QMRF Title: Oral ToxiciTy Reporter (OTTR).

Date: 2nd May 2025.

1. QSAR Identifier

1.1 QSAR Identifier (Title): Oral ToxiciTy Reporter (OTTR): Prediction of acute oral toxicity.

1.2 Other related models: Dow models for the cholinergic targets (nAChR, mAChR, AChE) and mitochondrial respiration.

1.3 Software coding the model: Random Forest machine-learning model built in KNIME using conjoint MACCS and mechanistic fingerprints as the sole covariate.

2. General Information

2.1 Date of QMRF: 2nd May 2025.

2.2 QMRF author(s) and contact details: Sanjeeva J. Wijeyesakere and M. Sue Marty. The Dow Chemical Company, Midland MI.

2.3 Date of QMRF update(s): N/A.

2.4 QMRF updates(s): N/A.

2.5 Model developers and contact details: Sanjeeva J. Wijeyesakere, Tyler Auernhammer, Amanda Parks and Dan Wilson. The Dow Chemical Company, Midland MI.

2.6 Date of model development and/or publication: March 2023.

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

Wijeyesakere, S. J., Auernhammer, T., Parks, A. & Wilson, D. Profiling mechanisms that drive acute oral toxicity in mammals and its prediction via machine learning. Toxicol Sci 193, 18–30 (2023). DOI: 10.1093/toxsci/kfad025

2.8 Availability of information about the model: Data curation and model assessment are detailed in the above-mentioned publication (Wijeyesakere et al., 2023; listed in section 2.7).

2.9 Availability of another QMRF for exactly the same model: No.

3. Defining the endpoint - OECD Principle 1

3.1 Species: Predictions are made for the acute oral toxicity of the test compound in the rat.

3.2 Endpoint: Acute oral toxicity (Mechanism, LD_{50} and associated GHS classification).

3.3 Comment on endpoint:

3.4 Endpoint units: The model makes binary (yes/no) predictions to align compounds to a defined mechanism followed by a prediction of its acute oral toxicity (LD_{50}) in mg/kg bw. GHS classification is based on defined LD_{50} cut-offs and is unitless.

3.5 Dependent variable: *In vivo* rat acute oral toxicity (LD_{50}) data.

3.6 Experimental protocol: *In vivo* acute oral toxicity.

3.7 Endpoint data quality and variability: Data curation and sources are detailed in the relevant publication (Wijeyesakere et al., 2023; listed in section 2.7).

4. Defining the algorithm - OECD Principle 2

4.1 Type of model: Machine-learning model that predicts acute oral toxicity based on a custom structural/mechanistic fingerprint (detailed in the relevant publication (Wijeyesakere et al., 2023; listed in section 2.7)).

4.2 Explicit algorithm: Random-forest machine learning model as implemented in KNIME ver. 4.6.1.

4.3 Descriptors in the model: 2D MACCS structural fingerprints conjoint with mechanistic fingerprint.

4.4 Descriptor selection: *A priori* selection.

4.5 Algorithm and descriptor generation: Calculated within KNIME using the CDK fingerprints node which are conjoint with the custom mechanistic fingerprint.

4.6 Software name and version for descriptor generation: KNIME ver. 5.4.3.

4.7 Chemicals/Descriptors ratio: N/A.

5. Defining the applicability domain - OECD Principle 3

5.1 Description of the applicability domain of the model: Single organic molecules with defined 2D structures not including As, B, Sn, Si, Zn, Cd, Zr, Al, Hg, Pt, Au, Ag, Pb, Bi, Ti, Os, Mo, Se, Po, Fe, Ni, Sb, V, Cr, Co, In, Te, Ru, Sc, Re, Cu and having molecular weights ≤ 1500 Da (based on properties of compounds in training set).

5.2 Method used to assess the applicability domain: Manually defined by the model developer based on chemicals in training set.

5.3 Software name and version for applicability domain assessment: N/A.

5.4 Limits of applicability: N/A.

5. Internal validation - OECD Principle 4

6.1 Availability of the training set: Yes – 6,234 chemicals available in the cited publication (Wijeyesakere et al., 2023; listed in section 2.7).

6.2 Available information for the training set:

Chemical identifier: Yes – CAS and/or DTXSID, InChI Key and/or Name (when available).

SMILES: Yes.

LD_{50}: Yes

GHS Classification: Yes

Formula: No.

InChI: No.

MOL file: No.

6.3 Data for each descriptor variable for the training set: MACCS fingerprints can be calculated in the CDK node from the provided SMILES structure.

6.4 Data for the dependent variable for the training set: Rat acute oral LD_{50} (mg/kg bw)

6.5 Other information about the training set: Detailed in the cited publication (Wijeyesakere et al., 2023; listed in section 2.7).

6.6 Pre-processing of data before modelling: Canonicalization of SMILES structures, aromatization of rings and calculation of MACCS and conjoint MACCS/mechanistic fingerprints for each structure based on its [SMILES] structure.

6.7 Statistics for goodness-of-fit: N/A

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

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

Model built with 90% of actives and controls, with 10% being reserved as a test set (repeated 5 times) for identifying highly acutely toxic compounds (those with LD_{50} values ≤ 300 mg/kg bw):

Concordance measure	Mean value
Sensitivity	76.6%
Specificity	70.8%
Balanced accuracy	73.7%

6.10 Robustness - Statistics obtained by Y-scrambling: N/A.

6.11 Robustness - Statistics obtained by bootstrap:

Concordance statistics from production model using all curated data (out-of-bag estimates) for identifying highly acutely toxic compounds (those with LD_{50} values ≤ 300 mg/kg bw):

Concordance measure	Mean value
Sensitivity	93.0%
Specificity	69.3%
Balanced accuracy	81.2%

6.12 Robustness - Statistics obtained by other methods: N/A.

7 External validation - OECD Principle 4

7.1 Availability of the external validation set: Yes (dataset of ω-fluoro alcohols with mouse acute oral LD_{50} data available in cited publication listed above (section 2.7; Wijeyesakere *et al.,* 2023).

7.2 Available information for the external validation set:

Chemical identifier: CAS and Name.

SMILES: Yes.

Formula: Yes.

INChI: No.

MOL file: No.

7.3 Data for each descriptor variable for the external validation set: No.

7.4 Data for the dependent variable for the external validation set: Mouse acute oral LD_{50} (mg/kg bw).

7.5 Other information about the external validation set: N/A.

7.6 Experimental design of test set: N/A.

7.7 Predictivity - Statistics obtained by external validation: N/A: Qualitative Assessment of model against ω-fluoro alcohols is detailed in the cited publication (Wijeyesakere et al., 2023; listed in section 2.7).

7.8 Predictivity - Assessment of the external validation set: N/A.

7.9 Comments on the external validation of the model: Model correctly flagged ω-fluoro alcohols predicted to act via an aconitase mode of action, with high acute toxicity ($LD_{50} \leq 50$ mg/kg bw) (detailed in the cited publication (Wijeyesakere et al., 2023; listed in section 2.7))).

8 Providing a mechanistic interpretation - OECD Principle 5

8.1 Mechanistic basis of the model: Mechanistic sub-models are trained on *in vitro* bioactivity data and can be used to align novel compounds to a mechanism of toxicity. These can then be combined with in vivo LD_{50} data to develop a model to predict acute oral LD_{50} estimates.

8.2 A priori or a posteriori mechanistic interpretation: The mechanistic basis of the model was developed *a priori* by manually reviewing the compiled *in vivo* data to assign a mechanism prior to developing the machine-learning model.

8.3 Other information about the mechanistic interpretation: All references supporting the development and mechanistic basis for the model are detailed in the references detailed above (Wijeyesakere et al., 2023; listed in section 2.7).

9 Miscellaneous information

9.1 Comments: This model can be used to screen large databases or single compounds for their acute oral lethality potential. As detailed in the cited publication (Wijeyesakere *et al.,* 2023), this model is designed to optimize sensitivity to minimize the likelihood of "false-negative" predictions.

9.2 Bibliography:

Wijeyesakere, S. J., Auernhammer, T., Parks, A. & Wilson, D. Profiling mechanisms that drive acute oral toxicity in mammals and its prediction via machine learning. Toxicol Sci 193, 18–30 (2023). DOI: 10.1093/toxsci/kfad025