United States
Environmental Protection
Agency

Laying the Groundwork for High-Throughput Occupational Exposure Modeling: Data and Models

Jeff Minucci, Tom Purucker, Kristin Isaacs, John Wambaugh, and
Katherine Phillips
U.S. EPA Office of Research and Development
February 25, 2021

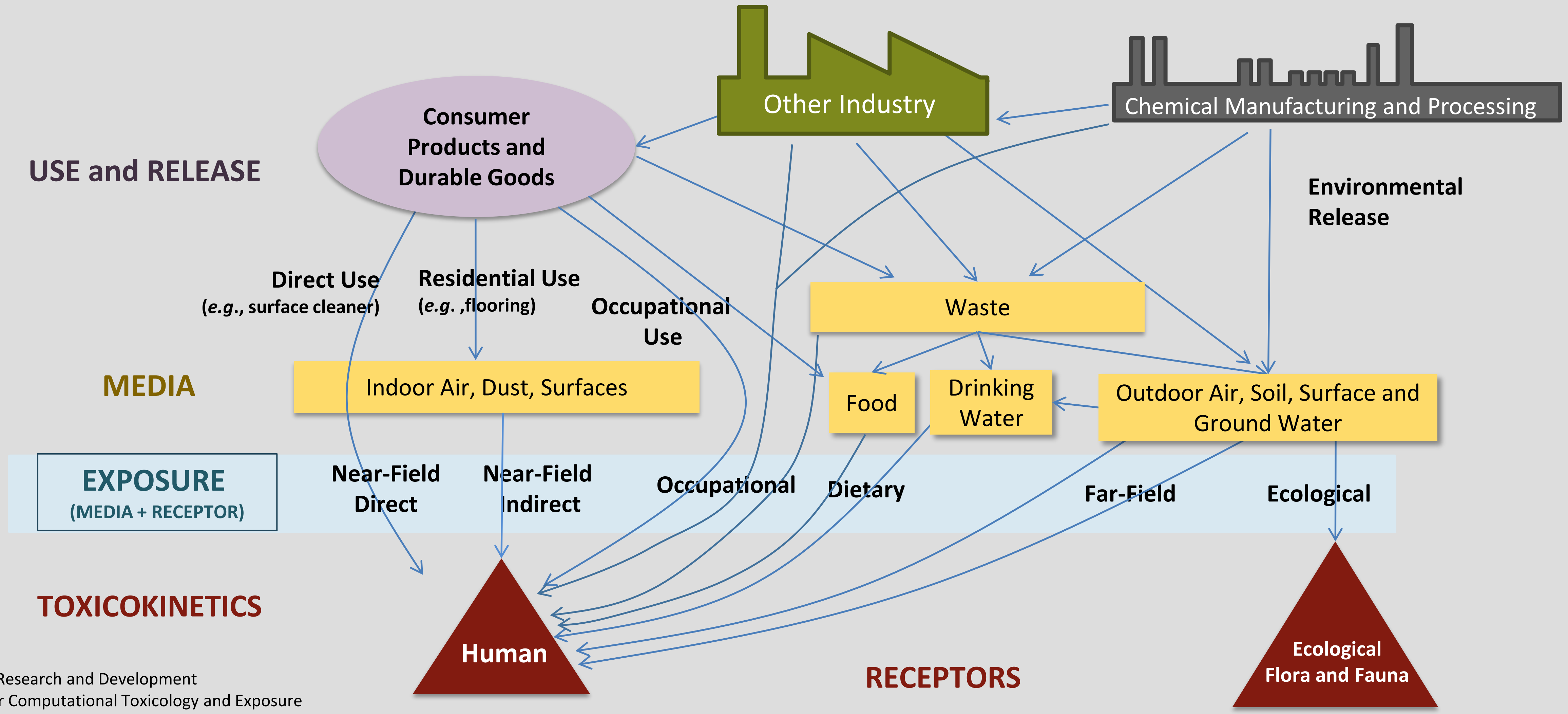
United States Environmental Protection Agency

Disclaimer: The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency

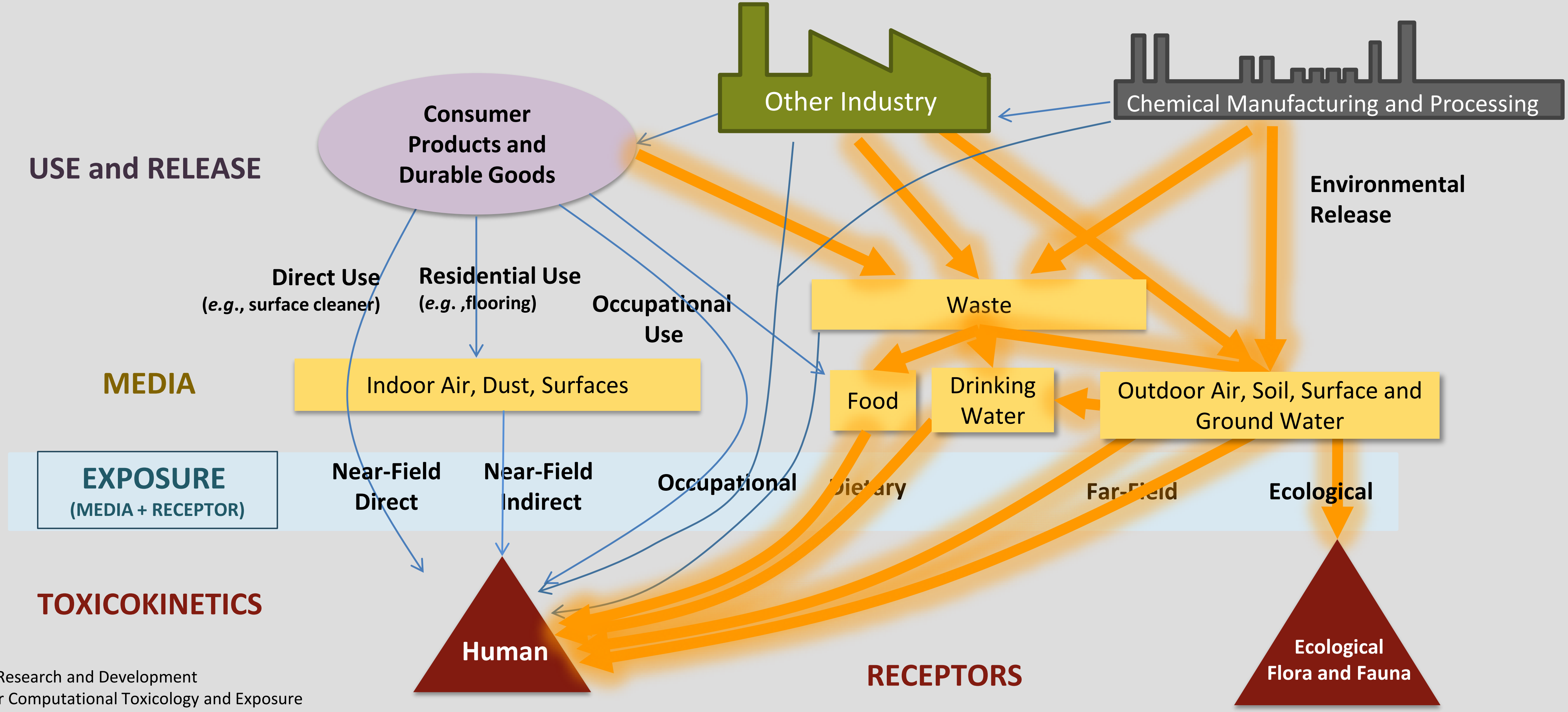
Outline

- Background
- Adapting Existing Models for High-throughput Use
- Organizing Occupational Exposure Data
- Predicting Chemical Concentrations in Workplaces
- Next Steps

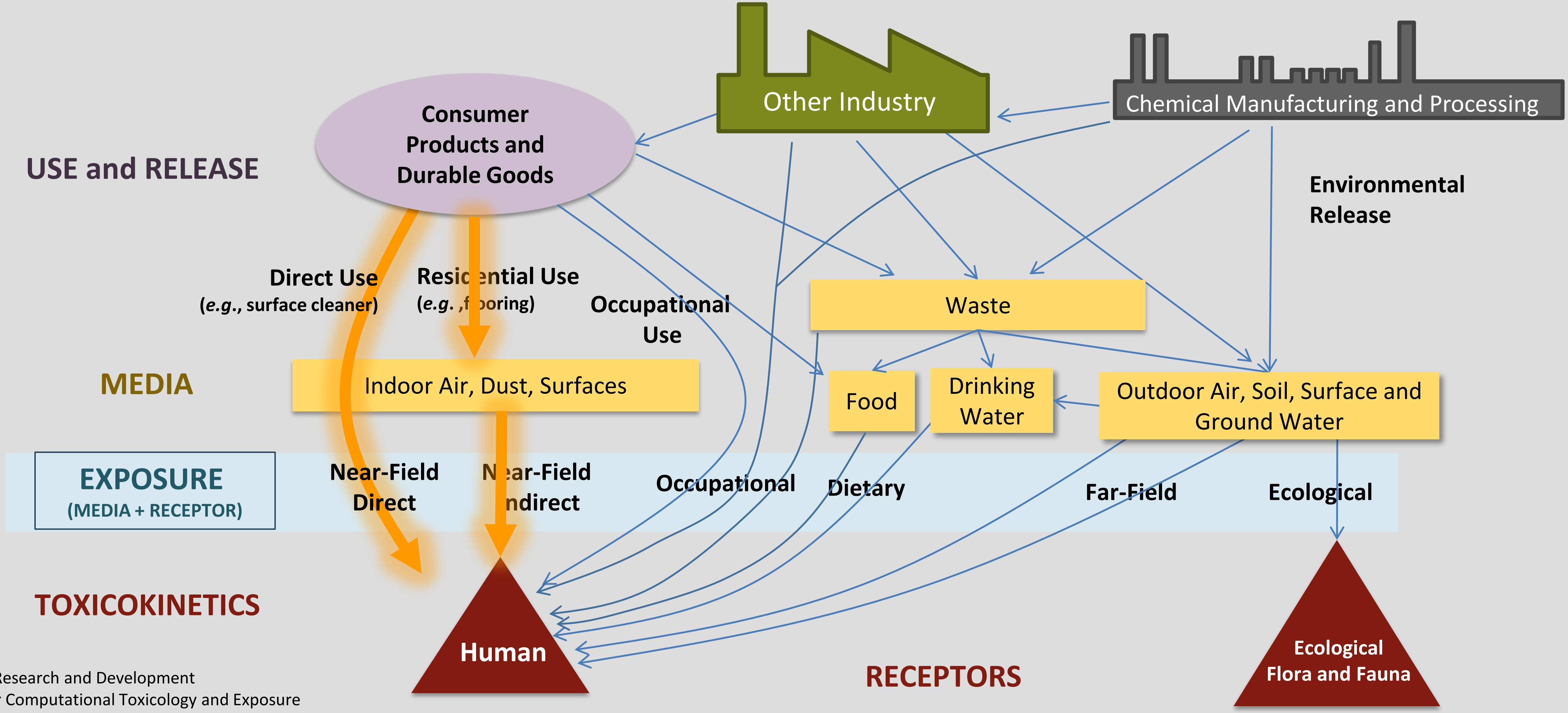
Exposure Pathways



Ambient Pathways

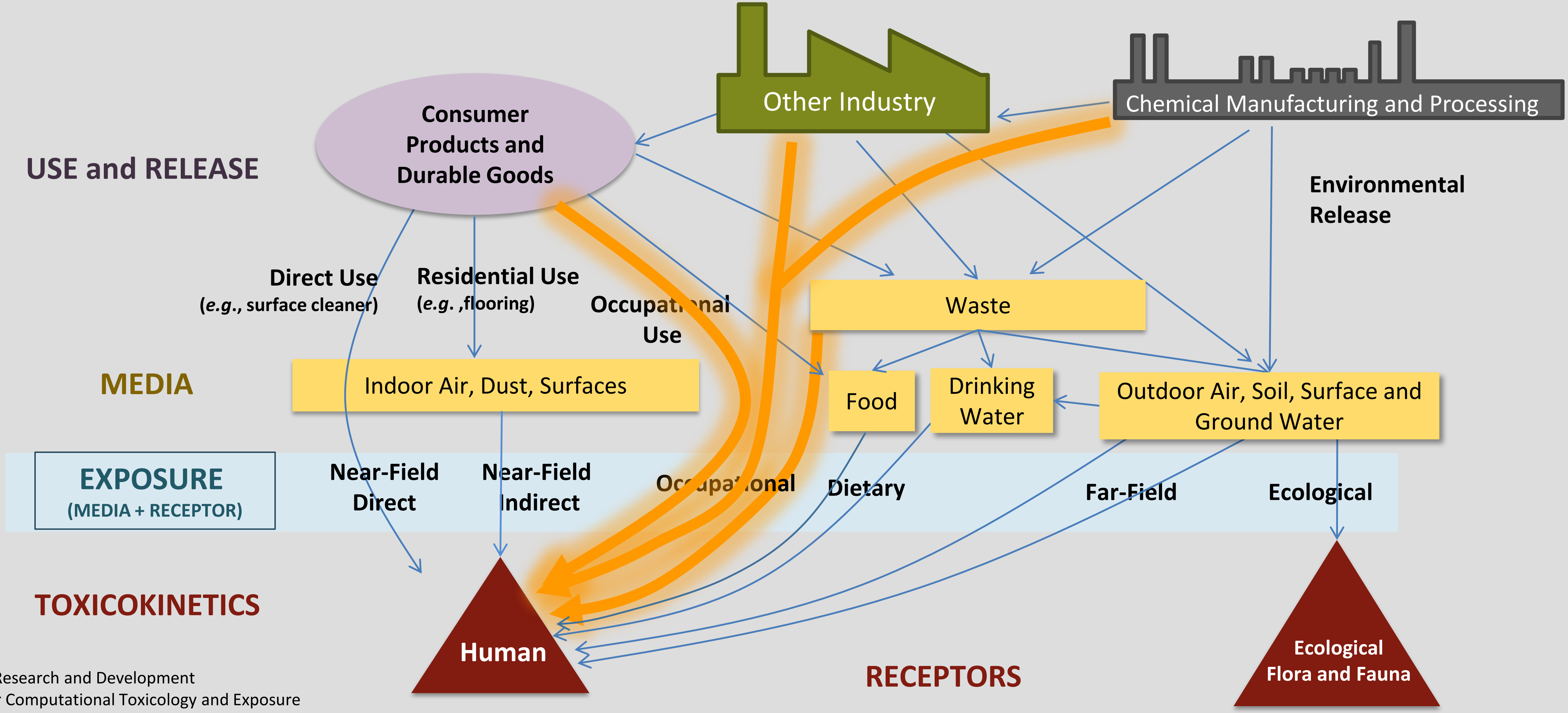


Consumer Pathways



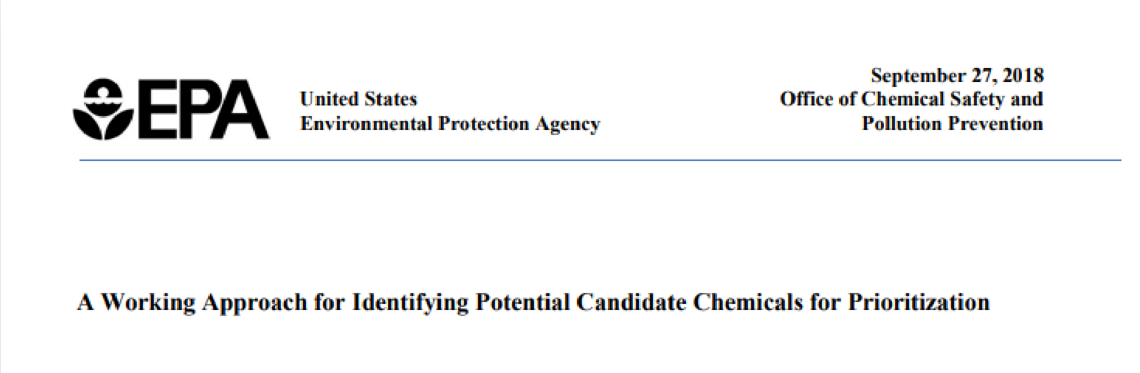
United States Environmental Protection Agency

Occupational Pathways



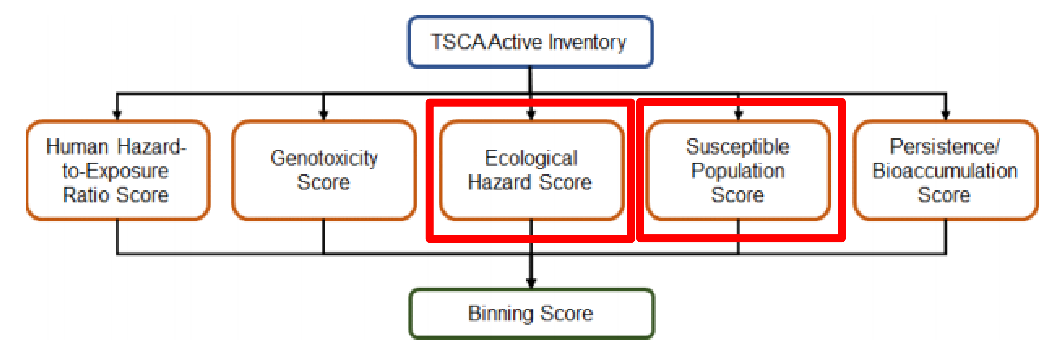
HT Pathway Predictions are Needed for Decision Workflows

- Human exposure pathway predictions for thousands of chemicals are currently integrated into working approaches for identifying potential candidates for prioritization under TSCA
  - Consumer, dietary, and ambient predictions currently integrated into consensus predictions
- Efforts are underway to incorporate high-throughput ecological exposure predictions for integration with ecological hazard data
- It is also ultimately desirable to predict occupational exposures in a high-throughput manner for use in such workflows
  - TSCA directs EPA to address potentially exposed or susceptible sub-populations, defined as *a group of individuals within the general population identified by the Administrator who, due to either greater susceptibility or greater exposure, may be at greater risk than the general population of adverse health effects from exposure to a chemical substance or mixture, such as infants, children, pregnant women, **workers**, or the elderly.*


United States
Environmental Protection Agency

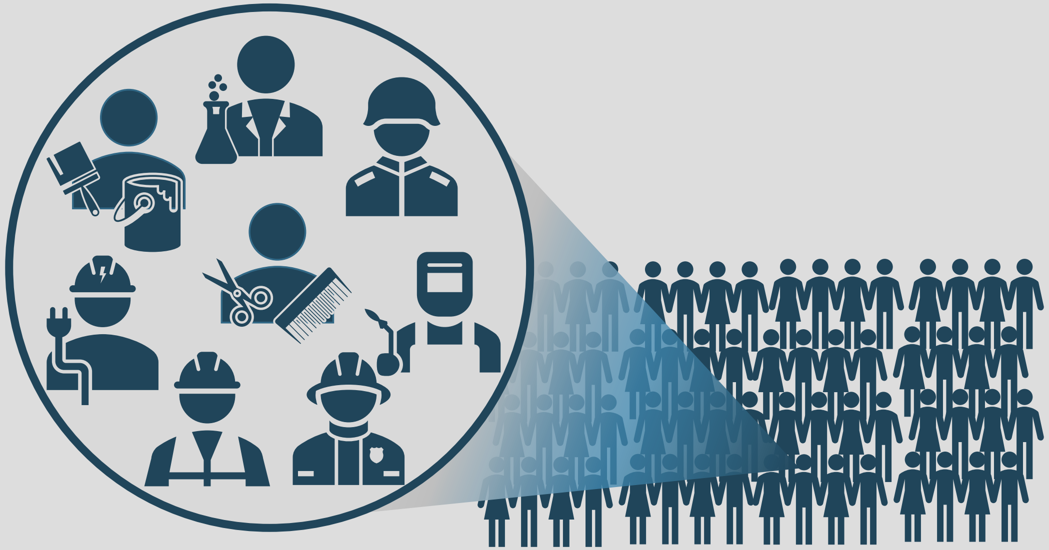
September 27, 2018
Office of Chemical Safety and
Pollution Prevention

A Working Approach for Identifying Potential Candidate Chemicals for Prioritization



Occupational Exposure

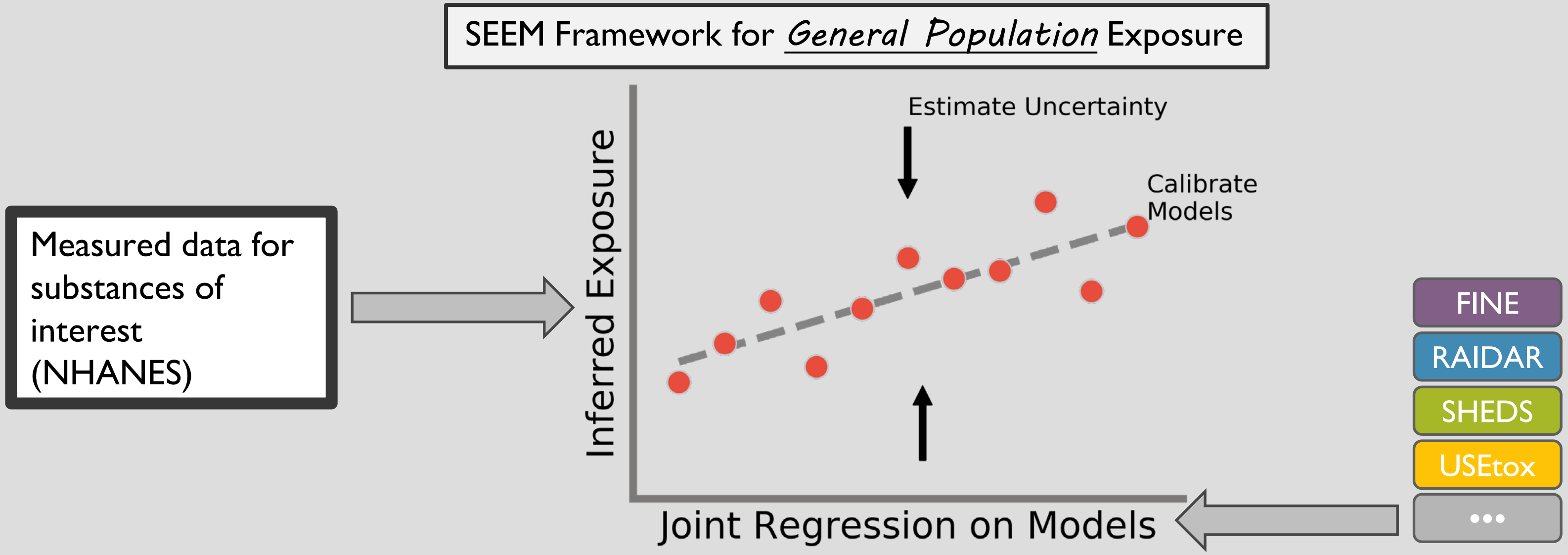
- EPA has made great strides in evaluating exposure models for the general population which can assess thousands of chemicals quickly
- Occupational exposure requires considering different exposure scenarios and different chemicals workers are exposed to across many different occupations
- Potential exposures often classified by different occupational classes



EPA United States Environmental Protection Agency

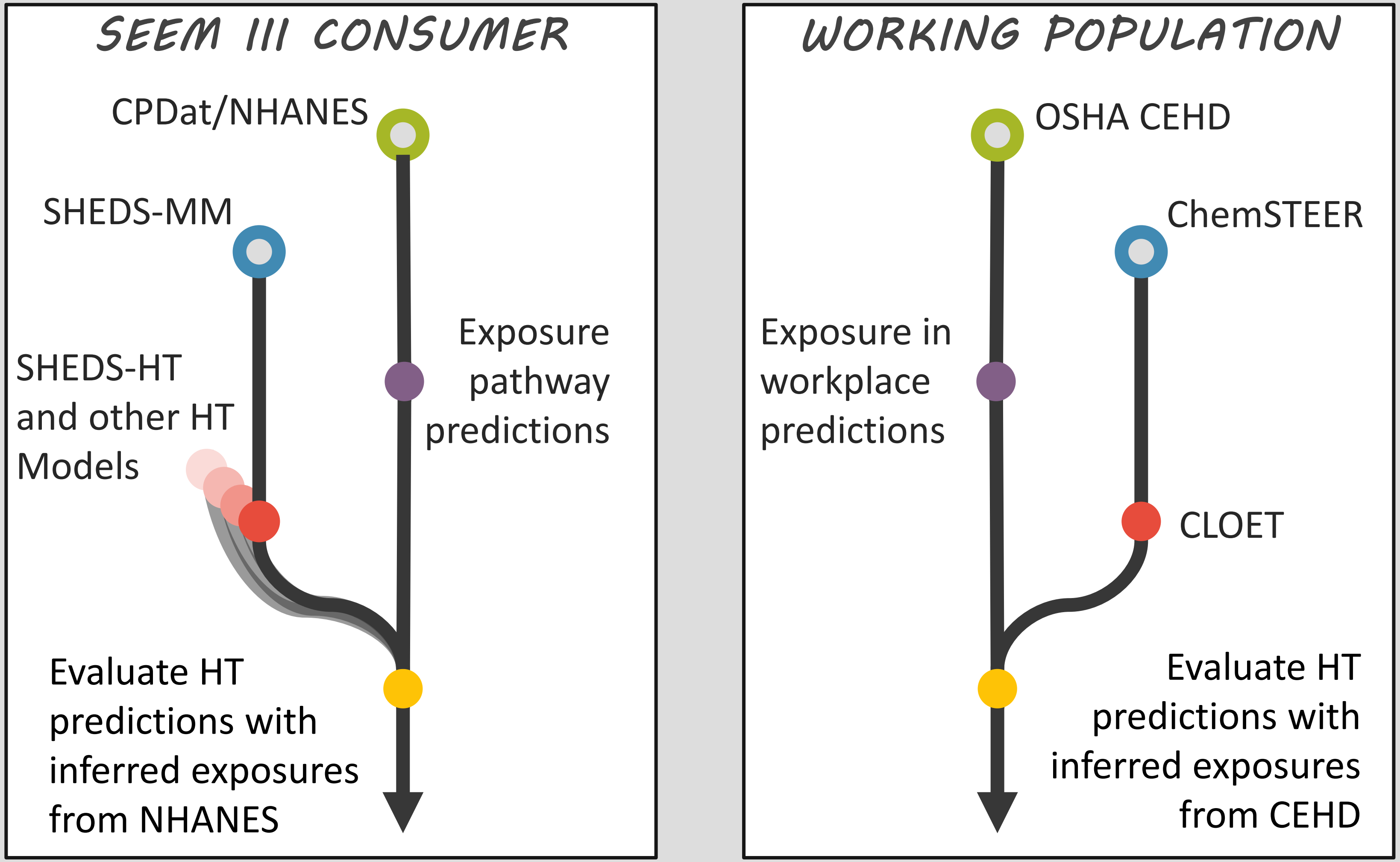
Consensus Exposure Modeling

- Multiple high-throughput exposure models can be run for a given population assuming general consumer, dietary, and far-field exposures
- Results are coupled through the Systematic Empirical Evaluation of Models (SEEM) framework



Wambaugh, et al, *J. Environ. Sci. & Tech.* 2013, 2014
Ring et al, *J. Environ Sci. & Tech.* 2019

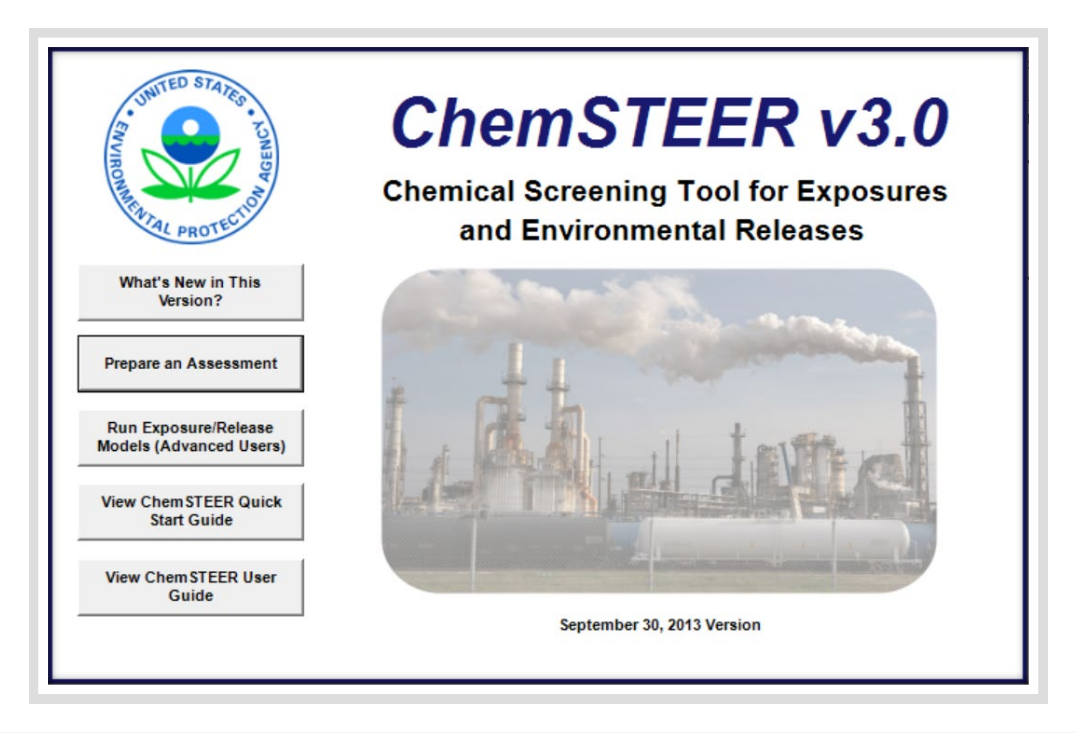
Path to Occupational Exposure



United States Environmental Protection Agency

ChemSTEER

- Developed by EPA to estimate workplace exposures and environmental releases
- Requires manual input of information
- 6 dermal exposure models
  - 1-hand dermal contact with liquid
  - 2-hand dermal contact with liquid
  - 2-hand dermal immersion in liquid
  - Direct 2-hand dermal contact with solids
  - 2-hand dermal contact with container surfaces
  - User defined
- 11 inhalation exposure models
  - Small volumes handling
  - PEL-limiting for substance specific particulates
  - Total PNOR PEL-limiting
  - Respirable PNOR PEL-limiting
  - Automobile OEM Spray Coating
  - Automobile Refinish Spray Coating
  - Automobile Spray Coating
  - UV Roll Coating
  - User defined
  - Mass balance
  - PEL-limiting for substance specific vapors



Phillips et al, *in preparation*

United States Environmental Protection Agency

CLOET

High-throughput Command Line Occupational Exposure Tool (CLOET) a command line tool that allows calculation of ChemSTEERv3.0 exposure models.

Written in base Python 3.0, so it depends on no external packages.

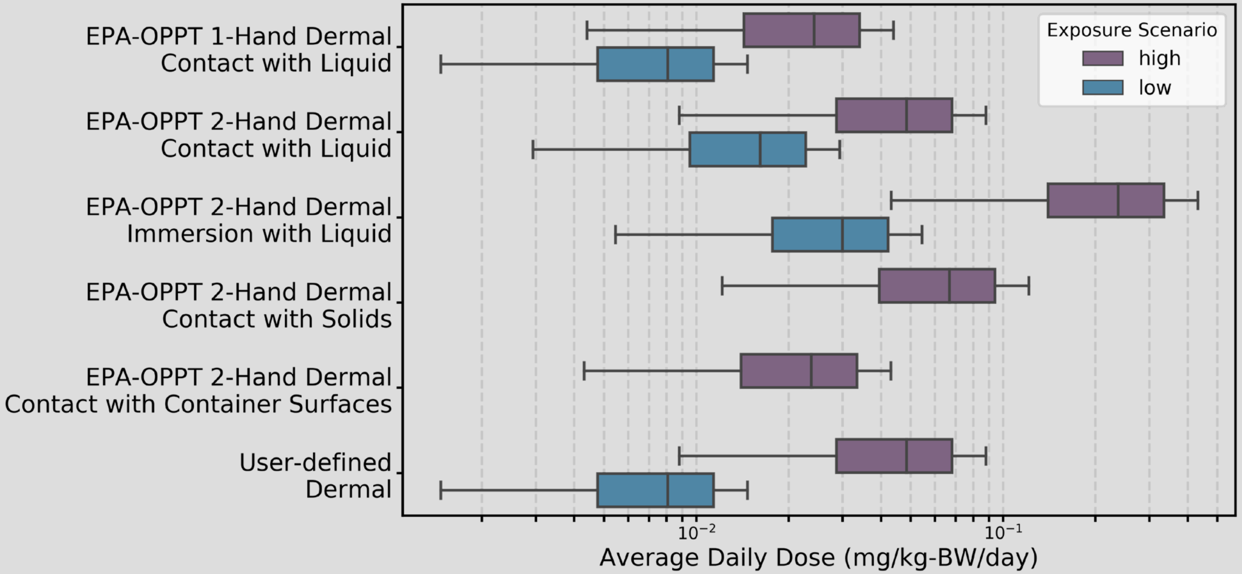
Multiple scenarios for each model have been run and tested against ChemSTEER GUI to test for model fidelity.

```
>>> import cloet
>>> e_derm = (cloet.dermal
              .one_hand_liquid_contact(Yderm=0.5))
>>> e_derm.inputs
{'ED': 1,
 'NWexp': 1,
 'NS': 1,
 'EY': 40,
 'BW': 70,
 'ATc': 70,
 'AT': 40,
 'S': 535,
 'Qu': 2.1,
 'FT': 1,
 'Yderm': 0.5}
>>> e_derm.outputs
{'NW': 1,
 'LADD': 0.012563600782778865,
 'ADD': 0.021986301369863015,
 'APDR': 8.025,
 'Dexp': 561.75}
```

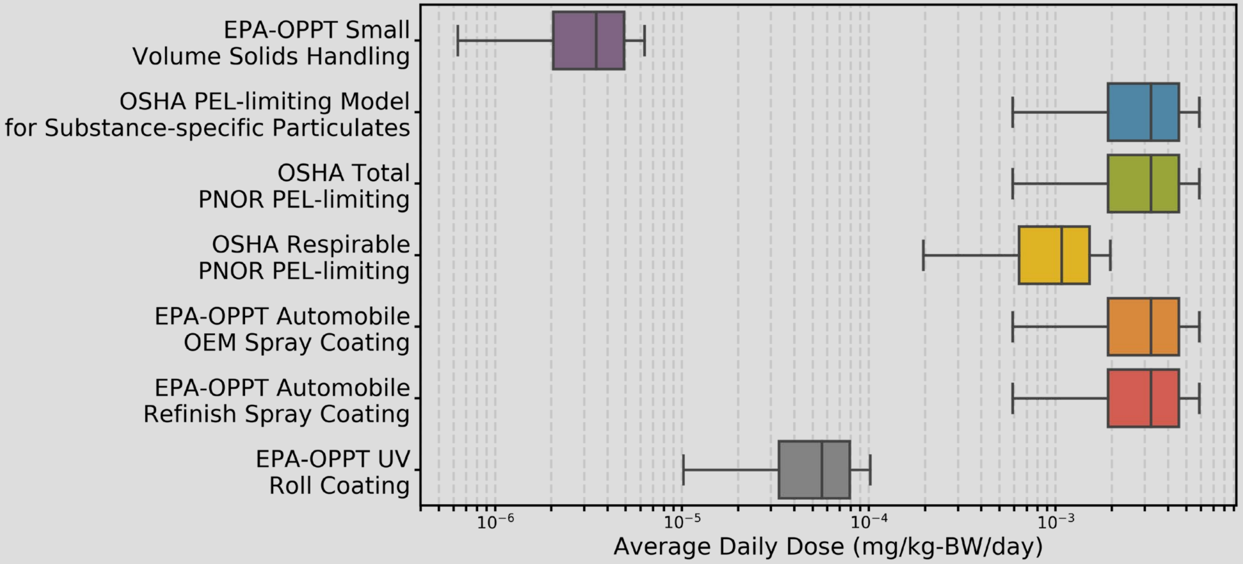
Phillips et al, *in preparation*

Chemical Agnostic ChemSTEER Models

DERMAL MODELS



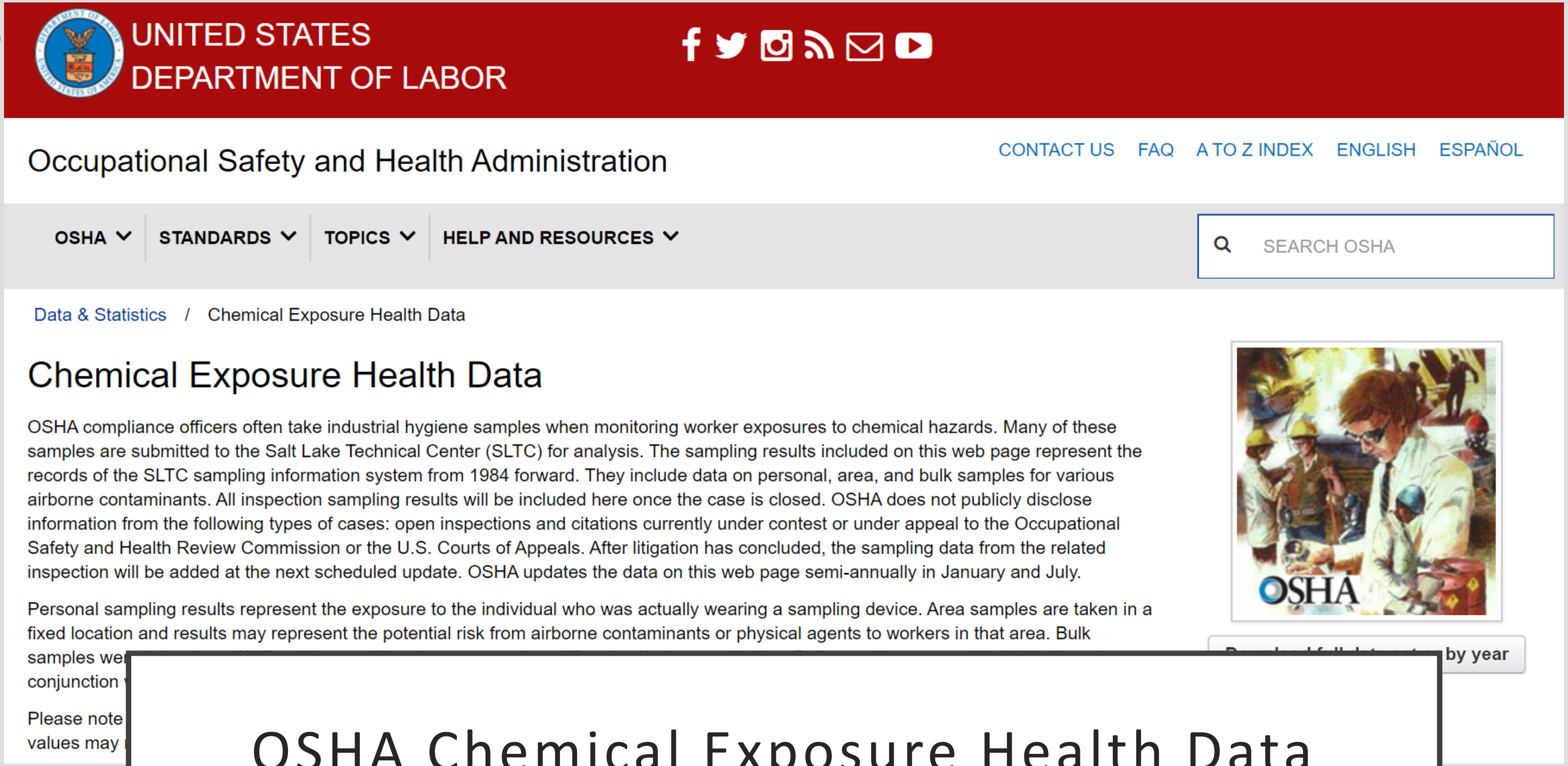
INHALATION MODELS



Concentrations were varied from 0.1 to 1 for all chemical agnostic models

Phillips et al, *in preparation*

OSHA Chemical Exposure Health Data



- Data available from 1989 – 2018
- One record is one sample taken from one work site for one chemical
- Inhalation Samples: 1.3 million samples; ~1000 substances; both area and personal sampling
- Dermal Samples: ~200,000 samples, ~70 substances, dermal wipes

Minucci et al, *in preparation*

EPA United States Environmental Protection Agency

Organizing OSHA Data

- Standardized and unified all years into one dataset
- Cleaned missing or unlabeled data
- Converted all air samples to *mg/m*^{3}
- Matched OSHA substance names to substances in DSSTox database by synonym searching on the EPA CompTox Dashboard.
- Substances that returned no match were searched manually in PubChem for synonyms
- Converted old SIC/NAICS codes to 2017 NAICS codes

Organizing OSHA Data

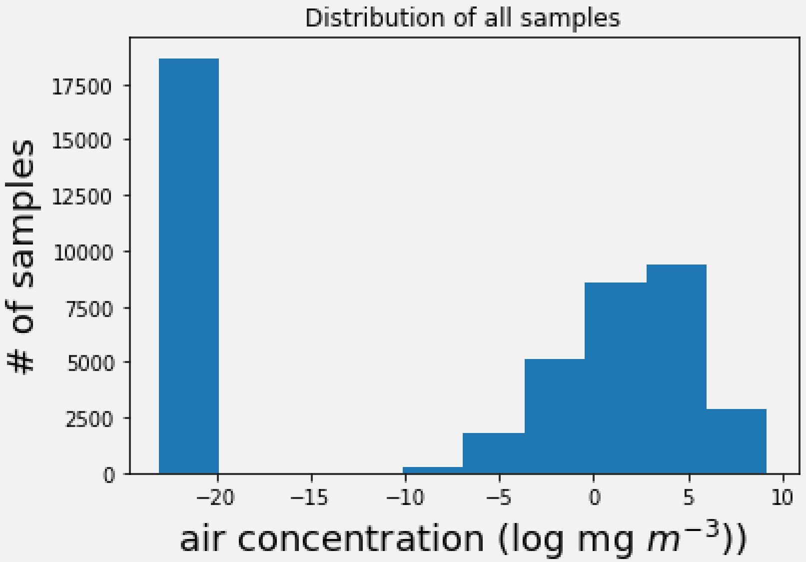
- Standardized and unified all years into one dataset
- Cleaned missing or unlabeled data
- Converted all air samples to *mg/m³*
- Matched OSHA substance names to substances in DSSTox database by synonym searching on the EPA CompTox Dashboard.
- Substances that returned no match were searched manually in PubChem for synonyms
- Converted old SIC/NAICS codes to 2017 NAICS codes

United States Environmental Protection Agency

Caveats to OSHA Data

- OSHA data is ***<u>not</u>*** a random sampling of every workplace. These sampling efforts typically only occur when someone is suspicious of a violation in a workplace. For this reason, the measurements can tend to be higher than average both in concentration and in frequency of detection.
- OSHA collected multiple measured values in a single sampling effort; to capture a “worst case” scenario, these records were aggregated to the maximum measured value (for example, measurements of 0, 0, and 12 mg/m^3 in a single inspection are aggregated to 12 mg/m^3 for that inspection)
- Concentrations for naturally occurring compounds in air (CO_2, O_2, N_2, etc.) are included in measurements even though they typically do, and should, exist in environments at high concentrations.

EPA United States Environmental Protection Agency



Challenge:

Many samples (40%) were 0 or below detection limit. Many modeling techniques are unreliable in this case

Solution:

A two-stage statistical model

OSHA Chemical Exposure Health Data

- Data available from 1989 – 2018
- One record is one sample taken from one work site for one chemical
- Inhalation Samples: 1.3 million samples; ~1000 substances; both area and personal sampling
- Dermal Samples: ~200,000 samples, ~70 substances, dermal wipes

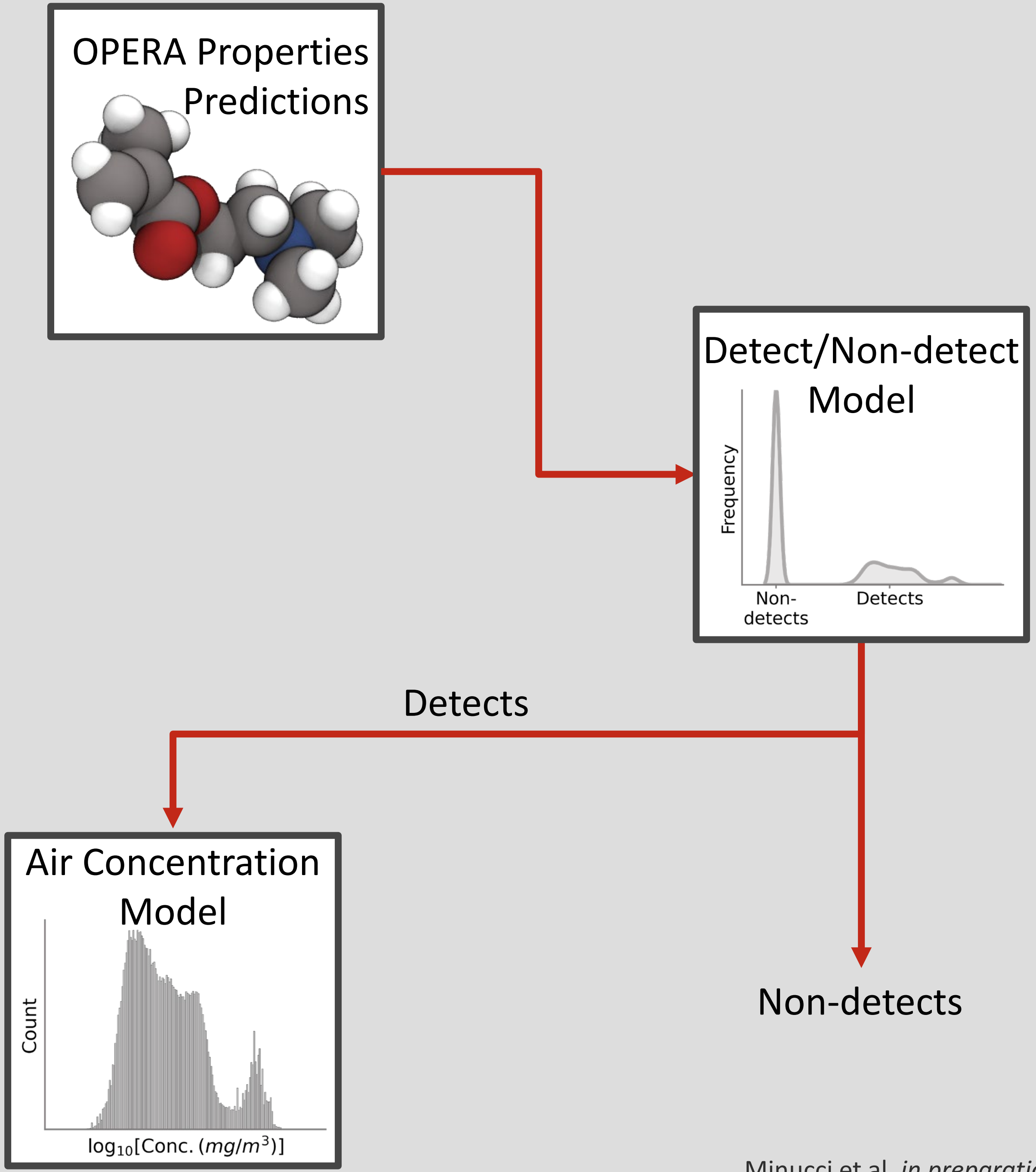
Minucci et al, *in preparation*

United States Environmental Protection Agency

Two-Stage Model

Using Bayesian Hierarchical Regression, we can construct a model where, knowing nothing about a chemical other than its structure, we can predict:

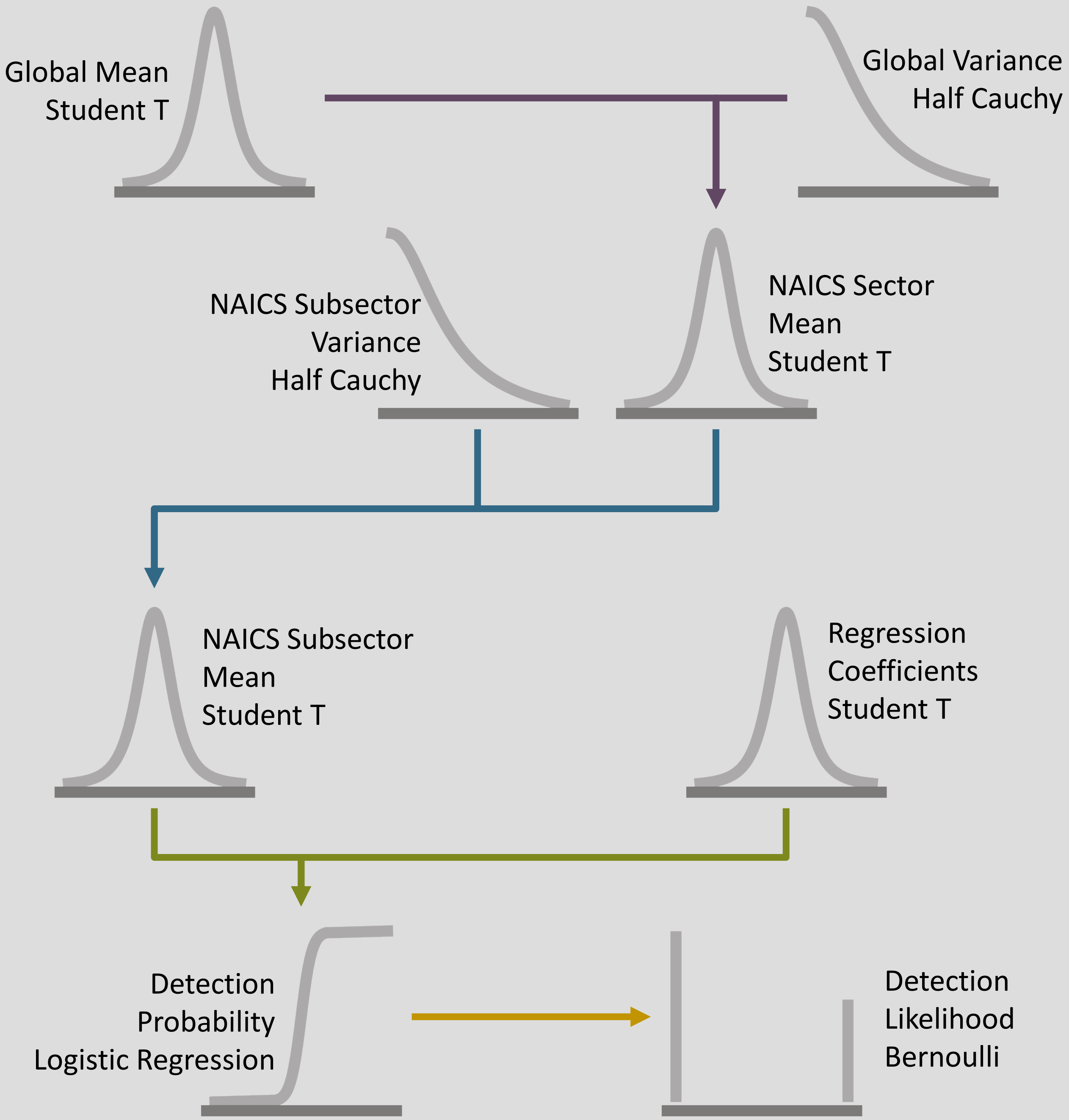
- the likelihood of a chemical being detected by OSHA's air measurement methods
- they likely concentration of chemical detected in an air sample



Minucci et al, *in preparation*

Bayesian Hierarchical Models

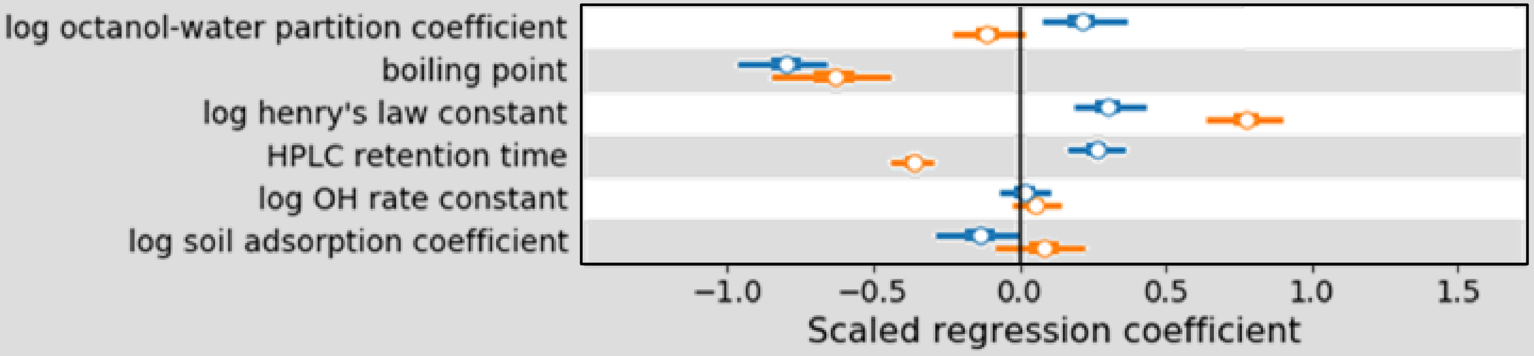
- Bayesian Hierarchical Regression allows us to organize our predictions (either detect/non-detect or concentration) by NAICS Sector and/or Subsector
- When data is lacking, at the Subsector level, we can aggregate up to the Sector prediction
- OPERA physicochemical property distributions across NAICS sector and subsectors are included as input distributions to the models in addition to OSHA data



Minucci et al, *in preparation*

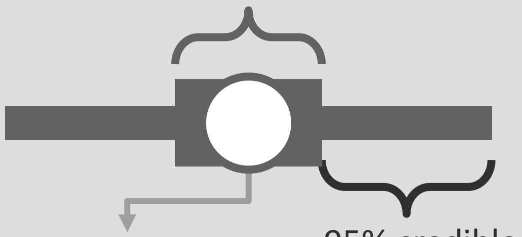
Physicochemical Effects

Detect/Non-Detect Model   Air Concentration Model



- correlation with probability of detection

+ correlation with probability of detection



Substances are more likely to be ***detected*** in the air of a workplace with

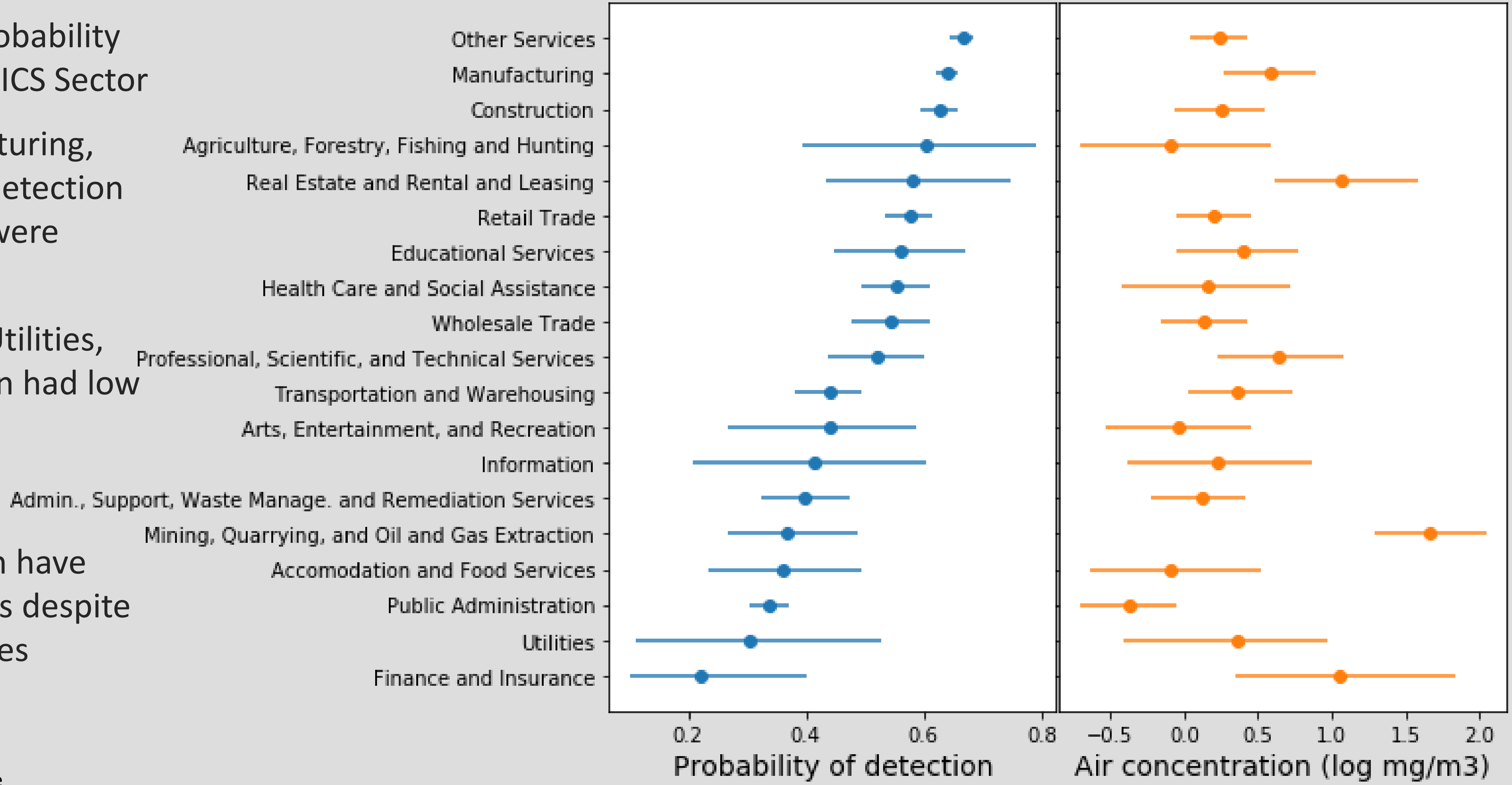
- Low
  - Boiling Point
- High
  - Octanol-water Coefficient (logP)
  - Henry's law constant (loghl)
  - HPLC retention time (rt)

<u>**Note**</u>: the properties listed are all ***predicted*** properties from the OPERA suite.

Minucci et al, *in preparation*

Detection and Concentration by NAICS Sectors

- Variation in detection probability and concentration by NAICS Sector
- Other Services, Manufacturing, Construction have high detection rates – many chemicals were sampled in these sectors
- Finance and Insurance, Utilities, and Public Administration had low detection probabilities
- Mining and Oil extraction have concentration predictions despite low detection probabilities



United States Environmental Protection Agency

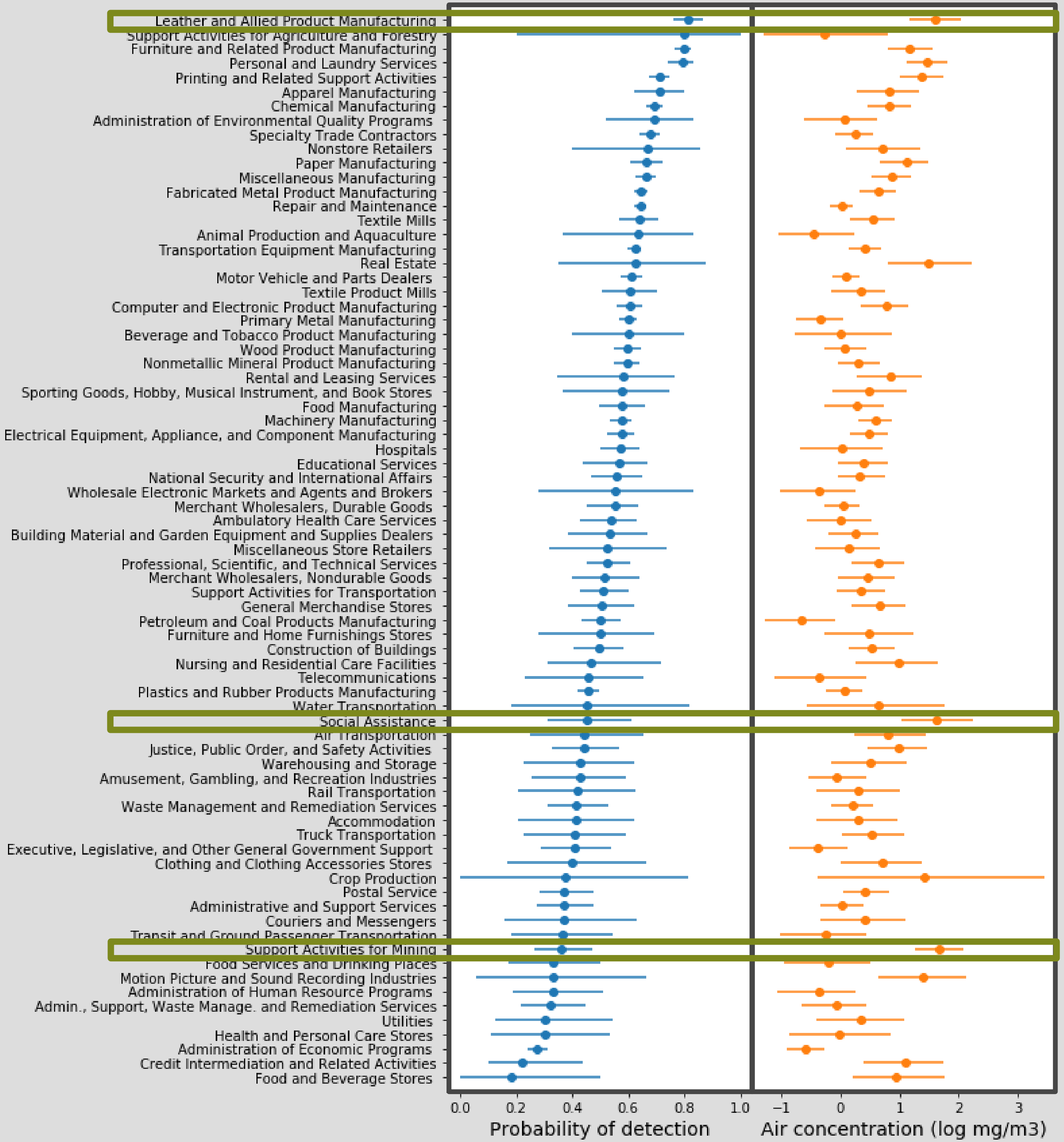
Predictions by NAICS Subsectors

- Highest likelihood of detection
  - Leather and Allied Product Manufacturing
  - Support Activities for Agriculture and Forestry
  - Furniture and Related Products Manufacturing
- Lowest likelihood of detection
  - Food and Beverage Stores
  - Credit Information and Related Activities
  - Administration of Economic Programs

EPA United States Environmental Protection Agency

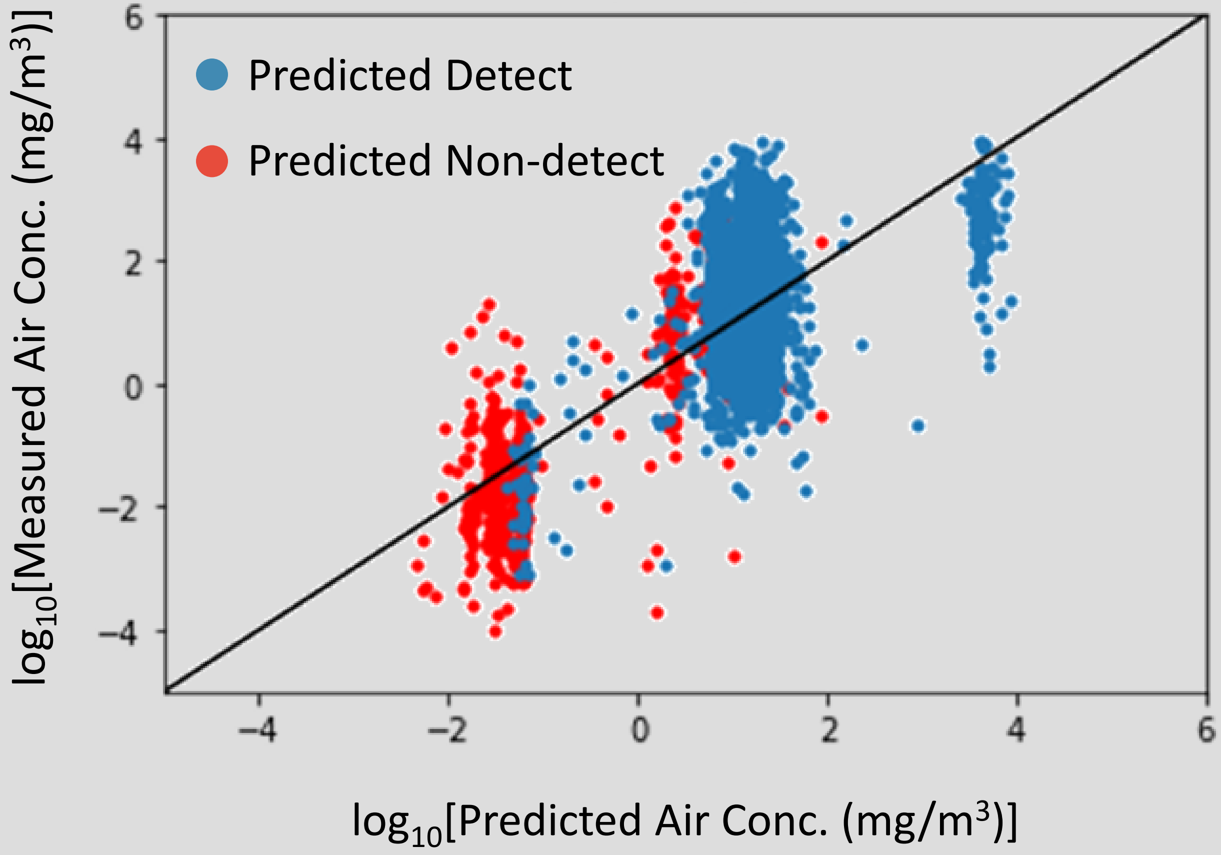
Predictions by NAICS Subsectors

- Highest likelihood of detection
  - Leather and Allied Product Manufacturing
  - Support Activities for Agriculture and Forestry
  - Furniture and Related Products Manufacturing
- Highest predicted concentration (if detected)
  - Support Activities for Mining
  - Social Assistance
  - Leather and Allied Product Manufacturing



United States Environmental Protection Agency

True Positive and False Negative Concentration Predictions of Test Set



Model Validation

CEHD Data
584 Substances
47229 Samples

Testing Data
58 Substances
5641 Samples

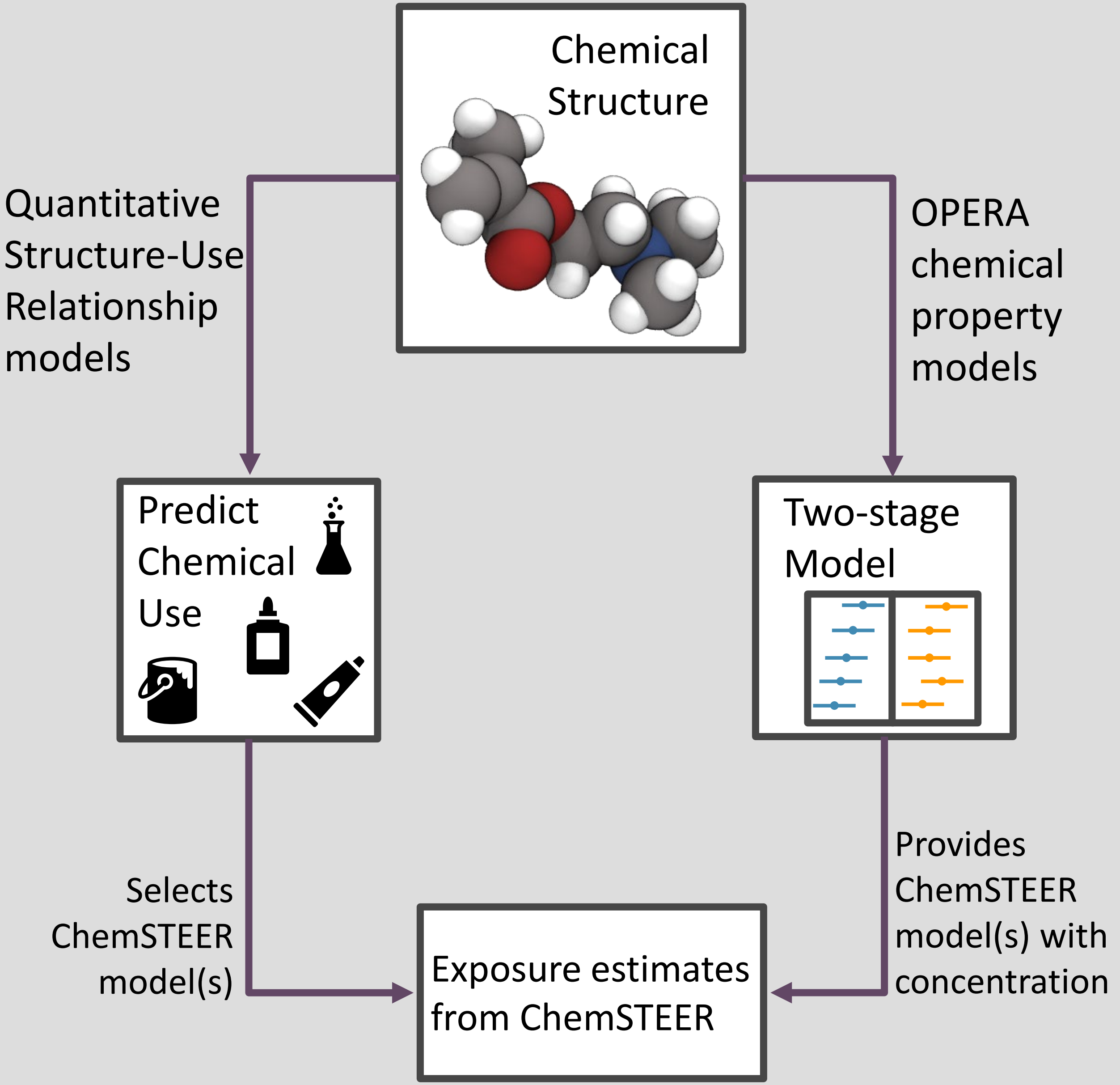
Training Data
526 Substances
41588 Samples

	Predicted Non-detect	Predicted Detect
Actual Non-detect	21.5%	17.1%
Actual Detect	7.4%	54.0%

United States
Environmental Protection
Agency

Next Steps

- Use detect/non-detect and concentration models to predict concentration of chemicals
- Use QSUR-models to predict functional use (technical function) and sector of use of chemical
- Use sector of use and concentration to choose which ChemSTEER models apply to which chemicals in which sectors and get exposure estimates



Acknowledgements

US EPA

Jeff Minucci
Tom Purucker
John Wambaugh
Tony Williams
Peter Egeghy
Cody Addington

Former Trainees

Ashley Jackson