Machine Learning Models: Regulatory Application, Acceptance, and Implementation

Nicole C. Kleinstreuer
NICEATM Deputy Director

ICCVAM CoP
23rd January, 2018
Outline

• Regulatory Needs & Challenges

• Current and Future Applications:
  – QSAR models (EPA, FDA)
  – TSCA Prioritization
  – Endocrine: CERAPP/CoMPARA
  – Skin Sensitization: Defined Approaches
  – Acute Oral Toxicity: Predictive Models
Environmental Chemical Disease Contributions

- **Pesticides**
  - Cancer, neurodegenerative diseases, thyroid

- **Consumer products**
  - Neurological, developmental, systemic

- **Air pollutants**
  - Childhood ADHD, autism, allergic asthma

- **Drinking water contaminants**
  - Systemic effects, cancer, neurological

- **Endocrine Disruptors**
  - Developmental impairment, decreased fertility, cancer

.......and many others.......
Chemicals >> Data

- 80+ million substances synthetized
- 140,000 chemicals in commerce (plus mixtures, natural products and metabolites)
- Less than 10% tested
Curated Legacy Data
e.g. REACH, ToxRefDB, ICE

Omics technologies
e.g. transcriptomics, metabolomics, exposomics

High-Throughput Screening
e.g. ToxCast, Tox21

High-Content Imaging
e.g. EuToxRisk
Current Regulatory Use of ML: Structure Based Models

- EPA/OPPT: Predictive Methods to Assess Hazard under TSCA
  - EcoSAR, OncoLogic, EPISuite

- FDA/CDER, CDRH: Genotoxicity and Carcinogenicity
  - Bacterial mutagenicity (expert-rule based & statistical)
EPA/NCCT Decision Support Tools
Deliver Data and Models

Comptox Chemistry Dashboard
https://comptox.epa.gov/dashboard/

ToxCast Dashboard
https://actor.epa.gov/dashboard/

RapidTox Dashboard

Internal Beta
RapidTox: Prioritization Workflow

[Diagram of RapidTox Prioritization Workflow]

- Chemical List:
  - TSCA
  - OPP Inerts

- To run prioritization, select the chemical set, the allowable data domains, and update the weights. Then select the Recalculate button and go to the prioritization tab. You can then sort by the different prioritization types.

- Recalculate
- Export Table

- Human Health
  - Acute
    - In vivo
    - Q SAR

  - Subchronic
  - Chronic

  - DevTox
  - ReproTox
  - Cancer
  - Mutagenicity
  - Neurotox

- Endocrine
  - Estrogen Agonist
    - In vitro
    - Q SAR

  - Estrogen Antagonist
    - In vitro
    - Q SAR

  - Androgen Agonist
    - In vitro
    - Q SAR

  - Androgen Agonist
    - In vitro
    - Q SAR

- Ecological
  - Fish Acute Tox
    - In vivo
    - Q SAR

  - Crustacea Acute Tox
    - Q SAR

  - Algae Tox
    - Q SAR

  - Fish ReproTox
    - Q SAR

Rusty Thomas, EPA/NCCT
Environmental Endocrine Disruptors

Developmental Effects of Endocrine-Disrupting Chemicals in Wildlife and Humans

Theo Colborn,1 Frederick S. vom Saal,2 and Ana M. Soto3

1W. Alton Jones Foundation and World Wildlife Fund, Washington, DC, 20037 USA; 2Division of Biological Sciences and John M. Dalton Research Center, University of Missouri, Columbia, MO 65211 USA; 3Department of Anatomy and Cellular Biology, Tufts University, Boston, MA 02111 USA

1993, Environmental Health Perspectives

Legislative Mandates:
1996 Federal Food, Drug and Cosmetic Act
1996 Safe Drinking Water Act Amendments

U.S. EPA EDSP
Evolution of the Endocrine Disruptor Screening Program

EDSP Tier 1 Testing: for the purposes of prioritization and screening, identify chemicals with the potential to disrupt estrogen, androgen, or thyroid hormone receptor signaling.

Mismatch between resources needed for EDSP Tier 1 testing and the number of chemicals to be tested
  - ~$1M per chemical for Tier 1
  - 11 low-throughput & animal tests

New Approach: EDSP + Tox21 = EDSP21
  - Pathway-based predictive models (HTS in vitro assays)
  - Validate to replace selected Tier 1 screening assays
  - Train QSAR models to prioritize chemical universe
ER/AR QSAR Models

- Training set (ToxCast): 1677 chemical structures

- CERAPP: Global collaborative project for ER
  - 17 international groups participated
  - Individual and consensus models
  - Mansouri et al. 2016 EHP

- Prediction Set (EDSP):
  - 32,464 chemical structures
  - 5-10% predicted to be ER-active: Prioritize for further testing

- COMPARA: Global collaborative project for AR
  - 34 international groups participating
  - Mansouri et al. 2018 in prep
Most models predict most chemicals as inactive

757 chemicals have >75% positive concordance

CERAPP:
Only a small fraction of chemicals are prioritized for further testing
EDSP Priority: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP) (SGT)

Humans are potentially exposed to tens of thousands of man-made chemicals in the environment. It is well known that some environmental chemicals mimic natural hormones and disrupt endocrine systems.

EDSP dashboard: http://actor.epa.gov/edsp21/
Skin Sensitization

“Allergic Contact Dermatitis”

Accounts for 10-15% of all occupational disease (Anderson et al. 2010)

Major testing requirement for cosmetics, pesticides, industrial chemicals, etc.
U.S. Agency Requirements/Considerations

<table>
<thead>
<tr>
<th>Pesticides</th>
<th>Reference Method</th>
<th>Classification Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial chem</td>
<td>LLNA</td>
<td>NS S</td>
</tr>
</tbody>
</table>

| Household Products | LLNA | NS S SS |

| Dermatological Products | GPMT* | NS |

*human data preferred
Accuracy Against Human Clinical Data (~150 chems)

**LLNA**

- Hazard: 72%-82%
- Potency: 54% - 60%

**GPMT / Buehler**

- Hazard: ~72%
- Potency: ~60%

Reproducibility of Multiple Tests (~100 chems)

- Hazard: ~78%
- Potency: ~62%
Global Skin Sensitization Project

- Objective: analysis of available non-animal defined approaches (DAs)
- Collaboration with Cosmetics Europe
  - 128 substance dataset
  - LLNA (mouse) and human data
  - Curation/generation of
    - *in vitro* cell-based data that maps to AOP
    - *in silico* computer predictions, chemical structural features & properties
- Analyze non-animal DAs in an open source and transparent way
- Evaluate performance against the LLNA and human hazard/potency categories

Spectrum of 128 substances

Kleinstreuer et al. 2018 Crit Rev Tox in press
Research article

Prediction of skin sensitization potency using machine learning approaches

Qingda Zang, Michael Paris, David M. Lehmann, Shannon Bell, Nicole Kleinstreuer, and Warren Casey

ABSTRACT: The replacement of animal tests by computer models for the identification of potential dermal irritants and sensitizers is an important goal for regulatory authorities. We have previously built a prediction model for the Local Lymph Node Assay (LLNA), an in vivo test that uses animal data to assess skin sensitization potential. Here we use machine learning techniques to predict the potency of skin sensitizers and a model for predicting the potency of skin sensitizers is developed. The predicted LLNA potency values are compared with experimental data from the literature and show good agreement. The model performs well on an independent test set, and the results suggest that machine learning techniques can be used to predict the potency of skin sensitizers with high accuracy. The model can be used to identify potential skin sensitizers early in the development process and to prioritize further testing.

Multivariate models for prediction of human skin sensitization hazard

Judy Strickland, Qingda Zang, Michael Paris, David M. Lehmann, David Allen, Neepa Choksi, Joanna Matheson, Abigail Jacobs, Warren Casey, and Nicole Kleinstreuer

ABSTRACT: One of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) priorities is the development and evaluation of non-animal approaches to identify and predict skin sensitizers. The use of machine learning techniques for the prediction of skin sensitization hazard is an important step towards the replacement of animal tests. In this study, we developed multivariate models for predicting skin sensitization hazard using machine learning techniques. The models were trained on a dataset of skin sensitization data and tested on an independent test set. The results show that the models perform well in predicting skin sensitization hazard, and that machine learning techniques can be used to predict skin sensitization hazard with high accuracy. The models can be used to identify potential skin sensitizers early in the development process and to prioritize further testing.

Integrated decision strategies for skin sensitization hazard

Judy Strickland, Qingda Zang, Nicole Kleinstreuer, Michael Paris, David M. Lehmann, Neepa Choksi, Joanna Matheson, Abigail Jacobs, Anna Lowit, David Allen, and Warren Casey

ABSTRACT: One of the priorities of the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) is the development and evaluation of non-animal approaches for skin sensitization testing. The use of machine learning techniques for the prediction of skin sensitization hazard is an important step towards the replacement of animal tests. In this study, we developed integrated decision strategies for skin sensitization hazard using machine learning techniques. The strategies were trained on a dataset of skin sensitization data and tested on an independent test set. The results show that the strategies perform well in predicting skin sensitization hazard, and that machine learning techniques can be used to predict skin sensitization hazard with high accuracy. The strategies can be used to identify potential skin sensitizers early in the development process and to prioritize further testing.
Different Modeling Approaches

- Meta models
  - Regression equations
  - Ordinary differential equation
- Consensus model
  - 2 out of 3 WoE
  - Sequential Testing Strategy with defined decision criteria after each step
- Bayesian Networks
- Artificial Neural Networks
- Support vector machine

Prediction

Silvia Casati, JRC
## Types of Information Sources

<table>
<thead>
<tr>
<th>Method Type</th>
<th>Description</th>
</tr>
</thead>
</table>
| Adopted in *chemico, in vitro*  | 2 out of 3 WoE – BASF  
| methods                         | STS sequential strategy – Kao                                               |
| Adopted in *chemico, in vitro*  | ITS battery system -Kao                                                     |
| method(s)                       | SSWG SVM model - ICCVAM                                                    |
|                                 | Non-testing pipeline approach - G. Patlewicz                               |
|                                 | Decision Strategy – L'Oréal                                                |
|                                 | Bayesian Network – P&G                                                      |
| Adopted in *chemico, in vitro*  | Non-standard Method(s)                                                      |
| method(s)                       | Phys-chem properties                                                       |
|                                 | Artificial Neural Network model - Shiseido                                  |
| Adopted in *chemico, in vitro*  | Non-standard Method(s)                                                      |
| method(s)                       | In silico prediction(s)                                                     |
| Adopted in *chemico, in vitro*  | STS- RIVM                                                                  |
| method(s)                       | Sensitizer potency prediction- Givaudan                                     |
| Modified adopted test methods   | SARA model for risk assessment - Unilever                                  |
| *In silico* prediction(s)       | Consensus model - JRC                                                      |

*Silvia Casati, JRC*
Defined Approach Evaluation

Most non-animal defined approaches evaluated so far perform **better** than the LLNA at predicting human skin sensitization hazard and potency.

(And when compared to the LLNA, are equivalent in performance to the LLNA at predicting itself.)
International Harmonization

• OECD proposal (SPSF) submitted November 2016
  – Co-led by U.S., EU, and Canada
  – Create an international performance based test guideline for non-animal defined approaches to skin sensitization testing
  – Achieve widespread replacement of mouse test

• National coordinators (WNT) voted unanimously to include the project in OECD workplan, April 2017

• Special session of the WNT met in December 2017 to review progress and discuss next steps
  – Achieved consensus on evaluation framework for consideration and assessment of DAs
Rat oral acute toxicity: LD50 Database

- Multiple existing resources containing rat oral acute toxicity LD50 data were mined and merged

<table>
<thead>
<tr>
<th>Data source</th>
<th>Number of LD50 values</th>
<th>Number of unique chemicals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECHA ChemProp</td>
<td>5,533</td>
<td>2,136</td>
</tr>
<tr>
<td>NLM HSDB</td>
<td>3,981</td>
<td>2,205</td>
</tr>
<tr>
<td>JRC AcutoxBase</td>
<td>637</td>
<td>138</td>
</tr>
<tr>
<td>NLM ChemIDplus</td>
<td>13,072</td>
<td>12,977</td>
</tr>
<tr>
<td>NICEATM PAI</td>
<td>364</td>
<td>293</td>
</tr>
<tr>
<td>OECD eChemPortal</td>
<td>10,119</td>
<td>2,290</td>
</tr>
</tbody>
</table>

Total: 34,511 LD50 values 16,307 chemicals

- LD50 data comprised point estimates as well as limit tests

Agnes Karmaus, ILS/NICEATM
Development of Predictive Models for Acute Oral Toxicity

- International modeling community invited to build models to predict acute oral systemic toxicity
- ICCVAM agencies informed model endpoints
- Training and test data derived from large dataset compiled by NICEATM and EPA/NCCT
  - 11,992 QSAR-ready structures (75% training, 25% test)
  - Quantitative & qualitative evaluation
  - Models will be integrated to yield consensus predictions

https://ntp.niehs.nih.gov/go/tox-models
Endpoints to be Modeled

Participants are asked to develop models for any/all of the following endpoints identified based on regulatory needs provided by ICCVAM agencies:

1. Very toxic (< 50 mg/kg vs. all others)
2. Nontoxic (>2000 mg/kg vs. all others)
3. LD50 point estimate
4. EPA hazard categories (n=4)
5. GHS hazard categories (n=5)*

*GHS categories 5 and “not classified” are combined into one category
Modeling Approach

• Modelers are encouraged to consider different modeling approaches
  – Machine learning, global/local, hybrid/consensus models, etc.

• Models could include any variety of data inputs:
  – Chemical features/structure classes, physiochemical properties, product use categories, production volumes, in vitro data (measured or predicted), etc.
Evaluation Criteria

The OECD QSAR validation principles to be considered as guidance:

1. A defined endpoint
2. An unambiguous algorithm
3. A defined domain of applicability
4. Appropriate measures of goodness-of–fit, robustness and predictivity
5. Mechanistic interpretation, if possible
Timeline

• **November 17, 2017:** Release of Training Data to the public.

• **December 15, 2017:** Release of Prediction Data to the public.

• **February 9, 2018:** Deadline for submission of model results and documentation to NICEATM.

• **March 9, 2018:** Organizing Committee finalizes selection of models to be invited for platform presentations and notifications are sent to presenters.

• **April 11-12, 2018:** Predictive Models for Acute Oral Systemic Toxicity Workshop, NIH Natcher Conference Center, Bethesda, MA.

[https://ntp.niehs.nih.gov/go/tox-models](https://ntp.niehs.nih.gov/go/tox-models)
Summary

• Toxicology data can be synthesized and modeled effectively using machine learning approaches.
  – Also: exposure, use case, systematic review, etc.

• Machine learning models (i.e. QSARs) have already achieved limited acceptance in the regulatory space.

• Additional education, training, and communication will facilitate more widespread adoption.

https://ntp.niehs.nih.gov/pubhealth/evalatm/natl-strategy/
Acknowledgments

• Gino Scarano
• Jennifer Goode
• Paul Browne
• Agnes Karmaus
• Grace Patlewicz
• Rusty Thomas
• Kamel Mansouri
• Richard Judson
• Sebastian Hoffmann
• Silvia Casati

• ICCVAM partners
• Cosmetics Europe STTF
• ILS/NICEATM
• EURL ECVAM/JRC
• Health Canada
• ICATM partners
Extra Slides
Previous CoP webinars

• Incorporating Chemical Information: Resources, Limitations, and Characterizing the Domain of Applicability for 21st Century Toxicity Testing (January 24, 2017)
  
  • https://ntp.niehs.nih.gov/go/commprac-2017

• Fundamentals of Using Quantitative Structure-Activity Relationship Models and Read-across Techniques in Predictive Toxicology (January 26, 2016)
  
  • https://ntp.niehs.nih.gov/go/commprac-2016
Big biological data in toxicology

<table>
<thead>
<tr>
<th>Name</th>
<th>General Information</th>
<th>Data description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubChem</td>
<td>Over 50 million compounds, over 700,000 bioassays, over 13 billion data points</td>
<td>Toxicity, genomics and literature data</td>
</tr>
<tr>
<td>ChEMBL</td>
<td>Over 600,000 compounds, 3.3 million bioassay readout data</td>
<td>Literature toxicity data</td>
</tr>
<tr>
<td>ACToR</td>
<td>The toxicity results from 100 various data resources</td>
<td>Both in vitro and in vivo toxicity data</td>
</tr>
<tr>
<td>ToxNET</td>
<td>Over 50,000 environmental compounds from 16 different resources</td>
<td>Both in vitro and in vivo toxicity data</td>
</tr>
<tr>
<td>SEURAT</td>
<td>Over 5,500 cosmetic-type compounds in the current COSMOS database web portal</td>
<td>Animal toxicity data</td>
</tr>
<tr>
<td>REACH</td>
<td>816,048 studies for 9,800 substances and 3,600 study types</td>
<td>Data submitted in EU chemical legislation, made machine-readable by Luechtefeld et al. 2016a (this issue)</td>
</tr>
<tr>
<td>CTD</td>
<td>Over 13,000 compounds, over 32,000 genes, over 6000 diseases</td>
<td>Compound, gene and disease relationships</td>
</tr>
<tr>
<td>CEBS</td>
<td>About 10,000 toxicity bioassays from various sources</td>
<td>Gene expression data</td>
</tr>
<tr>
<td>DrugMatrix</td>
<td>About 600 drug molecules and 10,000 genes</td>
<td>Gene expression data</td>
</tr>
<tr>
<td>Cmap</td>
<td>About 1,300 compounds and 7,000 genes</td>
<td>Gene expression data</td>
</tr>
</tbody>
</table>

Zhu et al. 2016 ALTEX
Impact of Variability on Hazard Classification

LD50 (log10(mg/kg))

GHS I
GHS II
GHS III
GHS IV
GHS V

EPA I
EPA II
EPA III
EPA IV

test_type
- experimental value
- limit test (max, less than)
- limit test (min, greater than)