

# Using Propensity Scores

Lucy D'Agostino McGowan

Wake Forest University

2021-09-01

# Propensity scores

Matching

Weighting

Stratification

Direct Adjustment

...

# Estimands

Greifer, N., & Stuart, E. A. (2021). Choosing the estimand when matching or weighting in observational studies. arXiv preprint arXiv:2106.10577.

Table 2. Summary of estimands and methods for estimating them.

Estimand	Target population	Example research question	Matching methods	Weighting methods
ATT	Treated patients	Should medical providers withhold treatment from those currently receiving it?	Pair matching (e.g., nearest neighbor, optimal) without a caliper (11) Full matching (12) Fine stratification (13)	Standardized mortality ratio weights (2)
ATU	Untreated (control) patients	Should medical providers extend treatment to those not currently receiving it?	Same as ATT	Same as ATT
ATE	Full sample/ population	Should a specific policy be applied to all eligible patients?	Full matching (12) Fine stratification (13)	Inverse probability weights (14,15)
ATO	Clinical equipoise	Should those at clinical equipoise receive treatment? Is there an effect of the treatment for some patients?	Caliper matching (11,16) Coarsened exact matching (17,18) Cardinality matching (19)	Overlap weights (20) Matching weights (21) Weight trimming (22)

Notes: ATT - average treatment effect in the treated; ATU - average treatment effect in the untreated; ATE - average treatment effect in the population; ATO - average treatment effect in the overlap

# 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)

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

# Matching in R (ATT)

```
1 matched_data <- get_matches(m, id = "i")
2 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, NA, ...
$ modth     <dbl> NA, NA, 1, NA, NA, NA, NA, NA, NA, N...
$ dadth     <dbl> NA, NA, 3, NA, 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, ...
^
```



# Target estimands

## Average Treatment Effect among the Controls (ATC)

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

# Matching in R (ATC)

```
1 library(MatchIt)
2 m <- matchit(qsmk ~ sex +
3     race + age + I(age^2) + education +
4     smokeintensity + I(smokeintensity^2) +
5     smokeyrs + I(smokeyrs^2) + exercise + active +
6     wt71 + I(wt71^2),
7     data = nhefs_complete,
8     estimand = "ATC")
9 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(smokeintensity^2), smokeyrs, I(smokeyrs^2), exercise, active, wt71, I(wt71^2)

# Target estimands

**Average Treatment Effect among the Matched (ATM)**

# Matching in R (ATM)

```
1 library(MatchIt)
2 m <- matchit(qsmk ~ sex +
3     race + age + I(age^2) + education +
4     smokeintensity + I(smokeintensity^2) +
5     smokeyrs + I(smokeyrs^2) + exercise + active +
6     wt71 + I(wt71^2),
7     data = nhefs_complete,
8     link = "linear.logit",
9     caliper = 0.1)
10 m
```

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

# Matching in R (ATM)

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(smokeintensity^2), smokeyrs, I(smokeyrs^2), exercise, active, wt71, I(wt71^2)

# Matching in R (ATM)

```
1 matched_data <- get_matches(m, id = "i")
2 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, NA, ...
$ modth     <dbl> NA, NA, 1, NA, NA, NA, NA, NA, NA, N...
$ dadth     <dbl> NA, NA, 3, NA, 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, ...
^
```

## ***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.**

# Propensity scores

Matching

**Weighting**

Stratification

Direct Adjustment

...



# Target estimands: ATE

Average Treatment Effect (ATE)

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

$$1 \cdot (z / p) + ((1 - z) / (1 - p))$$

# Target estimands: ATT & ATC

Average Treatment Effect Among the Treated (ATT)

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

$$1 \quad ((p * z) / p) + ((p * (1 - z)) / (1 - p))$$

# Target estimands: ATT & ATC

Average Treatment Effect Among the Controls (ATC)

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

$$1 \quad ( ((1 - p) * z) / p ) + ( ((1 - p) * (1 - z)) / (1 - p) )$$

# Target estimands: ATM & ATO

Average Treatment Effect Among the Evenly Matchable (ATM)

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

```
1 pmin(p, 1 - p) / (z * p + (1 - z) * (1 - p))
```

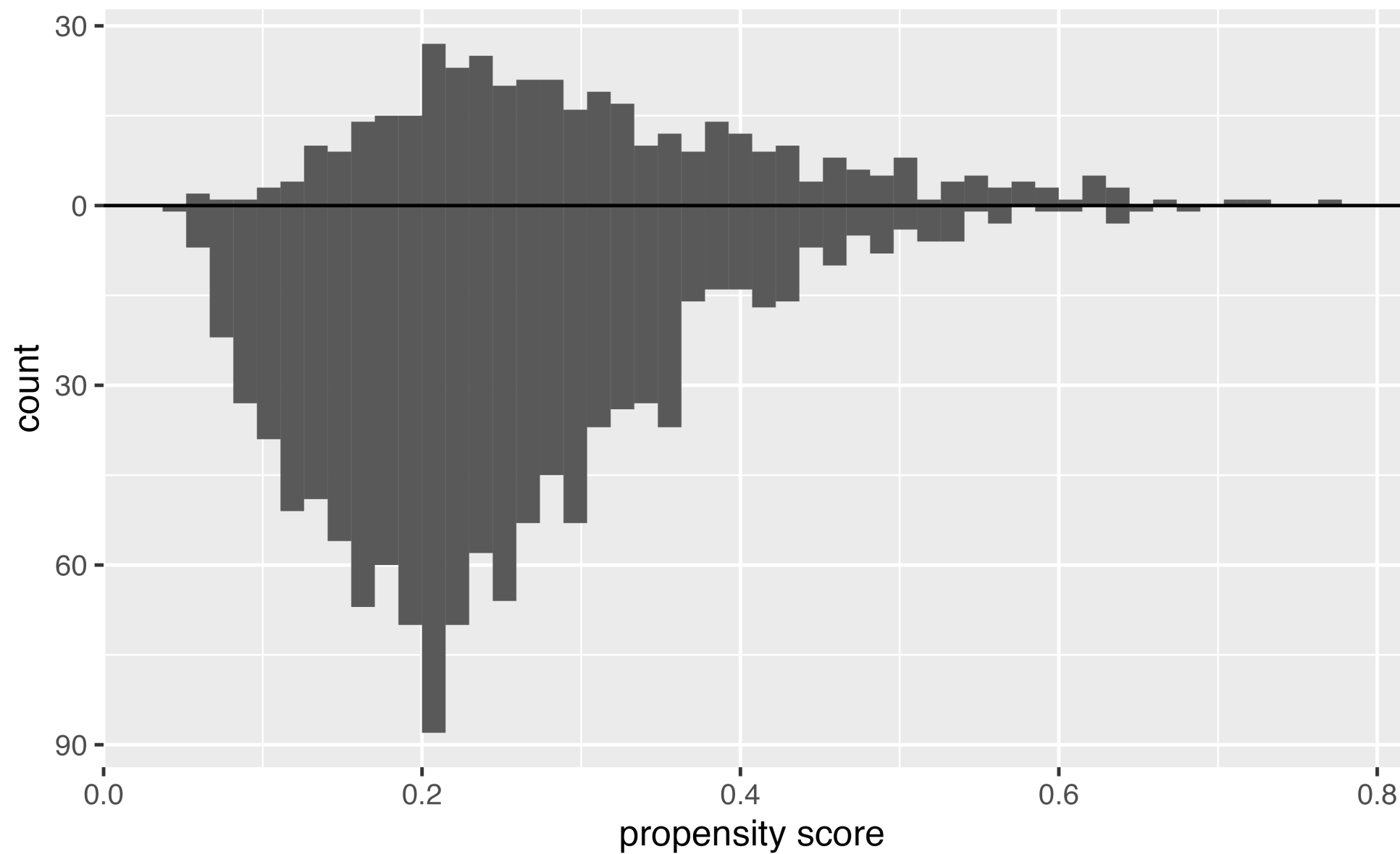
# Target estimands: ATM & ATO

Average Treatment Effect Among the Overlap Population

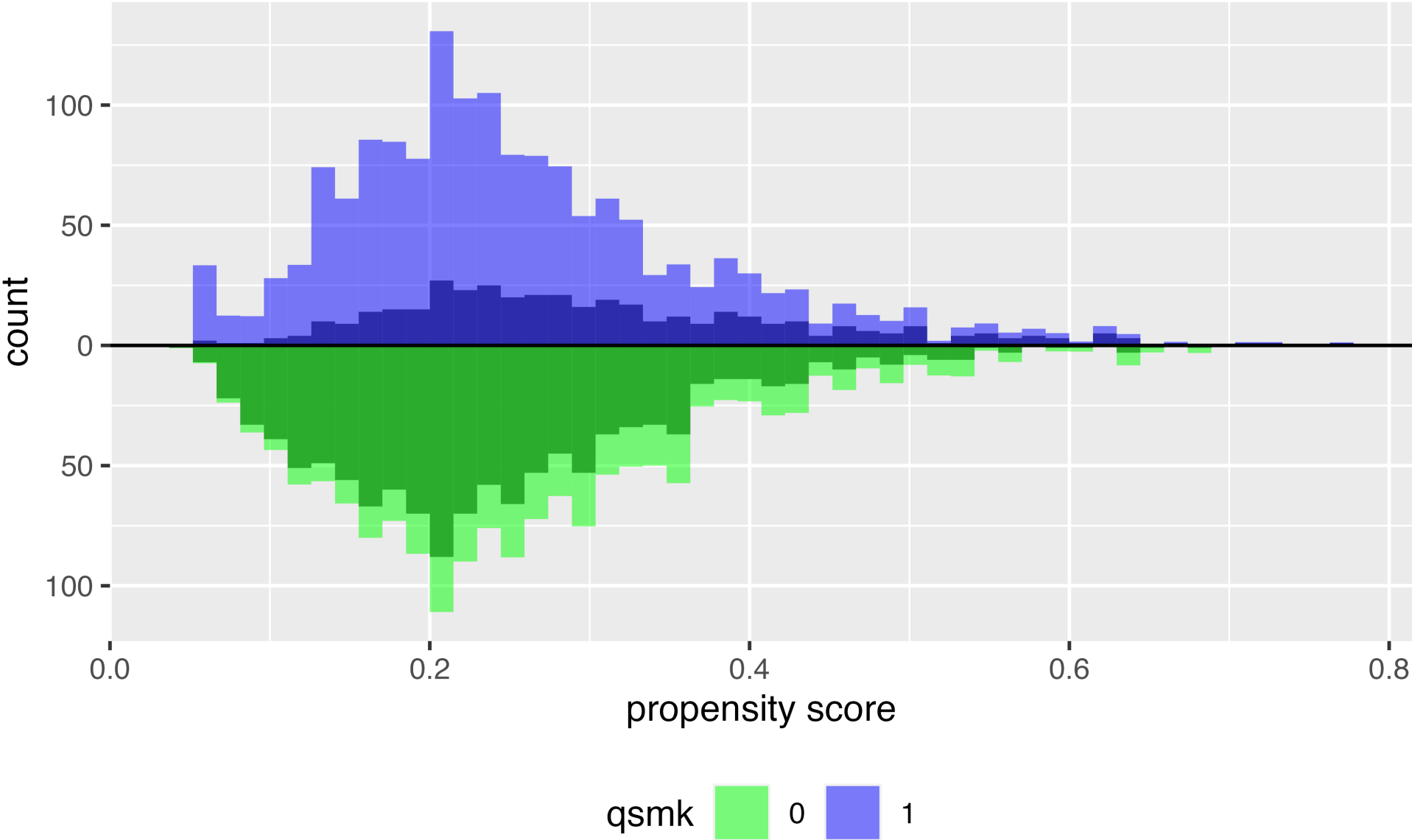
$$w_{ATO} = (1 - p_i)Z_i + p_i(1 - Z_i)$$

$$1 \quad (1 - p) * z + p * (1 - z)$$

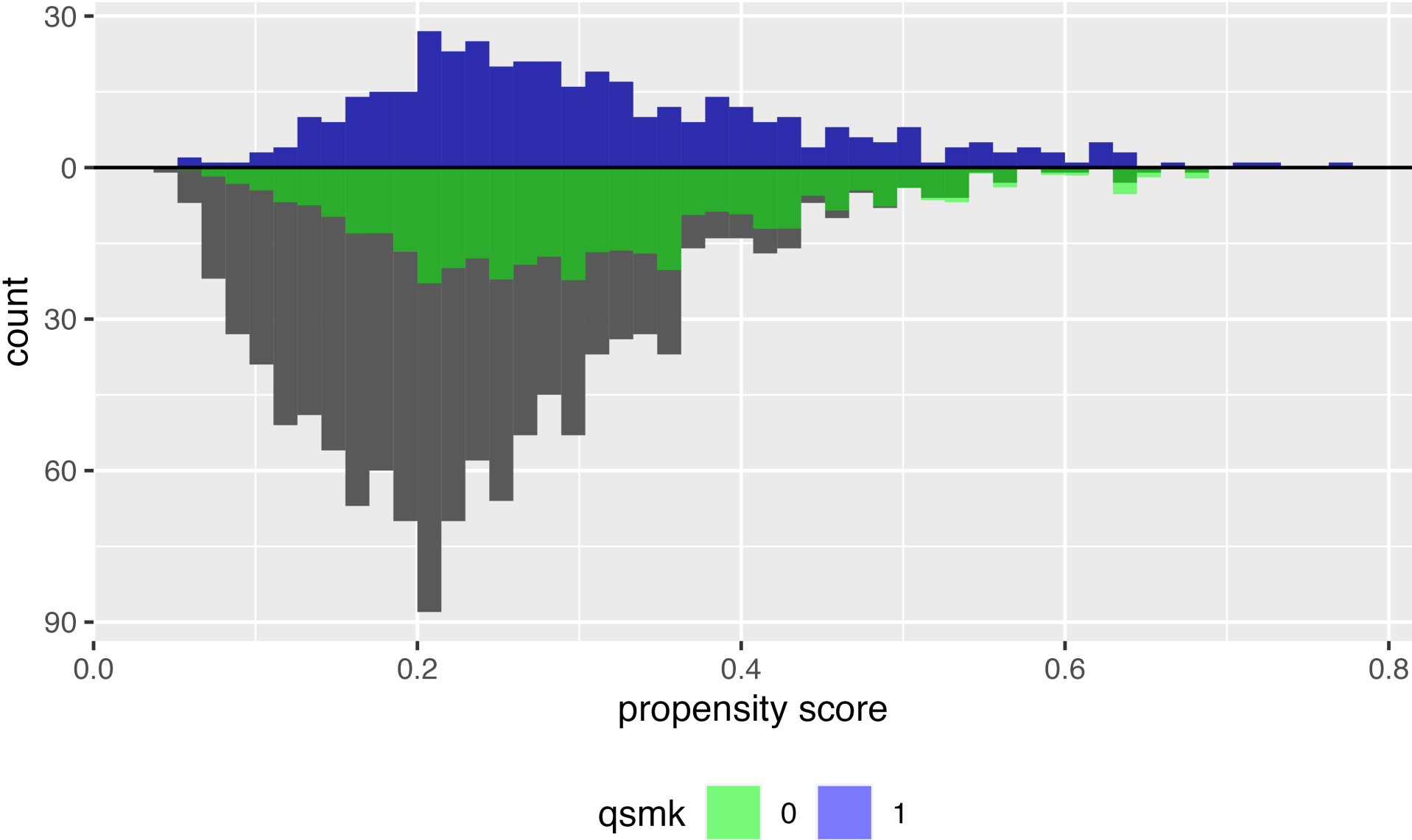
# Histogram of propensity scores



# ATE

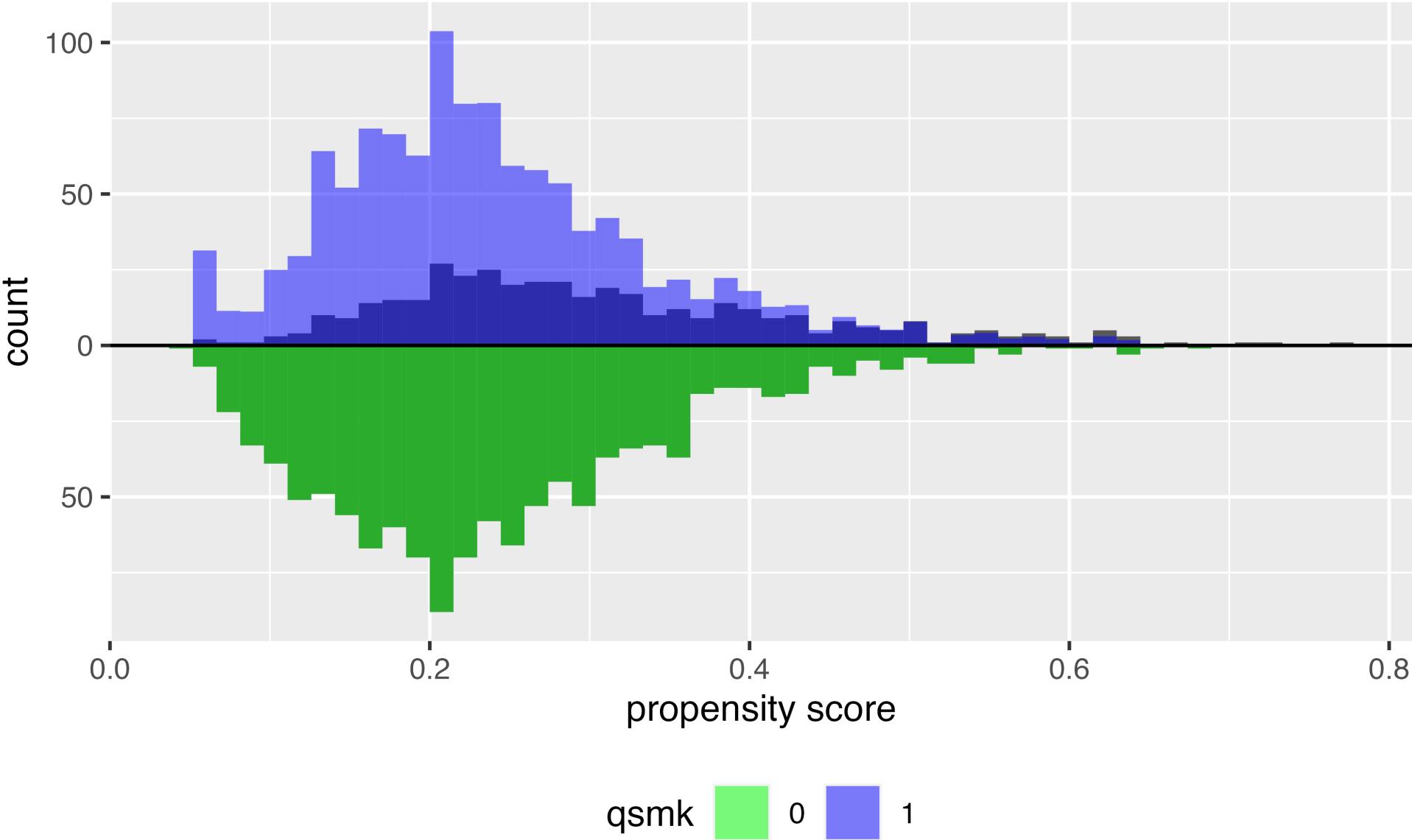


# ATT

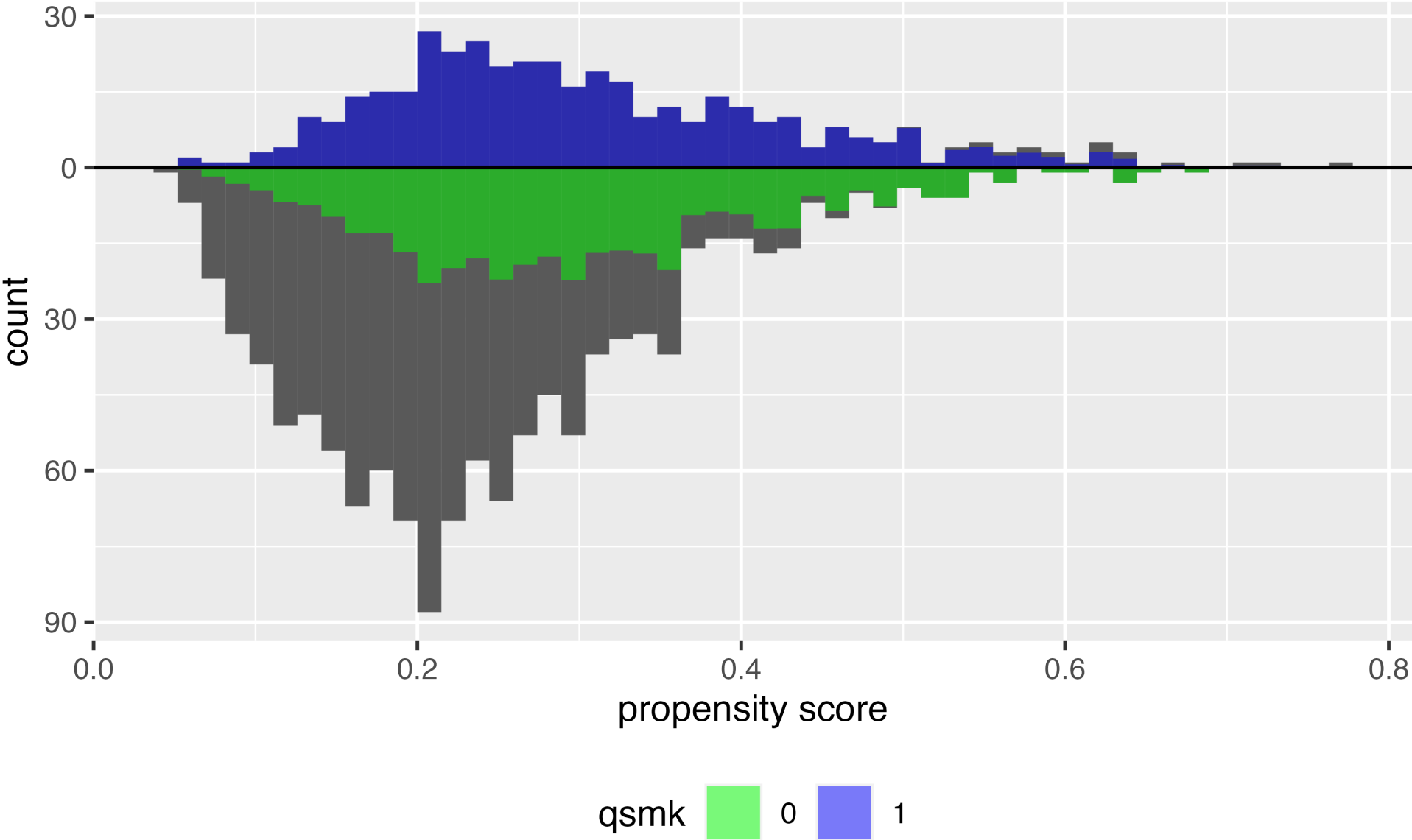




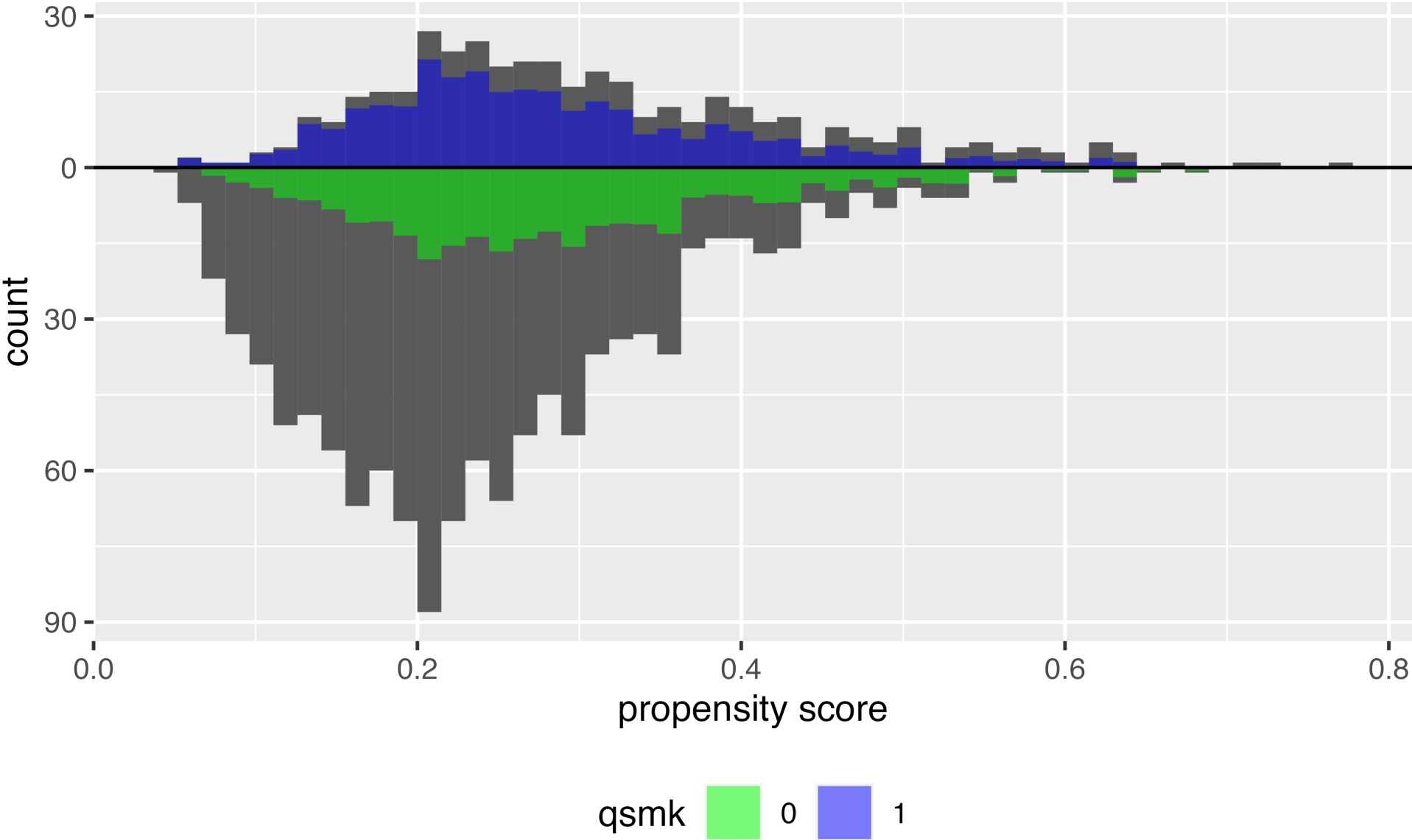
# ATC



# ATM



# ATO



# ATE in R



Average Treatment Effect (ATE)  $w_{ATE} = \frac{Z_i}{p_i} + \frac{1-Z_i}{1-p_i}$

```
1 library(propensity)
2 df <- propensity_model |>
3   augment(type.predict = "response", data = nhfs_complete) |>
4   mutate(w_ate = wt_ate(.fitted, qsmk))
```

## ***Your Turn***

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

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

