BIS 631 – Advanced Topics in Causal Inference Methods

Lecture 5
Understanding Causal Mechanisms: Causal Mediation Analysis

Laura Forastiere

Department of Biostatistics

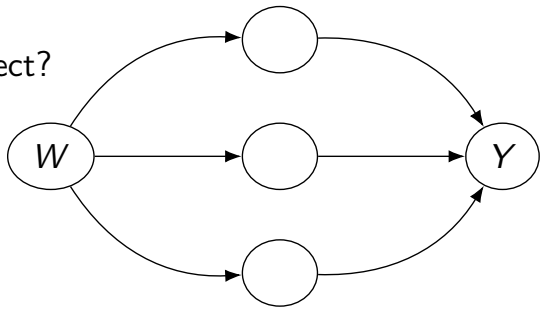
Yale School of Public Health

February 23, 2022

Concepts

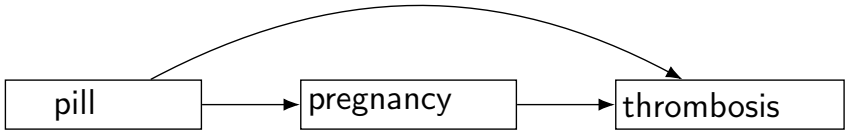
Mediation: The question of interest

- What is the mechanism that drives a particular causal effect?
  - How do we get from cause to effect?
  - What are the pathways?
  - What variables are in the pathway between the treatment and the final outcome?
  - What are the variables that are affected by the treatment and in turn have an effect on the final outcome, contributing to the total effect of the treatment on the outcome?



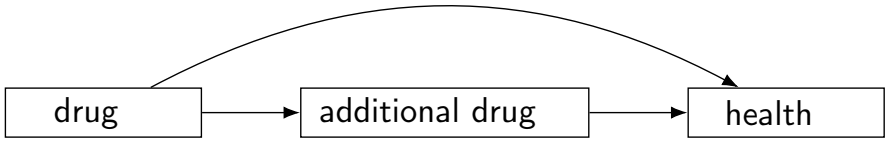
- Mediators: variables that are affected by the treatment and in turn have an effect on the final outcome, contributing to the total effect of the treatment on the outcome

Mediation: Examples



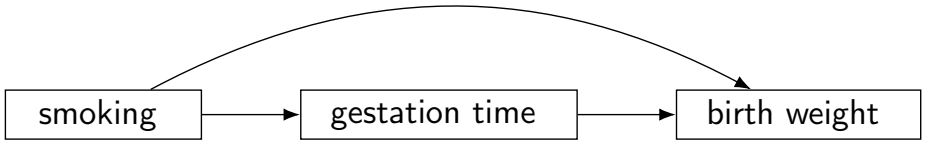
- To what extent is the causal effect of birth-control pill on thrombosis in women mediated by the effect of being on the contraceptive pill on pregnancy? (Pearl, 2001)

Mediation: Examples



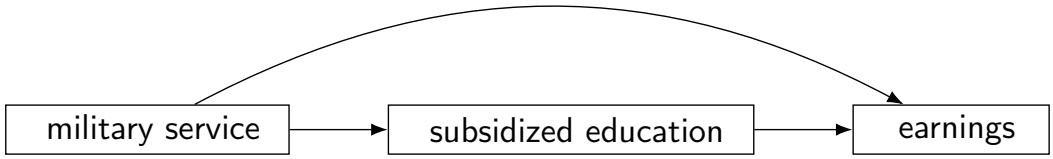
- To what extent is the causal effect of birth-control pill on thrombosis in women mediated by the effect of being on the contraceptive pill on pregnancy? (Pearl, 2001)
- To what extent is the causal effects of a new drug having side-effects mediated by the effect of taking additional medication to counter its side-effects? (Pearl, 2001)

Mediation: Examples



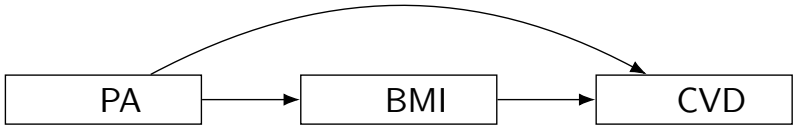
- To what extent is the causal effect of birth-control pill on thrombosis in women mediated by the effect of being on the contraceptive pill on pregnancy? (Pearl, 2001)
- To what extent is the causal effects of a new drug having side-effects mediated by the effect of taking additional medication to counter its side-effects? (Pearl, 2001)
- To what extent does smoking during pregnancy affect low birth weight through a shorter gestation time? (Flores & Flores-Lagunes, 2009)

Mediation: Examples



- To what extent is the causal effect of birth-control pill on thrombosis in women mediated by the effect of being on the contraceptive pill on pregnancy? (Pearl, 2001)
- To what extent is the causal effects of a new drug having side-effects mediated by the effect of taking additional medication to counter its side-effects? (Pearl, 2001)
- To what extent does smoking during pregnancy affect low birth weight through a shorter gestation time? (Flores & Flores-Lagunes, 2009)
- To what extent is the effect of military service on veterans' earnings channelled by subsidized higher education (Angrist & Chen, 2008)

Mediation: Examples



- To what extent is the causal effect of birth-control pill on thrombosis in women mediated by the effect of being on the contraceptive pill on pregnancy? (Pearl, 2001)
- To what extent is the causal effects of a new drug having side-effects mediated by the effect of taking additional medication to counter its side-effects? (Pearl, 2001)
- To what extent does smoking during pregnancy affect low birth weight through a shorter gestation time? (Flores & Flores-Lagunes, 2009)
- To what extent is the effect of military service on veterans' earnings channelled by subsidized higher education (Angrist & Chen, 2008)
- To what extent is the effect of physical activity (PA) on preventing cardiovascular diseases (CVD) mediated by a reduction in BMI, that is, to what extent can PA still prevent CVD even if it fails to prevent obesity (Sjöander et al., 2009; Schwartz et al., 2011)

Motivations for Studying Mediation

Scientific purpose:

- Scientific understanding and explanation
  E.g. To what extent is the causal effect of birth-control pill on thrombosis in women mediated by the effect of being on the contraceptive pill on pregnancy?
- Confirmation or refutation of theory
  E.g. Does physical activity affect CVD mainly by reducing BMI?

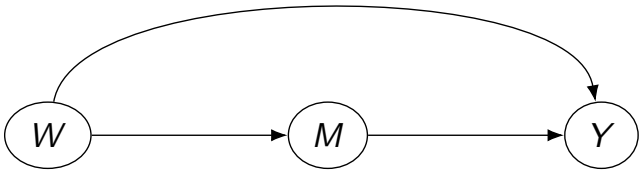
Intervention development: intervening on mediator or refining the intervention

- Limiting the detrimental effects of exposure by intervening on a mediator
  E.g. Can we eliminate the effects of antipsychotic medication on mortality by preventing the primary mechanism for mortality?
- Improving components of an intervention to target beneficial mechanisms
  E.g. Improving subsidized education for veterans might reduce the detrimental effect of military serving on earnings.
- Eliminating costly ineffective components of an intervention
  E.g. Does a CBT intervention improve depressive symptoms only through antidepressant use?
- Understanding why an intervention failed
  E.g. Did the intervention not affect the mediator, or does the mediator not affect the outcome, or was the direct effect in the opposite direction of the mediated effect??

Causal Mediation: Effect decomposition

- We cannot test for all causal mechanisms
- We can only test the decomposition of the total causal effect in few hypothesized pathways

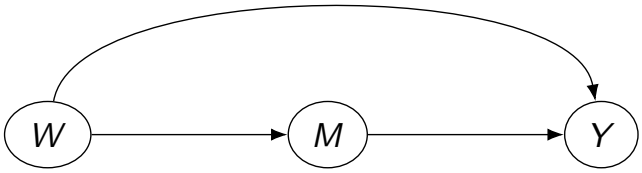
- W: treatment/exposure
- M: an intermediate post-treatment variable of interest, which we call MEDIATOR or intermediate variable
- Y: outcome



- In mediation, we are interested in assessing the the extent to which the effect of W on Y is mediated by an intermediate variable M and to what extent it is not

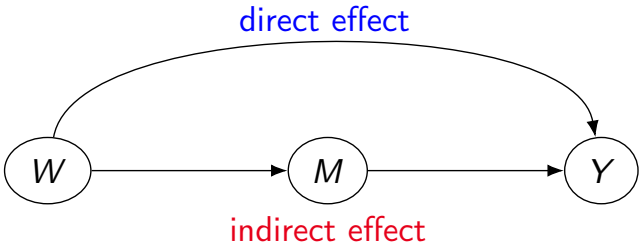
Causal Mediation: Effect decomposition

- We are interested in disentangling:
  - the extent to which the intervention W affects the outcome Y through a change in the mediator M which is turn affects the outcome Y
  - the extent to which the effect of W on Y is not mediated by M, that is, the part of the effect of W on Y that is not due to a change in M caused by the treatment, but can be due to other pathways



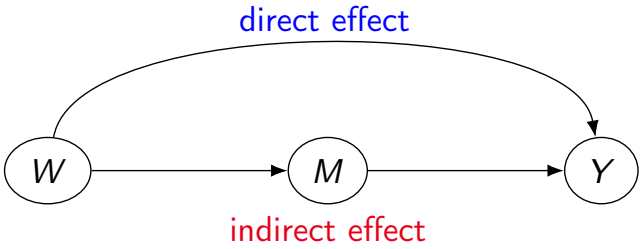
Causal Mediation: Effect decomposition

- We are interested in disentangling:
  - the extent to which the intervention W affects the outcome Y through a change in the mediator M which is turn affects the outcome $Y \longrightarrow$ MEDIATED or INDIRECT EFFECT
  - the extent to which the effect of W on Y is not mediated by M, that is, the part of the effect of W on Y that is not due to a change in M caused by the treatment, but can be due to other pathways $\longrightarrow$ NET or DIRECT EFFECT



Causal Mediation: Effect decomposition

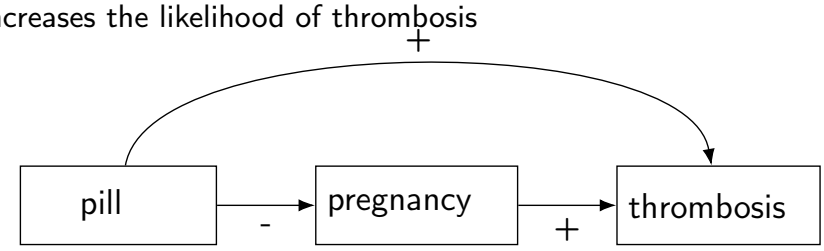
- We are interested in disentangling:
  - the extent to which the intervention W affects the outcome Y through a change in the mediator M which is turn affects the outcome $Y \longrightarrow$ MEDIATED or INDIRECT EFFECT
  - the extent to which the effect of W on Y is not mediated by M, that is, the part of the effect of W on Y that is not due to a change in M caused by the treatment, but can be due to other pathways $\longrightarrow$ NET or DIRECT EFFECT



- Causal Mediation:
  - Which causal estimands answer these research questions?
  - What assumptions allow identification of these estimands?
  - How do we estimate these causal estimands

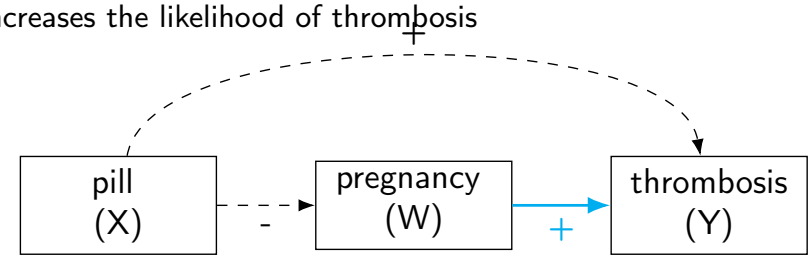
Covariate Adjustment vs Mediation

- Remember the Pill-Pregnancy-Thrombosis example
  - The consumption of birth control pill increases the likelihood of thrombosis
  - The consumption of birth control pill reduces the likelihood of pregnancy
  - Pregnancy increases the likelihood of thrombosis



Covariate Adjustment vs Mediation

- Remember the Pill-Pregnancy-Thrombosis example
  - The consumption of birth control pill increases the likelihood of thrombosis
  - The consumption of birth control pill reduces the likelihood of pregnancy
  - Pregnancy increases the likelihood of thrombosis

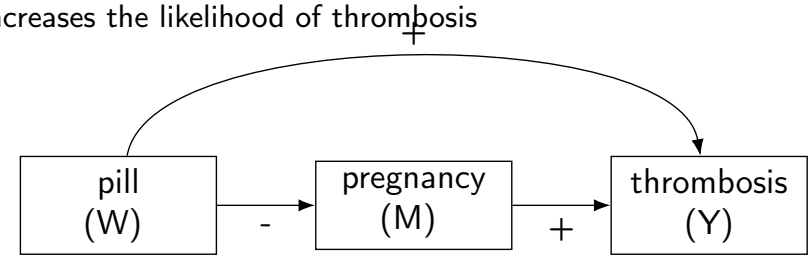


Covariate Adjustment:

- We are interested in the effect of pregnancy on thrombosis
- Given that the pill reduces pregnancy and affects thrombosis, the consumption of the pill can confound the effect of pregnancy on thrombosis
- Pregnant women are less likely to be on pill, which reduces their likelihood of thrombosis, while non-pregnant women are more likely to be on pill, with a greater risk of thrombosis. This might lead to an underestimated effect
- We must control for the consumption of the birth control pill

Covariate Adjustment vs Mediation

- Remember the Pill-Pregnancy-Thrombosis example
  - The consumption of birth control pill increases the likelihood of thrombosis
  - The consumption of birth control pill reduces the likelihood of pregnancy
  - Pregnancy increases the likelihood of thrombosis

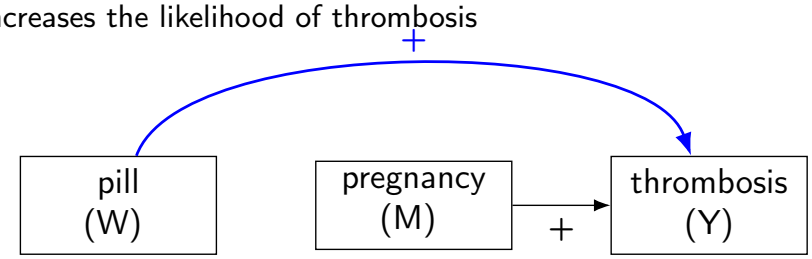


Mediation

- We are interested in the effect of the pill on thrombosis and in disentangling the effect through and not through pregnancy
- The total effect is a mixture of the physiological effect of the pill (through increased plasma fibrinogen) and the effect through a reduction in pregnancy

Covariate Adjustment vs Mediation

- Remember the Pill-Pregnancy-Thrombosis example
  - The consumption of birth control pill increases the likelihood of thrombosis
  - The consumption of birth control pill reduces the likelihood of pregnancy
  - Pregnancy increases the likelihood of thrombosis



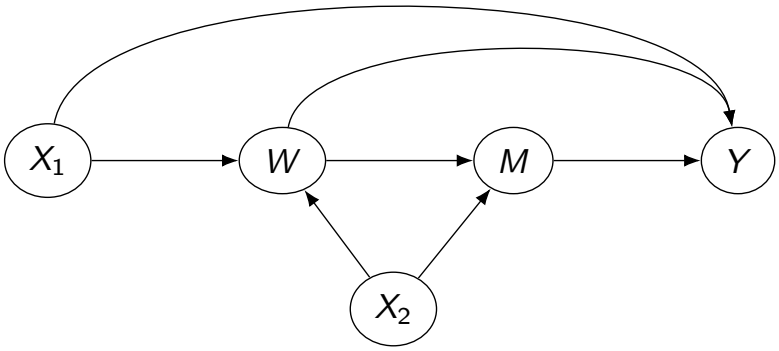
Mediation

- We are interested in the effect of the pill on thrombosis and in disentangling the effect through and not through pregnancy
- The total effect is a mixture of the physiological effect of the pill (through increased plasma fibrinogen) and the effect through a reduction in pregnancy
- If we wanted to estimate the physiological direct effect of the pill on thrombosis, we would need to untangle the mixture and isolate the net (or direct) effect
- This corresponds to asking the question of what would be the effect of the pill on thrombosis if we could somehow prevent it from reducing pregnancy, without altering the pill components

Traditional Approach to Mediation Analysis

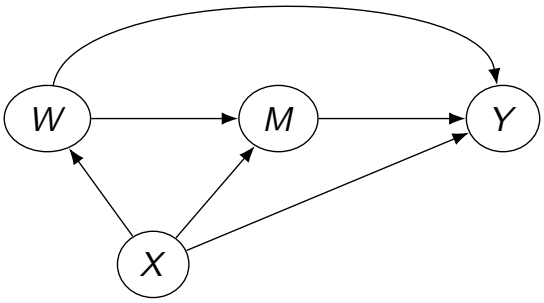
Traditional Approach to Mediation Analysis

(Baron and Kenny, 1986)



Traditional Approach to Mediation Analysis

(Baron and Kenny, 1986)



Traditional Approach to Mediation Analysis

(Baron and Kenny, 1986)

Difference Method:

- The difference method for mediation analysis, used with some frequency in much epidemiologic and social science research, consists first of controlling for the mediator to get the direct effect and subtract the direct effect from the total effect to get the indirect effect
- First regress the outcome Y on the exposure W and confounding factors X:

$$\mathbb{E}[Y_i|W_i = w, \mathbf{X}_i = \mathbf{x}] = \phi_0 + \phi_1 w + \phi_2^T \mathbf{x}$$

  ϕ_1 is the total effect
- Then we include the mediator M in the regression

$$\mathbb{E}[Y_i|W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x}$$

  θ_1 is the direct effect
- If the coefficients ϕ_1 and θ_1 differ then some of the effect is thought to be mediated and the following estimates are often used:

$$\text{Direct effect } = \theta_1$$

$$\text{Indirect effect } = \phi_1 - \theta_1$$

Traditional Approach to Mediation Analysis

(Caffo et al., 2008)

- Example: Caffo et al. (2008) consider the extent to which the effect of cumulative lead dose, W, on cognitive function, Y , is mediated by brain volumes, M.
- Controlling for age, education, smoking, and alcohol consumption, the authors obtained an estimate for the total effect of lead dose of 5.00 point decline (95% CI: -8.57, -1.42) in executive functioning cognitive test scores per 1μg/g increase in peak tibia lead exposure
- When also controlling for the mediator, brain volumes, the estimate of the "direct effect" of lead exposure becomes a decline of 3.79 points (95% CI:-7.40, -0.18)
- This gives an estimate of the indirect effect of 5.00 - 3.79 = 1.21 (P= 0.01)

Traditional Approach to Mediation Analysis

(Baron and Kenny, 1986)

Product Method:

- Another standard method, used more commonly in the social sciences is sometimes referred to as the "product method" (Baron and Kenny, 1986):
- First regress the mediator M on the exposure W and confounding factors X:

$$\mathbb{E}[M_i = m | W_i = w, \mathbf{X}_i = \mathbf{x}] = \beta_0 + \beta_1 w + \beta_2^T \mathbf{x}$$

- Then regress the outcome Y on the exposure W, the mediator M and confounding factors X:

$$\mathbb{E}[Y_i | W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x}$$

- The effects are estimated as follows

$$\text{Direct effect} = \theta_1$$

$$\text{Indirect effect} = \beta_1 \times \theta_2$$

Traditional Approach to Mediation Analysis

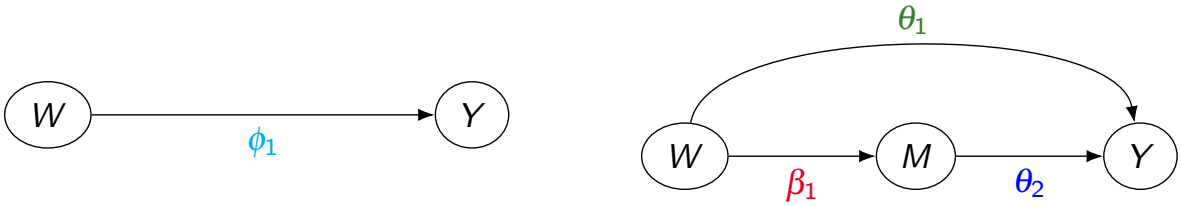
(Baron and Kenny, 1986)

- Given the mediator regression and the two outcome regressions

$$\mathbb{E}[Y_i|W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \phi_0 + \phi_1 w + \phi_2^T \mathbf{x} \quad \text{Difference method}$$

$$\mathbb{E}[M_i = m|W_i = w, \mathbf{X}_i = \mathbf{x}] = \beta_0 + \beta_1 w + \beta_2^T \mathbf{x} \quad \text{Product method}$$

$$\mathbb{E}[Y_i|W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x} \quad \text{Difference + Product methods}$$



- The effects are estimated as follows in the two methods

$$\text{Direct effect} = \theta_1$$

$$\text{Difference Method: Indirect effect} = \phi_1 - \theta_1$$

$$\text{Product Method: Indirect effect} = \beta_1 \times \theta_2$$

Traditional Approach to Mediation Analysis

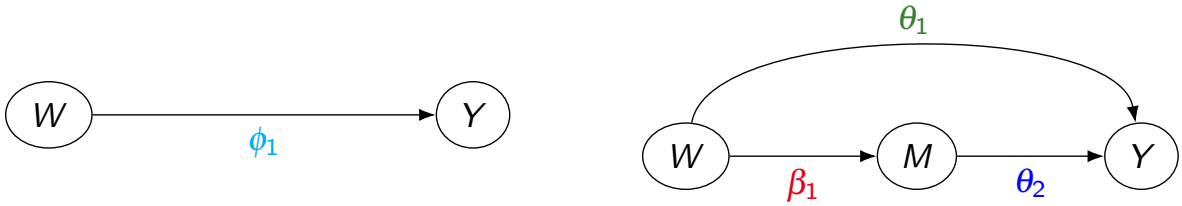
(Baron and Kenny, 1986)

- Given the mediator regression and the two outcome regressions

$$\mathbb{E}[Y_i|W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \phi_0 + \phi_1 w + \phi_2^T \mathbf{x} \quad \text{Difference method}$$

$$\mathbb{E}[M_i = m|W_i = w, \mathbf{X}_i = \mathbf{x}] = \beta_0 + \beta_1 w + \beta_2^T \mathbf{x} \quad \text{Product method}$$

$$\mathbb{E}[Y_i|W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x} \quad \text{Difference + Product methods}$$



- The effects are estimated as follows in the two methods

$$\text{Direct effect } = \theta_1$$

$$\text{Difference Method: Indirect effect } = \phi_1 - \theta_1$$

$$\text{Product Method: Indirect effect } = \beta_1 \times \theta_2$$

- Under joint normality product and difference estimators coincide (McKinnon, 2005)

Traditional Approach to Mediation Analysis

(Baron and Kenny, 1986)

The standard approaches to mediation analysis (difference method and product method), based on just including the mediator in the regression, are subject to three important limitations:

PROBLEM 1: Definition of direct and indirect effects is model driven. We need a causal definition that is not tied to a model

PROBLEM 2: Baron and Kenny (1986) claim that the difference and product methods are unbiased for the direct and indirect effects if we control for all exposure-outcome confounders. More detailed identifying assumptions are needed, also considering confounders of the mediator-outcome relationship

PROBLEM 3: They presuppose no interactions between the effects of the exposure and the mediator on the outcome, and presuppose the absence of non-linearities

PROBLEM 3: Exposure-Mediator Interaction and Non-lineairities

The standard approach presupposes no interactions between the effects of the exposure and the mediator on the outcome:

$$\mathbb{E}[Y_i|W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_3 \mathbf{x}$$

- This can lead to invalid conclusions
- To see why, suppose M is binary and the true model is:

$$\mathbb{E}[Y_i|W_i = w, M_i = m\mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_3 wm + \theta_4^T \mathbf{x}$$

  with $\theta_1 = 0.5$ and $\theta_3 = -1.0$, so that the sign of the direct effect of the exposure is different when the mediator is absent (+0.5) versus present (-0.5).
- If we fit the model without θ_3 we might estimate a value of θ_1 close to 0 because of averaging
- Then, under both the 'difference method' and 'product method' we would conclude that almost all of the effect of the exposure on the outcome is mediated
- But this would be completely an artifact of the interaction term $\theta_3 wm$ that was ignored
- Even in cases in which W had no effect on M (no mediation), but there is an interaction between the effects of W and M on Y, if we neglect this interaction, we might get to the same conclusion that almost all of the effect of the exposure on the outcome was mediated by M!

PROBLEM 3: Exposure-Mediator Interaction and Non-lineairities

- Even if we include an interaction term in the regression model the usual measures of direct and indirect effect break down
- Product Method and Difference Method do not yield the same result when exposure-mediator interaction is present. It is unclear how to handle the interaction coefficient
- Product Method and Difference Method estimators for direct and indirect effects are not robust to misspecification when mediator and/or outcome are binary

Potential Outcome Framework for Causal Mediation Analysis

Potential Outcome Framework for Causal Mediation Analysis

In what follows we will:

1. Consider the causal ("counterfactual") definitions of direct and indirect effects for mediation analysis
2. Introduce and discuss the no unmeasured confounding assumptions required for identification
3. Describe estimators for these counterfactual direct and indirect effect quantities (e.g. VanderWeele and Vansteelandt 2009, 2010; Valeri and VanderWeele 2013; Imai 2010; Lange et al. 2012)

Potential Outcomes

- Let W be a binary treatment and M a binary, discrete or continuous mediator

Potential Outcomes

- Let W be a binary treatment and M a binary, discrete or continuous mediator
- Mediators as post-treatment variables have potential 'outcomes':
  - $M_i(w)$: the potential value that the mediator would take for unit i when intervening to set the treatment W to w.

Potential Outcomes

- Let W be a binary treatment and M a binary, discrete or continuous mediator
- Mediators as post-treatment variables have potential 'outcomes':
  - $M_i(w)$: the potential value that the mediator would take for unit i when intervening to set the treatment W to w.
- Potential outcomes can be indexed by both the treatment and the mediator:
  - $Y_i(w, m)$: the potential value that the outcome would take for individual i when intervening to set the treatment W_i to w and the mediator M_i to m
  - This corresponds to a hypothetical intervention that we conceive on both the treatment and the mediator, regarded as an additional treatment

Potential Outcomes

- Let W be a binary treatment and M a binary, discrete or continuous mediator
- Mediators as post-treatment variables have potential 'outcomes':
  - $M_i(w)$: the potential value that the mediator would take for unit i when intervening to set the treatment W to w.
- Potential outcomes can be indexed by both the treatment and the mediator:
  - $Y_i(w, m)$: the potential value that the outcome would take for individual i when intervening to set the treatment W_i to w and the mediator M_i to m
  - This corresponds to a hypothetical intervention that we conceive on both the treatment and the mediator, regarded as an additional treatment
  - We can conceive an intervention that sets the mediator not to a fixed value but to specific value for each unit, in particular to $M_i(w^\star)$
  - $Y_i(w, M_i(w^\star))$: the potential value that the outcome would take for individual i when intervening to set the treatment W_i to w and the mediator M_i to $M_i(w^\star)$

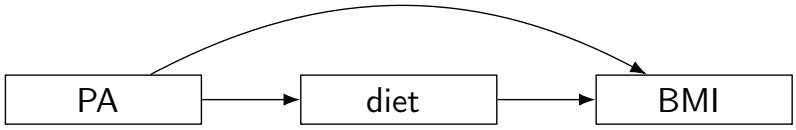
Potential Outcomes

- Let W be a binary treatment and M a binary, discrete or continuous mediator
- Mediators as post-treatment variables have potential 'outcomes':
  - $M_i(w)$: the potential value that the mediator would take for unit i when intervening to set the treatment W to w.
- Potential outcomes can be indexed by both the treatment and the mediator:
  - $Y_i(w, m)$: the potential value that the outcome would take for individual i when intervening to set the treatment W_i to w and the mediator M_i to m
  - This corresponds to a hypothetical intervention that we conceive on both the treatment and the mediator, regarded as an additional treatment
  - We can conceive an intervention that sets the mediator not to a fixed value but to specific value for each unit, in particular to $M_i(w^\star)$
  - $Y_i(w, M_i(w^\star))$: the potential value that the outcome would take for individual i when intervening to set the treatment W_i to w and the mediator M_i to $M_i(w^\star)$
  - $Y_i(w) = Y_i(w, M_i(w))$

Potential Outcomes

- Let W be a binary treatment and M a binary, discrete or continuous mediator
- Mediators as post-treatment variables have potential 'outcomes':
  - $M_i(w)$: the potential value that the mediator would take for unit i when intervening to set the treatment W to w.
- Potential outcomes can be indexed by both the treatment and the mediator:
  - $Y_i(w, m)$: the potential value that the outcome would take for individual i when intervening to set the treatment W_i to w and the mediator M_i to m
  - This corresponds to a hypothetical intervention that we conceive on both the treatment and the mediator, regarded as an additional treatment
  - We can conceive an intervention that sets the mediator not to a fixed value but to specific value for each unit, in particular to $M_i(w^\star)$
  - $Y_i(w, M_i(w^\star))$: the potential value that the outcome would take for individual i when intervening to set the treatment W_i to w and the mediator M_i to $M_i(w^\star)$
  - $Y_i(w) = Y_i(w, M_i(w))$
- Note: this notation is valid only under no-interference for both the mediator and the outcome
- The consistency assumption connects the potential outcomes to the observed outcomes:
  - W_i: observed treatment taken by unit i
  - $M_i^{obs} = M_i = M_i(W_i)$
  - $Y_i^{obs} = Y_i = Y_i(W_i, M_i(W_i)) = Y_i(W_i)$

Potential Outcomes Example



- W_i is exercise, physical activity, M_i is diet, Y_i is BMI
- The idea is that PA affects BMI because exercise reduces your weight, but PA will also affect your diet because exercise might make you hungry and eat more and this might reduce the effect of PA on BMI
- Take two values of W and M: w= 'run 10 km/day' and m='eat 1500 kcals/day'
- $Y_i(w, m)$ is the weight you would have if we forced you to run 10 km/day and eat 1500 kcals/day
- $M_i(w)$ is the diet you would have if we forced you to run 10 km/day, e.g., you would eat more, say, 2500 kcals/day.
- $Y_i(w) = Y_i(w, M_i(w))$: is the weight you would have if we forced you to run 10 km/day but didn't intervene on your diet, letting you eat what you would eat in that situation (if you ran 10 km/day), i.e., 2500 kcals a day.

Causal Estimands

Total Effect

(Robins and Greenland 1992; Pearl 2001)

- (Individual) Total Effect: the effect of receiving the treatment vs control for unit i, without intervening on the mediator, i.e., letting the mediator take the value it would take under the corresponding condition

$$TE_i = Y_i(1) - Y_i(0) = Y_i(1, M_i(1)) - Y_i(0, M_i(0))$$

- (Average) Total Effect: the average effect of receiving the treatment vs control:

$$TE = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1, M_i(1)) - Y_i(0, M_i(0))]$$

- The total effect is the effect of the treatment through all causal pathways, including M.
- TE corresponds to the ATE (or τ), estimated without taking the mediators into account
- In mediation analysis, given a mediator M, we wish to disentangle this (average) total effect into the direct effect not through the mediator and the indirect effect through the mediator

Controlled Direct Effects

(Robins and Greenland 1992; Pearl 2001)

- (Individual) Controlled Direct Effect: Effect of W on Y intervening to fix the mediator to a specific value m

$$CDE_i(m) = Y_i(1, m) - Y_i(0, m) \qquad w = 0, 1$$

E.g. The effect for unit i of running 10 km/day (vs not running) if we fixed his/her diet to $m = 1500 kcals/day$.

CDE is potentially different for each level of the mediator m.

Controlled Direct Effects

(Robins and Greenland 1992; Pearl 2001)

- (Individual) Controlled Direct Effect: Effect of W on Y intervening to fix the mediator to a specific value m

$$CDE_i(m) = Y_i(1, m) - Y_i(0, m) \qquad w = 0, 1$$

  E.g. The effect for unit i of running 10 km/day (vs not running) if we fixed his/her diet to $m = 1500 kcals/day$.

  CDE is potentially different for each level of the mediator m.

- We could define a similar 'controlled' indirect effect, given by the contrast between the potential outcomes if we intervened on the mediator and set it to m vs m', while keeping the exposure equal to w.

$$CIE_i(w) = Y_i(w, m) - Y_i(w, m') \qquad w = 0, 1$$

  This looks like the effect of the mediator M (seen as the treatment). Because the treatment W comes before the mediator M and affects M and Y, it can be seen as a confounder to control for. For this reason, this type of causal effect is not typically considered in mediation analysis.

- Also: TE $\neq$ CDE + CIE
- The controlled direct effect is still of policy interest

A Priori or Cross-World Conterfactuals

(Robins and Greenland 1992; Pearl 2001)

- The PO framework to mediation analysis has defined two causal effects, natural direct and indirect effects, that have can answer the causal questions of interest and have the benefit of decomposing the total effect.
- They are based on the definition of potential outcomes of the form:
  - $Y_i(w, M_i(w^\star))$: the potential outcome for each unit if we intervened to set the treatment to the value w and we intervened to fix the mediator to the value it would have taken if the treatment had been set to $w^\star$
  - When the treatment is binary we have: $Y_i(0, M_i(0))$, $Y_i(0, M_i(1))$, $Y_i(1, M_i(0))$ and $Y_i(1, M_i(1))$
  - When $w \neq w^\star$, they are called 'a priori' conterfactuals or 'cross-world' counterfactuals, because they are never observable, we need to see each unit in two different states of the world simultaneously
- These a-priori counterfactual are controversial and have led some researchers to dismiss mediation analysis altogether

Natural Direct and Indirect Effects

(Robins and Greenland 1992; Pearl 2001)

- (Individual) Natural Direct Effect: Individual effect of W (of begin exposed to treatment vs control) on Y, intervening to fix the mediator to the value it would have taken if W had been set to w

$$NDE_i(w) = Y_i(1, M_i(w)) - Y_i(0, M_i(w)) \qquad w = 0, 1$$

- (Individual) Natural Indirect Effect: Individual effect on the outcome Y of intervening to set the mediator to what it would have been if W were $w = 1$ in contrast to what it would have been if W were $w = 0$, while intervening to fix the treatment W to w

$$NIE_i(w) = Y_i(w, M_i(1)) - Y_i(w, M_i(0)) \qquad w = 0, 1$$

Natural Direct and Indirect Effects

(Robins and Greenland 1992; Pearl 2001)

- (Individual) Natural Direct Effect: Individual effect of W (of begin exposed to treatment vs control) on Y, intervening to fix the mediator to the value it would have taken if W had been set to w

$$NDE_i(w) = Y_i(1, M_i(w)) - Y_i(0, M_i(w)) \qquad w = 0, 1$$

  E.g. $NDE_i(0) = Y_i(1, M_i(0)) - Y_i(0, M_i(0))$: Effect on BMI for unit i of running 10 km/day vs not to running, when diet is kept to what it would be if he/she didn't run at all. In this way we are subtracting from the total effect of PA the effect through a change in diet which could reduce the effect

- (Individual) Natural Indirect Effect: Individual effect on the outcome Y of intervening to set the mediator to what it would have been if W were $w = 1$ in contrast to what it would have been if W were $w = 0$, while intervening to fix the treatment W to w

$$NIE_i(w) = Y_i(w, M_i(1)) - Y_i(w, M_i(0)) \qquad w = 0, 1$$

  E.g. $NIE_i(1) = Y_i(1, M_i(1)) - Y_i(1, M_i(0))$: Compares the potential BMI if we forced one to run 10 km/day and let him eat what he would eat when running 10 km/day to the BMI if we still forced one to run 10 km/day but we kept his diet to what it would be if he didn't run at all.

Average Direct and Indirect Effects

(Robins and Greenland 1992; Pearl 2001)

We have considered individual effects but we will generally consider average effects as these can be identified and estimated:

- (Average) Controlled Direct Effect: Average effect of W on Y intervening to fix the mediator to a specific value m

$$CDE(m) = \mathbb{E}[Y_i(1, m) - Y_i(0, m))] \qquad w = 0, 1$$

- (Average) Natural Direct Effect: Average effect of W (of moving from control to treatment) on Y intervening to fix the mediator to the value it would have taken if W had been set to w

$$NDE(w) = \mathbb{E}[Y_i(1, M_i(w)) - Y_i(0, M_i(w))] \qquad w = 0, 1$$

- (Average) Natural Indirect Effect: Average effect on the outcome Y of intervening to set the mediator to what it would have been if W were $w = 1$ in contrast to what it would have been if W were $w = 0$

$$NIE(w) = \mathbb{E}[Y_i(w, M_i(1)) - Y_i(w, M_i(0))] \qquad w = 0, 1$$

- These definitions do not rule out interactions between the effects of the exposure and the mediator on the outcome $\Rightarrow CDE(m) \neq CDE(m')$, $NDE(0) \neq NDE(1)$, and $NIE(0) \neq NIE(1)$

Controlled and Natural Direct Effects

(Robins and Greenland 1992; Pearl 2001)

- The natural direct effect compares treatment level $W = 1$ to $W = 0$ with the mediator set to what it would have been if the treatment had been kept fixed
- One way to think about the natural direct effect is that it is equal to the controlled direct effect $CDE(m)$ with m fixed to $M_i(0)$
- In general, the NDE will be different from the CDE.
- The difference between the CDE and the NDE is that in the CDE the mediator is fixed at a hypothetical value, not the potential value corresponding to some treatment, and this value is the same for all units
  - CDE(m): set M_i to m for all units
  - NDE(w): set M_i to $M_i(w)$ for all units
- CDE is identified under weaker conditions than the NDE.
- If there is no interaction between the exposure and the mediator, $CDE(m) = NDE(w)$, for every m and w

Effect Decomposition

(Robins and Greenland 1992; Pearl 2001)

- The total causal effect and the natural indirect and direct causal effects are related
- TE = NDE + NIE

$$\begin{aligned} TE &= NDE(w) + NIE(1-w) \\ &= NDE(0) + NIE(1) = NDE(1) + NIE(0) \end{aligned}$$

Proof.

$$\begin{aligned} TE = \mathbb{E}[Y_i(1) - Y_i(0)] &= \mathbb{E}[Y_i(1, M_i(1)) - Y_i(0, M_i(0))] \\ &= \mathbb{E}[Y_i(1, M_i(1))] + \mathbb{E}[Y_i(1, M_i(0))] - \mathbb{E}[Y_i(1, M_i(0))] - \mathbb{E}[Y_i(0, M_i(0))] \\ &= NIE(1) + NDE(0) \\ &= \mathbb{E}[Y_i(1, M_i(1))] + \mathbb{E}[Y_i(0, M_i(1))] - \mathbb{E}[Y_i(0, M_i(1))] - \mathbb{E}[Y_i(0, M_i(0))] \\ &= NDE(1) + NIE(0) \end{aligned}$$

□

- The fact that we can decompose the total effect of treatment into the sum of a direct and indirect effect if very important to social science researchers.

Pure and Total Direct and Indirect Effects

(Robins and Greenland 1992; Pearl 2001)

- NDE(0) and NIE(0) and called 'pure' (natural) direct and indirect effects
- NDE(1) and NIE(1) and called 'total' (natural) direct and indirect effects
- $NDE(0) = \mathbb{E}[Y_i(1, M_i(0)) - Y_i(0, M_i(0))]$ and $NIE(1) = \mathbb{E}[Y_i(1, M_i(1)) - Y_i(1, M_i(0))]$ are the most commonly used to describe mediation

Example

Let W be a binary treatment and M a binary mediator

Individual	$M_i(0)$	$M_i(1)$	$Y_i(0,0)$	$Y_i(1,0)$	$Y_i(0,1)$	$Y_i(1,1)$
1	0	1	0	1	0	1
2	1	1	0	1	0	0
3	0	1	0	0	0	1

Example

Let W be a binary treatment and M a binary mediator

Individual	$M_i(0)$	$M_i(1)$	$Y_i(0,0)$	$Y_i(1,0)$	$Y_i(0,1)$	$Y_i(1,1)$
1	0	1	0	1	0	1
2	1	1	0	1	0	0
3	0	1	0	0	0	1

- For individual 1 (No indirect effect, Total effect is completely direct, No exposure-mediator interaction):

$$TE_1 = Y_1(1) - Y_1(0) = Y_1(1, M_1(1)) - Y_1(0, M_1(0) = Y_1(1,1) - Y_1(0,0) = 1 - 0 = 1$$
$$CDE_1(m = 0) = Y_1(1,0) - Y_1(0,0) = 1 - 0 = 1$$
$$CDE_1(m = 1) = Y_1(1,1) - Y_1(0,1) = 1 - 0 = 1$$
$$NDE_1(0) = Y_1(1, M_1(0)) - Y_1(0, M_1(0)) = Y_1(1,0) - Y_1(0,0) = 1 - 0 = 1$$
$$NIE_1(1) = Y_1(1, M_1(1)) - Y_1(1, M_1(0)) = Y_1(1,1) - Y_1(1,0) = 1 - 1 = 0$$

Example

Let W be a binary treatment and M a binary mediator

Individual	$M_i(0)$	$M_i(1)$	$Y_i(0,0)$	$Y_i(1,0)$	$Y_i(0,1)$	$Y_i(1,1)$
1	0	1	0	1	0	1
2	1	1	0	1	0	0
3	0	1	0	0	0	1

- For individual 2 (Total effect is zero, exposure-mediator interaction):

$$TE_2 = Y_2(1) - Y_2(0) = Y_2(1, M_2(1)) - Y_2(0, M_2(0) = Y_2(1,1) - Y_2(0,1) = 0 - 0 = 0$$
$$CDE_2(m = 0) = Y_2(1,0) - Y_2(0,0) = 1 - 0 = 1$$
$$CDE_2(m = 1) = Y_2(1,1) - Y_2(0,1) = 0 - 0 = 0$$
$$NDE_2(0) = Y_2(1, M_2(0)) - Y_2(0, M_2(0)) = Y_2(1,1) - Y_2(0,1) = 0 - 0 = 0$$
$$NIE_2(1) = Y_2(1, M_2(1)) - Y_2(1, M_2(0)) = Y_2(1,1) - Y_2(1,1) = 0 - 0 = 0$$

Example

Let W be a binary treatment and M a binary mediator

Individual	$M_i(0)$	$M_i(1)$	$Y_i(0,0)$	$Y_i(1,0)$	$Y_i(0,1)$	$Y_i(1,1)$
1	0	1	0	1	0	1
2	1	1	0	1	0	0
3	0	1	0	0	0	1

- For individual 3 (Total effect is completely mediated, exposure-mediator interaction):

$$TE_3 = Y_3(1) - Y_3(0) = Y_3(1, M_3(1)) - Y_3(0, M_3(0) = Y_3(1,1) - Y_3(0,0) = 1 - 0 = 1$$
$$CDE_3(m = 0) = Y_3(1,0) - Y_3(0,0) = 0 - 0 = 0$$
$$CDE_3(m = 1) = Y_3(1,1) - Y_3(0,1) = 1 - 0 = 1$$
$$NDE_3(0) = Y_3(1, M_3(0)) - Y_3(0, M_3(0)) = Y_3(1,0) - Y_3(0,0) = 0 - 0 = 0$$
$$NIE_3(1) = Y_3(1, M_3(1)) - Y_3(1, M_3(0)) = Y_3(1,1) - Y_3(1,0) = 1 - 0 = 1$$

Example

Let W be a binary treatment and M a binary mediator

Individual	$M_i(0)$	$M_i(1)$	$Y_i(0,0)$	$Y_i(1,0)$	$Y_i(0,1)$	$Y_i(1,1)$
1	0	1	0	1	0	1
2	1	1	0	1	0	0
3	0	1	0	0	0	1

- Average effects

$$TE = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1, M_i(1)) - Y_i(0, M_i(0)] = (1+0+1)/3 = 2/3$$
$$CDE(m=0) = \mathbb{E}[Y_i(1,0) - Y_i(0,0)] = (1+1+0)/3 = 2/3$$
$$CDE(m=1) = \mathbb{E}[Y_i(1,1) - Y_i(0,1)] = (1+0+1)/3 = 2/3$$
$$NDE(0) = \mathbb{E}[Y_i(1, M_i(0)) - Y_i(0, M_i(0))] = (1+0+0)/3 = 1/3$$
$$NIE(1) = \mathbb{E}[Y_i(1, M_i(1)) - Y_i(1, M_i(0))] = (0+0+1)/3 = 1/3$$

Average Direct and Indirect effects for General Treatment

Similar concepts apply to a general treatment W, comparing treatment levels $W = w$ to $W = w^\star$:

- Average Controlled Direct Effect: Average effect of treatment level w vs $w^\star$ on Y intervening to fix the mediator to a specific value m

$$CDE(w, w^\star; m) = \mathbb{E}[Y_i(w, m) - Y_i(w^\star, m)]$$

- Average Natural Direct Effect: Average effect of treatment level w vs $w^\star$ on Y intervening to fix the mediator to the value it would have taken if W had been set to $w^\star$

$$NDE(w, w^\star; w^\star) = \mathbb{E}[Y_i(w, M_i(w^\star)) - Y_i(w^\star, M_i(w^\star))]$$

- Average Natural Indirect Effect: Average effect on the outcome Y of intervening to set the mediator to what it would have been if W were set to w in contrast to what it would have been if W were to $w^\star$, while intervening to fix the treatment W to w

$$NIE(w, w^\star; w) = \mathbb{E}[Y_i(w, M_i(w)) - Y_i(w, M_i(w^\star))]$$

Wrong use of controlled direct effect

- The difference between the total effect and controlled direct effect is not a mediated effect
- It is not uncommon to see subtracting controlled direct effects from total effects to get the indirect effect

$$TE - CDE(m)$$

- This coincides with the natural indirect effect $NIE(w)$, for every w, only if there is no interaction between the exposure and the mediator, and hence $CDE(m) = NDE(w)$, for every w. If there is an interaction, the result is not a mediated effect
- From a policy perspective we could ask: How much of the effect would remain if we were to intervene on the mediator and fix it to some value. If there is an interaction, in general, what we get is not a mediated effect.

Proportion mediated measure

- We may often want to know how much of the effect is mediated; the proportion mediated measure is sometimes used for this
- If we let TE be the total effect, then the proportion mediated measure is defined by:

$$PM(w) = \frac{NIE(w)}{TE}$$

- When there is interaction this differs from the proportion eliminated by fixing the mediator

$$PE(m) = \frac{TE - CDE(m)}{TE}$$

- The proportion eliminated is more relevant for policy (fixing m)
- The proportion mediated more relevant for etiology (through a pathway)

Missing data issue in mediation analysis

- These counterfactual definitions of direct and indirect effects are theoretically appealing
- But they are counterfactual definitions and we are not in general able to observe all the counterfactuals needed to calculate these effects
- Consider binary W and M:

Individual	W_i	$M_i(0)$	$M_i(1)$	$Y_i(0,0)$	$Y_i(1,0)$	$Y_i(0,1)$	$Y_i(1,1)$
1	0	0	?	0	?	?	?
2	1	?	1	?	?	?	0
3	1	?	1	?	?	?	1

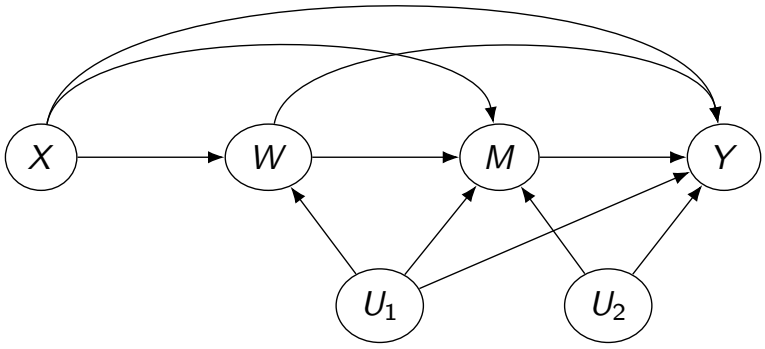
Identification

Identification of Direct and Indirect Effects: Sequential Ignorability

- What assumptions are needed to identify the direct and indirect effects?
- Imai et al. (2010) propose an assumption called **sequential ignorability (SI) assumption**
  - Similar to earlier assumptions from Pearl (2001).
  - Somewhat different from other uses of sequential ignorability by Robins and others.
  - Imai's version is now the most commonly used

$$(1) \quad \{\mathbf{Y_i}(\mathbf{w}^\star, \mathbf{m}), \mathbf{M_i}(\mathbf{w})\} \perp\!\!\!\perp \mathbf{W_i} | \mathbf{X_i} \qquad \forall \mathbf{w}^\star, \mathbf{w}, \mathbf{m}$$

$$(2) \quad \mathbf{Y_i}(\mathbf{w}^\star, \mathbf{m}) \perp\!\!\!\perp \mathbf{M_i}(\mathbf{w}) | \mathbf{X_i}, \mathbf{W_i} = \mathbf{w} \qquad \forall \mathbf{w}^\star, \mathbf{w}, \mathbf{m}$$



Identification of Direct and Indirect Effects: Sequential Ignorability

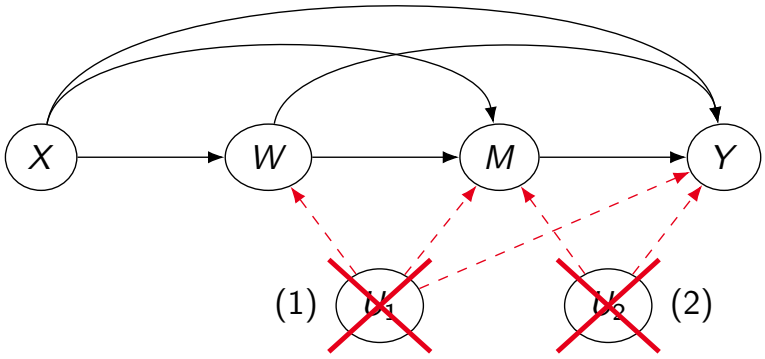
- What assumptions are needed to identify the direct and indirect effects?
- Imai et al. (2010) propose an assumption called **sequential ignorability (SI) assumption**
  - Similar to earlier assumptions from Pearl (2001).
  - Somewhat different from other uses of sequential ignorability by Robins and others.
  - Imai's version is now the most commonly used

$$(1) \quad \{\mathbf{Y_i}(\mathbf{w}^\star, \mathbf{m}), \mathbf{M_i}(\mathbf{w})\} \perp\!\!\!\perp \mathbf{W_i} | \mathbf{X_i} \qquad \forall \mathbf{w}^\star, \mathbf{w}, \mathbf{m}$$

No unmeasured exposure-outcome and exposure-mediator confounding given **X**

$$(2) \quad \mathbf{Y_i}(\mathbf{w}^\star, \mathbf{m}) \perp\!\!\!\perp \mathbf{M_i}(\mathbf{w}) | \mathbf{X_i}, \mathbf{W_i} = \mathbf{w} \qquad \forall \mathbf{w}^\star, \mathbf{w}, \mathbf{m}$$

No unmeasured mediator-outcome confounding given **X**



Identification of Direct and Indirect Effects: Sequential Ignorability - No exposure-induced mediator-outcome confounding

- The sequential ignorability assumption proposed by Imai et al (2010), by conditioning on the same set of covariates in both sub-assumptions, it implicitly makes an additional assumptions:

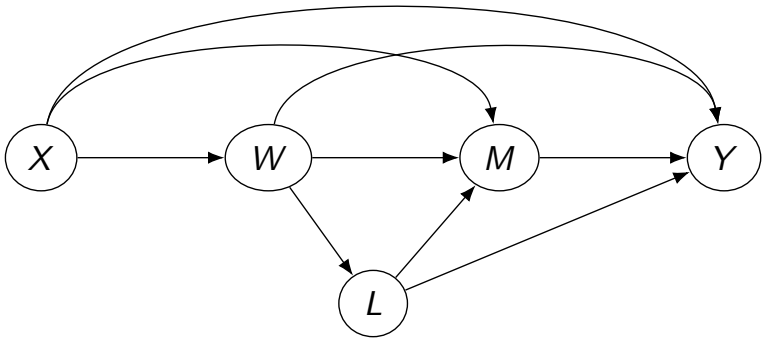
$$(1) \quad \{Y_i(w^\star, m), M_i(w)\} \perp\!\!\!\perp W_i | \mathbf{X}_i$$

No unmeasured exposure-outcome and exposure-mediator confounding given **X**

$$(2) \quad Y_i(w^\star, m) \perp\!\!\!\perp M_i(w) | \mathbf{X}_i, W_i = w$$

No unmeasured mediator-outcome confounding given **X**

No exposure-induced mediator-outcome confounding given **X**



Identification of Direct and Indirect Effects: Sequential Ignorability - No exposure-induced mediator-outcome confounding

- The sequential ignorability assumption proposed by Imai et al (2010), by conditioning on the same set of covariates in both sub-assumptions, it implicitly makes an additional assumptions:

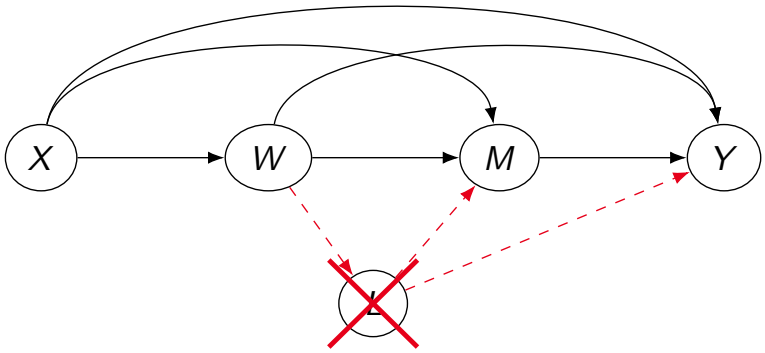
$$(1) \quad \{Y_i(w^\star, m), M_i(w)\} \perp\!\!\!\perp W_i | \mathbf{X}_i$$

No unmeasured exposure-outcome and exposure-mediator confounding given **X**

$$(2) \quad Y_i(w^\star, m) \perp\!\!\!\perp M_i(w) | \mathbf{X}_i, W_i = w$$

No unmeasured mediator-outcome confounding given **X**

No exposure-induced mediator-outcome confounding given **X**



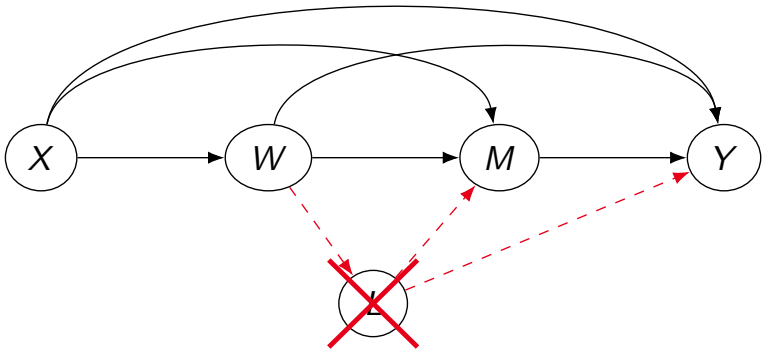
Identification of Direct and Indirect Effects: Sequential Ignorability - No exposure-induced mediator-outcome confounding

- The sequential ignorability assumption proposed by Imai et al (2010), by conditioning on the same set of covariates in both sub-assumptions, it implicitly makes an additional assumptions:

$$(1) \quad \{Y_i(w^\star, m), M_i(w)\} \perp\!\!\!\perp W_i | \mathbf{X}_i$$

$$(2) \qquad Y_i(w^\star, m) \perp\!\!\!\perp M_i(w) | \mathbf{X}_i \qquad \text{(Pearl, 2001)}$$

No exposure-induced mediator-outcome confounding given $\mathbf{X}$



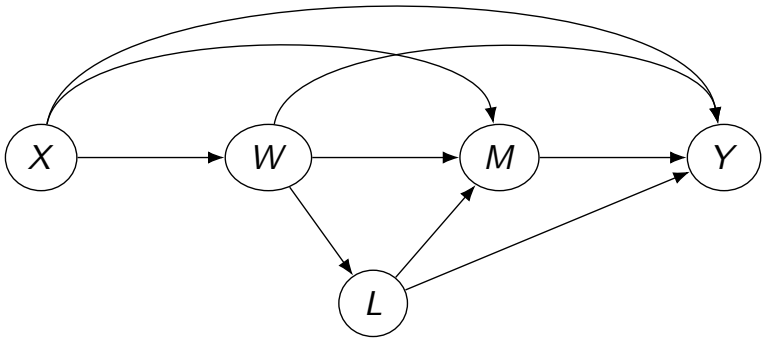
Identification of Direct and Indirect Effects: Sequential Ignorability - No exposure-induced mediator-outcome confounding

- The sequential ignorability assumption proposed by Imai et al (2010), by conditioning on the same set of covariates in both sub-assumptions, it implicitly makes an additional assumptions:

$$(1) \quad \{Y_i(w^\star, m), M_i(w)\} \perp\!\!\!\perp W_i | \mathbf{X}_i$$

$$(2) \qquad Y_i(w, m) \perp\!\!\!\perp M_i | \mathbf{X}_i, W_i = w, L_i$$

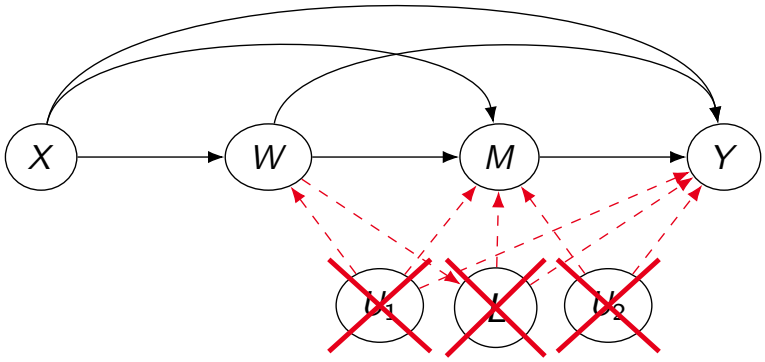
- Later, we will relax the no exposure-induced mediator-outcome confounding assumption



Identification of Direct and Indirect Effects: Sequential Ignorability - Alternative Version

- An alternative version of the sequential ignorability assumption can be written as follows:

$$(A) \quad M_i(w) \perp\!\!\!\perp W_i | \mathbf{X}_i \quad (B) \quad Y_i(w^\star, m) \perp\!\!\!\perp W_i | \mathbf{X}_i$$
$$(C) \quad Y_i(w, m) \perp\!\!\!\perp M_i | \mathbf{X}_i, W_i = w$$
$$(D) \quad Y_i(w^\star, m) \perp\!\!\!\perp M_i(w) | \mathbf{X}_i$$



Identification under Sequential Ignorability

(Imai et. al, 2010)

- Under the **sequential ignorability assumption** (and SUTVA), $CDE(m)$, for all m, $NDE(w)$ and $NIE(w)$, for all w, are identified. Specifically, we have the following identification results.

Controlled Direct Effect

$$CDE(m) = \mathbb{E}[Y_i(1,m) - Y_i(0,m)] = \mathbb{E}[Y_i|W_i = 1, M_i = m, \mathbf{X}_i] - \mathbb{E}[Y_i|W_i = 0, M_i = m, \mathbf{X}_i]$$

Proof.

$$\begin{aligned} CDE(m) &= \mathbb{E}[Y_i(1,m)] - \mathbb{E}[Y_i(0,m)] = \mathbb{E}_X[\mathbb{E}[Y_i(1,m)|\mathbf{X}_i]] - \mathbb{E}_X[\mathbb{E}[Y_i(0,m)|\mathbf{X}_i]] \\ &= \mathbb{E}[Y_i(1,m)|W_i = 1, \mathbf{X}_i] - \mathbb{E}[Y_i(0,m)|W_i = 0, \mathbf{X}_i] \\ &\text{(due to SI (1), i.e., ignorability of the treatment)} \\ &= \mathbb{E}[Y_i(1,m)|W_i = 1, M_i = m, \mathbf{X}_i] - \mathbb{E}[Y_i(0,m)|W_i = 0, M_i = m, \mathbf{X}_i] \\ &\text{(due to SI (2), i.e., ignorability of the mediator)} \\ &= \mathbb{E}[Y_i|W_i = 1, M_i = m, \mathbf{X}_i] - \mathbb{E}[Y_i|W_i = 0, M_i = m, \mathbf{X}_i] \\ &\text{(due to consistency)} \end{aligned}$$

- Note that identification of $CDE(m)$ does not require $M_i(w) \perp\!\!\!\perp W_i|\mathbf{X}_i$ (no unmeasured exposure-mediator confounding), which is part of SI (1) (SI (A)). It only requires SI (B) and (C) □

Identification under Sequential Ignorability

(Imai et. al, 2010)

- Under the **sequential ignorability assumption** (and SUTVA), $CDE(m)$, for all m, $NDE(w)$ and $NIE(w)$, for all w, are identified. Specifically, we have the following identification results.

Natural Direct Effect

$$\begin{aligned} NDE(w) &= \mathbb{E}[Y_i(1, M_i(w)) - Y_i(0, M_i(w))] \\ &= \sum_{\mathbf{x}} \sum_{m} \big(\mathbb{E}[Y_i | W_i = 1, M_i = m, \mathbf{X}_i = \mathbf{x}] - \mathbb{E}[Y_i | W_i = 0, M_i = m, \mathbf{X}_i = \mathbf{x}]\big) \\ &\qquad \times Pr(M_i = m | W_i = w, \mathbf{X}_i = \mathbf{x}) Pr(\mathbf{X}_i = \mathbf{x}) \end{aligned}$$

Natural Indirect Effect

$$\begin{aligned} NIE(w) &= \mathbb{E}[Y_i(w, M_i(1)) - Y_i(w, M_i(0))] \\ &= \sum_{\mathbf{x}} \sum_{m} \mathbb{E}[Y_i | W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] \\ &\qquad \times \big(Pr(M_i = m | W_i = 1, \mathbf{X}_i = \mathbf{x}) - Pr(M_i = m | W_i = 0, \mathbf{X}_i = \mathbf{x})\big) Pr(\mathbf{X}_i = \mathbf{x}) \end{aligned}$$

- These formulas are usually referred to as **g-formulas**.

Identification under Sequential Ignorability

(Imai et. al, 2010)

Proof. (1)

$$\mathbb{E}[Y_i(w, M_i(w^\star))|\mathbf{X}_i] = \sum_m \mathbb{E}[Y_i(1,m)|M_i(w^\star) = m, \mathbf{X}_i] Pr(M_i(w^\star) = m|\mathbf{X}_i)$$

(due to the law of total expectation)

$$= \sum_m \mathbb{E}[Y_i(w,m)|\mathbf{X}_i, W_i = w^\star, M_i(w^\star) = m] Pr(M_i(w^\star) = m|\mathbf{X}_i)$$

(due to SI (1,2)$\Rightarrow$ $Y_i(w,m) \perp\!\!\!\perp W_i|\mathbf{X}_i, M_i(w^\star) = m$)

$$= \sum_m \mathbb{E}[Y_i(w,m)|\mathbf{X}_i, W_i = w^\star] Pr(M_i(w^\star) = m|\mathbf{X}_i)$$

(due to SI (2), i.e., ignorability of the mediator)

$$= \sum_m \mathbb{E}[Y_i(w,m)|\mathbf{X}_i, W_i = w] Pr(M_i(w^\star) = m|\mathbf{X}_i, W_i = w^\star)$$

(due to SI (1), i.e., ignorability of the treatment)

$$= \sum_m \mathbb{E}[Y_i(w,m)|\mathbf{X}_i, W_i = w, M_i(w) = m] Pr(M_i(w^\star) = m|\mathbf{X}_i, W_i = w^\star)$$

(due to SI (2), i.e., ignorability of the mediator)

$$= \sum_m \mathbb{E}[Y_i|\mathbf{X}_i, W_i = w, M_i(w) = m] Pr(M_i) = m|\mathbf{X}_i, W_i = w^\star)$$

(due to consistency)

Identification under Sequential Ignorability

Proof. (2)

$$\mathbb{E}[Y_i(w, M_i(w^\star))|\mathbf{X}_i] = \sum_m \mathbb{E}[Y_i(1,m)|M_i(w^\star)=m, \mathbf{X}_i] Pr(M_i(w^\star)=m|\mathbf{X}_i)$$

(due to the law of total expectation)

$$= \sum_m \mathbb{E}[Y_i(w,m)|\mathbf{X}_i] Pr(M_i(w^\star)=m|\mathbf{X}_i)$$

(due to the cross-world assumption SI (D) $Y_i(w,m) \perp\!\!\!\perp M_i(w^\star)|\mathbf{X}_i$)

$$= \sum_m \mathbb{E}[Y_i(w,m)|\mathbf{X}_i, W_i=w] Pr(M_i(w^\star)=m|\mathbf{X}_i, W_i=w^\star)$$

(due to SI (1), i.e., ignorability of the treatment)

$$= \sum_m \mathbb{E}[Y_i(w,m)|\mathbf{X}_i, W_i=w, M_i(w)=m] Pr(M_i(w^\star)=m|\mathbf{X}_i, W_i=w^\star)$$

(due to SI (2), i.e., ignorability of the mediator)

$$= \sum_m \mathbb{E}[Y_i|\mathbf{X}_i, W_i=w, M_i=m] Pr(M_i)=m|\mathbf{X}_i, W_i=w^\star)$$

(due to consistency)

□

Identification under Sequential Ignorability

(Imai et. al, 2010)

- With a continuous mediator and possibly continuous covariates we have:

Natural Direct Effect

$$
\begin{aligned}
NDE(w) &= \mathbb{E}[Y_i(1, M_i(w)) - Y_i(0, M_i(w))] \\
&= \int_{\mathbf{x}} \int_{m} \left(\mathbb{E}[Y_i | W_i = 1, M_i = m, \mathbf{X}_i = \mathbf{x}] - \mathbb{E}[Y_i | W_i = 0, M_i = m, \mathbf{X}_i = \mathbf{x}]\right) \\
&\qquad \times dF_M(m | W_i = w, \mathbf{X}_i = \mathbf{x}) dF_X(\mathbf{x}) \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_i | W_i = 1, M_i, \mathbf{X}_i] \ W_i = w, \mathbf{X}_i]] - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_i | W_i = 0, M_i, \mathbf{X}_i] \ W_i = w, \mathbf{X}_i]]
\end{aligned}
$$

Natural Indirect Effect

$$
\begin{aligned}
NIE(w) &= \mathbb{E}[Y_i(w, M_i(1)) - Y_i(w, M_i(0))] \\
&= \int_{\mathbf{x}} \int_{m} \mathbb{E}[Y_i | W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] \\
&\qquad \times \left(dF_M(m | W_i = 1, \mathbf{X}_i = \mathbf{x}) - dF_M(m | W_i = 0, \mathbf{X}_i = \mathbf{x})\right) dF_X(\mathbf{x}) \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_i | W_i = w, M_i, \mathbf{X}_i] \ W_i = 1, \mathbf{X}_i]] - \mathbb{E}[\mathbb{E}[\mathbb{E}[Y_i | W_i = w, M_i, \mathbf{X}_i] \ W_i = 0, \mathbf{X}_i]]
\end{aligned}
$$

Identification under Sequential Ignorability

(Imai et. al, 2010)

- With a binary mediator and binary treatment:

Natural Indirect Effect

$$
\begin{aligned}
NIE(w) &= \mathbb{E}[Y_i(w, M_i(1)) - Y_i(w, M_i(0))] \\
&= \sum_{\mathbf{x}} \big(\mathbb{E}[Y_i | W_i = w, M_i = 1, \mathbf{X}_i = \mathbf{x}] - \mathbb{E}[Y_i | W_i = w, M_i = 0, \mathbf{X}_i = \mathbf{x}]\big) \\
&\qquad \times \big(Pr(M_i = 1 | W_i = 1, \mathbf{X}_i = \mathbf{x}) - Pr(M_i = 1 | W_i = 0, \mathbf{X}_i = \mathbf{x})\big) Pr(\mathbf{X}_i = \mathbf{x}) \\
&= (\text{Effect of M on Y}) \times (\text{Effect of W on Y})
\end{aligned}
$$

Connection with Product Method

- Remember the Product Method (Baron and Kenny, 1986)

$$(1) \quad \mathbb{E}[M_i = m | W_i = w, \mathbf{X}_i = \mathbf{x}] = \beta_0 + \beta_1 w + \beta_2^T \mathbf{x}$$

$$(2) \quad \mathbb{E}[Y_i | W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x}$$

$$\text{Indirect effect } = \beta_1 \times \theta_2 \qquad \text{Direct effect } = \theta_1$$

Connection with Product Method

- Remember the Product Method (Baron and Kenny, 1986)

$$(1) \quad \mathbb{E}[M_i = m | W_i = w, \mathbf{X}_i = \mathbf{x}] = \beta_0 + \beta_1 w + \beta_2^T \mathbf{x}$$

$$(2) \quad \mathbb{E}[Y_i | W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x}$$

$$\text{Indirect effect } = \beta_1 \times \theta_2 \qquad \text{Direct effect } = \theta_1$$

- We can show that under SUTVA, SI and the models (1) and (2) being correct (without interaction between w and m), the natural direct and indirect effects, identified by the g-formula, are finally identified as in the product method, that is:

$$NIE(1) = NIE(0) = \beta_1 \times \theta_2 \qquad NDE(1) = NDE(0) = \theta_1$$

Connection with Product Method

- Remember the Product Method (Baron and Kenny, 1986)

$$(1) \quad \mathbb{E}[M_i = m | W_i = w, \mathbf{X}_i = \mathbf{x}] = \beta_0 + \beta_1 w + \beta_2^T \mathbf{x}$$

$$(2) \quad \mathbb{E}[Y_i | W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] = \theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x}$$

$$\text{Indirect effect } = \beta_1 \times \theta_2 \qquad \text{Direct effect } = \theta_1$$

- We can show that under SUTVA, SI and the models (1) and (2) being correct (without interaction between w and m), the natural direct and indirect effects, identified by the g-formula, are finally identified as in the product method, that is:

$$NIE(1) = NIE(0) = \beta_1 \times \theta_2 \qquad NDE(1) = NDE(0) = \theta_1$$

- This is easy to see under a binary mediator and binary treatment, which implied the following identification result

$$\begin{aligned} NIE(w) = \sum_{\mathbf{x}} & \big(\mathbb{E}[Y_i | W_i = w, M_i = 1, \mathbf{X}_i = \mathbf{x}] - \mathbb{E}[Y_i | W_i = w, M_i = 0, \mathbf{X}_i = \mathbf{x}]\big) \\ & \times \big(Pr(M_i = 1 | W_i = 1, \mathbf{X}_i = \mathbf{x}) - Pr(M_i = 1 | W_i = 0, \mathbf{X}_i = \mathbf{x})\big) Pr(\mathbf{X}_i = \mathbf{x}) \\ = & \text{ (Effect of M on Y)} \times \text{(Effect of W on Y)} \end{aligned}$$

Connection with Product Method

Proof.

$$
\begin{aligned}
NIE(w) &= \mathbb{E}[Y_i(w, M_i(1)) - Y_i(w, M_i(0))] \\
&= \sum_{\mathbf{x}} \sum_{m} \mathbb{E}[Y_i | W_i = w, M_i = m, \mathbf{X}_i = \mathbf{x}] \\
&\qquad \times \big(Pr(M_i = m | W_i = 1, \mathbf{X}_i = \mathbf{x}) - Pr(M_i = m | W_i = 0, \mathbf{X}_i = \mathbf{x}) \big) Pr(\mathbf{X}_i = \mathbf{x}) \\
&= \sum_{\mathbf{x}} \sum_{m} \Big((\theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x}) Pr(M_i = m | W_i = 1, \mathbf{X}_i = \mathbf{x}) \\
&\qquad - (\theta_0 + \theta_1 w + \theta_2 m + \theta_4^T \mathbf{x}) Pr(M_i = m | W_i = 0, \mathbf{X}_i = \mathbf{x}) \Big) Pr(\mathbf{X}_i = \mathbf{x}) \\
&= \sum_{\mathbf{x}} (\theta_0 + \theta_1 w + \theta_4^T \mathbf{x}) - (\theta_0 + \theta_1 w + \theta_4^T \mathbf{x}) + \\
&\qquad \sum_{\mathbf{x}} \theta_2 \left(\sum_{m} m Pr(M_i = m | W_i = 1, \mathbf{X}_i = \mathbf{x}) - \sum_{m} m Pr(M_i = m | W_i = 0, \mathbf{X}_i = \mathbf{x}) \right) Pr(\mathbf{X}_i = \mathbf{x}) \\
&= \sum_{\mathbf{x}} \theta_2 \left(\mathbb{E}[M_i | W_i = 1, \mathbf{X}_i] - \mathbb{E}[M_i | W_i = 0, \mathbf{X}_i] \right) Pr(\mathbf{X}_i = \mathbf{x}) \\
&= \sum_{\mathbf{x}} \theta_2 \Big((\beta_0 + \beta_1 + \beta_2^T \mathbf{x}) - (\beta_0 + \beta_2^T \mathbf{x}) \Big) Pr(\mathbf{X}_i = \mathbf{x}) = \theta_2 \times \beta_1
\end{aligned}
$$

Product Method Estimator

- Given the previous considerations, we know that under the models (1) and (2), also known as Linear Structural Equation Models (LSEM), the following estimators are unbiased:

$$\widehat{NDE} = \hat{\theta}_1 \qquad \widehat{NIE} = \hat{\beta}_1 \times \hat{\theta}_2$$

- The standard errors (and variances) of the estimators can be derived from the standard errors of the coefficients:
  - The Variance of the direct effect is equal to the variance of the coefficient θ_1

$$\widehat{Var}[\widehat{NDE}] = \widehat{Var}[\hat{\theta}_1]$$

  - Using the delta method, the variance of $\widehat{NIE} = \hat{\beta}_1 \times \hat{\theta}_2$ can be approximated by:

$$\widehat{Var}[\widehat{NIE}] \approx \hat{\theta}_2^2 \widehat{Var}[\hat{\beta}_1] + \hat{\beta}_1^2 \widehat{Var}[\hat{\theta}_2]$$

    under the assumption that $Cov[\hat{\theta}_2, \hat{\beta}_1] \approx 0$

References

Baron RM, Kenny DA. The moderator-mediator variable distinction in social psycho- logical research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986; 51:1173-1182.

Ding, P., & Vanderweele, T. J. (2016). Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. Biometrika, 103(2), 483–490.

Hafeman, D.M. (2009). "Proportion explained": a causal interpretation for standard measures of indirect effect? Am J Epidemiol. 2009;170(11):1443-1448.

Hong, G. (2010).Ratio of mediator probability weighting for estimating natural direct and indirect effects. Proceedings of the American Statistical Association, Biometrics Section, Alexandria, VA: American Statistical Association, 2401–2415.

Imai, K., Keele, L., Yamamoto, T. (2010a). Identification, inference, and sensitivity analysis for causal mediation effects. Statistical Science, 25:51-71.

Imai, K., Keele, L., Tingley, D. (2010b). A general approach to causal mediation analysis. Psychological Methods, 15:309-334.

Imai, K., Keele, L., Tingley, D., Yamamoto, T. (2010c). Causal mediation analysis using R. In: H.D. Vinod (ed.), Advances in Social Science Research Using R. New York: Springer (Lecture Notes in Statistics), p.129-154.

Lange, T., S. Vansteelandt, and M. Bekaert (2012). A simple approach for estimating natural direct and indirect effects. American Journal of Epidemiology, 176, 190–195.

MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. Evaluation Review 1993; 17:144-158.

Pearl, J. (2001). Direct and indirect effects. In *Proc. 17th Conf. Uncertainty in Artificial Intelligence* (eds. J. S. Breese & D. Koller), 411–420. Morgan Kaufman, S. Francisco, CA.

References

Robins JM, Greenland S. (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology 3, 143-155.

Smith LH, VanderWeele TJ (2019). Mediational E-values: Approximate Sensitivity Analysis for Unmeasured Mediator-Outcome Confounding. Epidemiology. Nov;30(6):835-837.

Tchetgen Tchetgen, E.J. (2013). A note on formulae for causal mediation analysis in an odds ratio context. Epidemiologic Methods, 2:21-32.

Tchetgen Tchetgen EJ (2013). Inverse odds ratio-weighted estimation for causal mediation analysis. Statistics in medicine. 32: 4567 –4580.

Valeri, L. and VanderWeele, T.J. (2013) Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological Methods, 18:137-150.

VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. Epidemiology. Jan;20(1):18-26.

VanderWeele, T.J. (2015). Explanation in causal inference: Methods for mediation and interaction.

VanderWeele, T.J., Asomaning, K., Tchetgen Tchetgen, E.J., Han, Y., Spitz, M.R., Shete, S., Wu, X., Gaborieau, V., Wang, Y., McLaughlin, J., Hung, R.J., Brennan, P., Amos, C.I., Christiani, D.C. and Lin, X. (2012). Genetic variants on 15q25.1, smoking and lung cancer: an assessment of mediation and interaction. American Journal of Epidemiology, 75:1013-1020.

VanderWeele, T.J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. Statistics and Its Interface 2:457-468.

References

VanderWeele, T.J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis with a dichotomous outcome. American Journal of Epidemiology, 172:1339-1348.

VanderWeele TJ, Vansteelandt S (2014). Mediation analysis with multiple mediators. Epidemiologic Methods. 2(1): 95 - 115.

VanderWeele TJ, Tchetgen Tchetgen EJ (2017). Mediation analysis with time varying exposures and mediators. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 79(3): 917 - 938.

Vansteelandt S, Bekaert M, Lange T (2012). Imputation Strategies for the Estimation of Natural Direct and Indirect Effects. Epidemiologic Methods. 1(1): 131 - 158.