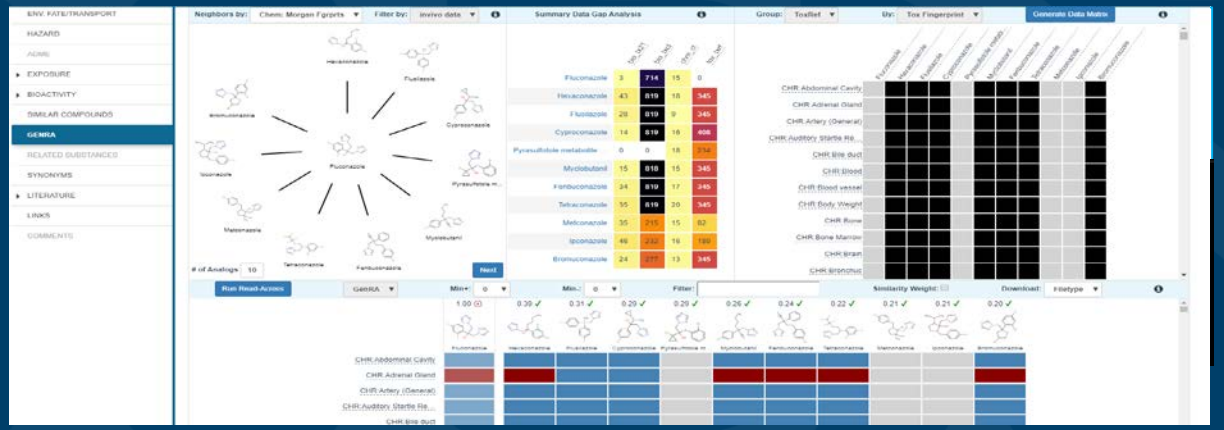
Generalised Read-Across GenRA, research, implementation and practical application



George Helman, Imran Shah, Antony Williams, Jeff Edwards, Jeremy Dunne, Grace Patlewicz
National Center for Computational Toxicology (NCCT), US EPA

Outline

- Definitions
- Landscape of read-across guidance & tools
- Re-thinking the read-across problem
- Summary remarks
- Acknowledgements

Definitions: Chemical grouping approaches

A chemical category is a group of chemicals whose physico-chemical and human health and/or environmental toxicological and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristics).

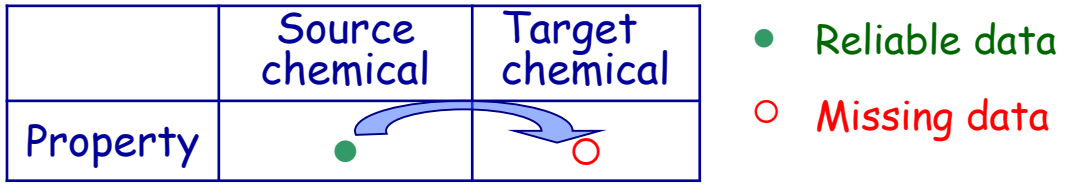
- Read-across describes one of the techniques for filling data gaps in either the analogue or category approaches i.e. not to be confused with the "analogue approach"

- "Analogue approach" refers to grouping based on a very limited number of chemicals (e.g. target substance + source substance)

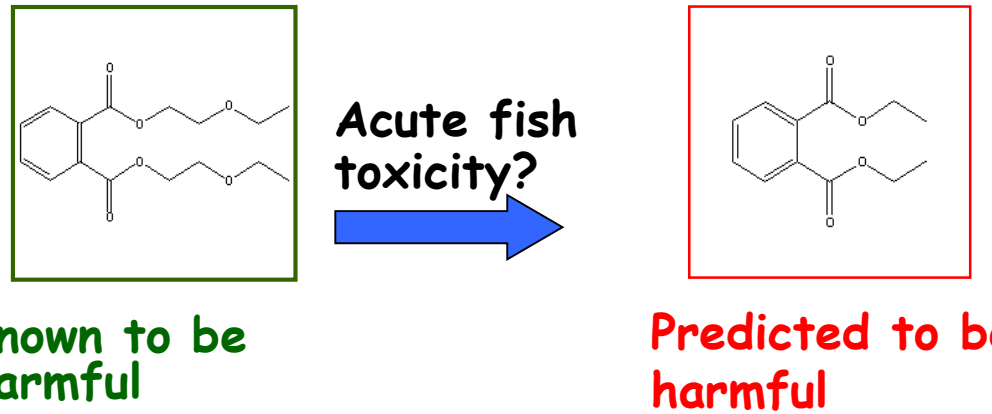
- "Category approach" is used when grouping is based on a more extensive range of analogues (e.g. 3 or more members)

Definitions: Read-across

Known information on the property of a substance (source) is used to make a prediction of the same property for another substance (target) that is considered "similar" i.e. endpoint & often study specific



- Reliable data
- Missing data



Known to be harmful

Predicted to be harmful

Landscape of read-across – 'Guidance'

- Intended to address:
- 1) the development of read-across
  - i.e. the process of deriving an analogue/category approach to facilitate a read-across prediction
  - technical regulatory guidance (OECD grouping document (2014), ECHA (Chapter R6, (2008)) and many publications in the scientific literature (Wu et al., 2010; ECETOC, 2012; Wang et al., 2012, Patlewicz et al., 2013)
- 2) the assessment (evaluation) of the read-across justification
  - technical regulatory guidance (ECHA RAAF, 2015,2017; OECD IATA templates) and publications in the scientific literature (Blackburn and Stuard, 2014; Patlewicz et al., 2015; Schultz et al., 2015)

Issues surrounding the consistency and concordance of the different guidance available

Landscape of read-across tools

- A number of different tools exist both in the public domain and commercially
- Examples include EPA's AIM, OECD Toolbox, JRC Toxmatch, Leadscope, Molecular Networks ToxQPS, ToxRead, CBRA..

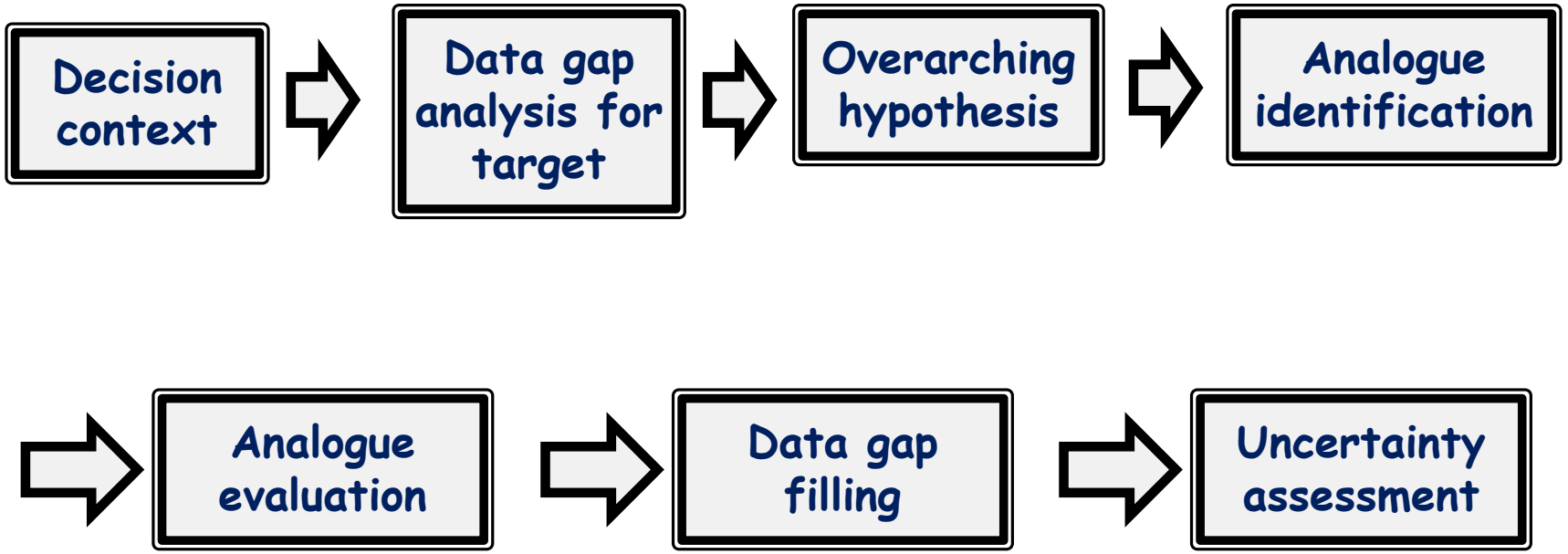
Difficult to compare and contrast these tools in terms of their utility

Need a consistent framework/workflow to understand their scope and utility and for what decision context(s) they might be useful for

Re-thinking the read-across problem

- Objective 1. Define the category (read-across) workflow
- Objective 2. Understand the scope and capability of existing read-across tools
- Objective 3. Identify an objective means of quantifying the performance of read-across and quantifying the uncertainties – Generalised Read-across (GenRA)
- Objective 4: Propose a harmonised hybrid read-across workflow
- Objective 5. Extend the approach to fold in expert driven considerations but in an objective manner

Objective 1: Defining the category (read-across) workflow



Patlewicz et al., 2013; ECETOC, 2012

Objective 2: Scope and capability of 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

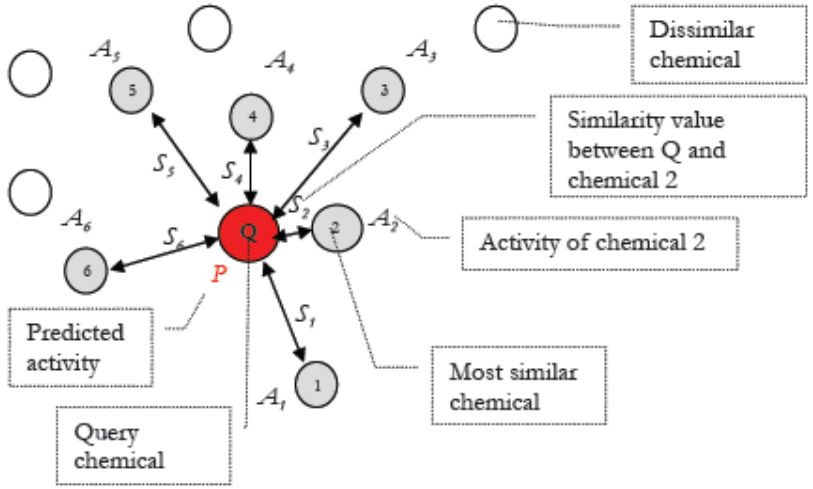
ABSTRACT

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.

Objective 3: GenRA (Generalised Read-Across)

- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and/or bioactivity descriptors
- Goal: to systematically evaluate read-across performance and uncertainty using available data
- The approach enabled a performance baseline for read-across predictions of toxicity effects within specific study outcomes to be established



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

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^{β} = activity of c_j in β

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

k = up to k nearest neighbours

GenRA analysis workflow

I. Data

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

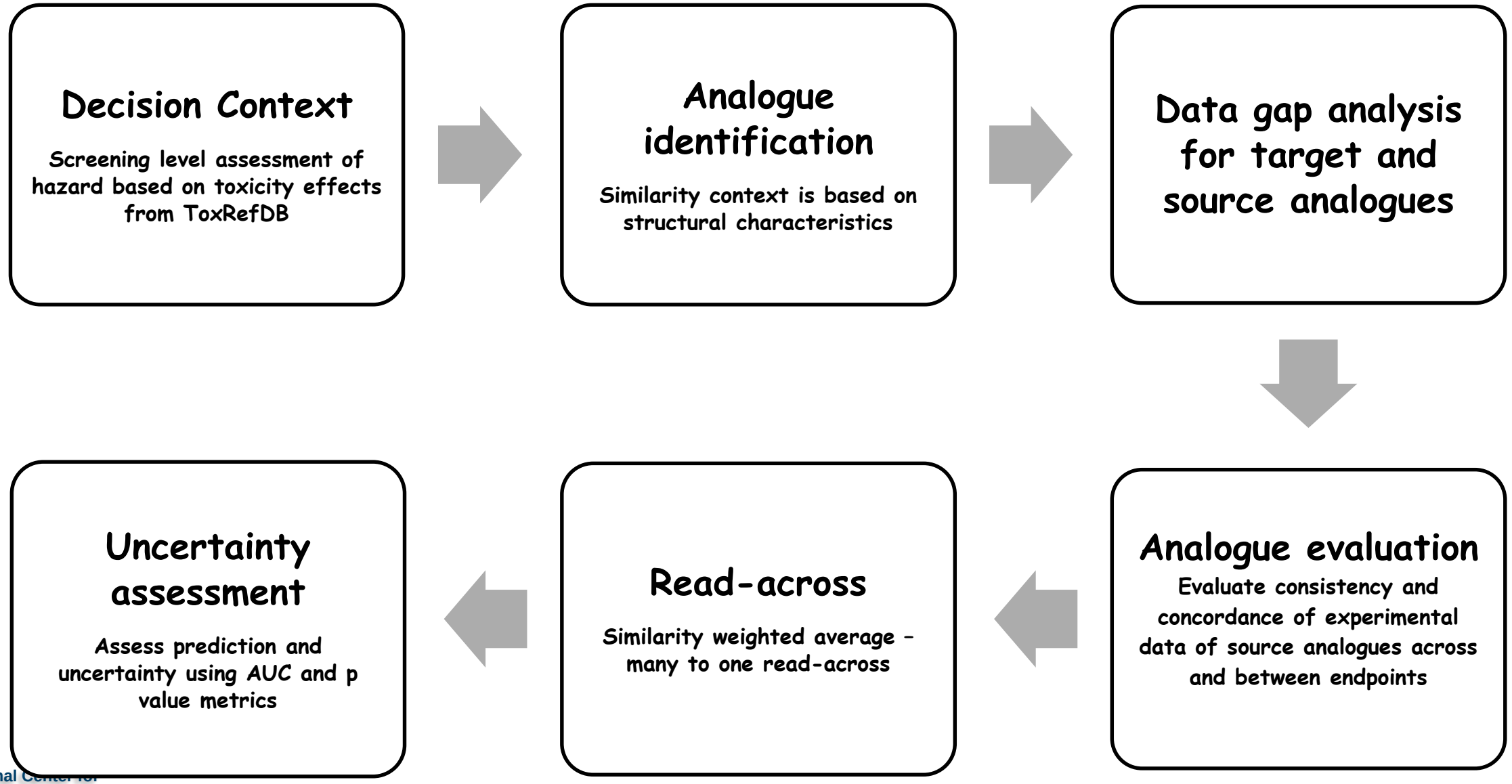
II. Define Local neighborhoods

Us 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 neighbor hoods
Evaluate impact descriptors (chm, bio, bc) on prediction
Quantify uncertainty

Objective 3: Read-across workflow in GenRA



Objective 3: GenRA tool in reality

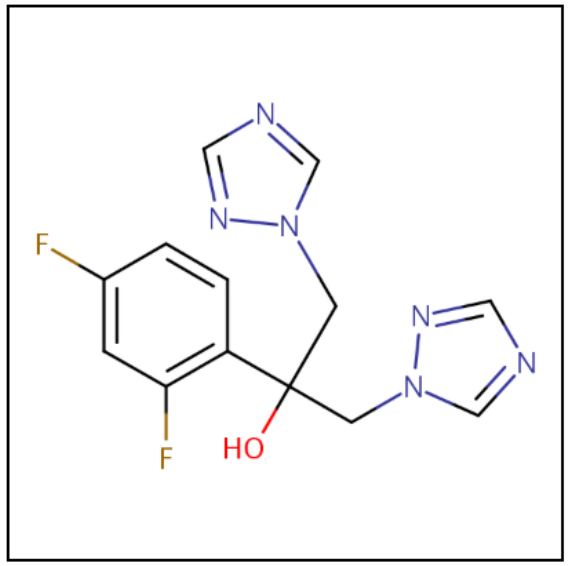
- **Integrated into the EPA CompTox Chemistry dashboard as a new addition**

Fluconazole

86386-73-4 | DTXSID3020627

Searched by DSSTox Substance Id.

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS



Wikipedia

Fluconazole is an antifungal medication used for a number of fungal infections. This includes candidiasis, blastomycosis, coccidiodomycosis, cryptococcosis, histoplasmosis, dermatophytosis, and pityriasis versicolor. It is also used to prevent candidiasis in those who are at high risk such as following organ transplantation, low birth weight babies, and those with low blood neutrophil counts. It is given either by mouth or by injection into a vein.

Common side effects include vomiting

...

Read more

Intrinsic Properties

Molecular Formula: $C_{13}H_{12}F_2N_8O$ Mol File

Find All Chemicals

Average Mass: 306.277 g/mol Isotope Mass Distribution

Monoisotopic Mass: 306.104065 g/mol

Structural Identifiers

Linked Substances

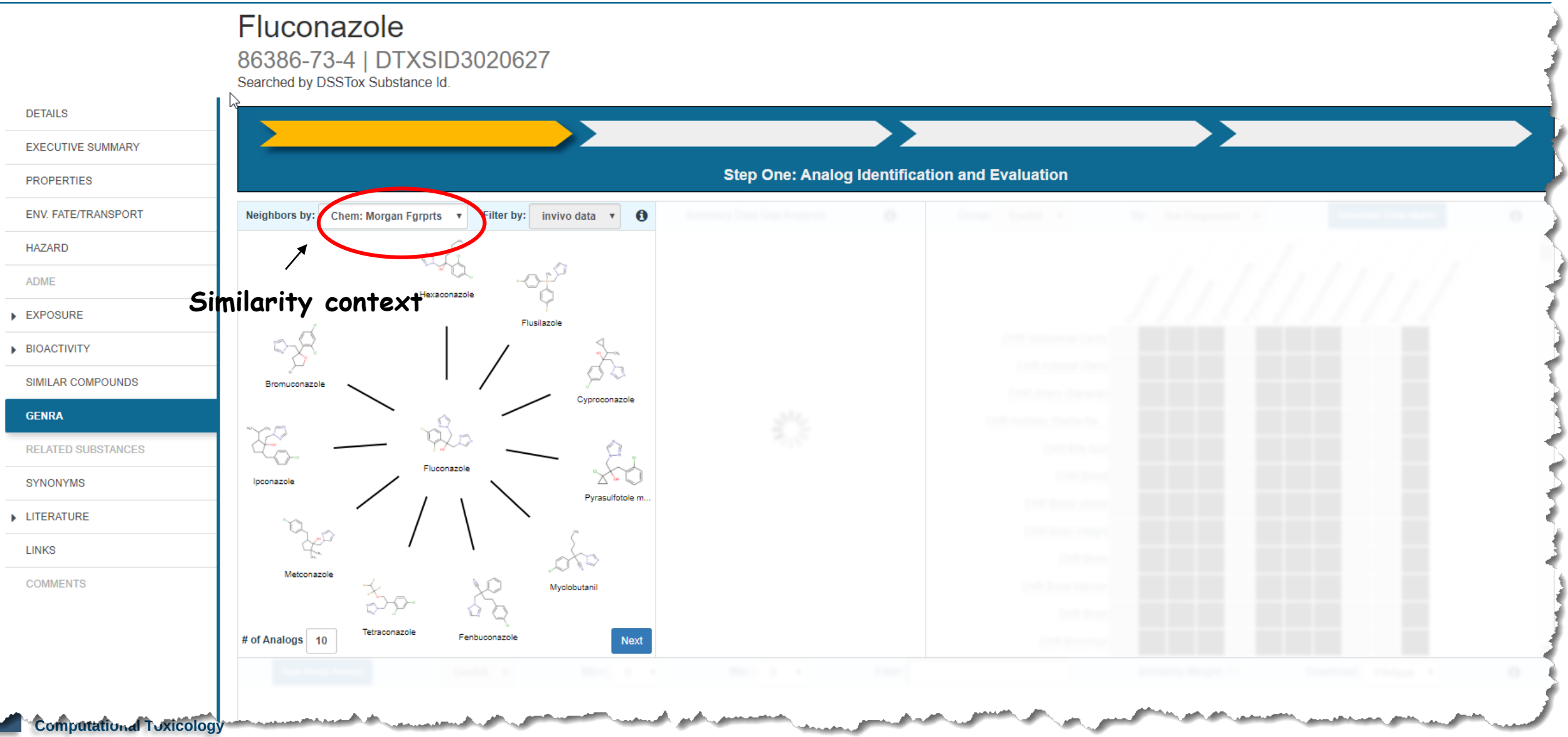
Presence in Lists

Record Information

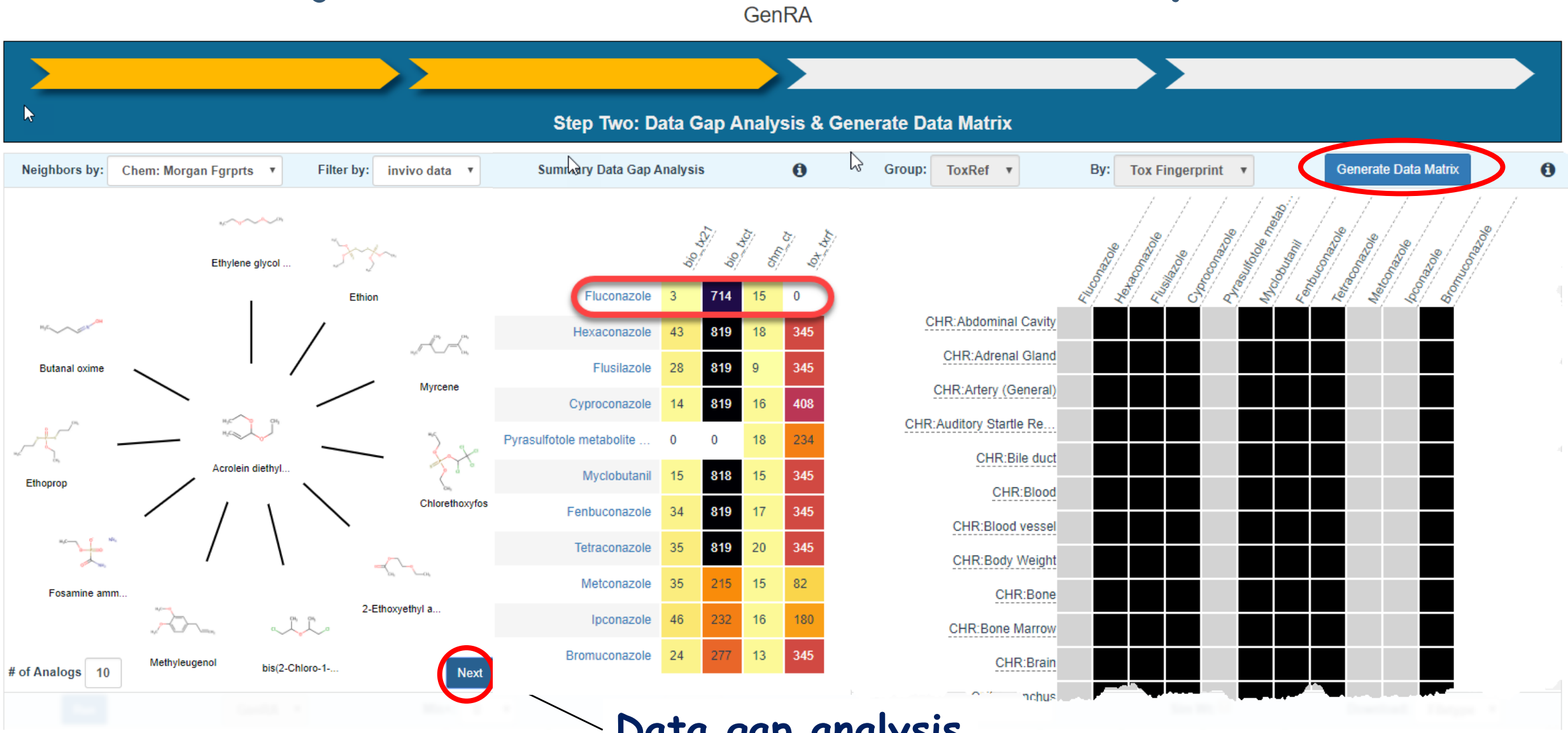
Quality Control Notes

Objective 3: GenRA tool in reality

- Structured as a workflow

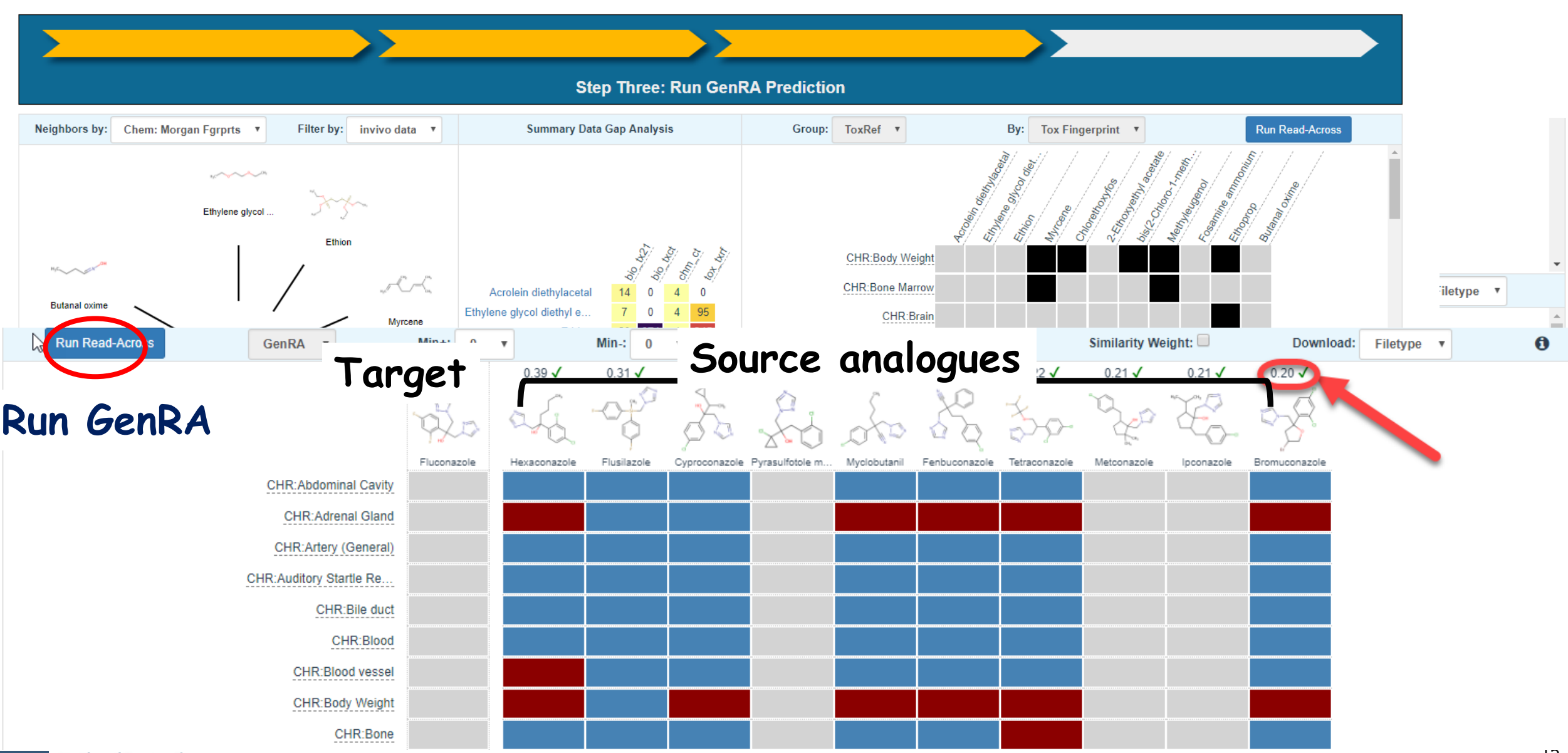


Objective 3: GenRA tool in reality



Data gap analysis

Objective 3: GenRA tool in reality



DEMONSTRATION

Objective 2: Extending the suite of read-across tools but addressing an unmet need

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	Public release date August 2018

Objective 4: A harmonised hybrid read-across workflow

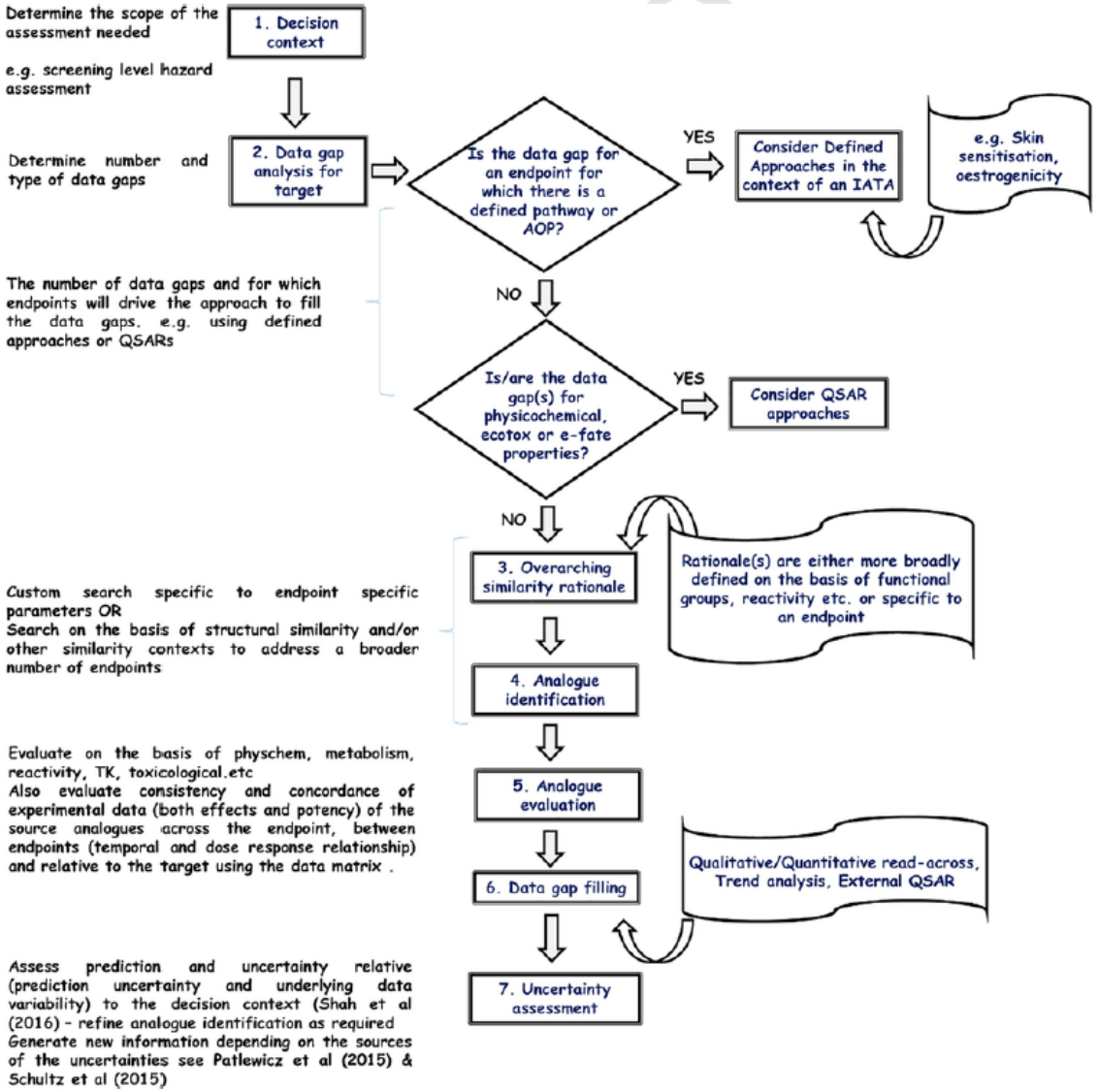


Fig. 9. A harmonised hybrid development and assessment framework.

Folding in the learnings in GenRA to inform and update a harmonised workflow

Patlewicz et al., 2018

Objective 4: A harmonised hybrid read-across workflow

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

Objective 5: GenRA – Next Steps

- Ongoing research:
- Summarising and aggregating the toxicity effect predictions to guide end users – what are the effects to be concerned about and which effect predictions are we most confident about
- Consideration of other information to define and refine the analogue selection – e.g. physicochemical similarity, metabolic similarity, reactivity similarity…
  - EPA New Chemical Categories
  - Quantifying the impact of physicochemical similarity on read-across performance
- 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 PODs

Objective 5: Refinements to the GenRA approach

Decision Context

Screening level assessment of hazard based on toxicity effects from ToxRefDB

Similarity contexts	GenRA
Structure similarity	√
Physicochemical	Subject of this study
Bioactivity e.g. ToxCast	-
Reactivity	-
Metabolic	-
Toxicokinetic	-

Data gap analysis for target and source analogues

Analogue evaluation

Evaluate consistency and concordance of experimental data of source analogues across and between endpoints

Read-across

Similarity weighted average – many to one read-across

Uncertainty assessment

Assess prediction and uncertainty using AUC and p value metrics

Physchem Similarity Context

- Important context of similarity in read-across
- Models "bioavailability"
- Properties selected: Lipinski Rule of 5 (LogP, MW, # HB donors/acceptors)
- Two approaches investigated as a means to identify source analogs and evaluate their predictive performance relative to GenRA:

Approach 1: "Filter"

Subcategorise from a set of analogues identified based on structural similarity

Common approach

Approach 2: "Search Expansion"

"Frontload" both structure and physchem into analogue identification

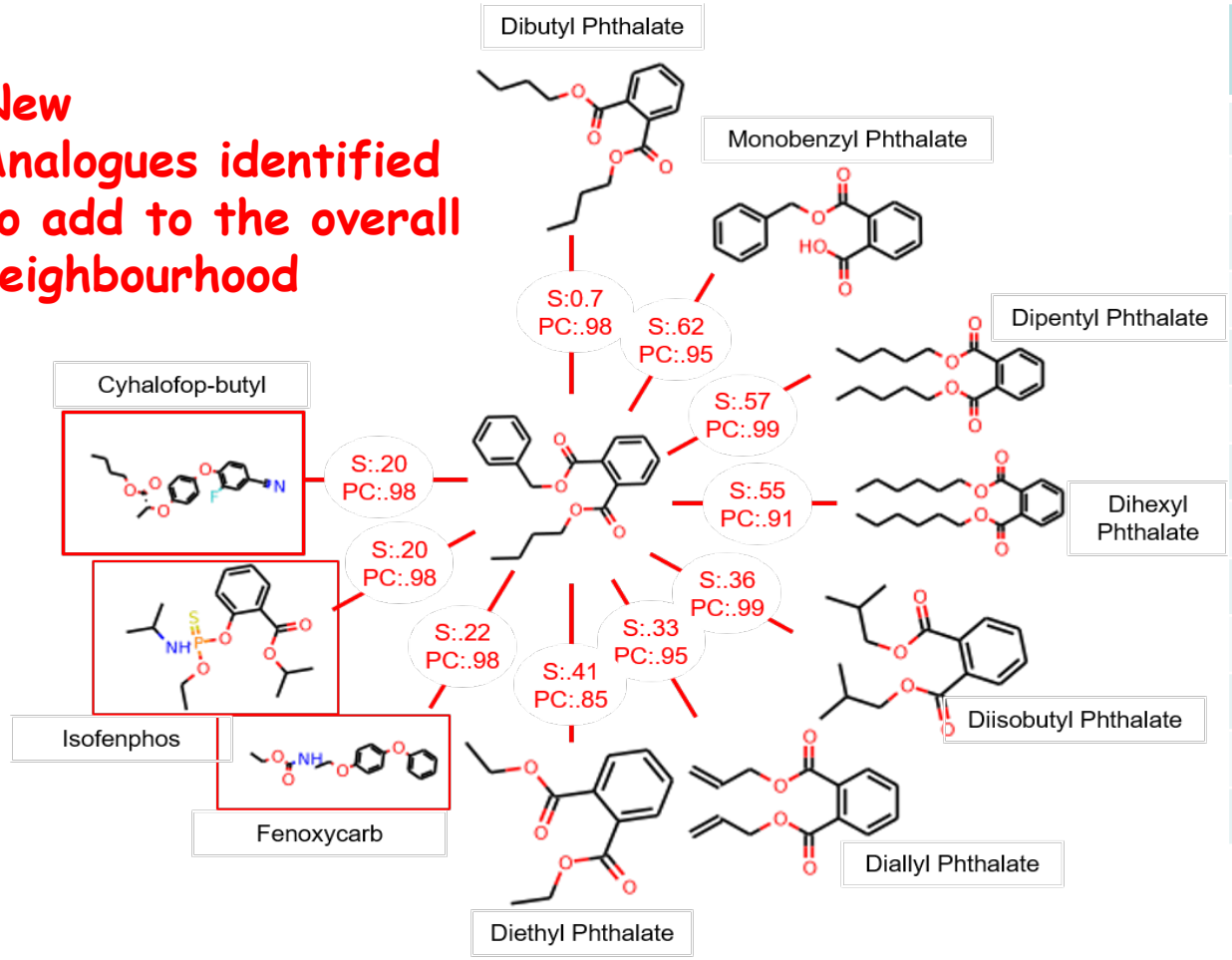
Novel approach

Helman et al., 2018

Case Study: Butyl Benzyl Phthalate

Approach 2: Search Expansion

New Analogues identified to add to the overall neighbourhood

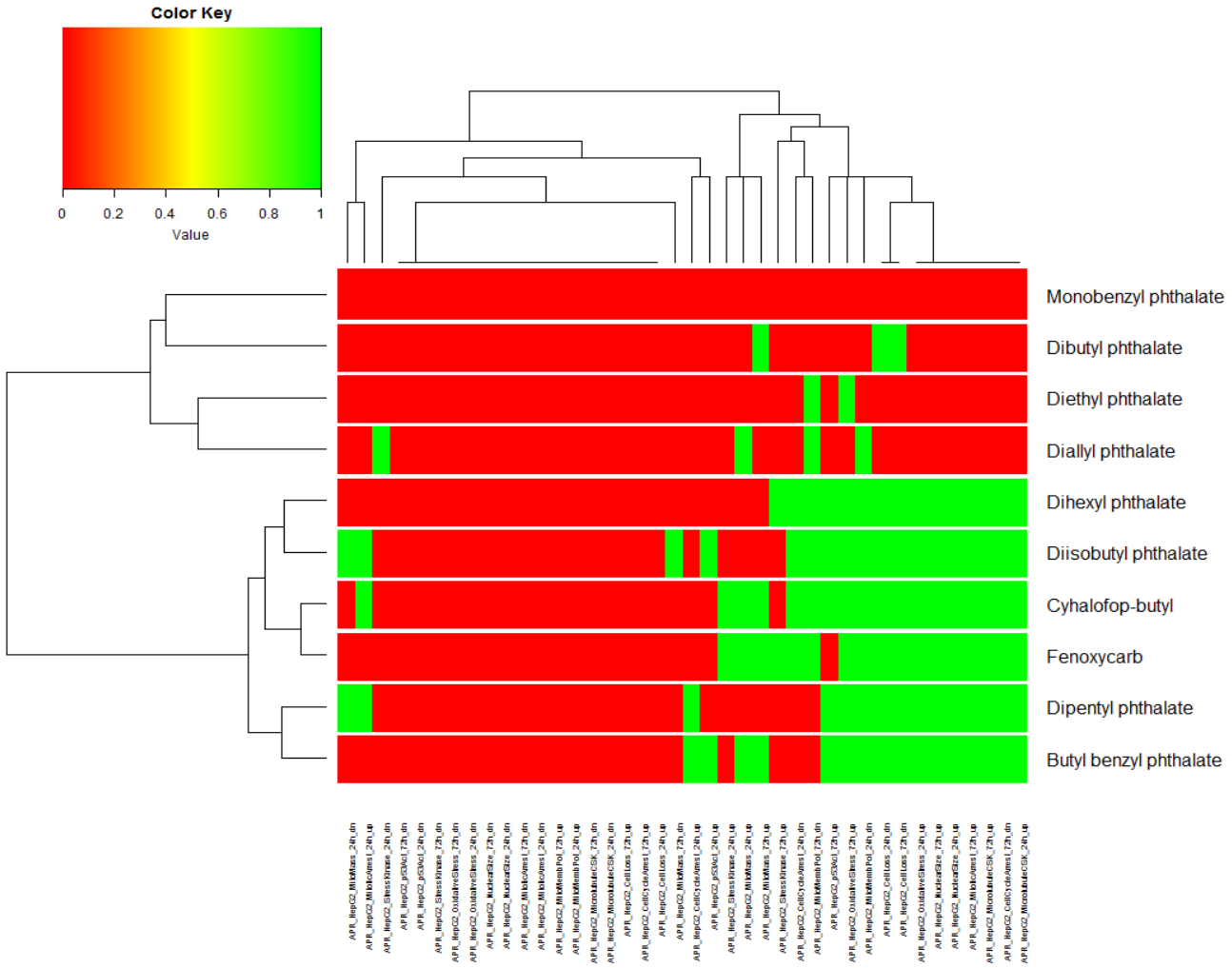


Endpoint	Baseline Prediction	Structure + Pchem Prediction
Body Weight	.78	.79
Clinical Chemistry	.27	.60
Food Consumption		
Hematology		
Kidney		
Liver		
Mortality		
Pancreas		
Prostate		
Skin		
Spleen		
Tissue NOS		
Urinary Bladder	0	0

- Adding phys-chem to similarity search overturns incorrect predictions for 2 endpoints
- Improves many others

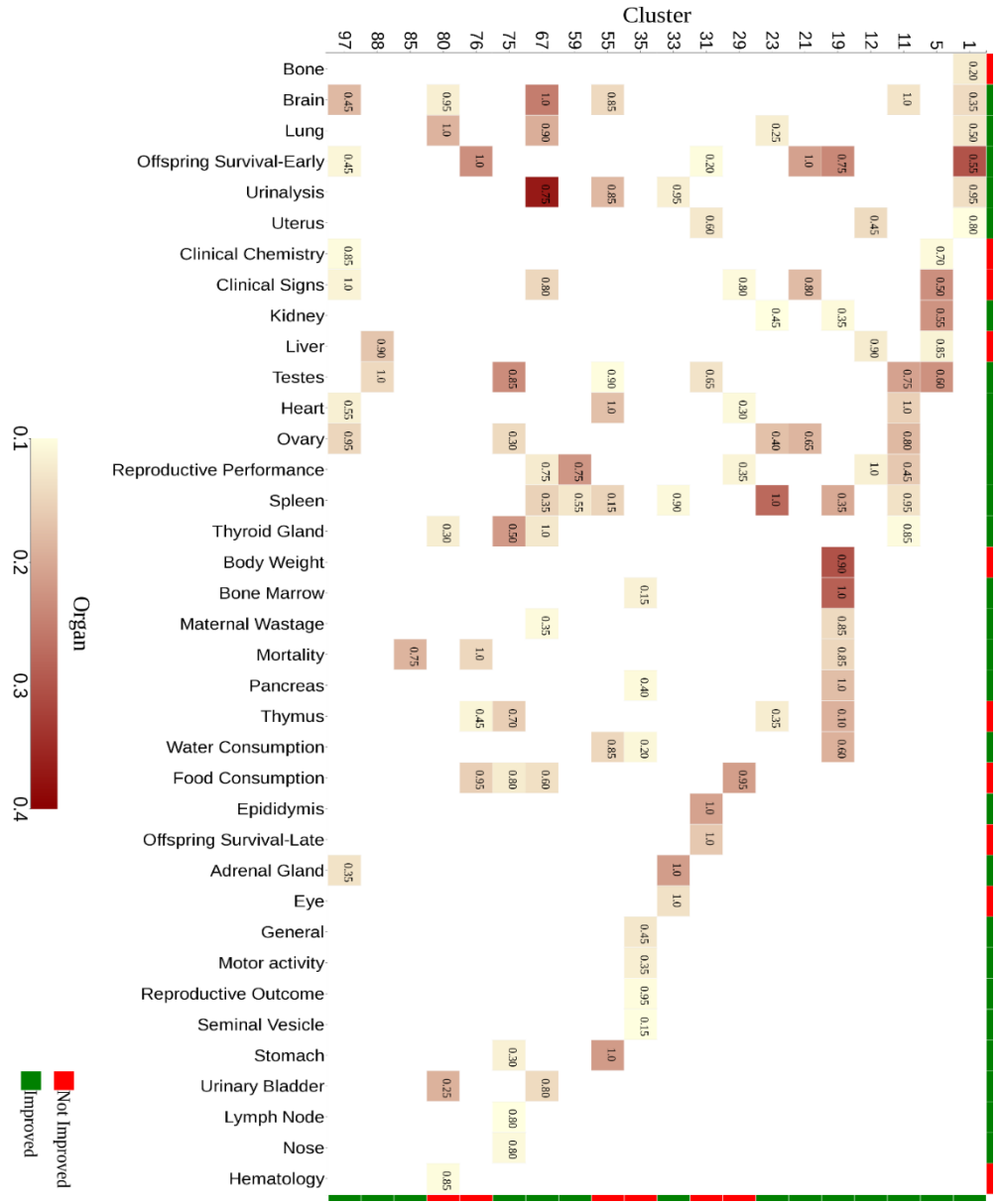
Case Study: Butyl Benzyl Phthalate

Approach 2: Search Expansion



- Are the non phthalate analogues plausible from a biological similarity context?
- Heatmap of ToxCast bioactivity profiler from one (Apredica) technology
- From a qualitative perspective – these non phthalates exhibit similarity wrt their bioactivity profile to the target and other source phthalates

"Search expansion" in practice



1) Identify target chemical

2) Perform Data gap analysis

3) Use cluster/organ key to guide selection of the optimal physicochemical threshold to use in source analogue identification for a specific toxicity effect of interest

Helman et al., 2018

"S

ELSEVIER
Computational Toxicology
Available online 23 July 2018
In Press, Corrected Proof

Extending the Generalised Read-Across approach (GenRA): A systematic analysis of the impact of physicochemical property information on read-across performance

George Helman [a, b], Imran Shah [b], Grace Patlewicz [b]
Show more

https://doi.org/10.1016/j.comtox.2018.07.001

Highlights

- GenRA approach is summarised in the context of the category workflow.
- The impact of physicochemical information on read-across performance was assessed in 2 ways: filtering and search expansion.
- Search expansion resulted in an up to 9% improvement in read-across performance for 10 of the 50 data rich target organs.
- Results are summarised on a neighbourhood (chemical category) basis.
- A case study substance is used to compare and contrast the read-across performance using the 2 approaches.

Flucona
86386-73-
Searched by DSS

DETAILS
EXECUTIVE SUMMARY
PROPERTIES
ENV. FATE/TRANSPORT
HAZARD
ADME
EXPOSURE
BIOACTIVITY
SIMILAR COMPOUNDS
GENRA
RELATED SUBSTANCES
SYNONYMS
LITERATURE
LINKS
COMMENTS

Phys
Stru

Neigh

Bromuconazole
Ipconazole
Metconazole

# of Analogs 10

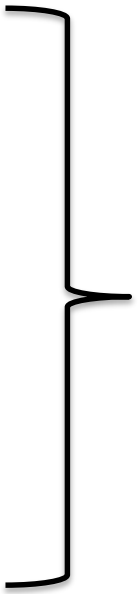
(w1),
dependent
nterest

Objective 5: Refinements to the GenRA approach

- Transitioning GenRA from binary predictions to 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 oral toxicity data

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

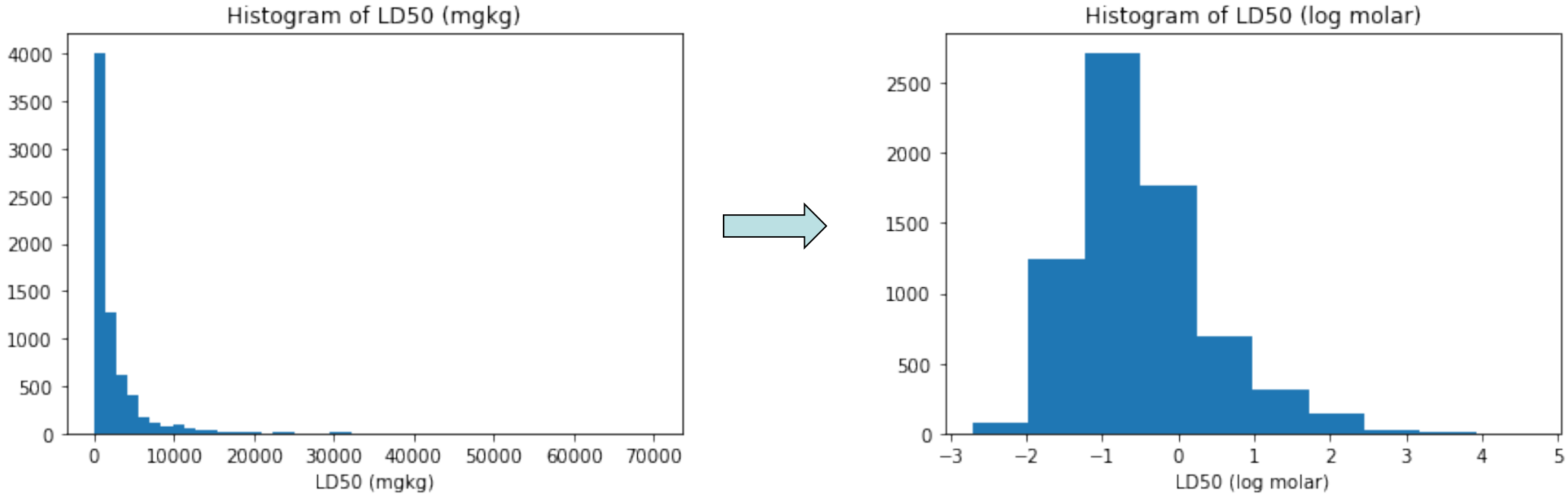
Require unique LD50 values with mg/kg units

15,688 chemicals total
21,200 LD50 values

Preprocessing for modelling

11,992 chemicals
16,209 LD50 values

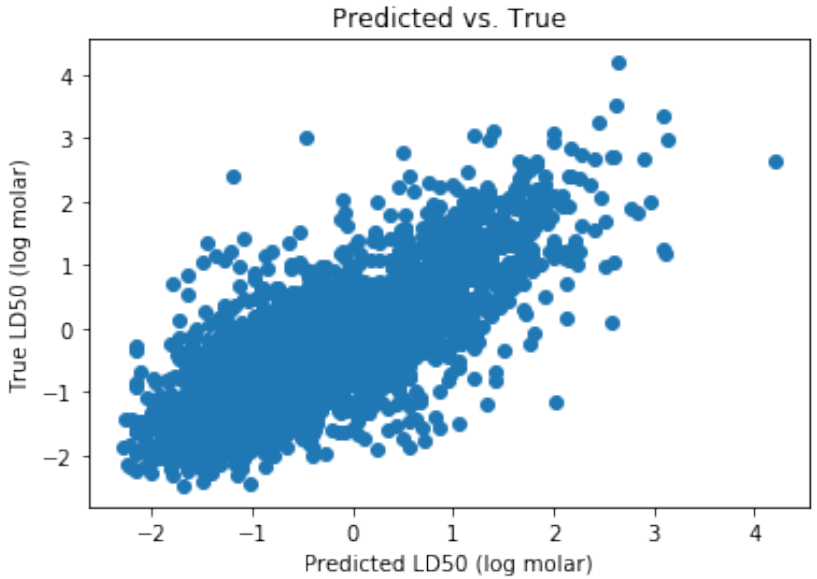
Exploratory Data Analysis

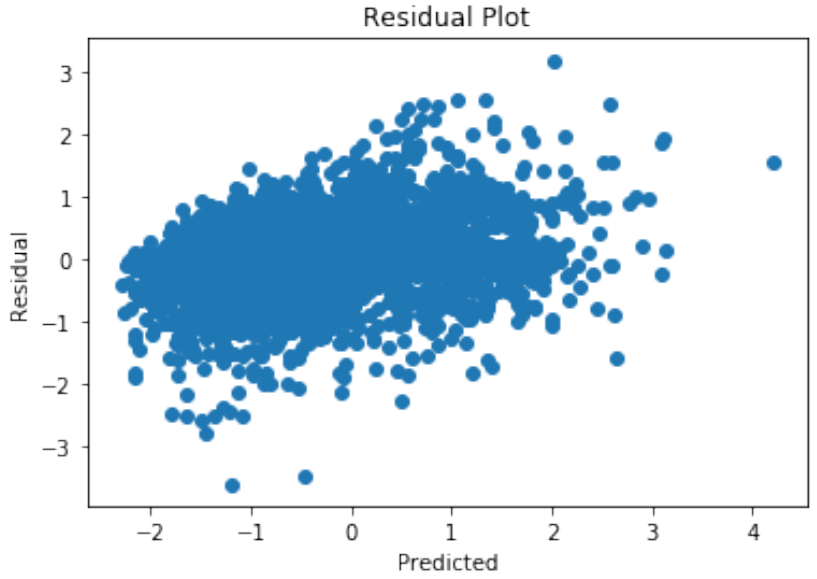


- Untransformed data highly skewed with extreme outliers
- Log molar transformation looks approximately normal

GenRA approach applied

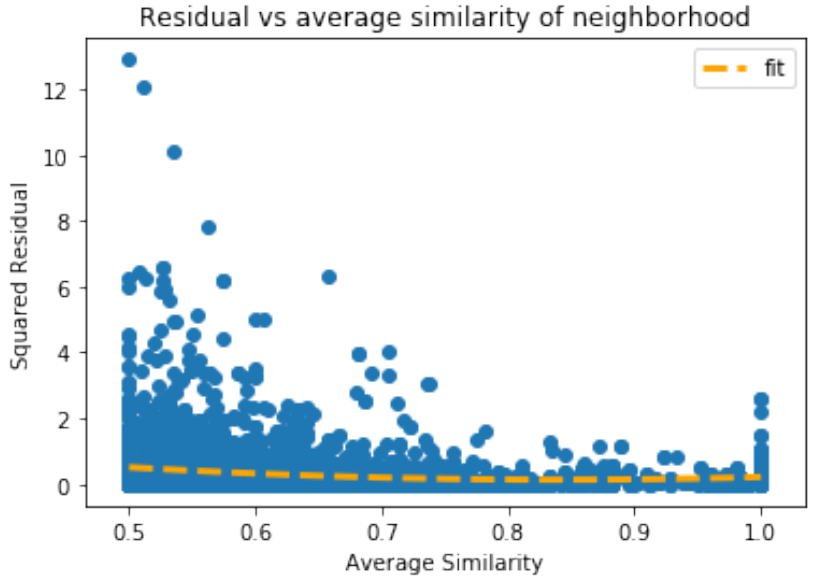
- Search for a maximum of 10 nearest neighbours on entire dataset
- Use a similarity threshold of 0.5

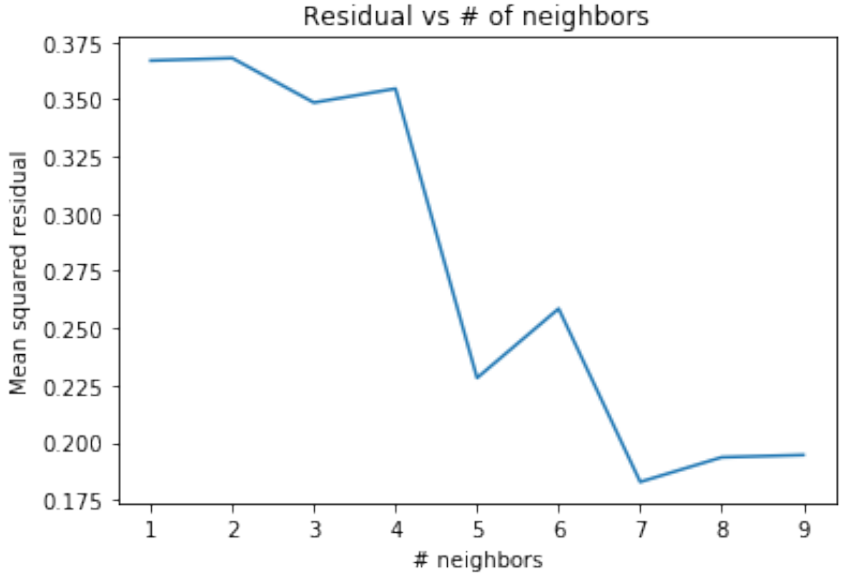




- R^2 = 0.61
- RMSE = 0.58
- A few outliers, but not too extreme
- Residuals clustered around zero with no obvious patterns

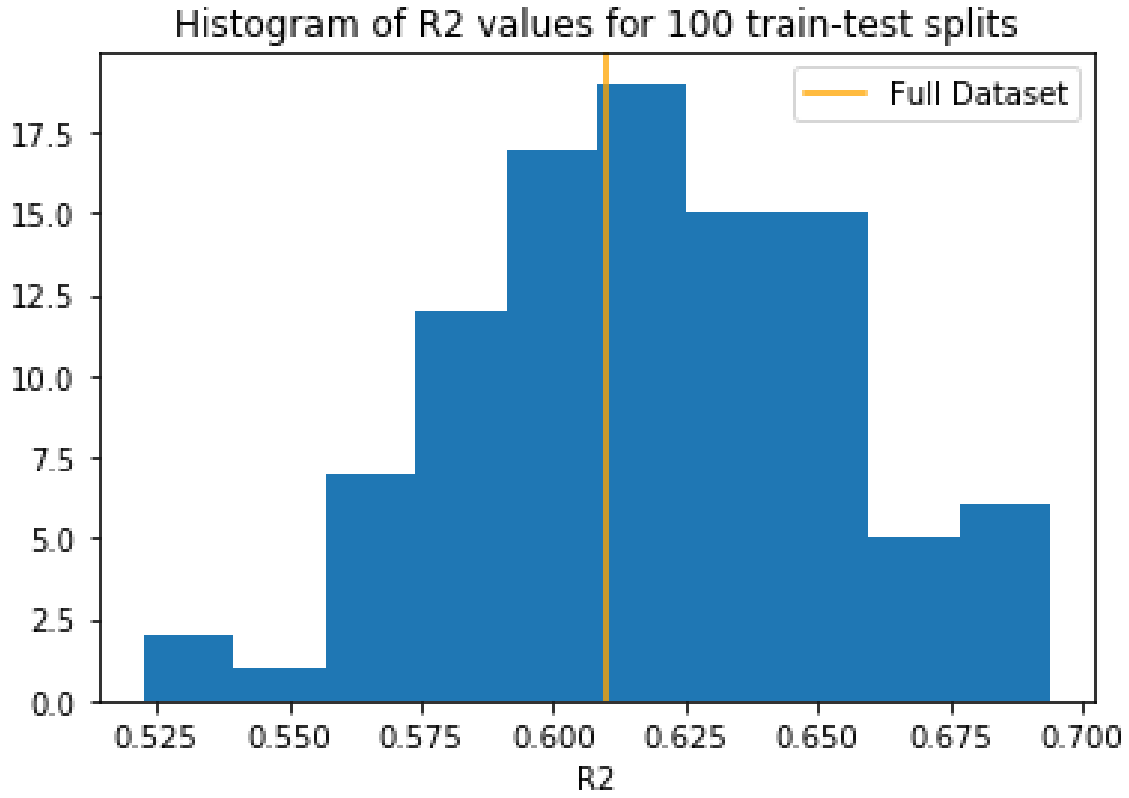
GenRA approach applied cont.





- Outliers tend to be for dissimilar neighbourhoods
- Increasing similarity of the neighbourhood leads to better predictions
- More neighbours in the neighbourhood also leads to better predictions.

Evaluation of the approach



- 75-25 train-test splits
- R^2 values range from 0.52 to 0.69
- GenRA performs strongly and robustly on this acute tox data set.

Helman et al., in preparation

Summary remarks

- Provided a perspective of the state of the science
- Outlined our research direction of read-across and how this fits within the context of the overall landscape of read-across
- Demonstrated the latest addition to the CompTox dashboard – GenRA
- Presented highlights of on-going analysis

Acknowledgements

- NCCT
- Imran Shah
- George Helman
- Prachi Pradeep
- Tony Williams
- Jeff Edwards
- Jeremy Dunne
- NCCT Development team
- Chris Grulke
- Reeder Sams
- Katie-Paul Friedman
- Rusty Thomas

- NCEA
- Jason Lambert
- Lucy Lizarraga

- Mark Cronin LJMU