	QMRF identifier (JRC Inventory): **Q17-452-0061**
	QMRF Title: **Lazar model for human maximum recommended daily dose**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Lazar model for human maximum recommended daily dose

1.2.Other related models:

1.3.Software coding the model:

lazar

lazar Lazy Structure- Activity Relationships. See

https://github.com/opentox/lazar/tree/b2b9cc5d14d10a62025486c5b3abbbeb06bf0ec6

info@in-silico.ch

https://lazar.in-silico.ch

2.General information

2.1.Date of QMRF:

21 September 2017

2.2.QMRF author(s) and contact details:

Christoph Helma in silico toxicology gmbh Rastatterstr. 41, CH-4057 Basel info@in-silico.ch www.in-silico.ch

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Christoph Helma in silico toxicology gmbh Rastatterstr. 41, CH-4057 Basel info@in-silico.ch www.in-silico.ch

2.6.Date of model development and/or publication:

2017

2.7.Reference(s) to main scientific papers and/or software package:

[1]Maunz A, Gütlein M, Rautenberg M, Vorgrimmler D, Gebele D and Helma C (2013) lazar: a modular predictive toxicology framework. Front. Pharmacol. 4:38

http://dx.doi.org/10.3389/fphar.2013.00038

[2]Helma C, Gebele D, Rautenberg M (2017) Lazar, software available at https://lazar.in-silico.ch,source code available at

https://github.com/opentox/lazar/tree/b2b9cc5d14d10a62025486c5b3abbbeb06bf0ec6

https://doi.org/10.5281/zenodo.215483

2.8.Availability of information about the model:

Prediction interface and validation results available at https://lazar.in-silico.ch

2.9.Availability of another QMRF for exactly the same model:

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Human

3.2.Endpoint:

4.14.Repeated dose toxicity 452Chronic Toxicity Studies

3.3.Comment on endpoint:

3.4.Endpoint units:

mmol/kg-bw/day

3.5.Dependent variable:

Maximum Recommended Daily Dose

3.6.Experimental protocol:

3.7.Endpoint data quality and variability:

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Regression

4.2.Explicit algorithm:

modified k-nearest neighbour regression (local random forest), see https://github.com/opentox/lazar/tree/b2b9cc5d14d10a62025486c5b3abbbeb06bf0ec6

4.3.Descriptors in the model:

MP2D fingerprints (Bender et al. 2004)

4.4.Descriptor selection:

Correlation with dependent variable (Pearson p <= 0.05)

4.5.Algorithm and descriptor generation:

lazar

4.6.Software name and version for descriptor generation:

lazar, submitted version: https://github.com/opentox/lazar/tree/b2b9cc5d14d10a62025486c5b3abbbeb06bf0ec6

4.7.Chemicals/Descriptors ratio:

variable (local regression models)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

No predictions are made for query compounds without similar structures in the training data. Similarity is determined as the Tanimoto coefficient of Molprint 2D fingerprints with a threshold of 0.2. Predictions based on a low number and/or very dissimilar neighbours or on neighbours with conflicting experimental measurements should be treated with caution.

5.2.Method used to assess the applicability domain:

Number and similarity of training set compounds (part of the main lazar algorithm): At least 2 compounds with similarity > 0.2 are required for a prediction. A warning is issued, if the closest neighbour has a similarity < 0.5.

5.3.Software name and version for applicability domain assessment:

lazar, submitted version: https://github.com/opentox/lazar/tree/b2b9cc5d14d10a62025486c5b3abbbeb06bf0ec6

5.4.Limits of applicability:

Compounds without similar substances in the training dataset

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula: Yes

INChI: Yes

MOL file: Yes

6.3.Data for each descriptor variable for the training set:

No

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

Original data from: http://www.epa.gov/comptox/dsstox/sdf_fdamdd.html

6.6.Pre-processing of data before modelling:

-log10 transformation

6.7.Statistics for goodness-of-fit:

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

3 independent 10-fold crossvalidations:

Num folds: 10

Num instances: 1214

Num unpredicted: 252

RMSE: 0.835

MAE: 0.598

R^2: 0.542

Num folds: 10

Num instances: 1214

Num unpredicted: 245

RMSE: 0.836

MAE: 0.596

R^2: 0.545

Num folds: 10

Num instances: 1214

Num unpredicted: 252

RMSE: 0.844MAE: 0.599

R^2: 0.536

6.10.Robustness - Statistics obtained by Y-scrambling:

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

No

7.2.Available information for the external validation set:

CAS RN:

Chemical Name:

Smiles:

Formula:

INChI:

MOL file:

7.3.Data for each descriptor variable for the external validation set:

Unknown

7.4.Data for the dependent variable for the external validation set:

Unknown

7.5.Other information about the external validation set:

7.6.Experimental design of test set:

7.7.Predictivity - Statistics obtained by external validation:

7.8.Predictivity - Assessment of the external validation set:

7.9.Comments on the external validation of the model:

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Compounds with similar structures (neighbours) are assumed to have similar activities as the query compound.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori for individual predictions

8.3.Other information about the mechanistic interpretation:

Hypothesis about biochemical mechanisms can be derived from individual predictions by inspecting neighbours and relevant descriptors.
Neighbors are compounds that are similar in respect to a certain endpoint and it is likely that compounds with high similarity act by similar mechanisms as the query compound. Links at the webinterface prove an easy access to additional experimental data and literature citations for the neighbours and the query structure.
Please note that lazar predictions are based on neighbours. Descriptors are only used for the calculation of similarities.

9.Miscellaneous information

9.1.Comments:

Public model interface: https://lazar.in-silico.ch
Source code: https://github.com/opentox/lazar/tree/b2b9cc5d14d10a62025486c5b3abbbeb06bf0ec6
Docker image: https://hub.docker.com/r/insilicotox/lazar/

9.2.Bibliography:

[1]Helma C, Gebele D, Rautenberg M (2017) Lazar, software available at https://lazar.in-silico.ch,source code available at https://github.com/opentox/lazar/tree/b2b9cc5d14d10a62025486c5b3abbbeb06bf0ec6

https://doi.org/10.5281/zenodo.215483
[2]Helma C, Rautenberg M and Gebele D (2017) Nano-Lazar: Read across Predictions for Nanoparticle Toxicities with Calculated and Measured Properties. Front. Pharmacol. 8:377 https://dx.doi.org/10.3389%2Ffphar.2017.00377
[3]Lo Piparo et al (2014), Automated and reproducible read-across like models for predicting carcinogenic potency. Regulatory Toxicology and Pharmacology.70 (1) 370-378 https://doi.org/10.1016/j.yrtph.2014.07.010
[4]Maunz A and Helma C (2008). Prediction of chemical toxicity with local support vector regression and activity-specific kernels. SAR and QSAR in Environmental Research 19 (5-6) 413-431 http://dx.doi.org/10.1080/10629360802358430
[5]Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Molecular Diversity 10 (2), 147–158 http://dx.doi.org/10.1007/s11030-005-9001-5
[6]Bender et al (2004) Molecular similarity searching using atom environments, information-based feature selection, and a nave bayesian classifier. J. Chem. Inf. Comput. Sci. 44 (1) 170–178 https://doi.org/10.1021/ci034207y

9.3.Supporting information:

Training set(s)Test set(s)Supporting information

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q17-452-0061

10.2.Publication date:

2017-10-10

10.3.Keywords:

Maximum Recommended Daily Dose;human;Lazar;

10.4.Comments:

To be entered by JRC