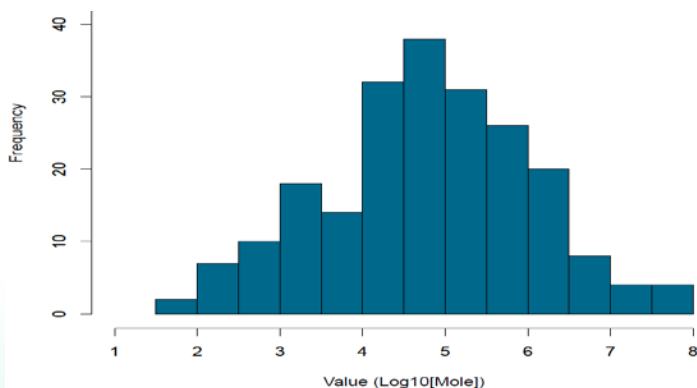
*All models were shown to perform significantly better than chance

Model Predictions Include Training Set-Based Uncertainties

- Average model error based on training set can be applied to new predictions
  - Provides uncertainty range around predictions

Toxicity value	Consensus model Q^2	Consensus model prediction absolute error (log) Average (90% CI)
RfD	0.48	**0.67 (0.05, 1.69)**
NOAEL	0.51	**0.63 (0.043, 1.62)**
BMD NC	0.34	**0.84 (0.08, 1.84)**

Unknown Chemical Z with Descriptor Matrix X

Predicted Value Y

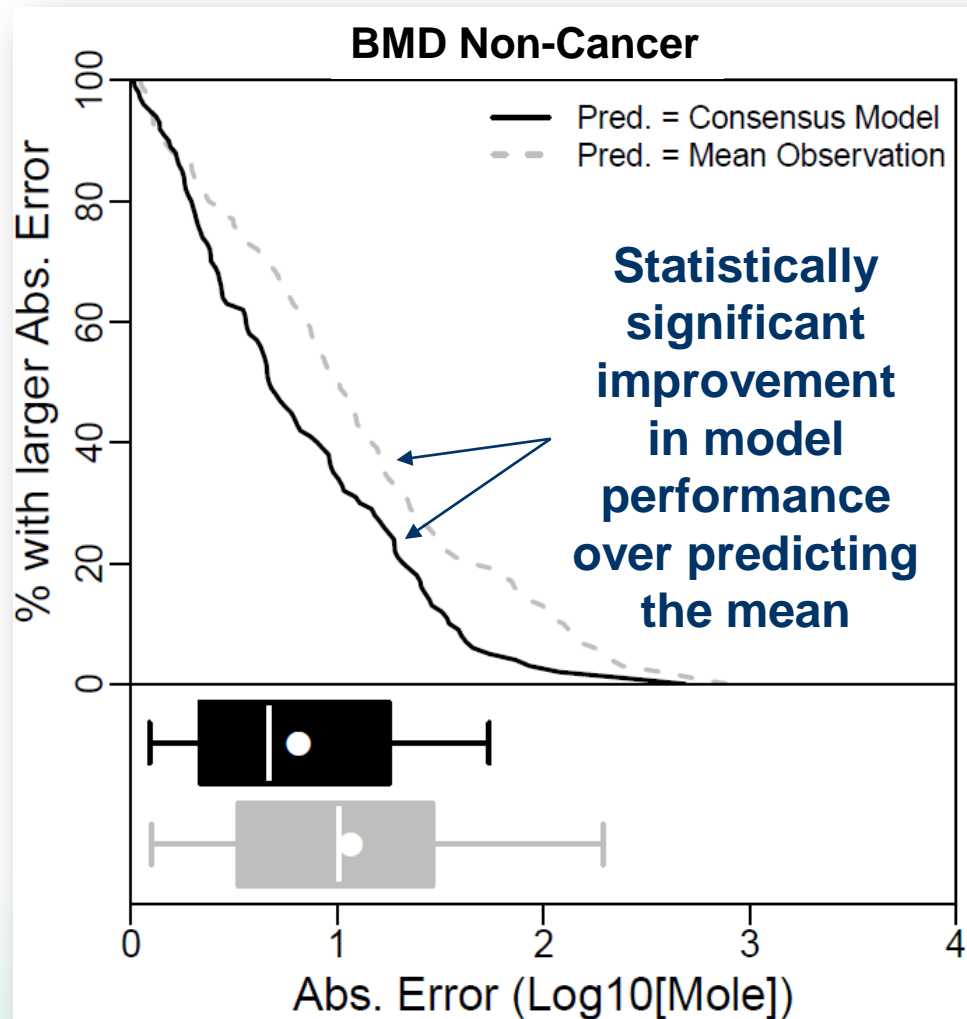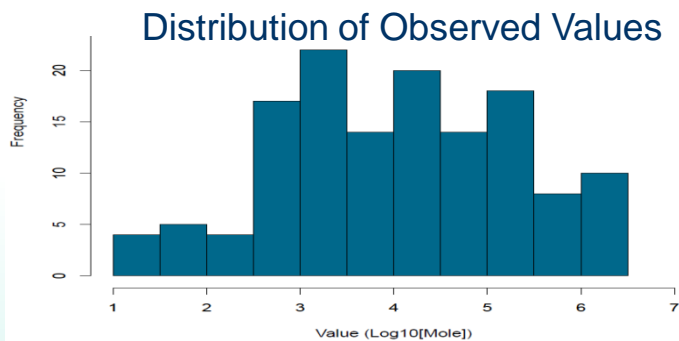|	
RfD	**y ± 0.67**
NOAEL	**y ± 0.63**
BMD NC	**y ± 0.84**

Even Models with Low Predictivity Provide Information

Toxicity value (# of compounds)	Consensus model Q^2	p-value for improvement over average
RfD (668)	0.48	< 0.0001
NOAEL (487)	0.51	< 0.001
BMD NC (136)	0.34	0.014
BMDL NC (136)	0.26	0.12
OSF (300)	0.43	< 0.0001
CPV (223)	0.38	< 0.0001
RfC (149)	0.55	< 0.001
IUR (148)	0.38	< 0.001

Distribution of Observed Values

RfD NOAEL

Pred. = Consensus Model
Pred. = Mean Observation

Statistically significant improvement in model performance over predicting the mean

Even Models with Low Predictivity Provide Information

Toxicity value (# of compounds)	Consensus model Q²	p-value for improvement over average
RfD (668)	0.48	< 0.0001
NOAEL (487)	0.51	< 0.001
BMD NC (136)	0.34	0.014
BMDL NC (136)	0.26	0.12
OSF (300)	0.43	< 0.0001
CPV (223)	0.38	< 0.0001
RfC (149)	0.55	< 0.001
IUR (148)	0.38	< 0.001

Distribution of Observed Values

BMD Non-Cancer

Pred. = Consensus Model
Pred. = Mean Observation

Statistically significant improvement in model performance over predicting the mean

QSAR Models In the Context of Baseline Expectations of Model Uncertainty

- Uncertainty around model predictions can be benchmarked against ability to predict rat chronic lowest effect levels (LEL) from rat subchronic LELs or other models.

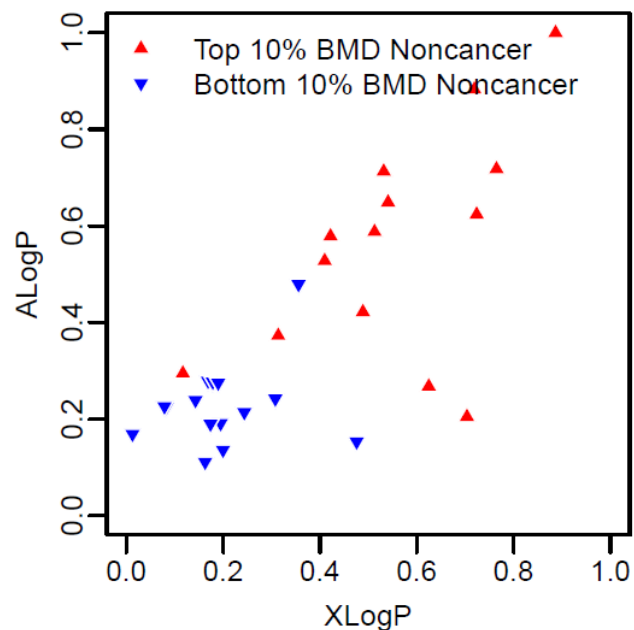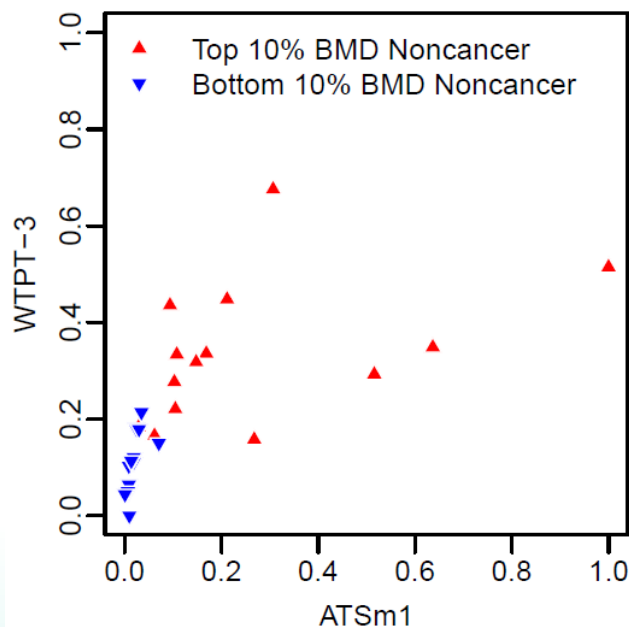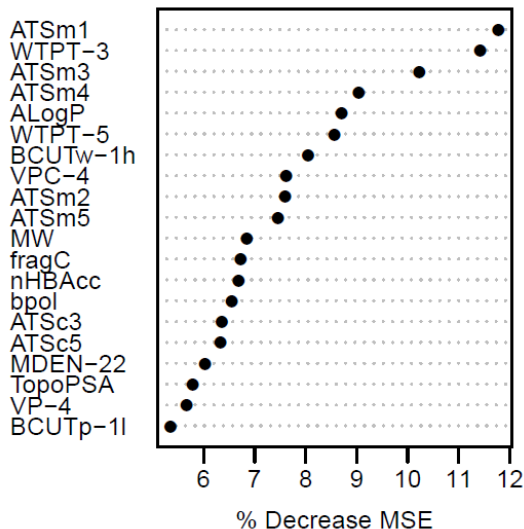- Consensus models reduce uncertainty around predictions compared to other model types.
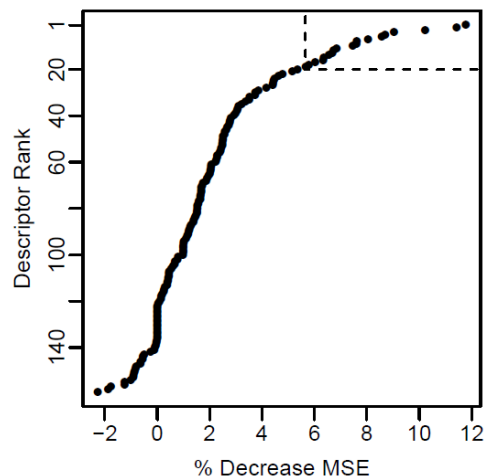
Uncertainy Around Model Predictions
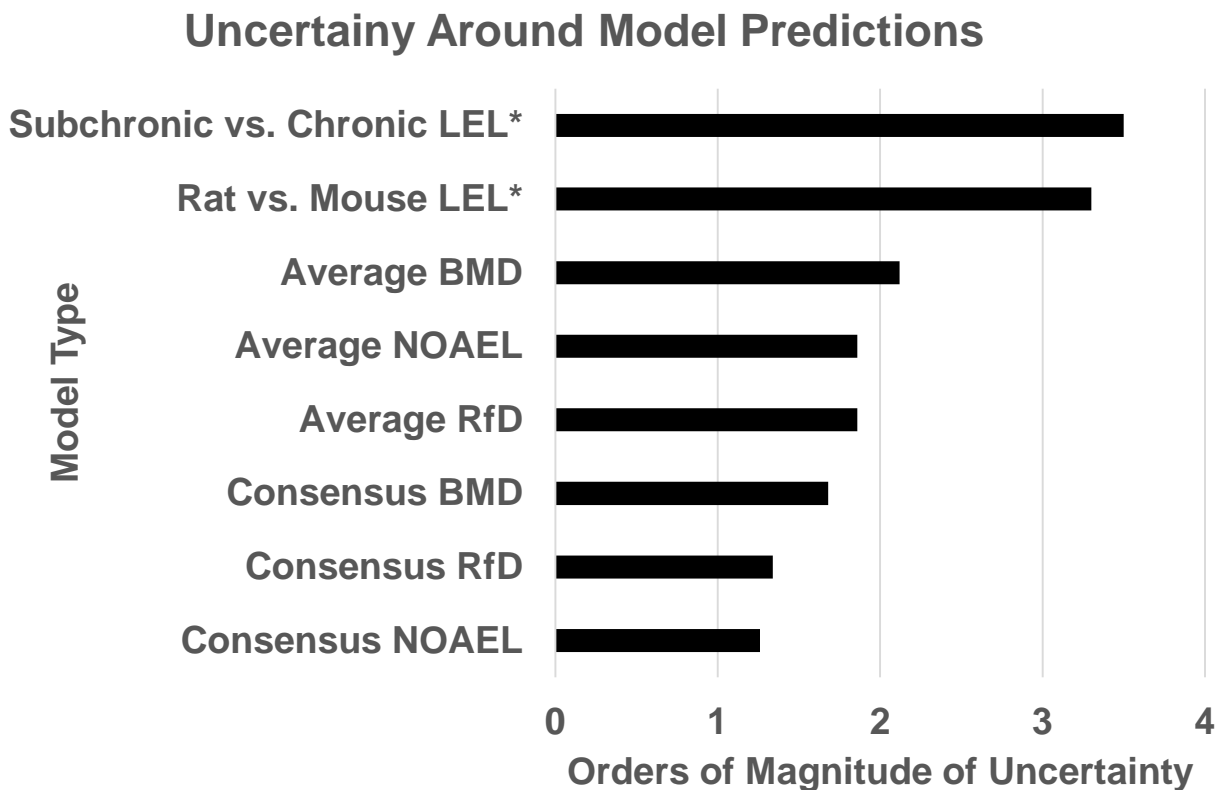
*As reported in previous analyses, source: Matt Martin, Personal Communication

Mechanistic Interpretation Possible

- Based on descriptor sets used
- Trends can highlight areas for interpretation

Summary

- Standardized BMDs can be calculated in an efficient manner

- QSAR models can be built to predict quantitative values of interest to decision makers

- Uncertainty around model predictions can be quantified

- The results of these models can be presented in the context of baseline expectations

Limitations

- Assumptions inherent in aggregating various systemic toxicity endpoints into one dataset
  - The more the homogenous the better, but balanced against need for robust training sets
- Limited *in vivo* data for model building
  - However, efforts underway to extract additional quantitative dose-response data from ToxRefDB animal studies

NOAEL = 1000 mg/kg
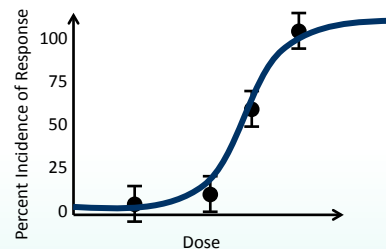
When Could This Be Used?

Addressing the "no data–no hazard" challenge:

- Generating screening values for risk decisions
- Predicting potential toxicity to inform risk management of emerging contaminants
- Prioritizing chemicals for grouping, testing, and assessment
- Evaluating alternatives in green chemistry/sustainable design
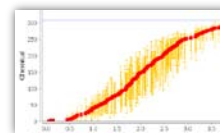
Inputs:

Available for almost all chemicals

Outputs:

Quantitative Predictions

IUR BMD RfC CPV OSF LOAEL POD BMDL RfD NOAEL

Case Study Application – Accidental Releases

- Site-specific sampling shows presence of chemical with no existing toxicity values

- Decision makers need to make quick decisions

- The models presented here could serve as another resource to stakeholders predicting potential hazard of compounds lacking *in vivo* data

ToxValue.org – Home of the Conditional Toxicity Value Predictor

Online Portals Can Be Used to Make Predictions

Step 1: Enter Compound Information

Available at ToxValue.org and ChemBench

Online Portals Can Be Used to Make Predictions

Step 2: Confirm and Select Compounds

Available at ToxValue.org and ChemBench

Online Portals Can Be Used to Make Predictions

Step 3: Select Models

Models are hosted on ChemBench

Online Portals Can Be Used to Make Predictions

Retrieve Predictions

To get future updates to the site: conditionaltoxvalue@gmail.com

Available at ToxValue.org and ChemBench

References

- Models referenced in this presentation can be found at the following sites
    - http://www.toxvalue.org/
    - https://chembench.mml.unc.edu/home
- Cdk descriptors: http://wiki.qspr-thesaurus.eu/w/CDK
- Ferrari T., Gini G. 2010. An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chemistry Central Journal, 4(Suppl 1):S2.
- ISIDA descriptors: http://infochim.u-strasbg.fr/spip.php?rubrique49
- Lo Piparo E, Maunz A, Helma C, Vorgrimmler D, Schilter B. 2014. Automated and reproducible read-across like models for predicting carcinogenic potency. Regul Toxicol Pharmacol. 70(1):370-8. doi: 10.1016/j.yrtph.2014.07.010. Epub 2014 Jul 15.
- McLellan, M. R., Ryan, M. D., & Breneman, C. M. 2011. Rank Order Entropy: why one metric is not enough. Journal of Chemical Information and Modeling, 51(9), 2302–2319. http://doi.org/10.1021/ci200170k
- Politi R, Rusyn I, Tropsha A. 2014. Prediction of binding affinity and efficacy of thyroid hormone receptor ligands using QSAR and structure-based modeling methods, Toxicology and Applied Pharmacology, 280(1):177-189, ISSN 0041-008X, http://dx.doi.org/10.1016/j.taap.2014.07.009.
- Tropsha, A., Gramatica, P. and Gombar, Vijay K. 2003. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR Comb. Sci., 22: 69–77. doi: 10.1002/qsar.200390007
- Venkatapathy R, Wang CY, Bruce RM, Moudgal C. 2009. Development of quantitative structure-activity relationship (QSAR) models to predict the carcinogenic potency of chemicals I. Alternative toxicity measures as an estimator of carcinogenic potency. Toxicol Appl Pharmacol. 2009 Jan 15;234(2):209-21. doi: 10.1016/j.taap.2008.09.028. Epub 2008 Oct 15.
- Wignall JA, Shapiro AJ, Wright FA, Woodruff TJ, Chiu WA, Guyton KZ, Rusyn I. 2014. Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. Environ Health Perspect 122:499–505; http://dx.doi.org/10.1289/ehp.1307539

SOT FDA Colloquia on Emerging Toxicological Science Challenges in Food and Ingredient Safety