

# p8122\_yc4384\_HW6

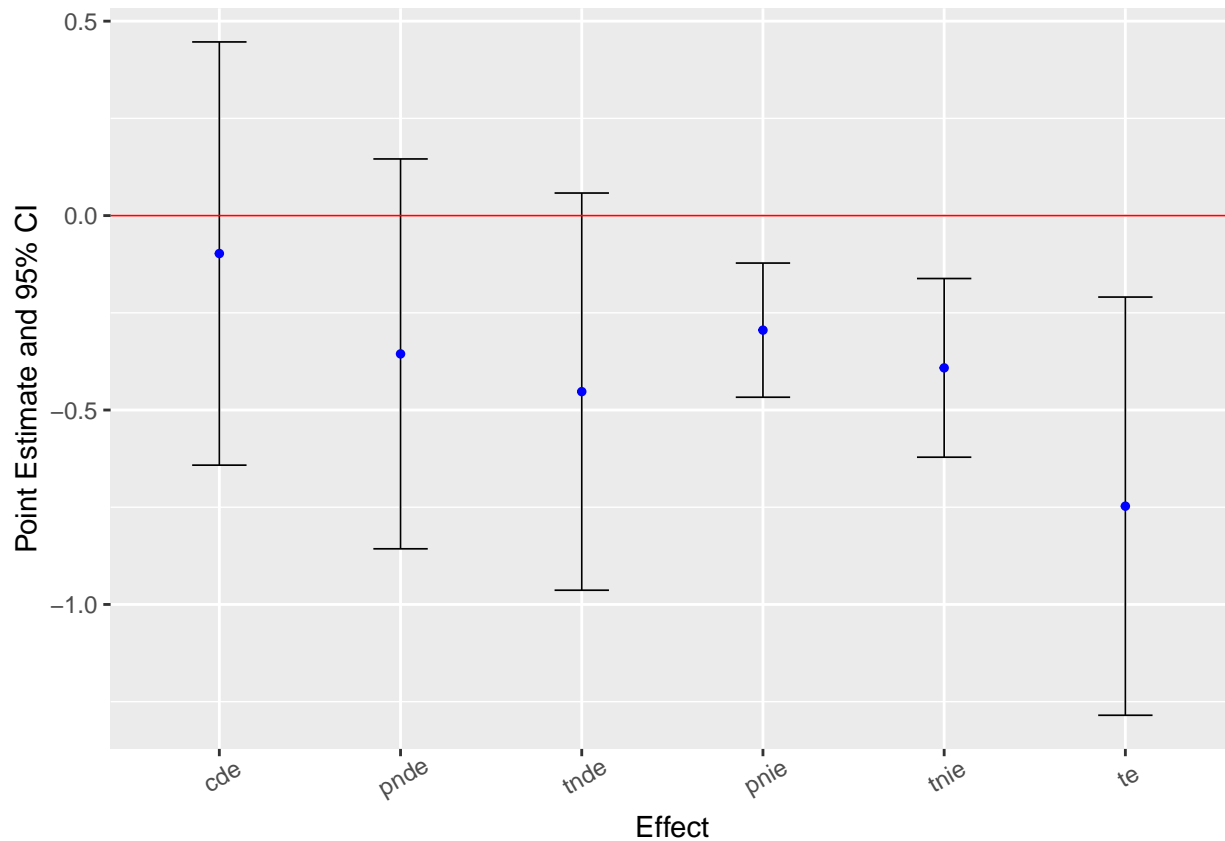
yc4384\_Yangyang\_Chen

2024-12-09

## Question 1

Mediation Analysis

```
## run mediation analysis with interaction using cmest()
mediation.int.m1 <- cmest(data = data, model = "rb",
  outcome = "cognitive_raw",
  exposure = "ln_mn_c",
  mediator = "birthlength_c", EMint = TRUE,
  basec = c("female", "approxage", "protein_c"),
  mreg = list("linear"), yreg = "linear",
  a = 1, astar = 0, mval = list(1),
  estimation = "paramfunc", inference = "delta", full=FALSE)
#summary(mediation.int.m1)
ggcmest(mediation.int.m1) +
  ggplot2::theme(axis.text.x = ggplot2::element_text(angle = 30, vjust = 0.8))
```



```
summary(mediation.int.m1)
```

```
## Causal Mediation Analysis
##
## # Outcome regression:
##
## Call:
## glm(formula = cognitive_raw ~ ln_mn_c + birthlength_c + ln_mn_c *
##     birthlength_c + female + approxage + protein_c, family = gaussian(),
##     data = getCall(x$reg.output$yreg)$data, weights = getCall(x$reg.output$yreg)$weights)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    56.15984    1.36760   41.064 < 2e-16 ***
## ln_mn_c        -0.37497    0.25480   -1.472  0.14176
## birthlength_c    0.84189    0.12446    6.764 3.81e-11 ***
## female         -0.08317    0.48416   -0.172  0.86367
## approxage       0.12095    0.05800    2.085  0.03756 *
## protein_c       0.15433    0.05357    2.881  0.00414 **
## ln_mn_c:birthlength_c 0.27738    0.12058    2.300  0.02185 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 28.95798)
##
## Null deviance: 16763  on 499  degrees of freedom
```

```

## Residual deviance: 14276  on 493  degrees of freedom
## AIC: 3110.8
##
## Number of Fisher Scoring iterations: 2
##
## # Mediator regressions:
##
## Call:
## glm(formula = birthlength_c ~ ln_mn_c + female + approxage +
##      protein_c, family = gaussian(), data = getCall(x$reg.output$mreg[[1L]])$data,
##      weights = getCall(x$reg.output$mreg[[1L]])$weights)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.377127   0.495002   0.762 0.446500
## ln_mn_c      -0.349862   0.090862  -3.850 0.000133 ***
## female       -0.275253   0.174827  -1.574 0.116028
## approxage    -0.006579   0.021002  -0.313 0.754205
## protein_c     0.098298   0.018891   5.203 2.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 3.79841)
##
##      Null deviance: 2048.0  on 499  degrees of freedom
## Residual deviance: 1880.2  on 495  degrees of freedom
## AIC: 2093.2
##
## Number of Fisher Scoring iterations: 2
##
## # Effect decomposition on the mean difference scale via the regression-based approach
##
## Closed-form parameter function estimation with
## delta method standard errors, confidence intervals and p-values
##
##      Estimate Std.error 95% CIL 95% CIU    P.val
## cde -0.09759   0.27764 -0.64176  0.447 0.725209
## pnde -0.35562   0.25576 -0.85691  0.146 0.164393
## tnde -0.45267   0.26057 -0.96337  0.058 0.082347 .
## pnle -0.29455   0.08802 -0.46706 -0.122 0.000819 ***
## tnle -0.39159   0.11721 -0.62132 -0.162 0.000835 ***
## te   -0.74721   0.27436 -1.28494 -0.209 0.006459 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (cde: controlled direct effect; pnle: pure natural direct effect; tnle: total natural direct effect;
##
## Relevant variable values:
## $a
## [1] 1
##
## $astar

```

```
## [1] 0
##
## $mval
## $mval[[1]]
## [1] 1
##
##
## $basecval
## $basecval[[1]]
## [1] 0.5
##
## $basecval[[2]]
## [1] 23.00725
##
## $basecval[[3]]
## [1] -0.1868748
```

Therefore,

- Natural Direct Effect: -0.453
- Natural Indirect Effect: -0.392
- Total Effect: -0.747

## Question 2

### Interpretation of Effects

1. Natural Direct Effect (NDE): -0.453

$$NDE = \sum_{m,c} E[Y|A = a, m, c] - E[Y|A = a^*, m, c]P(M = m|A = a^*, c)P(c) = -0.453$$

- This represents the portion of the effect of manganese exposure on child cognitive development that is not mediated by birth length (fetal growth). Specifically, for a one-unit increase in log-transformed manganese exposure (holding birth length constant), the cognitive development score decreases by approximately 0.45 units.

2. Natural Indirect Effect (NIE): -0.392

$$NIE = \sum_{m,c} E[Y|A = a, m, c]P(M = m|A = a, c) - P(M = m|A = a^*, c)P(c)$$

- This represents the portion of the effect that is mediated through birth length. In other words, the cognitive development score decreases by approximately 0.39 units as a result of the effect of manganese exposure on birth length, which subsequently affects cognitive development.

3. Total Effect (TE): -0.747

- The total effect represents the overall impact of manganese exposure on cognitive development. This combines both the direct and indirect effects, with the cognitive development score decreasing by approximately 0.75 units for a one-unit increase in manganese exposure.

### Assumptions for Causal Interpretation

- No unmeasured exposure-outcome confounding given ( C ):

$$Y_{am} \perp A \mid C$$

- No unmeasured mediator-outcome confounding given ( C ):

$$Y_{am} \perp M \mid C, A$$

- No unmeasured exposure-mediator confounding given ( C ):

$$M_a \perp A \mid C$$

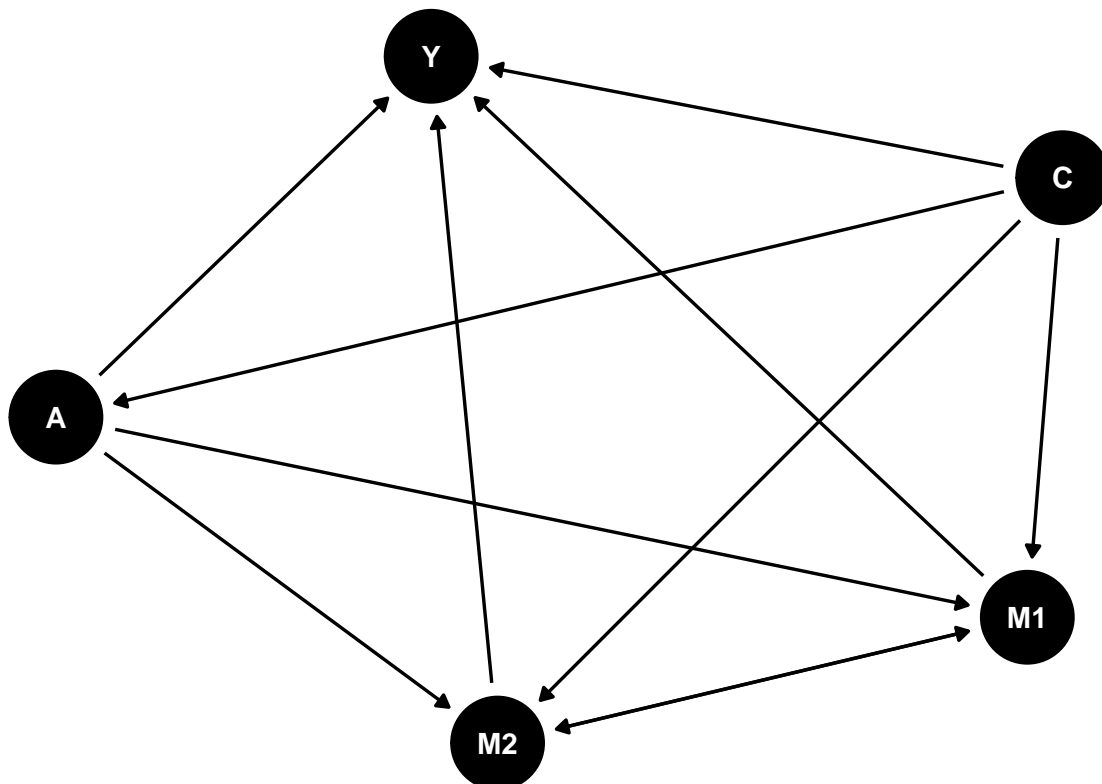
- No effect of exposure that confounds the mediator-outcome relationship:

$$Y_{am} \perp M_{a^*} \mid C$$

### Question 3

(a)

```
dag = dagify(  
  Y ~ A + M1 + M2 + C,  
  M1 ~ A + M2 + C,  
  M2 ~ A + M1 + C,  
  A ~ C  
)  
ggdag(dag) + theme_dag()
```



With the following notations:

- A: log and centered manganese levels
- M1: centered birth length
- M2: birth weight
- Y: Bayley cognitive score
- C: sex, age, protein intake score

**(b) Does adjustment for birthweight lead to a violation of the fourth identifiability assumption? Why or why not?**

- i. The Fourth Identifiability Assumption: This assumption states that there should be no unmeasured confounders for the mediator-outcome relationship conditional on the exposure and covariates (i.e.,  $Y_{a,m} \perp M^* | C$  ).
- ii. Adjustment for birthweight ( $M^*$ ) could **violate** this assumption if:
  - Birthweight is a collider: Conditioning on birthweight might open a backdoor path, introducing spurious associations between manganese exposure and cognitive development. This could bias the effect estimates.
  - Birthweight and birth length are highly correlated: Adjusting for one mediator could interfere with the independent pathways through which the other mediator operates.
  - Unmeasured confounders between birthweight and cognitive development: If such confounders exist, adjusting for birthweight would amplify bias because the assumption of no confounding would be violated.