Introduction

08.19.2019

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

Million dollar question

- The necessity to assess exposure to mixtures is now well-recognized
- US EPA, NRC and NIEHS all agree

Million dollar question

- The necessity to assess exposure to mixtures is now well-recognized
- US EPA, NRC and NIEHS all agree

How can we represent the compexity of reality in a (single) statistical model?

How do we deal with exposure to mixtures?

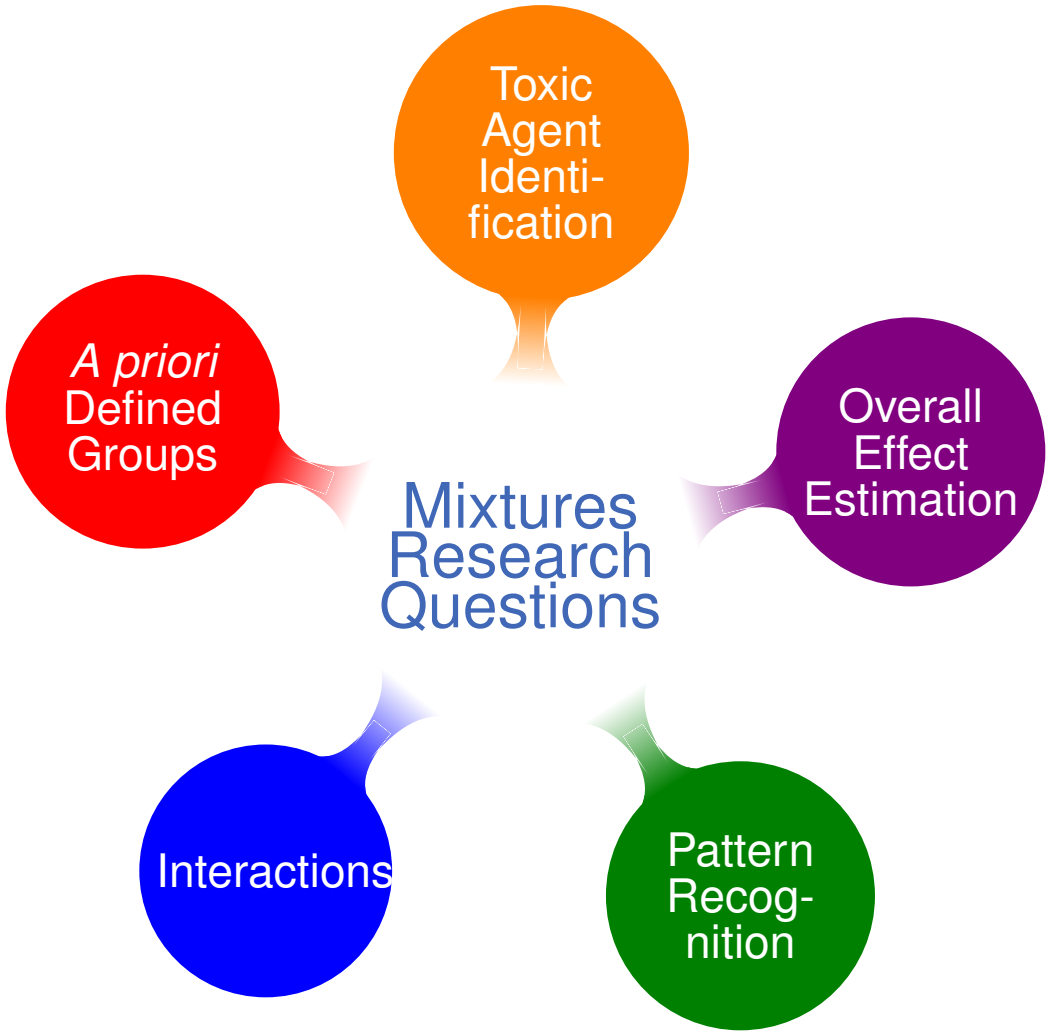
- This is still a very open question
- Existing methods have limitations
- There have been several workshops held by EPA and NIEHS to address this issue
- The most recent NIEHS workshop (2015) concluded that
  1. Although some methods performed better than others the presented estimated associations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

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`

Potential questions in mixtures analyses

For mixtures analyses the selected method depends on the primary research question



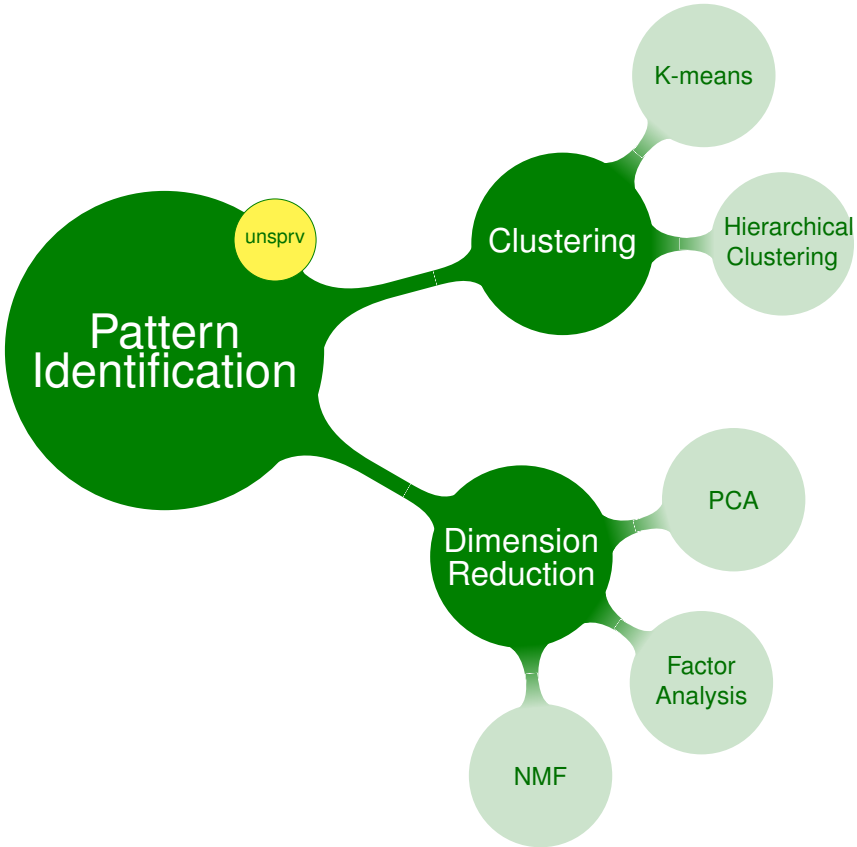
Bird's-eye (over)view of existing mixtures methods

Some considerations

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

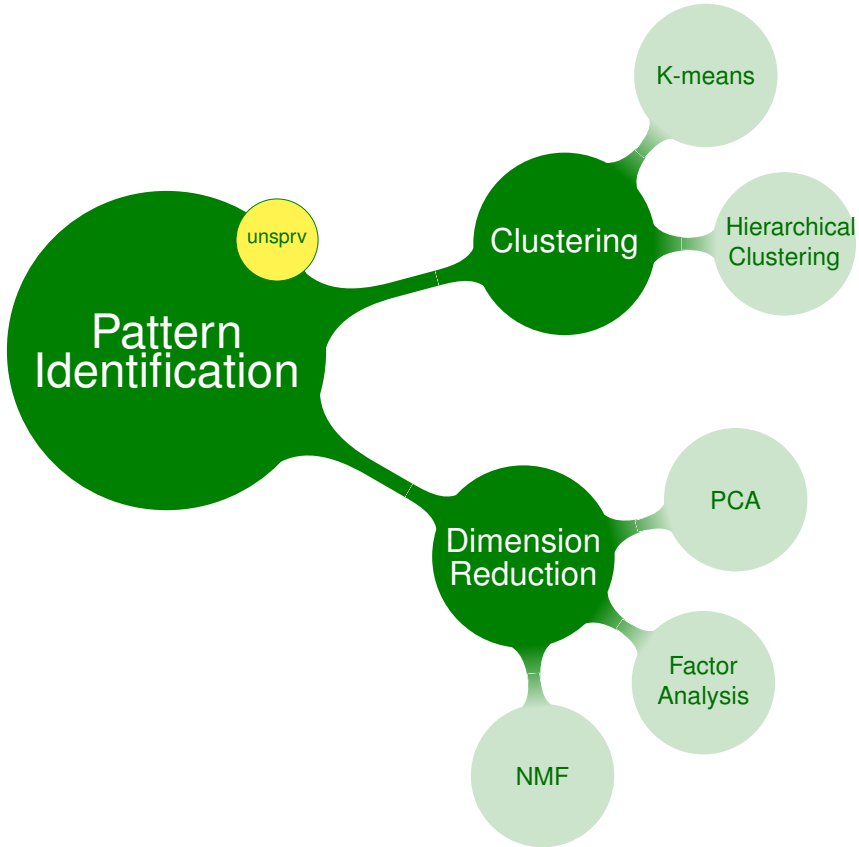
Exposure pattern recognition

- Why should we care about identifying exposure patterns to chemicals in a population?
  - Sources
  - Behaviors
- If we link these patterns to (multiple) adverse health outcomes
  - → Efficient regulations
  - → Targeted interventions



Exposure pattern recognition (cont'd)

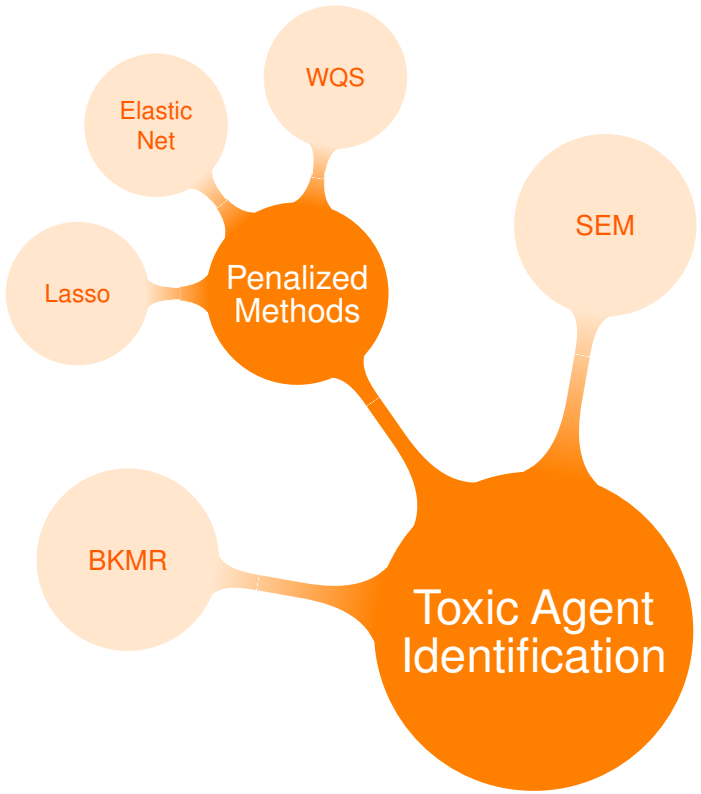
- Class of unsupervised methods
  - Solution independent of any outcome
- But supervised extensions exist (for most)
- Which version to use depends on the research question
  - Inform regulatory action, interventions
  - Better understand biological pathways



Identifying toxic agents

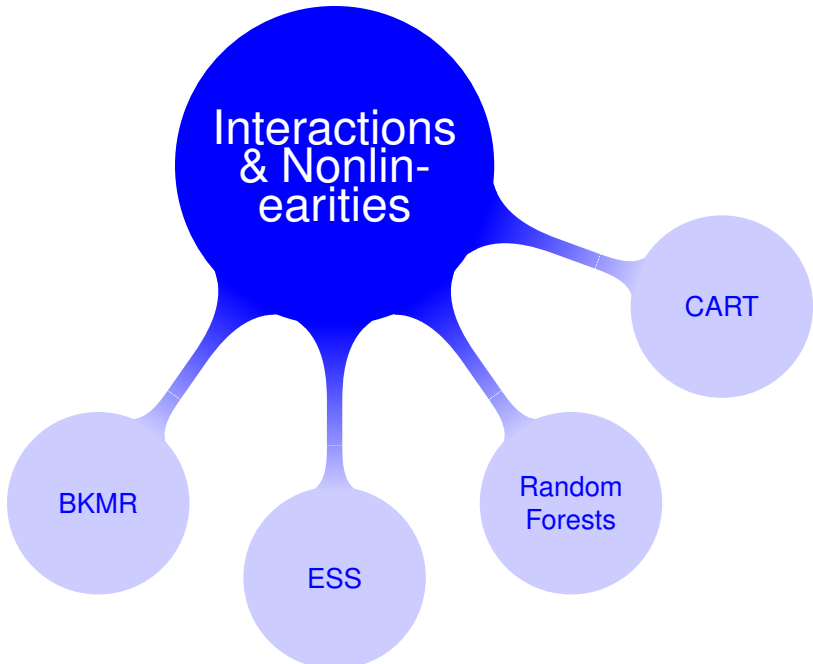
aka “bad actors”

- Which chemical(s) in my mixture are related to the outcome?
- Estimate chemical-specific *independent* effects
- While accommodating the (potentially very) high correlations among mixture members



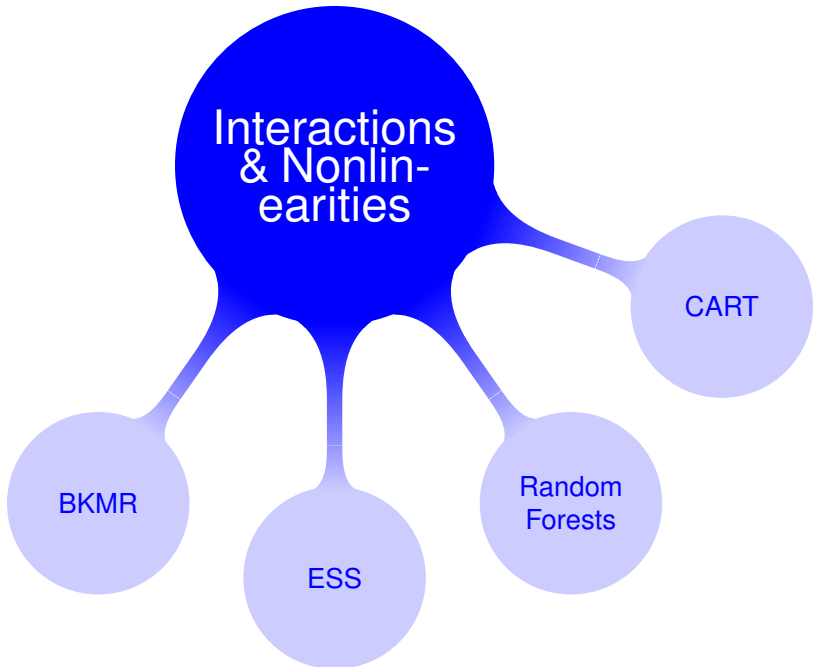
Interactions & non-linear relationships

- These are actually two different classifications of potential research questions
  1. Interactions among mixture members?
  2. Non-linear exposure – response curves?
- But methods tend to do both



Non-linear relationships

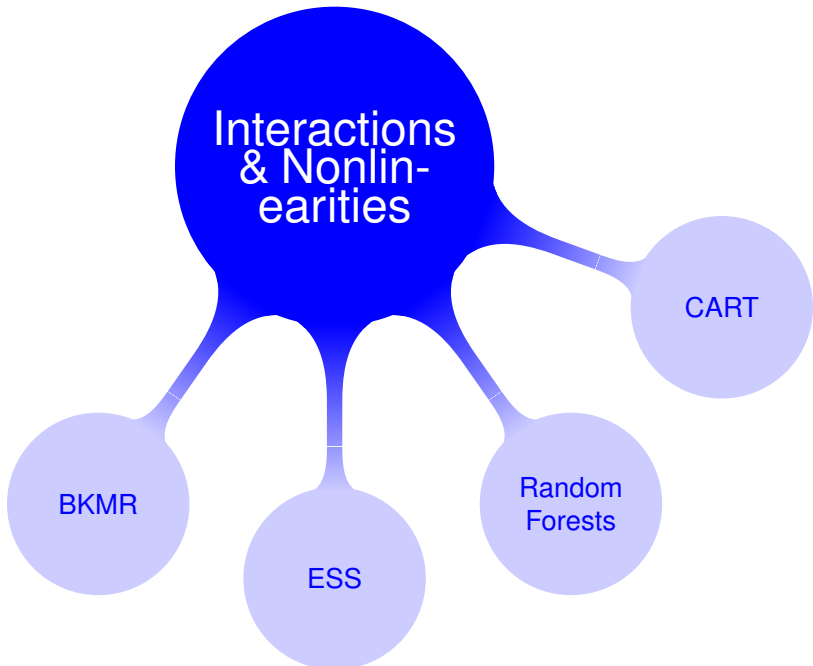
- Because linearity is just an assumption · · ·



Interactions among mixture members

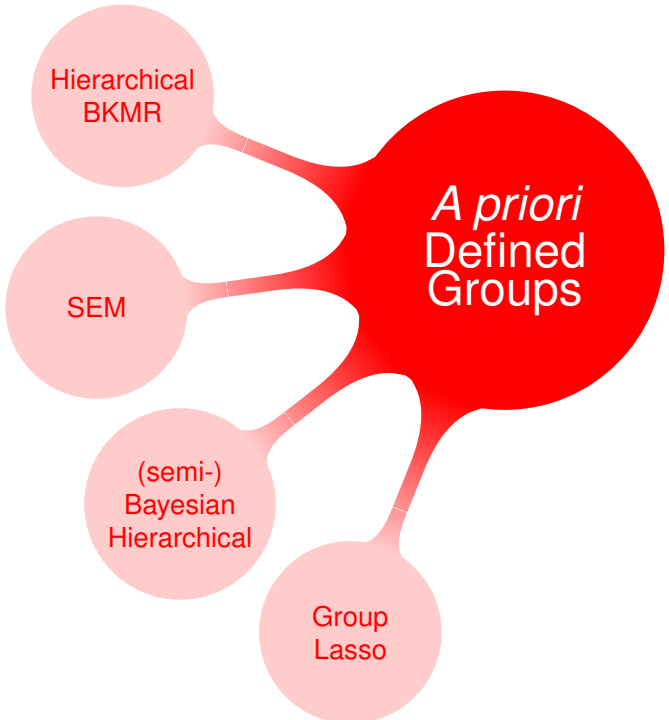
- If there is reason to believe that the combined health effect is greater (or less) than the sum of independent effects
  - Potential synergism
- Most methods can accommodate *a priori* defined interactons
  - Need to hard code
  - Dimensionality $\cdots$

$\rightarrow$ Semi- or non-parametric methods preferred



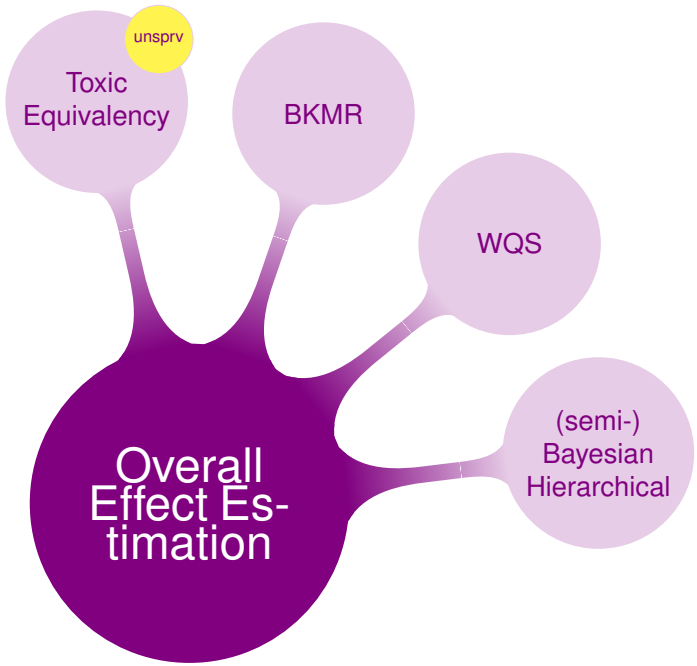
A priori defined groups

- We might have some prior knowledge or hypothesis on
  - How chemicals group naturally in the environment
  - Might share pathway to toxicity
- Methods exist to allow estimation both of group and within-group effects



Overall Mixture Effect

- Finally, we may want to estimate the overall mixture effect
- As chemical concentrations in the mixture increase, do we observe corresponding changes in the outcome?



Comparing results across methods

- Generally a good practice
  - Especially if complementary methods
  - Sensitivity analyses to assess robustness of results
- 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!

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 (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

The platform we'll be using in the labs

https://rstudio.cloud/ 

- Looks like R studio but runs on the cloud
- Might take a bit to load – please load at the break before the first lab
  - or now
- We have shared instructions, but if you have any questions please ask at the break before the first lab!
- Most files are .rmd instead of .r $\rightarrow$ the code is in the gray chunks
- You also have the option to follow using the .html files
  - choose option "view in web browser"

Thank you!

Questions?

mk3961@cumc.columbia.edu