International Collaboration to Build Predictive Models for Acute Oral Toxicity

Kamel Mansouri
LEAD COMPUTATIONAL CHEMIST, ILS IN SUPPORT OF NICEATM

SOT 2019 – Abstract: 2563

Disclaimer: ILS staff provide technical support for NICEATM, but do not represent NIEHS, NTP, or the official positions of any federal agency. (the author declares no conflict of interest)
Overview

• Project scope: acute oral toxicity
  – Regulatory use of these data
  – Endpoints selected for predictive modeling
  – Compiling inventory of rat acute oral LD50
  – Establishing training, evaluation, and prediction sets
  – Evaluation of submitted models

• International contributors

• Generation of consensus predictions

• Current status and public release
Scoping Regulatory Needs

ICCVAM Acute Toxicity Workgroup

- Identifies federal agency requirements, needs, and decision contexts for using acute systemic toxicity data
Agency-Based Modeling Endpoint Selection

**Binary Models**
- EPA Categories:
  - I (≤ 50 mg/kg)
  - II (>50 ≤ 500 mg/kg)
  - III (>500 ≤ 5000 mg/kg)
  - IV (>5000 mg/kg)
- GHS Categories:
  - I (≤ 5 mg/kg)
  - II (>5 ≤ 50 mg/kg)
  - III (>50 ≤ 300 mg/kg)
  - IV (>300 ≤ 2000 mg/kg)
  - NC (> 2000 mg/kg)

**Continuous Model**
- Point estimates of LD50 values

**Categorical Models**
- EPA Categories:
  - I (≤ 50 mg/kg)
  - II (>50 ≤ 500 mg/kg)
  - III (>500 ≤ 5000 mg/kg)
  - IV (>5000 mg/kg)

**Packing Group**
- OSHA Hazard

- Highly toxic (≤50 mg/kg)
- Toxic (>50-5000 mg/kg)
- Nontoxic (>2000 mg/kg)
Available data for modeling

Rat oral LD50s:
16,297 chemicals total
34,508 LD50 values

15,688 chemicals total
21,200 LD50 values

QSAR-ready standardization
Desalted, stereochemistry stripped,
tautomers and nitro groups standardized,
valence corrected, structures neutralized

11,992 chemicals with accurate structures

• Very toxic endpoint: 11,886 entries (binary, 0/1)
• Non-toxic endpoint: 11,871 entries (binary, 0/1)
• EPA endpoint: 11,755 entries (categorical, 4 categories)
• GHS endpoint: 11,845 entries (categorical, 5 categories)
• LD50 endpoint: 8,908 entries (continuous values)
Aim of the workflow:

- Combine different procedures and ideas
- Minimize the differences between the structures used for prediction
- Produce a flexible free and open source workflow to be shared

Wedebyle et al. Danish EPA Environmental Project No. 1503, 2013
Mansouri et al. (http://ehp.niehs.nih.gov/15-10267/)
Establishing Modeling Dataset

- **Training and evaluation sets:**
  - 11,992 chemicals from the final inventory of chemicals with QSAR-ready structures having rat oral acute toxicity data were split into training and test sets:
    - 75% training set: 8,994 chemicals
    - 25% evaluation set: 2,998 chemicals
  - The same training and test chemicals across all endpoints
  - Similar distributions and variability for values and categories
  - Similar distribution of chemical structures sources
Establishing Modeling Dataset

- **Prediction set:**

**Included Lists:**

- ToxCast/Tox21
- EDSP
- TSCA
- Substances on the market (EPA Dashboard list)

After QSAR-ready standardization:

48137 structures to be predicted (including the evaluation set)
Consortium:

- Participants from around the globe representing academia, industry, and government contributed.
Submitted Models

• GHS categories: 23 models
• Consortium Comprised 35 Participants/Groups
  • Very Toxic: 32 models
  • Non-toxic: 33 models
  • EPA categories: 26 models
  • LD50: 25 models

Total: 139 models
Evaluation procedure

Qualitative evaluation:
- Documentation
- Defined endpoint
- Unambiguous algorithm
- Availability of code
- Applicability domain definition
- Availability of data used for modeling
- Mechanistic interpretation

Quantitative evaluation:
- Goodness of fit: training statistics
- Evaluation set predictivity: statistics on the evaluation set
- Robustness: balance between (Goodness of fit) & (Test set predictivity)

\[ S = 0.3 \ast (\text{Goodness of fit}) + 0.45 \ast (\text{Test set predictivity}) + 0.25 \ast (\text{Robustness}) \]

Categorical models (binary and multi-class):

- **Goodness of fit**
  \[ \text{Goodness of fit} = 0.7 \ast (BA_{Tr}) + 0.3 \ast (1 - |Sn_{Tr}|) \]

- **Test set predictivity**
  \[ \text{Test set predictivity} = 0.7 \ast (BA_{Tst}) + 0.3 \ast (1 - |Sn_{Tst}|) \]

- **Robustness**
  \[ \text{Robustness} = 1 - |BA_{Tr} - BA_{Tst}| \]

Continuous models:

- **Goodness of fit**
  \[ \text{Goodness of fit} = R^2_{Tr} \]

- **Test set predictivity**
  \[ \text{Test set predictivity} = R^2_{Tst} \]

- **Robustness**
  \[ \text{Robustness} = 1 - |R^2_{Tr} - R^2_{Tst}| \]
Evaluation results

Quantitative evaluation

[Graph showing evaluation scores for various categories]
Coverage of the models

Distribution of the number of models/chemical
Generation of Consensus Predictions

- Models passing qualitative evaluation (requirement for transparency; description of approach was sufficient)

- Integrating only *in-domain* predictions across chemicals in the prediction set (48,137 chemicals) for each model, respectively
  - Categorical models: weighted majority rule
  - Continuous model: weighted average

Kleinstreuer et al., Predictive models for acute oral systemic toxicity: A workshop to bridge the gap from research to regulation, Computational Toxicology, 2018, https://doi.org/10.1016/j.comtox.2018.08.002.
### Performance Assessment

#### Consensus Model Statistics

<table>
<thead>
<tr>
<th></th>
<th>Very Toxic</th>
<th>Non-Toxic</th>
<th>EPA</th>
<th>GHS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Eval</td>
<td>Train</td>
<td>Eval</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.87</td>
<td>0.67</td>
<td>0.93</td>
<td>0.70</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
<td>0.88</td>
</tr>
<tr>
<td>Balanced Accuracy</td>
<td>0.93</td>
<td>0.81</td>
<td>0.94</td>
<td>0.79</td>
</tr>
<tr>
<td>In vivo Balanced Accuracy</td>
<td>0.81</td>
<td>0.89</td>
<td>0.82</td>
<td>0.79</td>
</tr>
</tbody>
</table>

The consensus predictions perform just as well as replicate *in vivo* data do at predicting oral acute toxicity outcome.
**Consensus concordance**

**Distributions of the concordance between models**

- **VT**
- **NT**
- **LD50**
- **EPA**
- **GHS**
Consensus implementation

Generalized CATMoS models: datasets

- LD50: 28954
- VT: 23767
- NT: 30971
- EPA: 25487
- GHS: 25720

- High concordance
  - Proportional distribution of:
    - LD50 values
    - VT/NT classes
    - EPA/GHS categories

- Split into 75% training and 25% test set
- Calculate PaDEL & CDK2 descriptors
- Dimensionality reduction (missing values & low variance)
- Feature selection (most relevant descriptors for each endpoint)
Consensus implementation

Generalized CATMoS models: new predictions

\[ d_1 \neq 0 \]
\[ w_i = f(d_i) \]
\[ \text{Pred}_i = f(w_i, N_i) \]

\( d_1 = 0 \)
\( \text{Pred}_i = N_i \)

New chemical to be predicted
Nearest neighbors \((N_i)\)

\( d_i \): Euclidean distance based on the selected descriptors for each endpoint

Automated, weighted-endpoint dependent read-across: weighted kNN
Consensus implementation

Generalized CATMoS models: statistics

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Descriptors</th>
<th>Training (5-f CV)</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT (BA)</td>
<td>21</td>
<td>0.79</td>
<td>0.77</td>
</tr>
<tr>
<td>NT (BA)</td>
<td>11</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>EPA (BA)</td>
<td>15</td>
<td>0.79</td>
<td>0.81</td>
</tr>
<tr>
<td>GHS (BA)</td>
<td>15</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>LD50 (Q²,R²)</td>
<td>23</td>
<td>0.79</td>
<td>0.81</td>
</tr>
</tbody>
</table>
OPERA Standalone app

Command line
- Free, opensource & open-data
- Single chemical and batch mode
- Multiple platforms (Windows and Linux)
- Embeddable libraries (java, C, C++, Python)

Graphical user interface

https://github.com/NIEHS/OPERA

https://doi.org/10.1186/s13321-018-0263-1
New in OPERA2:

- Physchem properties:
  - General structural properties
  - pKa
  - Log D

- Toxicity endpoints
  - ER activity (CERAPP)
    https://ehp.niehs.nih.gov/15-10267/
  - AR activity (CoMPARA)
    https://doi.org/10.13140/RG.2.2.19612.80009
  - Acute toxicity (CATMoS)
    https://doi.org/10.1016/j.comtox.2018.08.002

- ADME properties
  - Plasma fraction unbound (FuB)
  - Intrinsic clearance (Clint)
OPERA predictions on EPA’s CompTox dashboard

https://comptox.epa.gov/dashboard

Mansouri et al. OPERA models
(https://doi.org/10.1186/s13321-018-0263-1)

Williams et al. CompTox Chemistry Dashboard
(https://doi.org/10.1186/s13321-017-0247-6)
The “3C” Concept at Work!

- Success of the project was due in great part to the use of the 3C concept as well as up-front and continuous engagement of regulators in the process.

[Image: A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States]

https://ntp.niehs.nih.gov/go/natl-strategy
Acknowledgements

THANK YOU!

• ICCVAM Acute Toxicity Workgroup

• EPA/NCCT
  – Grace Patlewicz
  – Jeremy Fitzpatrick

• ILS/NICEATM
  – Agnes Karmaus
  – Dave Allen
  – Shannon Bell
  – Patricia Ceger
  – Judy Strickland
  – Amber Daniel

• NTP/NICEATM
  – Nicole Kleinstreuer
  – Warren Casey

Feedback welcome: Kamel Mansouri (kmansouri@ils-inc.com)

Technical support was provided by ILS under NIEHS contract HHSN273201500010C.