Transitioning towards objective read-across approaches: Generalised Read-across (GenRA)

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

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

• Implementation of GenRA in the CompTox Chemicals Dashboard

• Refinement of the GenRA approach: work in progress & current applications

• Summary Remarks
• Acknowledgements
• References
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.

<table>
<thead>
<tr>
<th>Property</th>
<th>Source chemical</th>
<th>Target chemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute toxicity?</td>
<td><img src="image1.png" alt="Chemical Structure" /></td>
<td><img src="image2.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>Known to be harmful</td>
<td>Predicted to be harmful</td>
<td></td>
</tr>
<tr>
<td>Reliable data</td>
<td>○ Missing data</td>
<td></td>
</tr>
</tbody>
</table>
A harmonised hybrid read-across workflow

- Many frameworks/workflows for the development and evaluation of read-across e.g. OECD grouping guidance, ECHA R6, RAAF etc.
- Can be challenging to orient what the similarities and differences are between these workflows?
- Where new approach methods fit?
- How to scale up read-across approach for large numbers of substances in a systemic and reproducible manner?
Selected read-across tools

• Several ‘read-across’ tools exist
• How do they compare and contrast?
• Which tool or combination of tools can be used for a specific decision context?
## Selected read-across tools

<table>
<thead>
<tr>
<th>Tool</th>
<th>AIM</th>
<th>ToxMatch</th>
<th>AMBIT</th>
<th>OECD Toolbox</th>
<th>CBRA</th>
<th>ToxRead</th>
<th>GenRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analogue identification</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Analogue Evaluation</td>
<td>NA</td>
<td>X</td>
<td>X by other tools available</td>
<td>X</td>
<td>X</td>
<td>X For Ames &amp; BCF</td>
<td>NA</td>
</tr>
<tr>
<td>Data gap analysis</td>
<td>NA</td>
<td>X</td>
<td>X Data matrix can be exported</td>
<td>X Data matrix viewable</td>
<td>NA</td>
<td>NA</td>
<td>X Data matrix can be exported</td>
</tr>
<tr>
<td>Data gap filling</td>
<td>NA</td>
<td>X</td>
<td>User driven</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Uncertainty assessment</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>X</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Availability</td>
<td>Free</td>
<td>Free</td>
<td>Free</td>
<td>Free</td>
<td>Free</td>
<td>Free</td>
<td>Free</td>
</tr>
</tbody>
</table>
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

\[ y_i^{\beta, \alpha} = \frac{\sum_j^k S_{ij} \alpha x_j^{\beta}}{\sum_j^k S_{ij}} \]

Jaccard similarity:

\[ s_{ij} = \frac{\sum_l (x_{il} \land x_{jl})}{\sum_l (x_{il} \lor x_{jl})} \]

\[ \alpha \sqsubseteq \{chm, bio, bc\} \]

\[ \beta \sqsubseteq \{bio, tox\} \]

\( y_i = \) predicted activity of chemical (c_i)

\( x_j^\beta = \) activity of c_j in \( \beta \)

\( s_{ij}^\alpha = \) Jaccard similarity between \( x_i^\alpha, x_j^\alpha \)

\( k = \) up to k nearest neighbours
Read-across workflow in GenRA v1.0

Decision Context
Screening level assessment of hazard based on toxicity effects from ToxRefDB v1

Analogue identification
Similarity context is based on structural characteristics

Data gap analysis for target and source analogues

Uncertainty assessment
Assess prediction and uncertainty using AUC and p value metrics

Read-across
Similarity weighted average - many to one read-across

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

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

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

George Helman¹,², Imran Shah², Antony J. Williams³, Jeff Edwards¹, Jeremy Dunne³ and Grace Patlewicz²²

¹Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA; ²National 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.
• 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)...

  - 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

  - Analyses with quantitative data – acute rat oral toxicity (LD50 values), ToxRefDB v (LOAEL values)

  - Overall ‘global’ performance was reasonable – categories of chemicals could be readily identified where performance improved)

![Graphs showing true vs. predicted values for different endpoints](image)
GenRA - Applications

• Opportunities to refine how read-across is conducted within Provisional Peer review toxicity values (PPRTVs) assessments
• Exploring how GenRA can be used to inform read-across of PFAS within defined structural categories
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 & current applications
Acknowledgements

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

- Helman G, Shah I, Patlewicz G. Transitioning the Generalised Read-Across approach (GenRA) to quantitative predictions: A case study using acute oral toxicity data. *Computational Toxicology* 2019, *in press*