Integrating methods for environmental mixtures in mediation analysis: a statistical tool for evaluating environmental health disparities

Andrea Bellavia

Department of Biostatistics, Department of Environmental Health
Harvard T.H. Chan School of Public Health
abellavi@hsph.harvard.edu

Society for Epidemiologic Research, June 2018

Acknowledgements

- Tamarra James-Todd
- Paige Williams

- Susanna Mitro

Mediation for Environmental Health Disparities

- Health disparities are differences in health status that systematically have an adverse effect on the health of less-advantaged populations
- Groups that are considered to be disadvantaged may depend on culture place, sociological, economic or anthropological constructs

To quantify a given disparity we fit a statistical model for testing the association between the social construct X and the outcome Y

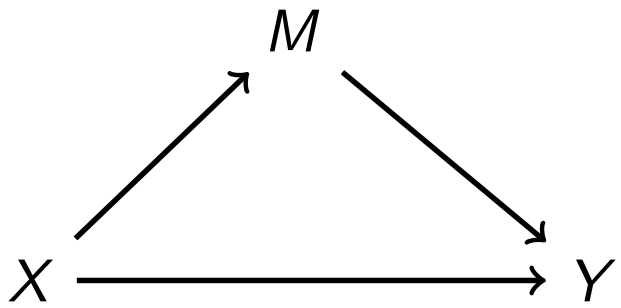
$$X \longrightarrow Y$$

The main characteristic of X is to be non-modifiable. Vulnerable populations can be identified, but interventions and recommendations are more challenging

Reducing Health Disparities

- **If the construct is un-modifiable, how can we reduce the disparity?**
- We would like to develop effective public health recommendations, as well as to implement interventions
- One possibility is to identify proximal risk factors that (partly) explain why these groups show enhanced vulnerability to a specific set of diseases

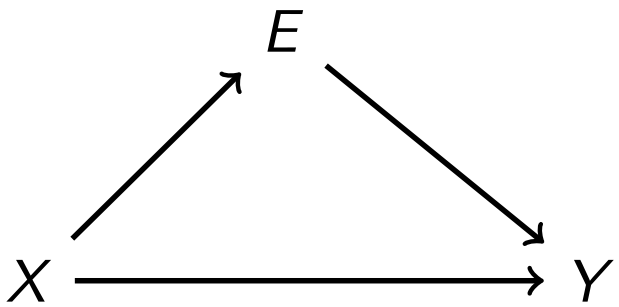
Specifically, the aim is to identify one or more modifiable factors that 1) are risk factors for Y, and 2) are unequally distributed over levels of X



Statistical methods such as mediation analysis are the common techniques used to evaluate the contribution of such factors M in determining the X-Y disparity (Naimi et al., 2012)

Environmental Health Disparities

A growing body of literature suggests that environmental factors may be important contributors to several health disparities (that is: Let's refer to M from the previous graph as an environmental factor E)



Example - Race/ethnicity and diabetes

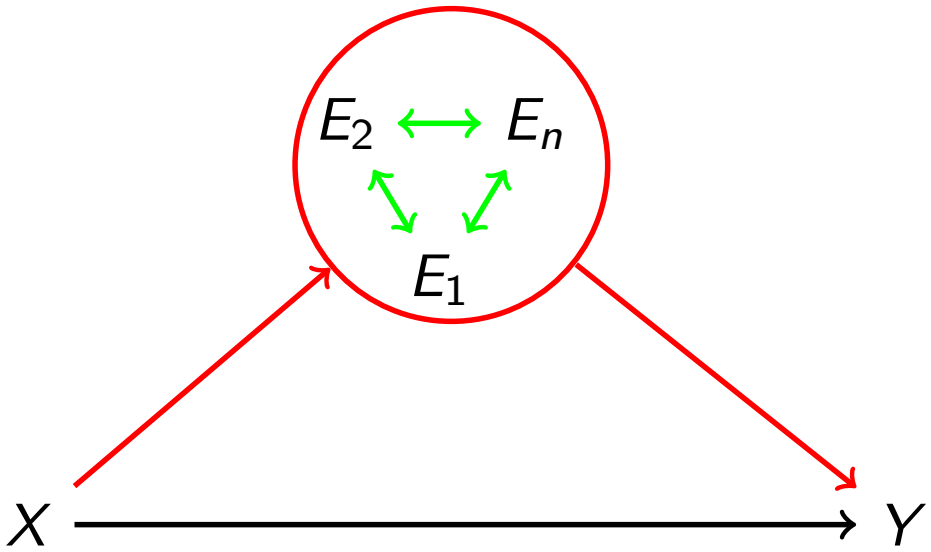
- In the US, racial/ethnic disparities in the prevalence of diabetes have been consistently documented
- Recent studies show a link between high exposure to certain classes of environmental chemicals and diabetes. (Huang et al., 2014)
- Racial/ethnic differences in the exposure to these chemicals have also been observed (James-Todd et al., 2014)
- **Does the higher exposure to chemicals (partly) explain the higher prevalence of diabetes in specific racial/ethnic subgroups?**

- So far, little has been done to quantify the contribution of environmental factors in health disparities
- One of the reasons for this literature gap is the additional number of challenges that evaluating environmental factors involves

Environmental mixtures

- Humans are generally exposed to a mixture of environmental factors
- Several studies (Taylor et al, 2012) recommend evaluating exposure to environmental mixtures rather than one exposure at the time
- Such reasoning extends to the evaluation of environmental health disparities. A potential contribution to the disparity is more likely given by a mixture of environmental components.

- Nevertheless, the different components of the mixture may have different chemical/biological properties and therefore different effects on Y.
- Ideally, we would like to both identify the overall contribution of the mixture, as well as the specific contribution of its components
- Furthermore, interactions between mixture components should be taken into account


E_2
E_n
E_1
X
Y

Methods

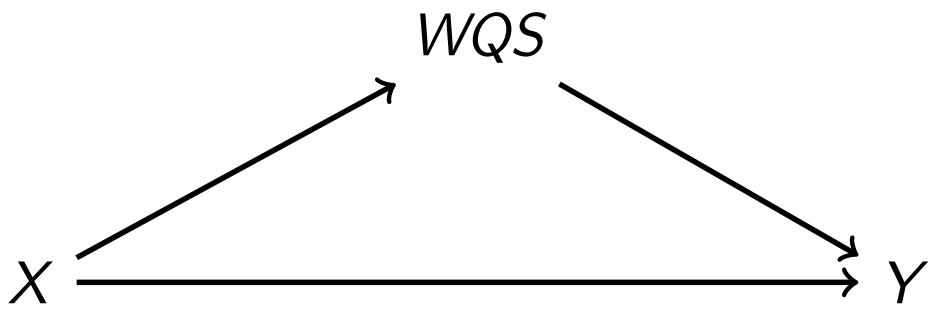
- The most intuitive way of integrating the environmental mixture as a mediator of the $X - Y$ association would be to conduct mediation analysis within the multiple mediators parametric framework
- However, as the number of mixture components increases, Our analyses may get seriously hampered by problems of overfitting
- Moreover, components of the mixture are often highly correlated, thus making collinearity another treat for the analyses

Methods for mixtures that accommodate these features should therefore be considered. A broad review can be found in Taylor et al., 2012:

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

Summary methods - Weighted Quantile Sum

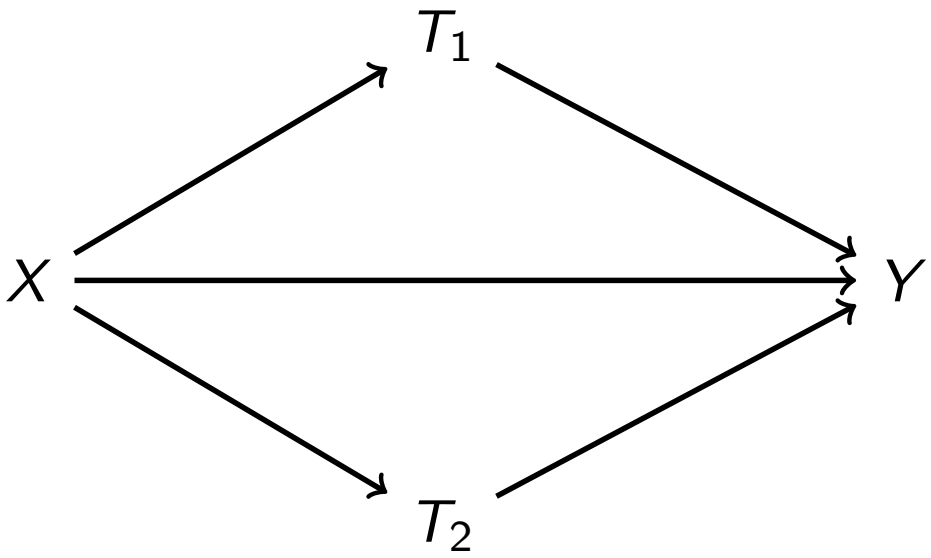
- Summary methods can be used to reduce the mixture to a single score (ie to a single mediator)



- A recommended method to create a score is the Weighted Quantile Sum regression approach (Carrico et al. 2015)
- The main advantage is that the indirect effect will represent the overall indirect effect due to the mixture
- However, retrieving the specific contribution is not straightforward

Classification methods - Principal component analysis

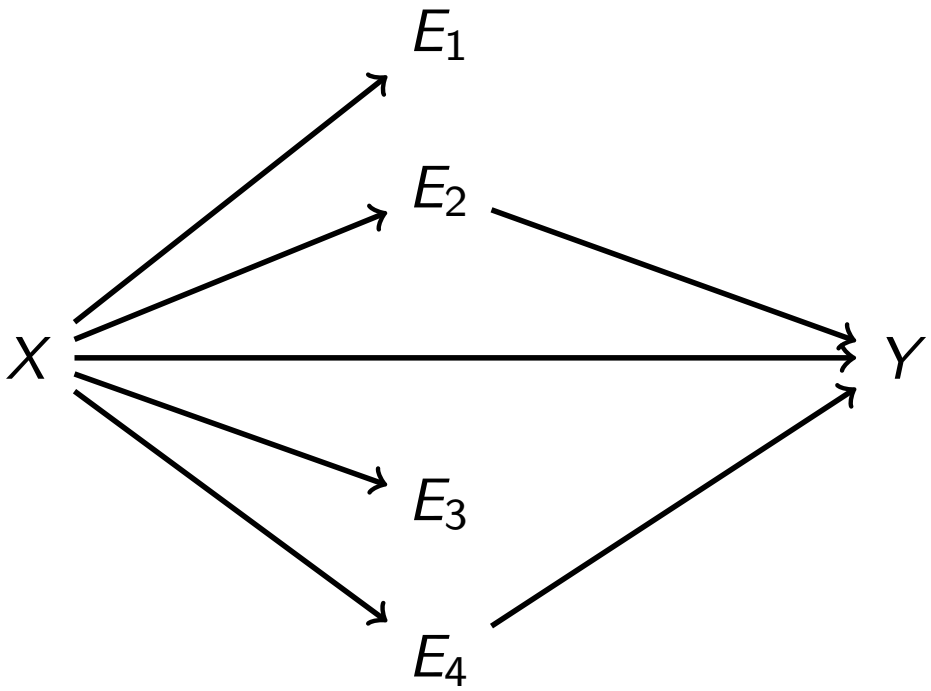
- Classification methods can be used to reduce the dimension of the mixture when meaningful groups can be identified
- PCA often represents a suitable classification methods as it results in summary factors that are uncorrelated.
- Scores from PCA can be integrated in a multiple mediation context as presented in the following figure:



- The main advantage of classification approaches based on the correlation structure is that these will likely identify the "real-world" exposure patterns
- To further identify the specific contribution of mixture components, PCA results can be used as a preliminary step for hierarchical modeling (Correia and Williams, 2016)
- One limitation of this (and of the previous) approach is that interactions between mixture components can not be identified

Selection methods and alternative methods for mixtures

- When the expected effects of the mixture components on the outcome present additional complexities (eg interactions, non-linearities ...), or in general when one is more interested in the specific contribution of individual components, selection methods (eg Lasso, regression trees) should be pursued.
- These can be used in a two-stage procedure, by using such methods to evaluate the $E - Y$ association, and only integrate relevant components in the final mediation model


E_1
E_2
X
Y
E_3
E_4

- The main advantage of this approach is that any method can be potentially used, including novel approaches such as Bayesian Kernel Machine Regression (BKMR) that provides a flexible non-parametric modeling that allows for both interactions and non-linearities (Bobb et al., 2012)
- On the other hand, the selection will only be based on the $E - Y$ association and not on the entire mediation model

Summary

- Environmental factors are potential contributors to the development of health disparities
- Few studies have focused on (or are even designed for studying) environmental disparities. A methodological framework for their investigation is not established
- Mediation analysis can be useful for evaluating environmental health disparities, but this framework should take into account specific features such as the exposure to mixtures
- Methods for mixtures are available and can integrated in a mediation analysis context
- The most suitable method may depend on the question of interest and on the data structure

Current and future research

- We are currently conducting simulation studies to test the performances of all the presented methods
- All discussed mixture methods (and others) are available in R, but are very limited in other software
- Applications!

References

- Bellavia A et al. Multiple mediators approach to study environmental chemicals as determinants of health disparities. Environmental Epidemiology. 2018 Jun 1;2(2):e015.
- Bobb JF et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics. 2014 Dec 22;16(3):493-508.
- Carrico C et al. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. Journal of Agricultural, Biological, and Environmental Statistics. 2015 Mar 1;20(1):100-20.
- Correia K, Williams PL. A hierarchical modeling approach for assessing the safety of exposure to complex antiretroviral drug regimens during pregnancy. Statistical methods in medical research. 2017 Jan.
- Huang T et al. Gender and racial/ethnic differences in the associations of urinary phthalate metabolites with markers of diabetes risk: National Health and Nutrition Examination Survey 2001?2008. Environmental Health. 2014 Dec;13(1):6.
- Naimi AI et al. Mediation analysis for health disparities research. American journal of epidemiology. 2016 Aug 3;184(4):315-24.
- Taylor KW et al. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: lessons from an innovative workshop. Environmental health perspectives. 2016 Dec;124(12):A227.