

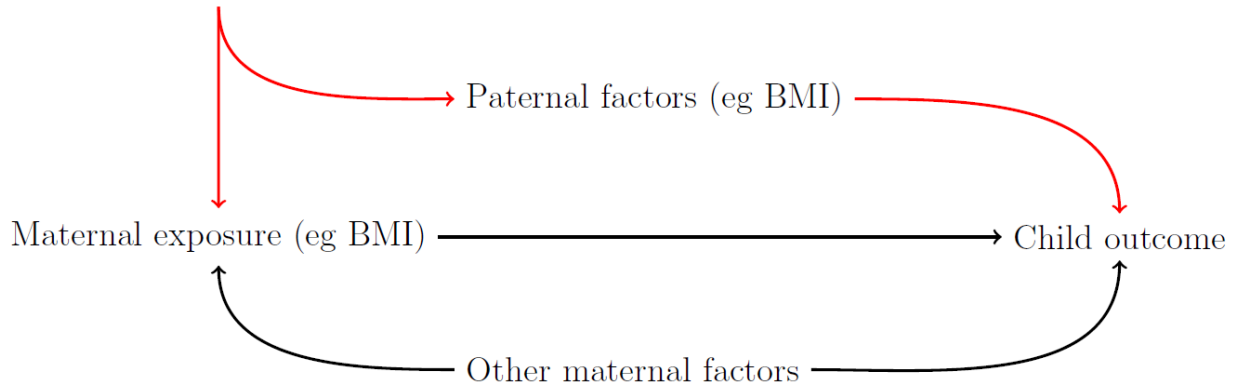
# Paternal bias

June 2019 (last update)

## Example - paternal BMI, maternal BMI, birth weight

Paternal factors are potential confounders of the effects of maternal exposures on offspring outcomes. Suppose we are interested in evaluating the effect of maternal BMI ( $X_m$ ) on offspring Birth weight ( $Y$ ). Paternal BMI ( $X_p$ ) may have an independent effect on the outcome (animal studies have first demonstrated that the sperm carries epigenetic content, and epidemiological studies are assessing several associations at the population level), and is associated with the maternal exposure (e.g. parents have similar diet, levels of physical activity, exposure to environmental factors). Thus,  $X_p$  is a potential confounder of the association between  $X_m$  and  $Y$ , as presented in the Figure below.

Shared unmeasured antecedent



Differently from maternal confounders, paternal factors are generally un-accounted for in population-based studies. We present some simulated data to evaluate the impact of omitting paternal factors when evaluating the effect of maternal exposures on offspring health.

n.b. The DAG above shows a direction from paternal factors to maternal exposures as for the common representation of confounders. However, the arrow does not assume a causal effect of paternal obesity on maternal BMI, but rather an association likely due to common antecedent (e.g. shared diet or other behavioral aspects like physical activity, or common socio/demographic characteristics).

## Simulation setting

We simulate a dataset of  $n=10,000$  pregnancies assuming a true effect of maternal BMI on BW ( $X_m \rightarrow Y$ ) of 25 grams (for any 5-units of BMI). For simplifying interpretation we assume: - paternal BMI to be a binary confounder of obese vs non-obese (with a prevalence of obesity of 40%) - the association between paternal and maternal BMI ( $X_p \rightarrow X_m : \gamma$ ) to be equal to 5 (i.e. women with obese partners have, on average, 5-point higher BMI as compared to women with non-obese partners) - the independent effect of paternal obesity on birth weight  $X_p \rightarrow Y : \delta$  to be 5 grams

```
correct<-c(rep(0),100)
estimated<-c(rep(0),100)
for (i in 1:100) {
  n<-10000
```

```

x_p<-rbinom(n,1,.4)
x_m<-rnorm(n, 27+5*x_p, 4)
y <-rnorm(n, 3000+5*x_m+5*x_p,5)
reg <- glm(y ~ x_m+x_p, family = "gaussian")
reg_unad <- glm(y ~ x_m, family = "gaussian")
correct[i]<- summary(reg)$coefficients[2]
estimated[i]<- summary(reg_unad)$coefficients[2]}
mean(correct)

```

```
## [1] 5.0016
```

```
mean(estimated)
```

```
## [1] 5.274425
```

The true effect is overestimated by 5%.

### a) Varying $\delta$ (true effect of paternal BMI on BW)

Here is how the bias increases by keeping  $\gamma = 5$  but varying  $\delta$  from 5 to 50 grams

```

a<-seq(5, 50, by = 5)
correct_res<-c(0)
estimated_res<-c(0)
for (j in a){
  correct<-c(rep(0),100)
  estimated<-c(rep(0),100)
  for (i in 1:100) {
    n<-10000
    x_p<-rbinom(n,1,.4)
    x_m<-rnorm(n, 27+5*x_p, 4)
    y <-rnorm(n, 3000+5*x_m+j*x_p,5)
    reg <- glm(y ~ x_m+x_p, family = "gaussian")
    reg_unad <- glm(y ~ x_m, family = "gaussian")
    correct<- summary(reg)$coefficients[2]
    estimated<- summary(reg_unad)$coefficients[2]}
  correct_res[which(a==j)]<-mean(correct)*5
  estimated_res[which(a==j)]<-mean(estimated)*5
}
correct_res

```

```
## [1] 25.05855 24.99995 25.00237 24.99896 25.04700 25.05387 24.93484
```

```
## [8] 25.01102 25.03216 24.98938
```

```
estimated_res
```

```
## [1] 26.40037 27.71725 29.10410 30.30425 31.83459 33.12942 34.57695
```

```
## [8] 35.95526 37.26394 38.88385
```

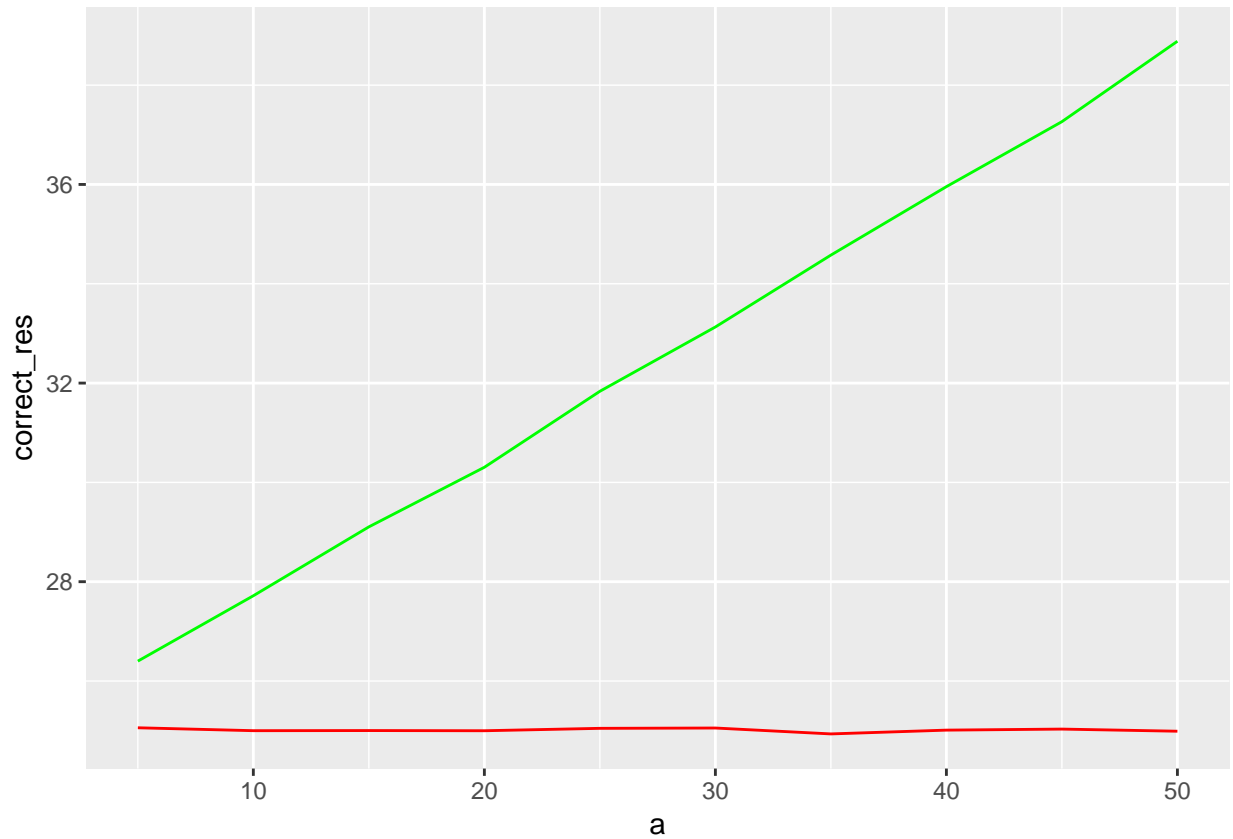
Plot

```

df <- data.frame(a,correct_res,estimated_res)

ggplot(df, aes(a)) +
  geom_line(aes(y=correct_res), colour="red") +
  geom_line(aes(y=estimated_res), colour="green")

```



## b) Varying $\gamma$ (the association between maternal and paternal BMI)

Here, on the other hand, is the bias by keeping  $\delta = 25$  but varying  $\gamma$  from 1 to 5

```
a<-seq(1, 5, by = 1)
correct_res<-c(0)
estimated_res<-c(0)
for (j in a){
  correct<-c(rep(0),100)
  estimated<-c(rep(0),100)
  for (i in 1:100) {
    n<-10000
    x_p<-rbinom(n,1,.4)
    x_m<-rnorm(n, 27+j*x_p, 4)
    y <-rnorm(n, 3000+5*x_m+25*x_p,5)
    reg <- glm(y ~ x_m+x_p, family = "gaussian")
    reg_unad <- glm(y ~ x_m, family = "gaussian")
    correct<- summary(reg)$coefficients[2]
    estimated<- summary(reg_unad)$coefficients[2]}
  correct_res[which(a==j)]<-mean(correct)*5
  estimated_res[which(a==j)]<-mean(estimated)*5
}
correct_res

## [1] 24.98873 24.87576 25.06093 25.09663 24.87729
```

```
estimated_res
```

```
## [1] 26.58729 28.52128 29.80914 31.06759 31.83714
```

Same loop, but now with  $\delta = 50$

```
a<-seq(1, 5, by = 1)
correct_res<-c(0)
estimated_res<-c(0)
for (j in a){
  correct<-c(rep(0),100)
  estimated<-c(rep(0),100)
  for (i in 1:100) {
    n<-10000
    x_p<-rbinom(n,1,.4)
    x_m<-rnorm(n, 27+j*x_p, 4)
    y <-rnorm(n, 3000+5*x_m+50*x_p,5)
    reg <- glm(y ~ x_m+x_p, family = "gaussian")
    reg_unad <- glm(y ~ x_m, family = "gaussian")
    correct<- summary(reg)$coefficients[2]
    estimated<- summary(reg_unad)$coefficients[2]}
  correct_res[which(a==j)]<-mean(correct)*5
  estimated_res[which(a==j)]<-mean(estimated)*5
}
correct_res
```

```
## [1] 24.88219 25.01700 24.91425 25.03436 25.08860
```

```
estimated_res
```

```
## [1] 28.29999 32.60496 35.27498 37.03043 38.63897
```

Same loop, but now with  $\delta = 5$

```
a<-seq(1, 5, by = 1)
correct_res<-c(0)
estimated_res<-c(0)
for (j in a){
  correct<-c(rep(0),100)
  estimated<-c(rep(0),100)
  for (i in 1:100) {
    n<-10000
    x_p<-rbinom(n,1,.4)
    x_m<-rnorm(n, 27+j*x_p, 4)
    y <-rnorm(n, 3000+5*x_m+5*x_p,5)
    reg <- glm(y ~ x_m+x_p, family = "gaussian")
    reg_unad <- glm(y ~ x_m, family = "gaussian")
    correct<- summary(reg)$coefficients[2]
    estimated<- summary(reg_unad)$coefficients[2]}
  correct_res[which(a==j)]<-mean(correct)*5
  estimated_res[which(a==j)]<-mean(estimated)*5
}
correct_res
```

```
## [1] 24.95234 24.97819 25.10619 24.92922 24.94016
```

```
estimated_res
```

```
## [1] 25.32503 25.71663 26.09590 26.16680 26.31225
```

### c) Adding an interaction between maternal and paternal BMI

In case of the further presence of a positive interaction between the two exposures, the bias will be even larger. Here, for example, are results with an interaction effect of 5 grams

```
a<-seq(5, 50, by = 5)
correct_res<-c(0)
estimated_res<-c(0)
for (j in a){
  correct<-c(rep(0),100)
  estimated<-c(rep(0),100)
  for (i in 1:100) {
    n<-10000
    x_p<-rbinom(n,1,.4)
    x_m<-rnorm(n, 27+5*x_p, 4)
    y <-rnorm(n, 3000+5*x_m+j*x_p+5+x_m*x_p,5)
    reg <- glm(y ~ x_m+x_p, family = "gaussian")
    reg_unad <- glm(y ~ x_m, family = "gaussian")
    correct<- summary(reg)$coefficients[2]
    estimated<- summary(reg_unad)$coefficients[2]}
  correct_res[which(a==j)]<-mean(correct)*5
  estimated_res[which(a==j)]<-mean(estimated)*5
}
correct_res
```

```
## [1] 27.11037 27.00266 27.03131 27.09999 27.06772 27.03680 27.03456
```

```
## [8] 26.90441 27.06673 27.08014
```

```
estimated_res
```

```
## [1] 36.70542 37.85923 39.21179 40.60771 42.39970 43.21720 45.13292
```

```
## [8] 46.14894 47.59139 48.50342
```