The Systematic Empirical Evaluation of Models (SEEM) framework:
High Throughput Exposure Science for Chemical Decision Making

John F. Wambaugh

National Center for Computational Toxicology
Office of Research and Development
United States Environmental Protection Agency
Research Triangle Park, North Carolina 27711

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

https://orcid.org/0000-0002-4024-534X
• Park *et al.* (2012): At least 3221 chemical signatures in pooled human blood samples, many appear to be exogenous

• A tapestry of laws covers the chemicals people are exposed to in the United States (Breyer, 2009)

• Different testing requirements exist for food additives, pharmaceuticals, and pesticide active ingredients (NRC, 2007)
Most other chemicals, ranging from industrial waste to dyes to packing materials, are covered by the Toxic Substances Control Act (TSCA).

Thousands of chemicals on the market were “grandfathered” in without assessment.

“Tens of thousands of chemicals are listed with the Environmental Protection Agency (EPA) for commercial use in the United States, with an average of 600 new chemicals listed each year.”

U.S. Government Accountability Office
• TSCA was updated in June, 2016 to allow more rapid evaluation of chemicals (Frank R. Lautenberg Chemical Safety for the 21st Century Act)

• New approach methodologies (NAMs) are being considered to inform prioritization of chemicals for testing and evaluation (Kavlock et al., 2018)

• EPA has released a “A Working Approach for Identifying Potential Candidate Chemicals for Prioritization” (September, 2018)
Risk = Hazard x Exposure

High throughput screening (Dix et al., 2006, Collins et al., 2008) + in vitro-in vivo extrapolation (IVIVE, Wetmore et al., 2012, 2015) can predict a dose (mg/kg bw/day) that might be adverse.

Need methods to forecast exposure for thousands of chemicals (Wetmore et al., 2015).

High throughput models exist to make predictions of exposure via specific, important pathways such as residential product use and diet.

NRC (1983)
Translation of high-throughput data into risk-based rankings is an important application of exposure data for chemical priority-setting. Recent advances in high-throughput toxicity assessment, notably the ToxCast and Tox21 programs... and in high-throughput computational exposure assessment... have enabled first-tier risk-based rankings of chemicals on the basis of margins of exposure...”

“...The committee sees the potential for the application of computational exposure science to be highly valuable and credible for comparison and priority-setting among chemicals in a risk-based context.”
Limited Available Data for Exposure Estimation

Most chemicals lack public exposure-related data beyond production volume (Egeghy et al., 2012)
What Do We Know About Exposure?

Biomonitoring Data

- Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey (NHANES) provides an important tool for monitoring public health
- Large, ongoing CDC survey of US population: demographic, body measures, medical exam, biomonitoring (health and exposure), ...
- Designed to be representative of US population according to census data
- Data sets publicly available (http://www.cdc.gov/nchs/nhanes.htm)
- Includes measurements of:
  - Body weight
  - Height
  - Chemical analysis of blood and urine
What Do We Know About Exposure? Exposure Models

• Human chemical exposures can be coarsely grouped into “near field” sources that are close to the exposed individual (consumer or occupational exposures) ‘far-field’ scenarios wherein individuals are exposed to chemicals that were released or used far away (ambient exposure) (Arnot et al., 2006).

• A model captures knowledge and a hypothesis of how the world works (MacLeod et al., 2010)

• EPA’s EXPOsure toolBOX (EPA ExpoBox) is a toolbox created to assist individuals from within government, industry, academia, and the general public with assessing exposure
  • Includes many, many models
  https://www.epa.gov/expobox

“Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations... The only question of interest is ‘Is the model illuminating and useful?’” George Box
Forecasting Exposure is a Systems Problem

- **USE and RELEASE**
  - Consumer Products and Durable Goods
  - Direct Use (e.g., surface cleaner)
  - Residential Use (e.g., flooring)
  - Occupational Use

- **MEDIA**
  - Indoor Air, Dust, Surfaces
  - Food
  - Drinking Water
  - Outdoor Air, Soil, Surface and Ground Water

- **TARGET**
  - Human
  - Ecological Flora and Fauna

Figure from Kristin Isaacs
Source-based Exposure Pathways

USE and RELEASE

- Consumer Products and Durable Goods
- Near-Field Direct
- Near-Field Indirect
- Residential Use (e.g., flooring)
- Occupational Use (e.g., surface cleaner)

TARGET

- Human
- Ecological Flora and Fauna

EXPOSURE (MEDIA + TARGET)

- Indoor Air, Dust, Surfaces
- Food
- Drinking Water
- Outdoor Air, Soil, Surface and Ground Water
- Waste

MEDIA

- Other Industry
- Chemical Manufacturing and Processing
- Environmental Release

The exposure event itself is often unobservable

Figure from Kristin Isaacs
We can try to predict exposure by describing the process leading to exposure.
We can also infer exposure from monitoring data.
Evaluating Models with Monitoring Data

**USE and RELEASE**
- Consumer Products and Durable Goods
  - Direct Use (e.g., flooring)
  - Residential Use (e.g., flooring)
- Near-Field Direct
- Near-Field Indirect
- Occupational Use
- Food
- Drinking Water
- Waste
- Outdoor Air, Soil, Surface and Ground Water
- Chemical Manufacturing and Processing

**MEDIA**
- Indoor Air, Dust, Surfaces
- Ecological Flora and Fauna
- Ecological

**EXPOSURE (MEDIA + TARGET)**
- Occupational
- Dietary
- Far-Field

**TARGET**
- Human
- Media Samples
- Biomarkers of Exposure
- Inference ("Reverse Modeling")

**MONITORING DATA**
- Predictive Modeling
- Evaluating Models with Monitoring Data
- Inference ("Reverse Modeling")

Figure from Kristin Isaacs
Consensus Exposure Predictions with the SEEM Framework

- Different exposure models incorporate knowledge, assumptions, and data (MacLeod et al., 2010)
- We incorporate multiple models into consensus predictions for 1000s of chemicals within the Systematic Empirical Evaluation of Models (SEEM) (Wambaugh et al., 2013, 2014)

• Evaluation is similar to a sensitivity analysis: What models are working? What data are most needed?
Multiple regression models:

\[ \text{Log(Parent Exposure)} = a + m \times \text{log(Model Prediction)} + b \times \text{Near Field} + \varepsilon \]

\[ \varepsilon \sim \text{N}(0, \sigma^2) \]

Residual error, unexplained by the regression model.
Multiple regression models:

\[
\text{Log(Parent Exposure)} = a + m \times \text{log(Model Prediction)} + b \times \text{Near Field} + \varepsilon
\]

Not all models have predictions for all chemicals
- We can run SHEDS-HT (Isaacs et al., 2014) for ~2500 chemicals

What do we do for the rest?
- Assign the average value?
- Zero?
Those chemicals with “near-field” – proximate, in the home, sources of exposure – had much higher rates of exposure than those with sources outside the home (Wallace et al., 1986).

The only available “high throughput exposure models in 2013 were for far-field sources.
R² ≈ 0.5 indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates.

Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:
- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume
Heuristics of Exposure

Wambaugh et al. (2014) $R^2 \approx 0.5$ indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates.

Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume
“In particular, the assumption that 100% of [quantity emitted, applied, or ingested] is being applied to each individual use scenario is a very conservative assumption for many compound / use scenario pairs.”
What Do We Know About Chemical Use?
Chemicals and Products Database

Broad “index” of chemical uses

CPCat

MSDS Data

Occurrence and quantitative chemical composition

CPDat

Green Chemistry

The roles that chemicals serve in products

https://comptox.epa.gov/dashboard

Ingredient Lists

Occurrence data

Measured Data

Measurement of chemicals in consumer products

Functional Use Data

Slide from Kristin Isaacs
Collaboration on High Throughput Exposure Predictions
Jon Arnot, Deborah H. Bennett, Peter P. Egeghy, Peter Fantke, Lei Huang, Kristin K. Isaacs, Olivier Jolliet, Hyeong-Moo Shin, Katherine A. Phillips, Caroline Ring, R. Woodrow Setzer, John F. Wambaugh, Johnny Westgate

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Reference(s)</th>
<th>Chemicals</th>
<th>Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPA Inventory Update Reporting and Chemical Data Reporting (CDR) (2015)</td>
<td>US EPA (2018)</td>
<td>7856</td>
<td>All</td>
</tr>
<tr>
<td>United Nations Environment Program and Society for Environmental Toxicology and Chemistry toxicity model (USEtox) Industrial Scenario (2.0)</td>
<td>Rosenbaum et al. (2008)</td>
<td>8167</td>
<td>Far-Field Industrial</td>
</tr>
<tr>
<td>USEtox Pesticide Scenario (2.0)</td>
<td>Fantke et al. (2011, 2012, 2016)</td>
<td>940</td>
<td>Far-Field Pesticide</td>
</tr>
<tr>
<td>Risk Assessment IDentification And Ranking (RAIDAR) Far-Field (2.02)</td>
<td>Arnot et al. (2008)</td>
<td>8167</td>
<td>Far-Field Pesticide</td>
</tr>
<tr>
<td>Fugacity-based INdoor Exposure (FINE) (2017)</td>
<td>Bennett et al. (2004), Shin et al. (2012)</td>
<td>645</td>
<td>Residential</td>
</tr>
<tr>
<td>RAIDAR-ICE Near-Field (0.803)</td>
<td>Arnot et al., (2014), Zhang et al. (2014)</td>
<td>1221</td>
<td>Residential</td>
</tr>
<tr>
<td>USEtox Residential Scenario (2.0)</td>
<td>Jolliet et al. (2015), Huang et al. (2016,2017)</td>
<td>615</td>
<td>Residential</td>
</tr>
<tr>
<td>USEtox Dietary Scenario (2.0)</td>
<td>Jolliet et al. (2015), Huang et al. (2016), Ernstoff et al. (2017)</td>
<td>8167</td>
<td>Dietary</td>
</tr>
</tbody>
</table>
High-Throughput Stochastic Human Exposure and Dose Simulation Model (SHEDS-HT)

- We treat models like related assays and look for consensus while considering model appropriateness.
- High-throughput model for simulating population exposures to chemical in consumer products via multiple product types, scenarios, and routes.
- R package, code, and default input files for consumer products (derived from CPDat) are available:

https://github.com/HumanExposure/SHEDSHTRPackage

Isaacs et al. 2014
- We treat models like related assays and look for consensus while considering model appropriateness.
- United Nations Environment Program (UNEP) and Society for Environmental Toxicology and Chemistry (SETAC) toxicity model Version 2.0
- USEttox is a global scientific consensus fate, exposure and effect model.
- USEttox consists of a set of nested environmental compartments at indoor, urban, continental, and global scale.

Rosenbaum et al. 2008
We treat models like related assays and look for consensus while considering model appropriateness.

The Risk Assessment IDentification And Ranking (RAIDAR) model is an environmental fate and transport model linked with food web bioaccumulation models for representative ecological and agricultural targets and humans.
SEEM3

Total Chemical Intake Rate (mg/kg BW/day)

Pathway
- Consumer
- Dietary
- Far-Field Pesticides
- Far-Field Industrial
- Unknown

Chemical-Specific Pathway Relevancy ($\delta_{ij}$)

Yes/No

Predictors
- Average Unexplained ($a_{\text{consumer}}$)
- SHEDS-HT
- FINE
- RAIDAR-ICE
- USEtox
- Production Volume
- Average Unexplained ($a_{\text{dietary}}$)
- SHEDS-HT Dietary
- Production Volume
- USEtox
- RAIDAR
- Food Contact Substance Migration
- Average Unexplained ($a_{\text{FFpesticide}}$)
- Pesticide REDs
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained ($a_{\text{FFindustrial}}$)
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained ($a_0$, the grand mean)
Reverse Dosimetry (Tan et al., 2006)

- Median chemical intake rates (mg / kg body weight /day) were inferred from:
  - NHANES serum/blood either using HTTK clearance (Pearce et al., 2017)
  - Literature clearance estimates were used for methodologically challenging chemicals not suited to HTTK
Reverse Dosimetry (Tan et al., 2006)
**Total Chemical Intake Rate**

- **Pathway**
  - Consumer
  - Dietary
  - Far-Field Pesticides
  - Far-Field Industrial
  - Unknown

**Chemical-Specific Pathway Relevancy** ($\delta_{ij}$)

- Predictors
  - Average Unexplained ($a_{\text{consumer}}$)
    - SHEDS-HT
    - FINE
    - RAIDER-ICE
    - USEtox
    - Production Volume
    - Average Unexplained ($a_{\text{dietary}}$)
    - SHEDS-HT Dietary
    - Production Volume
    - USEtox
    - RAIDER
    - Food Contact Substance Migration
  - Average Unexplained ($a_{FF\text{pesticide}}$)
    - Pesticide REDs
    - USEtox
    - RAIDER
    - Stockholm Convention
    - Production Volume
    - Average Unexplained ($a_{FF\text{industrial}}$)
    - USEtox
    - RAIDER
    - Stockholm Convention
    - Production Volume
    - Average Unexplained ($a_0$, the grand mean)
Evaluation Data

Total Chemical Intake Rate (mg/kg BW/day)

Intake Rates Inferred from NHANES

- NHANES serum/blood either using HTTK clearance (Pearce et al., 2017)
- Literature clearance estimates were used for methodologically challenging chemicals not suited to HTTK

Ring et al., 2019
Total Chemical Intake Rate (mg/ kg BW/ day)

- Consumer
  - Yes/No
- Dietary
  - Yes/No
- Far-Field Pesticides
  - Yes/No
- Far-Field Industrial
  - Yes/No
- Unknown

Chemical-Specific Pathway Relevancy ($\delta_{ij}$)

Predictors
- Average Unexplained ($a_{\text{consumer}}$)
- SHEDS-HT
- FINE
- RAIDAR-ICE
- USEtox
- Production Volume
- Average Unexplained ($a_{\text{dietary}}$)
- SHEDS-HT Dietary
- Production Volume
- USEtox
- RAIDAR
- Food Contact Substance Migration
- Average Unexplained ($a_{\text{FFpesticide}}$)
- Pesticide REDs
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained ($a_{\text{FFindustrial}}$)
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained ($a_0$, the grand mean)
Likelihood of exposure via various source-based pathways is predicted from production volume, OPERA physico-chemical properties and ToxPrint structure descriptors.

Machine learning (Random Forest) – generates a chemical specific probability of exposure by that pathway (which is then used as a Bayesian prior).
Total Chemical Intake Rate (mg/ kg BW/ day)

Pathway
Consumer

Chemical-Specific Pathway Relevancy ($\delta_{ij}$)

Yes/No

Bayesian analysis via Markov Chain Monte Carlo assigns each chemical either a “Yes” or “No” according to predicted probability

If the pathway is no for a chemical, nothing is added to the intake rate

Predictors
Average Unexplained ($a_{\text{consumer}}$)
SHEDS-HT
FINE
RAIDAR-ICE
USEtox
Production Volume

Linear regression is used to estimate the average unexplained exposure (intercept) and loadings (slopes, or predictive ability) for each model

Model predictions are centered at zero – if there is no prediction for a chemical, the average value “zero” is added
Total Chemical Intake Rate (mg/ kg BW/ day)

- **Consumer**
  - Chemical-Specific Pathway Relevancy ($\delta_{ij}$)
  - Predictors:
    - Average Unexplained ($a_{consumer}$)
    - SHEDS-HT
    - FINE
    - RAIDAR-ICE
    - USEtox
    - Production Volume
- **Dietary**
  - Predictors:
    - Average Unexplained ($a_{dietary}$)
    - SHEDS-HT Dietary
    - Production Volume
    - USEtox
    - RAIDAR
    - Food Contact Substance Migration
- **Far-Field Pesticides**
  - Predictors:
    - Average Unexplained ($a_{FFpesticide}$)
    - Pesticide REDs
    - USEtox
    - RAIDAR
    - Stockholm Convention
    - Production Volume
    - Average Unexplained ($a_{FFindustrial}$)
- **Far-Field Industrial**
  - Predictors:
    - Average Unexplained ($a_{FFindustrial}$)
    - USEtox
    - RAIDAR
    - Stockholm Convention
    - Production Volume
    - Average Unexplained ($a_0$, the grand mean)
- **Unknown**
First Generation SEEM

Wambaugh et al., 2013

- Those chemicals with “near-field” – proximate, in the home, sources of exposure – had much higher rates of exposure than those with sources outside the home (Wallace et al., 1986)

- The only available “high throughput exposure models in 2013 were for far-field sources
The 3rd Gen. SEEM framework incorporates the previous models.

**SEEM I**

- **Total Chemical Intake Rate** (mg/kg BW/day)
  - **Pathway**
    - Consumer
    - Dietary
    - Far-Field Pesticides
    - Far-Field Industrial
    - Unknown
  - **Pathway Relevancy** ($\delta_{ij}$)
    - Yes
    - No

**Predictors**

- Average Unexplained ($a_{consumer}$)
- Average Unexplained ($a_{FFpesticide}$)
- Average Unexplained ($a_{FFindustrial}$)
- Average Unexplained ($a_0$, the grand mean)

We were unfair to USEtox and RAIDAR in that we judged them on all chemicals, not just those that with far-field sources.
Total Chemical Intake Rate (mg/kg BW/day)

Pathway
- Consumer
- Dietary
- Far-Field Pesticides
- Far-Field Industrial
- Unknown

Chemical-Specific Pathway Relevancy \( (\delta_{ij}) \)

Predictors
- Average Unexplained \( (a_{\text{consumer}}) \)
- SHEDS-HT
- FINE
- RAIDAR-ICE
- USEtox
- Production Volume
- Average Unexplained \( (a_{\text{dietary}}) \)
- SHEDS-HT Dietary
- Production Volume
- USEtox
- RAIDAR
- Food Contact Substance Migration
- Average Unexplained \( (a_{\text{FFpesticide}}) \)
- Pesticide REDs
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained \( (a_{\text{FFindustrial}}) \)
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained \( (a_0, \text{ the grand mean}) \)
Heuristics of Exposure

Wambaugh et al. (2014) R² \approx 0.5 indicates that we can predict 50% of the chemical to chemical variability in median NHANES exposure rates.

Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume
SEEM2

Total Chemical Intake Rate (mg/kg BW/day)

Pathway
- Consumer
- Far-Field Pesticides
- Far-Field Industrial
- Unknown

Pathway Relevancy ($\delta_{ij}$)
- Yes/No

Chemical-Specific Pathway Relevancy

Predictors
- Average Unexplained ($a_{\text{consumer}}$)
- Average Unexplained ($a_{\text{FFpesticide}}$)
- Average Unexplained ($a_{\text{FFindustrial}}$)
- Average Unexplained ($a_{0}$, the grand mean)

ACToR UseDB gave us chemical pathway predictions (Yes/No) and we estimated the average exposure for each pathway.

The 3rd Gen. SEEM framework incorporates the previous models.
Total Chemical Intake Rate (mg/ kg BW/ day)

Pathway
- Consumer
- Dietary
- Far-Field Pesticides
- Far-Field Industrial
- Unknown

Chemical-Specific Pathway Relevancy ($\delta_{ij}$)

Predictors
- Average Unexplained ($a_{\text{consumer}}$)
- SHEDS-HT
- FINE
- RAIDAR-ICE
- USEtox
- Production Volume
- Average Unexplained ($a_{\text{dietary}}$)
- SHEDS-HT Dietary
- Production Volume
- USEtox
- RAIDAR
- Food Contact Substance Migration
- Average Unexplained ($a_{\text{FFpesticide}}$)
- Pesticide REDs
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained ($a_{\text{FFindustrial}}$)
- USEtox
- RAIDAR
- Stockholm Convention
- Production Volume
- Average Unexplained ($a_0$, the grand mean)
Predicting Exposure Pathways

We use the method of Random Forests to relate chemical structure and properties to exposure pathway

<table>
<thead>
<tr>
<th>Sources of Positives</th>
<th>Sources of Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHANES Chemicals</td>
<td></td>
</tr>
<tr>
<td>Positives</td>
<td>Positives</td>
</tr>
<tr>
<td>Negatives</td>
<td>Negatives</td>
</tr>
<tr>
<td>OOB Error Rate</td>
<td>OOB Error Rate</td>
</tr>
<tr>
<td>Positives Error Rate</td>
<td>Positives Error Rate</td>
</tr>
<tr>
<td>Balanced Accuracy</td>
<td>Balanced Accuracy</td>
</tr>
<tr>
<td>Dietary</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2523</td>
</tr>
<tr>
<td>Near-Field</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>1622</td>
</tr>
<tr>
<td>Far-Field Pesticide</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>1480</td>
</tr>
<tr>
<td>Far Field Industrial</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>5089</td>
</tr>
</tbody>
</table>

Ring et al., 2019
## Most Important Predictors
(Scaled Gini Impurity, Louppe et al., 2013)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Dietary</th>
<th>Near-Field</th>
<th>Far-Field Pesticide</th>
<th>Far Field Industrial</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCCT_LogKAW</td>
<td>1.00</td>
<td>0.88</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>NCCT VP</td>
<td>0.84</td>
<td>1.00</td>
<td>0.99</td>
<td>0.83</td>
</tr>
<tr>
<td>NCCT MP</td>
<td>0.94</td>
<td>0.95</td>
<td>0.89</td>
<td>0.81</td>
</tr>
<tr>
<td>NCCT_LogKOA</td>
<td>0.85</td>
<td>0.89</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>Structure_MolWt</td>
<td>0.86</td>
<td>0.89</td>
<td>0.91</td>
<td>0.69</td>
</tr>
<tr>
<td>NCCT_BP</td>
<td>0.79</td>
<td>0.79</td>
<td>0.92</td>
<td>0.74</td>
</tr>
<tr>
<td>NCCT_HL</td>
<td>0.72</td>
<td>0.69</td>
<td>0.87</td>
<td>0.58</td>
</tr>
<tr>
<td>NCCT_BIODEG</td>
<td>0.74</td>
<td>0.53</td>
<td>0.85</td>
<td>0.65</td>
</tr>
<tr>
<td>NCCT_KOC</td>
<td>0.72</td>
<td>0.60</td>
<td>0.88</td>
<td>0.48</td>
</tr>
<tr>
<td>NCCT_LogP</td>
<td>0.73</td>
<td>0.58</td>
<td>0.80</td>
<td>0.50</td>
</tr>
<tr>
<td>NCCT_Csatw</td>
<td>0.72</td>
<td>0.56</td>
<td>0.79</td>
<td>0.52</td>
</tr>
<tr>
<td>NCCT_AOH</td>
<td>0.69</td>
<td>0.54</td>
<td>0.82</td>
<td>0.51</td>
</tr>
<tr>
<td>NCCT_WS</td>
<td>0.69</td>
<td>0.54</td>
<td>0.80</td>
<td>0.53</td>
</tr>
<tr>
<td>NCCT_BCF</td>
<td>0.69</td>
<td>0.56</td>
<td>0.79</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Ring et al., 2019
Pathway-Based Consensus Modeling of NHANES

- Machine learning models were built for each of four exposure pathways.

- Pathway predictions can be used for large chemical libraries.

- Use prediction (and accuracy of prediction) as a prior for Bayesian analysis.

- Each chemical may have exposure by multiple pathways.

Consensus Model Predictions

Intake Rate (mg/kg BW/day) Inferred from NHANES Serum and Urine

$R^2 = 0.816$

RMSE = 0.929

Ring et al., 2019
### Estimated Model Parameters

- Median parameter estimates from multivariate regression
- Standard deviation is reported in parentheses
- Statistically association indicated in bold

<table>
<thead>
<tr>
<th>Pathway Mean (Unexplained)</th>
<th>Dietary</th>
<th>Residential</th>
<th>Far-Field Pesticide</th>
<th>Far-Field Industrial</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.291 (0.319)</td>
<td>0.483 (0.292)</td>
<td>0.888 (0.26)</td>
<td>0.346 (0.302)</td>
<td>-0.104 (0.228)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NHANES Chemicals</th>
<th>All Chemicals</th>
<th>SHEDS Direct</th>
<th>SHEDS Indirect</th>
<th>FINE</th>
<th>Food Contact</th>
<th>REDS</th>
<th>RAIDAR</th>
<th>RAIDAR.ICE</th>
<th>USETox Pest</th>
<th>USETox Indust</th>
<th>USETox Res</th>
<th>USETox Diet</th>
<th>Production.Volume</th>
<th>Stockholm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>86.9%</td>
<td>0.187 (0.0635)</td>
<td>0.0405 (0.0688)</td>
<td>0.0159 (0.0496)</td>
<td>0.378 (0.134)</td>
<td>0.0287 (0.144)</td>
<td>-0.119 (0.0959)</td>
<td>-0.296 (0.142)</td>
<td>0.129 (0.0631)</td>
<td>-0.29 (0.135)</td>
<td>-0.0167 (0.117)</td>
<td>-0.599 (0.169)</td>
<td>-1.48 (0.256)</td>
<td>-1.94 (0.462)</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>45</td>
<td>88</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.22%</td>
<td>4.68%</td>
<td>1.58%</td>
<td>9.89%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Ring et al., 2019*
We extrapolate to predict relevant pathway(s), median intake rate, and credible interval for each of 479,926 chemicals.

Of 687,359 chemicals evaluated, 30% have less than a 50% probability for exposure via any of the four pathways and are considered outside the “domain of applicability.”

This approach identifies 1,880 chemicals for which the median population intake rates may exceed 0.1 mg/kg bodyweight/day.
We extrapolate to predict relevant pathway(s), median intake rate, and credible interval for each of 479,926 chemicals.

Of 687,359 chemicals evaluated, 30% have less than a 50% probability for exposure via any of the four pathways and are considered outside the “domain of applicability.”

This approach identifies 1,880 chemicals for which the median population intake rates may exceed 0.1 mg/kg bodyweight/day.

There is 95% confidence that the median intake rate is below 1 µg/kg BW/day for 474,572 compounds.
Haven’t Had Enough?
SOT 2019 Sunrise Mini-Course SR02
“Publicly Available Exposure Tools to Inform the Toxic Substances Control Act”
7 AM Sunday morning
you can register in person at the SOT meeting
ToxCast + ExpoCast Provide NAMs for Chemical Prioritization

ToxCast + HTTK can estimate doses needed to cause bioactivity

Exposure intake rates can be inferred from biomarkers (Wambaugh et al., 2014)

Estimated Equivalent Dose or Predicted Exposure (mg/kg BW/day)

Chemicals Monitored by CDC NHANES

Ring et al. (2017)
SEEM2: Life-stage and Demographic Specific Predictions

- We can calculate margin between bioactivity and exposure for specific populations

- Based on variation in toxicokinetics and exposure

To date SEEM3 predictions are only available for overall (“Total”) U.S. Population

Ring et al. (2017)
Conclusions

- We can make chemical-specific estimates of intake rate for hundreds of thousands of chemical
  - Only predicting median intake rate (and even that has large uncertainty)
  - Synthesizing as many models and other data as we can find

- Models incorporate Knowledge, Assumptions and Data (Macleod, et al., 2010)
  - The trick is to know which model to use and when
  - Machine learning models allow educated guesses

- We are using existing chemical data to predict pathways
  - Not all chemicals fit within the domain of applicability
  - Need better training data for random (non-targeted analysis of environmental media needed)

- Eventually we have got to go beyond NHANES
  - Current evaluation based upon 114 chemicals
  - Non-targeted analysis of blood may eventually be possible

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA
"Scientists should resist the demand to describe any model, no matter how good, as validated. Rather than talking about strategies for validation, we should be talking about means of evaluation."

Naomi Oreskes
ExpoCast Project (Exposure Forecasting)

**NCCT**
Chris Grulke
Greg Honda*
Richard Judson
Ann Richard
Risa Sayre*
Mark Sfeir*
Rusty Thomas
John Wambaugh
Antony Williams

**NRMRL**
Xiaoyu Liu

**NERL**
Cody Addington*
Namdi Brandon*
Alex Chao*
**Kathie Dionisio**
Peter Egeghy
Hongtai Huang*
**Kristin Isaacs**
Ashley Jackson*
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Seth Newton
Katherine Phillips
Paul Price
Jeanette Reyes*
Randolph Singh*
Marci Smeltz
Jon Sobus
John Streicher*
Mark Strynar
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
Barbara Wetmore

*Trainees

**NHEERL**
Linda Adams
Christopher Ecklund
Marina Evans
Mike Hughes
Jane Ellen Simmons
Tamara Tal

**NRMRL**
Xiaoyu Liu

**NERL**
Cody Addington*
Namdi Brandon*
Alex Chao*
**Kathie Dionisio**
Peter Egeghy
Hongtai Huang*
**Kristin Isaacs**
Ashley Jackson*
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Seth Newton
Katherine Phillips
Paul Price
Jeanette Reyes*
Randolph Singh*
Marci Smeltz
Jon Sobus
John Streicher*
Mark Strynar
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
Barbara Wetmore

*Trainees

**Collaborators**

Arnot Research and Consulting
Jon Arnot
Johnny Westgate
Institut National de l'Environnement et des Risques (INERIS)
Frederic Bois
Integrated Laboratory Systems
Kamel Mansouri
National Toxicology Program
Mike Devito
Steve Ferguson
Nisha Sipes
Ramboll
Harvey Clewell
ScitoVation
Chantal Nicolas
Silent Spring Institute
Robin Dodson
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Ayliward
Technical University of Denmark
Peter Fantke
Tox Strategies
Caroline Ring
Miyoung Yoon
Unilever
Beate Nicol
Cecilia Rendal
Ian Sorrell
United States Air Force
Heather Pangburn
Matt Linakis
University of California, Davis
Deborah Bennett
University of Michigan
Olivier Jolliet
University of Texas, Arlington
Hyeong-Moo Shin
References

“Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public.” Burwell et al. (2013):

“...this includes using machine-readable and open formats...” Burwell et al. (2013):

Machine learning models based on chemical structure and physico-chemical properties predict whether or not each pathway is relevant to a library of over 680,000 chemicals,

- Each individual model prediction will also be made available