Comparison of read-across in REACH registration dossiers with GenRA

G Patlewicz[a], P. Karamertzanis[b], M. Sannicola[b], K. Paul-Friedman[a], I. Shah[a]
[a]Center for Computational Toxicology and Exposure (CCTE), US EPA
[b]Computational Assessment and Alternative Methods, European Chemicals Agency (ECHA), Telakkakatu 6, 00150 Helsinki, Finland

Disclaimers

- The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the US EPA or the European Chemicals Agency (ECHA).

- Mention of or reference to commercial products does not imply any official US EPA or ECHA endorsement.

Outline

- **Read-Across and ongoing issues with read-across acceptance**

- **Generalised Read-Across (GenRA)**

- **Case study to compile expert driven read-across examples from REACH dossiers to evaluate similarity contexts and performance relative to GenRA**

- **Summary remarks**

Read-across

- **Read-across** describes the method of filling a data gap whereby a chemical with existing data values is used to make a prediction for a 'similar' chemical.
- Used within analogue and category approaches.
- A **target chemical** is a chemical which has a data gap that needs to be filled i.e. the subject of the read-across.
- A **source analogue** is a chemical that has been identified as an appropriate chemical for use in a read-across based on similarity to the target chemical and existence of relevant data.
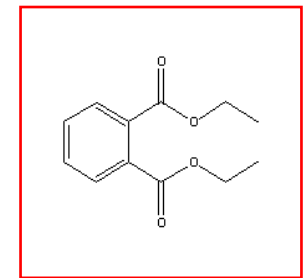
	Source chemical	Target chemical
Property	●	○

● Reliable data

○ Missing data
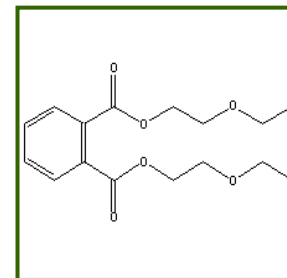
Known to be harmful

Acute toxicity?

Predicted to be harmful

Ongoing issues with read-across

- Although there is much technical guidance for developing read-across assessment, acceptance remains an issue.

- One issue hindering acceptance relates to what an acceptable level of uncertainty is for a read-across prediction.

- Many efforts have been undertaken to identify the sources of uncertainty in read-across, characterise them in a consistent manner and identify practical strategies to address and reduce those uncertainties.

- Notable in these efforts have been the development of frameworks/templates for the assessment of read-across & evaluating the utility of New Approach Methods (NAMs).

- Quantifying uncertainty and performance of read-across is a need as are ways to better characterise different similarity contexts (metabolism, reactivity etc.)

- Generalised Read-Across (GenRA) attempts to quantify uncertainty and performance of read-across.

GenRA (Generalised Read-Across)

Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information

Imran Shah [a,*], Jie Liu [b,c], Richard S. Judson [a], Russell S. Thomas [a], Grace Patlewicz [a]

[a] National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA
[b] Department of Information Science, University of Arkansas at Little Rock, AR 72204, USA
[c] Oak Ridge Institute for Science Education Fellow, National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

ABSTRACT

Read-across is a popular data gap filling technique within category and analogue approaches for regulatory purposes. Acceptance of read-across remains an ongoing challenge with several efforts underway for identifying and addressing uncertainties. Here we demonstrate an algorithmic, automated approach to evaluate the utility of using in vitro bioactivity data ("bioactivity descriptors", from EPA's ToxCast program) in conjunction with chemical descriptor information to derive local validity domains (specific sets of nearest neighbors) to facilitate read-across for up to ten in vivo repeated dose toxicity study types. Over 3239 different chemical structure descriptors were generated for a set of 1778 chemicals and supplemented with the outcomes from 821 in vitro assays. The read-across prediction of toxicity for 600 chemicals with in vivo data was based on the similarity weighted endpoint outcomes of its nearest neighbors. The approach enabled a performance baseline for read-across predictions of specific study outcomes to be established. Bioactivity descriptors were often found to be more predictive of in vivo toxicity outcomes than chemical descriptors or a combination of both. This generalized read-across (GenRA) forms a first step in systemizing read-across predictions and serves as a useful component of a screening level hazard assessment for new untested chemicals.

© 2016 Published by Elsevier Inc.

- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and/or bioactivity descriptors (Shah et al, 2016)
- Goal: To establish an objective performance baseline for read-across and quantify the uncertainty in the predictions made

$$y_i^{\beta,\alpha} = \frac{\sum_j^k s_{ij}^\alpha x_j^\beta}{\sum_j^k s_{ij}^\alpha}$$

Jaccard similarity:

$$s_{ij} = \frac{\sum_l (x_{il} \wedge x_{jl})}{\sum_l (x_{il} \vee x_{jl})}$$
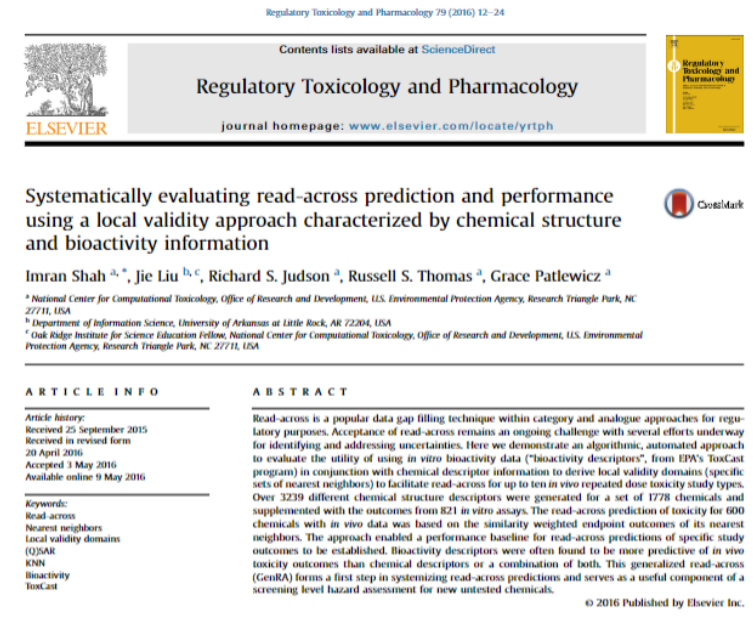
$\alpha \in \{chm, bio, bc\}$

$\beta \in \{bio, tox\}$

$y_i = $ predicted activity of chemical (c_i)

$x_j^\beta = $ activity of c_j in β

$s_{ij}^\alpha = $ Jaccard similarity between x_i^α, x_j^α

$k = $ up to k nearest neighbours

GenRA (Generalised Read-Across)

- Establish an objective performance baseline for read-across in making binary *in vivo* toxicity effect predictions.
- Have systematically evaluated the physicochemical similarity and mechanistic similarity (using targeted transcriptomic and High Throughput Screening data)
- Implemented GenRA into a web application. Version 3.2 is currently released at https://comptox.epa.gov/genra/
- One current focus is in compiling expert read-across examples to facilitate 1) an evaluation of GenRA performance and 2) explore how to quantify the different contributions arising from different similarity contexts.
- Identify a source of expert read-across examples is not trivial – few examples exist in the published literature. Sources of read-across include the EPA PPRTVs, OECD IATA case studies.
- Another source of examples are the published registration dossiers that include read-across that have been submitted to ECHA to satisfy the information requirements under REACH.

Case study

- Compile read-across examples that have been submitted to satisfy the information requirements under REACH from published registration dossiers

- Explore the similarity between target and source substances through the lens of different contexts

- Evaluate performance relative to GenRA

REACH Dossier information

- Downloaded the REACH study results which culminated in 26544 IUCLID dossiers.

- Queried the data to retrieve only dossiers where read-across had been performed for repeated dose toxicity studies by the oral route.

- Identified 3038 associations between registered substances (the targets) and source substances.

- All substances were then queried against the EPA's DSSTox database to retrieve DSSTox Substance identifier (DTXSID), name, CASRN and structural information (SMILES).

- DSSTox information was available for 2224 pairs of substances.

Registered (Target) Substances profile
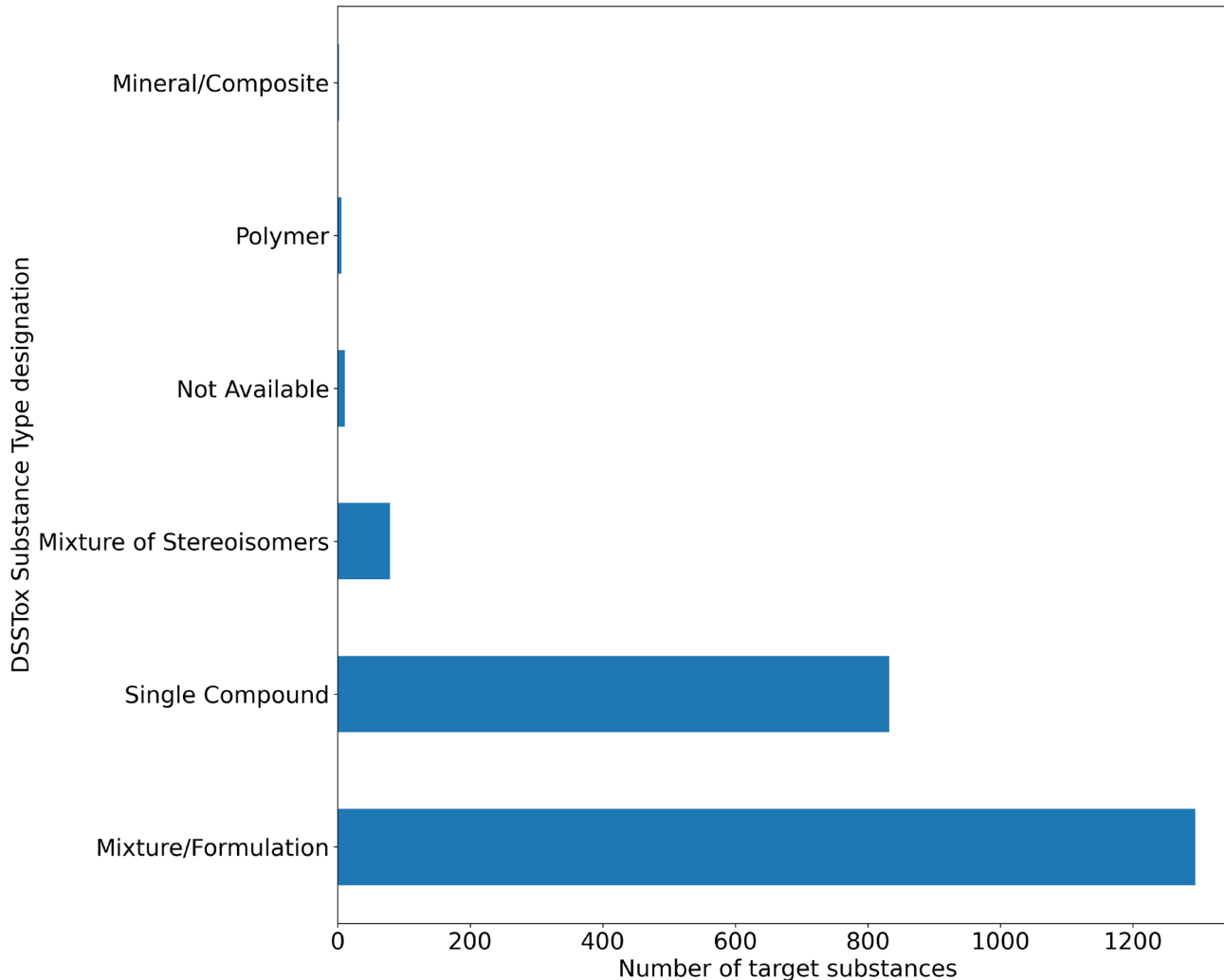
- **Over half (58%) of the registered substances (targets) were designated as mixtures/formulations by DSSTox**

- **Many of these 'mixtures/formulations' are UVCBs and included linear and branched saturated and unsaturated hydrocarbons e.g. C11-12 Alkenes, C14 Hydrocarbons, C11-16 branched and linear Alkanes as well as many Rosin and resin acids**

Focus of the case study

- Focus of the case study concentrated on target-source pairs that comprised organic substances that could be readily represented by a chemical structure (QSAR Ready SMILES).

- Of the associations first identified, 273 unique target-source pairs met these 2 conditions.

- These comprised 203 unique target substances and 179 unique source substances.

- 153 targets (75%) were associated with only 1 source analogue.

- 50 targets were associated with more than 1 source analogue.

- 18% of these were with 2 source analogues, whereas 7% had 3 or more source analogues.
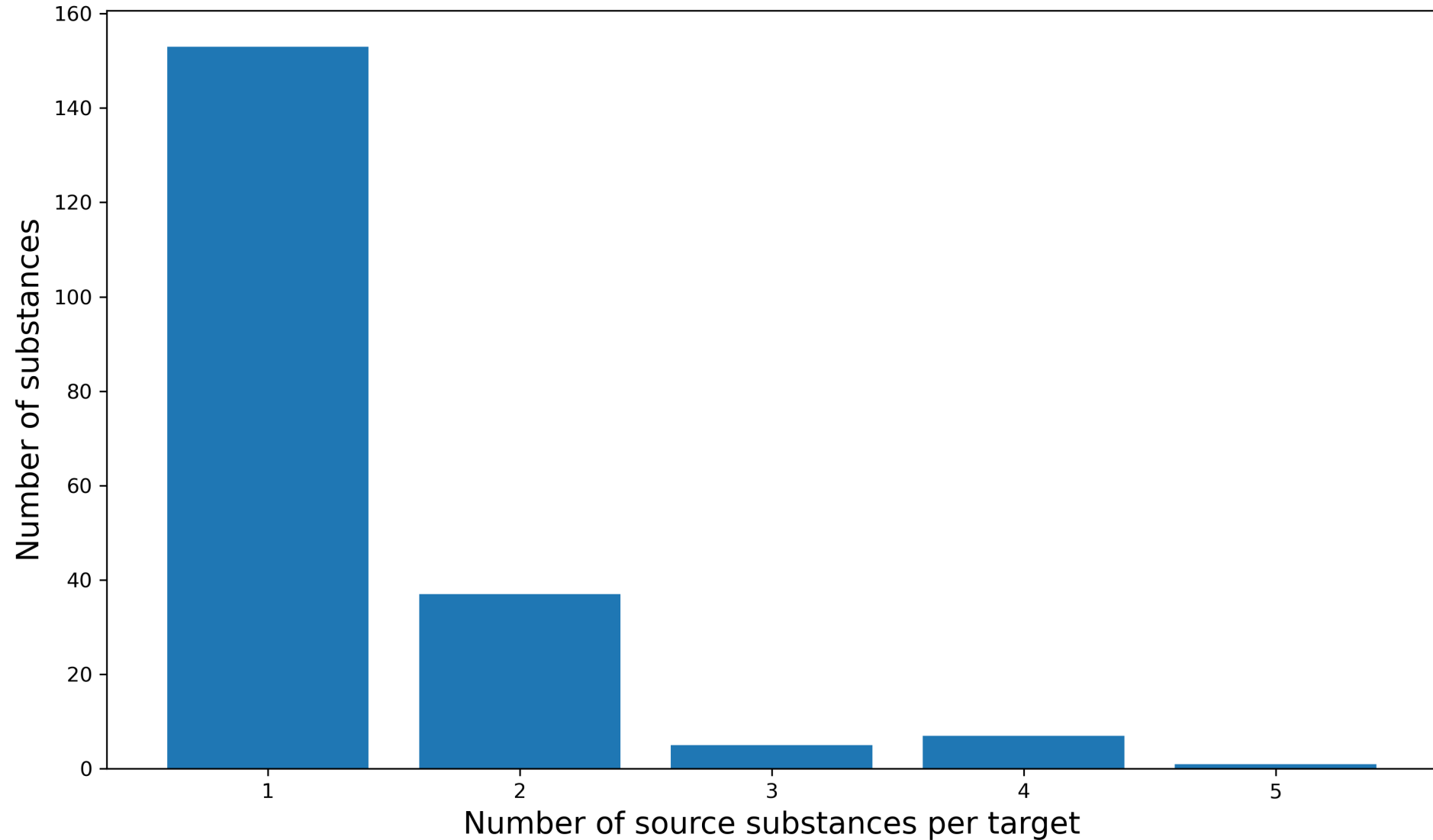
Similarity context evaluation

- **Structural similarity**
  - Derive Morgan chemical fingerprints (Morgan FP) and compute the pairwise Jaccard similarity.

- **Physicochemical similarity**
  - Estimate LogP, MW, number of Hydrogen donors and number of Hydrogen Acceptors.
  - Normalise based on the Lipinski Rule of 5 and calculate pairwise similarity using a Generalised Jaccard index.

- **Alert similarity**
  - Batch process the substances using default setting within Derek Nexus. Derive a binary fingerprint representation to reflect presence and absence of alerts for all substances. Compute pairwise Jaccard similarity on the basis of the alert fingerprint.

- **Metabolite similarity**
  - Generate predictions using the TIMES *in vitro* rat liver model for substances. Construct metabolic graphs and compute the Weisfeiler-Lehman Kernel as one measure of similarity, construct other measures of similarity using the transformation profile as a bit vector and the metabolites simulated as a third representation of metabolism information (Boyce et al 2022).

Structural Similarity relative to the Target

- Substances were characterised by Morgan chemical fingerprints and the pairwise similarities between the Target and Source analogues were computed
- For those analogue sets where there was more than 1 source analogue per target, a large variation is observed in the structural similarity.
- Source substances were not particularly structurally similar.
  - Likely challenge of identifying source substances with relevant data?
  - The which extent structural similarity is a determining factor in the analogue selection?
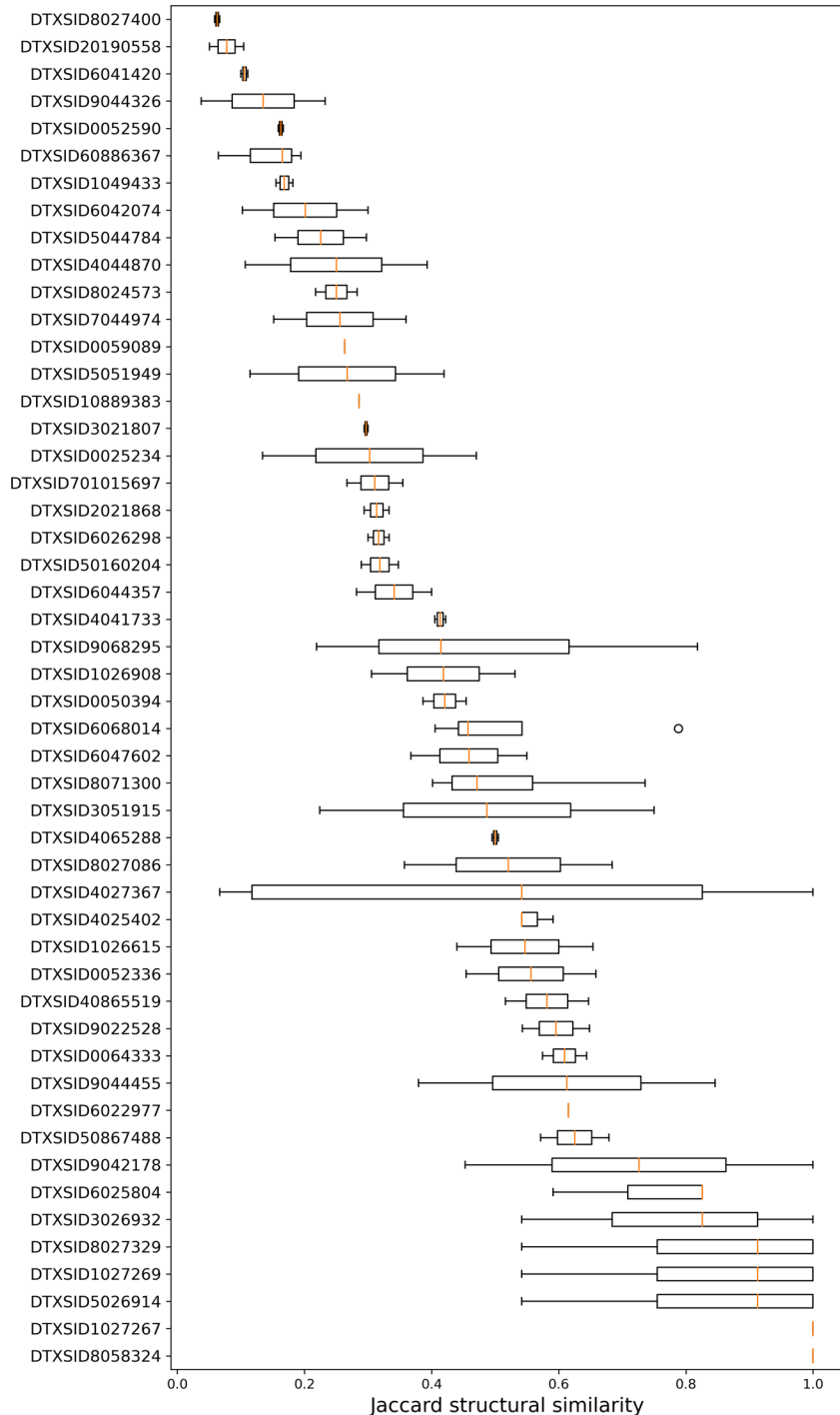
Example target substance with large variations in source structural similarity

- Target substance: 1-Decene

- Source substances ranged from 1-hexene (0.54), 1-octene (0.82) to 1-tetradecene and 1-octadecene (both 1).

- Target substance: 1-Tetradecene

- Some source substances appear plausible 1-hexene, 1-octadecene, 1-octene but 2 seem erroneous: 2-pentanone oxime, 1-[3-(Dimethylamino)propyl]urea! The latter have very low structural similarities 0.11 and 0.06 respectively.

Variation of pairwise similarities across source analogues
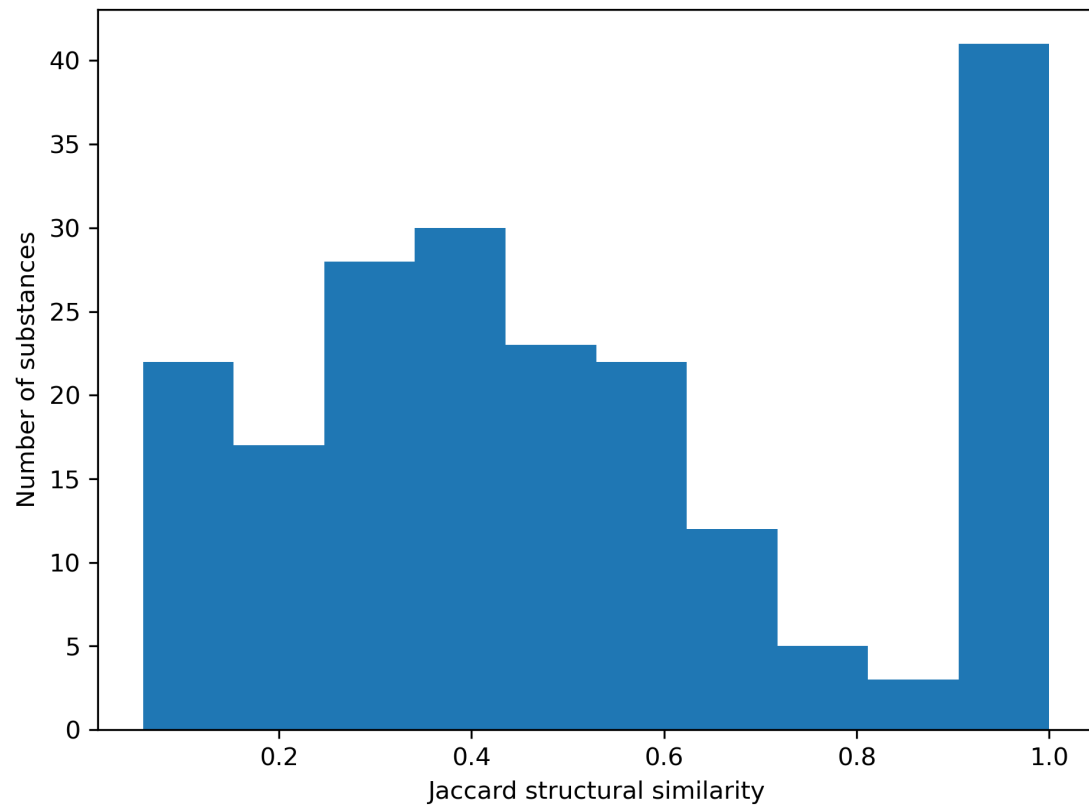
- Distribution in pairwise structural similarities based on Morgan FPs shows a large number of source substances with low pairwise similarities relative to their associated targets

Variation of pairwise physchem similarities across source analogues

- For targets with > 1 source analogue – the distribution in physchemical similarities based on Lipinski parameters
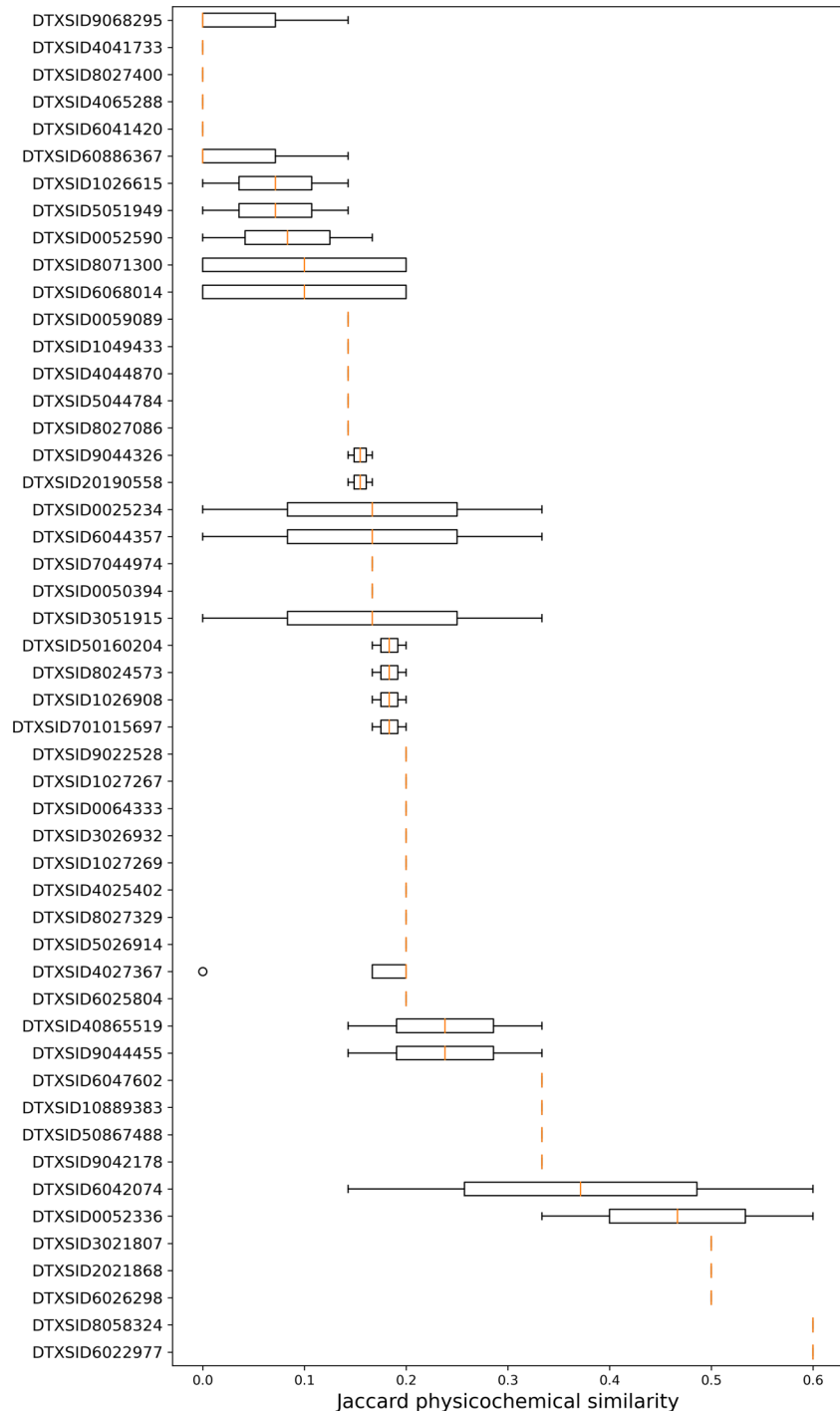
Variation of pairwise physchem similarities across source analogues
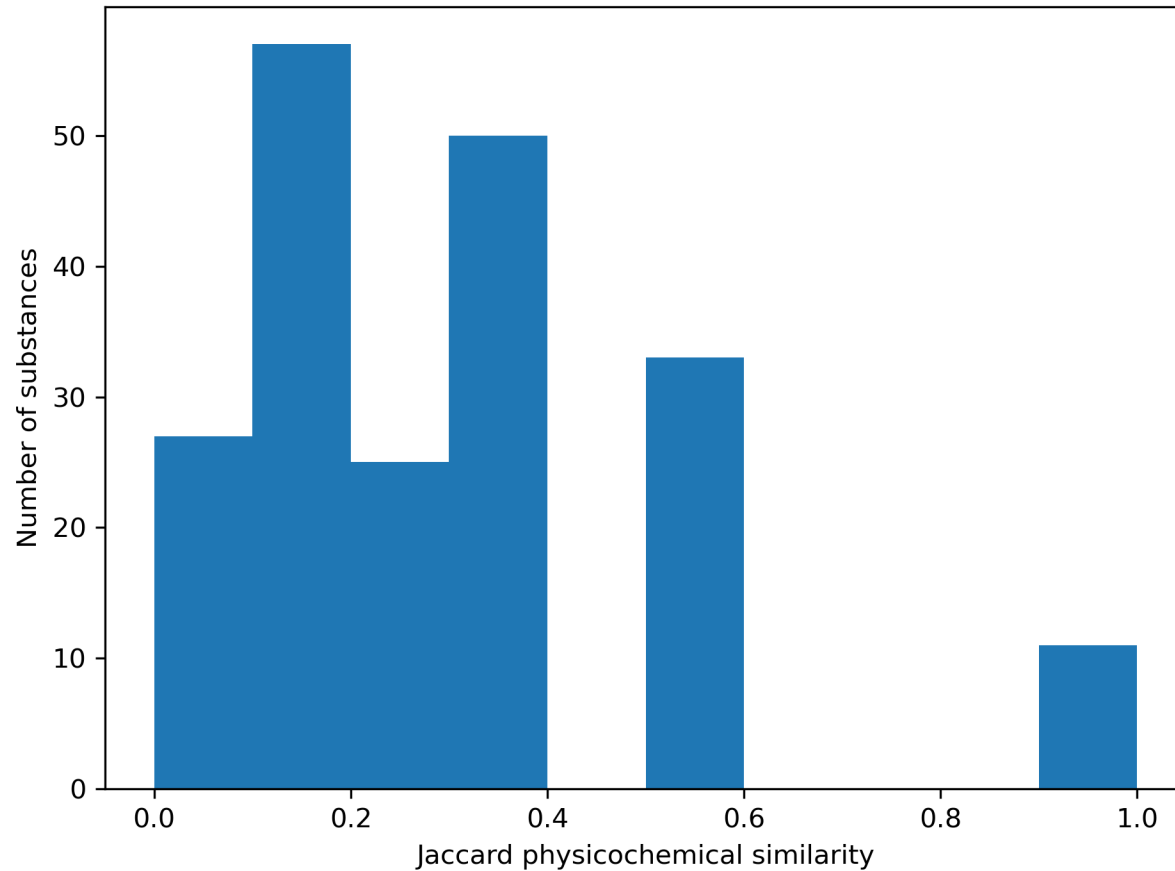
- Distribution in pairwise physicochemical similarities based on Lipinski parameters
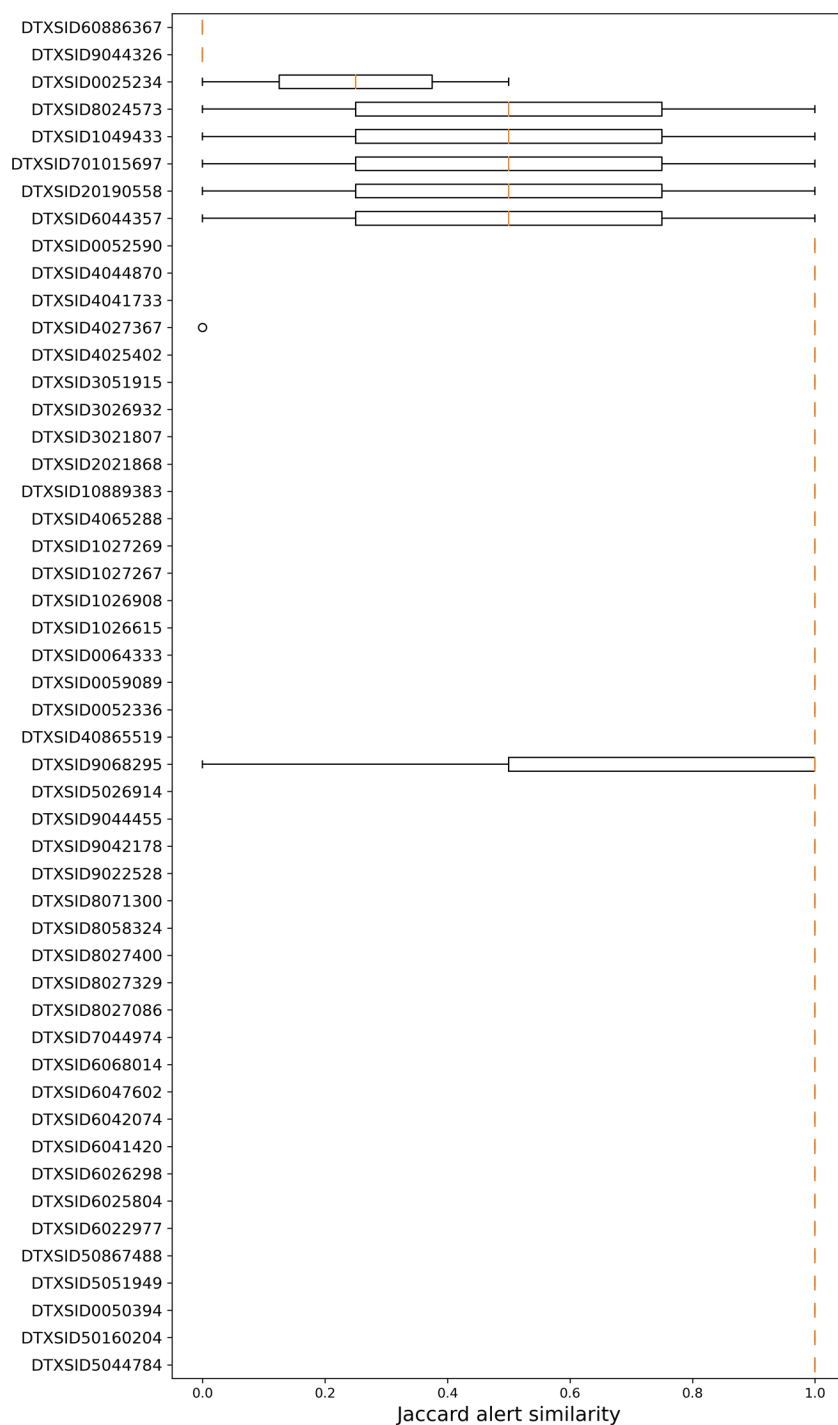
Variation of pairwise alert similarities across source analogues

- For those targets with > 1 source analogue – distribution in Derek alert profiles

- Highlights the sparsity of the number of alerts for the set of substances

Variation of pairwise alert similarities across source analogues
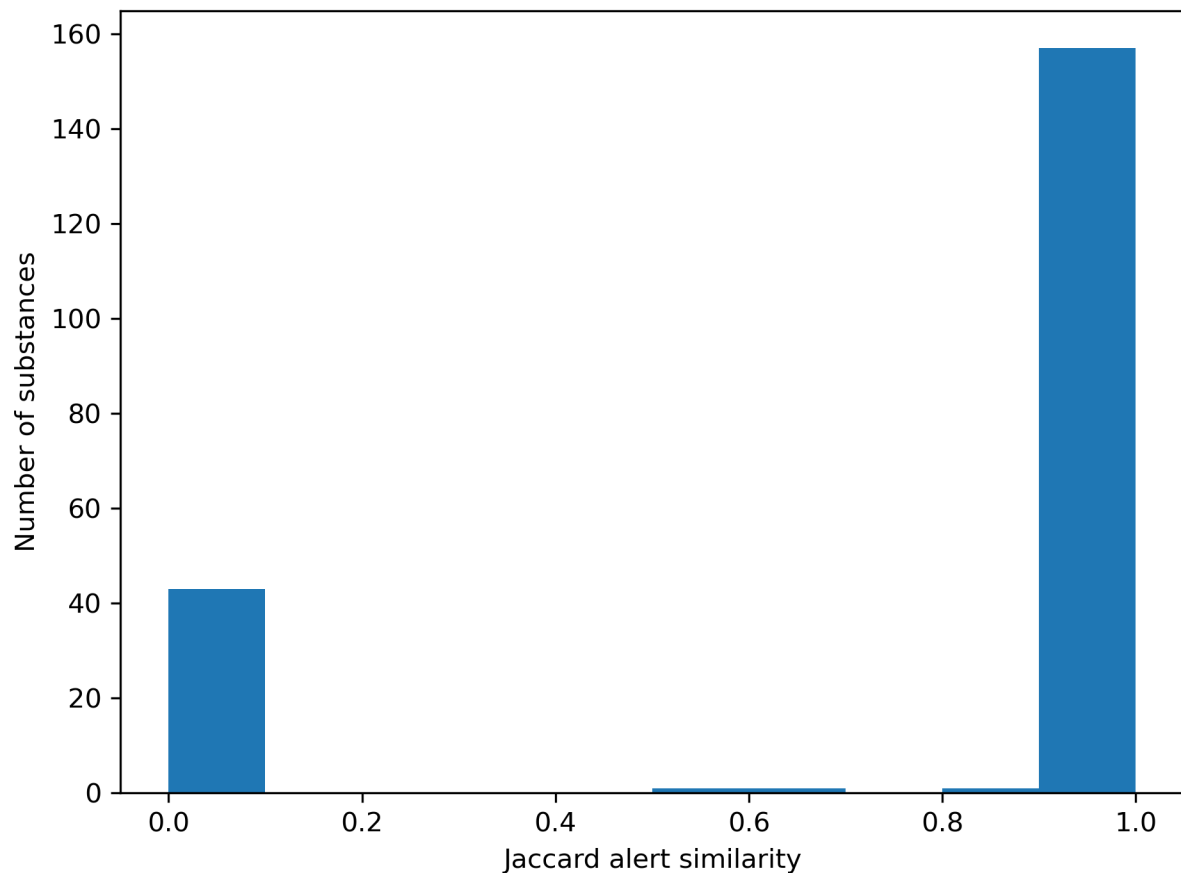
- **Distribution in Derek alert profiles when represented as a bit vector**

Number of alerts between source & target substances
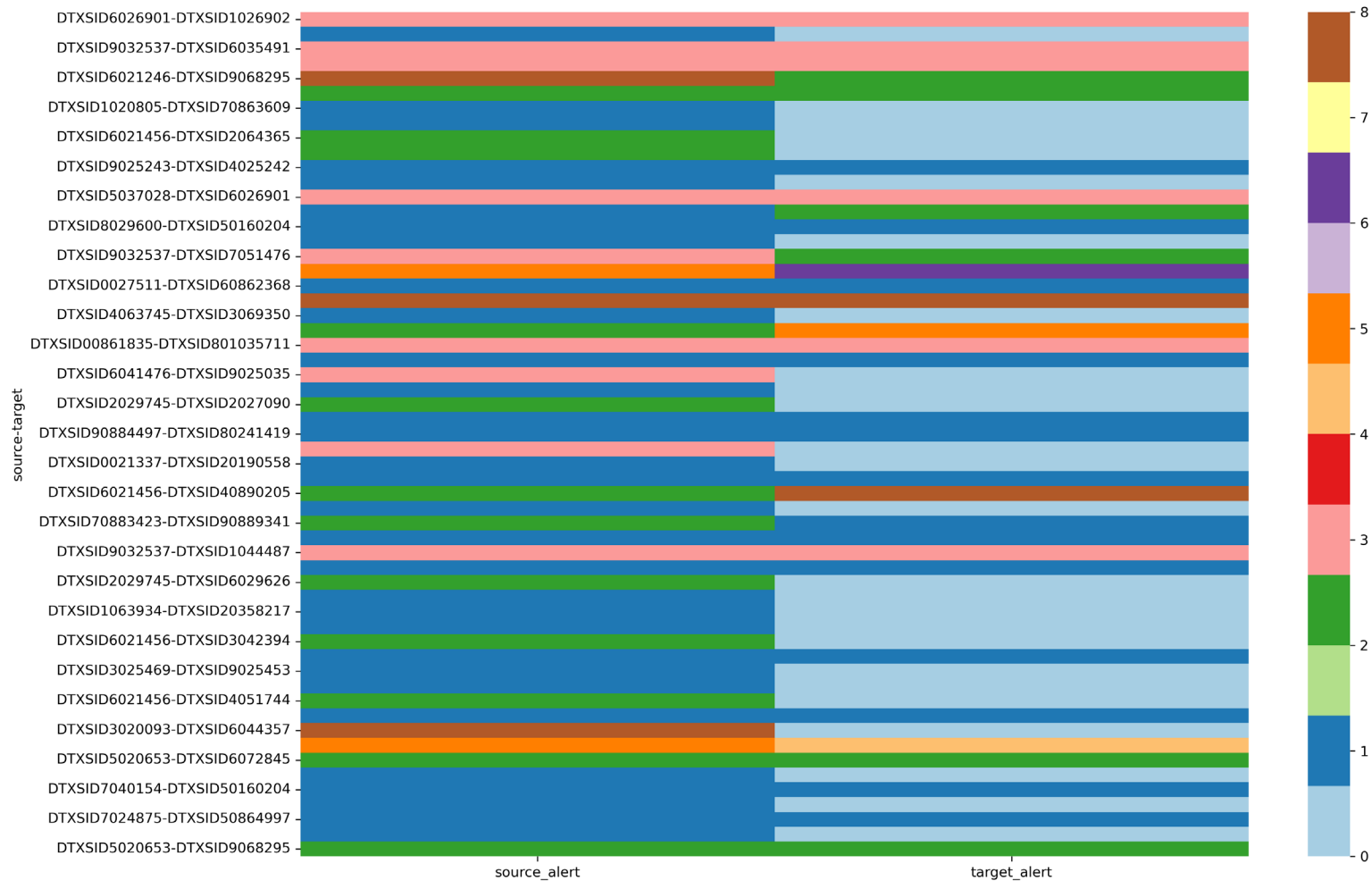
Of the target-source substances, majority flagged no alerts

Only 57 pairings were associated with alerts with the target substances typically flagging no alerts relative to the source substances or fewer alerts by count

Metabolic similarity

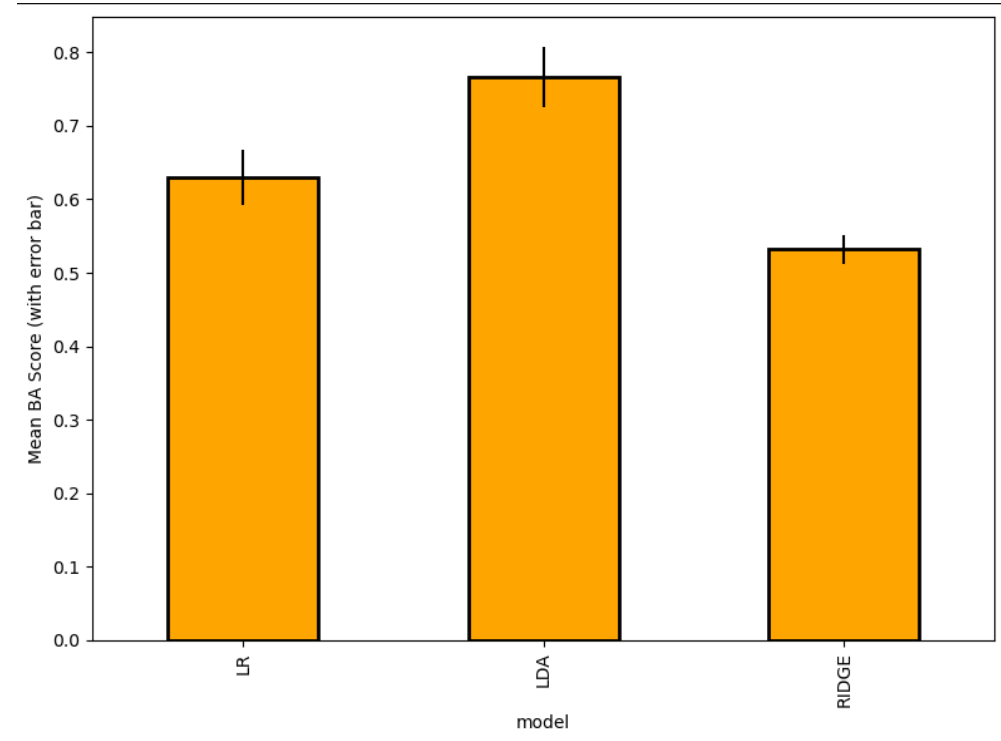- WL, transformation similarity, similarity in simulated metabolites
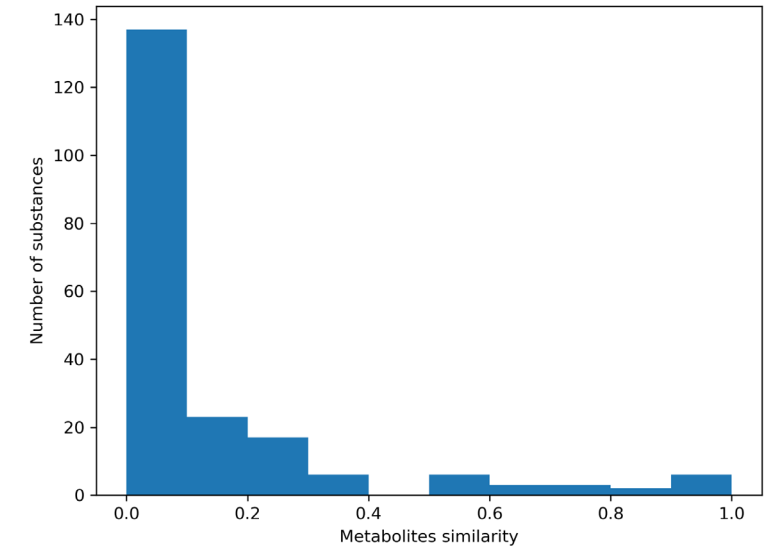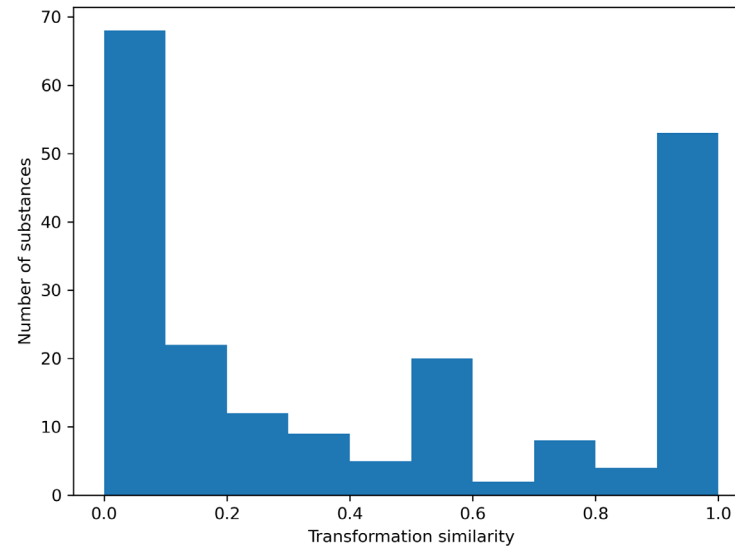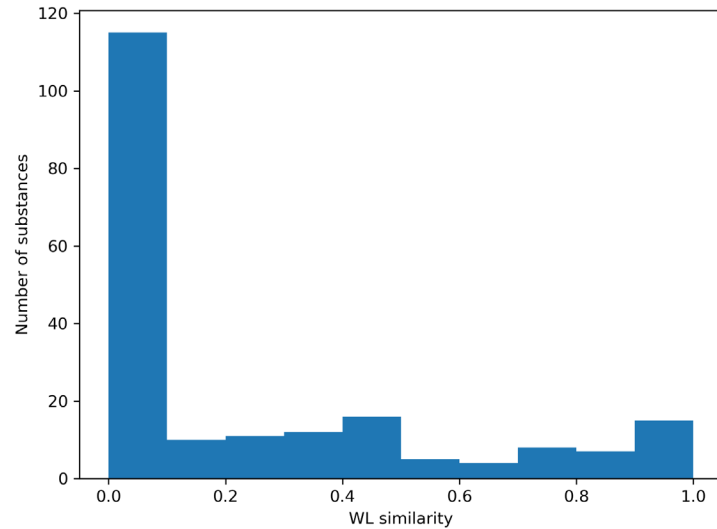
Quantifying the contribution of each similarity context

- Constructed a matrix for all targets and source substance combinations with their different similarity metrics as descriptors

- All actual target-source pairings were labelled as '1' and all other combinations as '0'

- Three different machine learning models (Logistic Regression, Ridge Regression and Linear Discriminant Analysis) were attempted to relate the similarity metrics to the labels. Models were trained to optimise for balanced accuracy

- Linear Discriminant Analysis gave rise to the best 10-fold stratified CV Balanced Accuracy (BA) (mean CV BA 0.77)

- Structural similarity, similarity in metabolites simulated were the most important features in the model

Quantifying the contribution of each similarity context

Comparison of mean CV balanced accuracy between the 3 models attempted

- Coefficients in the LDA
- 62.46 * structural similarity
- + 0.01 * alert similarity
- + 136.13 * metabolites similarity
- + 11.86 * transformation similarity
- + 7.68 * WL similarity
- + 4.33 * physicochemical similarity

Investigate the utility of GenRA to make toxicity predictions

- Derive a baseline model for toxicity predictions using GenRA

- This would provide a basis for comparison

- Searched the EPA Toxicity Values DB (ToxValDB v9.4) for all studies conducted by the oral route for which a NOAE(C)L, LOAE(C)L was available and where the units were in mg/kg-day

- 99406 studies were available for 7635 substances

- As a conservative approach, the 10th percentile of all studies on a per substance basis was computed irrespective of study type or POD type

- Each POD was then divided by the MW of the substance and the -log10 was calculated. This represented the modelled endpoint.

Investigate the utility of GenRA to make toxicity predictions

- Distribution of transformed POD values
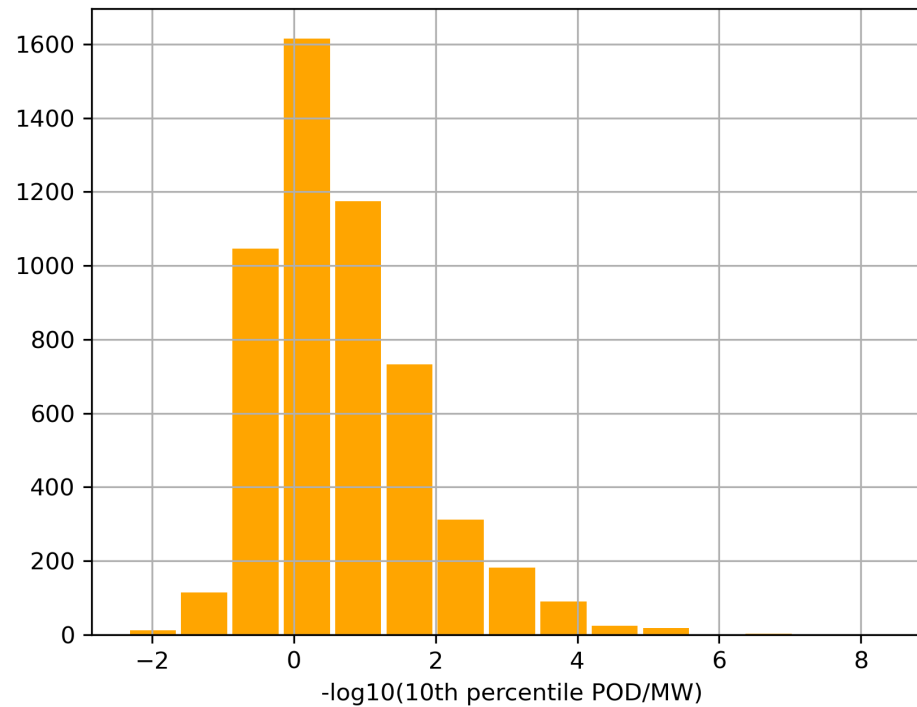
Investigate the utility of GenRA to make toxicity predictions

- A 5-fold CV approach was used with GenRA on the entire dataset to determine the optimal number of neighbours and similarity metric.

- Morgan Fingerprints were used as chemical fingerprint inputs.

- The Best CV R2 score was 0.383 with 6 neighbours

- Using a LOO approach, the GenRA model was applied to the entire dataset to predict the toxicity values of all chemicals.

Investigate the utility of GenRA to make toxicity predictions

- ## The R2 on the full dataset was 0.424, RMSE 0.85
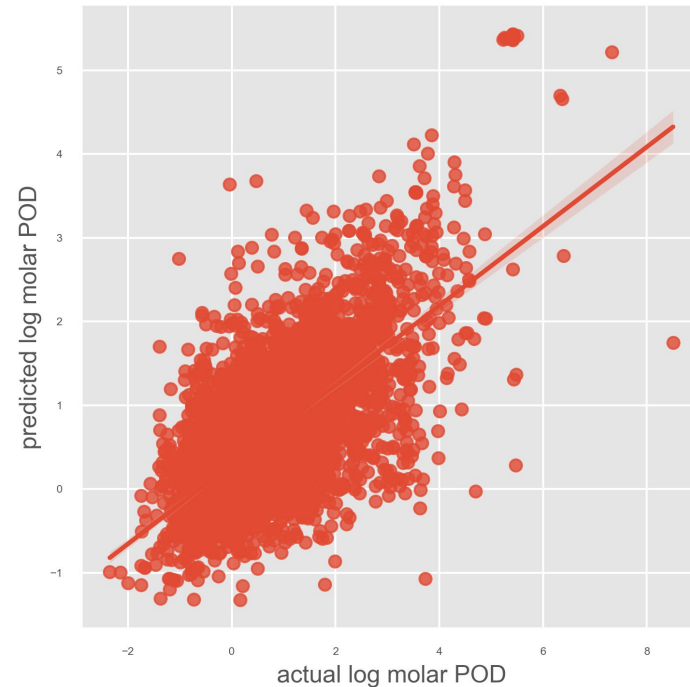
Given the variability in:
- replicate repeat dose studies (RMSE of approx. 0.4-0.6) (Pham *et al.*, 2020), and
- QSAR predictions of repeat dose toxicity (RMSE of 0.7-0.8 for external test set) (Pradeep *et al.*, 2020),

the RMSE reported here for GenRA toxicity predictions based on neighbours seem reasonable.

Investigate the utility of GenRA to make toxicity predictions

- Using 6 nearest neighbours, GenRA predictions were then made for each of the target substances in the ECHA REACH pairs dataset

- RDT oral data was then extracted for the source analogues that had been used to read-across for the target substances

Investigate the utility of GenRA to make toxicity predictions

- Variability in the experimental point of departure values for the source analogues themselves that were reported in the REACH dossiers
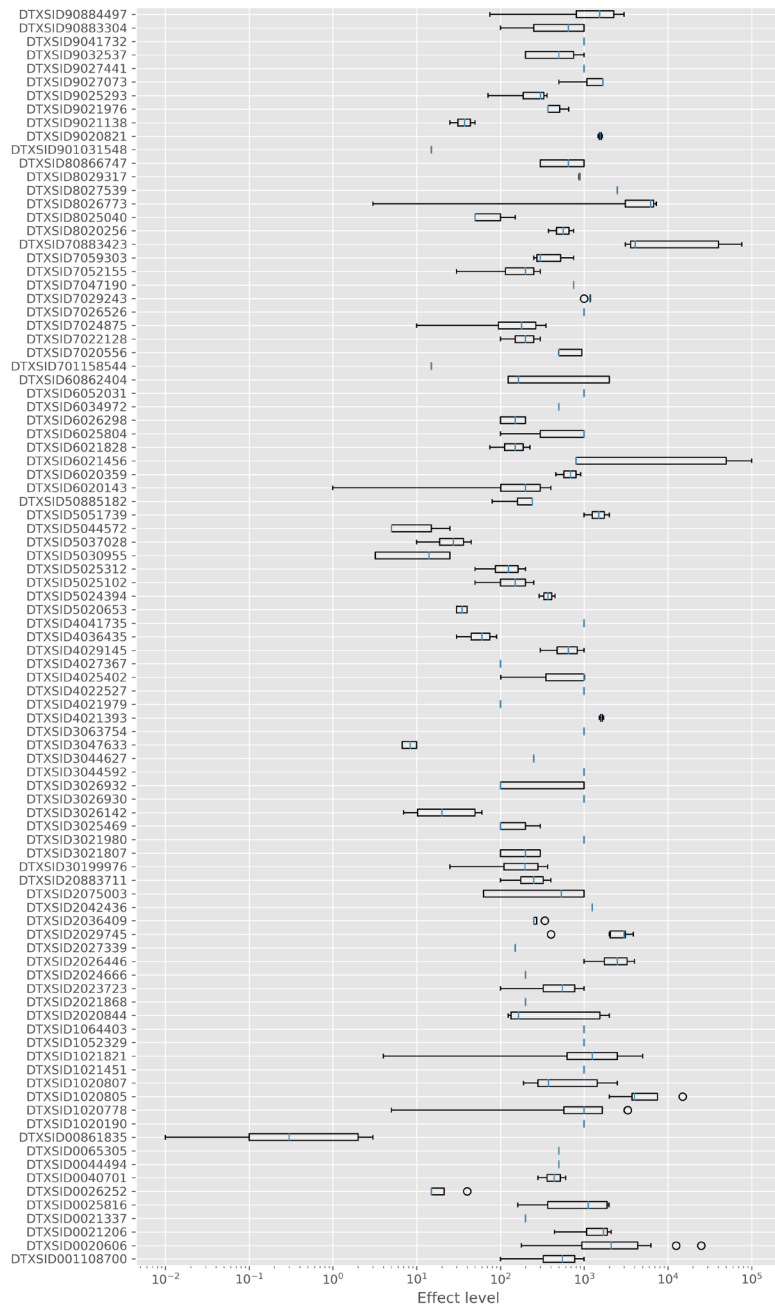
Investigate the utility of GenRA to make toxicity predictions

- The variation of the experimental data read across vs the GenRA predictions was explored for those targets with more than 1 source analogue

- Often the GenRA prediction was more conservative
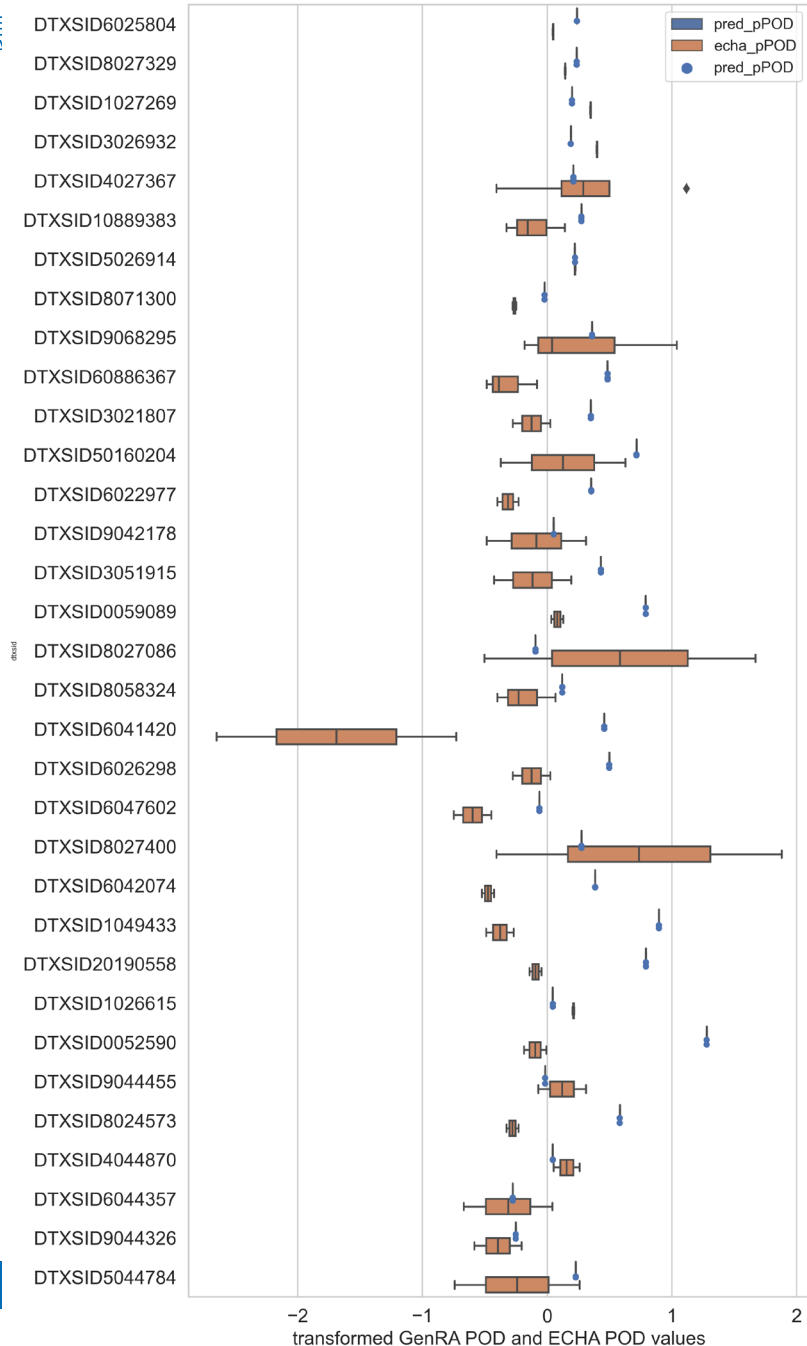
Summary

- Using IUCLID, the REACH Study Results, and the structural information in DSSTox it was possible to derive a large set of target/source read-across associations that were amenable for systematic analysis

- A significant percentage of target/source pairs appeared quite different when evaluating their pairwise similarities namely their structural similarity, physicochemical property similarity, alert similarity and metabolic similarity

- An attempt was made to quantify the contribution that each similarity context played by deriving a model that related the different similarities to the target-source pairs – structural similarity and similarity in the metabolites themselves played the largest roles in rationalising the source analogues.

Summary

- A model to predict the 10th percentile of point of departure outcomes from oral studies extracted from the Toxicity Values database (ToxValDB) was then undertaken to create a baseline model. The R2 of the model derived for the training set was 0.4.

- Predictions were made for the REACH target substances using GenRA. Often these predictions gave rise to more conservative points of departure relative to the ones reported in the registration dossiers.

- This dataset provided a means of evaluating & quantifying the uncertainties in read-across predictions at scale.

- Further work will consider the impact of bioactivity similarity and refining the GenRA POD model to consider different aggregations based on study type and point of departure.