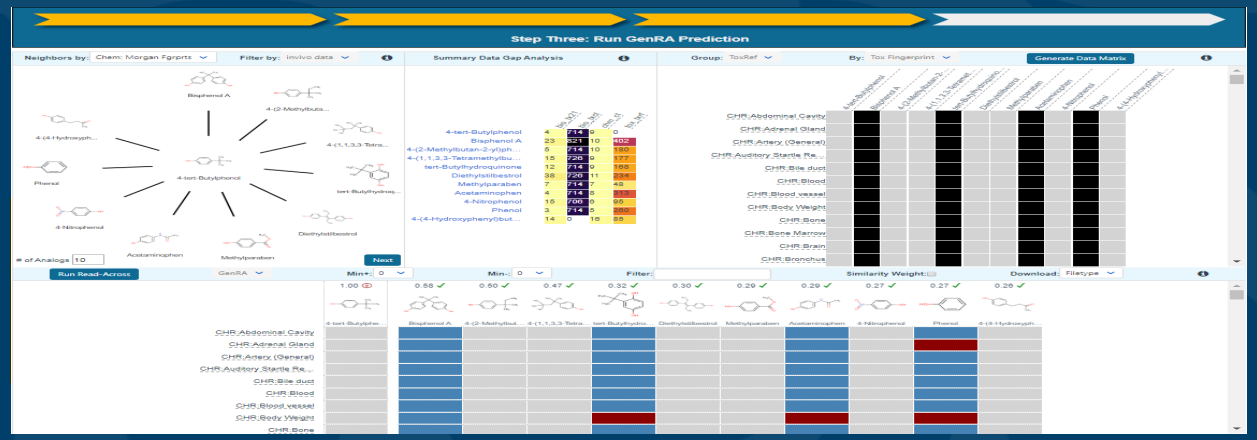
EPA United States Environmental Protection Agency

Transitioning Generalised Read-across (GenRA) towards quantitative predictions



George Helman, Imran Shah, Grace Patlewicz
National Center for Computational Toxicology (NCCT), US EPA

Conflict of Interest Statement

No conflict of interest declared.

Disclaimer:

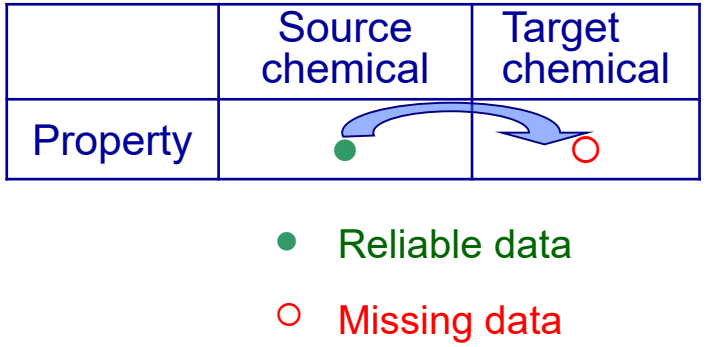
The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

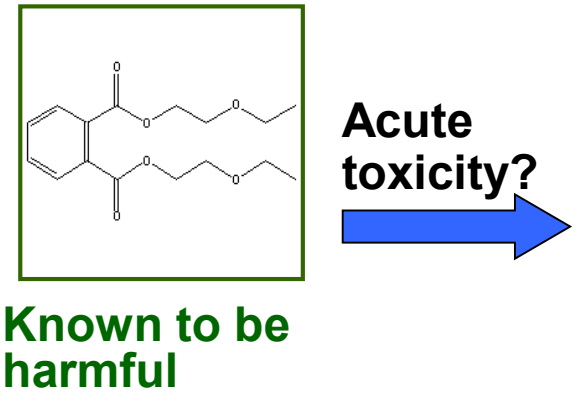
Outline

- Putting Read-across in context
- Overview of the Generalised Read-across (GenRA) approach
- Transitioning from 'qualitative' to 'quantitative' predictions
  - LD50 values from acute oral rodent toxicity studies
  - LOAEL values from repeated dose toxicity studies
  - Evaluation of predictions
- Summary Remarks
- Acknowledgements

Definitions: 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.
- 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.





A harmonised hybrid read-across workflow

United States Environmental Protection Agency

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com

Journal Cover Image

Navigating through the minefield of read-across frameworks: A commentary perspective

Grace Patlewicz[a, *], Mark T.D. Cronin[b], George Helman[a, c], Jason C. Lambert[d], Lucina E. Lizarraga[d], Imran Shah[a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency (US EPA), 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA
[b] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
[c] Oak Ridge Institute for Science and Education (ORISE), 1299 Bethel Valley Road, Oak Ridge, TN 37830, USA
[d] National Center for Evaluation Assessment (NCEA), US Environmental Protection Agency (US EPA), 26 West Martin Luther King Dr, Cincinnati, OH 45268, USA

A harmonised hybrid read-across workflow

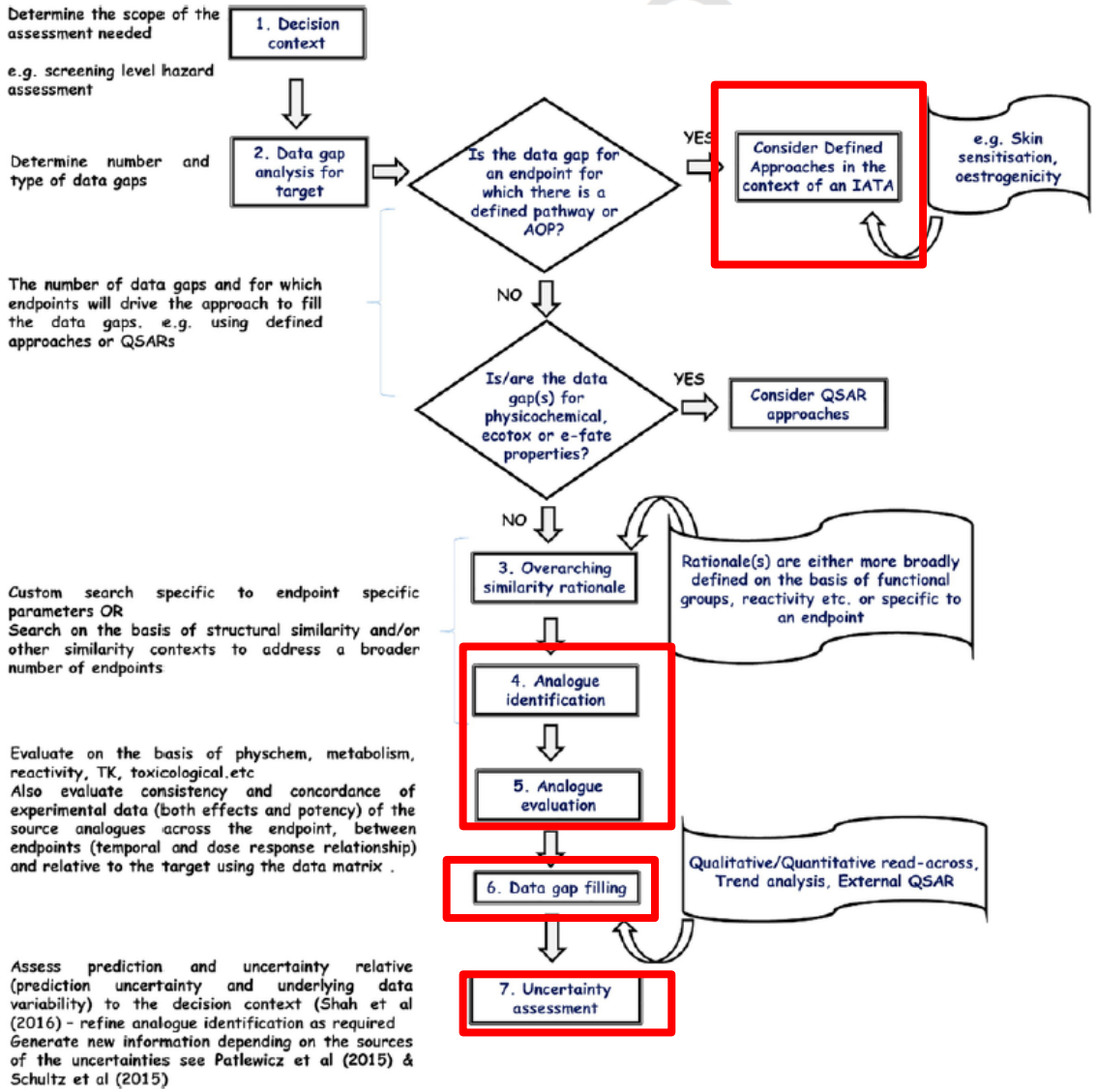


Fig. 9. A harmonised hybrid development and assessment framework.

Where do other NAM fit?
How should we transition to data-driven approaches?
What about characterising the uncertainty of the predictions made?

Patlewicz et al., 2018

Selected read-across tools

Computational Toxicology 3 (2017) 1–18

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox

Navigating through the minefield of read-across tools: A review of in silico tools for grouping

Grace Patlewicz [a,*], George Helman [a,b], Prachi Pradeep [a,b], Imran Shah [a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA
[b] Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA

ARTICLE INFO

Article history:
Received 29 March 2017
Received in revised form 22 May 2017
Accepted 25 May 2017
Available online 29 May 2017

Keywords:
Category approach
Analogue approach
Data gap filling
Read-across
(Q)SAR
Trend analysis
Nearest neighb

ABSTRACT

Read-across is a popular data gap filling technique used within analogue and category approaches for regulatory purposes. In recent years there have been many efforts focused on the challenges involved in read-across development, its scientific justification and documentation. Tools have also been developed to facilitate read-across development and application. Here, we describe a number of publicly available read-across tools in the context of the category/analogue workflow and review their respective capabilities, strengths and weaknesses. No single tool addresses all aspects of the workflow. We highlight how the different tools complement each other and some of the opportunities for their further development to address the continued evolution of read-across.

Published by Elsevier B.V.

Selected read-across tools

Tool	AIM	ToxMatch	AMBIT	OECD Toolbox	CBRA	ToxRead	GenRA
Analogue identification	X	X	X	X	X	X	X
Analogue Evaluation	NA	X	X by other tools available	X	X	X For Ames & BCF	NA
Data gap analysis	NA	X	X Data matrix can be exported	X Data matrix viewable	NA	NA	X Data matrix can be exported
Data gap filling	NA	X	User driven	X	X	X	X
Uncertainty assessment	NA	NA	NA	X	NA	NA	X
Availability	Free	Free	Free	Free	Free	Free	Free

GenRA (Generalised Read-Across)

- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and bioactivity descriptors (Shah et al, 2016)
- Generalised version of the Chemical-Biological Read-Across (CBRA) developed by Low et al (2013)
- Goal: To establish an objective performance baseline for read-across and quantify the uncertainty in the predictions made

Jaccard similarity:

GenRA v1.0 - Approach

United States Environmental Protection Agency

I. Data

1,778 Chemicals
3,239 Structure descriptors (chm)
820 Bioactivity hitcall (bio)
ToxCast

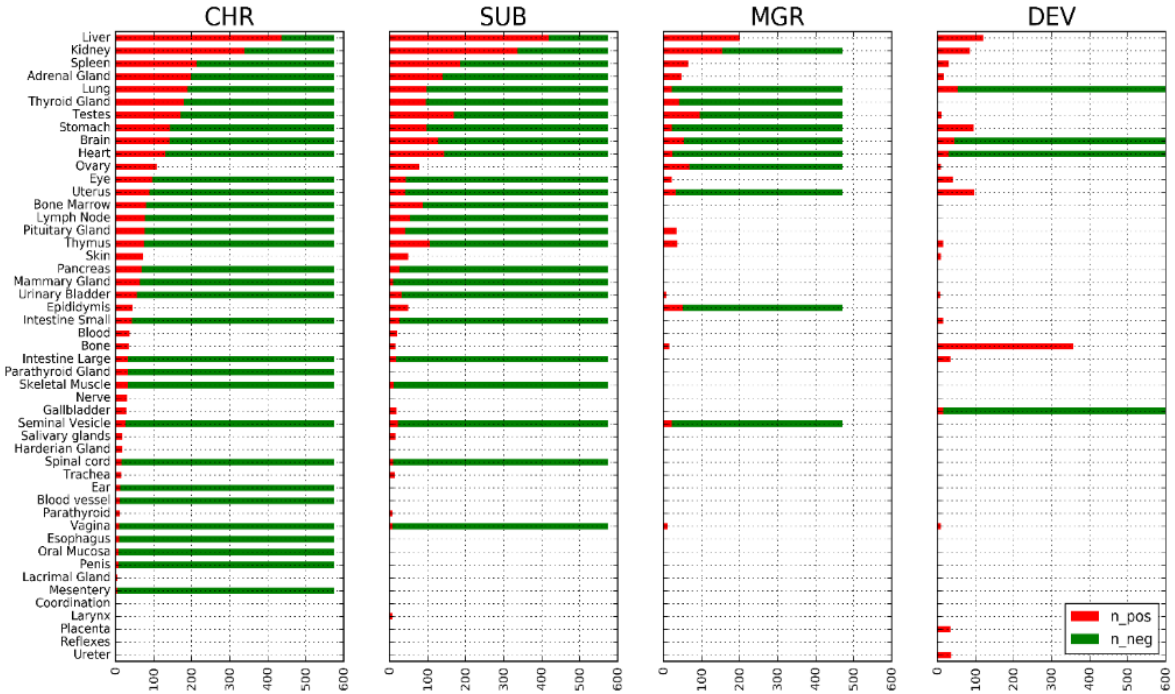
574 toxicity effects (tox)
ToxRefDB

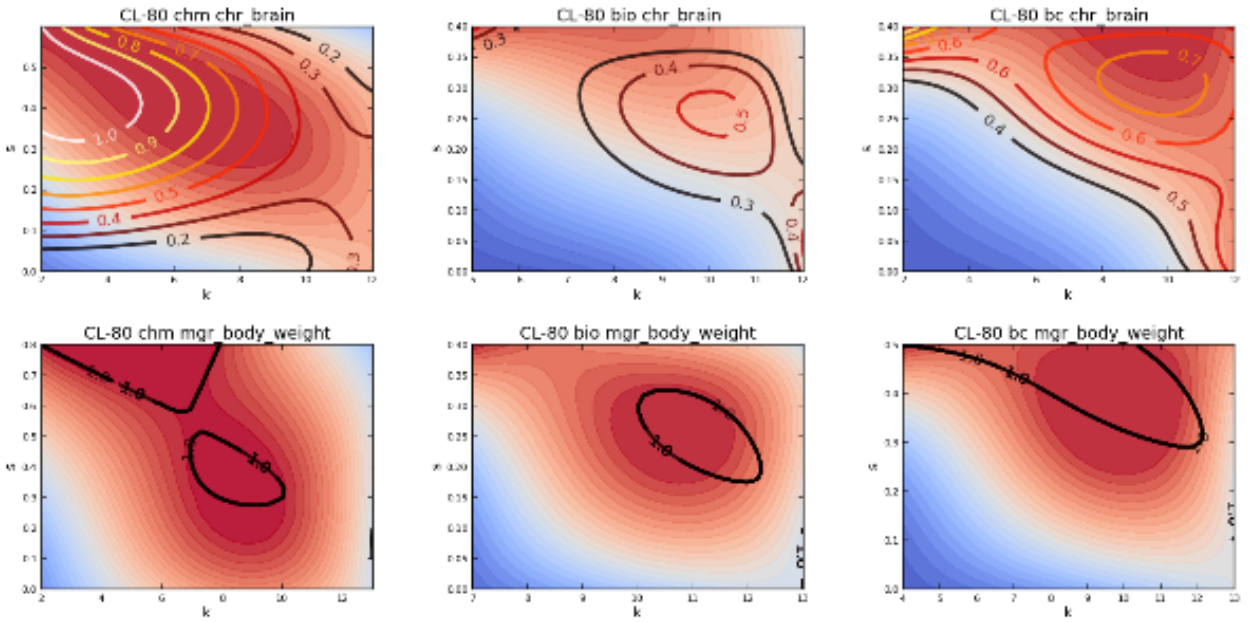
II. Define Local neighbourhoods

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

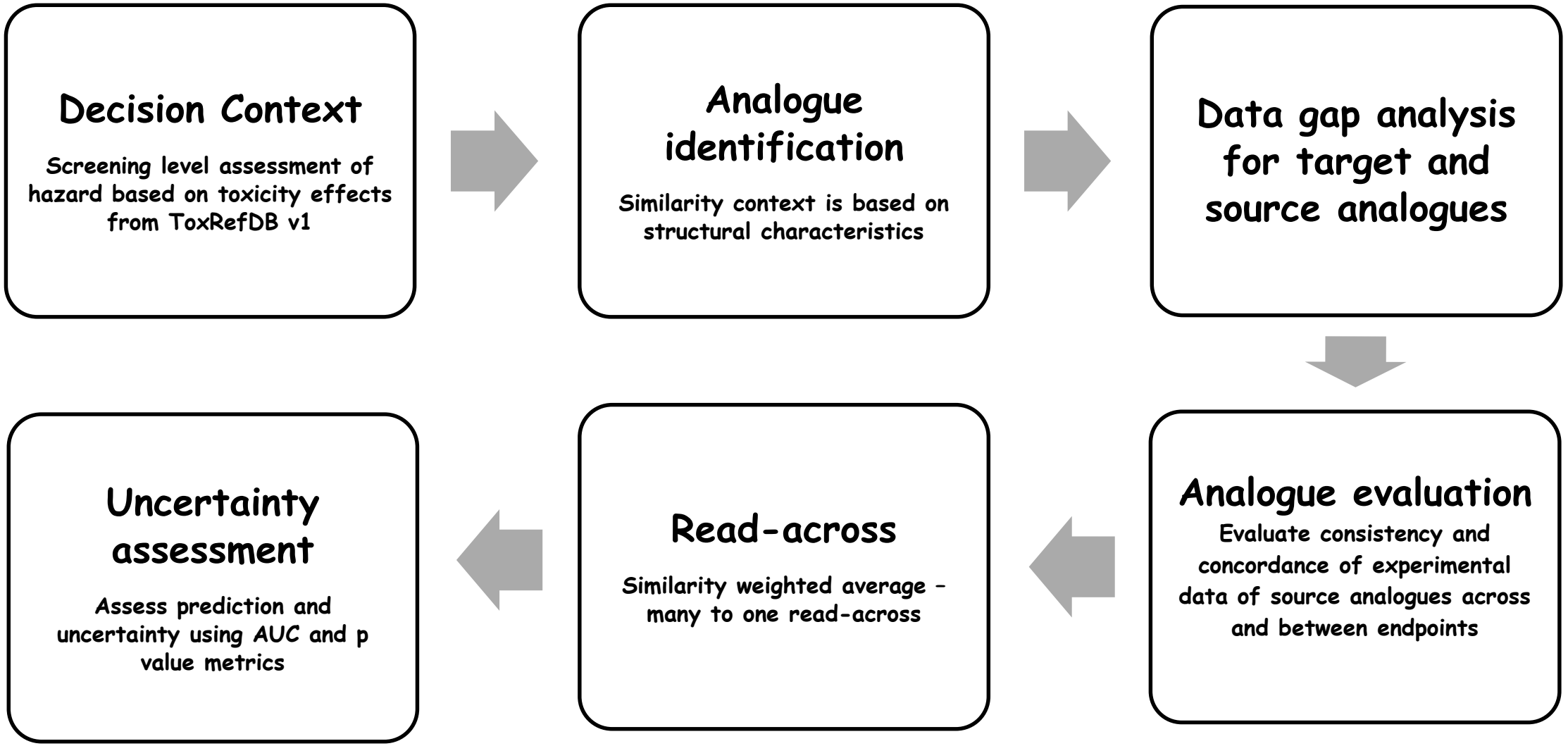
III. GenRA

Use GenRA to predict toxicity effects in local neighbourhoods
Evaluate impact of structural and/or bioactivity descriptors on prediction
Quantify uncertainty





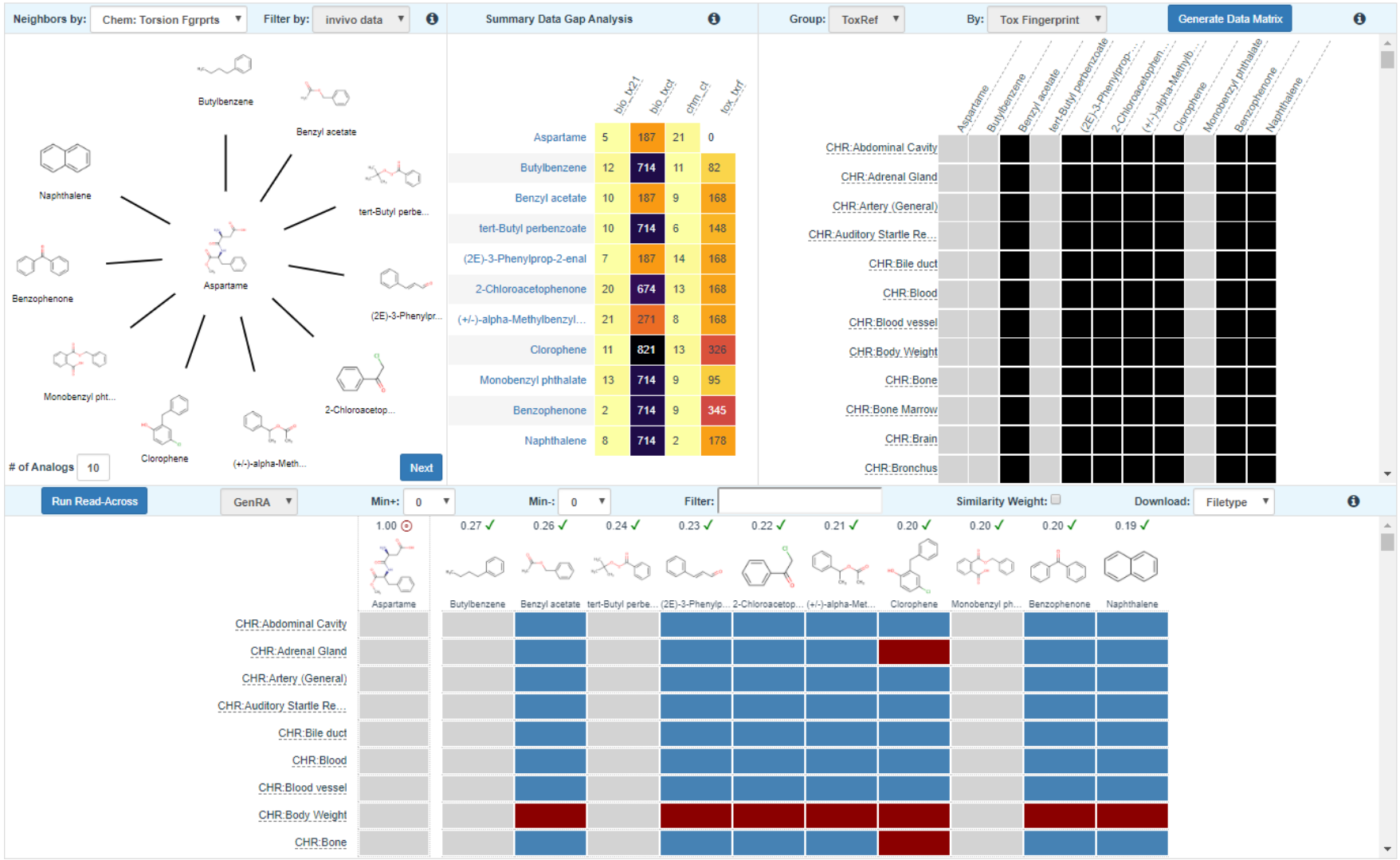
Read-across workflow in GenRA v1.0



Activity was translated into a binary score (1,0)

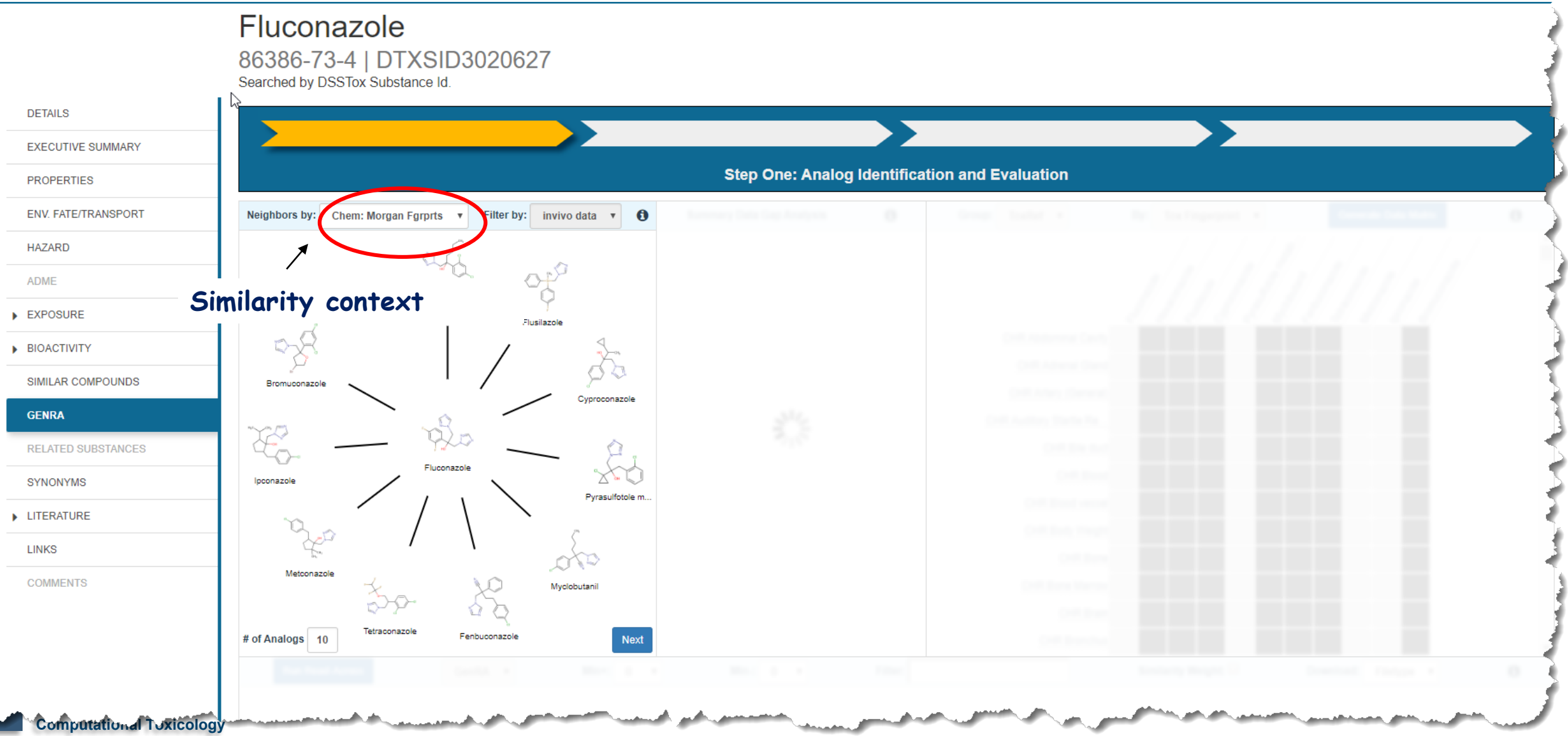
GenRA tool in reality

- Integrated into the EPA CompTox Chemicals dashboard

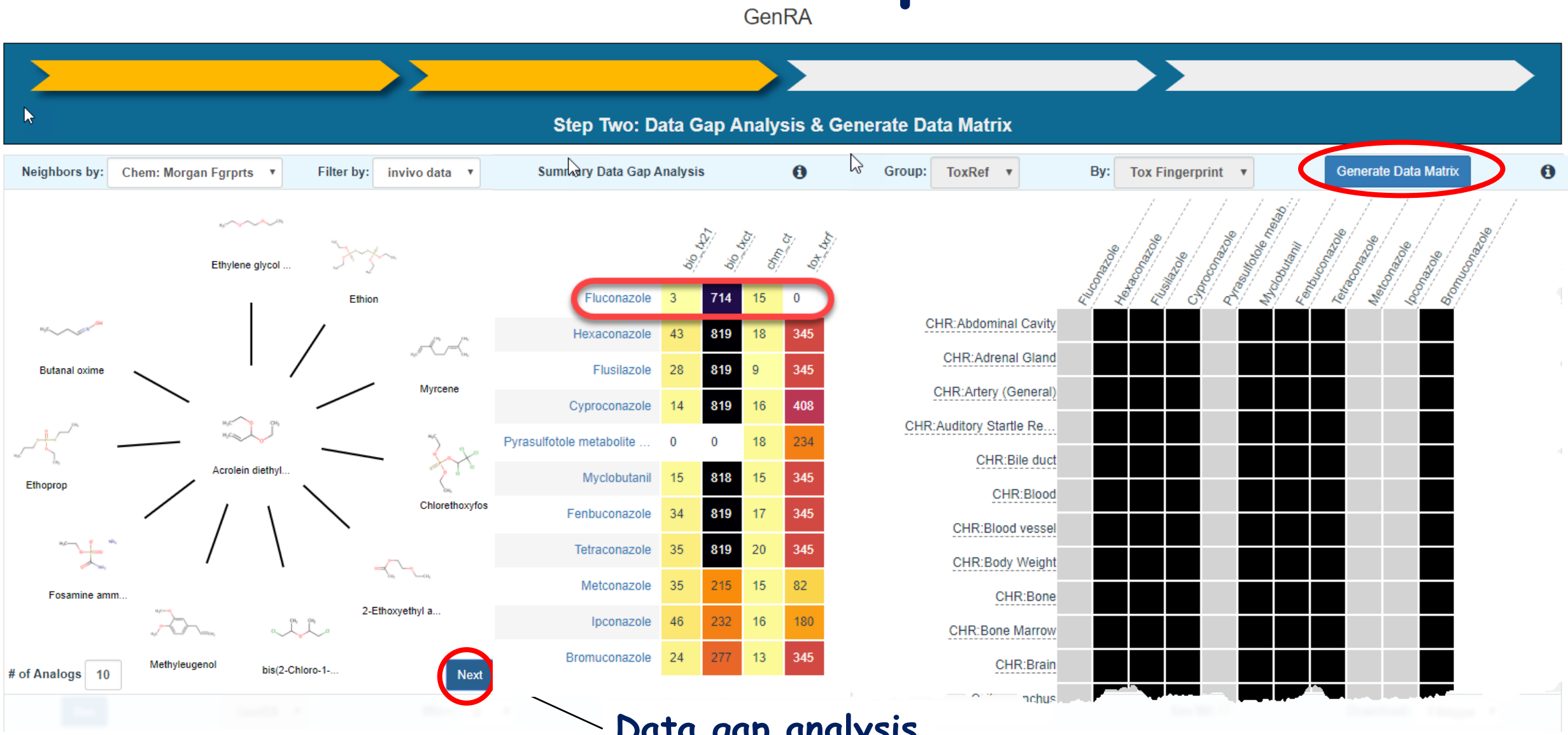


GenRA tool in practice

- Structured as a workflow

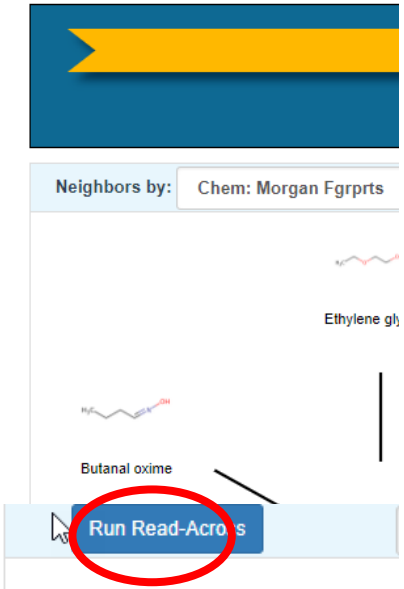


GenRA tool in practice



Data gap analysis

GenRA tool in practice



Run GenRA

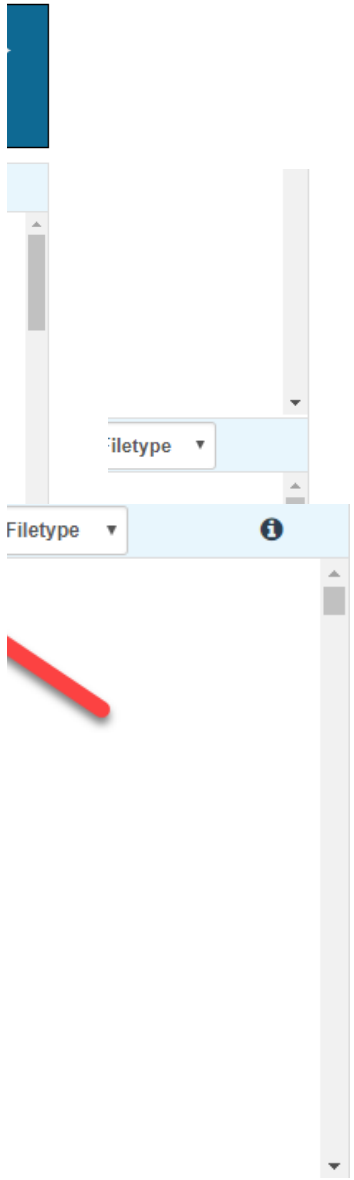
ALTEX preprint
published February 4, 2019
doi:10.14573/altex.1811292

Short Communication

Generalized Read-Across (GenRA): A workflow implemented into the EPA CompTox Chemicals Dashboard

George Helman[1,2], Imran Shah[2], Antony J. Williams[2], Jeff Edwards[2], Jeremy Dunne[2] and Grace Patlewicz[2]*

[1]Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA; [2]National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency, Research Triangle Park (RTP), NC, USA



Abstract

Generalized Read-Across (GenRA) is a data driven approach which makes read-across predictions on the basis of a similarity weighted activity of source analogues (nearest neighbors). GenRA has been described in more detail in the literature (Shah et al., 2016; Helman et al., 2018). Here we present its implementation within the EPA's CompTox Chemicals Dashboard to provide public access to a GenRA module structured as a read-across workflow. GenRA assists researchers in identifying source analogues, evaluating their validity and making predictions of *in vivo* toxicity effects for a target substance. Predictions are presented as binary outcomes reflecting presence or absence of toxicity together with quantitative measures of uncertainty. The approach allows users to identify analogues in different ways, quickly assess the availability of relevant *in vivo* data for those analogues and visualize these in a data matrix to evaluate the consistency and concordance of the available experimental data for those analogues before making a GenRA prediction. Predictions can be exported into a tab-separated value (TSV) or Excel file for additional review and analysis (e.g., doses of analogues associated with production of toxic effects). GenRA offers a new capability of making reproducible read-across predictions in an easy-to use-interface.

GenRA – Next Steps

- Ongoing research:

- Summarising and aggregating the toxicity effect predictions to guide end users – what effect predictions are we most confident about (digesting & interpreting the predictions more efficiently)

- Consideration of other information to define and refine the analogue selection & evaluation – e.g. physicochemical similarity, metabolic similarity, reactivity similarity, bioactivity similarity (transcriptomics similarity)…
  - EPA New Chemical Categories
  - Quantifying the impact of physicochemical similarity on read-across performance (Helman et al., 2018)

GenRA – Next Steps

- Dose response information to refine scope of prediction beyond binary outcomes
  - Transitioning from qualitative to quantitative predictions – how to apply and interpret GenRA in screening level hazard assessment
  - Starting with quantitative data – e.g. acute rat oral toxicity, ToxRefDB v2

Case study: Acute toxicity

- Transitioning GenRA to make quantitative predictions
- Investigated extending GenRA using the acute oral rat systemic toxicity data collected as part of the ICCVAM Acute toxicity workgroup
- NICEATM-NCCT effort to collate a large dataset of acute oral toxicity to evaluate the performance of existing predictive models and investigate the feasibility of developing new models

Acute toxicity: Dataset creation

Database Resource	Rows of Data (number of LD50 values)	Unique CAS
ECHA (ChemProp)	5533	2136
JRC AcutoxBase	637	138
NLM HSDB	4082	2238
OECD (eChemPortal)	10206	2314
PAI (NICEATM)	364	293
TEST (NLM ChemIDplus)	13689	13545

Rat oral LD50s:
16,297 chemicals total
34,508 LD50 values

Require unique LD50 values with mg/kg units

15,688 chemicals total
21,200 LD50 values

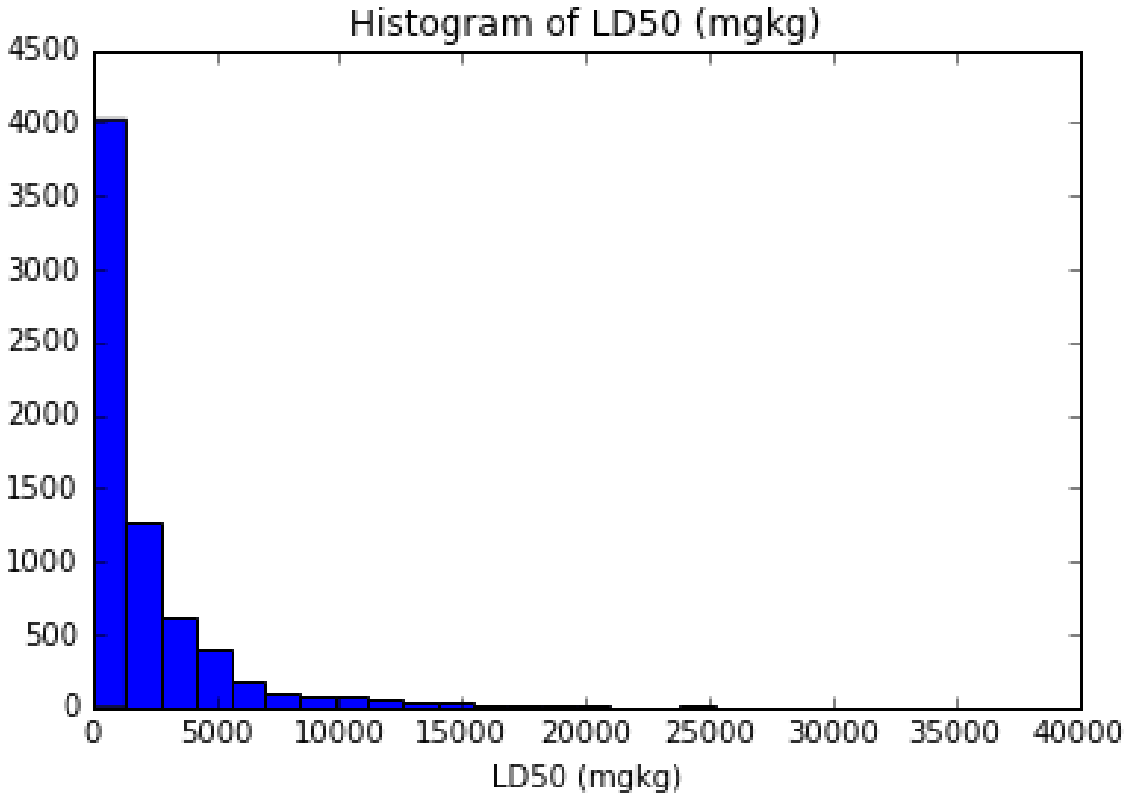
Preprocessing for modelling

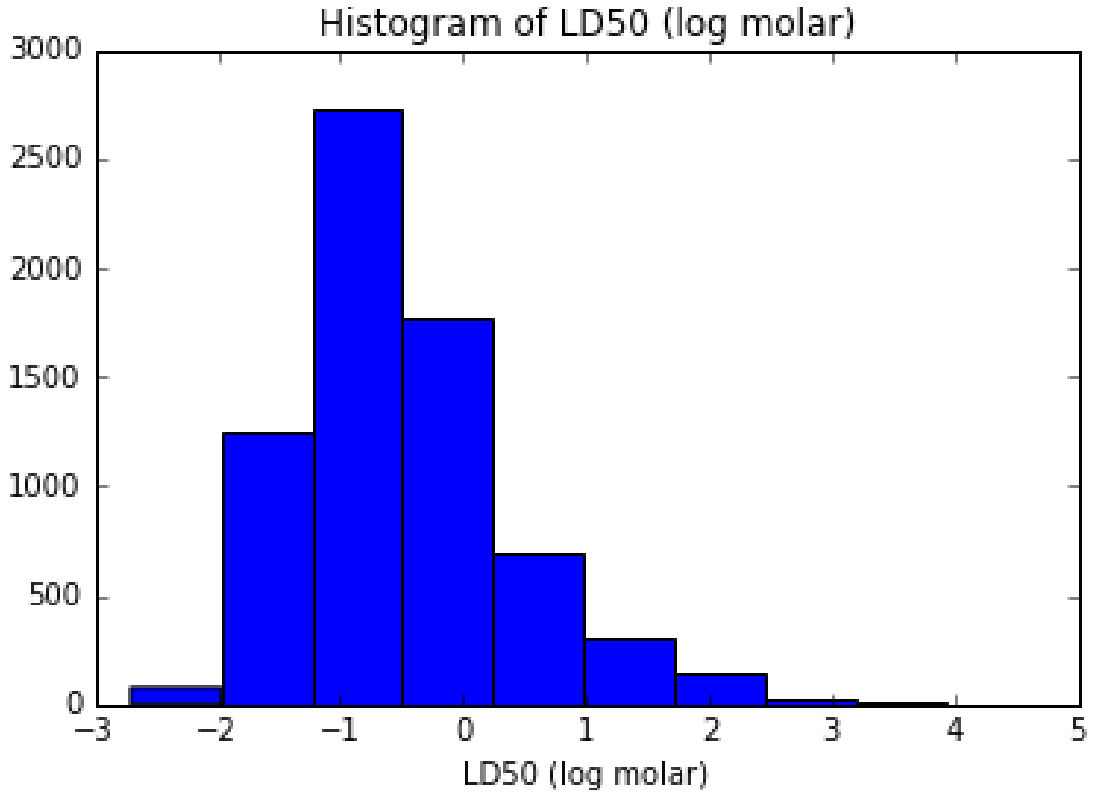
11,992 chemicals
16,173 LD50 values

Karmaus et al, 2018; Kleinstreuer et al., 2018

Exploratory Data Analysis

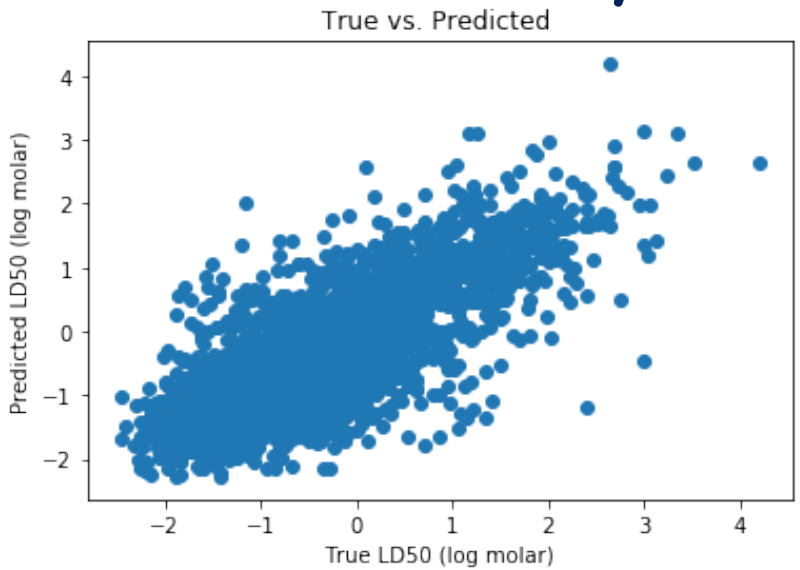
- Found DSSTox matches for 7011 substances
- Extracted MW values

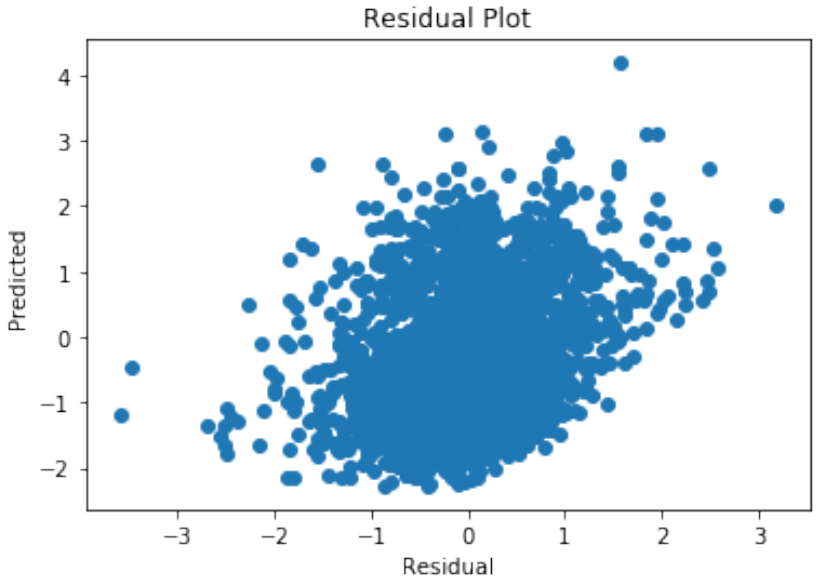




GenRA approach : Overall 'global' performance

- Search for a maximum of 10 nearest neighbours on entire dataset
- Use a min similarity threshold of 0.5

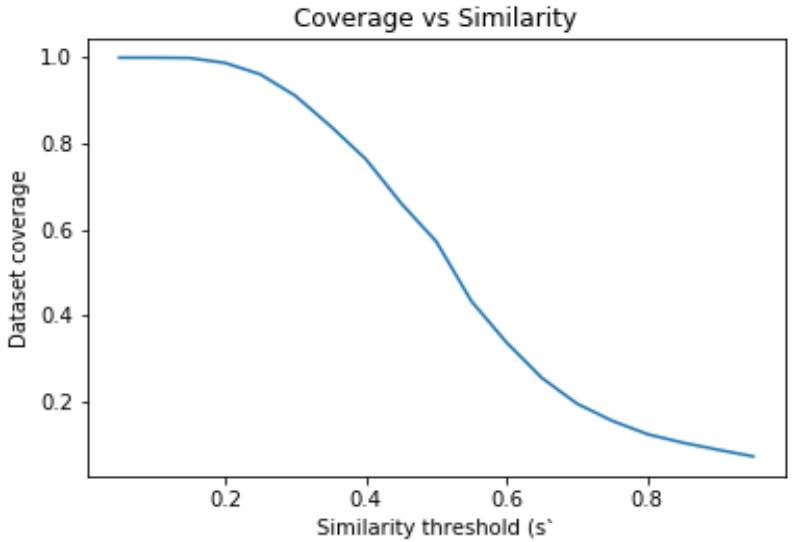


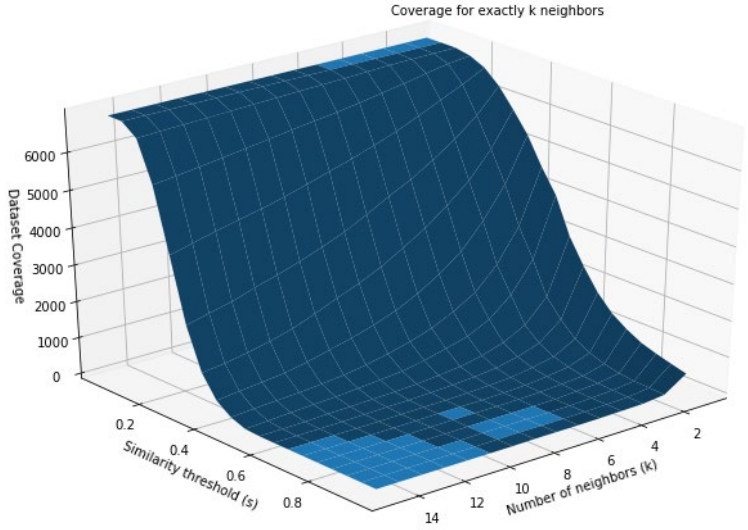


- Linear regression used to fit predicted and observed LD50 values
- R^2 = 0.61
- RMSE = 0.58
- A few outliers, but not too extreme
- Residuals clustered around zero with no obvious patterns

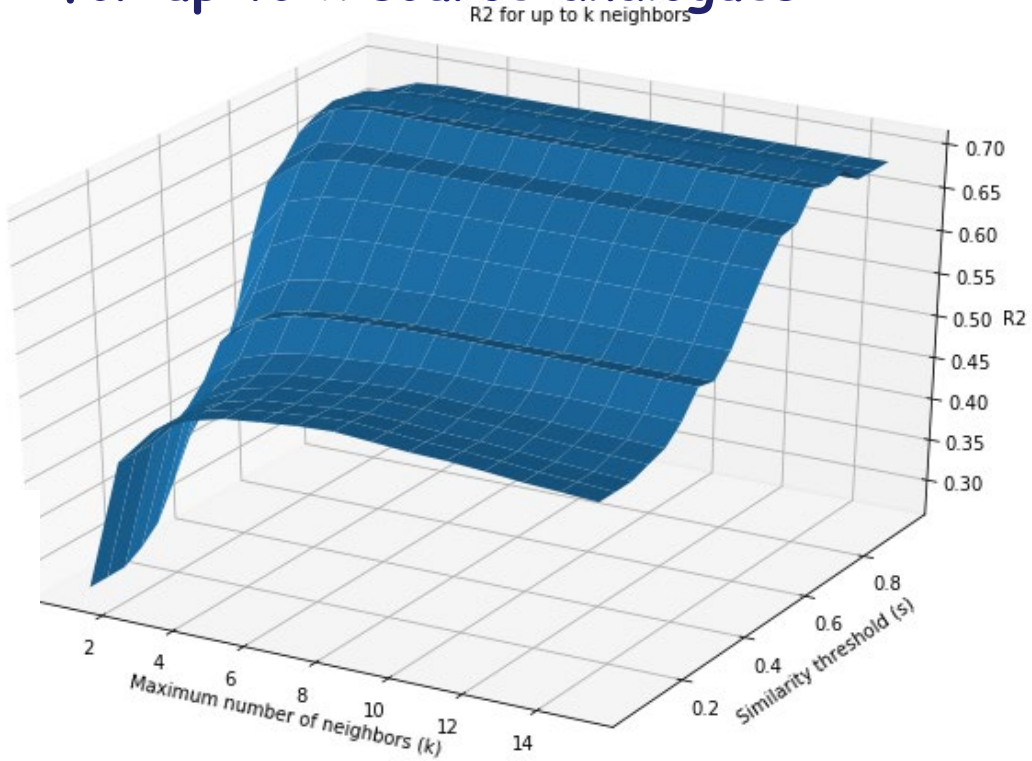
Coverage vs Similarity vs Performance

- Coverage vs Similarity



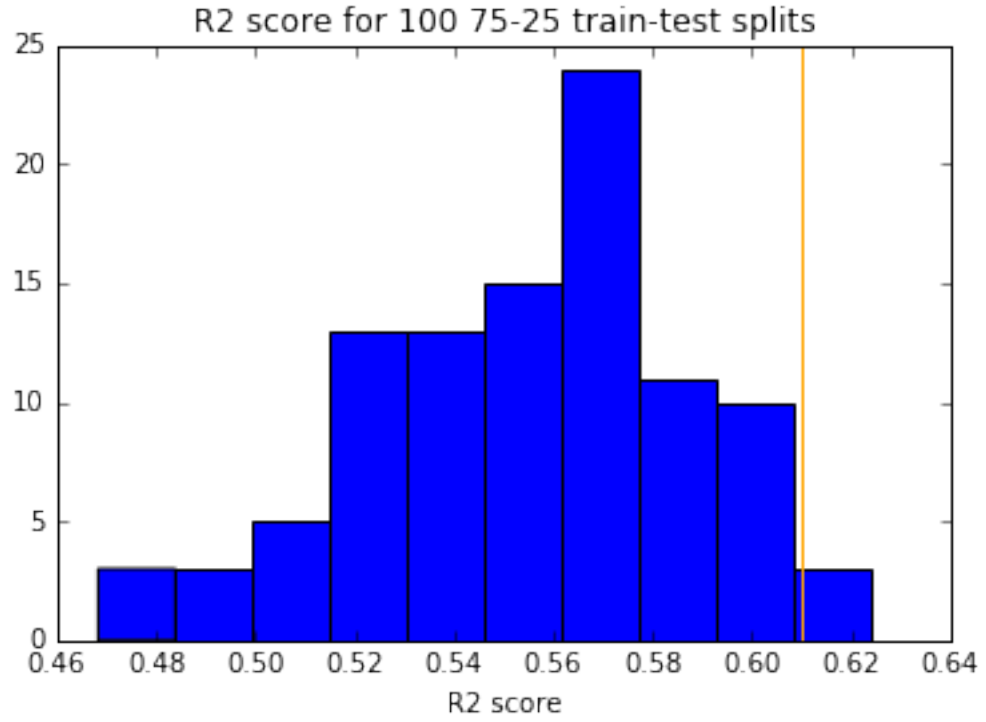


R2 for up to k source analogues



Based on the grid searches performed, k = 10, s = 0.5 were reasonable parameters to tradeoff coverage vs prediction accuracy

Monte Carlo Cross Validation



- Estimate confidence in R2
- 75-25 train-test splits

- R^2 values range from 0.46 to 0.62

- GenRA performs robustly on this acute tox data set

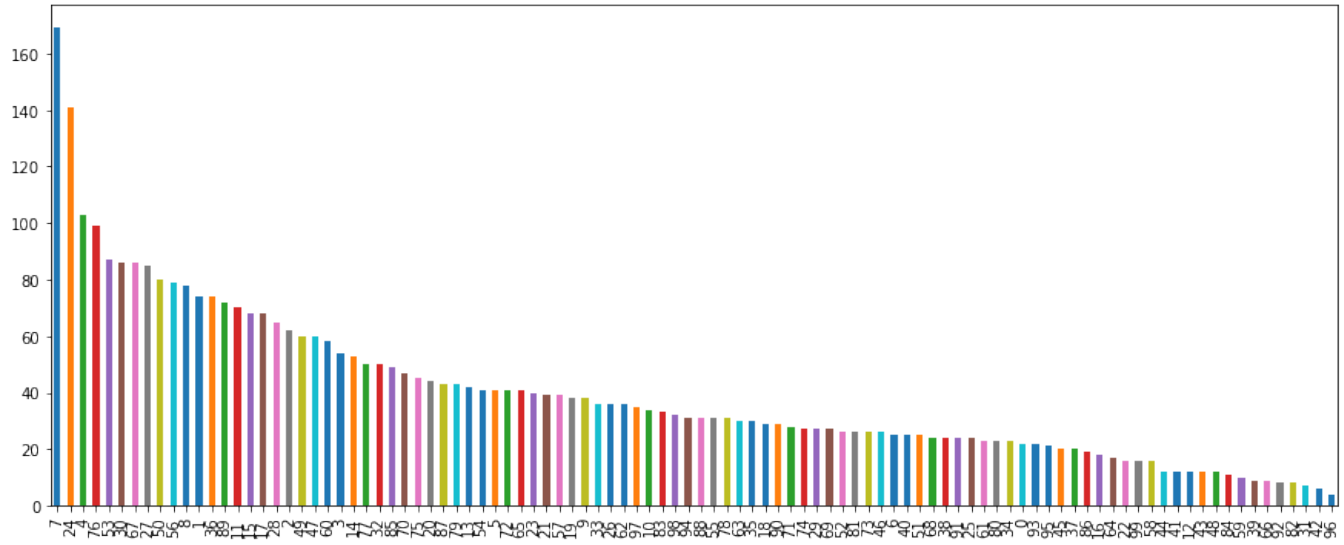
Helman et al. (2019)

Evaluating 'local' performance

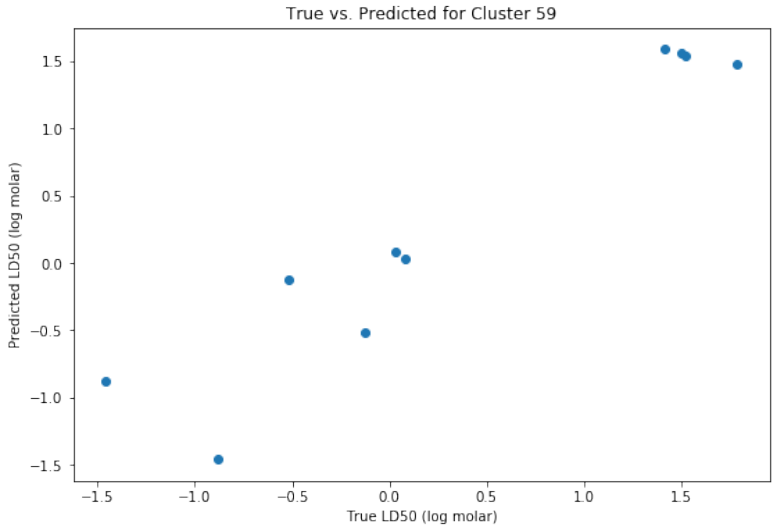
Clustered chemicals into 100 groups on the basis of ToxPrint fingerprints

Explored performance on the basis of individual clusters to gauge what sorts of chemicals resulted in significantly improved performance (R2) relative to the overall 'global' performance reported using 10 nearest neighbours and a similarity of 0.5

Average R2 values improved (R2>0.61) for 19 out of the 100 clusters, some up to 0.91

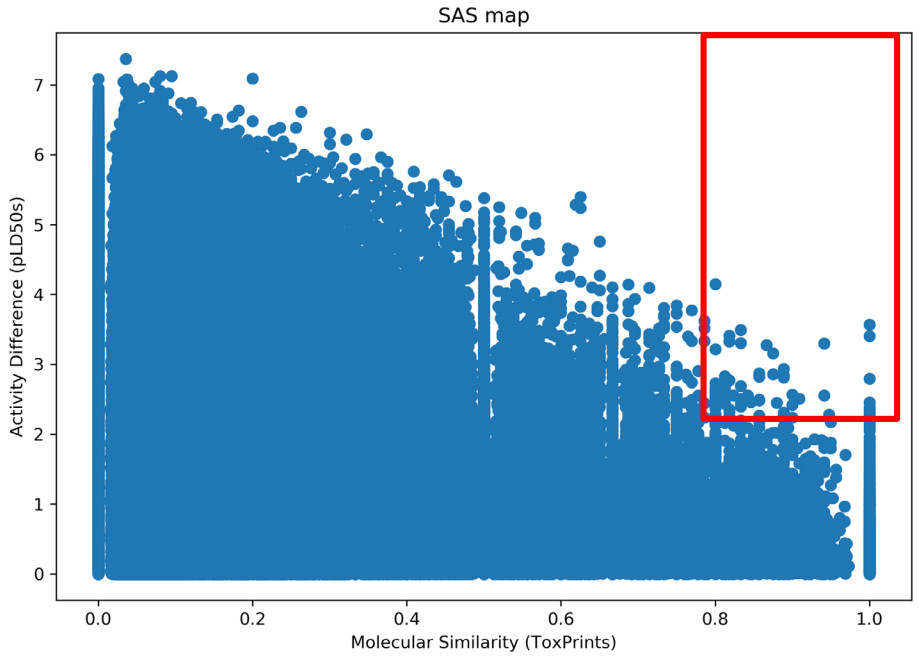


Carbamate containing substances



Structure-Activity similarity (SAS) map

- Are there pairs of substances that are very similar structurally with very high LD50 differences, so called activity cliffs



The number of chemical pairs that fell within the activity cliff quadrant was very low relative to the total number of chemical pairs captured.

This suggests that the chemical fingerprints were able to capture sufficient information to make robust predictions of acute oral toxicity.

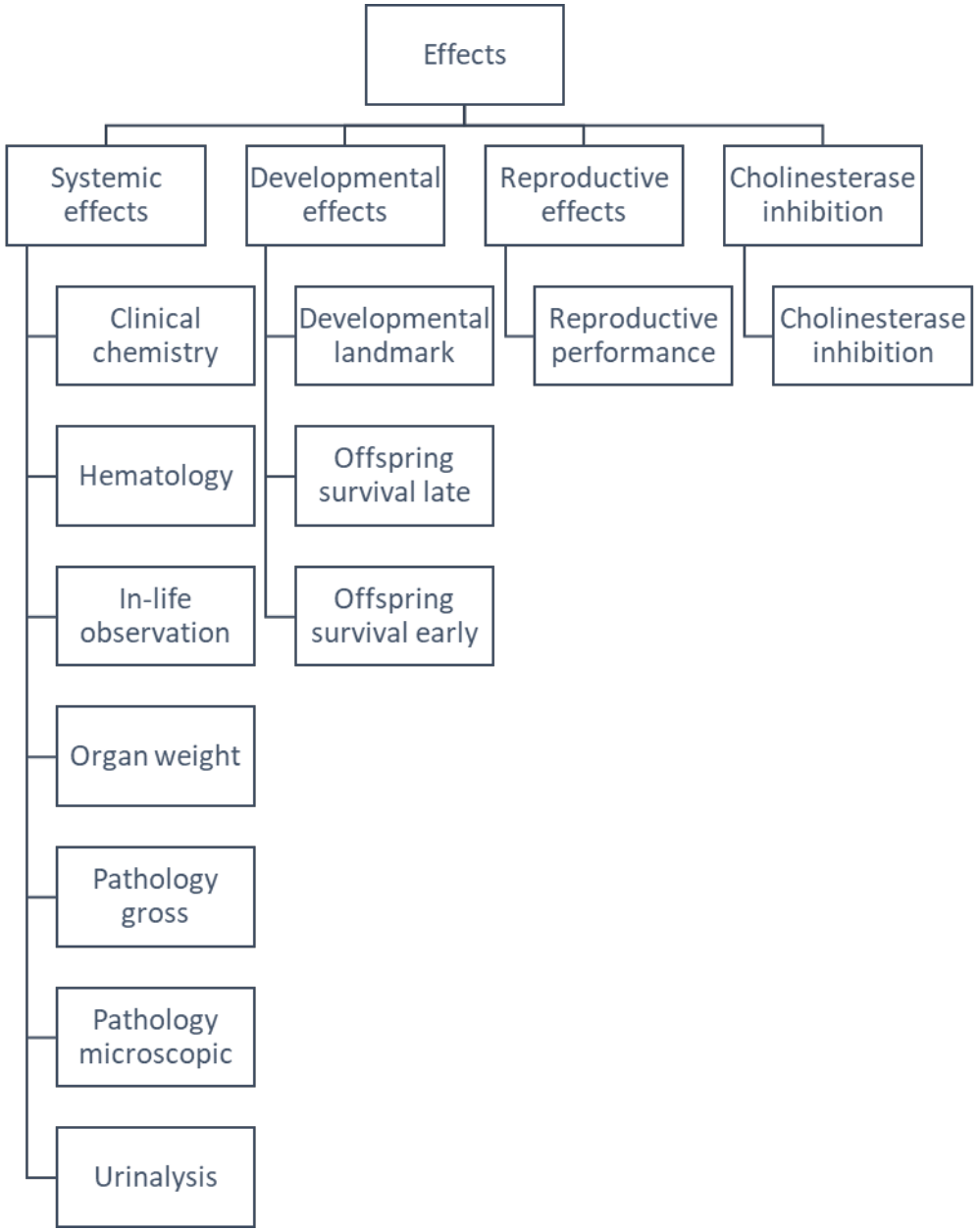
Take home messages – Part 1

- Initial GenRA (baseline) considered structural similarity and/or bioactivity to make binary predictions of toxicity
- Recent work has transitioned towards extending the GenRA approach to make quantitative predictions of toxicity
- This case study used the acute oral toxicity LD50 values collected as part of the ICCVAM ATWG and applied it to GenRA
- Using chemical fingerprints alone, a reasonable fit of R2 of 0.61 using k up to 10 and min s of 0.5
- This was a pragmatic set of parameters to balance performance with coverage
- On a 'local' level, 19 out of 100 clusters of chemicals were found to show much improved performance (up to a R2 of 0.91 in certain cases)

Case study – Repeated Dose toxicity

- How does GenRA perform using POD values from ToxRefDB 2.0.
- POD: Point of departure, or points on a dose-response curve corresponding to an observed effect level or no effect level
- POD types: LOAEL (lowest observed adverse effect level), NOAEL (no observed adverse effect level), LEL (lowest effect level), NEL (no effect level)
- 4 Broad endpoint categories: cholinesterase, developmental, reproductive, systemic effects
- 27,564 chemical level LOAEL values across 1049 substances

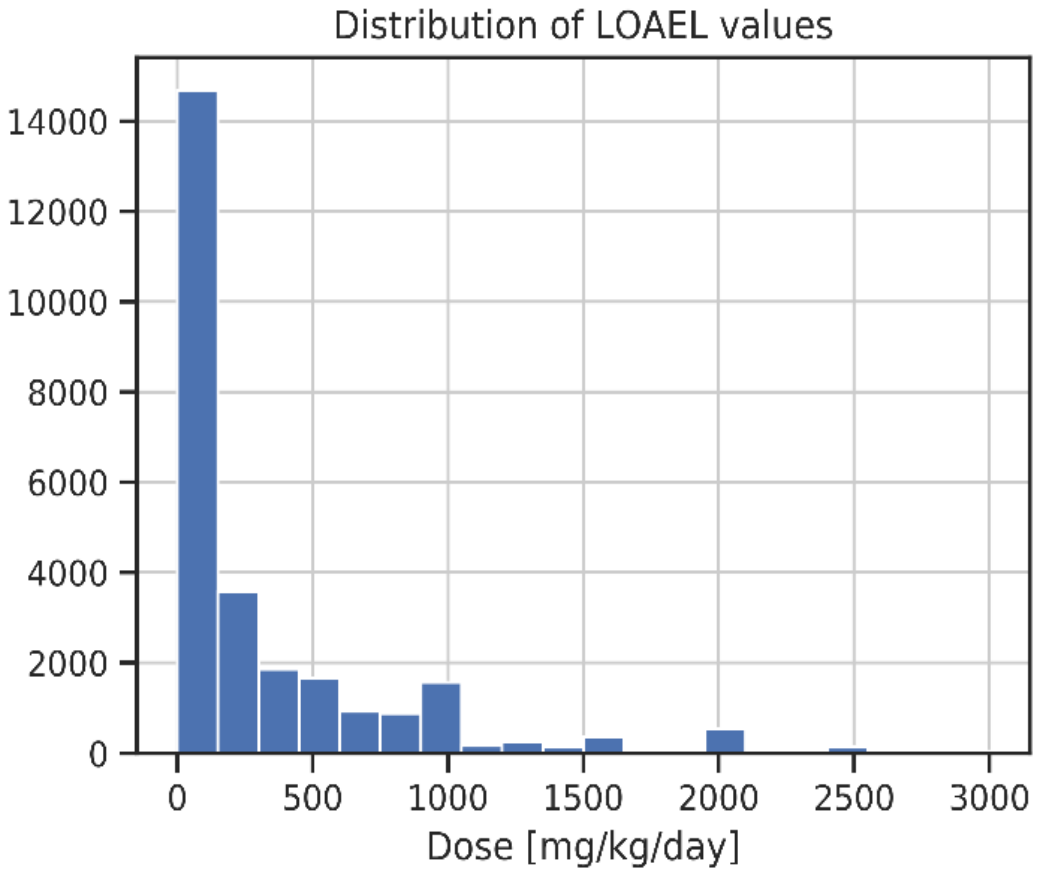
Overview of ToxRefDB v2.0 POD types

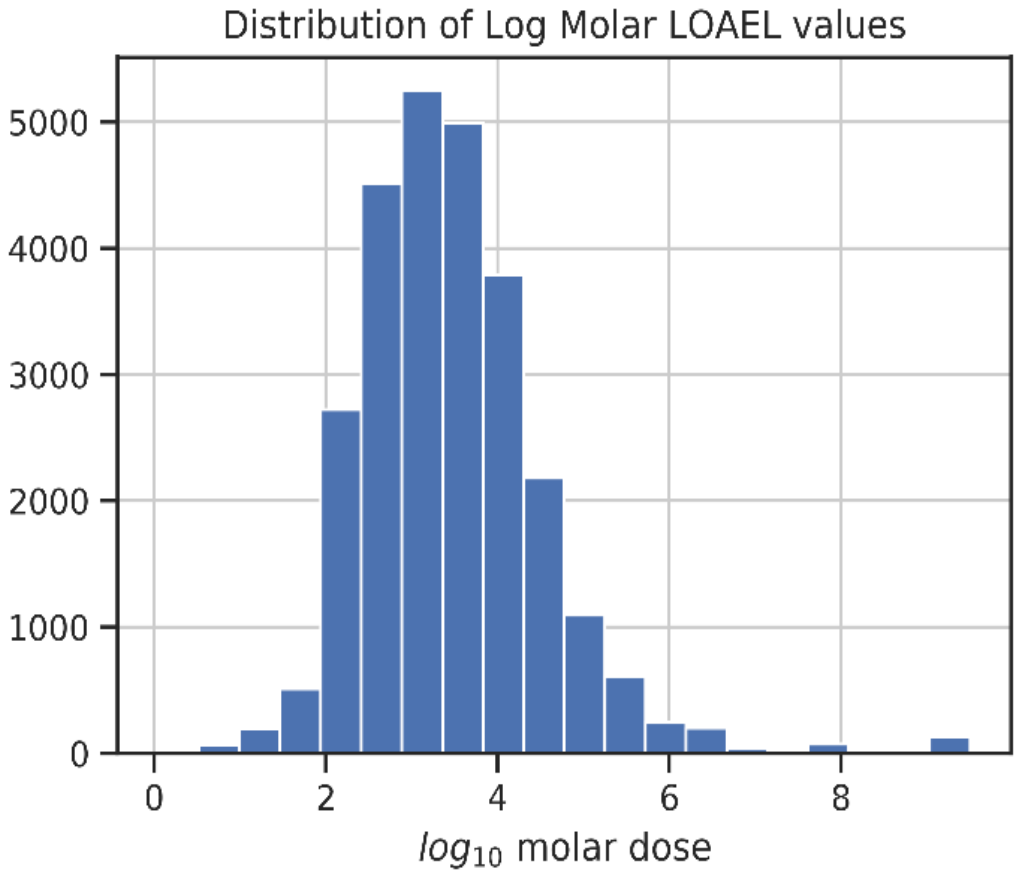


4 Endpoint Categories

14 endpoint types

Exploratory Data Analysis





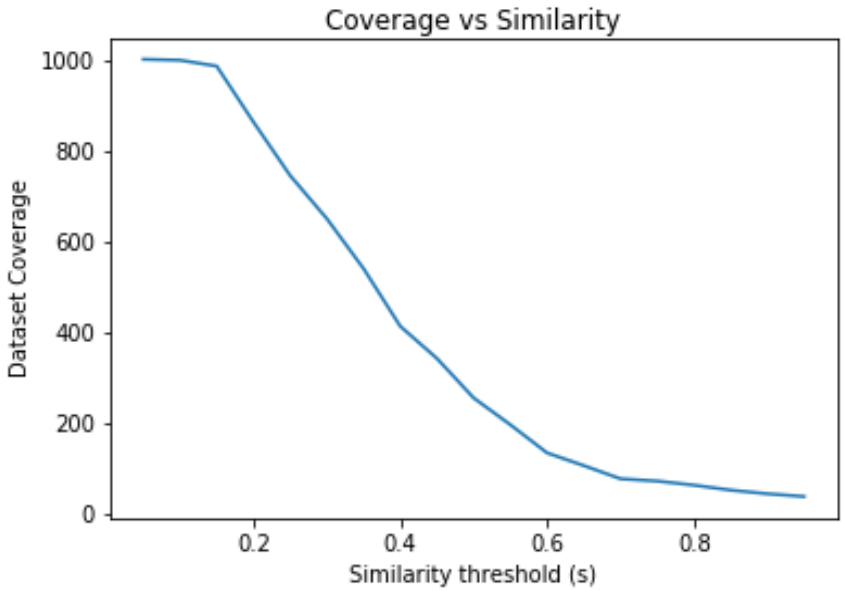
GenRA analysis approach

- For chemicals that contain multiple LOAEL values, aggregate them by taking the mean or the min and converting that to the log molar equivalents
- Use GenRA to predict LOAEL values using Morgan chemical fingerprints
- Search for a maximum of 10 nearest neighbours (k) with a min similarity (s) threshold of 0.05. Default values aimed at maximising the ability to predict LOAELs for as many chemicals as possible.
- Conduct a grid search over k (number of nearest neighbors) and s (similarity threshold) to find optimal values for R2
- Cluster analysis was performed to find local neighborhoods of chemicals where approach performs particularly well.

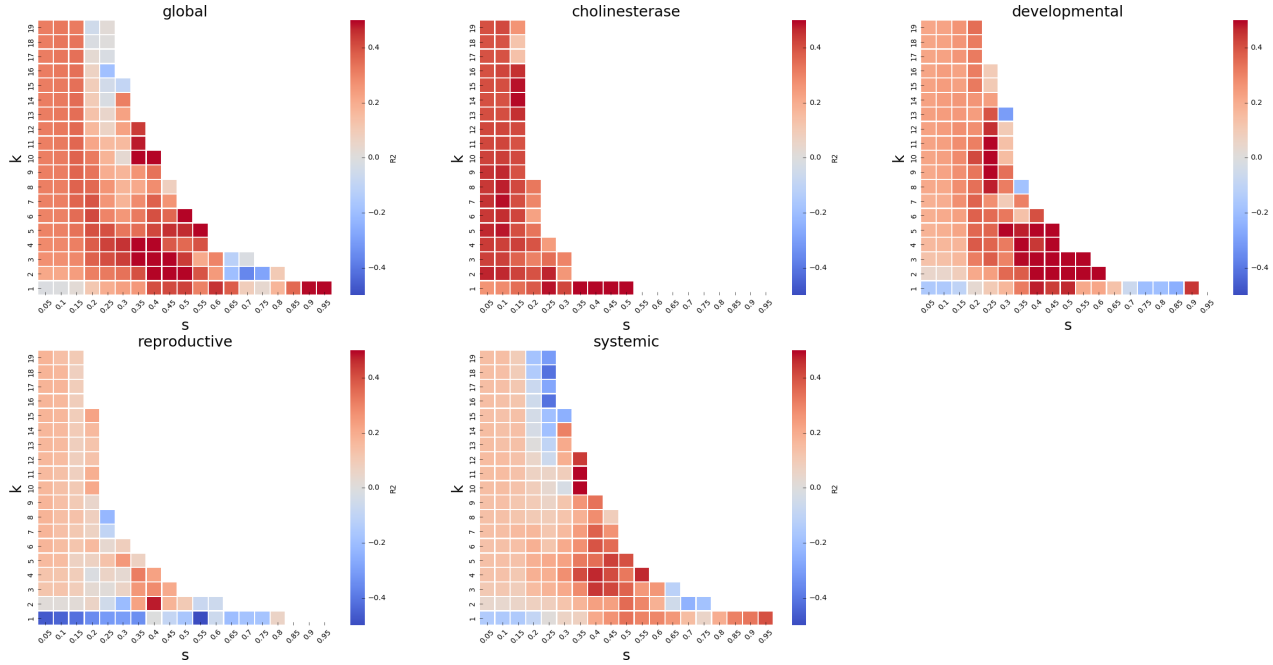
Coverage vs Similarity vs Performance

Coverage vs Similarity

The coverage of the data set is 92.3% for low values of s (s=0.10) as most chemicals in the dataset have at least one source analogue. The coverage decreases rapidly as s increases to 12.6% at s=0.60.

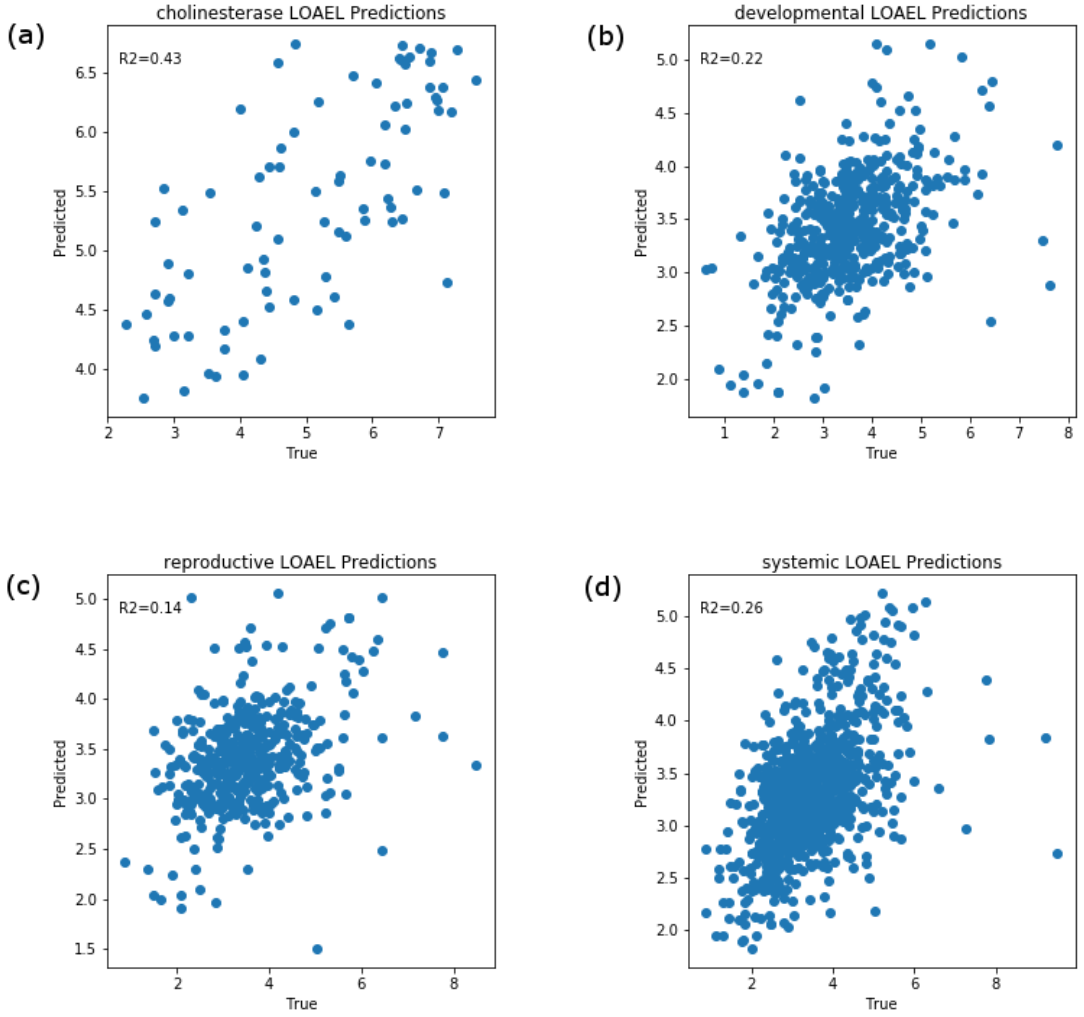


R2 for exactly k source analogues



Increasing values of k (0<k<30) are shown in the rows (from bottom to top), and increasing values of s (0<s<1) are shown in the columns (from left to right). The color of each cell corresponds to the R2 value for a specific hyperparameter (k,s) combination where the red/blue indicate high/low R2 values.

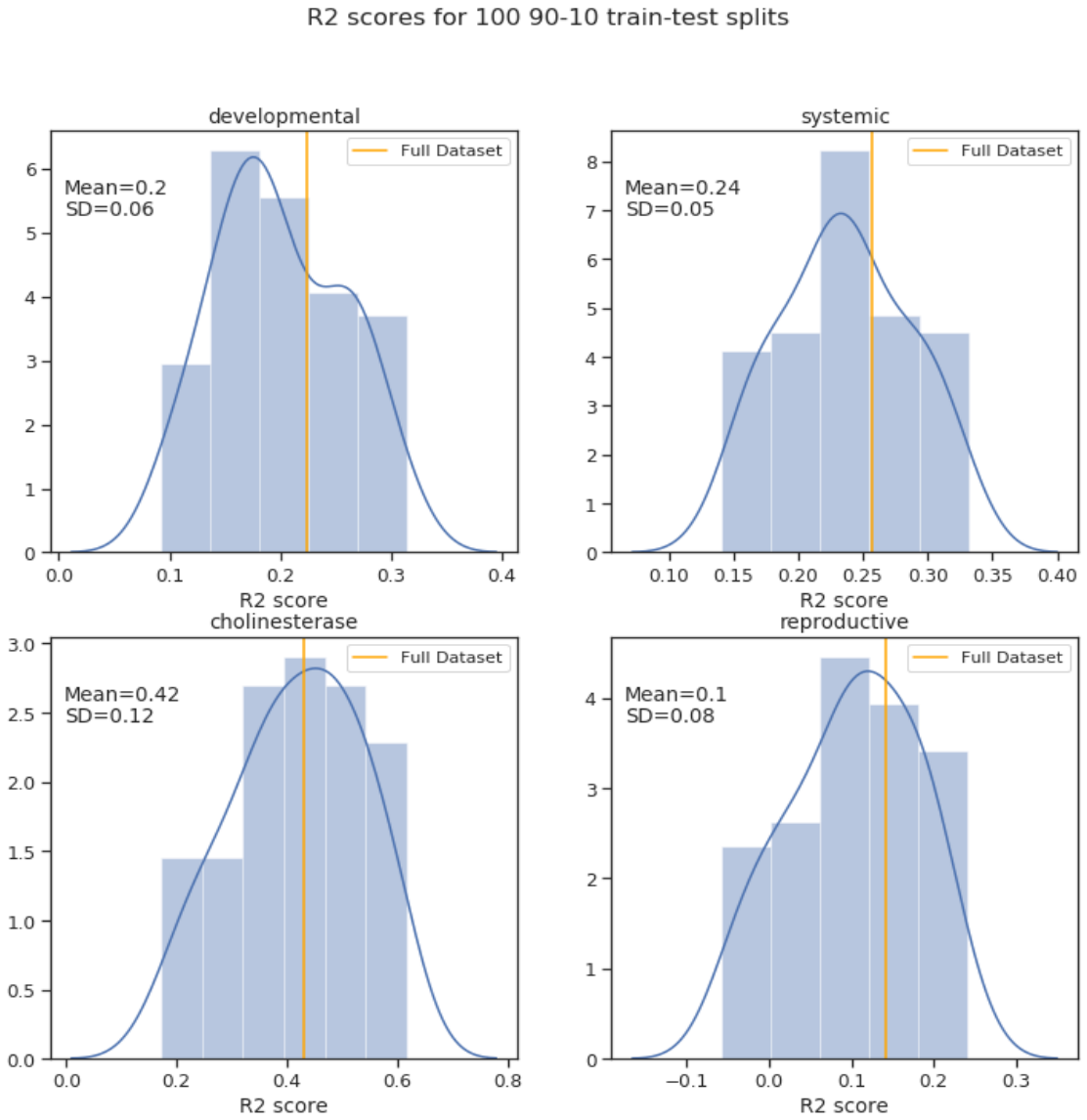
GenRA approach : Overall 'global' performance



GenRA Predictions using Morgan fingerprints with k=10 and s=0.05 (mean aggregated LOAELs)
Linear regression used to fit predicted and observed LOAEL values

Endpoint Category	R2
Cholinesterase	0.43
Developmental	0.22
Reproductive	0.14
Systemic	0.26

Monte Carlo Cross Validation



- Cross-validation testing
- 90-10 train-test splits

Endpoint	mean R2 & std
systemic toxicity	0.24±0.05
developmental toxicity	0.2±0.06
reproductive toxicity	0.1±0.08
cholinesterase inhibition	0.42±0.12

Evaluating 'local' performance

Clustered chemicals into 100 groups on the basis of Morgan fingerprints (re-used the clusters derived from GenRA v1.0)

Local GenRA predictions performed better than the global prediction by endpoint categories in 36 out of the 100 clusters.
Represent 22% (222/1014) of all chemicals

The average R2 values for systemic, developmental, reproductive effects and cholinesterase inhibition for these 36 clusters were 0.73, 0.66, 0.60 and 0.79, respectively.

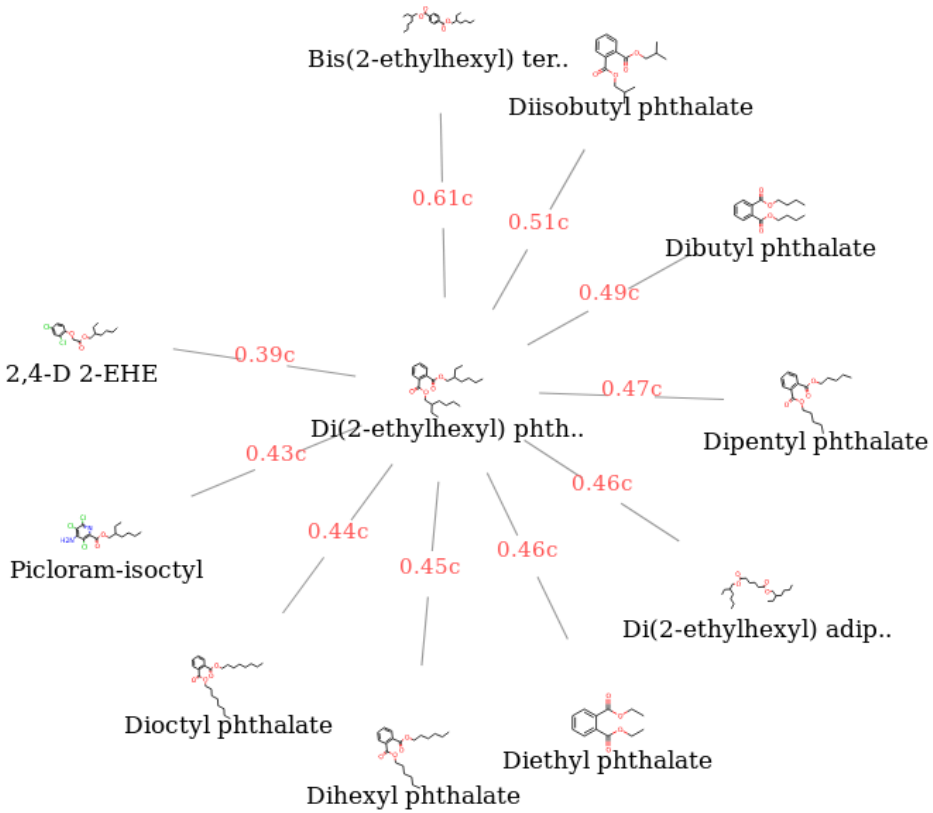
Example Predictions Di(2-ethylhexyl) phthalate

Log Molar (log mol/kg/day)

- Systemic prediction: 2.95
- Systemic measured: 3.00
- Developmental prediction: 2.95
- Developmental measured: 3.00
- Reproductive prediction: 3.04
- Reproductive measured: 3.00

Mg/kg/day

- Systemic prediction: 435.91
- Systemic measured: 388.64
- Developmental prediction: 436.73
- Developmental measured: 391.00
- Reproductive prediction: 359.65
- Reproductive measured: 391.00



Take home messages – Part 2

- This case study used the LOAEL values from ToxRefDB v2.0 and applied it to GenRA
- Using chemical fingerprints alone, fits of R2 from 0.14 to 0.43 for 4 endpoint categories (0.26 for systemic effects) were derived using k up to 10 and min s of 0.05
- On a 'local' level, 36 out of 100 clusters of chemicals were found to show much improved performance (up to a R2 of 0.76 in certain cases)

Summary

- Harmonised framework for read-across provides opportunities for NAM data
- GenRA developed is aligned with this framework
- Illustrated how GenRA baseline can been applied in practice
- Highlight ongoing research in extending the approach
  - transitioning to quantitative predictions of 'PODs' with 2 case studies – acute toxicity LD50 values and LOAELs from repeated dose toxicity studies

Acknowledgements

- Many but in particular..
- George Helman
- Imran Shah
- Tony Williams
- Jeff Edwards
- Jason Lambert
- Lucy Lizarraga
- Agnes Karmaus
- Nicole Kleinstreuer