Predictive Models for Acute Oral Systemic Toxicity

William H. Natcher Conference Center
National Institutes of Health, Bethesda, Maryland
Wednesday, April 11 – 8:30 a.m.-5:00 p.m.
Thursday, April 12 – 8:30 a.m.-4:15 p.m.

Breakout Group B: Interpretation, Characterization, and Extension (1)

- What types of approaches should be used to account for **variability** in the data or the **uncertainties** in the predictions?
  - Uncertainties should drive the collection of new data
  - Depends on decision context
  - Need feedback from users
  - How to identify sources of variability (e.g., species, gender, facility, chemical speciation)
  - Coefficient of variation, confidence interval, mean, median, mode
  - Modelers should model distributions rather than point estimates
  - Evaluate data quality of reference data
  - Consensus models need to propagate the uncertainty of constituent models; Bayesian model averaging approach recommended
  - Discordant and concordant outcomes in the consensus predictions
  - Expectations of model outputs should be addressed
    - Focus has been on statistical performance metrics but there are expectations on other outputs (i.e., mechanistic information) that may or may not be more difficult to provide

Breakout Group B: Interpretation, Characterization, and Extension (2)

- How should we communicate the variability in the reference data to establish a basis for performance of new approaches and set appropriate expectations?
  - Outline the workflow for treatment of the dataset
  - Include evaluation of other endpoints
  - Evaluate the sources

Breakout Group B: Interpretation, Characterization, and Extension (3)

- When is mechanistic information essential in order to accept predictions, and when is having a good performing model is enough?
  - Preference is always to have more data and mechanism is of interest
  - Mechanistic information is always important, but we don't have mechanistic information from the current animal study
  - Local domain interpretation is facilitated by mechanism

- What level of model performance is needed to provide sufficient confidence in the prediction?
  - When it performs to the level of the reference studies
  - Depends on the context (what decision is being made) and on the type of model as it relates to the purpose applied

Breakout Group B: Interpretation, Characterization, and Extension (4)

- How do we build confidence and trust in machine learning models for regulatory toxicology?
  - Evaluate the results with other information in a WOE approach
  - Evaluate applicability to similar chemicals
  - Check the predictions with new studies (test them) and update/retrain the models accordingly. Model changes must be supported by rationale for regulators to use.
  - Adopt a uniform lexicon
  - Agency/industry partnerships should develop and publish case studies
  - Need better communication between modelers and users, including documentation.

- What is the perspective on regulatory acceptance of "black box" or semi-black box prediction models (i.e. ANN, deep learning, etc)?
  - Model result needs to be defensible (need to understand biological relevance)
  - These may not be defensible. Regulators need to be able to know/explain how the model works, and what data streams are. Regulators need access to model experts.