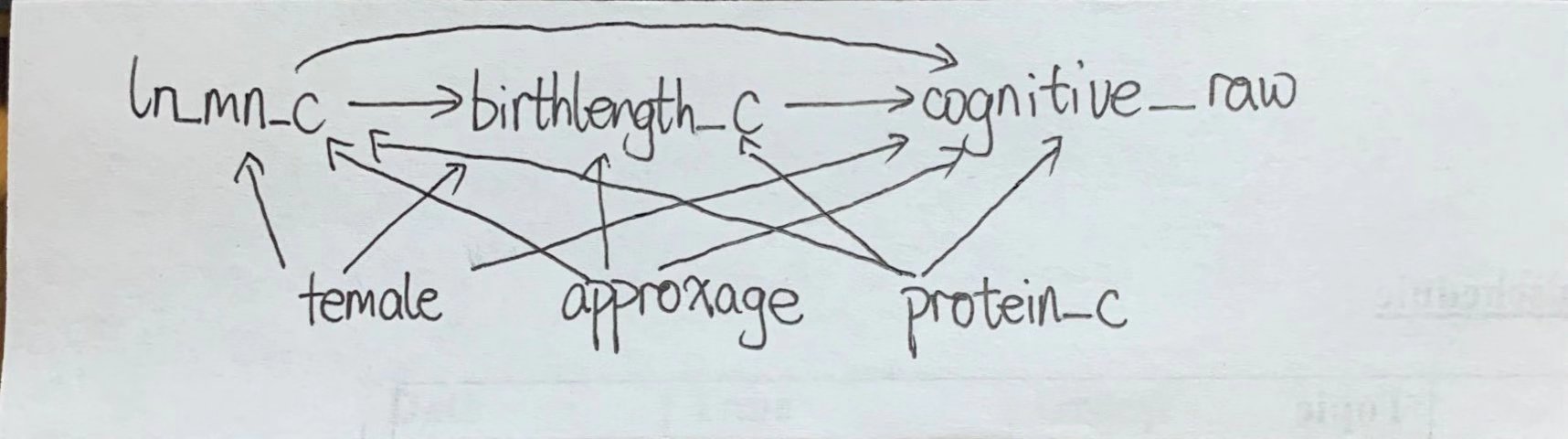
Final

Xinyi Lin

12/5/2019

## Part 1

### Question 1a



Question 1a

### Question 1b

## 25% 75%   
## -0.6099775 0.7058503

Regress Y on A and C:

##   
## Call:  
## lm(formula = cognitive\_raw ~ ln\_mn\_c + female + approxage + protein\_c,   
## data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.1104 -3.5195 0.2701 3.5397 15.4567   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.45956 1.43639 39.307 < 2e-16 \*\*\*  
## ln\_mn\_c -0.65445 0.26366 -2.482 0.0134 \*   
## female1 -0.35578 0.50731 -0.701 0.4834   
## approxage 0.11330 0.06094 1.859 0.0636 .   
## protein\_c 0.23859 0.05482 4.353 1.64e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.655 on 495 degrees of freedom  
## Multiple R-squared: 0.05556, Adjusted R-squared: 0.04793   
## F-statistic: 7.28 on 4 and 495 DF, p-value: 1.057e-05

Regress Y on A, M and C:

##   
## Call:  
## lm(formula = cognitive\_raw ~ ln\_mn\_c + birthlength\_c + female +   
## approxage + protein\_c, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.2818 -3.6632 0.0347 3.5259 14.7940   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.13384 1.37348 40.870 < 2e-16 \*\*\*  
## ln\_mn\_c -0.35228 0.25571 -1.378 0.16894   
## birthlength\_c 0.86369 0.12464 6.929 1.32e-11 \*\*\*  
## female1 -0.11804 0.48602 -0.243 0.80820   
## approxage 0.11898 0.05825 2.043 0.04161 \*   
## protein\_c 0.15369 0.05380 2.857 0.00446 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.405 on 494 degrees of freedom  
## Multiple R-squared: 0.1392, Adjusted R-squared: 0.1305   
## F-statistic: 15.98 on 5 and 494 DF, p-value: 1.311e-14

Regress M on A and C

##   
## Call:  
## lm(formula = birthlength\_c ~ ln\_mn\_c + female + approxage + protein\_c,   
## data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.1670 -1.2018 0.1061 1.1701 6.6815   
##   
## 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 \*\*\*  
## female1 -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  
##   
## Residual standard error: 1.949 on 495 degrees of freedom  
## Multiple R-squared: 0.08192, Adjusted R-squared: 0.0745   
## F-statistic: 11.04 on 4 and 495 DF, p-value: 1.382e-08

The total effect of one unit change in log and centered manganese level is -0.654 and the total effect of a change in log and centered manganese levels from 25th to the 75th percentile is -0.861.

The direct effect of one unit change in log and centered manganese level is -0.352 and the direct effect of a change in log and centered manganese levels from 25th to the 75th percentile is -0.464.

The indirect effect of difference method for one unit change in log and centered manganese level is -0.302 and the indirect effect of difference method for a change in log and centered manganese levels from 25th to the 75th percentile is -0.398.

The indirect effect of product method for one unit change in log and centered manganese level is -0.302 and the indirect effect of product method for a change in log and centered manganese levels from 25th to the 75th percentile is -0.39738.

### Question 1c

The model including interaction

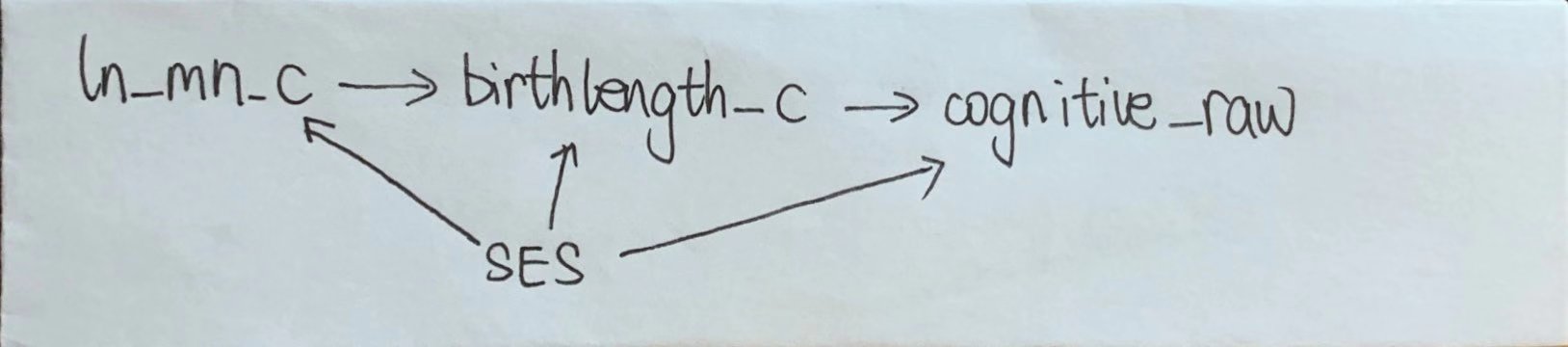
##   
## Call:  
## lm(formula = cognitive\_raw ~ ln\_mn\_c + birthlength\_c + ln\_mn\_c \*   
## birthlength\_c + female + approxage + protein\_c, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.965 -3.333 -0.041 3.528 15.159   
##   
## Coefficients:  
## Estimate Std. Error 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 \*\*\*  
## female1 -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  
##   
## Residual standard error: 5.381 on 493 degrees of freedom  
## Multiple R-squared: 0.1484, Adjusted R-squared: 0.138   
## F-statistic: 14.32 on 6 and 493 DF, p-value: 4.545e-15

The estimator for the interaction between the manganese exposure and birth length in a linear regression adjusted for the covariates is 0.277. The 95% confidence interval is (0.042, 0.512).

### Question 1d

If we believe there is interaction between the manganese exposure and birth length and include interaction term in our model, then the product and difference method estimators are not valid in this context as it is unclear how to handle the interaction coefficient and two methods provide different results. If we believe there is no interaction between the manganese exposure and birth length and there is no need to include interaction term in our model, then two estimators are valid.

### Question 2a



Question 2a

Identifiability assumptions:

1. No unmeasured exposure-outcome confounding given C
2. No unmeasured mediator-outcome confounding given C
3. No unmeasured exposure-mediator confounding given C
4. No effect of exposure that confounds the mediator-outcome relationship

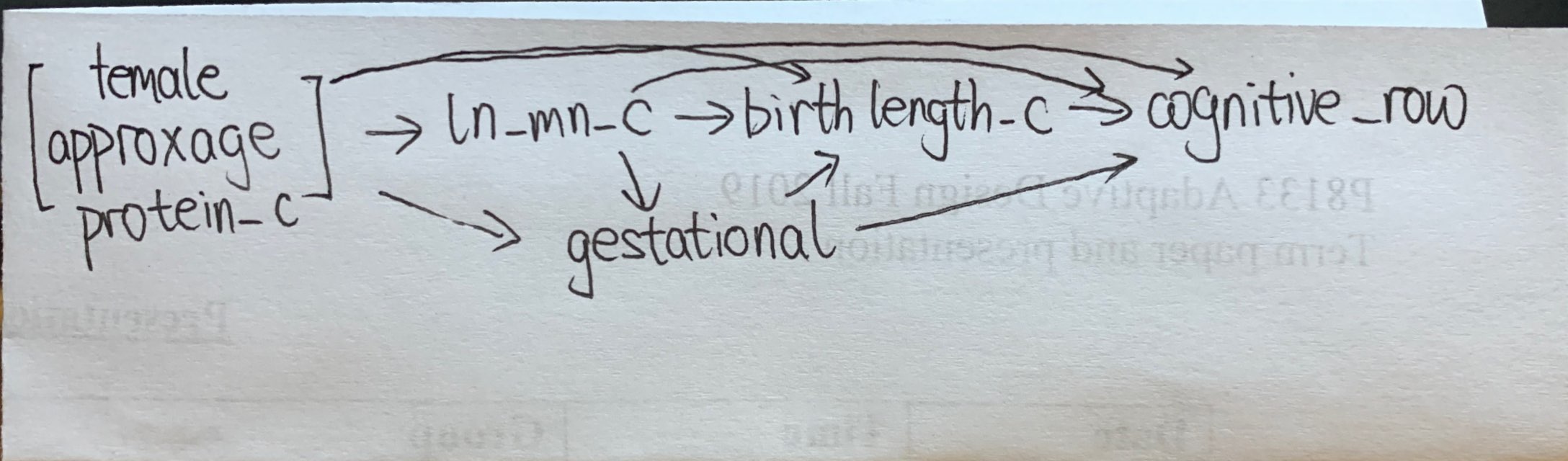
To estimate natural direct and indirect effects, we need above four assumptions. To estimate the controlled direct effect conditional on the covariates, we need assumption 1 and 2. As there are unmeasured confounder SES, assumptions 1,2 and 3 are violated, thus natural direct effect, natural indirect effect and controlled direct effects cannot be identified.

### Question 2b

Hypothesize: The direction of SES effect on exposure is negative. The direction of SES effect on mediator is positive. The direction of SES effect on outcome is positive.

As the direction of SES effect on exposure is negative and the direction of SES effect on outcome is positive, the direction of confounding bias for total effects is negative. As the direction of SES effect on mediator is positive, the direction of confounding bias for indirect effects is positive. Thus, the direction of confounding bias for direct effects is negative.

### Question 3a



Question 3a

### Question 3b

The natural direct and indirect effects are identified in the DAG, becuase following assumptions are not violated.

1. No unmeasured exposure-outcome confounding given C
2. No unmeasured mediator-outcome confounding given C
3. No unmeasured exposure-mediator confounding given C
4. No effect of exposure that confounds the mediator-outcome relationship

### Question 3c

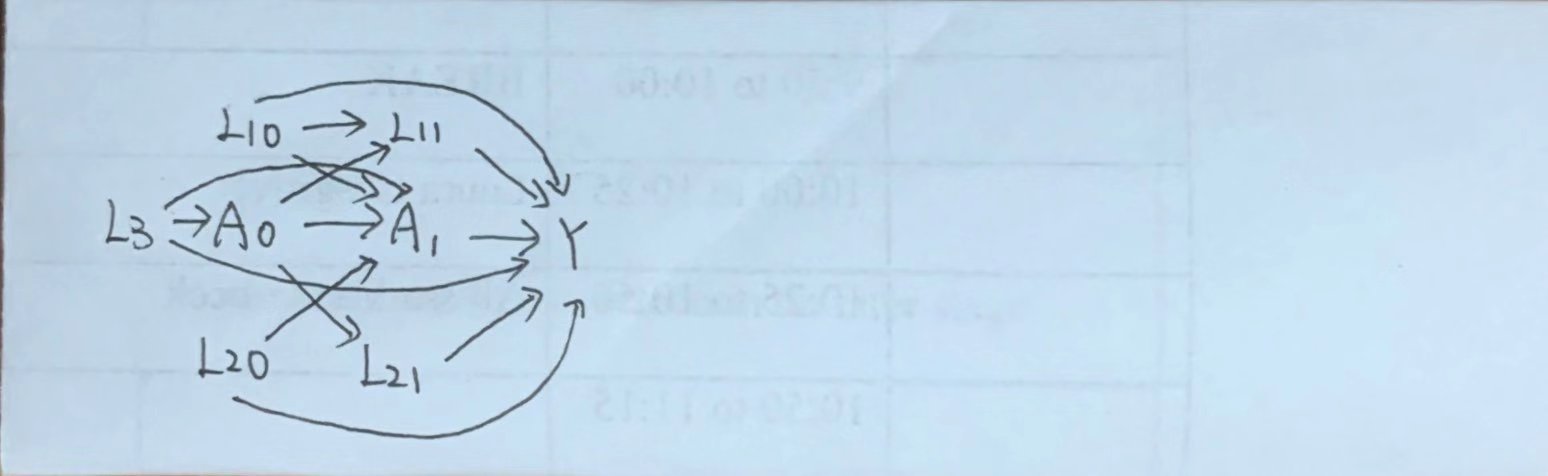
Then, direct effect is , indirect effect is .

## Part 2

### Question 1

1. Time point 0: L1, L2, L3, A
2. Time point 1: L1, L2, A
3. Time point 2: Y

### Question 2



Part 2 Question 2

### Question 3

The causal contrast means the causal effect comparing to .

### Question 4

In order to identify the causal contrast of interest, we need to know direct effect and indirect effec, thus following no unmeasured confounding assumptions need to hold:

1. exchangability, positivity, consistency and SUTVA.
2. At baseline: . At time point 1:
3. No unmeasured exposure-outcome confounding given C
4. No unmeasured mediator-outcome confounding given C
5. No unmeasured exposure-mediator confounding given C
6. No effect of exposure that confounds the mediator-outcome relationship

### Question 5

As we are interested in the treatment effect of overall treatment process influenced by both treatmengt at baseline and time point 1, I choose following marginal model:

Then causal contrast equals to . Using bootstrap, we can get estimates and 95% confidence interval as following:

|  |  |  |  |
| --- | --- | --- | --- |
| name | Estimate | CIL | CIU |
| beta | -4.1961 | -4.3905 | -3.9959 |
| beta0 | 0.0339 | -0.1163 | 0.1777 |
| beta1 | -0.3157 | -0.4656 | -0.1641 |
| Causal Contrast | -0.2818 | -0.5522 | -0.0266 |

So the estimate of causal contrast is -0.2818, the 95% confidence interval is (-0.5522, -0.0266).

### Question 6

The causal contrast means the causal effect comparing to .

### Question 7

In order to identify the causal contrast of interest, we need to know direct effect and indirect effec, thus following no unmeasured confounding assumptions need to hold:

1. exchangability, positivity, consistency and SUTVA.
2. At baseline: . At time point 1:
3. No unmeasured exposure-outcome confounding given C
4. No unmeasured mediator-outcome confounding given C
5. No unmeasured exposure-mediator confounding given C
6. No effect of exposure that confounds the mediator-outcome relationship

### Question 8

As we are interested in the treatment effect of overall treatment process influenced by both treatmengt at baseline and time point 1, I still choose following marginal model:

Then causal contrast equals to . According to bootstrap results, we can get the estimate of causal contrast is 0.0339, the 95% confidence interval is (-0.1163, 0.1777).

### Question 9

If we treat as mediator, than means the value of mediator given exposure is control which means . And means the value of mediator given exposure is treatment which means . The natural direct effect is and the natural indirect effect is .

As we already have the information of . If we assume following assumptions hold, which means the DAG in question 1 is valid, then these causal effects are identified.

1. No unmeasured exposure-outcome confounding given C
2. No unmeasured mediator-outcome confounding given C
3. No unmeasured exposure-mediator confounding given C
4. No effect of exposure that confounds the mediator-outcome relationship

**Appendix**

knitr::opts\_chunk$set(echo = FALSE)  
library(tidyverse)  
library(survey)  
# Part 1  
data1 = read.csv("./data1\_final.csv") %>%   
 mutate(female = as.factor(female))  
# Question 1b  
exp\_q = quantile(data1$ln\_mn\_c, c(0.25, 0.75))  
exp\_q  
exp.model = lm(cognitive\_raw~ln\_mn\_c+female+approxage+protein\_c, data = data1)  
summary(exp.model)  
med.model = lm(cognitive\_raw~ln\_mn\_c+birthlength\_c+female+approxage+protein\_c, data = data1)  
summary(med.model)  
MA.model = lm(birthlength\_c~ln\_mn\_c+female+approxage+protein\_c, data = data1)  
summary(MA.model)  
phi1 = -0.65445  
theta1 = -0.35228  
theta2 = 0.86369  
beta1 = -0.349862  
dif\_ind = phi1-theta1  
prd\_ind = theta2\*beta1  
exp\_chg = exp\_q[2]-exp\_q[1]  
# Question 1c  
int.model = lm(cognitive\_raw~ln\_mn\_c+birthlength\_c+ln\_mn\_c\*birthlength\_c+female+approxage+protein\_c, data = data1)  
summary(int.model)  
CIL = 0.277-1.96\*0.12  
CIU = 0.277+1.96\*0.12  
# Part 2  
data2 = read.csv("./data2\_final.csv")  
# create wide data  
AC\_data = data2 %>%  
 gather(key = "Lt", value = "value", c(L1,L2,A)) %>%  
 arrange(id) %>%  
 mutate(Lt = str\_c(Lt, t0)) %>%  
 select(id, Lt, value) %>%  
 spread(key = Lt, value = value)  
Y\_data = data2 %>%  
 select(id, Y) %>%  
 na.omit()  
L3\_data = data2 %>%   
 select(id,L3) %>%   
 na.omit()  
wide\_data = merge(merge(AC\_data, Y\_data),L3\_data) %>%   
 select(-c(A2,L12,L22))  
set.seed(123)  
nboots = 1000  
n\_sample = nrow(wide\_data)  
beta = rep(NA, nboots)  
beta0 = rep(NA, nboots)  
beta1 = rep(NA, nboots)  
CauEff = rep(NA, nboots)  
for (i in 1:nboots) {  
 S.b <- sample(1:n\_sample, size = n\_sample, replace = TRUE)  
 data.b <- wide\_data[S.b, ]  
 # Time point 0  
 glm.model0 = glm(A0~L3, data = data.b, family = binomial)  
 p0 = predict(glm.model0, type = "response")  
 w0 = ifelse(data.b$A0==1, 1/p0, 1/(1-p0))  
 # Time point 1  
 glm.model1 = glm(A1~L3+A0+L10+L20, data = data.b, family = binomial)  
 p1 = predict(glm.model1, type = "response")  
 w1 = ifelse(data.b$A1==1, 1/p1, 1/(1-p1))  
 w = w0\*w1  
   
 data.b$w = w  
 design = svydesign(ids = ~id, weights = ~w, data = data.b)  
 msm = svyglm(Y ~ A0 + A1, family = gaussian(link = "identity"), design = design)  
 beta[i] = msm$coef[1]  
 beta0[i] = msm$coef[2]  
 beta1[i] = msm$coef[3]  
 CauEff[i] = msm$coef[2] + msm$coef[3]  
}  
Estimate = c(mean(beta), mean(beta0), mean(beta1), mean(CauEff)) %>% round(4)  
CI = rbind(quantile(beta, probs = c(0.025, 0.975)),  
 quantile(beta0, probs = c(0.025, 0.975)),  
 quantile(beta1, probs = c(0.025, 0.975)),  
 quantile(CauEff, probs = c(0.025, 0.975))) %>% round(4)   
name = c("beta", "beta0", "beta1", "Causal Contrast")  
cbind(name, Estimate, CI) %>%   
 as.data.frame() %>%   
 mutate(CIL = `2.5%`,  
 CIU = `97.5%`) %>%   
 select(name, Estimate, CIL, CIU) %>%   
 knitr::kable()