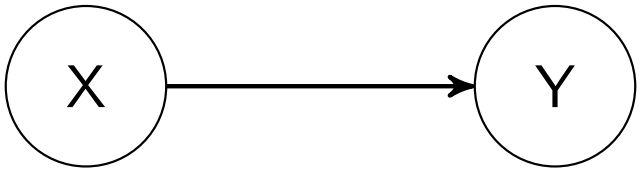
Mediation and interaction analysis for health disparities research

Andrea Bellavia

abellavi@hsph.harvard.edu

3 October 2016

Epidemiology, public health, and clinical research are largely about assessing and explaining exposure-outcome associations

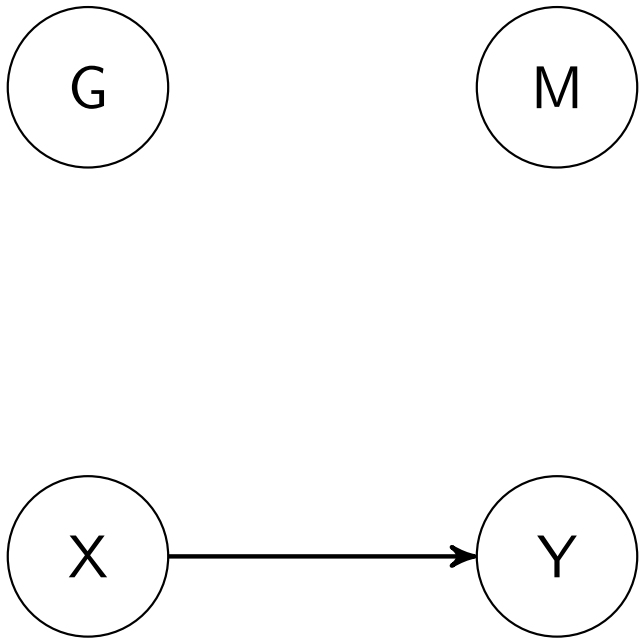


- **Observing a statistical association only represents the starting point**
- Addressing additional questions is required before implementing public health intervention and recommendations

Today we will focus on two specific questions:

- How is the association generated?
- Does the association only exist for specific subgroups?

Statistically/epidemiologically speaking, we are investigating the contribution of other covariates in the X-Y association



- These questions are particularly relevant when X is something we can't intervene on (i.e. non-modifiable exposures).
- Health disparities are defined as differences in health status that are systematically affecting groups of people based on their racial or ethnic group; religion; socioeconomic status; gender; age; mental health; cognitive, sensory, or physical disability; sexual orientation or gender identity; geographic locations; historical characteristics. (Healthy People 2020).
- **Addressing our additional questions may contribute to identify and implement public health intervention and recommendations to reduce/prevent health disparities**

Motivating example 1 - Race/ethnicity and diabetes

- In the US, racial/ethnic disparities in the prevalence of diabetes have been consistently documented (from Jackson, 1971).
- Recent studies show a link between high exposure to certain classes of environmental chemicals and diabetes. (James-Todd et al., 2012)
- 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?**
- Moreover, many of these chemicals are found in fast-food (Zota et al., 2016), and fast-food consumption differs across race/ethnicity (James et al., 2014).
- **What is the contribution of regular fast-food consumption in the reported disparity?**
- We want to quantify the contribution of chemicals exposure in the racial/ethnic disparity in diabetes, and to identify the proportion of disparity that could be reduced by implementing specific nutritional programmes.

Motivating example 2 - HIV infection at birth and adolescence health

- HIV-infected children born to women living with HIV experience delays in pubertal onset and sexual maturity, with higher risk subsequent risk of the related social and medical outcomes (Williams et al., 2013).
- Differences in growth (i.e lower height and higher BMI) are reported among children living with HIV (Arpadi et al., 2000).
- Growth is on of the main predictors of pubertal timing (Tanner, 1976).
- **Is the delay in puberty observed among HIV+ children due to differences in growth? To what extent?**
- Interventions aimed at improving growth during childhood and adolescence might prevent/reduce the disparity.

Motivating example 3 - Side effects in Schizophrenia clinical trial (CATIE)

- New-generations drugs have shown substantial improvements in the treatment of schizophrenia (Liberman et al., 2005).
- However, severe side-effects and excess weight gain have been observed (Clark et al, 2011)
- **Is drugs efficacy the same among patients with different side-effects?**
- Identifying subgroups of patients with different drug responses would allow personalising drug regimes.

Other potential fields of application

- Contribution of lifestyle factors in social epidemiology (e.g. The Stockholm Public Health Cohort)
- Interaction between genetic and lifestyle factors.
- Medical predecessors of common diseases (e.g. diabetes, hypertension, depression)
- Epigenetics (and genes-environment interactions).
- . . . many other

Investigating this contribution of third variables in an X-Y association generally requires merging multiple research fields. Interaction and mediation often require a **multidisciplinary approach**.

Mediation and Interaction

Interaction and mediation analysis are the statistical tools that address our two questions of interest:

- Interaction: **for whom** an effect occurs
- Mediation: **how** an effect occurs

The biological question comes before the statistical one

- Statistical methods to assess interaction and mediation (or even confounding), are generally simple (and often similar)
- **To correctly identify the potential role of each variable involved in the association is crucial.** Methods to conceptualise the causal pathway are available and increasingly recommended (e.g. DAGs)
- Failing to identify the correct DAG may lead to severe bias once we move to statistical analysis (i.e. the correct method is used to address a wrong question)

Data for the practical example

We will simulated data that recall Motivating Example 1 (slides 5-6). Using simulated data simplifies:

- Implementation and interpretation (i.e. no-unmeasured confounders; all assumptions are met)
- Reproducing the code to your own situation
- Sharing the same data

Code for generation data, also with all analyses, can be downloaded from:
http://abellavia.altervista.org/alterpages/files/WorkshopKI_OCT3.do

Simulated data

Description of simulated population

	Overall n=10000	Diabetes (Y=1) n=1063	No-Diabetes (Y=0) n=8937
Black-American (X), n	1863	275	788
Fast-food consumption (M1), n	3439	457	606
DiNP (M2), mean	11.2	11.1	11.9

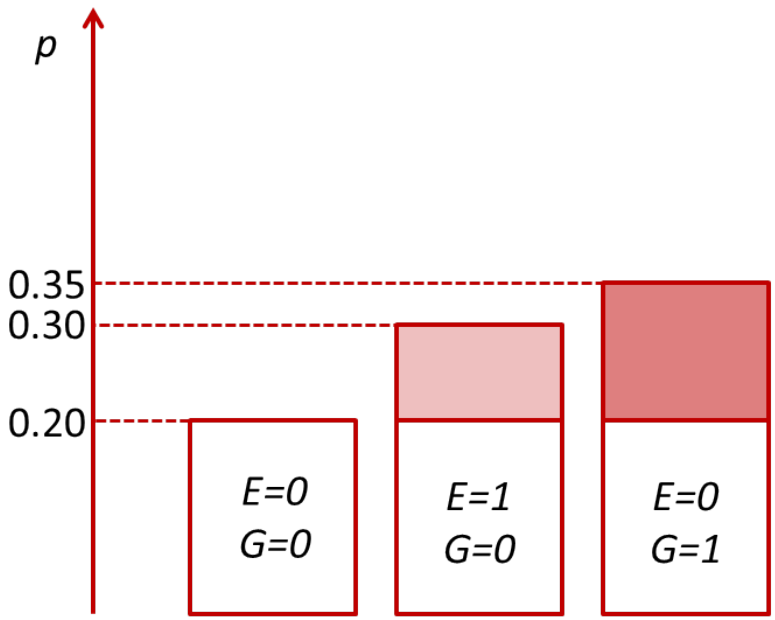
DiNP=di-isononyl phthalate

Concept of Interaction

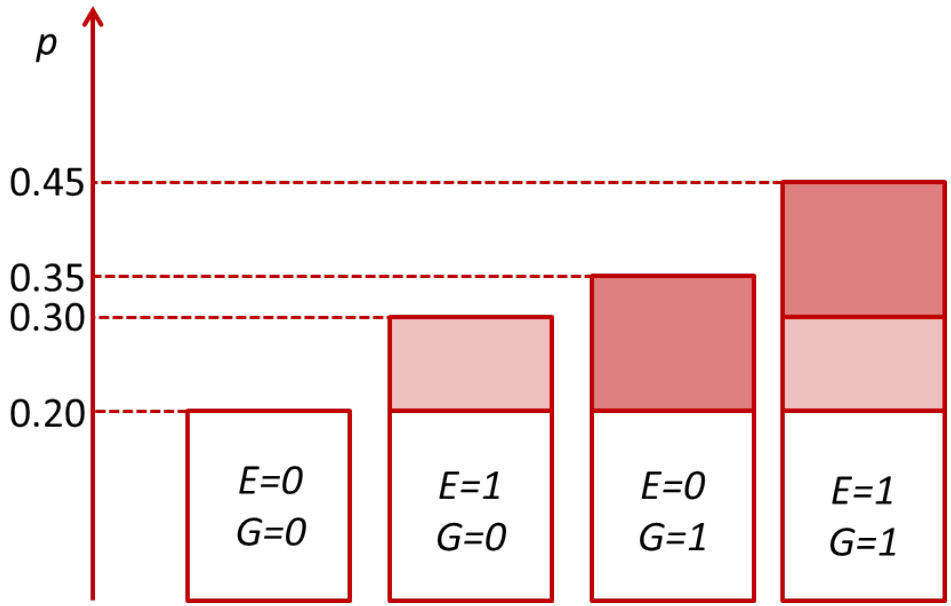
- The effect of a given exposure on the outcome of interest may depend on the presence or absence of another exposure
- Interaction and effect modification (EM) have a similar biological meaning.
- Vanderweele 2009: *On the distinction between Int. and EM.*
- Interaction can be defined on the additive or multiplicative scale

Dichotomous exposures

- Suppose we have a dichotomous outcome Y and two dichotomous exposures G and E
- We are interested in $P(Y = 1|G = g, E = e)$, the probability of Y given the two exposures



Additive Interaction

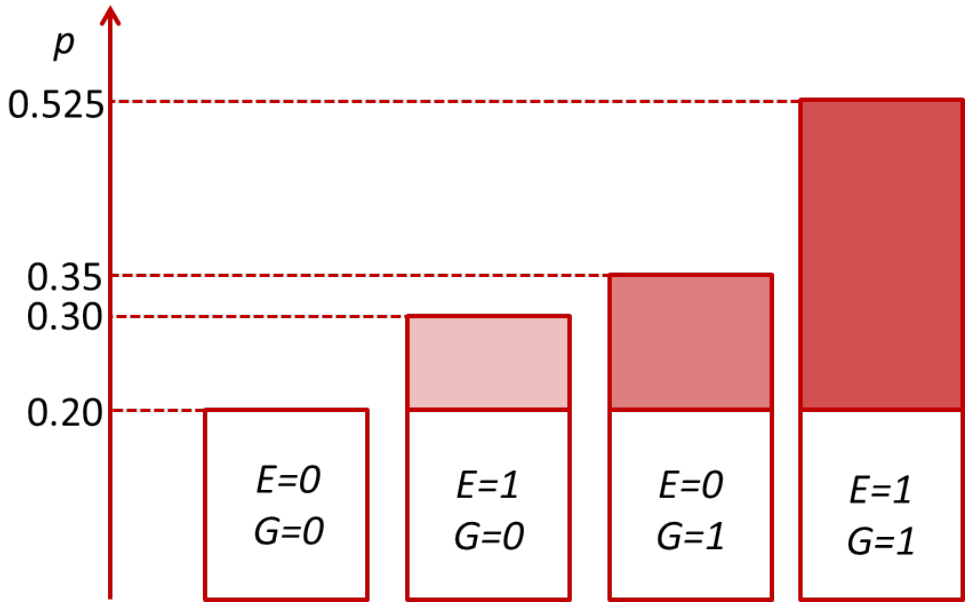


Definition of additive interaction

$$p_{11} = p_{10} + p_{01} - p_{00}$$

$$0.35 + 0.30 - 0.20 = 0.45$$

Multiplicative Interaction



Definition of multiplicative interaction

$$p_{11} = (p_{10} \times p_{01})/p_{00}$$

$$(0.35 \times 0.30)/0.20 = 0.525$$

Additive vs Multiplicative Interaction

- The sign of the interaction depends on the chosen scale. A common situation is to have negative interaction on the multiplicative scale, but positive interaction on the additive scale.
- In general, if both exposures have an effect on the outcome there should be an interaction on some scale.
- **Additive interaction is the more relevant public health measure. It can be used to identify a particular subgroup that should be treated**.
- For example, presence of additive interaction in the previous situation implies that the public health consequence of an intervention on E would be larger in the $G = 1$ group

Statistical interaction

- Statistical interaction is commonly assessed by including a product term in the regression model
- When dealing with dichotomous exposures, this method yields the same results of a simple stratification
- Whether additive or multiplicative interaction is evaluated depends on the chosen statistical model
- A statistical model on the linear scale can be used to evaluate statistical interaction on the additive scale

$$P(Y = 1|G = g, E = e) = \alpha_0 + \alpha_1 g + \alpha_2 e + \alpha_3 ge$$

- A statistical model on the log-linear scale can be used to evaluate statistical interaction on the multiplicative scale

$$\log[P(Y = 1|G = g, E = e)] = \beta_0 + \beta_1 g + \beta_2 e + \beta_3 ge$$

Interaction: Extensions and limitations

- To derive additive interaction when the statistical model is *log*-linear, measures such as the RERI can be used (Li & Chambless, 2007).
- In survival analysis, additive models can also be used (Rod et al., 2012; Bellavia et al., 2016)
- Interpreting a measure of interaction is simple and intuitive when dealing with dichotomous exposures, complicated with categorical, and extremely challenging with continuous
- There is no agreement on the definition of interaction, neither on the distinction between interaction and effect modification. One can also use the term *combined effect*

Practice

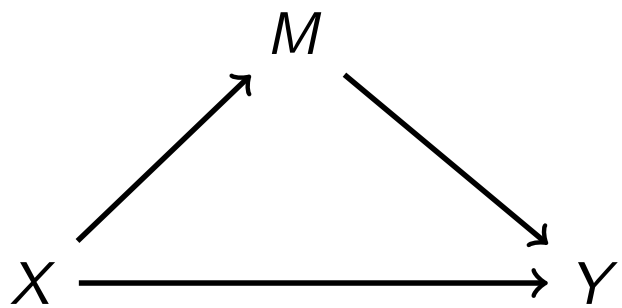
- Q1. Are DiNP levels higher among subjects with regular fast-food consumption?
- Q2. Is the association similar among black-Americans and other racial/ethnic groups? Address this question with two statistical methods.
- Q3. Imagine that only DiNP higher than 11.5 ug/l has negative health effects. Use a logistic regression to estimate the odds ratio of DiNP>11.5 between fast-food consumer vs non-consumer.
- Q4. Is there an interaction between race/ethnicity and fast-food consumption in predicting the odds of high DiNP? On both scales?

Practice

- Q1. Yes. $\beta = 1.15; 95\% CI : 1.07 - 1.24$
- Q2. Stratified analysis shows that the effect of fast-food consumption is higher among black-Americans ($\beta = 1.40$ vs $\beta = 0.98$). Linear regression, including an interaction term, will yield the same results, but additionally provide a statistical test for additive interaction ($\beta = 0.47; 95\% CI : 0.27 - 0.67$).
- Q3. OR=2.40; 95% CI: 2.21-2.61
- Q4. Multiplicative interaction: OR=1.69; 95% CI: 1.32-1.16. Additive interaction: RERI=7.8; 95% CI: 5.3-10.2
- **Public health recommendations to reduce fast-food consumption should be primarily implemented in the black-American population. There is a racial/ethnic disparity in the association between fast-food consumption and DiNP levels.**

Concept of Mediation

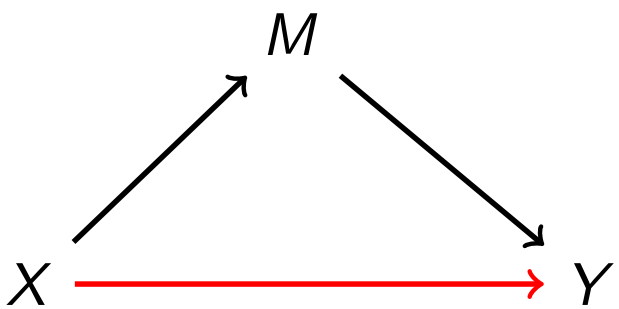
- A mediator is a covariate that mediates the association between X and Y



Part of the effect of X on Y is due to the fact that X causes M, which in turn causes Y

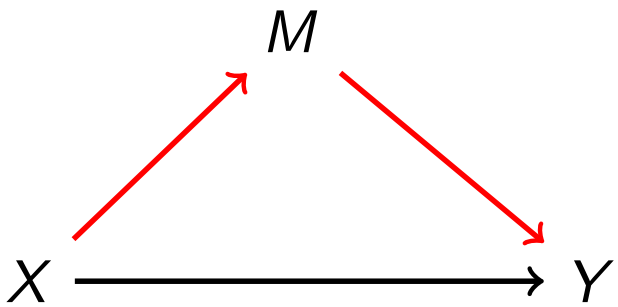
Concept of Mediation

We aim to disentangle the *total effect* of X on Y into a **direct effect** that goes through all possible pathways but M . . .



Concept of Mediation

. . . and an **indirect effect** that goes through M



Assessing Mediation - continuous outcomes

- The classical approach to Mediation Analysis (Baron & Kenny, 1986)

$$E[Y|x] = \alpha_0 + \alpha_1 \cdot x$$
$$E[Y|x, m] = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot m$$
$$E[M|x] = \gamma_0 + \gamma_1 \cdot x$$

- Once the three statistical models are correctly identified we can estimate direct and indirect effects:

$$\text{Direct Effect} = \beta_1$$
$$\text{Indirect Effect} = \alpha_1 - \beta_1 = \beta_2 \cdot \gamma_1$$

- Adjusting a model for a mediator gives an estimate of the direct effect rather than the total effect. **Crucial to distinguish confounders and mediators!**
- All models can be adjusted for potential confounders
- The product=difference statement is not valid under some scenarios (binary and survival outcomes, exposure-mediator interaction, missing values on mediators)
- A common measure to summarise results from a mediation model is the **proportion mediated**

$$PM = \frac{IndirectEffect}{TotalEffect}$$

Assessing Mediation - binary outcomes

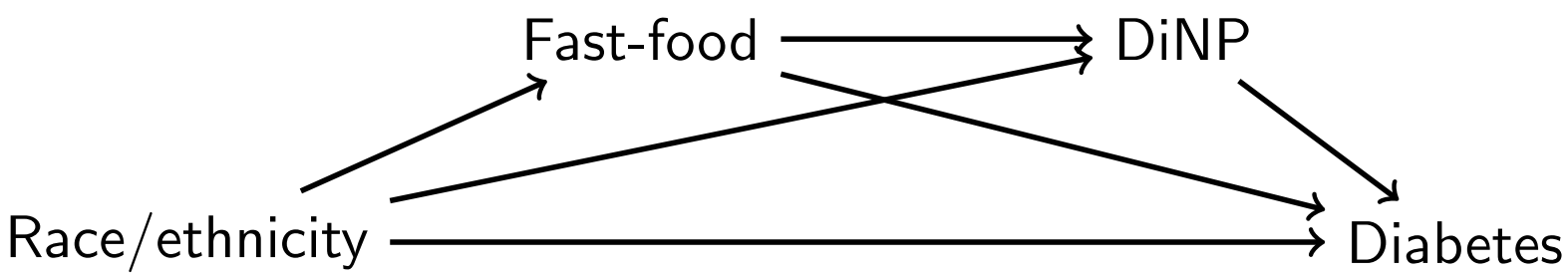
$$logit[Y|x] = \alpha_0 + \alpha_1 \cdot x$$
$$logit[Y|x, m] = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot m$$
$$E[M|x] = \gamma_0 + \gamma_1 \cdot x$$

- Once the three statistical models are correctly identified we can estimate direct and indirect effects:

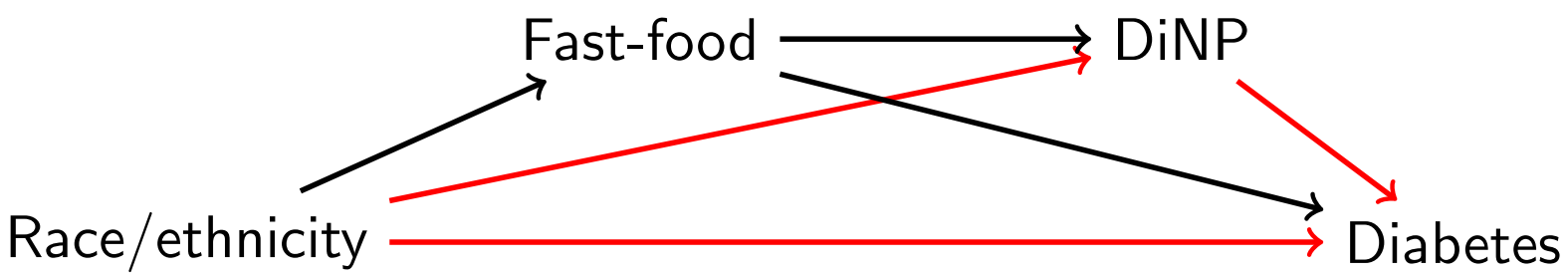
$$\text{Direct Effect} = exp(\beta_1)$$
$$\text{Indirect Effect} = exp(\beta_2 \cdot \gamma_1)$$

- The product method is recommended as the equivalence with the difference method only yields when the outcome is rare.

Practice



Practice - Binary outcome

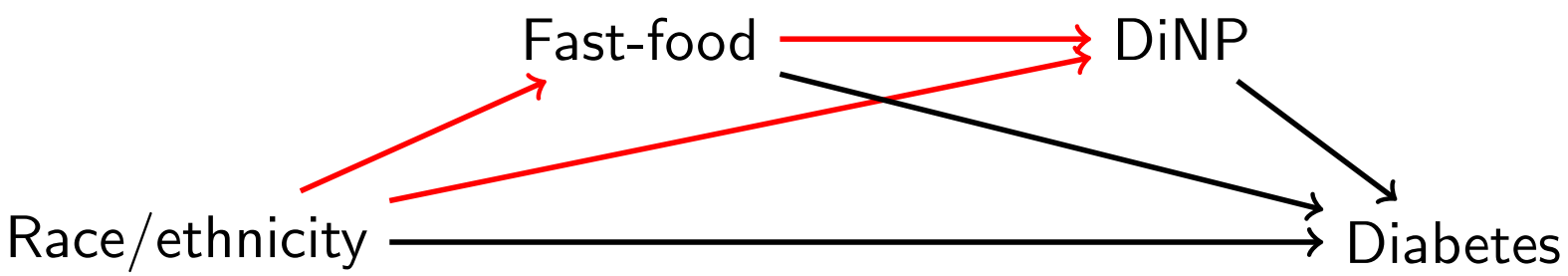


- Q1. Use a logistic regression model to estimate the racial/ethnic disparity in diabetes, and a linear regression model to estimate the racial/ethnic disparity in DiNP exposure.
- Q2. Further adjust the main model (i.e. diabetes as a function of race/ethnicity) for DiNP concentration to estimate the direct effect.
- Q3. Calculate the indirect effect with the product method (no CI required).
- Q4. Calculate the indirect effect with the difference method (no CI required). Why do they differ?
- Q5. Calculate the proportion of the total effect mediated by DiNP.

- Black-American have 60% higher odds of diabetes (OR=1.61; 95% CI: 1.39-1.87 - **total effect**) and have higher DiNP urinary concentration (β=1.80; 95% CI: 1.70-1.90)
- Further adjusting the main model (i.e. diabetes as a function of race/ethnicity) for DiNP concentration the main effect goes down to 19% (OR=1.19; 95% CI: 1.01-1.39 - **direct effect**)
- **Indirect effect** calculated with the product method: OR=1.37
- Indirect effect calculated with the difference method: OR=1.36 (the outcome is not rare)
- Proportion Mediated: 65%
- Assuming that there are no unmeasured confounders we can conclude that higher DiNP is responsible for 65% of the racial/ethnic disparity in diabetes.

- How can we reduce DiNP? One source of this chemicals is fast-food consumption. How much of the difference in DiNP exposure between black-Americans and other race/ethnicities would be reduced by eliminating fast-food consumption?

Practice - Continuous outcome



- Q1. Use a linear regression model to estimate the racial/ethnic disparity in DiNP, and a logistic regression model to estimate the racial/ethnic difference in fast-food consumption.
- Q2. Further adjust the main model (i.e. DiNP as a function of race/ethnicity) for fast-food consumption to estimate the direct effect.
- Q3. Calculate the indirect effect with the product method (no CI required).
- Q4. Calculate the proportion of the total effect mediated by fast-food consumption.
- Q5. Calculate the indirect effect with the difference method (no CI required). Is this equal to the one calculated in Q3? Why?

- Black-American have higher DiNP urinary concentration (β=1.80; 95% CI: 1.70-1.90 - **total effect**) and higher odds of fast-food consumption (OR=1.60; 95% CI: 1.45-1.78)
- Further adjusting the main model (i.e. DiNP as a function of race/ethnicity) for fast-food consumption the main effect goes down to 1.69 ug/l (β=1.69; 95% CI: 1.59-1.79 - **direct effect**)
- **Indirect effect** calculated with the product method: β=0.11 ug/l
- Proportion Mediated: 6%
- We can conclude that fast-food consumption is responsible for only 6% of the higher DiNP level among black-American. Other sources of exposure must be identified.

- What about the difference method?
- Indirect effect: 0.48 ug/l
- Proportion Mediated: 27%
- With continuous outcomes these were expected to be equal. What's going on?

The counterfactual approach to mediation analysis

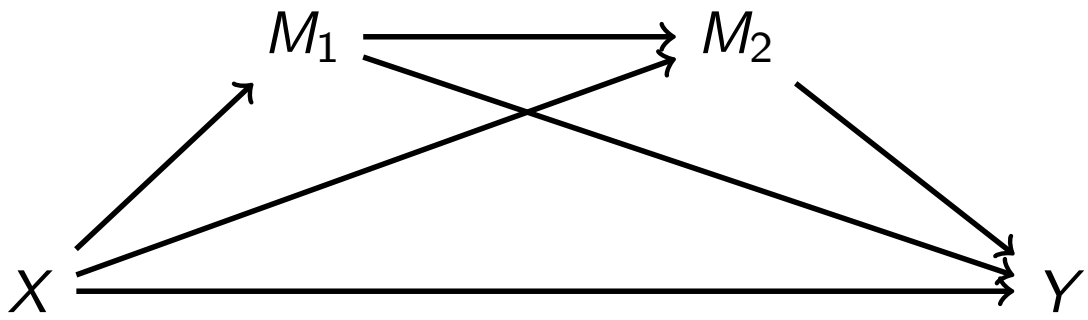
- The problem in the latter example is the strong **interaction** between race/ethnicity and fast-food consumption in predicting DiNP levels.
- The classical approach to mediation analysis does not allow to take into account exposure-mediators interaction. If this is present, results may be severely biased.
- Other limitations of the classical approach to mediation analysis include controlling for mediator-outcome confounding, and assessing non-linearities.
- **The counterfactual approach to mediation analysis was introduced to overcome these important limitations** (recently summarized in Vanderweele 2015).

Unification of mediation and interaction

- This approach defines direct and indirect effects in terms of the counterfactual intervention [i.e. fixing exposure and mediator to a predefined value (**controlled**), or fixing the exposure to a predefined value and the mediator to the value that naturally follows (**natural**)].
- Vanderweele (2014) also showed that by using the counterfactual approach **the total effect can be decomposed into four different components**: a direct effect (controlled), a proportion due to mediation alone (natural indirect effect), a proportion due to interaction alone (*reference interaction*), and a proportion due to both mediation and interaction (*mediated interaction*).

Multiple Mediators

- We have seen, in our hypothetical example, that racial/ethnic disparities in diabetes are mediated by high DiNP exposure.
- We have also seen that racial/ethnic disparities in DiNP levels are mediated by fast-food consumption.
- We may think of all the components in the same pathway.



We are interested in:

- The direct effect (not going through neither M_1 nor M_2)
- The effect mediated by M_1 only
- The effect mediated by M_2 only
- The effect mediated by both M_1 and M_2

- The basic approach to multiple mediators only requires including an interaction between the two mediators and specifying two additional statistical models (Vanderweele & Vansteelandt 2013)
- Binary and continuous outcomes can be evaluated
- When multiple mediators are of interest it may be difficult to specify correct statistical models for all of them and for their interactions. A weighting approach can be used in those situations
- Extension to the multiple mediators approach to other settings is not straightforward, and a simple framework to include exposure-mediators interactions is not yet available

Mediation: other extensions

- Causal Mediation Analysis (Robins & Greenland 1992, Pearl 2001)
- Binary and count outcomes (Valeri & Vanderweele 2013)
- Survival outcomes (Vanderweele 2011)
- Non-parametric approach (Imai et al. 2010)
- Time-varying Exposures and Mediators (van der Laan & Petersen 2008, Vanderweele 2009)

Mediation: software

- Stata: paramed
- R: mediation
- SAS: macro from Valeri & Vanderweele 2013

Summary

- Mediation and interaction are about **explaining** mechanisms that underlie exposure-outcome associations.
- Interaction provides information on **for whom** a given effect occurs
- Mediation provides insight on **how** an effect occurs
- Recent methodological developments allow evaluating interaction and mediation in a large variety of settings
- These methods can be applied with different data structures and to address different research questions, particularly those dealing when non-modifiable exposures.

Causality

- We have not touched the important topic of causality.
- In order to plan public health interventions and recommendations we require evidence in support of a causal association between X and Y, and we need to define interaction and mediation in causal terms.
- The counterfactual approach to mediation and interaction analysis defines all the **assumptions** required to interpret results in causal terms.
- The complexity is not in the statistical methods, rather in formulating those assumptions and in performing **sensitivity analyses**.

Causality - assumptions and conditions

- A temporal sequence of exposure, mediator, and outcome, is generally recommended. In health disparities research, however, we can often assume *a priori* that the exposure precedes the mediator (e.g. gender, race/ethnicity)
- Exposure-mediator, exposure-outcome, mediator-outcome, and mediator-mediator confounding must be all taken into account and are dealt in different ways. **Studies are often designed without thinking of mediator-outcome confounders**.
- Two assumptions of unmeasured confounding are required to causally interpret controlled direct effects.
- Two additional assumptions are required for the other effects.

Causality in health disparities research

- Good news: in this context the measure of interest is generally the controlled direct effect: only two assumptions are required.
- In health disparities research, the CDE is also referred to as **counterfactual disparity measure** (CDM), and represents the proportion of disparity that would remain if we intervene on the mediator (Naimi et al., 2015).
- Bad news: can we really talk about causality when dealing with characteristics such as gender and race? That is, are sex and race causes? There is a huge scientific debate on this, with no agreement (see for instance Vanderweele & Robinson, 2014; Glymour & Glymour, 2014).
- In any case, interpret mediation and interaction analysis in causal terms requires extreme caution.

Main referecens on mediation and interaction

- Baron, R. M., & Kenny, D. A. (1986). The moderator mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.
- Bellavia, A., Bottai, M., Orsini, N. (2016). Evaluating Additive Interaction Using Survival Percentiles. Epidemiology, 27(3), 360.
- Bind, M. A., Vanderweele, T. J., Coull, B. A., & Schwartz, J. D. (2016). Causal mediation analysis for longitudinal data with exogenous exposure. Biostatistics, 17(1), 122-134.
- Gamborg, M., Jensen, G. B., Srensen, T. I., & Andersen, P. K. (2011). Dynamic path analysis in life-course epidemiology. American journal of epidemiology, 173(10), 1131-1139.
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. Psychological methods, 15(4), 309.
- Lange, T., & Hansen, J. V. (2011). Direct and indirect effects in a survival context. Epidemiology, 22(4), 575-581.
- Li, R., & Chambless, L. (2007). Test for additive interaction in proportional hazards models. Annals of epidemiology, 17(3), 227-236.
- Pearl, J. (2001). Direct and indirect effects. In Proceedings of the seventeenth conference on uncertainty in artificial intelligence (pp. 411-420). Morgan Kaufmann Publishers Inc..
- Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology, 143-155.
- Rod, N. H., Lange, T., Andersen, I., Marott, J. L., & Diderichsen, F. (2012). Additive interaction in survival analysis: use of the additive hazards model. Epidemiology, 23(5), 733-737.
- Strohmaier, S., Rysland, K., Hoff, R., Borgan, ., Pedersen, T. R., & Aalen, O. O. (2015). Dynamic path analysisa useful tool to investigate mediation processes in clinical survival trials. Statistics in medicine, 34(29), 3866-3887.
- Tein, J. Y., & MacKinnon, D. P. (2003). Estimating mediated effects with survival data. In New developments in psychometrics (pp. 405-412). Springer Japan.
- VanderWeele, T. J. (2011). Causal mediation analysis with survival data. Epidemiology (Cambridge, Mass.), 22(4), 582.
- VanderWeele, T.J. (2015). Explanation in causal inference: methods for mediation and interaction. Oxford University Press.
- VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. Epidemiology, 20(6), 863-871.
- VanderWeele, T.J., & Vansteelandt, S. (2013). Mediation analysis with multiple mediators. Epidemiologic methods, 2(1), 95-115.
- VanderWeele, T. J. (2009). Marginal structural models for the estimation of direct and indirect effects. Epidemiology, 20(1), 18-26.
- VanderWeele, T. J. (2014). A unification of mediation and interaction: a 4-way decomposition. Epidemiology, 25(5), 749-761.
- Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposuremediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods, 18(2), 137.

Few references on causality in health disparity research

- Naimi AI et al. (2015) Mediation analysis for health disparities research. American Journal of Epidemiology
- Vanderweele TJ, Robinson WR. (2014) On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology
- Glymour C, Glymour MR. (2014) Commentary: race and sex are causes. Epidemiology

References from motivating examples

- Arpadi SM. (2000) Growth failure in children with HIV infection. J Acquir Immune Defic Syndr. 2000 Oct 1;25 Suppl 1:S37-42.
- Clark SL et al. (2011) Analysis of efficacy and side effects in CATIE demonstrates drug response subgroups and potential for personalized medicine. Schizophr Res. 2011 Nov;132(2-3):114-20
- Huang T. et al. (2014) 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. Environ Health. 2014 Feb 5;13(1):6
- Jackson WP. (1971) Racial and geographic factors in diabetes. Lancet
- James P. et al. (2014) Do minority and poor neighbourhoods have higher access to fast-food restaurants in the United States? Health Place. 2014 Sep;29:10-7
- James-Todd T., et al. (2012) Urinary phthalate metabolite concentrations and diabetes among women in the National Health and Nutrition Examination Survey (NHANES) 2001-2008. Environ Health Perspect. 2012 Sep;120(9):1307-13
- Lieberman JA et al. (2005)Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med. 2005 Sep 22;353(12):1209-23
- Tanner JM (1973) Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. Arch Dis Child. 1976 Mar;51(3):170-9.
- Williams PL et al., (2013) Pubertal onset in children with perinatal HIV infection in the era of combination antiretroviral treatment. AIDS. 2013 Jul 31;27(12):1959-70
- Zota A.R. et al. (2016). Recent Fast Food Consumption and Bisphenol A and Phthalates Exposures among the U.S. Population in NHANES, 2003-2010.Environ Health Perspect. 2016 Oct;124(10):1521-1528