Practical Considerations for Estimation of Oral Rat LD50

Todd Martin, US EPA
Difficulty with oral rat LD50 endpoint

- n-alcohols
  - Oral rat LD50
    - \[ y = -0.01x^2 + 0.22x + 0.79 \]
    - \[ R^2 = 0.38 \]
  - 96 hour Fathead minnow LC50
    - \[ y = -0.02x^2 + 0.72x - 0.82 \]
    - \[ R^2 = 1.00 \]

- n-amines
  - Oral rat LD50
    - \[ y = -0.00x^2 + 0.08x + 2.09 \]
    - \[ R^2 = 0.19 \]
  - 96 hour Fathead minnow LC50
    - \[ y = 0.60x^2 + 0.42x + 0.76 \]
    - \[ R^2 = 0.99 \]
Additional data to improve results

- ChemIDPlus lists effects for some chemicals for oral rat LD50

<table>
<thead>
<tr>
<th>rat</th>
<th>LD50 oral</th>
<th>Effect</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>930mg/kg</td>
<td></td>
<td>TREMOR</td>
</tr>
<tr>
<td></td>
<td>(930mg/kg)</td>
<td></td>
<td>CONVULSIONS OR EFFECT ON SEIZURE THRESHOLD</td>
</tr>
</tbody>
</table>

- There is a pest target species database:

- Both of these can help bin chemicals but may not improve external prediction accuracy since not available for new chemicals (have to predict them before can bin a chemical)
What were the major practical considerations identified when modelling – are there recommendations you could propose that would have helped resolve some/all of these?

- Should we fit models to continuous LD50 values or to binary/category endpoints? Or both depending on final goal?

Were there limits to appropriate methods for the different endpoints?

- It’s very difficult to fit a single regression model to the entire LD50 training set so need clustering or neighbor based method
QSAR Method Selection

What were the rationale to applying particular methods
- Needed a method that could be applied to very large dataset with lots of experimental error

What are advantages/disadvantages of making specific method choices?
- Hierarchical clustering
  - Slow for model building and prediction
  - Can correlate differences in toxicity for subsets of the data
  - Hard to make “external” predictions for training compounds
- Nearest neighbor
  - Very fast and easy to understand
  - Can’t quantify differences in toxicity between test compound and the neighbors
  - Can always make “external” predictions for training compounds
Applicability Domain

- How will I know if my chemistry is within the domain of applicability of your model?
  - Hierarchical clustering
    - Model ellipsoid, Rmax, and Fragment constraints
  - Nearest neighbor
    - Three most similar chemicals must exceed a minimum cosine similarity coefficient of 0.5

\[
h_{00} = X_o^T \left( X^T X \right)^{-1} X_0
\]

\[
distance_i = \sum_{j=1}^{d} \left( x_{ij} - C_j \right)^2
\]

\[
SC_{i,k} = \frac{\sum_{j=1}^{\#descriptors} x_{ij} x_{kj}}{\sqrt{\sum_{j=1}^{\#descriptors} x_{ij}^2 \cdot \sum_{j=1}^{\#descriptors} x_{kj}^2}}
\]
What elements do “difficult to predict” chemistries have in common?

- 16 of the top 20 worst predicted compounds contain **Nitrogen**
  - The nitrogens are usually in an aromatic ring or attached to one
- Usually it’s hard for machine-based methods to detect reactive substructures purely from descriptor values
Difficult to predict, cont.

* Numbers represent the prediction error in log10 (LD50 mg/kg)
Metabolism / bioavailability

Does metabolism (or bioavailability) play a role in the ability to predict acute toxicity?

- Usually QSAR models take that into account from the different functional groups in the molecule if the training set is big enough.
- May not be able to detect reactive substructures.
  - e.g. OH near C=C for an allylic alcohol.
Choice of Dependent Variable For Predicting Category Endpoints

- Additional data points (8953 vs 6706 chemicals for “very toxic” endpoint)
- Continuous endpoint (-log(LD50 mol/kg) has more concentration dependence information

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HC Cat. Input</td>
<td>0.515</td>
<td>0.794</td>
<td>0.527</td>
<td>0.659</td>
</tr>
<tr>
<td>HC LD&lt;sub&gt;50&lt;/sub&gt; input</td>
<td>0.467</td>
<td>0.821</td>
<td>0.487</td>
<td>0.659</td>
</tr>
<tr>
<td>NN Cat. Input</td>
<td>0.498</td>
<td>0.878</td>
<td>0.498</td>
<td>0.875</td>
</tr>
<tr>
<td>NN LD&lt;sub&gt;50&lt;/sub&gt; input</td>
<td>0.426</td>
<td>0.993</td>
<td>0.445</td>
<td>0.875</td>
</tr>
</tbody>
</table>

*AA = average accuracy over five categories, Cov. = fraction of chemicals predicted
Availability

How would end-users access your model and run it on new chemicals?

- **T.E.S.T. downloadable Java program**
  - [https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test](https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test)

- **WebTest:**
  - [https://comptox.epa.gov/dashboard/predictions/index](https://comptox.epa.gov/dashboard/predictions/index)
  - Via graphical user interface or API call

- Training set is probably slightly different
Mechanistic Considerations

- What was there coverage/evaluation of different mechanisms of action?
  - Look at pesticide data or effect data from ChemIDplus?

- What was the approach taken to determine mechanistic interpretations? What were the strengths and weaknesses of that approach? How does that compare/contrast with other approaches?
  - Didn’t investigate that since prior work didn’t show much promise
Is "structure-basis" (i.e., SAR or QSAR) sufficient for mechanistic interpretation?

- In my opinion there is so much experimental scatter that it’s hard to validate any sort of mechanistic interpretation

Should models be linked to adverse outcome pathways? How can we facilitate that?

- Are there AOPs assigned to enough chemicals?
- I am not sure that will increase external prediction accuracy aside from increasing confidence by toxicologists in the predicted values
Questions???

The views expressed in this presentation are those of the author and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency