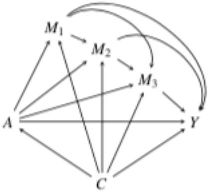
Application of a novel method for mediation analysis in the exposome context

Daniela Zugna

Exposome Data Challenge Event
29 April 2021

Multiple mediation analysis



- Mediation analysis aims at estimating to what extent the effect of an exposure on an outcome is explained by a set of mediators on the causal pathway between the exposure and the outcome.
- Counterfactual-based definitions of natural direct and indirect effects decompose the total effect of A on Y into the natural indirect effect, i.e. the effect explained by the mediators jointly, and the natural direct effect, i.e. the effect unexplained by the mediators.
- Extensions of natural effects to the multiple mediator setting are complicated by the complex (possibly unknown) confounding patterns among the different mediators.
- Most methods are applicable when the mediators can be causally ordered or the mediators do not exert causal effects on each others.

Novel estimation strategy

Received: 20 January 2020 | Revised: 28 October 2020 | Accepted: 11 November 2020
DOI: 10.1111/biom.13402

Biometrics WILEY

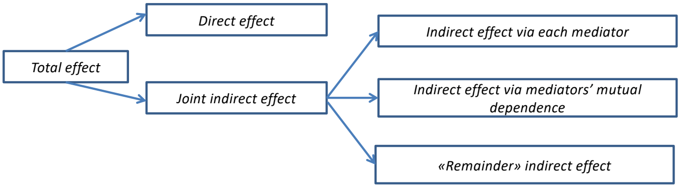
BIOMETRIC METHODOLOGY

Nonlinear mediation analysis with high-dimensional mediators whose causal structure is unknown

Wen Wei Loh[1] | Beatrijs Moerkerke[1] | Tom Loeys[1] | Stijn Vansteelandt[2,3]

- Specification of a (marginal structural) mean model for the outcome, and no models for the mediators.
- Mediators and outcome can be continuous or noncontinuous.
- Mediators can concurrently causally affect one another.
- Mediators can share hidden common causes.
- Decomposition of the joint indirect effect into separate indirect effects via each mediator is invariant to the presumed ordering of the mediator indices.

Interventional effects



- Interventional effects consider population-level (stochastic) interventions that set the value of the mediator to a random draw from its (counterfactual) distribution that depend only on exposure.
- Interventional indirect effect expresses the change in the outcome that would be seen if the distribution of mediator were shifted from what it would be if all subjects were exposed to what it would otherwise be.
- Interventional direct and indirect effects are identified under weaker assumptions than natural effects, especially when there is post- exposure confounding of the mediator(s)-outcome relation(s).

Definitions

$Y_{0(\tilde{M}_{11}\dots\tilde{M}_{p1})}$: potential outcome under A=0 when the mediator values are set to a random draw from the **joint** counterfactual distribution under A=1.

$Y_{0\tilde{M}_{11}\dots\tilde{M}_{p1}}$: potential outcome under A=0 when the mediator values are set to a random draw from the **marginal** counterfactual distribution under A=1.

Average potential outcomes:

$$E(Y_{0(\tilde{M}_{11}\dots\tilde{M}_{p1})}) = \int E(Y_{0m_1\dots m_p})dF_{M_{11}\dots M_{p1}}(m_1\dots m_p)$$

$$E(Y_{0\tilde{M}_{11}\dots\tilde{M}_{p1}}) = \int E(Y_{0m_1\dots m_p})dF_{M_{11}}(m_1)\dots dF_{M_{p1}}(m_p)$$

The joint and all marginal distributions for the (counterfactual) mediators are unconditional on all observed baseline covariates L.

Definitions

$$\overbrace{g\{E[Y_{1(\tilde{M}_{11}\ldots\tilde{M}_{p1})}]\} - g\{E[Y_{0(\tilde{M}_{10}\ldots\tilde{M}_{p0})}]\}}^{Total\ \ effect} =$$

$$\overbrace{g\{E[Y_{1(\tilde{M}_{11}\ldots\tilde{M}_{p1})}]\} - g\{E[Y_{0(\tilde{M}_{11}\ldots\tilde{M}_{p1})}]\}}^{Joint\ \ direct\ \ effect} + \overbrace{g\{E[Y_{0(\tilde{M}_{11}\ldots\tilde{M}_{p1})}]\} - g\{E[Y_{0(\tilde{M}_{10}\ldots\tilde{M}_{p0})}]\}}^{Joint\ \ indirect\ \ effect}$$

$$\overbrace{g\{E[Y_{0\tilde{M}_{10}\ldots\tilde{M}_{s-10}\tilde{M}_{s1}\tilde{M}_{s+10}\ldots\tilde{M}_{p1}}]\} - g\{E[Y_{0\tilde{M}_{10}\ldots\tilde{M}_{s-10}\tilde{M}_{s0}\tilde{M}_{s+10}\tilde{M}_{p0}}]\}}^{Indirect\ \ effect\ \ via\ \ each\ \ mediator\ \ Ms}$$

The effect through a mediator is the combined causal effect along all paths from exposure to the mediator, then directly to the outcome.

Assumptions

1. The effect of A on Y is unconfounded conditional on L:

$$Y_{am_1 \ldots m_p} \perp A | L \qquad \forall a, m_1, \ldots, m_p$$

2. There is sufficient information observed in L so that the association between any of the mediators and Y is unconfounded within levels of the covariates L:

$$Y_{am_1 \ldots m_p} \perp \{M_1 \ldots M_p\} | (A = a, L) \qquad \forall a, m_1, \ldots, m_p$$

3. The effect of A on Ms is unconfounded conditional on L:

$$\{M_{1a} \ldots M_{pa}\} \perp A | L \qquad \forall a$$

Estimation of joint effects

1. Fit the outcome model, conditional on exposure, mediators and covariates to the observed data $(h(a, m_1, \ldots, m_p, L))$
2. Construct the duplicated data for each individual.

$a^{(0)}$	$a^{(1)}$	$\{\tilde{M}_{1a^{(1)}}$	$\cdots$	$\tilde{M}_{pa^{(1)}}\}$	$\mathrm{E}\left(Y_{a^{(0)}\tilde{M}_{1a^{(1)}}\cdots\tilde{M}_{pa^{(1)}}}\middle\vert L\right)$
0	0	$\{\tilde{M}_{10}$	$\cdots$	$\tilde{M}_{p0}\}$	$\hat{h}(0, \tilde{M}_{10}, \ldots, \tilde{M}_{p0}, L)$
0	1	$\{\tilde{M}_{11}$	$\cdots$	$\tilde{M}_{p1}\}$	$\hat{h}(0, \tilde{M}_{11}, \ldots, \tilde{M}_{p1}, L)$
1	1	$\{\tilde{M}_{11}$	$\cdots$	$\tilde{M}_{p1}\}$	$\hat{h}(1, \tilde{M}_{11}, \ldots, \tilde{M}_{p1}, L)$

3. For each row, randomly draw the counterfactuals mediators values $\{M_{1a^1} \ldots M_{p^1}\}$ jointly from the observed group $A = a^1$.
4. Impute the expected potential outcomes.
5. Repeat previous steps in order to account for the variability in the counterfactuals mediator values, thereby obtaining the Monte Carlo averaged imputed potential outcomes for each individual.
6. Calculate the average imputed potential outcome across all individuals and estimates the direct effect and joint indirect effect.

Estimation of indirect effects via each mediator

$a^{(0)}$	$a^{(1)}$	$a^{(2)}$	$\cdots$	$a^{(p-1)}$	$a^{(p)}$	$\tilde{M}_{1a^{(1)}}$	$\tilde{M}_{2a^{(2)}}$	$\cdots$	$\tilde{M}_{p-1,a^{(p-1)}}$	$\tilde{M}_{pa^{(p)}}$	$\mathrm{E}\left(Y_{a^{(0)}\tilde{M}_{1a^{(1)}}\cdots\tilde{M}_{pa^{(p)}}}\middle\vert L\right)$
0	0	0	$\cdots$	0	0	$\tilde{M}_{10}$	$\tilde{M}_{20}$	$\cdots$	$\tilde{M}_{p-1,0}$	$\tilde{M}_{p0}$	$\hat{h}(0,\tilde{M}_{10},\ldots,\tilde{M}_{p0},L)$
0	1	0	$\cdots$	0	0	$\tilde{M}_{11}$	$\tilde{M}_{20}$	$\cdots$	$\tilde{M}_{p-1,0}$	$\tilde{M}_{p0}$	$\hat{h}(0,\tilde{M}_{11},\ldots,\tilde{M}_{p0},L)$
0	0	1	$\cdots$	0	0	$\tilde{M}_{10}$	$\tilde{M}_{21}$	$\cdots$	$\tilde{M}_{p-1,0}$	$\tilde{M}_{p0}$	$\hat{h}(0,\tilde{M}_{10},\ldots,\tilde{M}_{p0},L)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0	0	0	$\cdots$	1	0	$\tilde{M}_{10}$	$\tilde{M}_{20}$	$\cdots$	$\tilde{M}_{p-1,1}$	$\tilde{M}_{p0}$	$\hat{h}(0,\tilde{M}_{10},\ldots,\tilde{M}_{p0},L)$
0	0	0	$\cdots$	0	1	$\tilde{M}_{10}$	$\tilde{M}_{20}$	$\cdots$	$\tilde{M}_{p-1,0}$	$\tilde{M}_{p1}$	$\hat{h}(0,\tilde{M}_{10},\ldots,\tilde{M}_{p1},L)$
0	1	1	$\cdots$	1	1	$\tilde{M}_{11}$	$\tilde{M}_{21}$	$\cdots$	$\tilde{M}_{p-1,1}$	$\tilde{M}_{p1}$	$\hat{h}(0,\tilde{M}_{11},\ldots,\tilde{M}_{p1},L)$

1. The interventional indirect effect via each Ms corresponds to the difference between the estimands in rows s+1 and 1.
2. For each row and for each column, randomly draw the counterfactual mediator values M_{sa^s} unconditionally on the other mediators from the observed group $A = a^s$.
3. Impute the expected potential outcomes.
4. Repeat previous steps in order to account for the variability in the counterfactuals mediator values, thereby obtaining the Monte Carlo averaged imputed potential outcomes for each individual.
5. Calculate the average imputed potential outcome across all individuals and estimates the direct effect and joint indirect effect.

Non-randomly assigned exposure

- When A is not randomly assigned, the counterfactual mediator values cannot be sampled by selecting mediator values at random within each observed exposure group.
- Sampling the mediator values with probability proportional to the inverse of the conditional probability of receiving the observed exposure given the confounders:
- The probabilities of the observed exposures may be estimated using predictions from a logistic regression model fitted to the observed data, with exposure as dependent variables and covariates as predictors.

High dimensional mediators

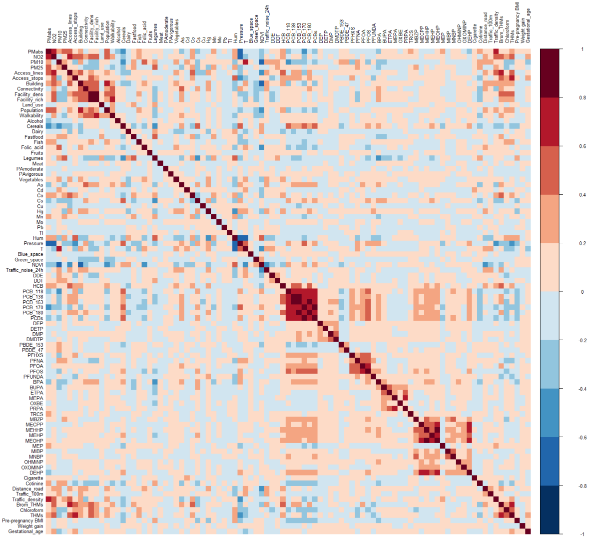
High-dimensional mediation settings (fewer observations than mediators and covariates):

- Indirect effects via each Ms:
  1. Outcome model: fit a penalized regression model for the outcome conditional on all mediators, exposure and covariates to the observed data (i.e. elastic net penalty). The coefficients for A and Ms should be unpenalized.
  2. Mediator model: fit a penalized regression model for Ms conditional on all other mediators, exposure and covariates to the observed data (i.e. elastic net penalty).
  3. The selected predictors in (1) and (2) are possible confounders of the Ms-Y relation and can be included in the new outcome model.
- Joint direct and indirect effect and mutual indirect effects:
  1. Outcome model: fit a separate penalized regression model for the outcome conditional on all mediators, treatment and covariates to the observed data.

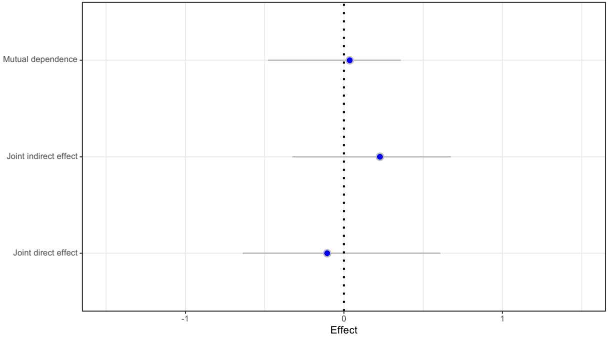
Motivating example

- Population under study (N=1301)
- Exposure variable: Socio economic position (low/medium vs high)
- Outcome: low birthweight (<2500 gr)
- Mediators (N=91) evaluated during pregnancy: air pollution, built environment, life style, metals, metereological, natural spaces, noise, organochlorines, organophosphate pesticides, phenols, phtalates, traffic, water DBPs, PBDE, PFAS
- Confounders: cohort, maternal age, ethnicity, parity, child's sex

Correlation matrix between mediators

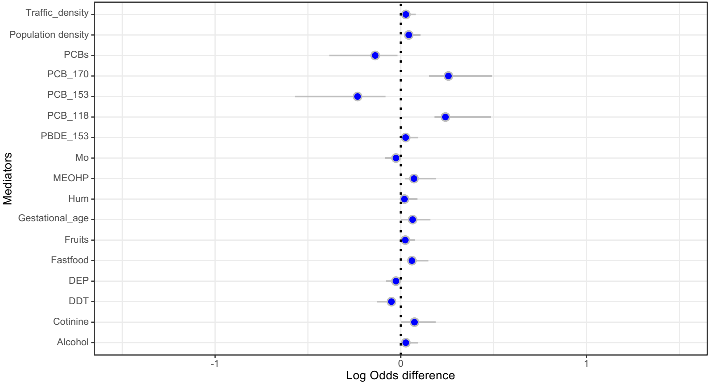


Joint direct and indirect effects

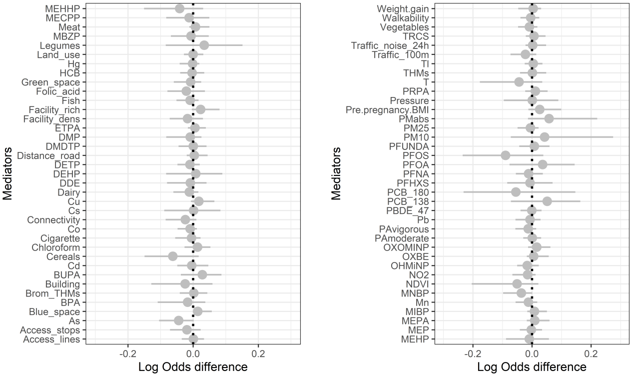


$$JDE = -0.10(-0.64, 0.61), JIE = 0.23(-0.32, 0.67), MIE = 0.04(-0.48, 0.36)$$

Estimated indirect effects via Ms



Estimated indirect effects via Ms



Area of future work

- Establishing theoretical properties to ensure valid inference of the proposed Monte Carlo-based estimators following mediator selection.
- Comparisons with high-dimensional mediation methods to select mediators.
- Assessing robustness to misspecification of the outcome model.