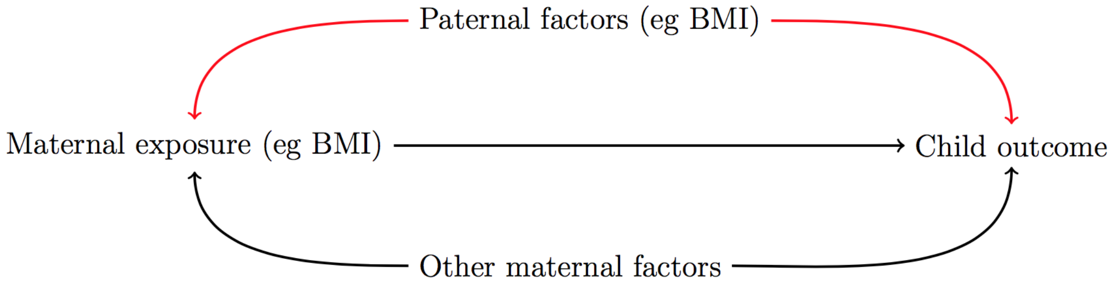
Paternal bias

February 2019

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.



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.

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, one point higher BMI as compared to women with non-obese partners) - the independent effect of paternal obesity on birth weight $X_p \rightarrow Y : \delta$ is 5 grams

```r
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%. Here is how the bias increases by keeping $\gamma = 5$ but varying λ 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
```

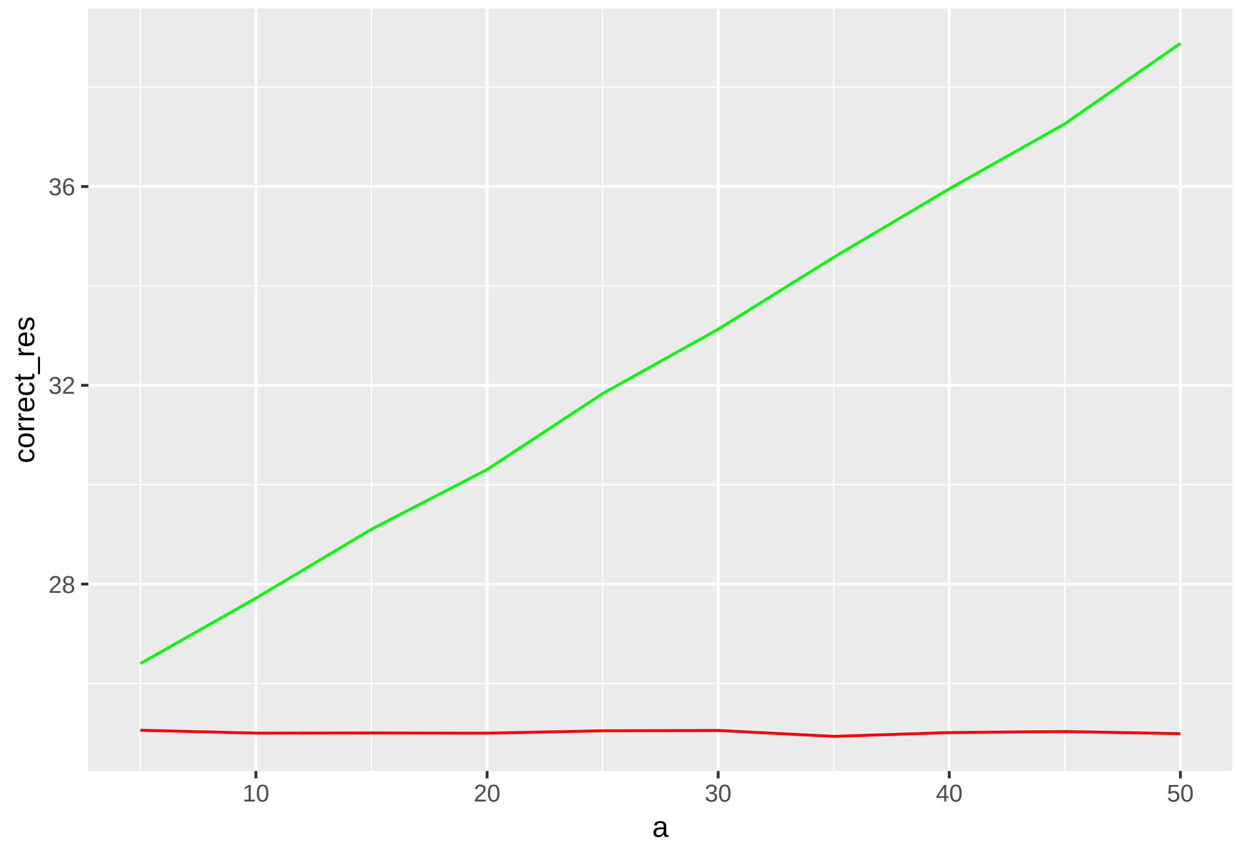
```
estimated_res
```

```
##  [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")
```



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] 26.96791 26.83002 27.02931 27.06532 26.77581 26.86939 27.01540
##  [8] 26.95590 26.99972 27.05342
```

```
estimated_res
```

```
##  [1] 36.49228 37.97308 39.13955 40.54758 42.08188 43.30697 45.08807
##  [8] 46.56581 47.44857 48.71811
```