

# Example analysis

Linear model, smoking is the exposure, birthweight in grams the outcome

```
data <- AF::clslowbwt
head(data)
```

	id	birth	smoke	race	age	lwt	bwt	low	lbw	smoker
1	1	1	1. Yes	3. Other	28	120	2865	0. No	0	1
2	1	2	1. Yes	3. Other	33	141	2609	0. No	0	1
3	2	1	0. No	1. White	29	130	2613	0. No	0	0
4	2	2	0. No	1. White	34	151	3125	0. No	0	0
5	2	3	0. No	1. White	37	144	2481	1. Yes	1	0
6	3	1	1. Yes	2. Black	31	187	1841	1. Yes	1	1

```
## propensity score fit
```

```
pwfit <- glm(smoker ~ race * age * lwt + I(age^2) + I(lwt^2), data = data,
             family = "binomial")
phat <- predict(pwfit, type = "response")
```

```
data$weight <- data$smoker / phat + (1 - data$smoker) / (1 - phat)
## outcome model
```

```
outfit <- glm(bwt ~ smoker * (race + age + lwt) + I(age^2) + I(lwt^2),
              data = data, family = "gaussian", weights = weight)
```

```
## dummy data, where we set X to 0 and 1
```

```
data0 <- data1 <- data
data0$smoker <- 0
data1$smoker <- 1
```

```

## predicted counterfactuals
Yhat0 <- predict(outfit, newdata = data0, type = "response")
Yhat1 <- predict(outfit, newdata = data1, type = "response")

ATEmean <- mean(Yhat1) - mean(Yhat0)
ATEmean

```

[1] -223.6736

```

## CAN Richard Funk?

## separate outcome models, unweighted

ofit1 <- glm(bwt ~ (race + age + lwt) + I(age^2) + I(lwt^2),
             data = subset(data, smoker == 1), family = "gaussian")

ofit0 <- glm(bwt ~ (race + age + lwt) + I(age^2) + I(lwt^2),
             data = subset(data, smoker == 0), family = "gaussian")

dr1 <- data$bwt * (data$smoker == 1) / phat -
      predict(ofit1, newdata = data1, type = "response") *
      ((data$smoker == 1) - phat) / phat
dr0 <- data$bwt * (data$smoker == 0) / (1 - phat) +
      predict(ofit0, newdata = data0, type = "response") *
      ((data$smoker == 1) - phat) / (1 - phat)

ATEfunk <- mean(dr1) - mean(dr0)
ATEfunk

```

[1] -225.5489

Bootstrap for confidence interval, not echoed

	2.5%	97.5%
	-485.73018	29.81936

**Logistic model, smoking is the exposure, low birthweight (birth weight < 2500g) the outcome**

```
## propensity score fit

pwfit <- glm(smoker ~ race * age * lwt + I(age^2) + I(lwt^2), data = data,
             family = "binomial")
phat <- predict(pwfit, type = "response")

data$weight <- data$smoker / phat + (1 - data$smoker) / (1 - phat)
## outcome model

outfit <- glm(lbw ~ smoker * (race + age + lwt) + I(age^2) + I(lwt^2),
              data = data, family = "binomial", weights = weight)
```

Warning in eval(family\$initialize): non-integer #successes in a binomial glm!

```
## dummy data, where we set X to 0 and 1

data0 <- data1 <- data
data0$smoker <- 0
data1$smoker <- 1

## predicted counterfactuals
Yhat0 <- predict(outfit, newdata = data0, type = "response")
Yhat1 <- predict(outfit, newdata = data1, type = "response")

ATEmean <- mean(Yhat1) - mean(Yhat0)
ATEmean
```

[1] 0.134932

```
RRmean <- mean(Yhat1) / mean(Yhat0)
RRmean
```

[1] 1.478392

```
## separate outcome models, unweighted

ofit1 <- glm(lbw ~ (race + age + lwt) + I(age^2) + I(lwt^2),
             data = subset(data, smoker == 1), family = "binomial")

ofit0 <- glm(lbw ~ (race + age + lwt) + I(age^2) + I(lwt^2),
             data = subset(data, smoker == 0), family = "binomial")

dr1 <- data$lbw * (data$smoker == 1) / phat -
  predict(ofit1, newdata = data1, type = "response") *
  ((data$smoker == 1) - phat) / phat
dr0 <- data$lbw * (data$smoker == 0) / (1 - phat) +
  predict(ofit0, newdata = data0, type = "response") *
  ((data$smoker == 1) - phat) / (1 - phat)

ATEfunk <- mean(dr1) - mean(dr0)
RRfunk <- mean(dr1) / mean(dr0)
ATEfunk
```

```
[1] 0.1372512
```

```
RRfunk
```

```
[1] 1.488187
```

Bootstrap for confidence interval, not echoed

```
      2.5%      97.5%
-0.0188878  0.2783308
```

```
      2.5%      97.5%
0.9394744  2.2549131
```