Navigating towards data-driven read-across approaches: Generalised Read Across (GenRA), a workflow module within the EPA CompTox Chemicals Dashboard

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

• Introduction to the EPA CompTox Chemicals Dashboard
• Read-across – existing resources
• Generalised Read-across (GenRA) approach
• GenRA implementation
• Practical application
• Ongoing research to enhance GenRA
• Acknowledgements
Earlier Dashboard Applications: Single architecture in development
EPA CompTox Chemicals Dashboard

• A publicly accessible website delivering access:
  - ~875,000 chemicals with related property data
  - Experimental and predicted physicochemical property data
  - Integration to “biological assay data” for 1000s of chemicals
  - Information regarding consumer products containing chemicals
  - Links to other agency websites and public data resources
  - “Literature” searches for chemicals using public resources
  - “Batch searching” for thousands of chemicals
  - DOWNLOADABLE Open Data for reuse and repurposing
CompTox Chemicals Dashboard: Landing Page

- Different entry points depending on domain of interest
CompTox Chemicals Dashboard:
Landing Page for a specific chemical

Bisphenol A
80-05-7 | DTXSID7020182
Searched by D3STox Substance Id.

Bisphenol A (BPA) is an organic synthetic compound with the chemical formula (CH₃)₄C₆H₄OH belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water. It has been in commercial use since 1937. BPA is a starting material for the synthesis of plastics, primarily.

Wikipedia

Intrinsic Properties

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes
CompTox Chemicals Dashboard: Executive Summary

Bisphenol A
80-05-7 | DTXSID7020182

Quantitative Risk Assessment Values
- IRIS values available
- No PPRITV values
- EPA RSL values available
- Minimum RID: 0.050 mg/kg-day (chronic, IRIS, oral, 8)
- No RIC calculated
- NOAEL: POD not calculated

Quantitative Hazard Values
- Minimum oral PCO: 3.8 mg/kg-day (reproductive, HPVIS, oral, 6)
- No inhalation POD values
- Lowest Observed Toxicity Equivalent Level: CYP1A1, CYP1A2, Tpo, ESR2, ESR1, ESR1, NR113, PPARA, NR112, Cyp2c11, MMP3, Er1

Cancer Information
- No cancer slope factor
- No inhalation unit risk value
- Carcinogenicity data available: University of Maryland carcinogenicity warning
- No genotoxicity (findings reported)

Reproductive Toxicology
- 200 Reproductive toxicity PDEs available

Regional Screening

<table>
<thead>
<tr>
<th>Class</th>
<th>THQ</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>risk-based SSL (mg/kg)</td>
<td>THQ = 0.1</td>
<td>5.8</td>
</tr>
<tr>
<td>GIABS (unspecified)</td>
<td>THQ = 1</td>
<td>1</td>
</tr>
<tr>
<td>GIABS (unspecified)</td>
<td>THQ = 0.1</td>
<td>1</td>
</tr>
<tr>
<td>Abs (unspecified)</td>
<td>THQ = 0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>RPEDO (mg/kg-day)</td>
<td>THQ = 0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>screening level (residential)</td>
<td>THQ = 0.1</td>
<td>320</td>
</tr>
<tr>
<td>screening level (industrial)</td>
<td>THQ = 0.1</td>
<td>4100</td>
</tr>
</tbody>
</table>
Generalised Read-Across (GenRA) as a workflow module
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>Reliable data</td>
<td><img src="source-chemical.png" alt="Chemical Structure" /></td>
<td><img src="target-chemical.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>Missing data</td>
<td><img src="source-chemical.png" alt="Chemical Structure" /></td>
<td><img src="target-chemical.png" alt="Chemical Structure" /></td>
</tr>
</tbody>
</table>

- Acute toxicity?
  - Known to be harmful
  - Predicted to be harmful
Frameworks for developing category/analogue approaches

OECD (2014)

Wu et al, 2010

Patlewicz et al, 2015
Frameworks for the assessment of read-across

READ ACROSS UNCERTAINTY EVALUATION QUESTIONNAIRE FOR:
Target chemical (SOI) = (list CAS#)

INSTRUCTIONS
Complete the Questionnaire. Answer the questions for each endpoint where SAR was conducted, and follow instructions listed in each section below. (In general, NO responses indicate potential areas of uncertainty in the proposed read across.)

<table>
<thead>
<tr>
<th>Questions</th>
<th>Responses by Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Repeat Dose Toxicity</td>
</tr>
</tbody>
</table>

Section 1: Chemical similarity between source (analogues) and target (SOI)

1. For each endpoint, list the CAS# of the source (analogues) contributing the critical study for the read across for each endpoint.
   - CAS#

2. What is the ‘suitability rating’ of the analogue?
   - Suitable
   - Unsuitable
   - Skip to next
   - (skip to next
   - (continue of the intake

3. Are any differences in functional groups and associated activities (if yes, be more reactive than the target)?
   - YES
   - NO
   - UNKNOWN
   - No Differences

Table 2
Scientific confidence considerations in read-across evaluation.

<table>
<thead>
<tr>
<th>Data issues</th>
<th>Similarity rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analogue/category approach</td>
<td>Similarity rationale/hypothesis that underpins the analogue/category approach</td>
</tr>
<tr>
<td>Quality of data for source analogues – e.g. Klimisch scores of 1 or 2</td>
<td>Concordance of effects and potency (if relevant) per endpoint</td>
</tr>
<tr>
<td>Completeness of data matrix – No of data gaps e.g. source analogue(s) have many data points to address, target substance has a handful of data gaps.</td>
<td>Presence or absence of adverse effects</td>
</tr>
<tr>
<td>Concordance of any available anchor data</td>
<td>Type of read-across (Qualitative, Quantitative, Trend Analysis)</td>
</tr>
<tr>
<td>Concordance of effects and potency (if relevant) across endpoints</td>
<td>Rationalization of why structural differences do not impact the toxicity</td>
</tr>
<tr>
<td></td>
<td>Structural similarity</td>
</tr>
<tr>
<td></td>
<td>Analogue validity</td>
</tr>
<tr>
<td></td>
<td>Analogue similarity with respect to general and endpoint specific considerations</td>
</tr>
</tbody>
</table>

Blackburn and Stuard (2014)

Read-across resources: Selected read-across tools

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

Grace Patlewicz, George Helman, Prachi Pradeep, Imran Shah

*National Center for Computational Toxicology (NCTR), Office of Research and Development, US Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27709, USA.
Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA

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

Grace Patlewicz\textsuperscript{a,*}, Mark T.D. Cronin\textsuperscript{b}, George Helman\textsuperscript{a, c}, Jason C. Lambert\textsuperscript{d}, Lucina E. Lizarraga\textsuperscript{d}, Imran Shah\textsuperscript{a}

\textsuperscript{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

\textsuperscript{b} School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

\textsuperscript{c} Oak Ridge Institute for Science and Education (ORISE), 1299 Bethel Valley Road, Oak Ridge, TN 37830, USA

\textsuperscript{d} National Center for Evaluation Assessment (NCEA), US Environmental Protection Agency (US EPA), 25 West Martin Luther King Dr, Cincinnati, OH 45268, USA
Read-across resources: Selected read-across examples/decision contexts

Exploring current read-across applications and needs among selected U.S. Federal Agencies

Grace Patlewicz"a, Lucina E. Lizarraga"a, Diego Rua"b, David G. Allen"b, Amber B. Daniel"b, Suzanne C. Fitzpatrick"b, Natalia Garcia-Reyero"a, John Gordon"b, Perti Hakkinen"b, Angela S. Howard"b, Agnes Karmaus"b, Joanna Matheson"a, Moiz Mumtaz"b, Andrea-Nicole Richardz"b, Patricia Ruiz"b, Louis Scarrano"b, Takashi Yamada"a, Nicole Kleinsteuber"b

"National Center for Computational Toxicology, U.S. Environmental Protection Agency, 1801 Alexander Dr., Research Triangle Park, NC, 27709, USA
"National Center for Environmental Assessment, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, OH, 45268, USA
"Center for Devices and Radiological Health, U.S. Food and Drug Administration, 10900 New Hampshire Avenue, Silver Spring, MD, 20902, USA
"EPA, P.O. Box 1325, Research Triangle Park, NC, 27709, USA
"Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, 5100 Paint Branch Parkway, College Park, MD, 20740, USA
"Environmental Laboratory, U.S. Army Engineer Research and Development Center, 2900 Blue Ridge Rd., Valdese, NC, 28690, USA
"U.S. Consumer Product Safety Commission, 5 Research Place, Rockville, MD, 20850, USA
"National Library of Medicine, 6707 Democracy Blvd., Bethesda, MD, 20892, USA
"Agency for Toxic Substances and Disease Registry, 1600 Clifton Rd., Atlanta, GA, 30333, USA
"Environmental Protection Agency, 1200 Pennsylvania Ave. NW, Washington, DC, 20460, USA
"Division of Risk Assessment, Biological Safety Research Center, National Institute of Health Sciences, 3-25-23, Toyama-ku, Kawasaki, Kawasaki, 210-0582, Japan
"National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, P.O. Box 12333, Research Triangle Park, NC, 27709, USA

More than just a 'REACH' regulatory context
ICCVAM Read-Across Workgroup

Predicting estrogen receptor activation by a group of substituted phenols: An integrated approach to testing and assessment case study

Francina Webster †, Matthew Gagné, Grace Patlewicz, Prachi Pradeep, Nicholas Trefiak, Richard S. Judson, Tara S. Barton-Maclaren

ICCVAM Read-Across Workgroup

OECD IATA Case study

Traditional approaches for chemical risk assessment cannot keep pace with the number of substances requiring assessment. Thus, in a global effort to expedite and moderate chemical risk assessment, New Approach Methodologies (NAMs) are being explored and developed. Included in this effort is the OECD Integrated Approaches for Testing and Assessment (IATA) program, which provides a forum for OECD member countries to develop and present case studies illustrating the application of NAM in various risk assessment contexts. Here, we present an IATA case study for the prediction of estrogenic potential of three target phenols: 4-tert-butyphenol, 2,4-di-tert-butyphenol and ortho-tert-butyphenol. Key features of this IATA include the use of two computational approaches for analogue selection for read-across, data collected from traditional and NAM sources, and a workflow to generate predictions regarding the target’s ability to bind the estrogen receptor (ER). Endocrine disruption may occur when a chemical substance mimics the activity of natural estrogen by binding to the ER and, if potency and exposure are sufficient, alters the function of the endocrine system to cause adverse effects. The data indicated that of the three target substances that were considered herein, 4-tert-butyphenol is a potential endocrine disruptor. Further, this IATA illustrates that the NAM approach explored is health protective when compared to its in vivo endpoints traditionally used for human health risk assessment.
A harmonised hybrid read-across workflow

- Where do other NAM fit?
- How should we transition to data-driven approaches? Limit subjectivity
- What about characterising the uncertainty of the predictions made?
- Generalisability/Scalability of read-across - coverage of read-across for inventories of chemicals?

Patlewicz et al., 2018
# 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</td>
<td>X by other tools available</td>
<td>X</td>
<td>X</td>
<td>X For Ames &amp; BCF</td>
</tr>
<tr>
<td>Data gap analysis</td>
<td>NA</td>
<td>X</td>
<td>X</td>
<td>X Data matrix can be exported</td>
<td>X Data matrix viewable</td>
<td>NA</td>
<td>NA</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>

- AIM: Analgo Identification Model
- ToxMatch: ToxMatch Toolbox
- AMBIT: Analysis of Mammalian Biotransformation
- OECD Toolbox: Organisation for Economic Co-operation and Development Toolbox
- CBRA: Chemicals Behaviour Resource Agent
- ToxRead: ToxRead
- GenRA: General Read-Across Assessment
GenRA

- **GenRA (Generalised Read-Across)**
- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and bioactivity descriptors
- Systematically evaluates read-across performance and uncertainty using available data

Jaccard similarity:
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

Read-across workflow in GenRA
GenRA tool in practice
GenRA tool in practice
GenRA tool in practice

Data gap analysis
GenRA tool in practice
GenRA in practice – step by step

- Analogue identification:
- Similarity based on Morgan chemical fingerprints and selecting a default of 10 source analogues
GenRA in practice – step by step

- Data matrix view of source analogues relative to target chemical
GenRA in practice – step by step

Look for commonality in profile across target effects

What are the most common effects across analogues
GenRA in practice - step by step

• Updated Data matrix view with GenRA predictions for target chemical

• Predictions are binary (yes/no) for toxicity effects within ToxRefDB v1 studies.
• Predictions summarised on a study level basis. Red: “positive” and Blue: “negative”.
GenRA in practice: Approach

Short Communication

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

George Helman1,4, Imran Shah2, Antony J. Williams3, Jeff Edwards3, Jeremy Dunne3 and Grace Palewicz*

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., 2018; 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 – Ongoing research

- Summarising and aggregating the toxicity effect predictions to guide end users
- 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
  - Quantifying the impact of reactivity similarity on read-across performance
GenRA – Ongoing research

• 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 (e.g. effect level or point-departure [NOAEL, LOAEL, etc.] predictions)

• Using quantitative data from acute rat oral toxicity, ToxRefDB v2 [1 manuscript submitted, 1 in internal clearance]
Take home messages

• Harmonised framework for read-across provides opportunities for expanded integration of NAM data

• GenRA developed is aligned with this framework

• Initial GenRA (baseline) considers structural similarity but current work has evaluated the quantitative impact of physicochemical similarity (as it relates to bioavailability) and transitioned to quantitative predictions of effect levels or PODs

• Illustrated how GenRA baseline can been applied in practice
Acknowledgements

• Many but in particular..
• NCCT
• Imran Shah
• George Helman
• Tony Williams
• Jeff Edwards
• Jason Lambert

• NCEA
• Lucy Lizarraga