# Using Propensity Scores

Lucy D'Agostino McGowan

Wake Forest University

2021-09-01 (updated: 2022-06-07)

# Propensity scores

- Matching
- Weighting
- Stratification
- Direct Adjustment
- ...

# Propensity scores

**Matching**

Weighting

Stratification

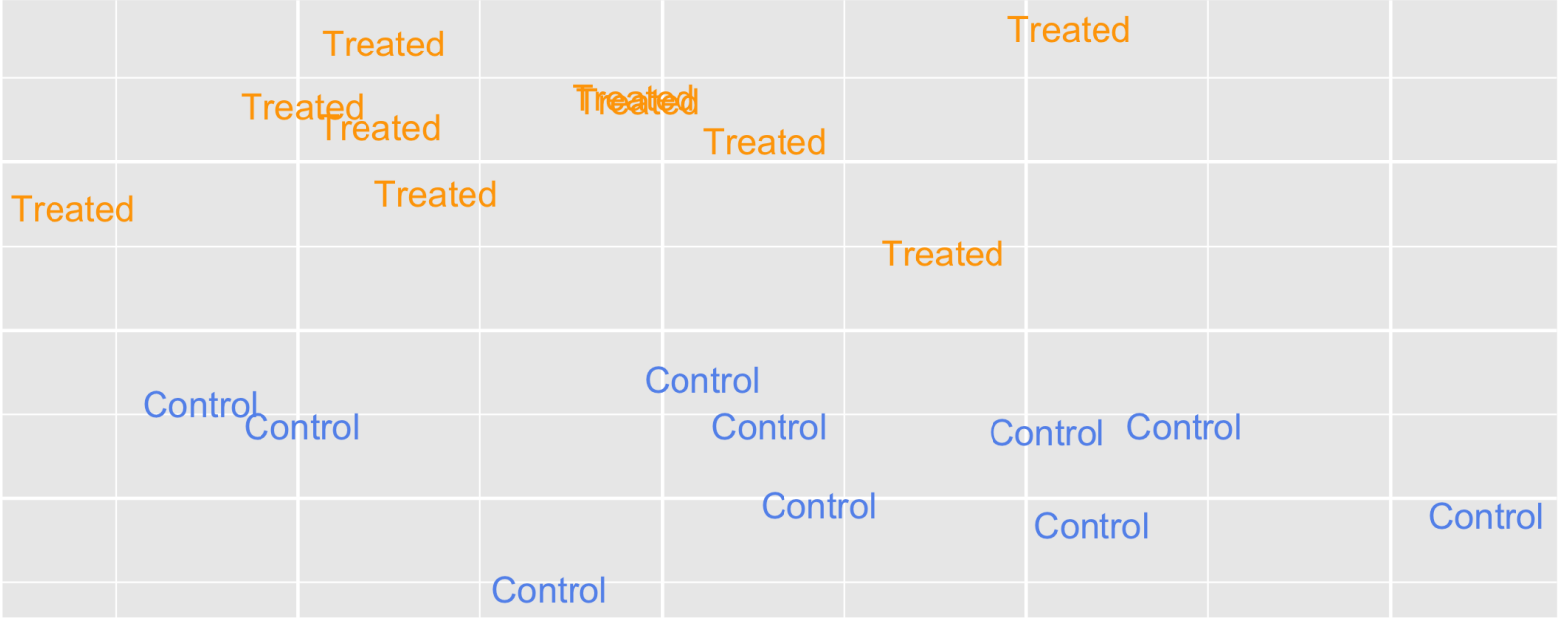
Direct Adjustment

...

# Target estimands

## Average Treatment Effect (ATE)

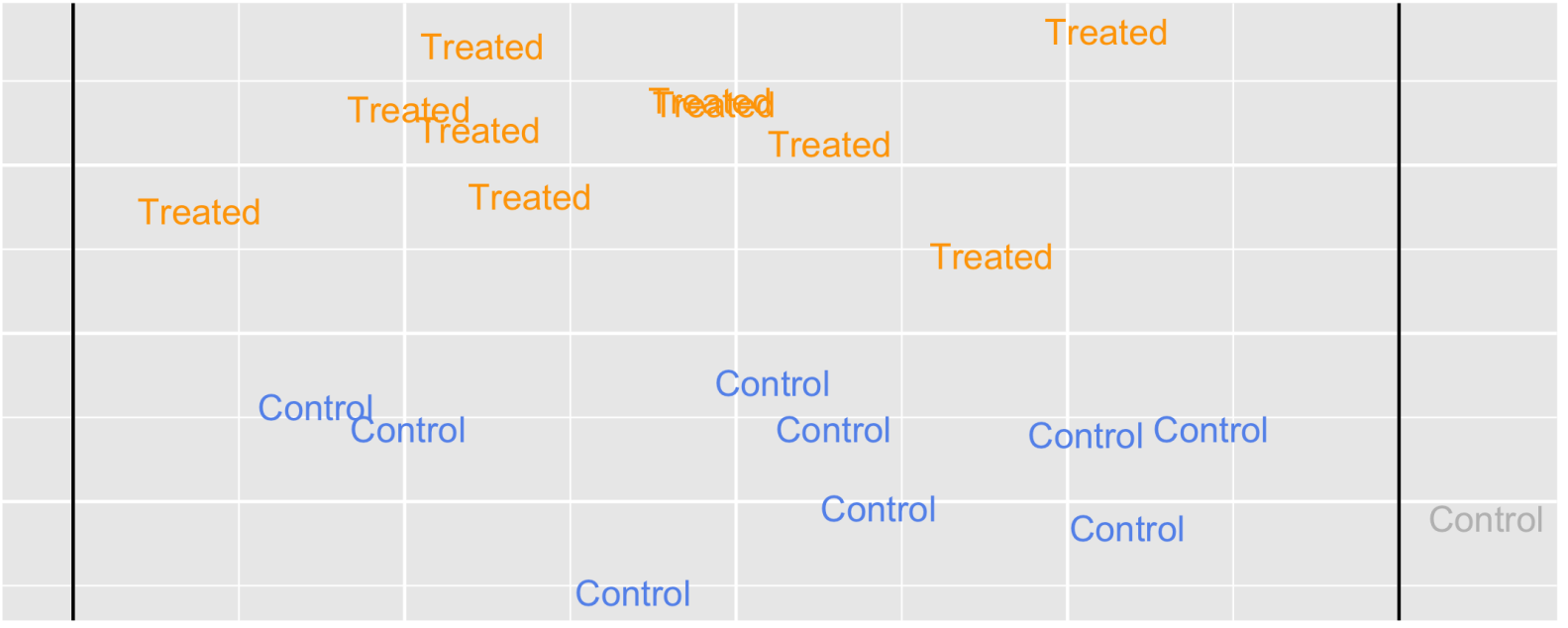
$$\tau = E[Y(1) - Y(0)]$$



# Target estimands

## Average Treatment Effect among the Treated (ATT)

$$\tau = E[Y(1) - Y(0)|Z = 1]$$



# Matching in R (ATT)

```
library(MatchIt)
m <- matchit(qsmk ~ sex +
    race + age + I(age^2) + education +
    smokeintensity + I(smokeintensity^2) +
    smokeyrs + I(smokeyrs^2) + exercise + active +
    wt71 + I(wt71^2),
  data = nhefs_complete)
m
```

```
## A matchit object
##  - method: 1:1 nearest neighbor matching without replacement
##  - distance: Propensity score
##              - estimated with logistic regression
##  - number of obs.: 1566 (original), 806 (matched)
##  - target estimand: ATT
##  - covariates: sex, race, age, I(age^2), education, smokeintensity, I
```

# Matching in R (ATT)

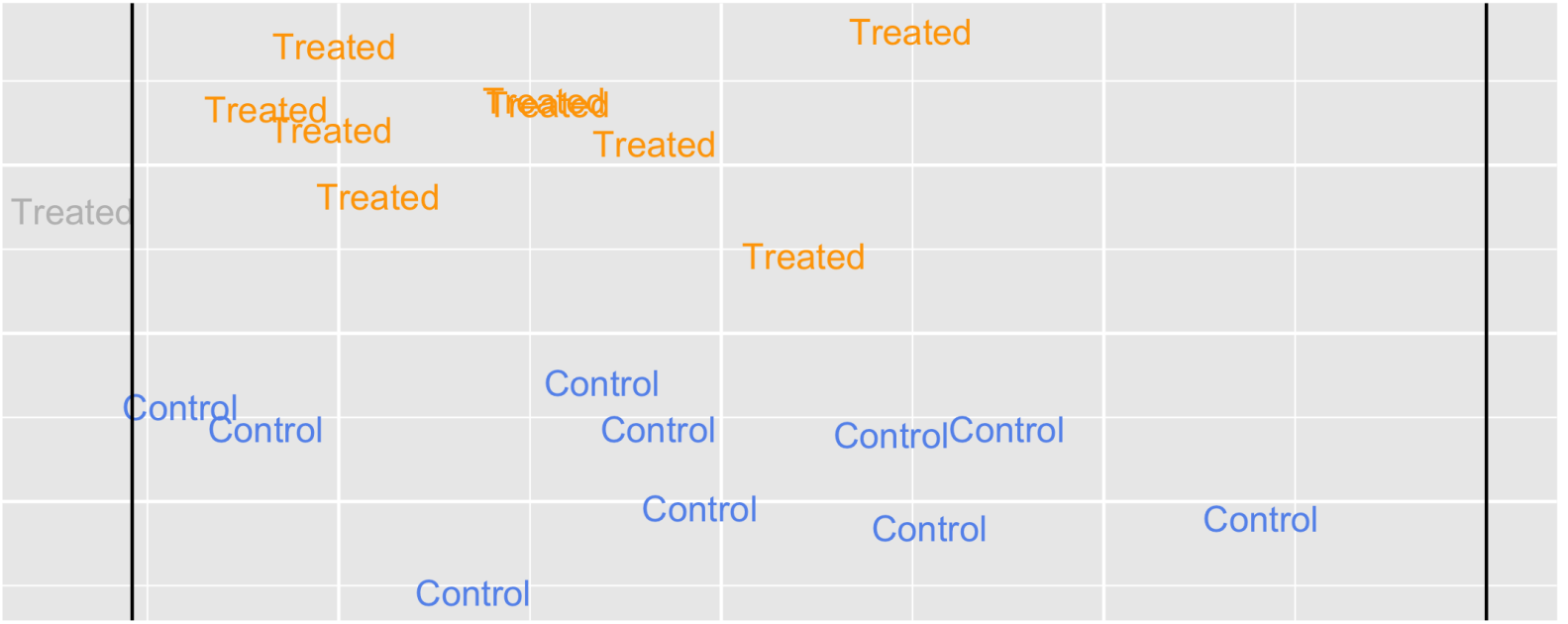
```
matched_data <- get_matches(m, id = "i")
glimpse(matched_data)
```

```
## Rows: 806
## Columns: 71
## $ i                <chr> "11", "1220", "15", "1082", "18"…
## $ subclass         <fct> 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6,…
## $ weights          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
## $ seqn             <dbl> 428, 23045, 446, 22294, 596, 140…
## $ qsmk             <dbl> 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1,…
## $ death            <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,…
## $ yrdth            <dbl> NA, NA, 88, NA, NA, NA, NA, NA, …
## $ modth            <dbl> NA, NA, 1, NA, NA, NA, NA, NA, N…
## $ dadth            <dbl> NA, NA, 3, NA, NA, NA, NA, NA, N…
## $ sbp              <dbl> 135, 159, 141, 113, 151, NA, 125…
## $ dbp              <dbl> 89, 91, 79, 73, 80, NA, 71, 85, …
## $ sex              <fct> 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0,…
## $ age              <dbl> 43, 49, 71, 36, 48, 51, 56, 40, …
## $ race             <fct> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,…
## $ income           <dbl> 19, 22, 17, 21, 18, 22, 20, 18, …
```

# Target estimands

## Average Treatment Effect among the Controls (ATC)

$$\tau = E[Y(1) - Y(0)|Z = 0]$$



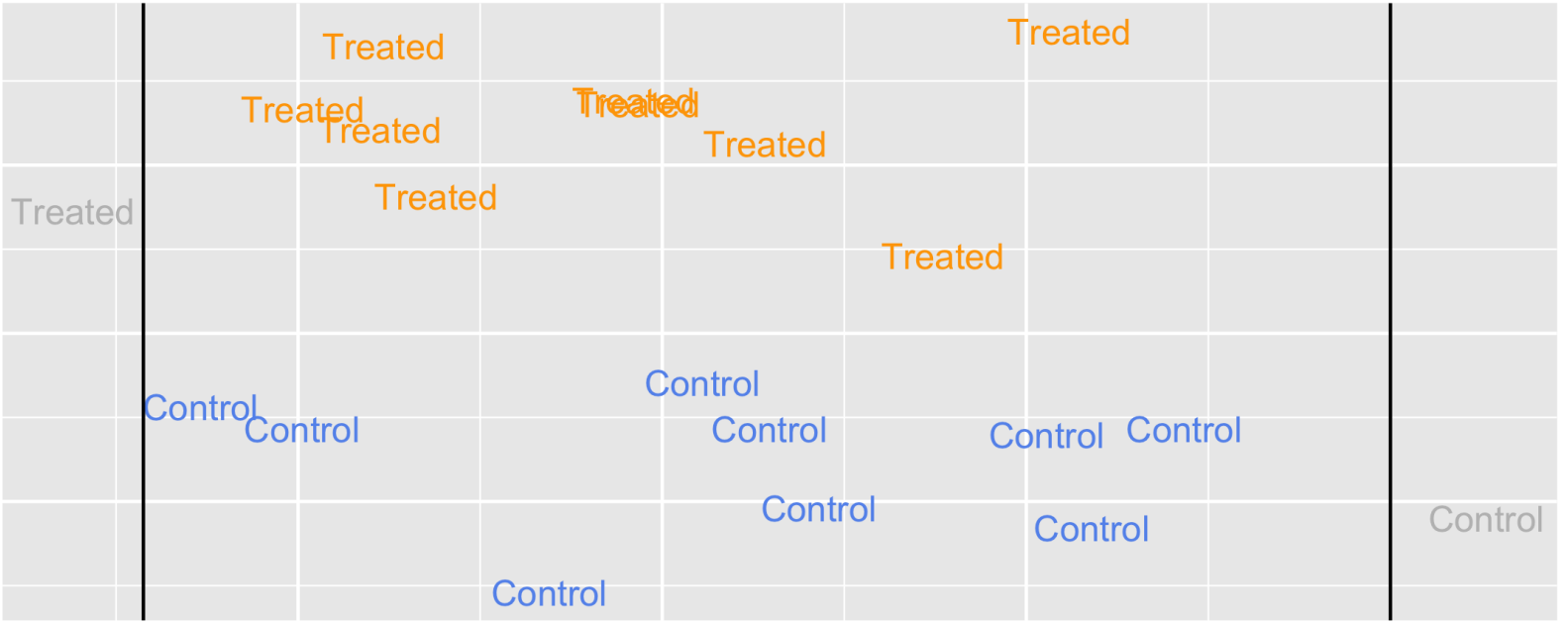
# Matching in R (ATC)

```
library(MatchIt)
m <- matchit(qsmk ~ sex +
    race + age + I(age^2) + education +
    smokeintensity + I(smokeintensity^2) +
    smokeyrs + I(smokeyrs^2) + exercise + active +
    wt71 + I(wt71^2),
  data = nhefs_complete,
  estimand = "ATC")
m
```

```
## A matchit object
##  - method: 1:1 nearest neighbor matching without replacement
##  - distance: Propensity score
##              - estimated with logistic regression
##  - number of obs.: 1566 (original), 806 (matched)
##  - target estimand: ATC
##  - covariates: sex, race, age, I(age^2), education, smokeintensity, I
```

# Target estimands

## Average Treatment Effect among the Matched (ATM)



# Matching in R (ATM)

```
library(MatchIt)
m <- matchit(qsmk ~ sex +
    race + age + I(age^2) + education +
    smokeintensity + I(smokeintensity^2) +
    smokeyrs + I(smokeyrs^2) + exercise + active +
    wt71 + I(wt71^2),
  data = nhefs_complete,
  link = "linear.logit",
  caliper = 0.1)
m
```

```
## A matchit object
##  - method: 1:1 nearest neighbor matching without replacement
##  - distance: Propensity score [caliper]
##                - estimated with logistic regression and linearized
##  - caliper: <distance> (0.063)
##  - number of obs.: 1566 (original), 780 (matched)
##  - target estimand: ATT
##  - covariates: sex, race, age, I(age^2), education, smokeintensity, I
```

# Matching in R (ATM)

```
library(MatchIt)
m <- matchit(qsmk ~ sex +
    race + age + I(age^2) + education +
    smokeintensity + I(smokeintensity^2) +
    smokeyrs + I(smokeyrs^2) + exercise + active +
    wt71 + I(wt71^2),
  data = nhefs_complete,
  link = "linear.logit",
  caliper = 0.1)
m
```

Observations with propensity scores (on the linear logit scale) within 0.1 (the caliper) will be discarded

# Matching in R (ATM)

```
matched_data <- get_matches(m, id = "i")
glimpse(matched_data)
```

```
## Rows: 780
## Columns: 71
## $ i                <chr> "11", "1220", "15", "1082", "18"…
## $ subclass         <fct> 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6,…
## $ weights          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
## $ seqn             <dbl> 428, 23045, 446, 22294, 596, 140…
## $ qsmk             <dbl> 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1,…
## $ death            <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,…
## $ yrdth            <dbl> NA, NA, 88, NA, NA, NA, NA, NA, …
## $ modth            <dbl> NA, NA, 1, NA, NA, NA, NA, NA, N…
## $ dadth            <dbl> NA, NA, 3, NA, NA, NA, NA, NA, N…
## $ sbp              <dbl> 135, 159, 141, 113, 151, NA, 125…
## $ dbp              <dbl> 89, 91, 79, 73, 80, NA, 71, 85, …
## $ sex              <fct> 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1,…
## $ age              <dbl> 43, 49, 71, 36, 48, 51, 56, 40, …
## $ race             <fct> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,…
## $ income           <dbl> 19, 22, 17, 21, 18, 22, 20, 18, …
```

# Your Turn

1. **Using the propensity scores you created in the previous exercise, create a "matched" data set using the ATM method with a caliper of 0.2.**

10:00

# Propensity scores

- Matching
- **Weighting**
- Stratification
- Direct Adjustment
- ...

# Target estimands

## Average Treatment Effect (ATE)

$$w_{ATE} = \frac{Z_i}{p_i} + \frac{1 - Z_i}{1 - p_i}$$

# Target estimands

## Average Treatment Effect Among the Treated (ATT)

$$w_{ATT} = \frac{p_i Z_i}{p_i} + \frac{p_i(1 - Z_i)}{1 - p_i}$$

## Average Treatment Effect Among the Controls (ATC)

$$w_{ATC} = \frac{(1 - p_i) Z_i}{p_i} + \frac{(1 - p_i)(1 - Z_i)}{(1 - p_i)}$$

# Target estimands

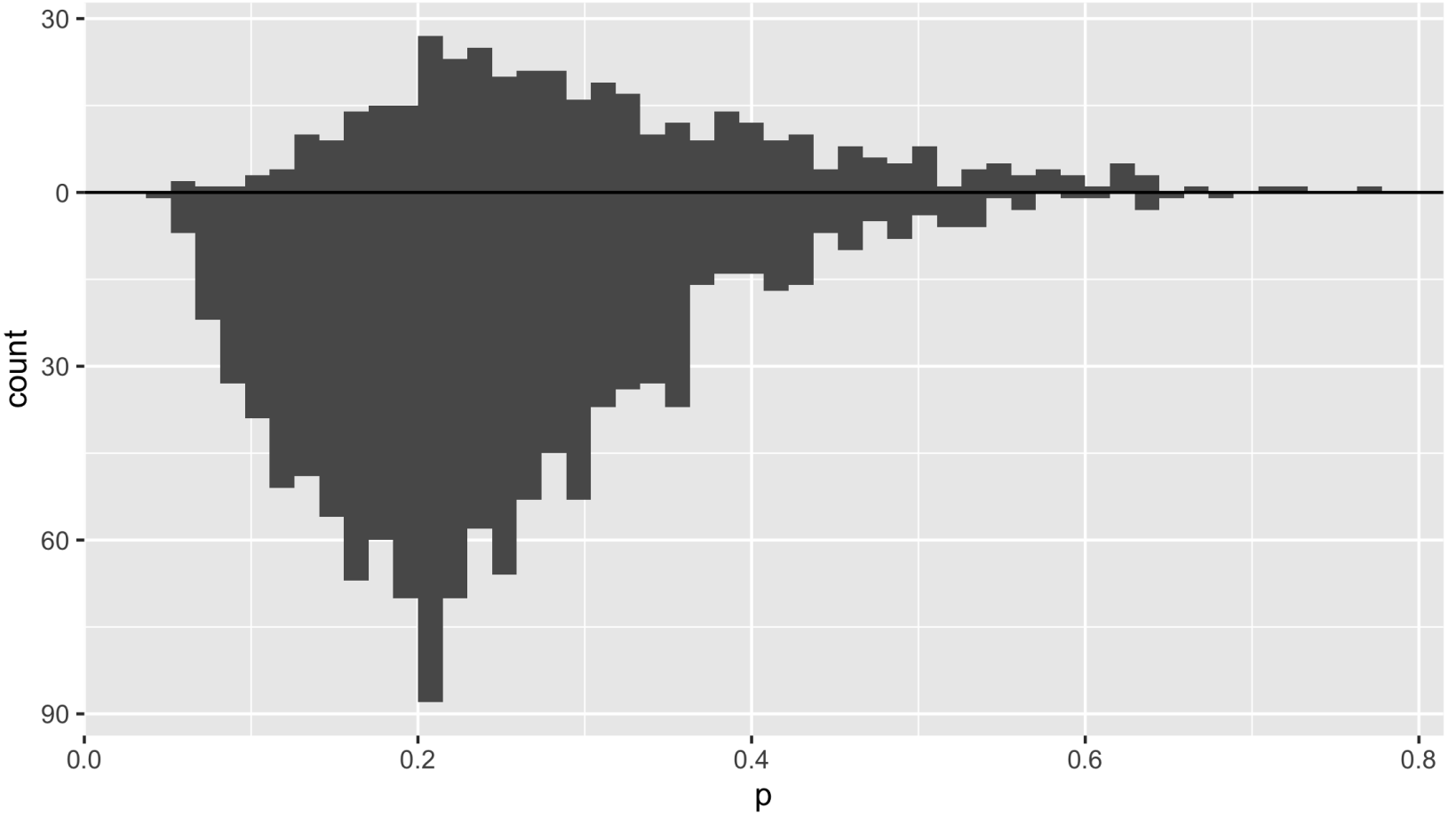
## Average Treatment Effect Among the Evenly Matchable (ATM)

$$w_{ATM} = \frac{\min\{p_i, 1-p_i\}}{Z_i p_i + (1-Z_i)(1-p_i)}$$

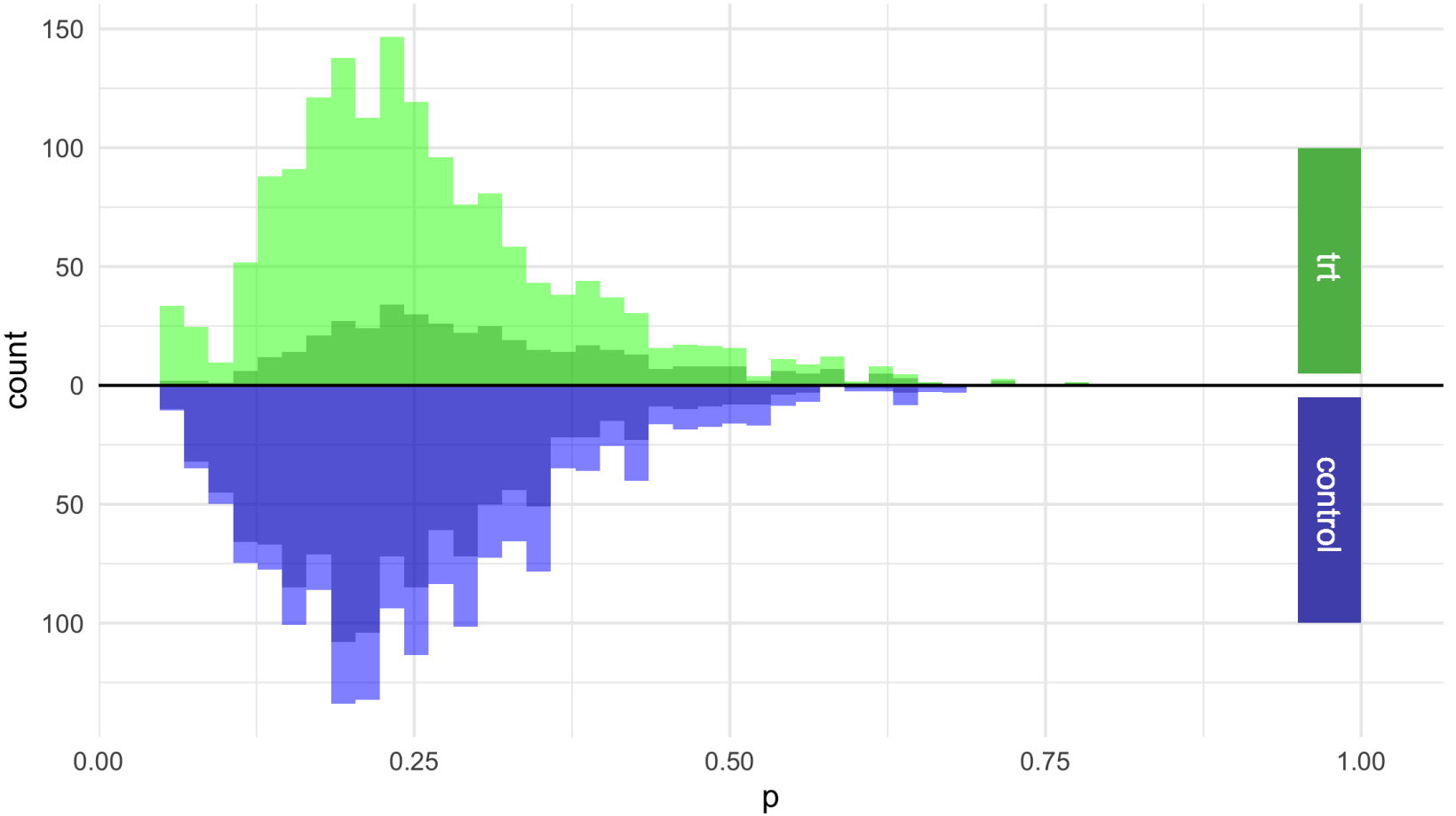
## Average Treatment Effect Among the Overlap Population

$$w_{ATO} = (1-p_i)Z_i + p_i(1-Z_i)$$

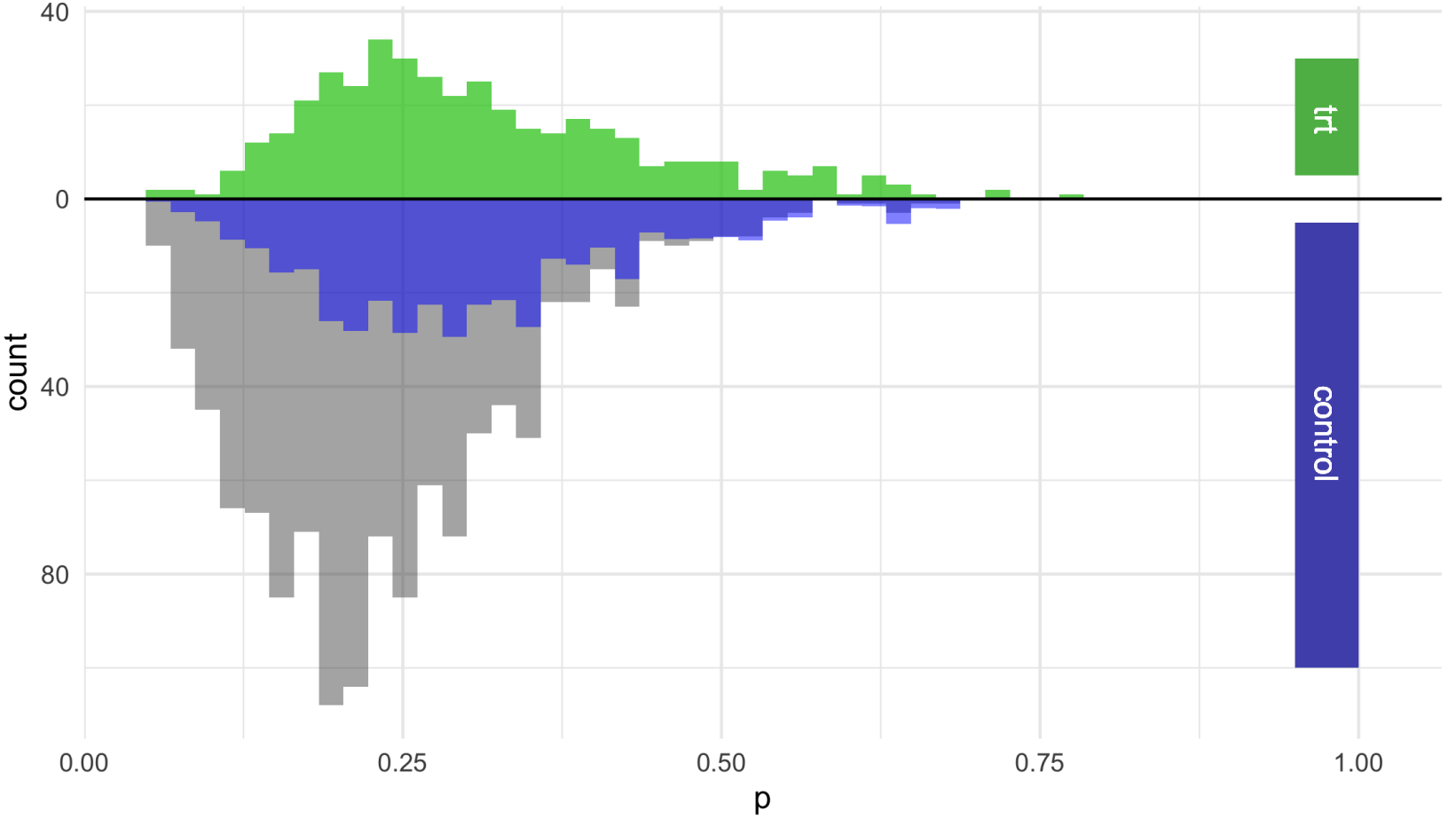
# Histogram of propensity scores



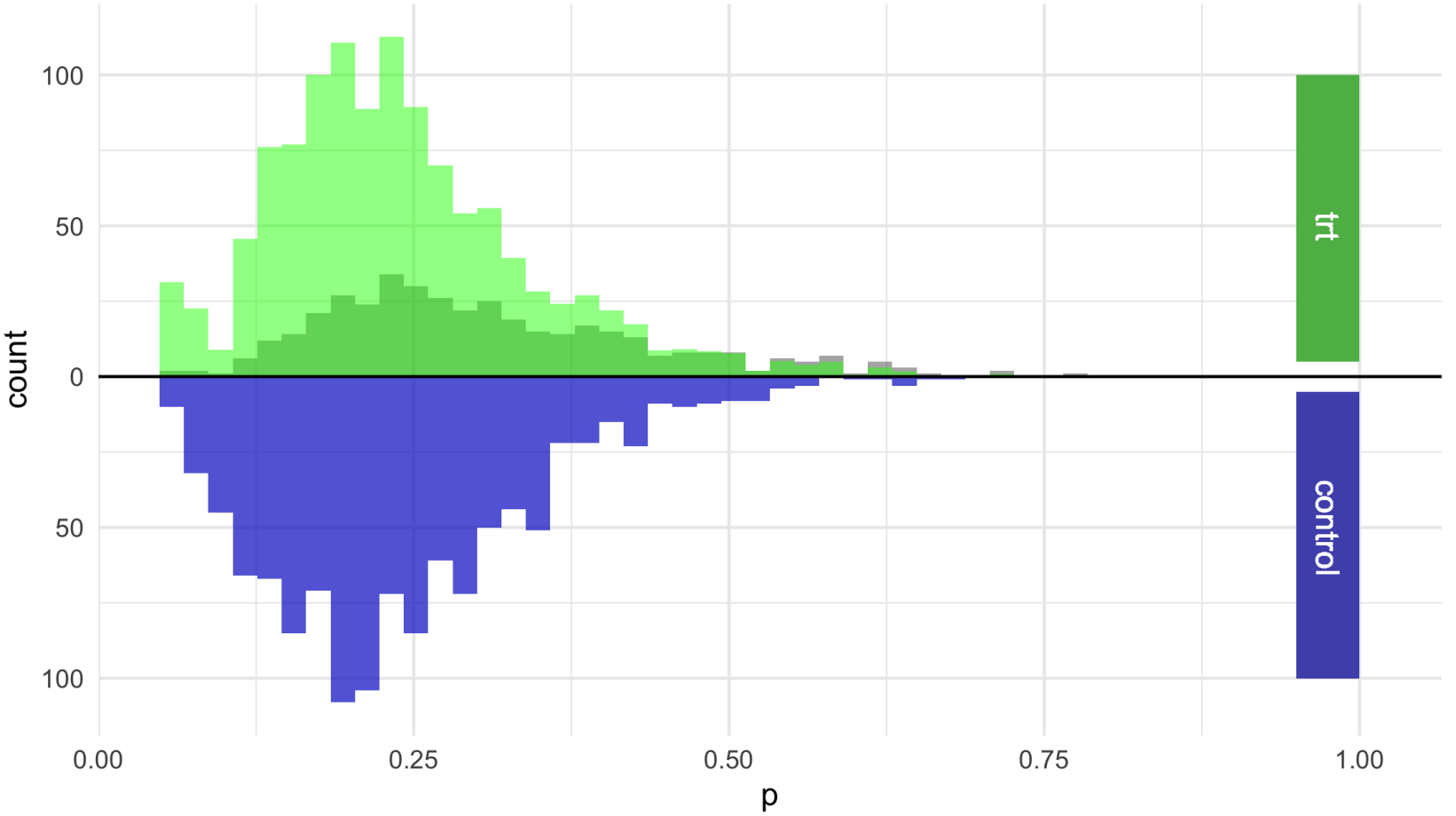
# ATE



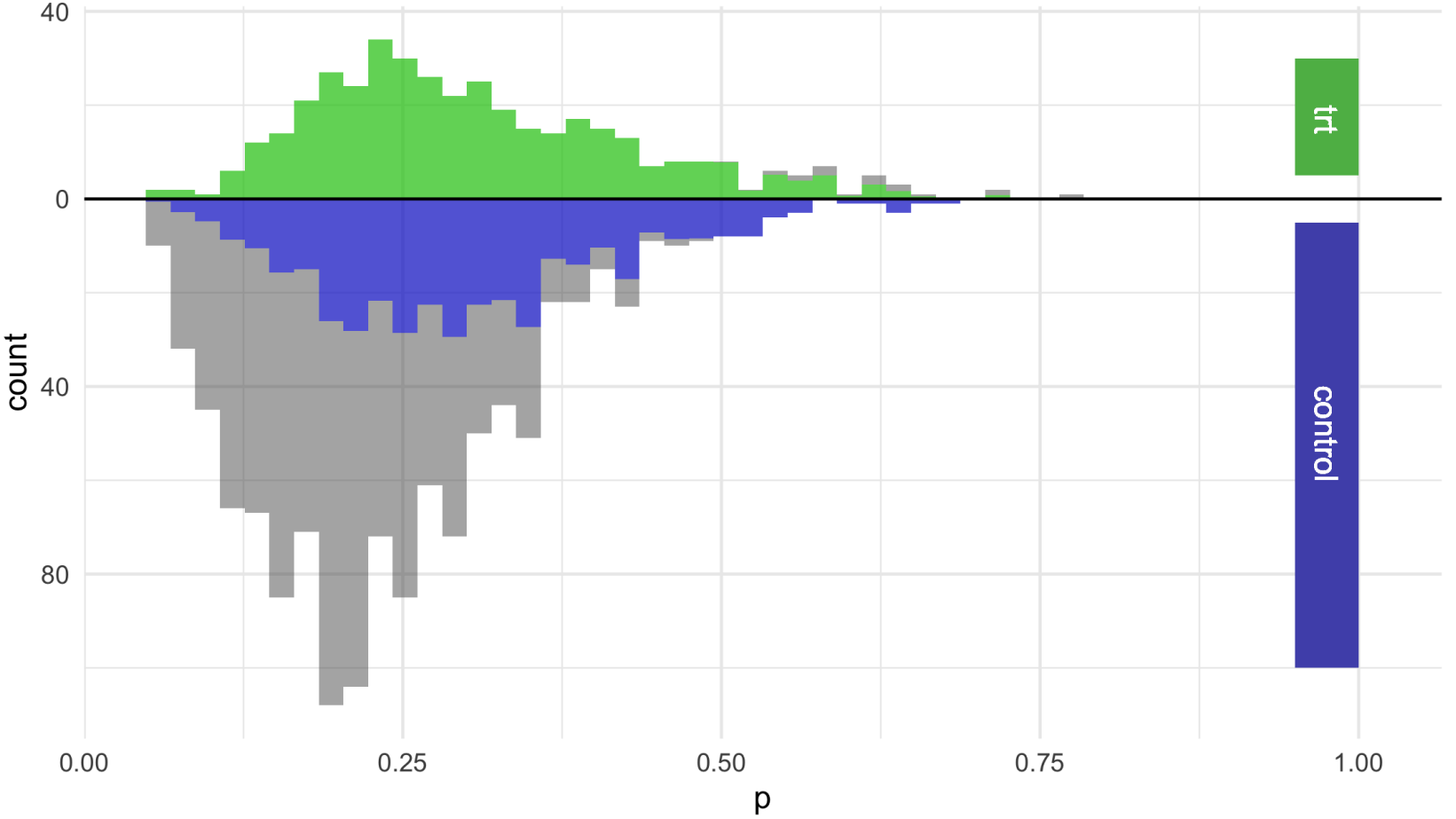
# ATT



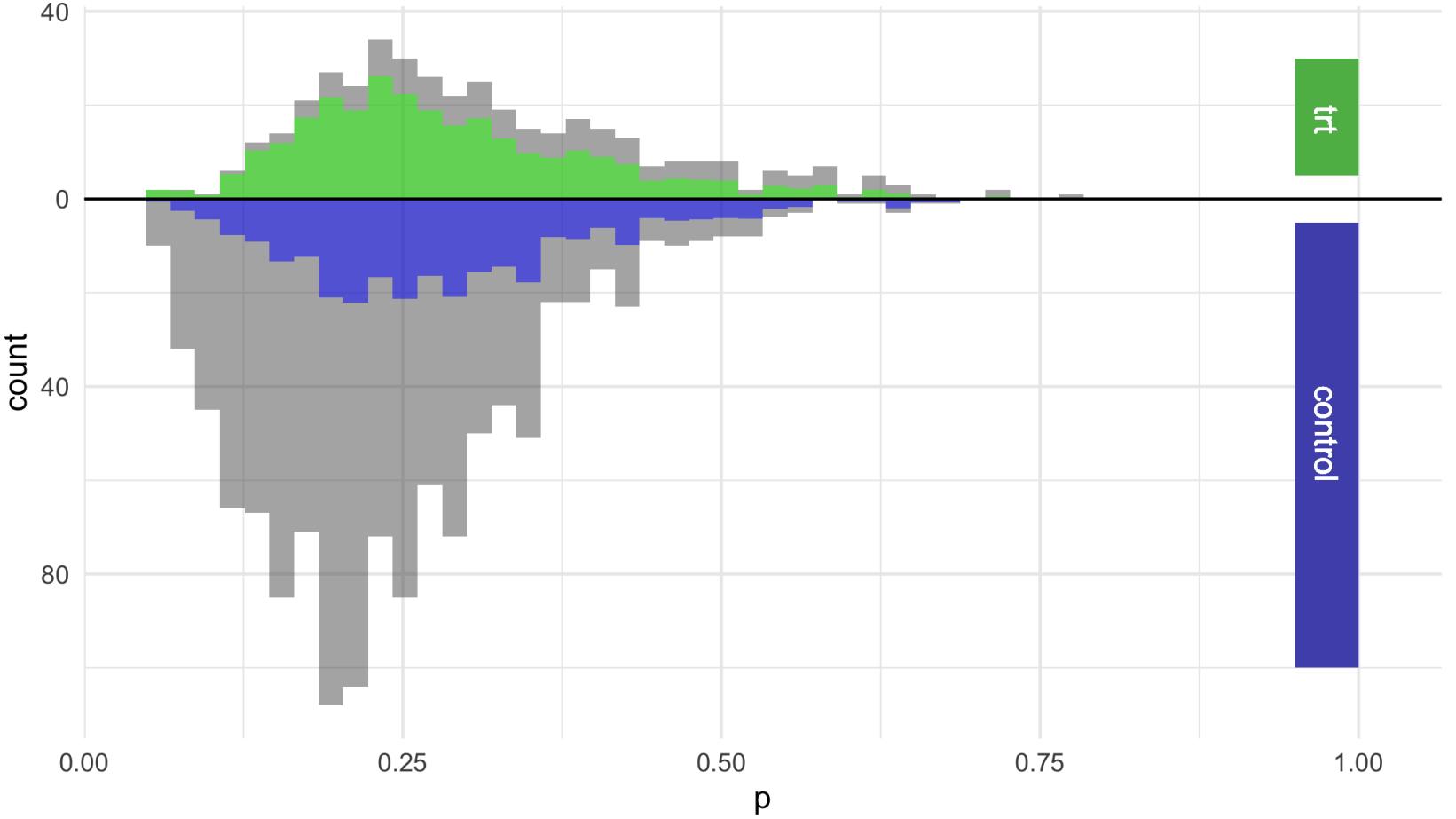
# ATC



# ATM



# ATO



# ATE in R

## Average Treatment Effect (ATE)

$$w_{ATE} = \frac{Z_i}{p_i} + \frac{1-Z_i}{1-p_i}$$

```
df <- propensity_model %>%
  augment(type.predict = "response", data = nhefs_complete) %>%
  mutate(w_ate = (qsmk / .fitted) +  ((1 - qsmk) / (1 - .fitted)))
```

# Your Turn

1. **Using the propensity scores you created in the previous exercise, add the ATE weights to your data frame** df

2. **Stretch: Using the same propensity scores, create ATM weights**

10:00