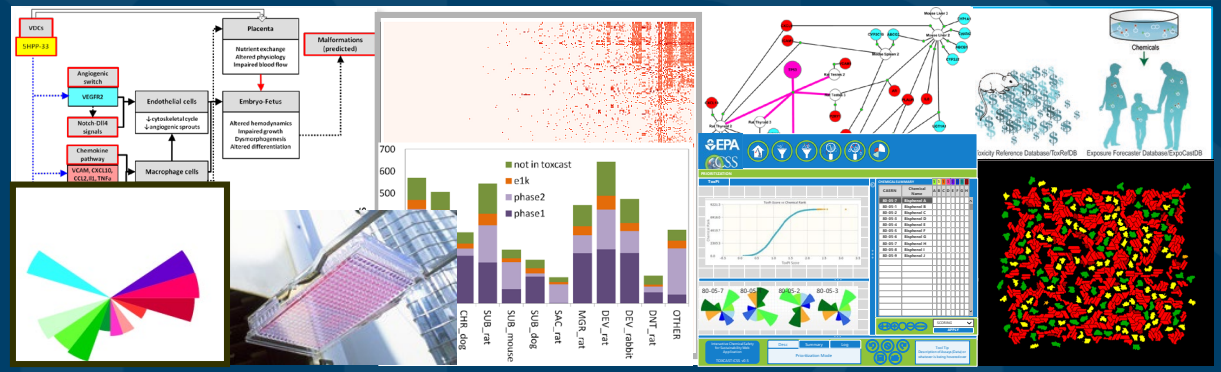
Transitioning GenRA to Quantitative Predictions: A Case Study Using ToxRefDB 2.0



George Helman[1,2], K Paul Friedman[2], Grace Patlewicz[2], Imran Shah[2]

[1]Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA
[2]National Center for Computational Toxicology, US EPA, RTP, NC, USA

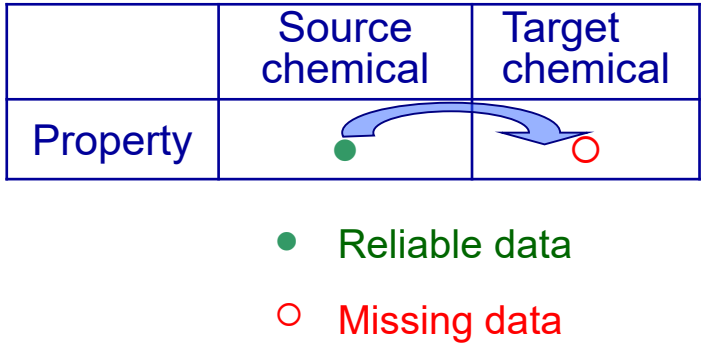
The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA

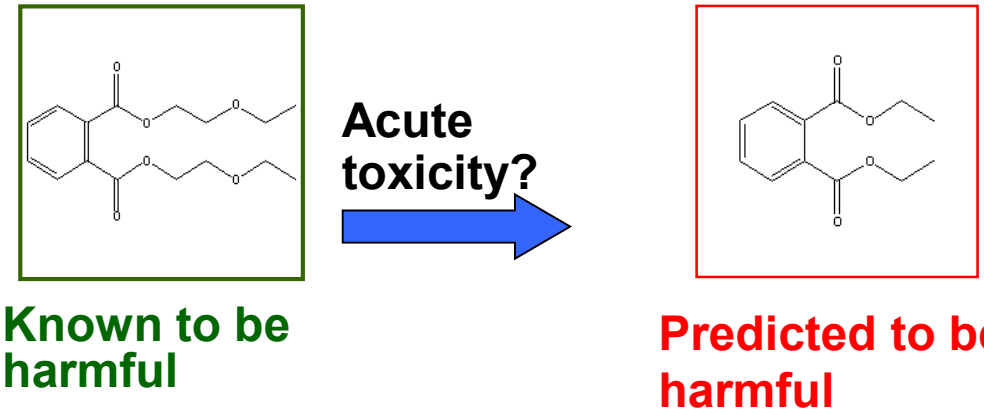
Outline

- Overview of approach
- Summary of ToxRef data
- Analysis
- Evaluation of predictions
- Future work + conclusions

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.





GenRA - Introduction

- GenRA (Generalized Read-Across) is a “local validity” approach predicting toxicity as a similarity-weighted activity of source analogues based on chemistry and/or bioactivity descriptors. (Shah et al, 2016)

- Generalized version of 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.

Methods

- GenRA is a similarity-weighted activity score of nearest neighbors

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

- Similarity calculated using Jaccard distance over Morgan chemical fingerprints
- Search for a maximum of 10 nearest neighbors on entire dataset.
- Use a similarity threshold of 0.5

Original Application

- Underlying data used was taken from ToxRefDB v1, a collection of repeated dose toxicity study types e.g. chronic, multigeneration, developmental, subchronic etc
- Toxicity effects within those study types were recorded as binary outcomes (0 for non-toxic, 1 for toxic)
- Toxicity effects were then predicted as binary outcomes (0 or 1)
- Dataset was clustered into local validity domains to find areas of chemical space where method performs best

Current Application

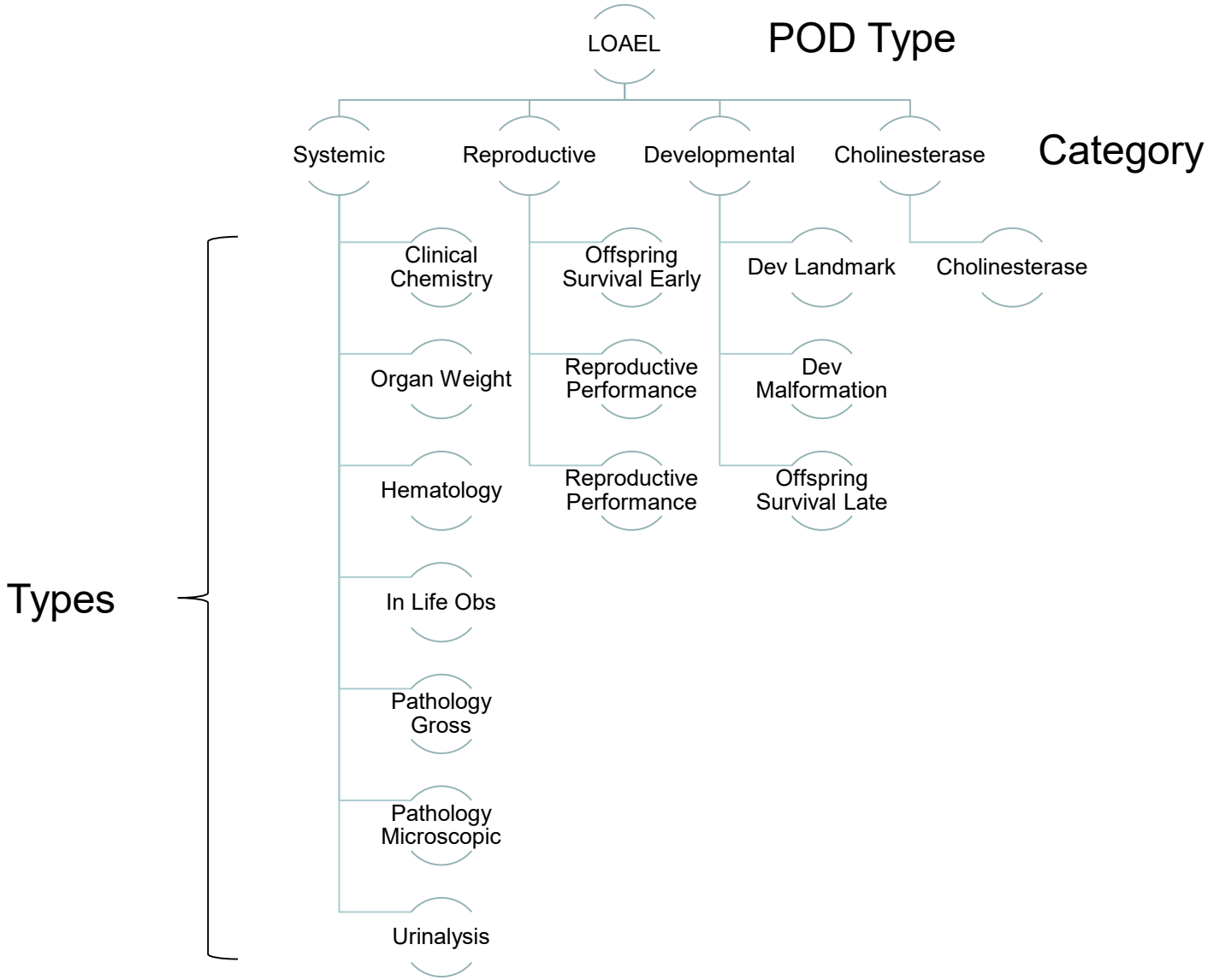
- We would like to test how GenRA performs on non-binary data using POD values from ToxRefDB 2.0.
- POD: Point of departure, or points on a dose-response curve corresponding to an observed effect level or no effect level
- 104,108 chemical level PODs across 1076 substances
- POD types: LOAEL (lowest observed adverse effect level), NOAEL (no observed adverse effect level), LEL (lowest effect level), NEL (no effect level)
- Endpoint categories: cholinesterase, developmental, reproductive, systemic
- 13 endpoint types
- 253 endpoint targets

EPA United States Environmental Protection Agency

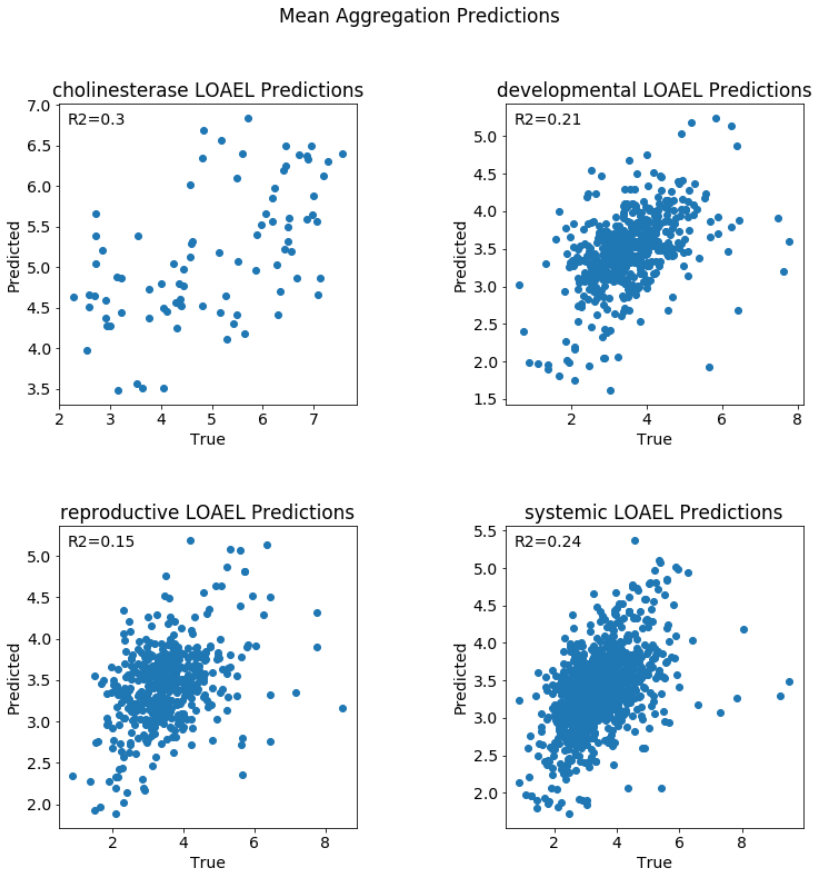
Approach

- We use GenRA to predict LOAEL values using Morgan fingerprints for similarity
- For chemicals that contain multiple LOAEL values, we aggregate them by taking the mean.
- We conduct a grid search over k (number of nearest neighbors) and s (similarity threshold) to find optimal values for R^2
- Cluster analysis was performed to find local neighborhoods of chemicals where approach performs particularly well.

Overview of ToxRef POD types



GenRA Predictions using Morgan fingerprints with k=10 and s=0.05

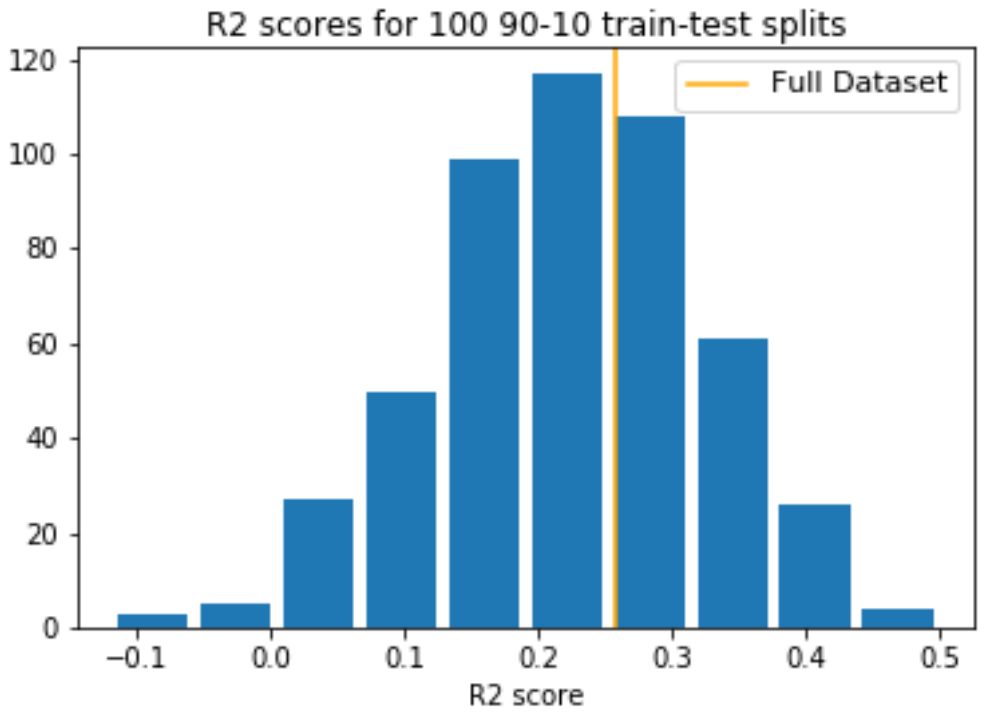


Endpoint Category	R2
Cholinesterase	0.26
Developmental	0.2
Reproductive	0.15
Systemic	0.24

Comparison with Other Methods

- Wignall et al 2018
  - Self-aggregated dataset of chronic toxicity values including RfDs, OSFs, CPVs, RfCs, and IURs for 2261 chemicals
  - Random forest regression
  - $0.2 < Q^2 < 0.45$ (depending on type of toxicity value)
- Helma et al 2018
  - Chronic rat LOAEL values for 826 chemicals
  - Local weighted random forest regression
  - $0.45 < R^2 < 0.47$ (cross validation)
- GenRA
  - LOAEL values for cholinesterase inhibition, developmental, reproductive, and systemic toxicity for 1064 chemicals
  - k-Nearest Neighbors with Morgan fingerprints
  - R2 = 0.24 (k=10, s=0.5, systemic endpoint)

Evaluation of GenRA Predictions



- Cross-validation testing
- Systemic endpoint
- 90-10 train-test splits
- R^2 values range from -0.04 to 0.43

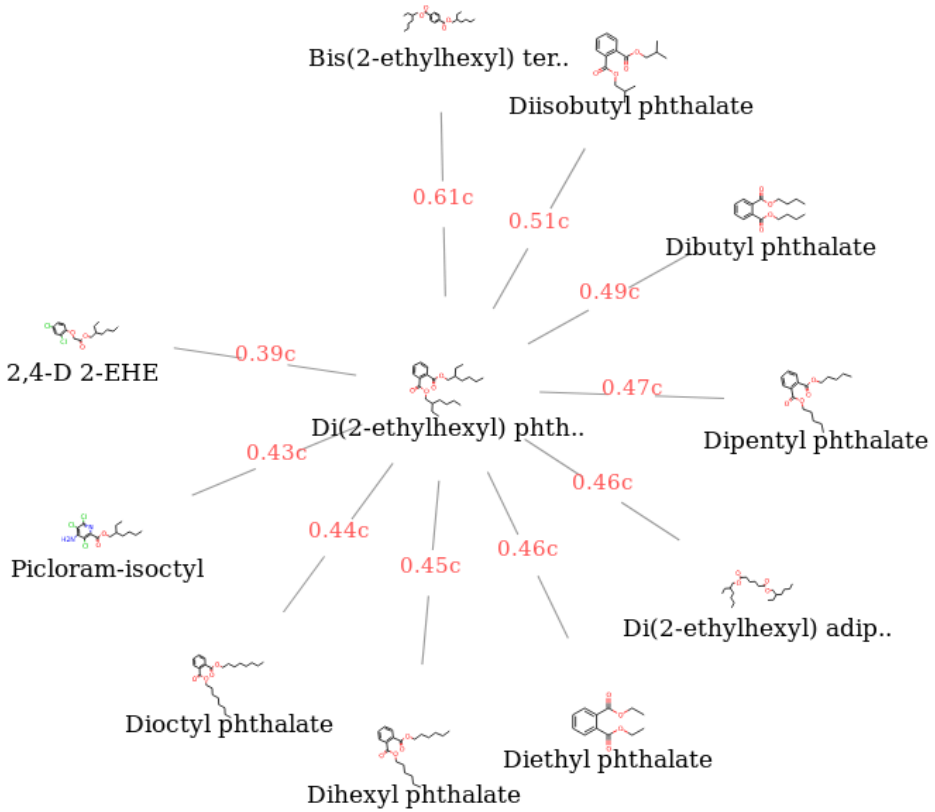
Example Predictions Di(2-ethylhexyl) phthalate

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



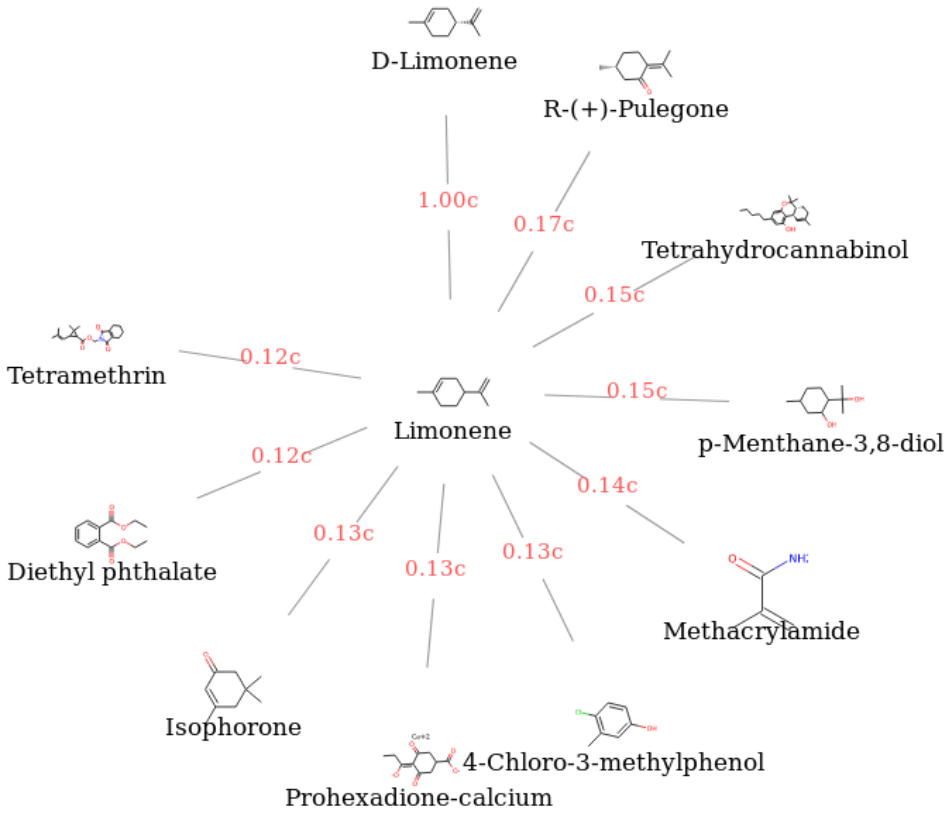
Example Predictions Limonene

Log Molar (log mol/kg/day)

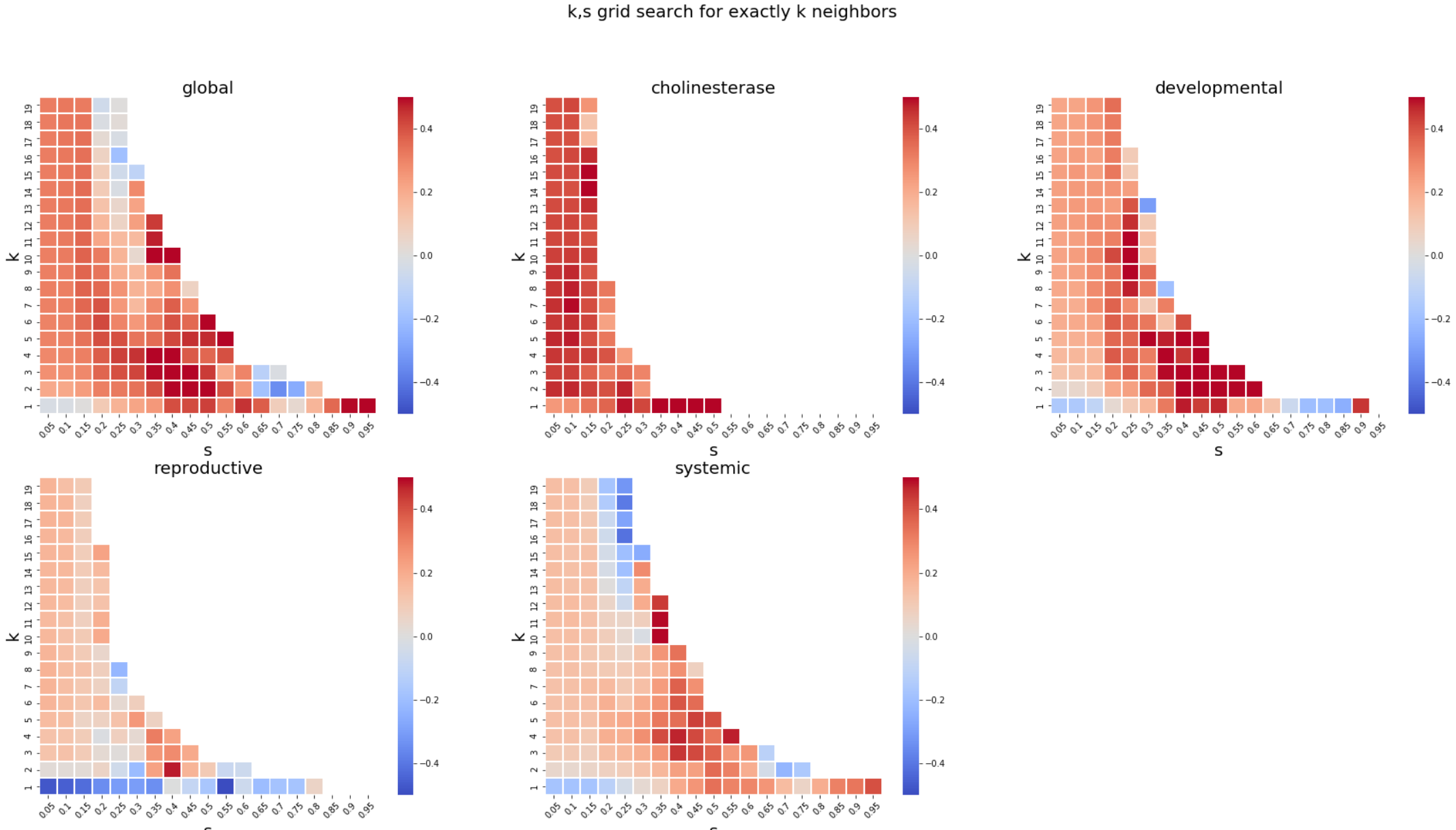
- Systemic prediction: 2.90
- Systemic measured: 2.44
- Developmental prediction: 3.86
- Developmental measured: 2.44

Mg/kg/day

- Systemic prediction: 172.45
- Systemic measured: 500.00
- Developmental prediction: 18.68
- Developmental measured: 500.00



Grid Search over k,s



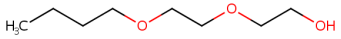
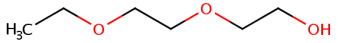
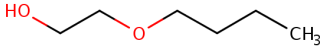
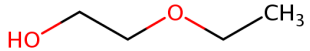
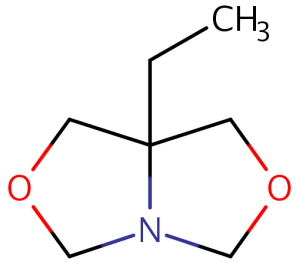
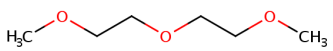
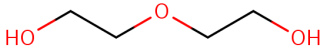
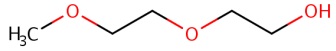
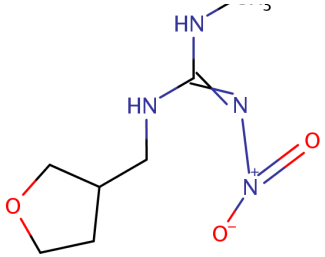
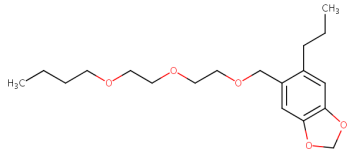
Cluster Analysis

- We try subsetting the data based on the chemical clusters discovered in the original GenRA manuscript in order to find local validity domains where GenRA predicts accurately

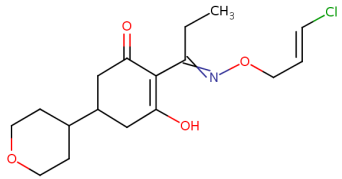
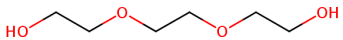
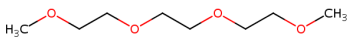
- Clusters discovered by k-means clustering

- We found 36/100 clusters that perform better than the global predictions by 3-fold on average.

Illustrative Example

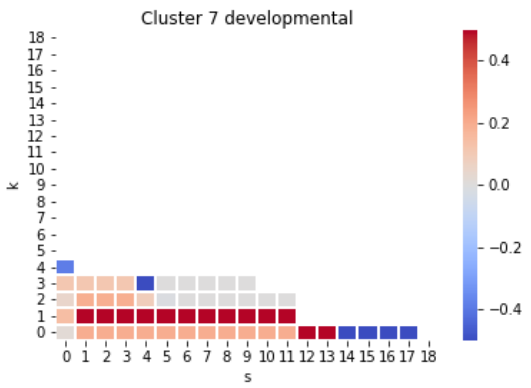
35 chemicals total,
mostly polyols and ethers

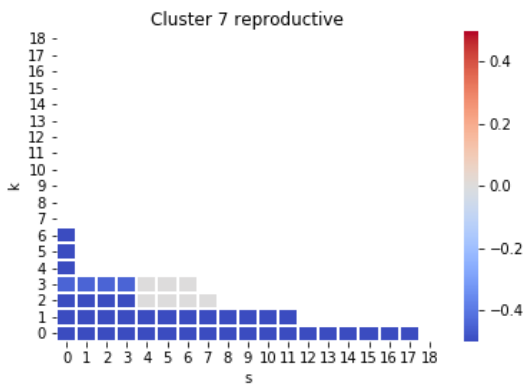




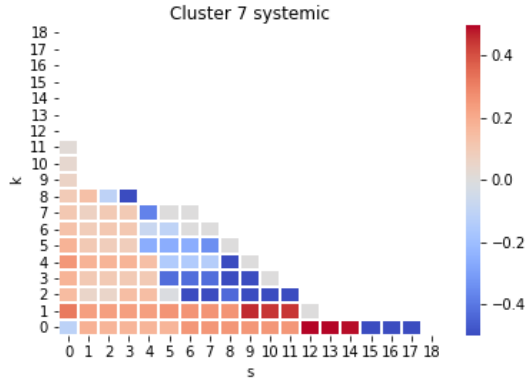
Illustrative example predictions

- Developmental LOAEL range: 120-11260 mg/kg
- Reproductive LOAEL range: 175-5175 mg/kg
- Systemic LOAEL range: 3-2795 mg/kg

- Developmental performance: R2 = 0.95 (k=1, s=0.65)
- Reproductive performance: R2 = 0.76 (k=7, s=0.20)
- Systemic performance: R2 = 0.73 (k=1, s=0.70)







Future Work + Conclusions

- We achieve a reasonable performance compared to other global methods

- Future Work
  - Use of different aggregations for consolidating multiple studies
  - Explore TTC (Threshold of Toxicological Concern) approach