lazar read across models for lowest adverse effect levels: A comparison of experimental variability with read across predictions

Introduction

Methods

Data

Mazzatorta dataset

Swiss dataset

Preprocessing

Missing and invalid SMILES Unfortunately no identifier is complete across all compound therefore we focused on SMILES. Missing SMILES were generated from other identifiers when available.

study type/ table

rat_chron mouse_chron multigen missing SMILES 35 27 31 invalid SMILES 9 6 9 corrected SMILES 44 33 40 Detailed tables:

https://docs.google.com/spreadsheets/d/14P8F-3iX5gr5FbN7oSeuwabUOr_xdDhhr5KwiUX6LXY/edit?usp=sharing

Dataset comparison

Structural composition

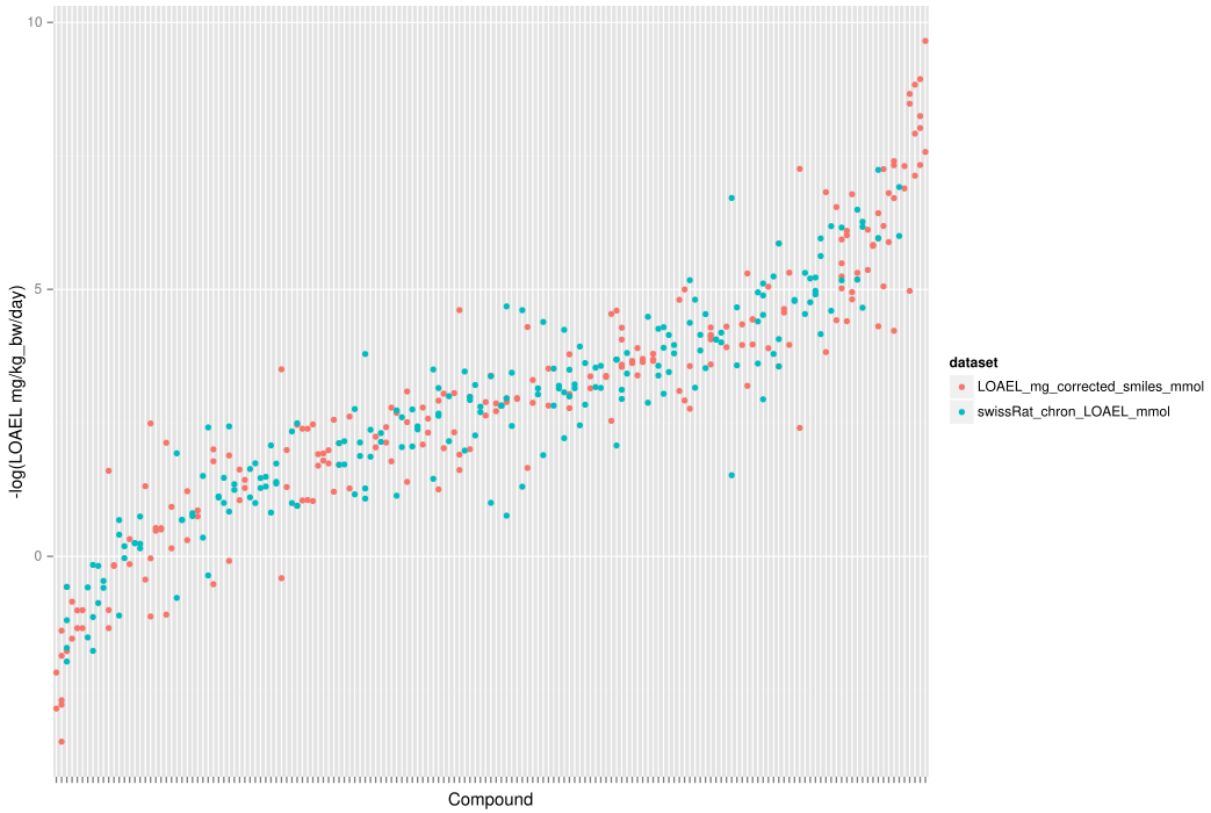
Ches-Mapper analysis

Distribution of functional groups

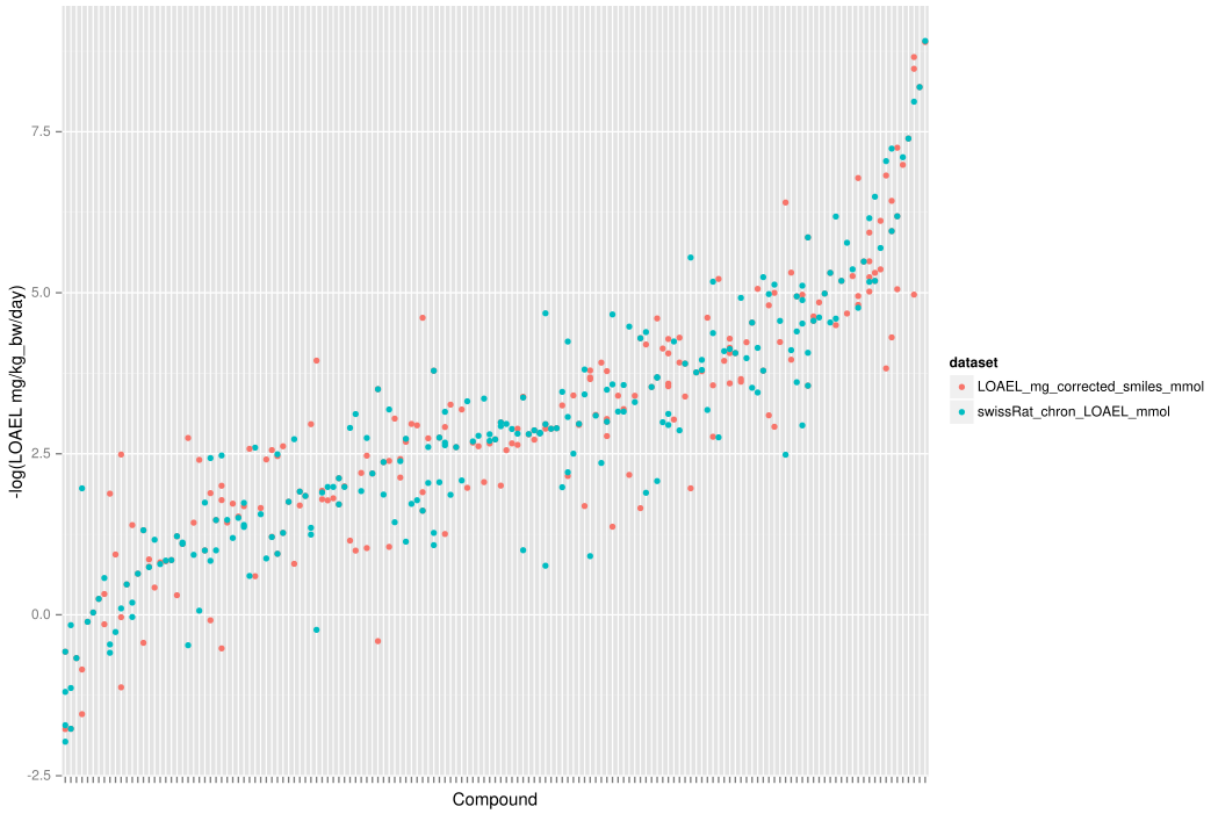
LOAEL values

Intra dataset variability

p-value: 0.4750771581019402

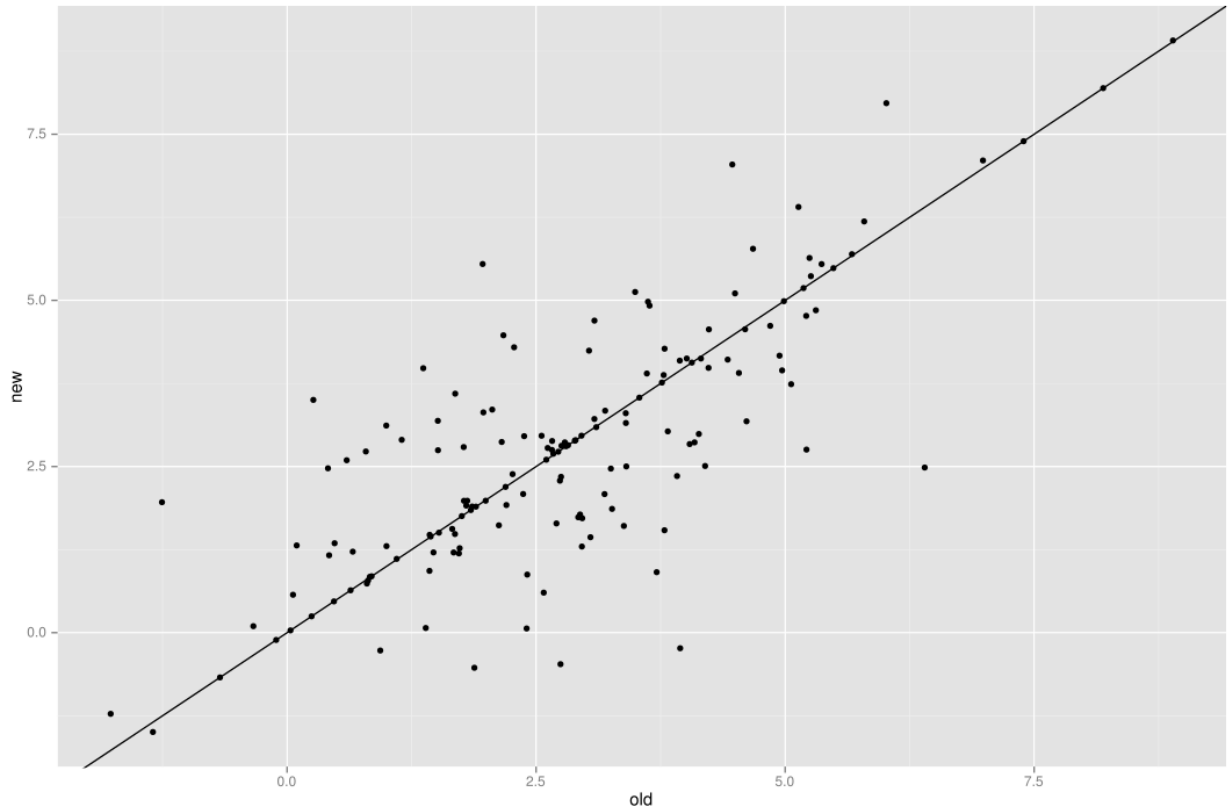


Inter dataset variability



Correlation between datasets

using means



with "identical" values

r^2: 0.6106457754533314 RMSE: 1.2228212261024438 MAE: 0.801626064534318

Algorithms

Fingerprints

- OB Fingerprints (add MNA)
- Fingerprint counts
- Physchem Descriptors (OB only?)

Feature selection

- none
- t-test (nonparam?) (qual)
- correlation (quant)

Similarity calculation

- tanimoto
- weighted tanimoto
- cosine
- wighted cosine

Regression

- weighted majority vote
- local linear regression