

Novel Methods for Estimating NOAEL Confidence Bounds and Optimizing Similarity Measures for Read-Across Workflows



C. Yang^{1,2*}; J.F. Rathman^{1,2}; T. Magdziarz¹; A. Mostrag¹
T. Kawamoto³; E.H. Theophilus⁴



¹ MN-AM, Nürnberg, Germany & Columbus, OH, USA; ² The Ohio State University, Columbus, OH, USA; ³ Kao Corporation, Tokyo, Japan; ⁴ Shiseido Americas Corporation, East Windsor, NJ, USA

*Chihai.yang@mn-am.com

OBJECTIVES

- Compare usefulness of different fingerprints in identifying similarity or dissimilarity of structures.
- Demonstrate that different thresholds are needed for various fingerprints when finding similar structures for a given query
 - Variety of fingerprints may be needed for handling different structures
- Test hypothesis that analogues with higher similarity scores (or analog quality) have similar toxic effects as target in Read-Across.
- Estimate the NOAEL bounds of a target structure based on data from a high-quality analogue.

FINGERPRINTS - COVERAGE

- A fingerprint set is suitable for a given collection of structures if all structures contain at least one fingerprint feature (100% coverage) and if the number of structural hits increases at a moderate rate with increasing number of randomly-sampled fingerprints.
- ToxPrints cover 90% at ~30% sampling, while the RDKit TFs cover 90% at ~1% sampling but without reaching 100% (838/839).

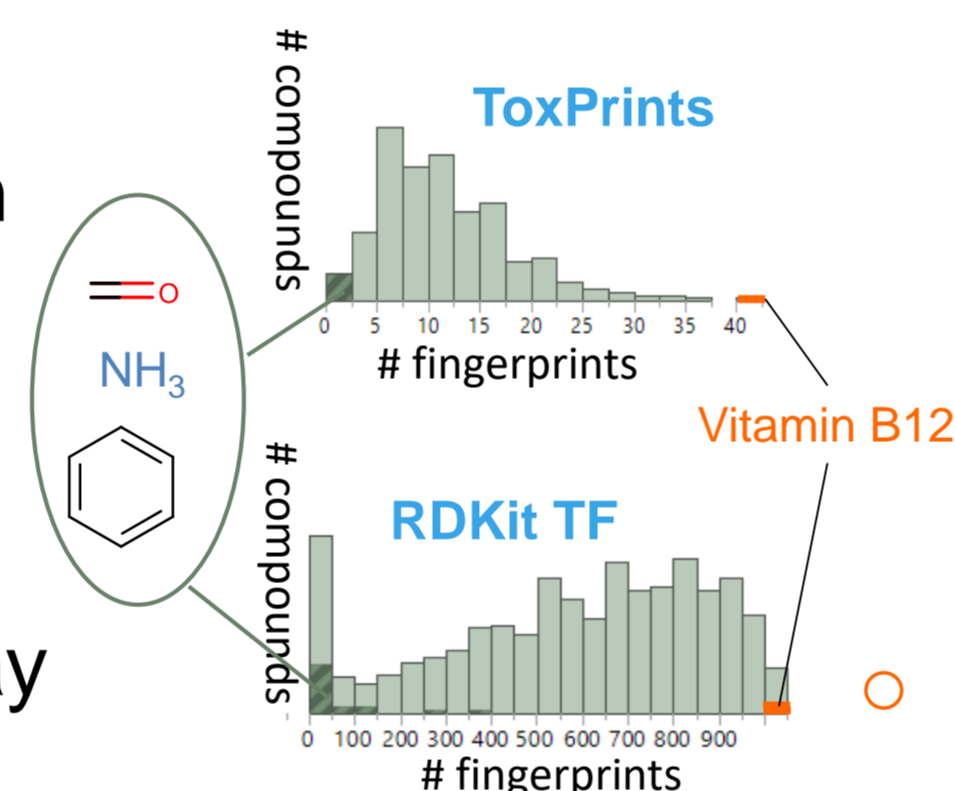
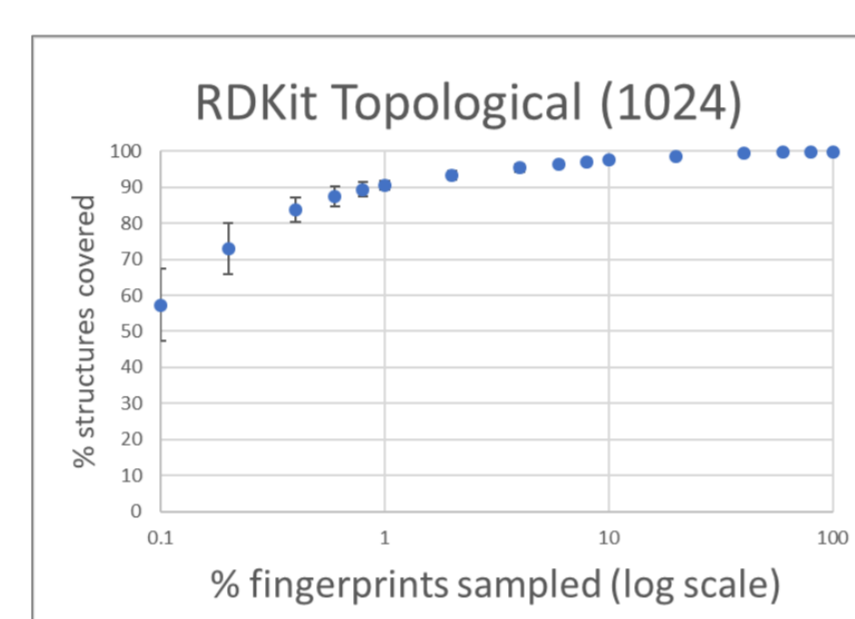
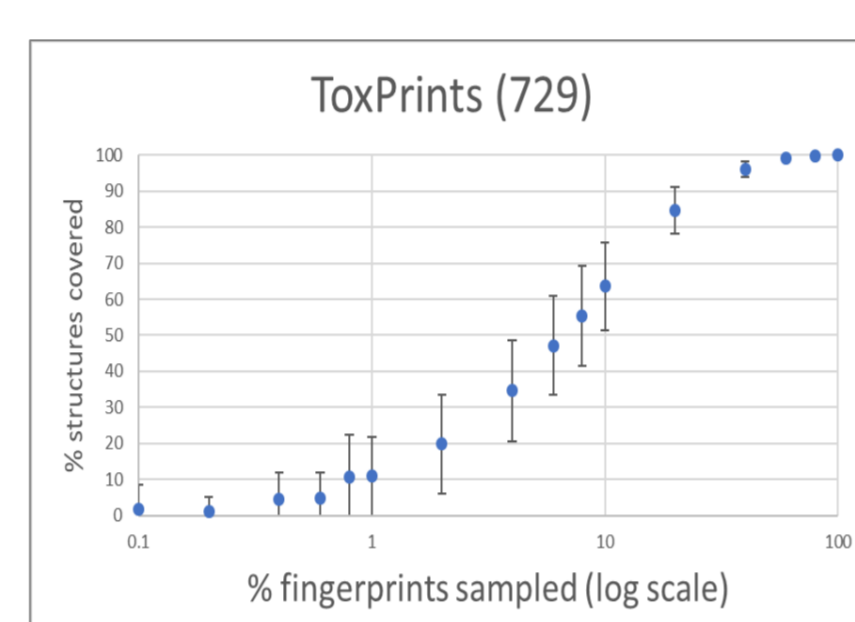
FINGERPRINTS - INFORMATION DENSITY

- A feature is considered more discriminating if it can distinguish subtle variations in structure. This resolution power is important for mechanistic interpretations.
- A small feature matching many diverse structures may not provide useful information.
- Features appearing only in a few specific structures may be discriminatory with high information density, but not relevant in most cases.

FINGERPRINTS

- RDKit Topological Fingerprints (RDKit TFs)
 - <https://www.rdkit.org/docs/>
 - Topological paths along bonds (1-7 bonds), 2048 bits (1024 fingerprints)
 - Similar to Daylight fingerprints
- ToxPrint* Fingerprints (ToxPrints)
 - <https://chemotyper.org/>
 - Molecular fragments encoding connectivity and topology, and properties of atoms, bonds and electronic systems.
 - Mechanistically designed for chemical reactivity and toxicity
 - Features designed to reflect categories.

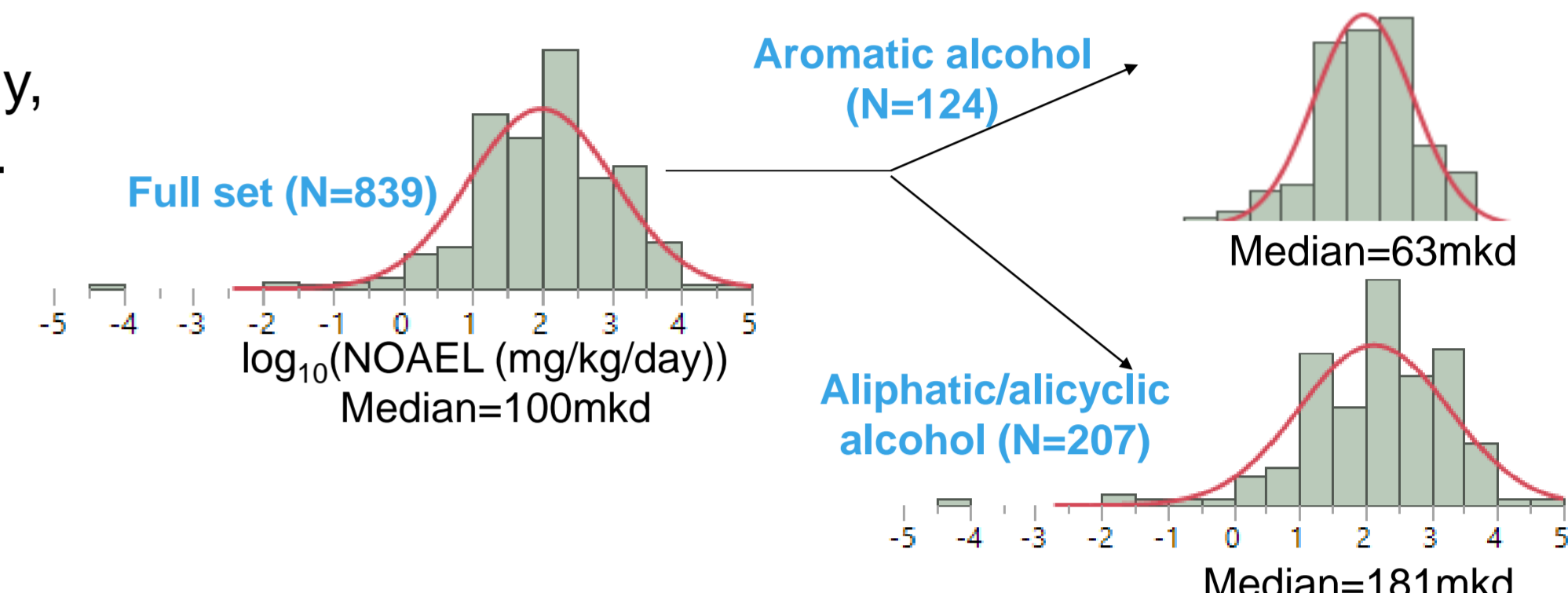
*C. Yang et al. J. Chem. Inf. Model. 2015, 55, 510–52.



SIMILARITY

NOAEL DATASET

- Cosmetics and food ingredients from COSMOS TTC and oRepeatTox (SCCS, EFSA, FDA CFSAN, and ECHA)
- A highly curated NOAEL dataset on repeated dose (RDT) and systemic effects from development & reproductive (DART) studies (934 test substances; 839 structures)



HYPOTHESIS

- “Analog quality” is determined from structural similarity calculated from fingerprints, profiles of relevant properties, and biological assays.
- The “analog quality” score, calculated from similarity measures relevant to the toxicity effect of interest, can be used to identify good analogue candidates for Read-Across.
- NOAEL values of analogues within a related category will generally be more similar.
- The uncertainty of the target NOAEL varies inversely with analog quality

NOAEL

NOAEL RANGE ESTIMATION

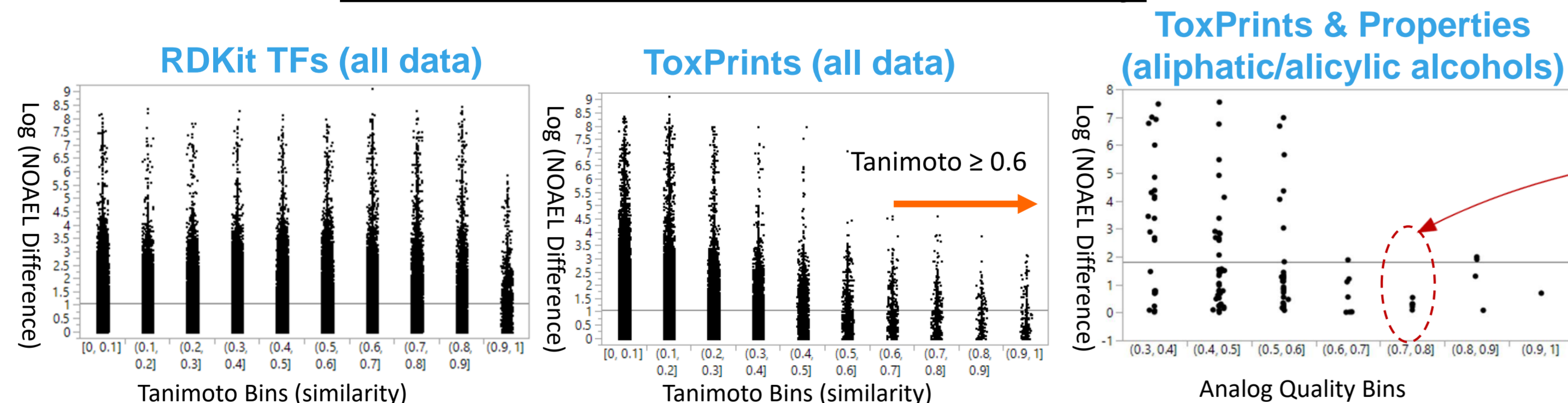
- NOAEL range is estimated by analyzing the quantiles in a specific similarity bin of the “NOAEL Difference vs. Similarity” plot. A confidence level and the interval for “NOAEL Difference” is determined.
- As an example, if the target-analogue pair has an analogue quality of **0.74**, the distribution of the **(0.7-0.8]** bin will be used, as in the case study below.

PAIR-WISE DISTRIBUTION ANALYSIS

- Similarities (Tanimoto coefficients) were calculated using RDKit TF or ToxPrints structure-based fingerprints for all pairs of structures in the COSMOS NOAEL set (839).
- The absolute difference between the NOAEL values for structures in each pair is quantified by

$$NOAEL\ diff = |\log_{10}(NOAEL_2) - \log_{10}(NOAEL_1)| = \left| \log_{10} \left(\frac{NOAEL_2}{NOAEL_1} \right) \right| \quad \text{Eq (1)}$$

NOAEL Difference vs. Similarity



- Differences between NOAEL values tend to be smaller for pairs of structures that are more similar. This is consistent with our hypothesis.
- Compared to RDKit TF, ToxPrints exhibit remarkably clear thresholds where the pairwise NOAEL differences become smaller. We thus chose ToxPrints for use in the Read-Across case study.

Considering Properties in Analogue Quality

- Property-based (Pearson correlation) similarities were calculated and combined with the Structure-based similarities for all pairs of structures.
- Overall similarities were obtained by averaging the structure- and property-based values for each pair:

$$analog\ quality = \sqrt{(Tanimoto)(Pearson + 1)} / 2 \quad \text{Eq (2)}$$

CASE STUDY: dihydro- α -terpineol

Target	Analogue 1	
<chem>CC1(C)CC(C)CC1</chem>	<chem>CC1(C)CC(C)CC1</chem>	
Analogue quality* 0.74		
Chronic RDT Study	Description	Rat, oral, 730 days
	Outcome	NOAEL = 750 mkd [†]
	Reliability	High (90% reliable)

- Target and analogue both belong to the alicyclic (non-primary) alcohol category
- Structure-only similarity between target (dihydro- α -terpineol) and analog-1 (menthol) is 0.8 by RDKit TFs and 0.5 by ToxPrints.
- Analogue quality is 0.74, based on structure- and property-based similarity, as well as study reliability.

[†] Properties: H-bond acceptors & donors, rotational bonds, number of atoms, MW, polarizability, complexity, diameter, volume, logS, logP, TPSA, heat formation, HOMO, LUMO

	Distribution: Structure Category	Distribution: Full Dataset
NOAEL-diff @ 90% Confidence	$\left \log_{10} \left(\frac{NOAEL_2}{NOAEL_1} \right) \right \leq 0.54$	$\left \log_{10} \left(\frac{NOAEL_2}{NOAEL_1} \right) \right \leq 1.92$
Analogue-1 NOAEL data	750 mg/kg-bw/day	750 mg/kg-bw/day
90% CI NOAEL range for target	216 – 2600 mg/kg-bw/day	9 – 62,000 mg/kg-bw/day

CONCLUSION

- Similarity thresholds meaningful for selecting high-quality analogues depends on types of fingerprints as well as the dataset.
- By applying probability bounds to a distribution of NOAEL values, a novel method was developed for estimating the limits of NOAEL values for a target compound based on the NOAEL value of an analogue and the analog quality.
- Performing this analysis within data subsets defined by structural or MOA categories provides tighter confidence intervals for the NOAEL of the target.