Balancing Machine Learning and Mechanistic Modeling

Nicole C. Kleinstreuer
NICEATM Deputy Director

19th September, SACATM, Washington DC
• Two “competing” approaches to modern toxicology/drug discovery:

1) Build testing strategies/models based exclusively on existing biological knowledge

2) Generate as much data as possible and let the machines sort it out

• Success lies in leveraging both approaches

• BUT this requires appropriate toolkits, resources and support infrastructure
Predictive Toxicology Vision

FAIR Resources

Big Data
(e.g. clinical records, HTS, systematic review)

Mechanistic Models

Assay Development, Data Generation

AOPs, Agent Based Models

Predictive Models

Machine Learning / Artificial Intelligence

Chemical, Target Prioritization

Experimentation
(e.g. 3D organoids, tissue chips, targeted bioassays)

Regulatory/Safety/Efficacy Decisions
## FAIR Principles

<table>
<thead>
<tr>
<th>Findable</th>
<th>A data object should be uniquely and persistently identifiable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessible</td>
<td>Data is accessible by authorized users (human and machine) through a well-defined protocol.</td>
</tr>
<tr>
<td>Interoperable</td>
<td>(Meta) data assigned to the data object is syntactically parse-able and semantically machine accessible.</td>
</tr>
<tr>
<td>Reusable</td>
<td>Data objects must comply with the above three principles and sufficiently documented to allow integration/linkage with other data sources.</td>
</tr>
</tbody>
</table>

[Wilkinson et al. 2016](https://www.force11.org/fairprinciples)
NIEHS Data Commons – initial stage

- Internal research data & metadata
- Capture, access, control, search, and sharing
- Engage external stakeholders
Interoperability Across Systems

Consistent & compatible web-APIs

CEBS

ICE

Data Commons

Others…

NIEHS Data Systems

Consistent data set access & retrieval

PubChem

Many others…

EPA Dashboard

NCATS BioPlanet
Goals of ICE

- Uphold FAIR principles for ICCVAM Data
- Provide intuitive access to high quality (curated) data and tools to support:
  - chemical evaluations,
  - data integration,
  - informatics analyses, and
  - model development
- Enable wider community to engage in the use of alternative and computational approaches for assessing chemical safety
31 October 2019: Release of Final Core Trustworthy Data Repositories Requirements 2020–2022

- R0: Context
- R1: Mission/Scope
- R2: Licenses
- R3: Continuity of Access
- R4: Confidentiality/Ethics
- R5: Organizational Infrastructure
- R6: Expert Guidance
- R7: Data Integrity and Authenticity
- R8: Appraisal
- R9: Documented Storage Procedures
- R10: Preservation Plan
- R11: Data Quality
- R12: Workflows
- R13: Data Discovery and Identification
- R14: Data Reuse
- R15: Technical Infrastructure
- R16: Security

https://www.coretrustseal.org/
What goes into ICE?

- Validation Studies
- Databases
- Published Data
- Computational Models
# What data are currently in ICE?

<table>
<thead>
<tr>
<th>Toxicity endpoint/ Data source</th>
<th>Assays</th>
<th># of chemicals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Oral Toxicity</td>
<td>Acute oral toxicity</td>
<td>10,348</td>
</tr>
<tr>
<td>Skin Sensitization</td>
<td>DPRA, hCLAT, KeratinoSens, LLNA, human potency, etc</td>
<td>578</td>
</tr>
<tr>
<td>Skin Irritation</td>
<td>Acute skin irritation/corrosion, 4h HPT</td>
<td>120</td>
</tr>
<tr>
<td>Eye Irritation</td>
<td>Acute eye irritation/corrosion (e.g, Draize eye), Vitrigel</td>
<td>183</td>
</tr>
<tr>
<td>Endocrine</td>
<td>AR/ER Pathway Models, Uterotrophic, AR/ER binding</td>
<td>1903</td>
</tr>
<tr>
<td>cHTS</td>
<td>ToxCast and Tox21 assays</td>
<td>9076</td>
</tr>
<tr>
<td>OPERA predictions</td>
<td>BP, HLC, KOA, BCF, LogP, MP, MW, VP, WS</td>
<td>705,666</td>
</tr>
<tr>
<td>Formulation data</td>
<td>Acute 6-pack</td>
<td>298 (747 formulations)</td>
</tr>
</tbody>
</table>
Call for Data

Currently of interest:

• *In vivo* data
  – Collections of data generated using regulatory guideline-like studies
  – Acute inhalation, skin and eye irritation/corrosion

• Toxicokinetic data
  – Collections of *in vivo* measurements
  – Data from *in vitro* assays aimed at informing modeling of chemical ADME within the body

ICE-support@niehs.nih.gov
Tools

Machine Learning

Chemical Space Characterization

IVIVE

\[ C_{ss} = \frac{\text{Dose}}{CL_r + CL_h} \]
**IVIVE**

**Tools**

\[ C_{ss} = \frac{\text{Dose}}{CL_r + CL_h} \]

**In Vivo Effects**

- Target plasma concentration
- Range of equivalent administered doses (EADs) (mg/kg/day)

**In Vitro HTS Assays**

- Activity concentration (AC)
- Equivalent

**Experimental measurement**

- Hepatic clearance, fraction unbound to plasma protein (fu), \( K_p \), gut absorption, etc

**One-compartment PK or PBPK models**

- Plasma concentration at dose of 1 mg/kg/day (\( C_{ss} \), \( C_{ss} \) *fu* or \( C_{max} \))

**QSAR or QPPR prediction**

**Reverse dosimetry**

- \( AC \ast \left( \frac{1 \text{ mg/kg/day}}{\text{plasma concentration at dose of 1 mg/kg/day}} \right) \)
• Overlay in vivo data
• Update the plots with assay filtering
• Download data to explore locally
Predicting Key Toxicokinetic Parameters

Multiple Machine Learning Models (SVM, DNN, XGB, etc.)

- ADME properties
  - Plasma fraction unbound (FuB)
  - Intrinsic clearance (Clint)

- Tissue partition coefficient inputs
  - pKa
  - Log D

Mansouri et al. 2019 Journal of Cheminformatics in press

https://github.com/NIEHS/OPERA

https://github.com/NIEHS/OPERA
Global Collaborative Projects

Applying machine learning to predict endpoints of regulatory importance

CERAPP
Collaborative Estrogen Receptor Activity Prediction Project (2015/16)

CoMPARA
Collaborative Modeling Project for Androgen Receptor Activity (2017/18)

CATMoS
Collaborative Acute Toxicity Modeling Suite (2018/19)

Endocrine Disruptor Screening Program (EDSP)

ICCVAM Acute Systemic Toxicity Workgroup

Mansouri et al. 2016 EHP 124:1023–1033
Mansouri et al. 2019 under revision at EHP
Kleinstreuer et al. 2018 Comp Tox; Mansouri et al. 2019 in prep
Manually Identifying Reference Data

- Systematic literature search of publically available data (e.g. PubMed, Scopus)
- Identify chemical activities measured in “guideline-like” uterotrophic studies
- Identify a subset of *in vivo* reference chemicals
  - Active chemicals verified in ≥2 independent studies
  - Inactive chemicals verified in ≥2 independent studies (with no positive results in any study)

*Ex: Uterotrophic Database*

*Kleinstreuer et al. EHP (2015)*
Automating Reference Data Identification

- Project with Oak Ridge National Labs (ORNL) and FDA CFSAN to apply text-mining (NLP) approaches & ML to identify high-quality data

- Semi-automated retrieval and evaluation of published literature (trained on uterotrophic database)

- Apply to developmental toxicity studies (with ICCVAM DARTWG)
  - Define literature search keywords, identify corpus
  - Extract/characterize study protocol details from regulatory guidelines: minimum criteria
  - Apply ML algorithms to identify high-quality studies, expert check
Study Extractions and Endpoint Mapping

- Extract study details from prenatal developmental toxicity guideline studies
  - NTP legacy studies
  - ECHA submissions (expert reviewed for quality)
- Map results to controlled vocabularies/ontologies
  - UMLS (ToxRefDBv2.0)
  - EPA/BfR DevTox DB
  - OECD Harmonized Templates
Flipping the Paradigm: Mechanistic Screening

X lbs./yr. commercial production

Initial Focus

https://ncats.nih.gov/tissuechip/chip

Mortality
Selected Causes of Death

Figure 2. Age-adjusted death rates for selected causes of death for all ages, by sex: United States, 2004–2014

 Courtesy of B. Berridge
Cardiovascular Health Effects Strategy

Map cardiovascular ‘failure modes’ and agents

Mine current databases for CV-relevant MOA data

Define primary mechanistic screening strategy

Define and test Tier II 3D cardiomyocyte and vascular modeling systems

Identify translational biomarkers of chronic CV human health effects

CV Evidence-based Testing Paradigm

Knowledge Integration

Chronic \textit{in vivo} Studies

Short term \textit{in vivo} Studies

\textit{In vitro} Studies

Data / Knowledge Mining

QSAR Profiling

Bioactivity Screening

Communicate

Regulatory and public health stakeholders

Contributing projects Partners and opportunities

NHLBI-NIEHS partnership

NCATS, IQ, NCTR partnerships

Adapted from B. Berridge - NTP BSC Presentation - Dec. 2018
## Cellular Events Linked to CV Failure Modes

### Drug actions on human receptors, ion channels, cellular processes

<table>
<thead>
<tr>
<th>Category</th>
<th>Events</th>
<th>Receptors/Channels</th>
<th>Pharmacological Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delta Vasoactivity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Delta Inotropy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Valvular injury/proliferation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Endothelial injury/coagulation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Systemic Hypertension
- Antagonism at α2-receptors
- Rebound phenomenon (α2-agonists and beta-blockers)
- Agonism at glucocorticoid receptors
- Inhibition of VEGF pathway
- Inhibition of monoamine oxidases (MAO)
- 11β-hydroxysteroid dehydrogenase type 2 inhibition

### Systemic Hypotension
- Antagonism at α1-receptors
- Ca2+ channel blockade
- Opening of K+ channels
- Inhibition or reversion of angiotensin—aldosterone axis
- Agonism at β2 receptors
- Agonism at α2-receptors
- Agonism at I1-receptors
- Stimulation of cGMP Synthesis
- Inhibition of phosphodiesterase 5

### Left ventricular (LV) dysfunction/heart failure
- Ca2+ channel blockade
- Na+ channel blockade
- Antagonism at β1-receptors
- Antagonism at β3-receptors
- Anthracyclines, cyclophosphamide (high dose), taxanes
- HER2 signaling inhibition
- VEGF signaling inhibition
- Tyrosine kinase inhibition (multikinase drugs)
- Proteasome inhibition

### Bradydysrhythmias
- Blockade of hERG channels
- Agonism at β-receptors
- Antagonism at M-receptors
- Antagonism at I1-receptors

### Tachydysrhythmias
- Blockade of hERG channels
- Agonism at β-receptors
- Antagonism at M-receptors
- Antagonism at α2-receptors
- Agonism at I1-receptors
- Stimulation of cGMP Synthesis
- Inhibition of phosphodiesterase 5

### Myocardial ischemia
- Agonism at β1-receptors (direct effects or indirect effect via endogenous catecholamines)
- Rebound phenomenon (nitrates, β-blockers)

### Myocarditis
- Autoimmune reactions (e.g., clozapine)
- Monoclonal antibodies targeting PD-1

### Impairment of cardiac valves
- Agonism at 5-HT2B receptors

### Pericardial disease induction
- Immune reaction (e.g., drugs inducing lupus erythematosus)

### Arterial
- Inhibition of cyclooxygenase 2
- VEGF targeting
- Agonism at erythropoietin receptors

### Venous
- Agonism at estrogenic receptors
- VEGF targeting
- Agonism at erythropoietin Receptors
CardioToxPi: HTS Assay Mapping
Vascular Development & Disruption

Adverse Outcome Pathway (AOP)

Vasculogenesis

Primary tubular network

Angiogenesis

Remodeling

Adverse outcomes:
- Placenta: Nutrient exchange, Altered physiology, Impaired blood flow
- Embryo-Fetus: Altered hemodynamics, Impaired growth, Dysmorphogenesis, Altered differentiation


AOP43: one of 28 AOPs included in the OECD work plan with status ‘open for citation & comment’ [https://aopwiki.org/wiki/index.php/Aop:43]
Mechanistic Models and Experimental Results

| 38 chemical test set: qualification of pVDC ToxPi across 9 endothelial behaviors |
|------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| A pVDC score from ToxCast dataset (ToxPi) | B HUVEC tubulogenesis (FICAM) | C tubulogenesis in synthetic matrices | D tubulogenesis in Matrigel | E nuCTNB biomarker (EndMT) | F endothelial cell migration | G sprouting assay (iPSC-derived endothelial cells) | H reporter zebrafish (ISV outgrowth) | I reporter zebrafish (hyaloid vascular network) | J HUVEC tubulogenesis (VALA) | K ANY (B to J) |
| Decane | 1,2,3-Trichloropropane | Pymetrozine | Methimazole | Imazamox | D-Mannitol | Methylparaben | Valproic acid | Tris(2-ethylhexyl) phosphate | PFOS | 1,2,4-Trichlorobenzene |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1,2,3-Trichloropropane | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pymetrozine | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Methimazole | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Imazamox | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| D-Mannitol | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Methylparaben | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Valproic acid | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tris(2-ethylhexyl) phosphate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PFOS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1,2,4-Trichlorobenzene | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TNP-470 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Reserpine | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sodium dodecylbenzenesulfonate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4-Nonylphenol, branched | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tris(2-chloroethyl) phosphate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2,4-Diaminotoluene | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tris[1,3-dichloro-2-propyl]phosphate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Oxetetracycline dihydrate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Celecoxib | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Quercetin | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C.I. Solvent Yellow 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Triclosan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bisphenol AF | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Docusate sodium | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| tert-Butylhydroquinone | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Haloperidol | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gadrabine | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Triclocarban | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pyridaben | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1-Hydroxypyrene | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Disulfiram | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Fluazinam | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bisphenol A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Phenolphthalein | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Octyl gallate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SNP-33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Saili et al 2019 Current Opinion in Toxicology
Predicting Toxicity of Mixtures

• How can we leverage machine learning, mechanistic modeling, and systems approaches to tackle complex problems such as predicting mixtures toxicity across heterogeneous populations?

• How do we build datasets that will allow models for mixtures toxicity against human health endpoints to be more effectively developed?
Ocular QSAR Mixture Models

- EPA_ANY = Category I, II, III vs Category IV
- EPA_IRR = Category I/II vs Category III/IV
- EPA_Corr = Category I vs Category II, III, IV

Sedykh et al. SOT 2019
Predictive Toxicology Vision

FAIR Resources

Big Data
(e.g. clinical records, HTS, systematic review)

Machine Learning / Artificial Intelligence

Predictive Models

Experimentation
(e.g. 3D organoids, tissue chips, targeted bioassays)

Mechanistic Models

AOPs, Agent Based Models

Assay Development, Data Generation

Regulatory/Safety/Efficacy Decisions
Predicting Human Toxicity

Mechanistic Models

Predictive Models

Experimentation

Big Data

3-D Tissue Construct Models
Acknowledgments

- ILS/NICEATM group
- Sciome collaborators
- ICCVAM agencies
- ICATM partners
- Brian Berridge (NTP)
- Shagun Krishna (Kelly)
- Robert Patton (ORNL)
- Jessica Wignall (ICF)
- Tom Knudsen (EPA)
- Todd Zurlinden (EPA)
- Kate Saili (EPA)
- Tony Williams (EPA)

Questions?