Introduction to Workshop

08.23.2018

Welcome!

- The goal of this workshop is not to give a comprehensive overview of all possible methods to analyze exposure to mixtures
- This is a very hot topic right now in the EH community
- New methods are being developed and adapted from other fields daily

Welcome!

- The goal of this workshop is not to give a comprehensive overview of all possible methods to analyze exposure to mixtures
- This is a very hot topic right now in the EH community
- New methods are being developed and adapted from other fields daily

- Instead, our goal is to give some examples of different approaches
  - That are used to answer different research questions
- Go over some tools in  to apply these methods
- Importantly, discuss what *types* of methods are appropriate for which research questions

What is a mixture?

- Actually, there is no strict definition
- According to NIEHS "`a mixture must have at least three independent chemicals or chemical groups`"
- Generally, exposure to a mixture indicates exposure to multiple "stressors" simultaneously
  - Chemical
  - Non-chemical (SES, diet, etc)

Why care about mixtures?

- We are exposed to hundreds (thousands?) of chemicals at any single time point
- Traditionally, epi studies have focused on single-chemical analyses
- This does not represent reality
- The combination of exposures to multiple chemicals likely induces different responses
  - Compared to exposure to each chemical independently
- The question becomes:

  `How can we represent the complexity of reality in a statistical model?`

How do we deal with exposure to mixtures?

- This is still a very open question
- Existing methods have limitations
- There have been several workshops held by EPA and NIEHS to address this issue
- The most recent NIEHS workshop (2015) concluded that
  1. Although some methods performed better than others the presented estimated associations were still quite variable and not in agreement
  2. The choice of method should depend on the research question

NIEHS Workshop – Methods used

Method	Category
Single chemical analysis	Classic linear regression (ordinary least squares)
Multiple regression	Classic linear regression (ordinary least squares)
Visualization, structural equation modeling (SEM), and principal component analysis (PCA)	Classification and prediction
Informed sparse PCA and segmented regression	Classification and prediction
Bayesian g-formula	Classification and prediction
PCA	Classification and prediction
Classification and regression trees (CART)	Classification and prediction
Bayesian profile regression	Classification and prediction
Random forest	Classification and prediction
Multivariate adaptive regression splines (MARS)	Classification and prediction
Bayesian non-parametric regression	Classification and prediction
Bayesian additive regression trees (BART) and negative sparse PCA (NSPCA)	Classification and prediction
Conformal predictions	Classification and prediction
Bayesian kernel machine regression (BKMR)	Exposure–response surface estimation
Building Bayesian networks	Exposure–response surface estimation
Exposure surface smoothing (ESS)	Exposure–response surface estimation
Modes of action (results presented for $Z = 0$ strata)	Other
Feasible solution algorithm (FSA)	Other
Exploratory data analysis (EDA)	Other
Novel approach and least-angle regression (LARS)	Variable selection
Machine learning	Variable selection
Two-step variable selection and least absolute shrinkage and selection operator (LASSO)	Variable selection
Two-step shrinkage-based regression	Variable selection
Factor mixture models	Variable selection
Subset and bootstrap	Variable selection
Variable selection regression (VSR)	Variable selection
Bayesian estimation of weighted sum	Variable shrinkage strategies
Shrinkage methods (LASSO/LARS)	Variable shrinkage strategies
Weighted quantile sum regression (WQS)	Variable shrinkage strategies
LASSO	Variable shrinkage strategies

Taylor et al. EHP 2016

PRIME

- In fact, after the 2015 workshop NIEHS had an RFA for robust methods development
- Powering Research Through Innovative Methods for Mixtures in Epidemiology (PRIME)
- Grants were awarded starting January 2018
- Six grantees with very different proposed methods
- One of the requirements was that all software developed under PRIME will be publicly available
- More here:
  `https://www.niehs.nih.gov/research/supported/exposure/mixtures/prime_program/index.cfm`

What research questions could we have?

1. Identification of "bad actors" (aka toxic agents)
2. Overall mixture effect
3. Synergistic/ antagonistic effects (interactions, thresholds)
4. Identification of exposure patterns
5. Shape of exposure-response curves

Based on our primary research question, different methods are appropriate

What research questions could we have? (cont'd)

- No single method outperforms all others for all potential questions
- Other considerations to keep in mind
  1. Interpretability
  2. Robustness (stable solutions)
  3. Computational scalability – as the dimensionality of our dataset increases (either N or p) some methods might start to fail
  4. Exploration vs. hypothesis testing
  5. Might not be a good idea to "blindly" use methods from other fields – may need to adjust them first

Comparing Results across Methods

- Generally a good practice
- Even if different methods address different questions, consistency in findings is always welcome
- If/when differences across methods are detected $\rightarrow$ keep in mind what the aim of each method is!
- Trying different methods and choosing the answer we like the best should *always* be avoided
  - I.e. no cherry-picking!

To supervise or not to supervise?

- Two large categories of methods
  1. Supervised
  2. Unsupervised
- Generally, "supervised" refers to the outcome of interest informing the solution
- Most methods are inherently supervised
  - E.g. variable selection
- But not all
  - E.g. clustering

To supervise or not to supervise? (cont'd)

- As always, it depends on the research question ···
  1. Do we want to inform policy?
  2. Or understand better certain biological pathways?
- If (1) then our goal is to identify common patterns of exposure, independent of outcomes*, on which we can act
  - Through regulatory action, interventions etc
  - E.g. Source apportionment in air pollution
  - *Or rather, to be assessed with many different outcomes
- If (2) then of course any method should include the outcome of interest

In Summary···

- During this workshop we will present information on a few different methods
- That are used to answer different research questions
- By no means a comprehensive list
- The goal is to have an open Discussion about how to quantify health effects of exposure to mixtures
- ···With some extra technical details ;)
- This is a very heterogeneous group in terms of experience with such methods

→ Please share your previous experience!

The example we'll be using during this workshop

- We wanted to use a real-life application for all the labs
  - Instead of simulated datasets
- We decided on the Mitro et al paper:
  1. High-dimensional exposure matrix
  2. Publicly available data (NHANES)
- A big Thank you! to Dr. Ami Zota (GW)
  - Provided the datasets and code to get the final dataset as it is exactly on the paper
  - For consistency

The example we'll be using during this workshop

- We wanted to use a real-life application for all the labs
  - Instead of simulated datasets
- We decided on the Mitro et al paper:
  1. High-dimensional exposure matrix
  2. Publicly available data (NHANES)
- A big Thank you! to Dr. Ami Zota (GW)
  - Provided the datasets and code to get the final dataset as it is exactly on the paper
  - For consistency
- Caveat: None of us is an expert in PCBs...
  - If you are, please help us out! :)

The example we'll be using (cont'd)

- For consistency with the Mitro et al paper and across our labs
- We kept all parameters as in the paper
  - Log-transformed outcome and exposures
  - Same list of confounders included in the models
- Only included variables with at least 60% $>$ LOD
- Since it is not the purpose of the workshop to discuss these choices, please refrain from asking such questions during the labs
- Happy to discuss these during the breaks
  - Although we might not be able to explain the choices the authors made

Thank you!

Questions?

mk3961@cumc.columbia.edu