

# **A framework to build scientific confidence in read-across results**

---

**Grace Patlewicz, PhD**

**National Center for Computational Toxicology (NCCT)**

**US Environmental Protection Agency (US EPA)**

**RTP, NC**

**[patlewicz.grace@epa.gov](mailto:patlewicz.grace@epa.gov)**

# Abbreviations/Definitions

---

- Target – substance of interest, data poor
- Source – analogue with data which will be used to make the read-across prediction
- PMN – Premanufacture notice
- PPRTV - Provisional Peer Reviewed Toxicity Values (for Superfund)
- Reaction domain – organic chemistry reaction mechanisms that characterise electrophilic chemicals
- GenRA – Generalised Read-across

# Talk Objectives

---

## Understanding:

- Workflow for category/analogue approaches
- Importance of the decision context
- Current read-across software tools – where within the category workflow they add most value
- Uncertainty assessment
- Future directions towards quantifying read-across performance and its associated uncertainties

# Workflow for category/analogue approach

---

1. Decision context
2. Data gap analysis
3. Overarching hypothesis
4. Analogue identification
5. Analogue evaluation
  - Data gap filling
6. Uncertainty assessment

# Workflow for category/analogue approach

---



1. Decision context

2. Data gap analysis

3. Overarching hypothesis

4. Analogue identification

5. Analogue evaluation

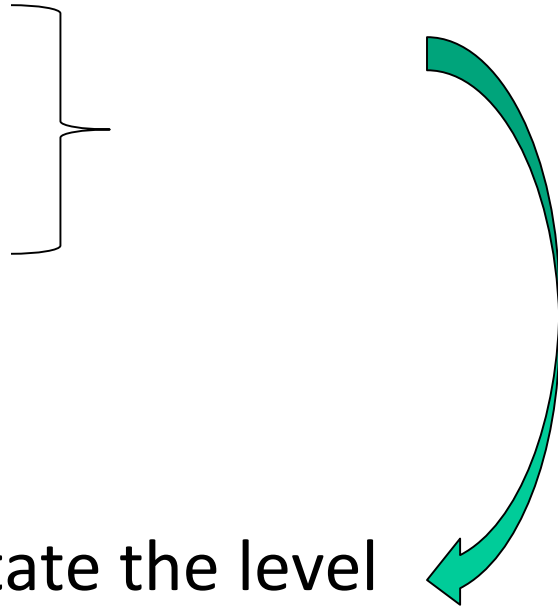
– Data gap filling

Read-Across  
Tools

6. Uncertainty assessment

# 1. Decision context

---

- Prioritisation e.g. PMN
  - Screening level hazard assessment
  - Risk Assessment e.g. PPRTV
- 
- Different decision contexts will dictate the level of uncertainty that can be tolerated

# Workflow for category/analogue approach

---

1. Decision context
2. Data gap analysis
3. Overarching hypothesis
4. Analogue identification
5. Analogue evaluation
  - Data gap filling
6. Uncertainty assessment

# Read-across Tools – An Illustrative List

Tool	OECD Toolbox	ToxMatch	AMBIT	ToxRead
Analogue identification	X	X	X	X
Analogue Evaluation	X	X	X To an extent by other predictive tools available	X For Ames & BCF
Data gap analysis	X Data matrix viewable		X Data matrix can be exported	
Availability	Free	Free	Free	Free



## 2. Data gap analysis

---

- Evaluating the completeness of the data matrix to identify specific data gaps for a target substance
- Depends on access to high quality study data
  - Study quality can be assessed using frameworks such as that proposed in Klimisch et al 1997
  - ToxRTool is a software tool that can facilitate such an assessment

## 2. Data gap analysis

---

- Read-across tools that allow data gaps to be quickly identified for the target chemical include:
  - AMBIT
  - OECD Toolbox














# Data matrix: AMBIT

Home > All assessments > This assessment > Versions >

Assessment identifier
Collect structures
Endpoint data used
Assessment details
Report














































Initial matrix
Working matrix
Final matrix

Identifiers
P-CHEM
ENV FATE
TOX

Download dataset as:














Showing from 1 to 5 in pages of 20
▼ entries
Previous
Next

Filter...

	CAS	Substance Name	ISUUUID	Data source	Tag	Diagram	4.7. Partition coefficient	5.1.2. Hydrolysis	7.2.1. Acute toxicity - oral
- 1 -  ▼		Hydroxypropyl met hacrylate	ECHA-3d7...	-	  	  	0.97 (Temperature = 20.0 °C, pH = 2.0 8.0)  0.9 (Temperature = 20.0 °C, pH = 7.0)  0.79 (Temperature = 20.0 °C, pH = 9.0)  0.48 (Temperature = 20.0 °C, pH = 11.0) 	ca. (pH = 4.0)  73.3_d (Temperature = 40.0 °C, pH = 7.0)  38.2_h (Temperature = 40.0 °C, pH = 9.0) 	LD50 > 4000_mg/kg bw (Species = rat)  LD50 > 5000_mg/kg bw (Species = rat)  LD50 > 2000_mg/kg bw (Species = rat)  LD50 = 11200_mg/kg bw (Species = rat)  LD50 = 6162_mg/kg bw (Species = mouse)  LD50 = 7965_mg/kg bw (Species = mouse) 
- 2 -  ▲ ▼		2-methoxyethyl acr ylate	ECHA-6e3...	-			0.9 (Temperature = 25.0 °C) 	84.6_h (Temperature = 25.0 °C, pH = 9.0)  24700_h (Temperature = 25.0 °C, pH = 7.0) 	LD50 = 404_mg/kg bw (Species = rat)  LD50 ca. 818_mg/kg bw (Species = rat)  LD50 ca. 0.81_mL/kg bw (Species = rat) 
- 3 -  ▲ ▼		2,3-epoxypropyl m ethacrylate	ECHA-2f8...	-	 	 	ca. 0.96 (Temperature = 25.0 °C, pH = ca.7.0) 	ca. 2.83_d (Temperature = 25.0 °C, pH = 4.0)  ca. 4.1 (Temperature = 25.0 °C, pH = 7.0)  ca. 1.9 (Temperature = 25.0 °C, pH = 3.0)  ca. 0.054 (Temperature = 25.0 °C, pH = 11.0)  ca. 3.66_d (Temperature = 25.0 °C, pH = 7.0) 	LD50 = 1050_mg/kg bw (Species = mouse)  LD50 = 390_mg/kg bw (Species = mouse)  LD50 = 697_mg/kg bw (Species = guinea pig)  LD50 = 451_mg/kg bw (Species = rat)  LD50 ca. 700_mg/kg bw (Species = rat) 

# Data matrix: OECD Toolbox

QSAR Toolbox 3.4.0.17 [Document]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Data Import Export Delete Tautomerize

Gather Import IUCLID5 Export IUCLID5 Database Inventory Database

Databases

Select All Unselect All Invert About

- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards

Structure

Substance Identity

- CAS Number: 27813-02-1
- Chemical IDs: EINECS:2486663
- Chemical Name: hydroxypropyl methacrylate, methacrylic acid, monoester with prop... 2-propenoic acid, 2-methyl-, monoeste... 2-hydroxypropyl 2-methylprop-2-enoate, methacrylic acid, monoester with 1,2-p... methacrylic acid, ester with 1,2-propan...
- Molecular Formula: C7H12O3
- Structural Formula: CC(O)COC(=O)C(C)=C

Physical Chemical Properties (1/32): M: 1.07E5 mg/L, 1.3E5 mg/L, 10.7 vol...

Environmental Fate and Transport (1/7): M: Calculation according to Mackay, L...

Ecotoxicological Information (1/13): M: 379 mg/L, 493 mg/L, 641 mg/L, 83...

Human Health Hazards (1/41): M: 1E3 mg/kg bw/day, 50 mg/kg bw/d...


Profile

- General Mechanistic
- Protein binding by OASIS v1.4
- Protein binding by OECD
- Endpoint Specific
- Protein binding alerts for skin sensitization by OA...

Inventories

Select All Unselect All Invert About

1 [target]



AN2

AN2 >> Michael addition to alpha, bet...

AN2 >> Michael addition to alpha, bet...

Michael addition

Michael addition >> Michael addition o...

Michael addition >> Michael addition o...

Michael addition

Michael addition >> Polarised Alkenes

Michael addition >> Polarised Alkenes...

Michael Addition

Michael Addition >> Michael addition o...

Michael Addition >> Michael addition o...

<https://www.qsartoolbox.org/>

# Steps 3 to 5 of the workflow

---

- Read-across tools that assist in identifying similar analogues and justifying their similarity for the endpoint of interest include:
  - OECD Toolbox
  - ToxMatch
  - ToxRead

# **Analogue identification and evaluation:**

## **OECD Toolbox**

- Define an endpoint specific category to predict e.g. skin sensitisation potential for a target chemical
- Overarching similarity rationale = same protein binding alerts
- Data matrix is updated to reflect target and potential source analogues

# Analogue identification and evaluation

QSAR Toolbox 3.4.0.17 [Document]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Define Define with metabolism Subcategorize Combine Clustering Delete Delete All

Grouping methods

- Predefined
  - Database Affiliation
  - Inventory Affiliation
  - OECD HPV Chemical Categories
  - Substance Type
  - US-EPA New Chemical Categories
- General Mechanistic
  - Biodeg BioHC half-life (Blowin)
  - Biodeg primary (Blowin 4)
  - Biodeg probability (Blowin 1)
  - Biodeg probability (Blowin 2)
  - Biodeg probability (Blowin 5)
  - Biodeg probability (Blowin 6)
  - Biodeg probability (Blowin 7)
  - Biodeg ultimate (Blowin 3)
  - DNA binding by OASIS v.1.4
  - DNA binding by OECD
  - DPRA Cysteine peptide depletion
  - DPRA Lysine peptide depletion

Defined Categories

- Document
  - [481] AN2<AND>AN2>> Michael addition to

Structure

Immunotoxicity

- Itiritation / Corrosion (101/275) M: not irritating, moderately irritating, n...
- Neurotoxicity (10/15)
- Photoinduced Toxicity
- Repeated Dose Toxicity (69/6204) M: 300 mg/kg bw/day (nominal), 0.5 mg/L M: 15 mg/kg bw/d...
- Sensitisation
- Respiratory Tract
- Skin
- In Chemico
- In Vitro (18/114)
- In Vivo
- Alternative Methods (1/1)
- Buehler Test (5/5)
- Intracutaneous and Topical S... (1/1)
- Reaction Pattern (1/1)
- ist (2/2)
- Complete Adjuvant Test (12/14)
- ig Local Lymph Node Assay (8/8)
- ig Maximisation Test (46/64) M: not sensitising,...
- Technique (4/6)
- Human Patch Test and Guinea Pig Mag... (1/1)
- LLNA
- EC3 (20/31)
- Maximization Test and Observations of ... (1/1)
- Miscellaneous (44/62)
- Modified Draize Test (1/1)
- Modified Maximization Test (1/1)
- Mouse Ear Swelling Test (4/4)
- Mouse Local Lymphnode Assay (LLNA) (1/1)
- Skin Sensitisation
- No Data (1/1)
- Open Epicutaneous Test (5/7)

Source substances

Target

Data gap

Endpoint specific

Similarity rationale

Protein binding by OASIS v1.4

Protein binding by OECD

Superfragments

- Toxic hazard classification by Cramer (ext)
- Toxic hazard classification by Cramer (orig)
- Ultimate biodeg
- Biodeg BioHC half-life (Blowin)
- Biodeg primary (Blowin 4)

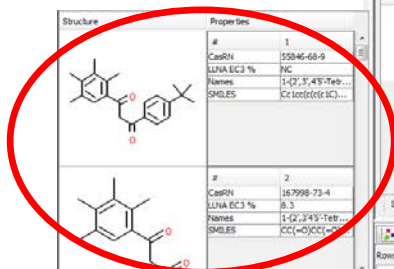
# Analogue identification and evaluation: Toxmatch

- Identify similar analogues on the basis of fingerprints from a predefined dataset e.g. skin sensitisation
- Filter analogues on the basis of a similarity index threshold

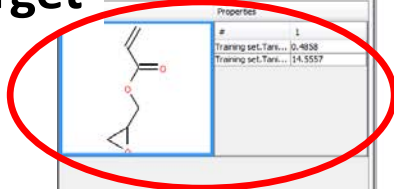


# Toxmatch

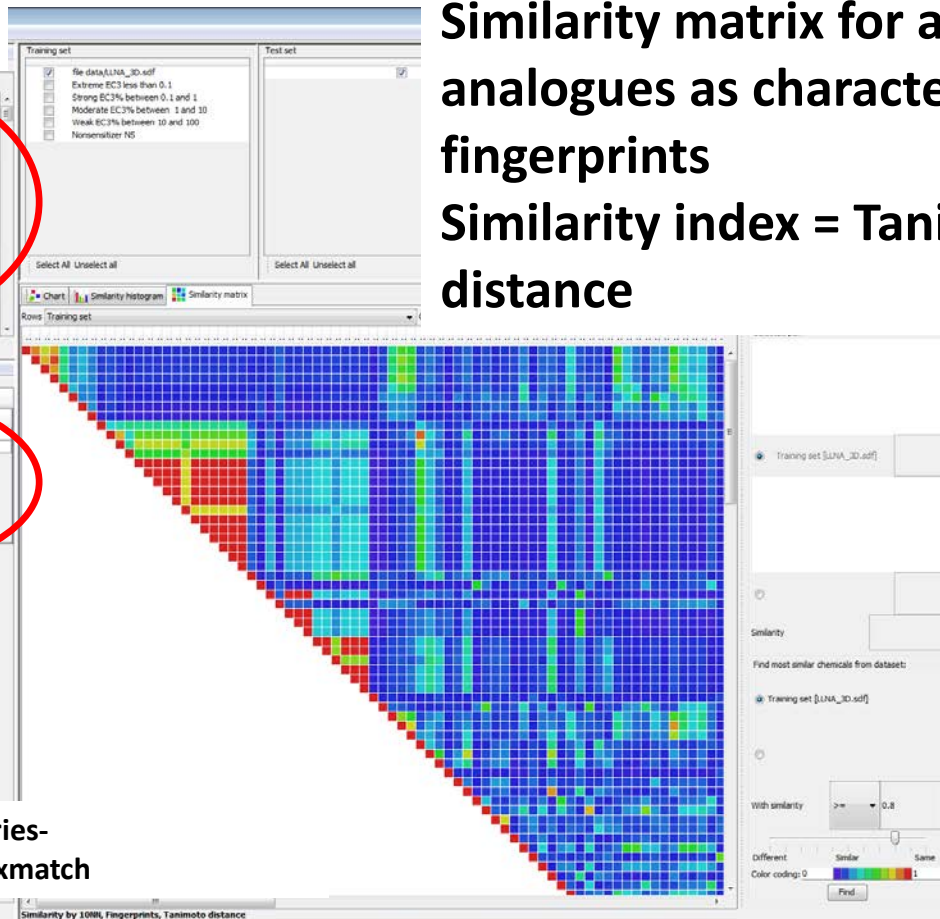
Source analogues



Target



Similarity matrix for all source analogues as characterised by fingerprints  
Similarity index = Tanimoto distance



# Toxmatch

**Target**

The screenshot displays the Toxmatch application window. On the left, the 'Structure' panel shows two chemical structures. The top structure is circled in red and labeled 'Target'. Below it, another structure is also circled in red. The 'Properties' panel for the target structure shows: # 1, Cactx: 818-01-1, LINA EC3 % 1.4, Names: [2-Hydroxyethyl ...], SMILES: OCCOC(=O)C=C. The 'Training set' panel on the right shows a list of training set files, including 'file: data\LINA\_3D.sdf'. The 'Test set' panel shows a list of test set files, including 'file: default.sdf'. The 'Similarity matrix' panel at the bottom right shows a similarity score of 0.763 for the target structure. A red arrow points from the 'Similarity matrix' panel to the 'Similar analogues' text box. Another red arrow points from the 'Similar analogues' text box to the 'Find' button in the 'Find most similar chemicals from dataset' section.

**Similar analogues within a similarity threshold of 0.5**

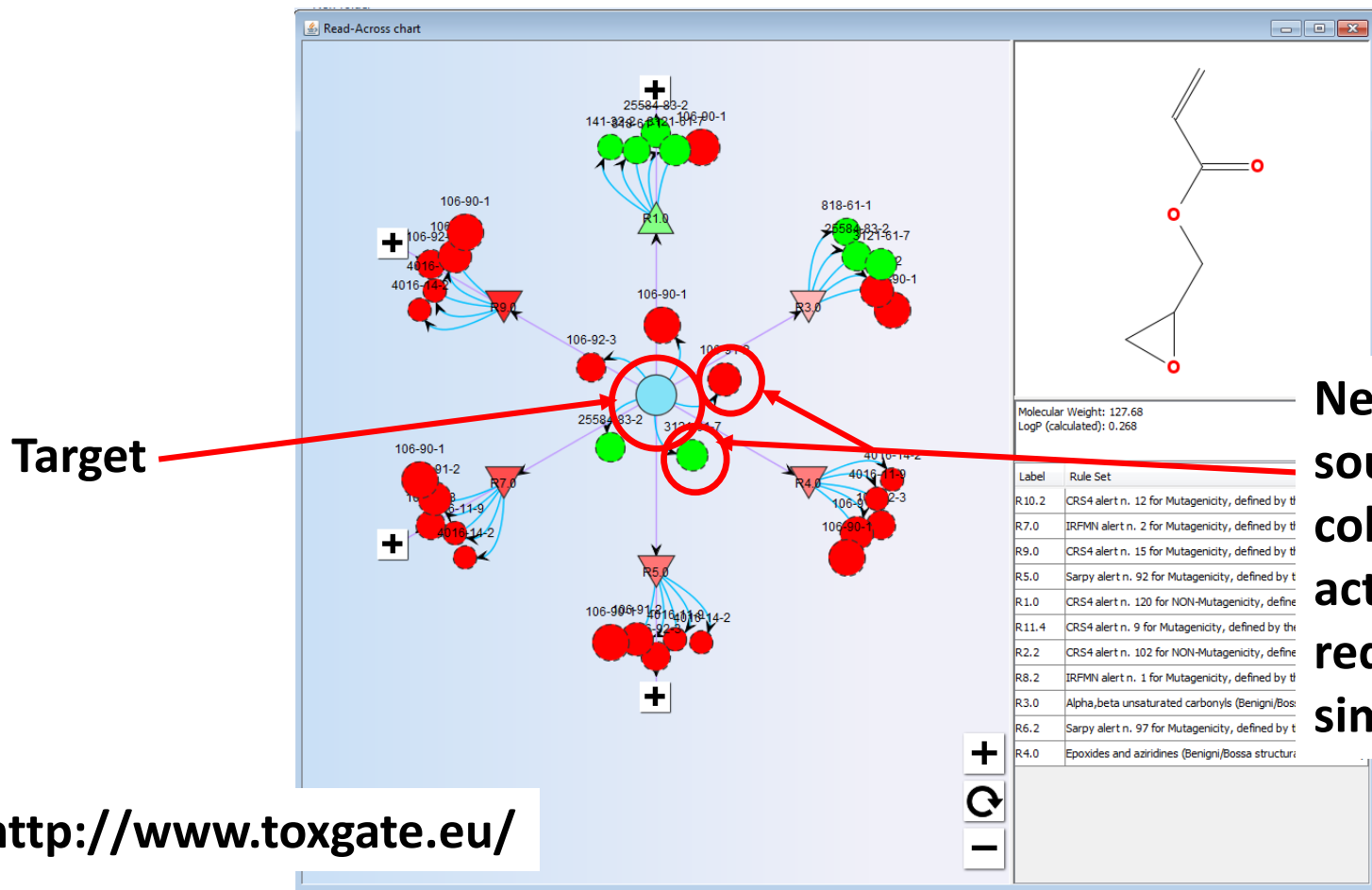
Similarity by 100%, Fingerprints, Tanimoto distance

# **Analogue identification and evaluation:**

## **ToxRead**

- Identify similar analogues on the basis of structural similarity and structural alerts
- Endpoints covered are mutagenicity and bioconcentration potential
- User defines number of source analogues

# ToxRead



**Neighbouring  
source analogues,  
colour coded by  
activity (positive =  
red) and by  
similarity index**

**<http://www.toxgate.eu/>**

# 6. Uncertainty assessment

---

- A number of publications exist that can guide the construction and assessment of categories and use of read-across
  - Guidance and examples (OECD, 2014; ECHA, 2008; ECETOC TR 116, 2012;)
  - Frameworks for identifying analogues e.g. Wu et al, 2010, Patlewicz et al, 2013
  - Frameworks for assessing read-across (Blackburn and Stuart, 2014, Patlewicz et al, 2015; Patlewicz et al, 2015; ECHA – RAAF, 2015; Schultz et al, 2015; Ball et al, 2016)

See references list for full citations

## 6. Sources of uncertainty

---

- Analogue or category approach? (# analogues)
- Completeness of the data matrix - no of data gaps
- Data quality for the underlying analogues for the target and source analogues
- Consistency of data across the data matrix – concordance of effects and potency across analogues

## 6. Sources of uncertainty (cont'd)

---

- Overarching hypothesis/Similarity rationale – how to identify similar analogues and justify their similarity for the endpoint of interest
- Address the dissimilarities and whether these are significant from a toxicological standpoint
- Presence vs absence of toxicity
- Toxicokinetics

# Strategies to evaluate and address uncertainties

## - addressing dissimilarities

---

- Evaluating whether structural differences of the source analogue may impact the toxicity relative to the target substance
- Are there specific structural alerts identified for the structural features that are not common between the target and source analogues?
  - e.g. Use of systems such as the OECD Toolbox, Derek Nexus can be helpful in identifying specific structural alerts



# Strategies to evaluate and address uncertainties

## - addressing dissimilarities (cont'd)

---

- Do the structural differences translate to significant differences to the metabolic pathway between source and target analogue that could result in differences in toxicity? e.g. Use of the OECD Toolbox's metabolic simulators or METEOR may prove helpful in exploring the metabolic pathways and their differences
- Do the structural differences result in significant differences to the physicochemical properties that could impart differences in bioavailability? e.g. Estimation of LogKow and MW can provide useful insights into potential differences in bioavailability

# Strategies to evaluate and address uncertainties – **toxicokinetics and metabolism**

---

- Toxicokinetics – including Metabolism
  - Underlying rationale presumes a metabolic transformation e.g. Source analogue => Target
  - Assumption is that this transformation is rapid and complete
  - What sort of practical approaches can be applied to demonstrate that such transformation occurs?

# Strategies to evaluate and address uncertainties

## – toxicokinetics and metabolism (cont'd)

---

- Predict likely metabolite(s) using in silico tools
  - e.g. OECD Toolbox, Meteor Nexus, MetaPrint 2D, TIMES, Catalogic
- Assessing metabolism through one or another experimental systems.
  - E.g. precision-cut tissue slices, subcellular fractions such as the microsomal fraction, primary cells (immortalized, in suspension, monolayers in culture), cell lines (continuous, liver-derived etc.)

# Read-across performance

---

- Uncertainty that can be tolerated depends on the decision context
  - However read-across acceptance relies on a subjective expert assessment
  - Uncertainty assessment is qualitative in nature
  - There is no objective measure of read-across performance
  - But there are efforts in progress
- (NB: previous presentation)**

# Quantifying uncertainty & Assessing performance of read-across

---

- GenRA (Generalised Read-Across) is a “local validity” approach
- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and bioactivity descriptors
- Systematically evaluates read-across performance and uncertainty using available data

Jaccard similarity:

# GenRA - Approach

---

## I. Data

1,778 Chemicals  
3,239 Structure descriptors (chm)  
820 Bioactivity assays (bio)  
ToxCast  
574 Apical outcomes (tox)  
ToxRefDB

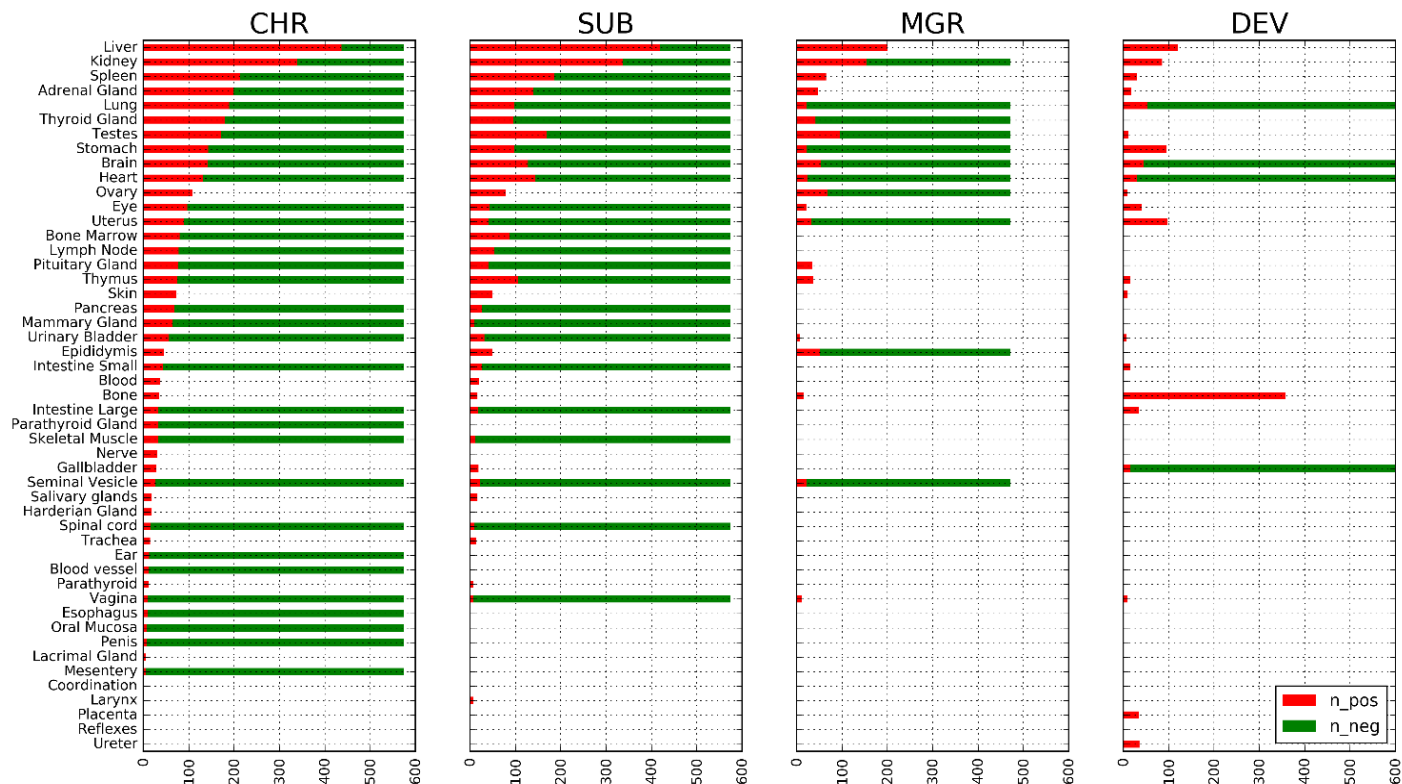
## II. Define Local neighborhoods

Use K-means analysis to group chemicals by similarity  
Use cluster stability analysis  
~ 100 local neighborhoods

## III. GenRA

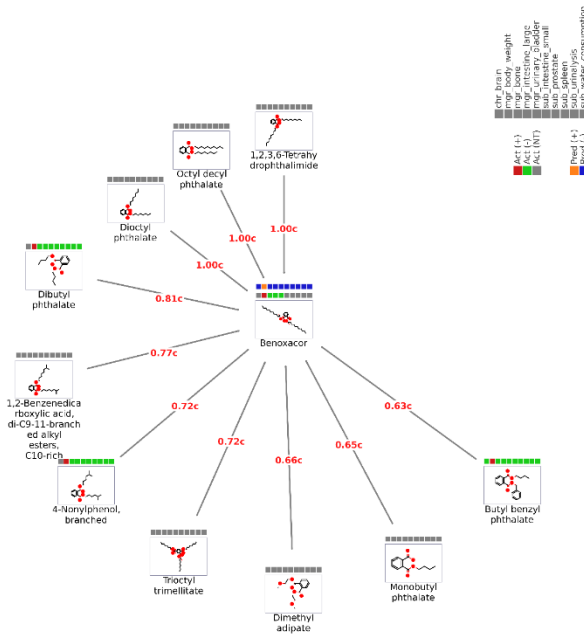
Use GenRA to predict apical outcomes in local neighborhoods  
Evaluate impact descriptors (chm, bio, bc) on prediction  
Quantify uncertainty

# GenRA - Toxicity Data from ToxRefDB



# GenRA – performance in each cluster

- Use GenRA to predict the similarity weighted toxicity scores for each
  - Toxicity type ( $\beta$ )
  - Descriptor = {chm,bio,bc} ( $\alpha$ )
  - No. of nearest neighbours ( $k$ )
  - Similarity score threshold (  $s_{ij}^{\alpha}$  )
- Calculate performance by comparing predicted  $y^{tox}$  and true  $x^{tox}$  for all chemicals using area under ROC curve (AUC)
- Results: {cluster,  $\alpha, \beta, k, s, AUC$ }







# GenRA – Insights and Next steps

---

- Bioactivity descriptors were often found to be more predictive of *in vivo* toxicity outcomes
- The approach enabled a performance baseline for read-across predictions of specific study outcomes to be established but was still context dependent on the endpoint and the chemical
- Next steps:
- Use of other chemical descriptor sets that encode more expert knowledge of SARs
- Incorporating TK information

# Conclusions

---

- Current workflow for developing category/analogue approaches follows a series of steps
- Decision context is a key consideration as this will drive the level of uncertainty that can be tolerated
- There are many sources of uncertainty and proposals to address these
- To move towards quantifying uncertainties we need to consider different approaches to structuring read-across
- An example is provided to illustrate some of the possibilities

# Talk Objectives

---

Understanding:

- Workflow for category/analogue approaches ✓
- Importance of the decision context ✓
- Current read-across software tools – where within the category workflow they add most value ✓
- Uncertainty assessment ✓
- Future directions towards quantifying read-across performance and its associated uncertainties ✓

# Acknowledgements

---

Imran Shah – US EPA

Rusty Thomas – US EPA

Nicholas Ball – The Dow Chemical Company

# References

---

## Data Quality

- Conrad JW, Jr, Becker RA. 2011. Environ Health Perspect. 119: a508–a509.
- <https://arasp.americanchemistry.com/Data-Quality-Evaluation.pdf>
- <https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/toxrtool>
- Samuel GO, et al 2016 Environ Int. 92-93:630-46.

# References

---

## Guidance and examples

- OECD, 2014:  
[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO\(2014\)4&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2014)4&doclanguage=en)
- ECETOC TR 116: <http://www.ecetoc.org/publication/tr-116-category-approaches-read-across-qsar/>

# References

---

## **Frameworks for identifying analogues:**

- **Wu S et al 2010. Regul Toxicol Pharmacol. 56(1):67-81.**
- **Patlewicz G et al 2013 Regul Toxicol Pharmacol. 67(1):1-12.**



# References

---

## Frameworks for assessing read-across:

- Blackburn K, Stuard SB. 2014 Regul Toxicol Pharmacol. 68(3):353-62.
- Patlewicz G et al 2014 ALTEX. 2014;31(4):387-96. Patlewicz G et al 2015 Regul Toxicol Pharmacol. 72(1):117-33.
- Schultz TW et al 2015 Regul Toxicol Pharmacol. 72(3):586-601.
- ECHA RAAF  
[https://echa.europa.eu/documents/10162/13628/raaf\\_en.pdf](https://echa.europa.eu/documents/10162/13628/raaf_en.pdf)
- Ball N et al 2016 ALTEX. 33(2):149-66.

# References

---

## **New approaches in read-across**

- **Low Y et al 2013 Chem Res Toxicol. 26(8):1199-208.**
- **Shah I et al 2016 Regul Toxicol Pharmacol. 2016 79:12-24.**
- **Zhu H et al 2016 ALTEX. 33(2):167-82.**