Leadscope Enterprise version of commercial MultiCASE MC4PC model A49 for teratogenic potential in humans

1. QSAR identifier

1.1 QSAR identifier (title)

Leadscope Enterprise version of commercial MultiCASE MC4PC model A49 for teratogenic potential in humans, Danish QSAR Group at DTU Food.

1.2 Other related models

MultiCASE CASE Ultra version of commercial MultiCASE MC4PC model A49 for teratogenic potential in humans, Danish QSAR Group at DTU Food.

SciMatics SciQSAR version of commercial MultiCASE MC4PC model A49 for teratogenic potential in humans, Danish QSAR Group at DTU Food.

1.3. Software coding the model

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

2. General information

2.1 Date of QMRF

January 2015.

2.2 QMRF author(s) and contact details

QSAR Group at DTU Food;

Danish National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Marianne Dybdahl;

National Food Institute at the Technical University of Denmark;

Sine Abildgaard Rosenberg;

National Food Institute at the Technical University of Denmark;

2.3 Date of QMRF update(s)

2.4 QMRF update(s)

2.5 Model developer(s) and contact details

MultiCASE Inc.;

23811 Chagrin Blvd Ste 305, Beachwood, OH, 44122, USA;

www.multicase.com

MultiCASE Inc. has kindly given their permission that remodelling of their training set from the commercial MC4PC A49 model in Leadscope was performed by:

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Danish QSAR Group at DTU Food;

National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

2.6 Date of model development and/or publication

January 2014.

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

Roberts, G., Myatt, G. J., Johnson, W. P., Cross, K. P., and Blower, P. E. J. (2000) LeadScope: Software for Exploring Large Sets of Screening Data. *Chem. Inf. Comput. Sci.*, 40, 1302-1314.

Cross, K.P., Myatt, G., Yang, C., Fligner, M.A., Verducci, J.S., and Blower, P.E. Jr. (2003) Finding Discriminating Structural Features by Reassembling Common Building Blocks. *J. Med. Chem.*, *46,* 4770-4775.

Valerio, L. G., Yang, C., Arvidson, K. B., and Kruhlak, N. L. (2010) A structural feature-based computational approach for toxicology predictions. *Expert Opin. Drug Metab. Toxicol.*, 6:4, 505-518.

2.8 Availability of information about the model

The training set is commercially available from MultiCASE Inc. The training set was originally compiled by MultiCASE Inc. and used to train their commercial MultiCASE MC4PC A49 model. The Danish QSAR Group bought this model from MultiCASE Inc. in 1999. Permission to remodel the training set in Leadscope was kindly granted by MultiCASE Inc. The model algorithm is proprietary from commercial software.

2.9 Availability of another QMRF for exactly the same model

3. Defining the endpoint

3.1 Species

Human (based on epidemiological, clinical and animal data).

3.2 Endpoint

QMRF 4. Human Health Effects

QMRF 4.16. In vivo pre-natal-developmental toxicity

3.3 Comment on endpoint

It was previously believed that the mammalian embryo, developed in the impervious uterus of the mother, was protected from all extrinsic factors. However, after the thalidomide disaster of the 1960s, it became apparent and more accepted that the developing embryo and fetus could be highly vulnerable to certain environmental agents that have negligible or non-toxic effects to adult individuals.
Birth defects are known to occur in 3-5 % of newborns. It is estimated that approximately 10% of all birth defects are caused by prenatal exposure to a teratogenic agent. A teratogen is defined as an agent that causes malformations of the embryo or fetus *in utero* and include radiation, infections, maternal metabolic imbalances (e.g. diabetes, folic acid deficiency), drugs (e.g. anticancer drugs, tetracyclines, many hormones, thalidomide), and environmental chemicals (e.g. mercury, lead, dioxins PBDEs, HBCD, tobacco smoke).
Many biological mechanisms are involved in developmental toxicity and this complicates the identification of teratogenic agents by simpler models (e.g. *in vitro*). Six main teratogenic mechanisms associated with medication use has been identified: folate antagonism, neural crest cell disruption, endocrine disruption, oxidative stress, vascular disruption and specific receptor- or enzyme-mediated teratogenesis (van Gelder *et al.* 2010). The time at which the embryo or fetus is exposed to the teratogenic agent also plays a role in the potential outcome as the critical window of exposure differs between the different organ systems.
The training set for this model was compiled by MultiCASE Inc. and is composed of data from the Teratogen Information System (TERIS) and a compilation in which the United States Food and Drug Administration (FDA) definitions were used to quantify potential for developmental toxicity from drugs used during pregnancy (Ghanooni *et al.* 1997). The training set consists of clinical, epidemiologic and animal data. Results from the animal data are extrapolated to humans (Jensen *et al.* 2008).

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Potential teratogenic in humans, positive or negative.

3.6 Experimental protocol

Positive compounds consists of compounds whose FDA classification is D (evidence of human fetal potential) or X (a clear adverse fetal effect in animals, humans or both) or for which the TERIS score assigned is high or moderately high. Negative compounds possess an FDA classification of A (no potential of developmental toxicity on the basis of controlled studies in women) or a TERIS score of "none" or "unlikely." (Ghanooni *et al.* 1997).

3.7 Endpoint data quality and variability

Only TERIS data for which data quality was reported fair, good or excellent were considered (Ghanooni *et al.* 1997).

4. Defining the algorithm

4.1 Type of model

A categorical (Q)SAR model based on structural features and numeric molecular descriptors.

4.2 Explicit algorithm

This is a categorical (Q)SAR model made by use of partial logistic regression (PLR). The specific implementation is proprietary within the Leadscope software.

4.3 Descriptors in the model

structural features,

aLogP,

polar surface area,

number of hydrogen bond donors,

Lipinski score,

number of rotational bonds,

parent atom count,

parent molecular weight,

number of hydrogen bond acceptors

4.4 Descriptor selection

Leadscope Predictive Data Miner is a software program for systematic sub-structural analysis of a chemical using predefined structural features stored in a template library, training set-dependent generated structural features (scaffolds) and calculated molecular descriptors. The feature library contains approximately 27,000 pre-defined structural features and the structural features chosen for the library are motivated by those typically found in small molecules: aromatics, heterocycles, spacer groups, simple substituents. Leadscope allows for the generation of training set-dependent structural features (scaffold generation), and these features can be added to the pre-defined structural features from the library and be included in the descriptor selection process. It is possible in Leadscope to remove redundant structural features before the descriptor selection process and only use the remaining features in the descriptor selection process. Besides the structural features Leadscope also calculates eight molecular descriptors for each training set structure: the octanol/water partition coefficient (alogP), hydrogen bond acceptors (HBA),

hydrogen bond donors (HBD), Lipinski score, atom count, parent compound molecular weight, polar surface area (PSA) and rotatable bonds. These eight molecular descriptors are also included in the descriptor selection process.

Leadscope has a default automatic descriptor selection procedure. This procedure selects the top 30% of the descriptors (structural features and molecular descriptors) according to X^2-test for a binary variable, or the top and bottom 15% descriptors according to *t*-test for a continuous variable. Leadscope treats numeric property data as ordinal categorical data. If the input data is continuous such as IC_{50} or cLogP data, the user can determine how values are assigned to categories: the number of categories and the cut-off values between categories. (Roberts *et al.*2000).

When developing this model, intermediate models with application of different modelling approaches in Leadscope were tried:

1. 'Single model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors for descriptor selection.
2. 'Single model' using both the Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.
3. 'Single model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation), with subsequent removal of redundant structural features, and calculated molecular descriptors for descriptor selection.
4. 'Composite model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors in the descriptor selection.
5. 'Composite model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.

Based on model performance as measured by a preliminary cross-validation the model developed using approach number 2. was chosen.

For this model scaffolds were generated by Leadscope for the training set structures and added to the Leadscope library of structural features. Descriptors were then automatically selected among the structural features and the eight molecular descriptors.

4.5 Algorithm and descriptor generation

For descriptor generation see 4.4.

After selection of descriptors the Leadscope Predictive Data Miner program performs partial least squares (PLS) regression for a continuous response variable, or partial logistic regression (PLR) for a binary response variable, to build a predictive model. By default the Predictive Data Miner performs leave-one-out or leave-groups-out (in the latter case, the user can specify any number of repetitions and percentage of structures left out) cross-validation on the training set depending on the size of the training set. In the cross-validation made by Leadscope the descriptors selected for the 'mother model' are used when building the validation submodels and they therefore have a tendency to be overfittet and give overoptimistic validation results.

In this model, because of the categorical outcome in the response variable, PLR was used to build the predictive model. For this model 196 descriptors were selected to build the model. These include 8 Leadscope calculated molecular descriptors, 124 hierarchy features, and 64 scaffolds. The 196 descriptors were distributed on 6 PLS factors.

4.6 Software name and version for descriptor generation

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

4.7 Descriptors/chemicals ratio

In this model 196 descriptors were used and distributed on 196 PLS factors. The training set consists of 323 compounds. The descriptor/chemical ratio is 1:1.65 (196:323).

5. Defining Applicability Domain

5.1 Description of the applicability domain of the model

The definition of the applicability domain consists of two components; the definition of a structural domain in Leadscope and the in-house further probability refinement algorithm on the output from Leadscope to reach the final applicability domain call.

1. Leadscope

For assessing if a test compound is within the structural applicability domain of a given model Leadscope examines whether the test compound bears enough structural resemblance to the training set compounds used for building the model (i.e. a structural domain analysis). This is done by calculating the distance between the test compound and all compounds in the training set (distance = 1 - similarity). The similarity score is based on the Tanimoto method. The number of neighbours is defined as the number of compounds in the training set that have a distance equal to or smaller than 0.7 with respect to the test compound. The higher the number of neighbours, the more reliable the prediction for the test compound. Statistics of the distances are also calculated. Effectively no predictions are made for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated in Leadscope.

2. The Danish QSAR group

In addition to the general Leadscope structural applicability domain definition the Danish QSAR group has applied a further requirement to the applicability domain of the model. That is only positive predictions with a probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted. Predictions within the structural applicability domain but with probability between 0.5 to 0.7 or 0.3 to 0.5 are defined as positives out of applicability domain and negatives out of applicability domain, respectively. When these predictions are wed out the performance of the model in general increases at the expense of reduced model coverage.

5.2 Method used to assess the applicability domain

Leadscope does not generate predictions for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated.

Only positive predictions with probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted.

5.3 Software name and version for applicability domain assessment

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

5.4 Limits of applicability

The Danish QSAR group applies an overall definition of structures acceptable for QSAR processing which is applicable for all the in-house QSAR software, i.e. not only Leadscope. According to this definition accepted structures are organic substances with an unambiguous structure, i.e. so-called discrete organics defined as: organic compounds with a defined two dimensional (2D) structure containing at least two carbon atoms, only certain atoms (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, and I), and not mixtures with two or more 'big components' when analyzed for ionic bonds (for a number of small known organic ions assumed not to affect toxicity the 'parent molecule' is accepted). Calculation 2D structures (SMILES and/or SDF) are generated by stripping off ions (of the accepted list given above). Thus, all the training set and prediction set chemicals are used in their non-ionized form. See 5.1 for further applicability domain definition.

6. Internal validation

6.1 Availability of the training set

No

6.2 Available information for the training set

6.3 Data for each descriptor variable for the training set

6.4 Data for the dependent variable for the training set

6.5 Other information about the training set

323 compounds are in the training set: 130 positives and 193 negatives.

6.6 Pre-processing of data before modelling

Only structures acceptable for Leadscope were used in the final training set. That is only discrete organic chemicals as described in 5.4 were used. In case of replicate structures, one of the replicates was kept if all the compounds had the same activity and all were removed if they had different activity. No further structures accepted by the software were eliminated (i.e. outliers).

6.7 Statistics for goodness-of-fit

Not performed.

6.8 Robustness – Statistics obtained by leave-one-out cross-validation

Not performed. (It is not a preferred measurement for evaluating large models).

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

A five times two-fold 50 % cross-validation was performed. This was done by randomly removing 50% of the full training set used to make the “mother model”, where the 50% contains the same ratio of positive and negatives as the full training set. A new model (validation submodel) was created on the remaining 50% using the same settings in Leadscope but with no information from the “mother model” regarding descriptor selection etc. The validation submodel was applied to predict the removed 50% (within the defined applicability domain for the submodel). Likewise, a validation submodel was made on the removed 50% of the training set and this model was used to predict the other 50% (within the defined applicability domain for this submodel). This procedure was repeated five times.

Predictions within the defined applicability domain for the ten validation submodels were pooled and Cooper’s statistics calculated. This gave the following results for the 59.4% (960/(5*323)) of the predictions which were within the applicability domain:

- Sensitivity (true positives / (true positives + false negatives)): 72.0%
- Specificity (true negatives / (true negatives + false positives)): 85.5%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 80.1%

6.10 Robustness - Statistics obtained by Y-scrambling

Not performed.

6.11 Robustness - Statistics obtained by bootstrap

Not performed.

6.12 Robustness - Statistics obtained by other methods

Not performed.

7. External validation

7.1 Availability of the external validation set

7.2 Available information for the external validation set

7.3 Data for each descriptor variable for the external validation set

7.4 Data for the dependent variable for the external validation set

7.5 Other information about the training 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

External validation not performed for this model.

8. Mechanistic interpretation

8.1 Mechanistic basis of the model

The global model identifies structural features and molecular descriptors which in the model development was found to be statistically significant associated with effect. Many predictions may indicate modes of action that are obvious for persons with expert knowledge for the endpoint.

8.2 A priori or posteriori mechanistic interpretation

A posteriori mechanistic interpretation. The identified structural features and molecular descriptors may provide basis for mechanistic interpretation.

8.3 Other information about the mechanistic interpretation

9. Miscellaneous information

9.1 Comments

The model is useful for identifying potential human developmental toxicants, as well as serving as a starting point for further mechanistic investigations (Ghanooni *et al.* 1997).

9.2 Bibliography

Ghanooni, M., Mattison, D. R., Zhang, Y.P., Macina, O. T., Rosenkranz, H. S., and Klopman, G. (1997) Structural determinants associated with risk of human developmental toxicity. *Am. J. Obstet. Gynecol*., 176:4, 799-805.

Jensen, G. E., Niemelä, J. R., Wedebye, E. B., and Nikolov, N. G. (2008) QSAR models for reproductive toxicity and endocrine disruption in regulatory use - a preliminary investigation. *SAR and QSAR in Environmental Research*, 19:7,631-641.

van Gelder, M.M.H.J., van Rooij, I.A.L.M., Miller, R.K., Zielhuis, G.A., de Jong-van den Berg, L.T.W, and Roeleveld, N. (2010) Teratogenic mechanisms of medical Drugs. *Human Reproduction Update*, 16:4, 378–394.

9.3 Supporting information