# Summary

*Baoyi Shi*

## 1 Supported Data Types and Functionalities

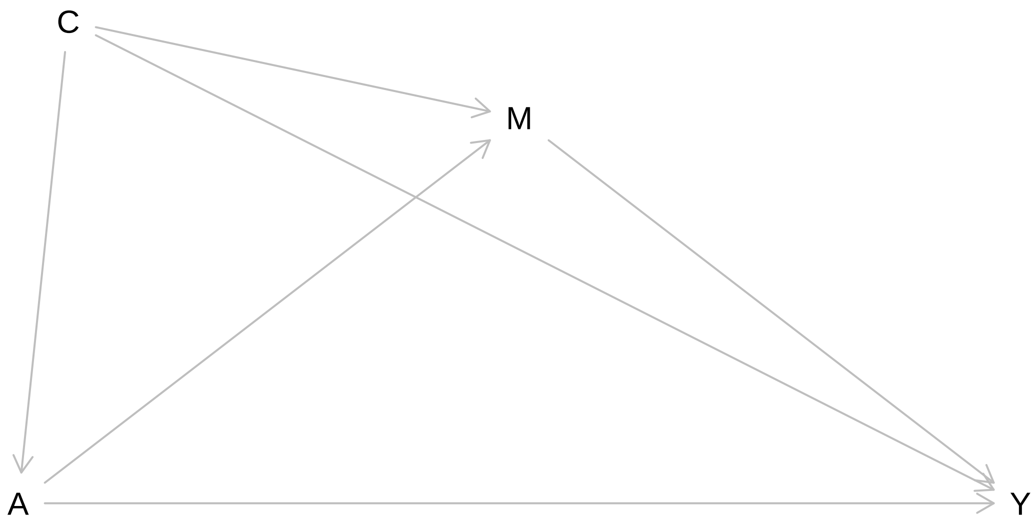
Table 1: Supported Data Types and Functionalities

| | rb | wb | msm | iorw | ne | g-formula |
|---|---|---|---|---|---|---|
| Linear Y | √ | √ | √ | √ | √ | √ |
| Logistic Y | √ | √ | √ | √ | √ | √ |
| Loglinear Y | √ | √ | √ | √ | √ | √ |
| Poisson Y | √ | √ | √ | √ | √ | √ |
| Quasipoisson Y | √ | √ | √ | √ | √ | √ |
| NegBin Y | √ | √ | √ | √ | × | √ |
| Coxph Y | √ | √ | √ | √ | × | √ |
| AFT Exp Y | √ | √ | √ | √ | × | √ |
| AFT Weibull Y | √ | √ | √ | √ | × | √ |
| Linear M | √ | √ | √ | √ | √ | √ |
| Logistic M | √ | √ | √ | √ | √ | √ |
| Categorical M | √ | √ | √ | √ | √ | √ |
| Any type M | ⊘ | √ | ⊘ | √ | √ | ⊘ |
| User-defined Y/M Models | √ | √ | √ | √ | × | √ |
| Continuous A | √ | ⊘ | ⊘ | ⊘ | √ | √ |
| Binary A | √ | √ | √ | √ | √ | √ |
| Categorical A | √ | √ | √ | √ | √ | √ |
| Single M | √ | √ | √ | √ | √ | √ |
| Multiple M | √ | √ | √ | √ | √ | √ |
| Pre-exposure Confounding | √ | √ | √ | √ | √ | √ |
| Post-exposure Confounding | ⊘ | ⊘ | √ | ⊘ | ⊘ | √ |
| 2-way Decomposition | √ | √ | √ | √ | √ | √ |
| 4-way Decomposition | √ | √ | √ | ⊘ | × | √ |
| Estimation: Closed-form Parameter Function | √* | ⊘ | ⊘ | √ | √ | ⊘ |
| Estimation: Direct Imputation | √ | √ | √ | ⊘ | × | √ |
| Inference: Delta Method** | √ | ⊘ | ⊘ | ⊘ | √ | ⊘ |
| Inference: Bootstrapping | √ | √ | √ | √ | √ | √ |

* Not available for multiple-mediator cases; only outputs conditional causal effects.
** Only available for closed-form parameter function estimation.
√ Available.
× Currently not available but applicable.
⊘ Neither available nor applicable.
rb: Regression-based Approach; wb: Weighting-based Approach; msm: Marginal; iorw: Inverse Odds Ratio Weighting Approach; ne: Natural Effect Model; g-formula: G-formula Approach.

# 2 Pre-treatment Confounding

## 2.1 DAG




In the DAG, A denotes the treatment, Y denotes the outcome, M denotes a set of mediators, and C denotes a set of pre-treatment covariates.

## 2.2 Estimand

**2-way decomposition in additive scale**

$CDE = E[Y_{am} - Y_{a^*m}]$

$PNDE = E[Y_{aM_a^*} - Y_{a^*M_a^*}]$

$TNDE = E[Y_{aM_a} - Y_{a^*M_a}]$

$PNIE = E[Y_{a^*M_a} - Y_{a^*M_a^*}]$

$TNIE = E[Y_{aM_a} - Y_{aM_a^*}]$

$TE = PNDE + TNIE$

$PM = \frac{TNIE}{PNDE+TE}$

**2-way decomposition in RR scale**

$rr^{CDE} = E[Y_{am}]/E[Y_{a^*m}]$

$rr^{PNDE} = E[Y_{aM_a^*}]/E[Y_{a^*M_a^*}]$

$rr^{TNDE} = E[Y_{aM_a}]/E[Y_{a^*M_a}]$

$rr^{PNIE} = E[Y_{a^*M_a}]/E[Y_{a^*M_a^*}]$

$rr^{TNIE} = E[Y_{aM_a}]/E[Y_{aM_a^*}]$

$rr^{TE} = rr^{PNDE} \times rr^{TNIE}$

$PM = \frac{rr^{PNDE} * (rr^{TNIE} - 1)}{rr^{TE} - 1}$

**4-way decomposition in additive scale**

$CDE = E[Y_{am} - Y_{a^*m}]$

$INT_{ref} = PNDE - CDE$

$INT_{med} = TNIE - PNIE$

$PIE = PNIE$

$prop^{CDE} = \frac{CDE}{TE}$

$prop^{INT_{ref}} = \frac{INT_{ref}}{TE}$

$prop^{INT_{med}} = \frac{INT_{med}}{TE}$

$prop^{PIE} = \frac{PIE}{TE}$

$overall^{PM} = \frac{PNIE+INT_{med}}{TE}$

$overall^{INT} = \frac{INT_{ref}+INT_{med}}{TE}$

$overall^{PE} = \frac{INT_{ref}+INT_{med}+PIE}{TE}$

**4-way decomposition in RR scale**

$err^{CDE} = (E[Y_{am} - Y_{a^*m}])/E[Y_{a^*M_a^*}]$

$err^{INT_{ref}} = rr^{PNDE} - 1 - err^{CDE}$

$err^{INT_{med}} = rr^{TNIE} * rr^{PNDE} - rr^{PNDE} - rr^{PNIE} + 1$

$err^{PIE} = rr^{PNIE} - 1$

$err^{TE} = err^{CDE} + err^{INT_{ref}} + err^{INT_{med}} + err^{PIE} = rr^{TE} - 1$

$prop^{err^{CDE}} = \frac{err^{CDE}}{err^{TE}}$

$prop^{err^{INT_{ref}}} = \frac{err^{INT_{ref}}}{err^{TE}}$

$prop^{err^{INT_{med}}} = \frac{err^{INT_{med}}}{err^{TE}}$

$prop^{err^{PIE}} = \frac{err^{PIE}}{err^{TE}}$

$overall^{PM} = \frac{err^{PIE}+err^{INT_{med}}}{err^{TE}}$

$overall^{INT} = \frac{err^{INT_{ref}}+err^{INT_{med}}}{err^{TE}}$

$overall^{PE} = \frac{err^{INT_{ref}}+err^{INT_{med}}+err^{PIE}}{err^{TE}}$

## 2.3 Estimation

**Closed-form Parameter Function Estimation**

Effect estimates are calculated using regression parameters.

**Direct Conterfacturals imputation Estimation**

1. Impute $Y_{am,i}$, $Y_{a^*m,i}$, $Y_{aMa,i}$, $Y_{a^*Ma^*,i}$, $Y_{aMa^*,i}$, and $Y_{a^*Ma,i}$ for each subject i.
2. Estimate $E[Y_{am}]$, $E[Y_{a^*m}]$, $E[Y_{aMa}]$, $E[Y_{a^*Ma^*}]$, $E[Y_{aMa^*}]$, and $E[Y_{a^*Ma}]$ by $\frac{\sum_{i=1}^{N} Y_{am,i}}{n}$, $\frac{\sum_{i=1}^{N} Y_{a^*m,i}}{n}$, $\frac{\sum_{i=1}^{N} Y_{aMa,i}}{n}$, $\frac{\sum_{i=1}^{N} Y_{a^*Ma^*,i}}{n}$, $\frac{\sum_{i=1}^{N} Y_{aMa^*,i}}{n}$ and $\frac{\sum_{i=1}^{N} Y_{a^*Ma,i}}{n}$ respectively.
3. Calculate causal effects using $E[Y_{am}]$, $E[Y_{a^*m}]$, $E[Y_{aMa}]$, $E[Y_{a^*Ma^*}]$, $E[Y_{aMa^*}]$, and $E[Y_{a^*Ma}]$.

## 2.4 Inference

**Delta Method**

Standard errors of effects are estimated using the standard errors of regression parameters and delta method based on the closed-form parameter function.

**Bootstrapping**

Bootstrap the data, refit the regression models, and calculate a bootstrap estimate. Repeat the bootstrapping K times and calculate the standard error of these K boostrap estimates for each estimand, which is the estimated standard error of this estimand.

## 2.5 Estimation Approaches Can be Used

### 2.5.1 Regression-based Approach

Reference:

https://www.ncbi.nlm.nih.gov/pubmed/23379553

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287269/

#### 2.5.1.1 Estimation

**Closed-form Parameter Function Estimation**

(1) Continuous Outcome and Continuous Mediator

Fit a simple linear regression model for the mediator:

$$E[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c \; (for\ continuous\ exposure)$$

$$E[M|a,c] = \beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a=k\} + \beta_2' c \; (for\ binary\ or\ categorical\ exposure)$$

Fit a simple linear regression model for the outcome:

$$E[Y|a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \; (for\ continuous\ exposure)$$

$$E[Y|a,m,c] = \theta_0 + \sum_{k=1}^{K} \theta_{1k} I\{a=k\} + \theta_2 m + \sum_{k=1}^{K} \theta_{3k} I\{a=k\} m + \theta_4' c \; (for\ binary\ or\ categorical\ exposure)$$

Closed-form parameter function estimators for the causal effects when the exposure is continuous:

$CDE = (\theta_1 + \theta_3 m)(a - a^*)$

$PNDE = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)\}(a - a^*)$

$TNDE = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a + \beta_2' c)\}(a - a^*)$

$PNIE = (\theta_2 \beta_1 + \theta_3 \beta_1 a^*)(a - a^*)$

$TNIE = (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*)$

Closed-form parameter function estimators for the causal effects when the exposure is binary or categorical:

$CDE = (\sum_{k=1}^{K} \theta_{1k} I\{a = k\} - \sum_{k=1}^{K} \theta_{1k} I\{a^* = k\}) + (\sum_{k=1}^{K} \theta_{3k} I\{a = k\} - \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\})m$

$PNDE = (\sum_{k=1}^{K} \theta_{1k} I\{a = k\} - \sum_{k=1}^{K} \theta_{1k} I\{a^* = k\}) + (\sum_{k=1}^{K} \theta_{3k} I\{a = k\} - \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\})(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a^* = k\} + \beta_2' c)$

$TNDE = (\sum_{k=1}^{K} \theta_{1k} I\{a = k\} - \sum_{k=1}^{K} \theta_{1k} I\{a^* = k\}) + (\sum_{k=1}^{K} \theta_{3k} I\{a = k\} - \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\})(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a = k\} + \beta_2' c)$

$PNIE = (\theta_2 + \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\})(\sum_{k=1}^{K} \beta_{1k} I\{a = k\} - \sum_{k=1}^{K} \beta_{1k} I\{a^* = k\})$

$TNIE = (\theta_2 + \sum_{k=1}^{K} \theta_{3k} I\{a = k\})(\sum_{k=1}^{K} \beta_{1k} I\{a = k\} - \sum_{k=1}^{K} \beta_{1k} I\{a^* = k\})$

(2)Continuous Outcome and Binary Mediator

Fit a logistic regression model for the mediator:

$$logitE[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c \ (for\ continuous\ exposure)$$

$$logitE[M|a,c] = \beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a = k\} + \beta_2' c \ (for\ binary\ or\ categorical\ exposure)$$

Fit a simple linear regression model for the outcome:

$$E[Y|a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \ (for\ continuous\ exposure)$$

$$E[Y|a,m,c] = \theta_0 + \sum_{k=1}^{K} \theta_{1k} I\{a = k\} + \theta_2 m + \sum_{k=1}^{K} \theta_{3k} I\{a = k\} m + \theta_4' c \ (for\ binary\ or\ categorical\ exposure)$$

Closed-form parameter function estimators for the causal effects when the exposure is continuous:

$CDE = (\theta_1 + \theta_3 m)(a - a^*)$

$PNDE = \{\theta_1 + \theta_3 \frac{exp(\beta_0 + \beta_1 a^* + \beta_2' c)}{1 + exp(\beta_0 + \beta_1 a^* + \beta_2' c)}\}(a - a^*)$

$TNDE = \{\theta_1 + \theta_3 \frac{exp(\beta_0 + \beta_1 a + \beta_2' c)}{1 + exp(\beta_0 + \beta_1 a + \beta_2' c)}\}(a - a^*)$

$PNIE = (\theta_2 + \theta_3 a^*)(\frac{exp(\beta_0 + \beta_1 a + \beta_2' c)}{1 + exp(\beta_0 + \beta_1 a + \beta_2' c)} - \frac{exp(\beta_0 + \beta_1 a^* + \beta_2' c)}{1 + exp(\beta_0 + \beta_1 a^* + \beta_2' c)})$

$TNIE = (\theta_2 + \theta_3 a)(\frac{exp(\beta_0 + \beta_1 a + \beta_2' c)}{1 + exp(\beta_0 + \beta_1 a + \beta_2' c)} - \frac{exp(\beta_0 + \beta_1 a^* + \beta_2' c)}{1 + exp(\beta_0 + \beta_1 a^* + \beta_2' c)})$

Closed-form parameter function estimators for the causal effects when the exposure is binary or categorical:

$CDE = (\sum_{k=1}^{K} \theta_{1k} I\{a = k\} - \sum_{k=1}^{K} \theta_{1k} I\{a^* = k\}) + (\sum_{k=1}^{K} \theta_{3k} I\{a = k\} - \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\})m$

$PNDE = (\sum_{k=1}^{K} \theta_{1k} I\{a = k\} - \sum_{k=1}^{K} \theta_{1k} I\{a^* = k\}) + (\sum_{k=1}^{K} \theta_{3k} I\{a = k\} - \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\}) \frac{exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a^* = k\} + \beta_2' c)}{1 + exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a^* = k\} + \beta_2' c)}$

$TNDE = (\sum_{k=1}^{K} \theta_{1k} I\{a = k\} - \sum_{k=1}^{K} \theta_{1k} I\{a^* = k\}) + (\sum_{k=1}^{K} \theta_{3k} I\{a = k\} - \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\}) \frac{exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a = k\} + \beta_2' c)}{1 + exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a = k\} + \beta_2' c)}$

$PNIE = (\theta_2 + \sum_{k=1}^{K} \theta_{3k} I\{a^* = k\})(\frac{exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a = k\} + \beta_2' c)}{1 + exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a = k\} + \beta_2' c)} - \frac{exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a^* = k\} + \beta_2' c)}{1 + exp(\beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a^* = k\} + \beta_2' c)})$

$TNIE = (\theta_2 + \sum_{k=1}^{K} \theta_{3k} I\{a=k\})(\frac{exp(\beta_0+\sum_{k=1}^{K}\beta_{1k}I\{a=k\}+\beta_2'c)}{1+exp(\beta_0+\sum_{k=1}^{K}\beta_{1k}I\{a=k\}+\beta_2'c)} - \frac{exp(\beta_0+\sum_{k=1}^{K}\beta_{1k}I\{a^*=k\}+\beta_2'c)}{1+exp(\beta_0+\sum_{k=1}^{K}\beta_{1k}I\{a^*=k\}+\beta_2'c)})$

(3)Continuous Outcome and Categorical Mediator

Fit a multinomial logistic regression model for the mediator:

$$log\frac{E[M=j|a,c]}{E[M=0|a,c]} = \beta_{0j} + \beta_{1j}a + \beta_{2j}'c, j=1,2,...,l\ (for\ continuous\ exposure)$$

$$log\frac{E[M=j|a,c]}{E[M=0|a,c]} = \beta_{0j} + \sum_{k=1}^{K}\beta_{1jk}I\{a=k\} + \beta_{2j}'c, j=1,2,...,l\ (for\ binary\ or\ categorical\ exposure)$$

Fit a simple linear regression model for the outcome:

$$E[Y|a,m,c] = \theta_0 + \theta_1 a + \sum_{j=1}^{l}\theta_{2j}I\{m=j\} + a\sum_{j=1}^{l}\theta_{3j}I\{m=j\} + \theta_4'c\ (for\ continuous\ exposure)$$

$$E[Y|a,m,c] = \theta_0 + \sum_{k=1}^{K}\theta_{1k}I\{a=k\} + \sum_{j=1}^{l}\theta_{2j}I\{m=j\} + \sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{m=j\}I\{a=k\} + \theta_4'c$$
$$(for\ binary\ or\ categorical\ exposure)$$

Closed-form parameter function estimators for the causal effects when the exposure is continuous:

$CDE = (\theta_1 + \sum_{j=1}^{l}\theta_{3j}I\{m=j\})(a-a^*)$

$PNDE = \{\theta_1 + \frac{\sum_{j=1}^{l}\theta_{3j}exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)}\}(a-a^*)$

$TNDE = \{\theta_1 + \frac{\sum_{j=1}^{l}\theta_{3j}exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)}\}(a-a^*)$

$PNIE = \frac{\sum_{j=1}^{l}(\theta_{2j}+\theta_{3j}a^*)exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)} - \frac{\sum_{j=1}^{l}(\theta_{2j}+\theta_{3j}a^*)exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)}$

$TNIE = \frac{\sum_{j=1}^{l}(\theta_{2j}+\theta_{3j}a)exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)} - \frac{\sum_{j=1}^{l}(\theta_{2j}+\theta_{3j}a)exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)}$

Closed-form parameter function estimators for the causal effects when the exposure is binary or categorical:

$CDE = (\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + (\sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{m=l\}I\{a=k\} - \sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{m=l\}I\{a^*=k\})$

$PNDE = (\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + \frac{\sum_{j=1}^{l}(\sum_{k=1}^{K}\theta_{3jk}I\{a=k\}-\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\})exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}$

$TNDE = (\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + \frac{\sum_{j=1}^{l}(\sum_{k=1}^{K}\theta_{3jk}I\{a=k\}-\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\})exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)}$

$PNIE = \frac{\sum_{j=1}^{l}(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\})exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)} - \frac{\sum_{j=1}^{l}(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\})exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}$

$$TNIE = \frac{\sum_{j=1}^{l}(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a=k\})exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)} - \frac{\sum_{j=1}^{l}(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a=k\})exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}{1+\sum_{j=1}^{l}exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}$$

(4)Binary Outcome and Continuous Mediator

Fit a simple linear regression model for the mediator:

$$E[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c \ (for\ continuous\ exposure)$$

$$E[M|a,c] = \beta_0 + \sum_{k=1}^{K}\beta_{1k}I\{a=k\} + \beta_2' c \ (for\ binary\ or\ categorical\ exposure)$$

Fit a logistic regression model for the outcome:

$$logitE[Y|a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \ (for\ continuous\ exposure)$$

$$logitE[Y|a,m,c] = \theta_0 + \sum_{k=1}^{K}\theta_{1k}I\{a=k\} + \theta_2 m + \sum_{k=1}^{K}\theta_{3k}I\{a=k\}m + \theta_4' c \ (for\ binary\ or\ categorical\ exposure)$$

Closed-form parameter function estimators for the causal effects when the exposure is continuous:

$OR^{CDE} = exp((\theta_1 + \theta_3 m)(a - a^*))$

$OR^{PNDE} = exp(\{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c + \theta_2\sigma^2)\}(a - a^*) + 0.5\theta_3^2\sigma^2(a^2 - a^{*2}))$

$OR^{TNDE} = exp(\{\theta_1 + \theta_3(\beta_0 + \beta_1 a + \beta_2' c + \theta_2\sigma^2)\}(a - a^*) + 0.5\theta_3^2\sigma^2(a^2 - a^{*2}))$

$OR^{PNIE} = exp((\theta_2\beta_1 + \theta_3\beta_1 a^*)(a - a^*))$

$OR^{TNIE} = exp((\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*))$

$comp^{CDE} = (exp(\theta_1(a-a^*)+\theta_3 am) - exp(\theta_3 a^* m))exp(\theta_2 m - (\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c) - 0.5(\theta_2+\theta_3 a^*)^2\sigma^2)$

Closed-form parameter function estimators for the causal effects when the exposure is binary or categorical:

$OR^{CDE} = exp((\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + (\sum_{k=1}^{K}\theta_{3k}I\{a=k\} - \sum_{k=1}^{K}\theta_{3k}I\{a^*=k\})m)$

$OR^{PNDE} = exp((\{\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + (\{\sum_{k=1}^{K}\theta_{3k}I\{a=k\} - \sum_{k=1}^{K}\theta_{3k}I\{a^*=k\})(\beta_0 + \sum_{k=1}^{K}\beta_{1k}I\{a^*=k\} + \beta_2' c + \theta_2\sigma^2)\} + 0.5\sigma^2(\sum_{k=1}^{K}\theta_{3k}^2 I\{a=k\} - \sum_{k=1}^{K}\theta_{3k}^2 I\{a^*=k\}))$

$OR^{TNDE} = exp((\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + (\sum_{k=1}^{K}\theta_{3k}I\{a=k\} - \sum_{k=1}^{K}\theta_{3k}I\{a^*=k\})(\beta_0 + \sum_{k=1}^{K}\beta_{1k}I\{a=k\} + \beta_2' c + \theta_2\sigma^2) + 0.5\sigma^2(\sum_{k=1}^{K}\theta_{3k}^2 I\{a=k\} - \sum_{k=1}^{K}\theta_{3k}^2 I\{a^*=k\}))$

$OR^{PNIE} = exp(\theta_2(\sum_{k=1}^{K}\beta_{1k}I\{a=k\} - \sum_{k=1}^{K}\beta_{1k}I\{a^*=k\}) + \sum_{k=1}^{K}\theta_{3k}I\{a^*=k\}(\sum_{k=1}^{K}\beta_{1k}I\{a=k\} - \sum_{k=1}^{K}\beta_{1k}I\{a^*=k\}))$

$OR^{TNIE} = exp(\theta_2(\sum_{k=1}^{K}\beta_{1k}I\{a=k\} - \sum_{k=1}^{K}\beta_{1k}I\{a^*=k\}) + \sum_{k=1}^{K}\theta_{3k}I\{a=k\}(\sum_{k=1}^{K}\beta_{1k}I\{a=k\} - \sum_{k=1}^{K}\beta_{1k}I\{a^*=k\}))$

$comp^{CDE} = (exp((\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + \sum_{k=1}^{K}\theta_{3k}I\{a=k\}m) - exp(\sum_{k=1}^{K}\theta_{3k}I\{a^*=k\}m))exp(\theta_2 m - (\theta_2+\sum_{k=1}^{K}\theta_{3k}I\{a^*=k\})(\beta_0+\sum_{k=1}^{K}\beta_{1k}I\{a^*=k\}+\beta_2' c) - 0.5(\theta_2+\sum_{k=1}^{K}\theta_{3k}I\{a^*=k\})^2\sigma^2)$

(5)Binary Outcome and Binary Mediator

Fit a logistic regression model for the mediator:

$$logitE[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c \ (for\ continuous\ exposure)$$

Fit a logistic regression model for the outcome:

$$logitE[M|a,c] = \beta_0 + \sum_{k=1}^{K} \beta_{1k} I\{a=k\} + \beta_2' c \; (for\ binary\ or\ categorical\ exposure)$$

Fit a logistic regression model for the outcome:

$$logitE[Y|a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \; (for\ continuous\ exposure)$$

$$logitE[Y|a,m,c] = \theta_0 + \sum_{k=1}^{K} \theta_{1k} I\{a=k\} + \theta_2 m + \sum_{k=1}^{K} \theta_{3k} I\{a=k\} m + \theta_4' c \; (for\ binary\ or\ categorical\ exposure)$$

Closed-form parameter function estimators for the causal effects when the exposure is continuous:

$OR^{CDE} = exp((\theta_1 + \theta_3 m)(a - a^*))$

$OR^{PNDE} = \frac{exp(\theta_1 a)\{1+exp(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\beta_2' c)\}}{exp(\theta_1 a^*)\{1+exp(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a^*+\beta_2' c)\}}$

$OR^{TNDE} = \frac{exp(\theta_1 a)\{1+exp(\theta_2+\theta_3 a+\beta_0+\beta_1 a+\beta_2' c)\}}{exp(\theta_1 a^*)\{1+exp(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a+\beta_2' c)\}}$

$OR^{PNIE} = \frac{\{1+exp(\beta_0+\beta_1 a^*+\beta_2' c)\}\{1+exp(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a+\beta_2' c)\}}{\{1+exp(\beta_0+\beta_1 a+\beta_2' c)\}\{1+exp(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a^*+\beta_2' c)\}}$

$OR^{TNIE} = \frac{\{1+exp(\beta_0+\beta_1 a^*+\beta_2' c)\}\{1+exp(\theta_2+\theta_3 a+\beta_0+\beta_1 a+\beta_2' c)\}}{\{1+exp(\beta_0+\beta_1 a+\beta_2' c)\}\{1+exp(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\beta_2' c)\}}$

$comp^{CDE} = \frac{exp(\theta_2 m)(exp(\theta_1(a-a^*)+\theta_3 am)-exp(\theta_3 a^* m))(1+exp(\beta_0+\beta_1 a^*+\beta_2' c))}{1+exp(\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*)}$

Closed-form parameter function estimators for the causal effects when the exposure is binary or categorical:

$OR^{CDE} = exp((\sum_{k=1}^{K} \theta_{1k} I\{a=k\} - \sum_{k=1}^{K} \theta_{1k} I\{a^*=k\}) + (\sum_{k=1}^{K} \theta_{3k} I\{a=k\} - \sum_{k=1}^{K} \theta_{3k} I\{a^*=k\})m)$

$OR^{PNDE} = \frac{exp(\sum_{k=1}^{K} \theta_{1k} I\{a=k\})\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c)\}}{exp(\sum_{k=1}^{K} \theta_{1k} I\{a^*=k\})\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a^*=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c)\}}$

$OR^{TNDE} = \frac{exp(\sum_{k=1}^{K} \theta_{1k} I\{a=k\})\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a=k\}+\beta_2' c)\}}{exp(\sum_{k=1}^{K} \theta_{1k} I\{a^*=k\})\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a^*=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a=k\}+\beta_2' c)\}}$

$OR^{PNIE} = \frac{\{1+exp(\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c)\}\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a^*=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a=k\}+\beta_2' c)\}}{\{1+exp(\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a=k\}+\beta_2' c)\}\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a^*=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c)\}}$

$OR^{TNIE} = \frac{\{1+exp(\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c)\}\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a=k\}+\beta_2' c)\}}{\{1+exp(\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a=k\}+\beta_2' c)\}\{1+exp(\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a=k\}+\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c)\}}$

$comp^{CDE} = \frac{exp(\theta_2 m)(exp((\sum_{k=1}^{K} \theta_{1k} I\{a=k\}-\sum_{k=1}^{K} \theta_{1k} I\{a^*=k\})+\sum_{k=1}^{K} \theta_{3k} I\{a=k\}m)-exp(\sum_{k=1}^{K} \theta_{3k} I\{a^*=k\}m))}{1+exp(\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c+\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a^*=k\})}$

$\frac{1+exp(\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c)}{1+exp(\beta_0+\sum_{k=1}^{K} \beta_{1k} I\{a^*=k\}+\beta_2' c+\theta_2+\sum_{k=1}^{K} \theta_{3k} I\{a^*=k\})}$

(6)Binary Outcome and Categorical Mediator

Fit a multinomial logistic regression model for the mediator:

$$log\frac{E[M=j|a,c]}{E[M=0|a,c]} = \beta_{0j} + \beta_{1j} a + \beta_{2j}' c, j = 1,2,...,l \; (for\ continuous\ exposure)$$

$$log\frac{E[M=j|a,c]}{E[M=0|a,c]} = \beta_{0j} + \sum_{k=1}^{K}\beta_{1jk}I\{a=k\} + \beta_{2j}'c, j=1,2,...,l\ (for\ binary\ or\ categorical\ exposure)$$

Fit a logistic regression model for the outcome:

$$logitE[Y|a,m,c] = \theta_0 + \theta_1 a + \sum_{j=1}^{l}\theta_{2j}I\{m=j\} + a\sum_{j=1}^{l}\theta_{3j}I\{m=j\} + \theta_4'c\ (for\ continuous\ exposure)$$

$$logitE[Y|a,m,c] = \theta_0 + \sum_{k=1}^{K}\theta_{1k}I\{a=k\} + \sum_{j=1}^{l}\theta_{2j}I\{m=j\} + \sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{m=j\}I\{a=k\} + \theta_4'c$$
$$(for\ binary\ or\ categorical\ exposure)$$

Closed-form parameter function estimators for the causal effects when the exposure is continuous:

$OR^{CDE} = exp((\theta_1 + \sum_{j=1}^{l}\theta_{3j}I\{m=j\})(a-a^*))$

$OR^{PNDE} = \frac{exp(\theta_1 a)\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a+\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)\}}{exp(\theta_1 a^*)\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a^*+\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)\}}$

$OR^{TNDE} = \frac{exp(\theta_1 a)\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a+\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)\}}{exp(\theta_1 a^*)\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a^*+\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)\}}$

$OR^{PNIE} = \frac{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a^*+\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)\}}{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a^*+\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)\}}$

$OR^{TNIE} = \frac{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a+\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)\}}{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\beta_{1j}a+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a+\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)\}}$

$comp^{CDE} = \frac{exp(\sum_{j=1}^{l}\theta_{2j}I\{m=j\})(1+\sum_{j=1}^{l} exp(\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c))(exp(\theta_1(a-a^*)+a\sum_{j=1}^{l}\theta_{3j}I\{m=j\})-exp(a^*\sum_{j=1}^{l}\theta_{3j}I\{m=j\}))}{1+\sum_{j=1}^{l} exp(\theta_{2j}+\theta_{3j}a^*+\beta_{0j}+\beta_{1j}a^*+\beta_{2j}'c)}$

Closed-form parameter function estimators for the causal effects when the exposure is binary or categorical:

$OR^{CDE} = exp((\sum_{k=1}^{K}\theta_{1k}I\{a=k\} - \sum_{k=1}^{K}\theta_{1k}I\{a^*=k\}) + (\sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{m=l\}I\{a=k\} - \sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{m=l\}I\{a^*=k\}))$

$OR^{PNDE} = \frac{exp(\sum_{k=1}^{K}\theta_{1k}I\{a=k\})\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)\}}{exp(\sum_{k=1}^{K}\theta_{1k}I\{a^*=k\})\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)\}}$

$OR^{TNDE} = \frac{exp(\sum_{k=1}^{K}\theta_{1k}I\{a=k\})\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)\}}{exp(\sum_{k=1}^{K}\theta_{1k}I\{a^*=k\})\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)\}}$

$OR^{PNIE} = \frac{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)\}}{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)\}}$

$OR^{TNIE} = \frac{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)\}}{\{1+\sum_{j=1}^{l} exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a=k\}+\beta_{2j}'c)\}\{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)\}}$

$comp^{CDE} = \frac{exp(\sum_{j=1}^{l}\theta_{2j}I\{m=j\})(1+\sum_{j=1}^{l} exp(\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c))}{1+\sum_{j=1}^{l} exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}$

$$\frac{(exp((\sum_{k=1}^{K}\theta_{1k}I\{a=k\}-\sum_{k=1}^{K}\theta_{1k}I\{a^*=k\})+\sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{a=k\}I\{m=j\})-exp(\sum_{j=1}^{l}\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\}I\{m=j\}))}{1+\sum_{j=1}^{l}exp(\theta_{2j}+\sum_{k=1}^{K}\theta_{3jk}I\{a^*=k\}+\beta_{0j}+\sum_{k=1}^{K}\beta_{1jk}I\{a^*=k\}+\beta_{2j}'c)}$$

**Direct Conterfacturals imputation Estimation**

1. Fit a regression model for Y on A, M and C.
2. Fit a regression model for each mediator in M on A and C.

Estimation Algorithm for $E(Y_{am})$:

3. For each subject i, simulate $Y_{am,i}$ by the predicted value of the outcome regression model under $A = a, M = m, C = C_i$.
4. Estimate $E[Y_{am}]$ by $\frac{\sum_{i=1}^{N} Y_{am,i}}{n}$.

Estimation Algorithm for $E[Y_{a1Ma2}]$:

3. For each subject i, simulate $M_{a2,i}$ by the predicted values of mediator regression models under $A = a2, C = C_i$.
4. For each subject i, simulate $Y_{a1Ma2,i}$ by the predicted value of the outcome regression model under $A = a1, M = M_{a2,i}, C = C_i$.
5. Estimate $E[Y_{a1Ma2}]$ by $\frac{\sum_{i=1}^{N} Y_{a1Ma2,i}}{n}$.

#### 2.5.1.2 Inference

When the estimands are estimated through closed-form parameter function estimation, their standard errors can be estimated by the delta method or bootstrapping; when the estimands are estimated through direct counterfactuals imputation estimation, their standard errors can be estimated by bootstrapping.

### 2.5.2 Weighting-based Approach

Reference:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287269/

#### 2.5.2.1 Estimation

**Direct Counterfactuals Imputation Estimation**

1. Fit a regression model for Y on A, M, C.
2. For $E[Y_{a^*m}]$, estimate it by taking a weighted average of the predicted Y values for subjects with $A = a^*$ if the subjects had had mediator $M = m$ rather than their own values of mediator and each subject i is given a weight $\frac{P(A=a^*)}{P(A=a^|c_i)}$.
3. For $E[Y_{am}]$, estimate it by taking a weighted average of the predicted Y values for subjects with $A = a$ if the subjects had had mediator $M = m$ rather than their own values of mediator and each subject i is given a weight $\frac{P(A=a)}{P(A=a|c_i)}$.
4. For $E[Y_{a^*Ma^*}]$, estimate it by taking a weighted average of the subjects with $A = a^*$ and each subject i is given a weight $\frac{P(A=a^*)}{P(A=a^*|c_i)}$.
5. For $E[Y_{aMa}]$, estimate it by taking a weighted average of the subjects with $A = a$ and each subject i is given a weight $\frac{P(A=a)}{P(A=a|c_i)}$.

6. For $E[Y_{aMa^*}]$, estimate it by taking a weighted average of the predicted Y values for subjects with $A = a^*$ if the subjects had had exposure $A = a$ rather than $A = a^*$ and each subject i is given a weight $\frac{P(A=a^*)}{P(A=a^*|c_i)}$.
7. For $E[Y_{a^*Ma}]$, estimate it by taking a weighted average of the predicted Y values for subjects with $A = a$ if the subjects had had exposure $A = a^*$ rather than $A = a$ and each subject i is given a weight $\frac{P(A=a)}{P(A=a|c_i)}$.

#### 2.5.2.2 Inference

Estimate standard errors of estimands by bootstrapping.

### 2.5.3 Inverse Odds Ratio Weighting Approach

Reference:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3954805/

https://www.ncbi.nlm.nih.gov/pubmed/25693776

#### 2.5.3.1 Estimation

**Closed-form Parameter Function Estimation**

1. Fit a regression model for A given M and C
2. Calculate weights for each subject, $w_i = \frac{f_{A|M,C}(A=0|M_i,C_i)}{f_{A|M,C}(A=A_i|M_i,C_i)}$.
3. Estimate the direct effect by a weighted regression model of Y on A and C using the weights calculated in 2. The estimated direct effect is the coefficient of A in this regression model.
4. Estimate the total effect by a regression model of Y on A and C. The estimated total effect is the coefficient of A in this regression model.
5. Calculate the indirect effect by subtracting the direct effect from the total effect.

#### 2.5.3.2 Inference

Estimate standard errors of estimands by bootstrapping.

### 2.5.4 Natural Effect Model

Incorporate the Medflex package: https://www.jstatsoft.org/article/view/v076i11

#### 2.5.4.1 Estimation

1. Fit a working model for Y on A, M and C.
2. For each subject i, expand the dataset by setting $x^*$ to be the observed exposure value, setting x to enumerate all potential exposure levels, and then imputing $Y_{x,M_{x^*},i}$ using the predicted value of Y under $A = x, M = M_i, C = C_i$. If the exposure is continuous, expand the dataset by setting x to be a number of draws(defaults to 5) from the conditional distribution of A given $C_i$.
3. Fit a natural effect model for Y on x,$x^*$ and C using the expanded dataset. The natural effect model should at least reflect the structure of the working model.

4. The coefficient of x captures the natural direct effect and the coefficient of $x^*$ captures the natural indirect effect.

#### 2.5.4.2 Inference

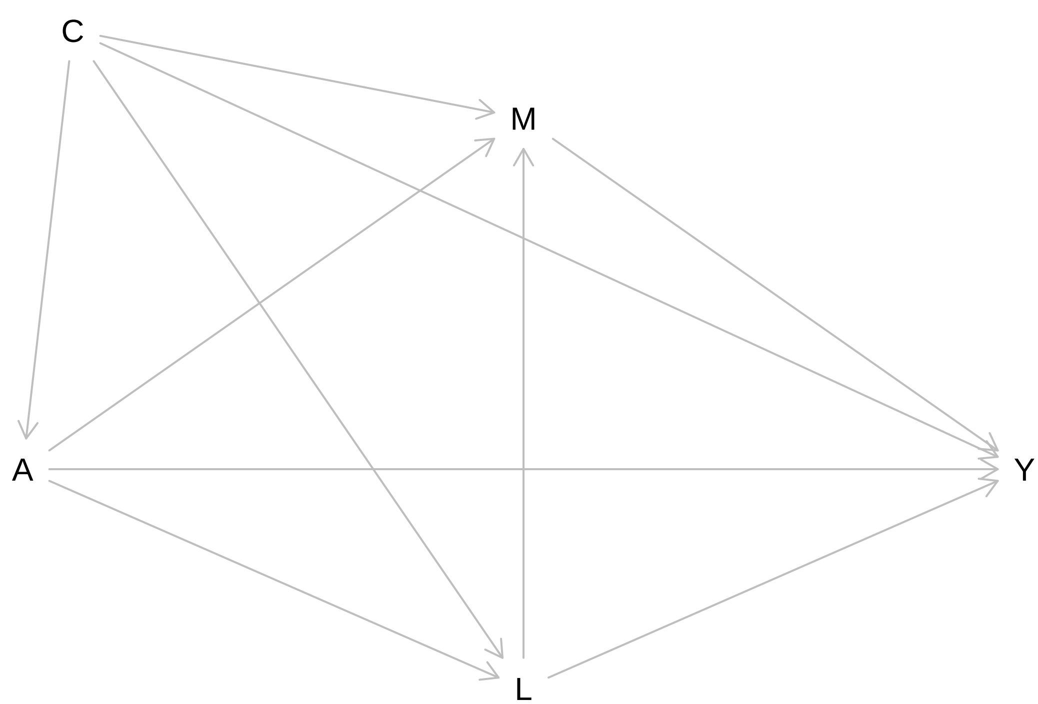
The standard errors of estimands can be estimated by the delta method or bootstrapping.

### 2.5.5 Other Approaches

Approaches talked about later can also be used.

# 3 Post-treatment Confounding

## 3.1 DAG




In the DAG, A denotes the treatment, Y denotes the outcome, M denotes a set of mediators and $M = (M_1, M_2, ..., M_k)$, C denotes a set of pre-exposure covariates and L denotes a set of post-exposure covariates and $L = (L_1, L_2, ..., L_s)$.

## 3.2 Estimand

**2-way decomposition in additive scale**

$CDE = E[Y_{am} - Y_{a^*m}]$

$rPNDE = E[Y_{aG^*_a} - Y_{a^*G^*_a}]$

$rTNDE = E[Y_{aG_a} - Y_{a^*G_a}]$

$rPNIE = E[Y_{a^*G_a} - Y_{a^*G_a^*}]$

$rTNIE = E[Y_{aG_a} - Y_{aG_a^*}]$

$rTE = rPNDE + rTNIE$

$PM = \frac{rTNIE}{rPNDE+rTE}$

**2-way decomposition in RR scale**

$rr^{CDE} = E[Y_{am}]/E[Y_{a^*m}]$

$rr^{rPNDE} = E[Y_{aG_a^*}]/E[Y_{a^*G_a^*}]$

$rr^{rTNDE} = E[Y_{aG_a}]/E[Y_{a^*G_a}]$

$rr^{rPNIE} = E[Y_{a^*G_a}]/E[Y_{a^*G_a^*}]$

$rr^{rTNIE} = E[Y_{aG_a}]/E[Y_{aG_a^*}]$

$rr^{rTE} = rr^{rPNDE} \times rr^{rTNIE}$

$PM = \frac{rr^{rPNDE}*(rr^{rTNIE}-1)}{rr^{rTE}-1}$

**4-way decomposition in additive scale**

$CDE = E[Y_{am} - Y_{a^*m}]$

$rINT_{ref} = rPNDE - CDE$

$rINT_{med} = rTNIE - rPNIE$

$rPIE = rPNIE$

$prop^{CDE} = \frac{CDE}{rTE}$

$prop^{rINT_{ref}} = \frac{rINT_{ref}}{rTE}$

$prop^{rINT_{med}} = \frac{rINT_{med}}{rTE}$

$prop^{rPIE} = \frac{rPIE}{rTE}$

$overall^{PM} = \frac{rPNIE+rINT_{med}}{rTE}$

$overall^{INT} = \frac{rINT_{ref}+rINT_{med}}{rTE}$

$overall^{PE} = \frac{rINT_{ref}+rINT_{med}+rPIE}{rTE}$

**4-way decomposition in RR scale**

$err^{CDE} = (E[Y_{am} - Y_{a^*m}])/E[Y_{a^*G_a^*}]$

$err^{rINT_{ref}} = rr^{rPNDE} - 1 - err^{CDE}$

$err^{rINT_{med}} = rr^{rTNIE} * rr^{rPNDE} - rr^{rPNDE} - rr^{rPNIE} + 1$

$err^{rPIE} = rr^{rPNIE} - 1$

$err^{rTE} = err^{CDE} + err^{rINT_{ref}} + err^{rINT_{med}} + err^{rPIE} = rr^{rTE} - 1$

$prop^{err^{CDE}} = \frac{err^{CDE}}{err^{rTE}}$

$prop^{err^{rINT_{ref}}} = \frac{err^{rINT_{ref}}}{err^{rTE}}$

$prop^{err^{rINT_{med}}} = \frac{err^{rINT_{med}}}{err^{rTE}}$

$prop^{err^{rPIE}} = \frac{err^{rPIE}}{err^{rTE}}$

$overall^{PM} = \frac{err^{rPIE} + err^{rINT_{med}}}{err^{rTE}}$

$overall^{INT} = \frac{err^{rINT_{ref}} + err^{rINT_{med}}}{err^{rTE}}$

$overall^{PE} = \frac{err^{rINT_{ref}} + err^{rINT_{med}} + err^{rPIE}}{err^{rTE}}$

## 3.3 Estimation Approaches Can be Used

### 3.3.1 Marginal Structual Model

Reference:

https://journals.lww.com/epidem/fulltext/2009/01000/marginal_structural_models_for_the_estimation_of.6.aspx

#### 3.3.1.1 Estimation

**Direct Counterfactuals Imputation Estimation**

Estimation Algorithm for $E(Y_{am})$:

1. Fit a weighted regression model for Y on A and M where each subject i is given a weight $\frac{P(A=a_i)}{P(A=a_i|C=c_i)} \frac{P(M_1=M_{1,i}|A=a_i)}{P(M_1=M_{1,i}|A=a_i,C=c_i,L=l_i)} \cdots \frac{P(M_k=M_{k,i}|A=a_i)}{P(M_k=M_{k,i}|A=a_i,C=c_i,L=l_i)}$. Then, simulate $Y_{am}$ by the predicted value of the weighted outcome regression model under $A = a, M = m$.
2. Estimate $E(Y_{am})$ by $Y_{am}$.

Estimation Algorithm for $E(Y_{a1Ga2})$:

1. Fit a weighted regression model for each mediator $M_p, p = 1, 2, ..., k$, on A and C where each subject i is given a weight $\frac{P(A=a_i)}{P(A=a_i|C=c_i)}$. Then, for each subject i, simulate the value of $M_{p,a2,i}$ by the predicted value of $M_p|A = a2, C = C_i$.
2. Fit a weighted regression model for Y on A, M and C where each subject i is given a weight $\frac{P(A=a_i)}{P(A=a_i|C=c_i)} \frac{P(M_1=M_{1,i}|A=a_i)}{P(M_1=M_{1,i}|A=a_i,C=c_i,L=l_i)} \cdots \frac{P(M_k=M_{k,i}|A=a_i)}{P(M_k=M_{k,i}|A=a_i,C=c_i,L=l_i)}$. Then, for each subject i, simulate the potential outcome $Y_{a1Ga2,i}$ by the predicted value of $Y|A = a1, M_1 = M_{1,a2,i}, M_2 = M_{2,a2,i}, ..., M_k = M_{k,a2,i}, C = C_i$.
3. Estimate $E(Y_{a1Ga2})$ by $\frac{\sum_{i=1}^{N} Y_{a1Ga2,i}}{n}$.

#### 3.3.1.2 Inference

Estimate standard errors of estimands by bootstrapping.

### 3.3.2 G-formula Approach

Reference:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5285457/

#### 3.3.2.1 Estimation

**Direct Counterfactuals Imputation Estimation**

Estimation Algorithm for $E(Y_{am})$:

1. Fit a model for E(Y|A,L,M,C). For subject i, simulate $Y_{am,i}$ by the predicted value of the outcome regression model under $A = a, L = L_i, M = m, C = C_i$.
2. Estimate $E(Y_{am})$ by $\frac{\sum_{i=1}^{N} Y_{am,i}}{n}$

Estimation Algorithm for $E(Y_{a1Ga2})$:

1. Fit a regression model for each post-exposure covariate $L_q|A, C), q = 1, 2, ..., s$, on A and C. Then, for each subject i, simulate the value of $L_{q,a1,i}$ by the predicted value of $L_q|A = a1, C = C_i$.
2. Fit a regression model for each mediator $M_p, p = 1, 2, ..., k$, on A, C and L. Then, for each subject i, simulate the value of $M_{p,a2,i}$ by the predicted value of $M_p|A = a2, C = C_i, L = L_i$.
3. Fit a regression model for Y on A, L, M, and C. Then, for each subject i, simulate the potential outcome $Y_{a1Ga2,i}$ by the predicted value of $Y|A = a1, L_1 = L_{1,a1,i}, L_2 = L_{2,a1,i}, ..., L_s = L_{s,a1,i}, M_1 = M_{1,a2,i}, M_2 = M_{2,a2,i}, ..., M_n = M_{k,a2,i}, C = C_i$.
4. Estimate $E(Y_{a1Ga2})$ by $\frac{\sum_{i=1}^{N} Y_{a1Ga2,i}}{n}$.

#### 3.3.2.2 Inference

Estimate standard errors of estimands by bootstrapping.