MultiCASE CASE Ultra model for maximum recommended daily dose (MRDD) in humans

1. QSAR identifier

1.1 QSAR identifier (title)

MultiCASE CASE Ultra model for maximum recommended daily dose (MRDD) in humans, Danish QSAR Group at DTU Food.

1.2 Other related models

Leadscope Enterprise model for maximum recommended daily dose (MRDD) in humans, Danish QSAR Group at DTU Food.

SciMatics SciQSAR model for maximum recommended daily dose (MRDD) in humans, Danish QSAR Group at DTU Food.

1.3. Software coding the model

MultiCASE CASE Ultra 1.4.6.6 64-bit.

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

Sine Abildgaard Rosenberg;

National Food Institute at the Technical University of Denmark;

Trine Klein Reffstrup;

National Food Institute at the Technical University of Denmark;

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.

2.3 Date of QMRF update(s)

2.4 QMRF update(s)

2.5 Model developer(s) and contact details

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

ebawe@food.dtu.dk

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

nign@food.dtu.dk

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

Klopman, G. (1992) MULTICASE 1. A Hierarchical Computer Automated Structure Evaluation Program. *Quant. Struct.-Act. Relat.,* 11*,* 176 - 184.

Chakravarti, S.K., Saiakhov, R.D., and Klopman, G. (2012) Optimizing Predictive Performance of CASE Ultra Expert System Models Using the Applicability Domains of Individual Toxicity Alerts. *J. Chem. Inf. Model*., 52, 2609 −2618.

Saiakhov, R.D., Chakravarti, S.K., and Klopman, G. (2013) Effectiveness of CASE Ultra Expert System in Evaluating Adverse Effects of Drugs. *Mol. Inf*., 32, 87 – 97.

2.8 Availability of information about the model

The training set is non-proprietary and was compiled from the Maximum Recommended Daily Dose (MRDD) Database which is publically available at the FDA/CDER Webpage (http://www.epa.gov/comptox/dsstox/sdf_fdamdd.html, accessed 9th of July 2013). 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 (phase 1 clinical trial).

3.2 Endpoint

QMRF 4. Human Health Effects

QMRF 4.14. Repeated dose toxicity

3.3 Comment on endpoint

The Maximum Recommended Daily Dose (MRDD) for a pharmaceutical is an estimated upper dose limit beyond which a drug's efficacy is not increased and/or undesirable adverse effects begin to outweigh beneficial effects. The MRDD is related to the No Observed Effect Level (NOEL) for non-pharmaceuticals (NOEL equals 1/10 MRDD), a dose at which there are no statistically or biologically significant increases in the frequency or severity of any effect between the exposed population and its appropriate control. Because of this relation this model can be used to estimate both the MRDD and NOEL values for a given compound.

Data for this model was compiled from FDA's Center for Drug Evaluation and Research, Office of Pharmaceutical Science, Informatics and Computational Safety Analysis Staff's Maximum Recommended Daily Dose (FDAMDD) database. Most of the MRDD values in the FDAMDD database were determined from pharmaceutical phase 1 human clinical trials that employed an oral route of exposure and daily treatments, usually for 3 - 12 months. The pharmaceuticals were given as single or divided dose treatment regimens to achieve desired pharmacological effects. In contrast, roughly 5% of the pharmaceuticals in the FDAMDD database were anti-neoplastics and anesthetics and these were administered intravenously and/or intramuscularly. When separate MRDDs were reported for different routes of exposure, only the oral MRDD was included in the database. In addition, some pharmaceuticals have different MRDD values for male and female adults, children, or elderly patients. In this situation only MRDD values for the average adult patient were used.

Pharmaceuticals that are administered orally are usually tested over a limited range of doses and have MRDDs reported as mg/day. The MRDDs were converted from the mg/day unit to mg/kg body weight (bw)/day based upon an average adult weighing 60 kg. In contrast, the dose unit for most antineoplastic drug MRDDs is reported as mg/m^2 which was converted to mg/kg bw/day using the formula mg/kg bw/day = $mg/m^2/37$ for an average adult. Additionally, a few drugs had MRDDs reported in parts per million (ppm) which were converted to mg/kg bw/day on the basis that 1000 ppm equals 25 mg/kg bw/day for an average 60 kg adult. MRDD values for the 1,220 chemicals in this training set range from 0.00001 to 1000 mg/kg bw/day (Matthews *et al.* 2004).

As data for this model is derived directly from human data it can be argued that the model predictions can give a more accurate estimate of human MRDD than data derived from repeat-dose tests in rodents.

To make a categorical model compounds with a MRDD value between 0.0167-2.69 mg/kg bw/day were defined as positive and compounds with MRDD values between 5.00-1000 mg/kg bw/day were defined as negative. Intermediate compounds were defined as marginal.

3.4 Endpoint units

CASE Unit, 45 for positive, 25 for marginal, and 10 or below for negatives.

3.5 Dependent variable

Maximum recommended Daily dose (MRDD) in humans, positive, marginal or negative.

3.6 Experimental protocol

Data originate from pharmaceutical phase 1 human clinical trials that employed an oral route of exposure and daily treatments, usually for 3 - 12 months. The pharmaceuticals were given as single or divided dose treatment regimens to achieve desired pharmacological effects. In contrast, roughly 5% of the pharmaceuticals in the FDAMDD database were anti-neoplastics and anesthetics and these were administered intravenously and/or intramuscularly (Matthews *et al.* 2004).

3.7 Endpoint data quality and variability

According to (http://www.epa.gov/comptox/dsstox/sdf_fdamdd.html) "Several features of DSSTox FDAMDD have the potential to impact on SAR analysis and should be taken into account in any future use of these data. Most prominent among these is the imprecise nature of the reported MRDD value, both in terms of the wide range of adverse or toxic effects that would be considered in assigning the MRDD, and in terms of the ambiguous chemical structure association with this dose measure. In DSSTox FDAMDD and the corresponding Source FDA MRDD database, there are several cases where a single Dose_MRDD_mg value is assigned to multiple related structural derivatives of a pharmaceutical, i.e., the same activity is assigned to multiple Structure/CASRN records in the database. In theory, an MRDD value will reflect the lowest dose of a drug producing adverse effects but for the FDA MRDD database this value has been derived from pooled clinical reports where more than one form of a drug may have been administered. When MRDD mg mass units are converted to mmol units for SAR analysis, a single Dose_MRDD_mg is converted to a range of mmol doses, taking into account the different molecular weights of the various drug derivatives. Assuming that these various drug derivatives have similar or equal molar potencies, the reported Dose_MRDD_mg could be presumed to reflect the dose of the smallest STRUCTURE_MolecularWeight derivative that would register as the highest molar content and, therefore, most potent for a given mass dose."

4. Defining the algorithm

4.1 Type of model

A categorical (Q)SAR model based on structural fragments and calculated molecular descriptors.

4.2 Explicit algorithm

This is a categorical (Q)SAR model composed of multiple local (Q)SARs made by use of stepwise regression. The specific implementation is proprietary within the MultiCASE CASE Ultra software.

4.3 Descriptors in the model

Fragment descriptors,

Distance descriptors,

Physical descriptors,

Electronic descriptors,

Quantum mechanical descriptors

4.4 Descriptor selection

Automated hierarchical selection (see 4.5).

4.5 Algorithm and descriptor generation

MultiCASE CASE Ultra is an artificial intelligence (AI) based computer program with the ability to learn from existing data and is the successor to the program MultiCASE MC4PC. The system can handle large and diverse sets of chemical structures to produce so-called global (Q)SAR models, which are in reality series of local (Q)SAR models. Upon prediction of a query structure by a given model one or more of these local models, as well as global relationships if these are identified, can be involved if relevant for the query structure. The CASE Ultra algorithm is mainly built on the MCASE methodology (Klopman 1992) and was released in a first version in 2011 (Chakravarti *et al.* 2012, Saiakhov *et al.* 2013).

CASE Ultra is a fragment-based statistical model system. The methodology involves breaking down the structures of the training set into all possible fragments from 2 to 10 heavy (non-hydrogen) atoms in length. The fragment generation procedure produces simple linear chains of varying lengths and branched fragments as well as complex substructures generated by combining the simple fragments.

A structural fragment is considered as a positive alert if it has a statistical significant association with chemicals in the active category. It is considered a deactivating alert if it has a statistically significant relation with the inactive category.

Once final lists of positive and deactivating alerts are identified, CASE Ultra attempts to build local (Q)SARs for each alert in order to explain the variation in activity within the training set chemicals covered by that alert. The program calculates multiple molecular descriptors from the chemical structure such as molecular orbital energies and two-dimensional distance descriptors. A stepwise regression method is used to build the local (Q)SARs based on these molecular descriptors. For each step a new descriptor (modulator) is

added if the addition is statistically significant and increases the cross-validated R2 (the internal performance) of the model. The number of descriptors in each local model is never allowed to exceed one fifth of the number of training set chemicals covered by that alert. If the final regression model for the alert does not satisfy certain criteria (R2 ≥ 0.6 and Q2 ≥ 0.5) it is rejected. Therefore, not all alerts will necessarily have a local (Q)SAR.

The collection of positive and deactivating alerts with or without a local (Q)SAR constitutes a global (Q)SAR model for a particular endpoint and can be used for predicting the activity of a test chemical.

More detailed information about the algorithm can be found in Chakravarti *et al.* (2012), Saiakhov *et al.* (2013).

4.6 Software name and version for descriptor generation

MultiCASE CASE Ultra 1.4.6.6 64-bit.

4.7 Descriptors/chemicals ratio

The program primarily uses fragment descriptors specific to a group of structurally related chemicals from the training set. Therefore estimation of the number of descriptors used in a specific model, which is a collection of local models as explained under 4.5, may be difficult. In general, we estimate that the model uses an order of magnitude less descriptors than there are observations. The number of descriptors in each local (Q)SAR model is never allowed to exceed one fifth of the number of training set chemicals covered by that alert (Saiakhov *et al.* 2013).

It should be noted that due to CASE Ultra's complex decision making scheme overfitting is rare compared to simpler linear models. Warnings are issued in case of statistically insufficient overall number of observations to produce a model, which is not the case in the present model.

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 in CASE Ultra and the in-house further refinement algorithm on the output from CASE Ultra to reach the final applicability domain call.

1. CASE Ultra

CASE Ultra recognizes unknown structural fragments in test chemicals that are not found in the training data and lists these in the output for a prediction. Fragments this way impose a type of global applicability domain for the overall model. The presence of more than three unknown structural fragments in the test chemical results in an 'out of domain' call in the program. (Chakravarti *et al.*2012, Saiakhov *et al.*2013).

For each structural alert, CASE Ultra uses the concept of so-called domain adherences and statistical significance.

The domain adherence for an alert in a query chemical depends on the similarity of the chemical space around the alert in the query chemical compared to the chemical space (in terms of frequencies of occurrences of statistically relevant fragments) of the training set chemicals used to derive the alert. The domain adherence value (between zero and one) is the ratio of the sum of the squared frequency of occurrence values of the subset of the fragments that are present in the test chemical and sum of the squared frequency of occurrence of all the fragments that constitute the domain of the alert in question. The more fragments of the domain of the alert in the test chemical the closer the domain adherence value is to 1. The value will never be zero as the alert itself is part of the alerts domain.

Furthermore, all alerts come with a measure of its statistical significance, and this depends on the number of chemicals in the training set which contained the alert and the prevalence within these of actives and inactives. (Chakravarti *et al.*2012).

2. In-house refinement algorithm to reach the final applicability domain call

The Danish QSAR group has applied a stricter definition of applicability domain for its MultiCASE CASE Ultra models.

An optimization procedure based on preliminary cross-validation is applied to further restrict the applicability domain for the whole model based on non-linear requirements for domain adherence and statistical significance, giving the following primary thresholds:

Domain adherence = 0.72 and significance = 70

Any positive prediction is required to contain at least one valid positive alert, namely an alert with statistical significance and domain adherence exceeding thresholds defined for the specific model.

The positive predictions for chemicals which only contain invalid positive alerts are considered 'out of domain' (in CASE Ultra these chemicals are predicted to be inactive).

Furthermore, only query chemicals with no unknown structural fragments are considered within the applicability domain, except for chemicals predicted 'positive', where one unknown fragment is accepted. Also no significant positive alerts are accepted for an inactive prediction.

5.2 Method used to assess the applicability domain

The applicability domain is assessed in terms of the output from CASE Ultra with the Danish QSAR group's further refinement algorithm on top as described in 5.1.

Because of the complexity of the system (see 5.1), the assessment of whether a test chemical is within the applicability domain of the model requires predicting the chemical with the specific model, and the application of the Danish QSAR group model-specific thresholds for domain adherence and significance.

This applicability domain was also applied when determining the results from the cross-validations (6.9).

5.3 Software name and version for applicability domain assessment

MultiCASE CASE Ultra 1.4.6.6 64-bit.

5.4 Limits of applicability

All structures are run through the DataKurator feature within CASE Ultra to check for compatibility with the program. Furthermore, 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 CASE Ultra. 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). Structures with less than two carbon atoms or containing atoms not in the list above (e.g. heavy metals) are rendered out as not acceptable for further QSAR processing. Calculation 2D structures (SMILES and/or SDF) are generated by stripping off accepted organic and inorganic ions. 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

Yes

6.2 Available information for the training set

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

1222 compounds are in the training set: 524 positives, 115 marginals, and 583 negatives.

6.6 Pre-processing of data before modelling

Data was originally collected from the FDAMDD database. Only compounds for which SMILES codes could be found and that had a structure acceptable for the commercial software 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.

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”, thereby splitting the full training set into two subsets A and B, each containing the same ratio of positives to negatives as the full training set. A new model (validation sub-model) was created on subset A without using any information from the “mother model” (regarding e.g. descriptor selection etc.). The validation sub-model was applied to predict subset B (within the CASE Ultra applicability domain for the validation sub-model and the in-house further refinement algorithm for the full model). Likewise, a validation sub-model was made on subset B and this model was used to predict subset A (within the CASE Ultra applicability domain for the validation sub-model and the in-house further refinement algorithm for the full model). This procedure was repeated five times.

Predictions within the defined applicability domain for the ten validation sub-models were pooled and Cooper's statistics calculated. This gave the following results for the 36.5% (2018/(5*1106)) of the predictions which were within the applicability domain:

- Sensitivity (true positives / (true positives + false negatives)): 69.4%
- Specificity (true negatives / (true negatives + false positives)): 92.5%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 82.5%

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 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

External validation has not been performed for this model.

8. Mechanistic interpretation

8.1 Mechanistic basis of the model

The model identifies statistically relevant substructures (i.e. alerts) and for each set of molecules containing a specific alert it further identifies additional parameters found to modulate the alert (e.g. logP and molecular orbital energies, etc.). Many predictions may indicate modes of action that are obvious for persons with expert knowledge about 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 can be used to predict the human MRDD in a categorical way: Positive means MRDD value between 0.0167-2.69 mg/kg bw/day, negative means MRDD values between 5.00-1000 mg/kg bw/day and marginal means intermediate in between 2.69 and 5.00 mg/kg bw/day. It can be argued that the predictions from this (Q)SAR model give a more accurate estimate of human MRDD/NOEL than those derived from animal toxicity studies, where multiple uncertainty/safety factors are necessary to compensate for incompatibility and uncertainty underlying the extrapolation of animal toxicity to humans.

9.2 Bibliography

Matthews, E.J, Kruhlak, N.L, Benz, R.D and Contrera, J.F. (2004) Assessment of the Health Effects of Chemicals in Humans: I. QSAR Estimation of the Maximum Recommended Daily Dose (MRDD) and No Effect Level (NOEL) of Organic Chemicals Based on Clinical Trial Data. *Current drug Discovery Technologies,* 1, 61-76.

Maximum recommended daily dose (MRDD) Database: http://www.epa.gov/comptox/dsstox/sdf_fdamdd.html. Accessed 9th of July 2013. According to which the data originates from:

- Blacow, N. W. *Martindale: The Extra Pharmacopoeia, 26th ed.*; The Pharmaceutical Press: London, **1972**.
- Wade, A. *Martindale: The Extra Pharmacopoeia, 28th ed.* ; The Pharmaceutical Press: London, **1982**.
- Reynolds, J. E. F. *Martindale: The Extra Pharmacopoeia, 30th ed.;* The Pharmaceutical Press: London, **1993**.
- Sweetman, S. C. *Martindale: The Complete Drug Reference*; Ed.; The Pharmaceutical Press: London. Electronic version, Thomson
- MICROMEDEX: Greenwood Village, CO. Edition expires **2003**.
- Arky, R. *The Physicians' Desk Reference, 49th ed.*; Medical Economics Company: New Jersey, **1995**.
- Arky, R. *The Physicians' Desk Reference, 53rd ed.*; Medical Economics Company: New Jersey, **1999**.

9.3 Supporting information