MAS_Prelim_Results

Devin Reeh

2025-05-27

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

I. Identify Treatment Assignment and Outcome Variables

This study defines treatment as school participation in either the free meal program or the Beyond the Bell (BTB) afterschool program, based on FRPM eligibility thresholds. The sole outcome examined is chronic absenteeism, measured as the percentage of students missing 10% or more of the school year. Clearly defining both treatment and outcome enables a causal comparison between schools just above and below the FRPM cutoff, where access to these programs changes. This setup supports the core research question: whether program participation leads to differences in absenteeism rates.

II. Estimate the Treatment Effect with Local Linear Regression

```
# 2. Sharp RD estimate
rd75 <- rdrobust(merged$chronic_absenteeism,
                 merged$frpm_rate,
                 c          = 75,
                 masspoints = "adjust")
```

```
## Warning in rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, :
## Mass points detected in the running variable.
```

```
summary(rd75)
```

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                8391
```

```
## BW type                       mserd
## Kernel                    Triangular
## VCE method                        NN
##
## Number of Obs.             2615          5776
## Eff. Number of Obs.         252           309
## Order est. (p)                1             1
## Order bias  (q)               2             2
## BW est. (h)               4.443         4.443
## BW bias (b)               7.426         7.426
## rho (h/b)                0.598         0.598
## Unique Obs.                 176           433
##
## =============================================================================
##        Method    Coef. Std. Err.        z     P>|z|      [ 95% C.I. ]
## =============================================================================
##    Conventional  10.368     1.734    5.978     0.000   [6.969 , 13.767]
##         Robust        -         -    5.446     0.000   [7.169 , 15.229]
## =============================================================================
```

III. Choose and Justify Bandwidth

```
# Load package
library(rdrobust)

# Declare models
m1  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 2, p = 1, masspoints = "adjus
m2  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 2, p = 2, masspoints = "adjus
m3  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 2, p = 3, masspoints = "adjus
m4  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 3, p = 1, masspoints = "adjus
m5  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 3, p = 2, masspoints = "adjus
m6  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 3, p = 3, masspoints = "adjus
m7  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 4, p = 1, masspoints = "adjus
m8  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 4, p = 2, masspoints = "adjus
m9  <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 4, p = 3, masspoints = "adjus
m10 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 5, p = 1, masspoints = "adjus
m11 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 5, p = 2, masspoints = "adjus
m12 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 5, p = 3, masspoints = "adjus
m13 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 6, p = 1, masspoints = "adjus
m14 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 6, p = 2, masspoints = "adjus
m15 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 6, p = 3, masspoints = "adjus

# Store models with correct metadata manually
model_info <- data.frame(
  name = paste0("m", 1:15),
  cutoff = 75,
  bandwidth = rep(2:6, each = 3),
  poly_order = rep(1:3, times = 5),
  stringsAsFactors = FALSE
)
```

```r
models <- list(m1, m2, m3, m4, m5, m6, m7, m8, m9, m10, m11, m12, m13, m14, m15)

# Extract results and build table
robust_results <- do.call(rbind, Map(function(model, info) {
  data.frame(
    cutoff = info$cutoff,
    bandwidth = info$bandwidth,
    poly_order = info$poly_order,
    coef = model$coef[1, 1],
    se = model$se[1, 1],
    pval = model$pv[1, 1],
    ci_lower = model$ci[1, 1],
    ci_upper = model$ci[1, 2]
  )
}, models, split(model_info, seq(nrow(model_info)))))

# Show result
print(robust_results)
```

```
##    cutoff bandwidth poly_order       coef        se         pval    ci_lower
## 1      75         2          1  -2.428590  2.811263 3.876546e-01   -7.938566
## 2      75         2          2 -31.283435  5.056396 6.135728e-10  -41.193789
## 3      75         2          3 -40.774795 10.145976 5.849172e-05  -60.660543
## 4      75         3          1   8.075415  2.070939 9.643238e-05    4.016449
## 5      75         3          2  -7.387688  3.629974 4.183172e-02  -14.502306
## 6      75         3          3 -42.998963  6.160372 2.953283e-12  -55.073070
## 7      75         4          1   9.984008  1.862172 8.254047e-08    6.334218
## 8      75         4          2   1.781048  2.765611 5.195767e-01   -3.639449
## 9      75         4          3 -13.602781  4.368021 1.844601e-03  -22.163945
## 10     75         5          1  10.174908  1.625776 3.887195e-10    6.988444
## 11     75         5          2   7.326720  2.440236 2.678041e-03    2.543945
## 12     75         5          3  -8.064500  3.650360 2.715838e-02  -15.219075
## 13     75         6          1   8.156506  1.441245 1.519510e-08    5.331717
## 14     75         6          2  10.897604  2.233584 1.066379e-06    6.519859
## 15     75         6          3  -2.008942  3.159919 5.249345e-01   -8.202270
##      ci_upper
## 1    3.0813847
## 2  -21.3730818
## 3  -20.8890472
## 4   12.1343814
## 5   -0.2730700
## 6  -30.9248554
## 7   13.6337982
## 8    7.2015449
## 9   -5.0416172
## 10  13.3613709
## 11  12.1094953
## 12  -0.9099248
## 13  10.9812952
## 14  15.2753484
## 15   4.1843856
```

```
bw <- rdbwselect(merged$chronic_absenteeism, merged$frpm_rate, c = 75)
```

```
## Warning in rdbwselect(merged$chronic_absenteeism, merged$frpm_rate, c = 75):
## Mass points detected in the running variable.
```

```
summary(bw)
```

```
## Call: rdbwselect
##
## Number of Obs.                8391
## BW type                      mserd
## Kernel                  Triangular
## VCE method                      NN
##
## Number of Obs.                2615           5776
## Order est. (p)                   1              1
## Order bias  (q)                  2              2
## Unique Obs.                    176            433
##
## =========================================================
##                   BW est. (h)      BW bias (b)
##            Left of c Right of c  Left of c Right of c
## =========================================================
##      mserd     4.443      4.443      7.426      7.426
## =========================================================
```

V. Interpret and Contextualize the Results

This means the optimal comparison window is ±4.443 percentage points around the cutoff — i.e., schools
with FRPM between 70.56% and 79.44%.

```
main <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 4.443, p = 1, masspoints = "
summary(main)
```

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                8391
## BW type                     Manual
## Kernel                  Triangular
## VCE method                      NN
##
## Number of Obs.                2615           5776
## Eff. Number of Obs.            252            309
## Order est. (p)                   1              1
## Order bias  (q)                  2              2
## BW est. (h)                  4.443          4.443
## BW bias (b)                  4.443          4.443
## rho (h/b)                    1.000          1.000
## Unique Obs.                   2615           5776
##
```

```
## ================================================================================
##           Method      Coef. Std. Err.            z      P>|z|      [ 95% C.I. ]
## ================================================================================
##     Conventional     10.368       1.734       5.979      0.000     [6.969 , 13.767]
##           Robust          -           -       1.992      0.046     [0.085 , 10.317]
## ================================================================================
```

VI. Visualize the Discontinuity (RDD Plots)

Identification: Quick Visual Check

- Schools with similar FRPM can have different absenteeism — typical heterogeneity.
- The dashed line at 75% is the policy threshold for program eligibility.
- A visible jump suggests program participation may increase absenteeism.

```r
library(rdrobust)

# Optimal bandwidth from rdbwselect (rounded for clarity)
rdplot(merged$chronic_absenteeism, merged$frpm_rate,
       c = 75, h = 4.443, p = 1,
       x.lim = c(65, 85),
       title = "RDD Plot at h = 4.443 (Optimal Bandwidth)",
       x.label = "FRPM Eligibility (%)",
       y.label = "Chronic Absenteeism (%)")
```

```
## [1] "Mass points detected in the running variable."
```
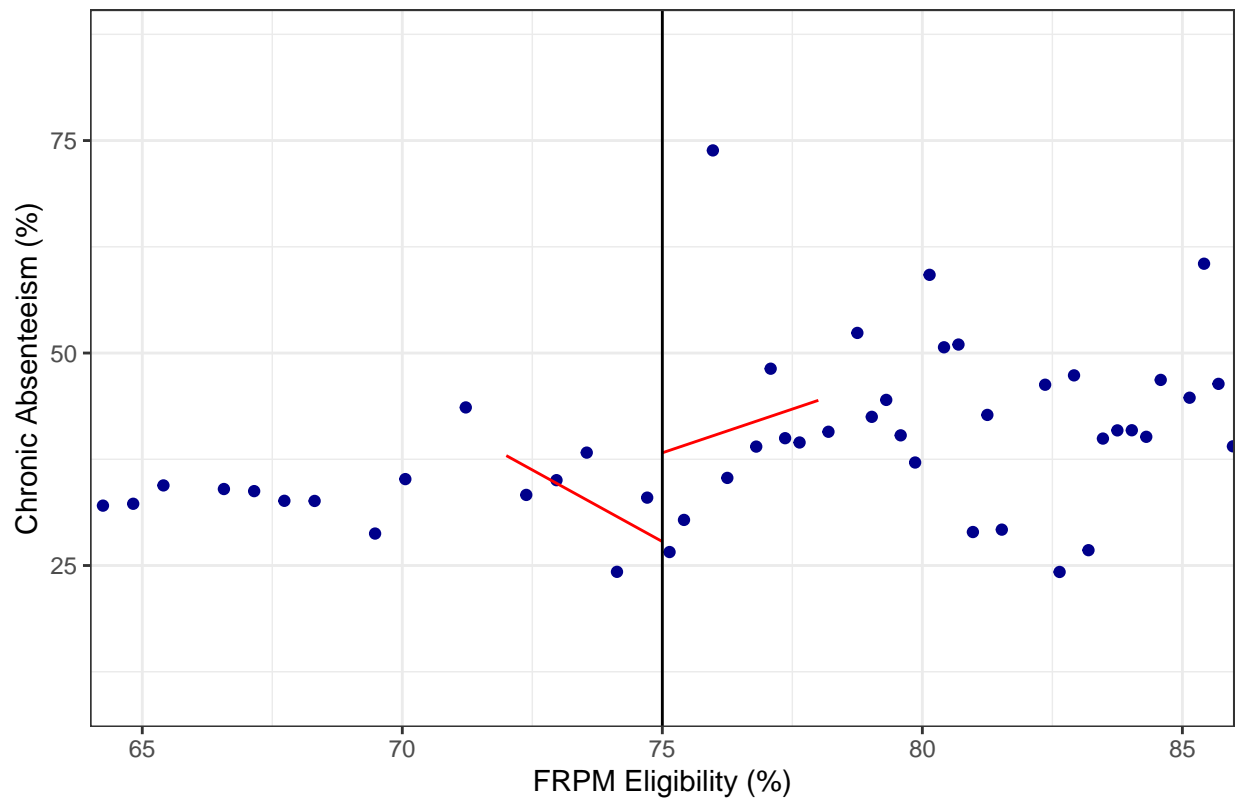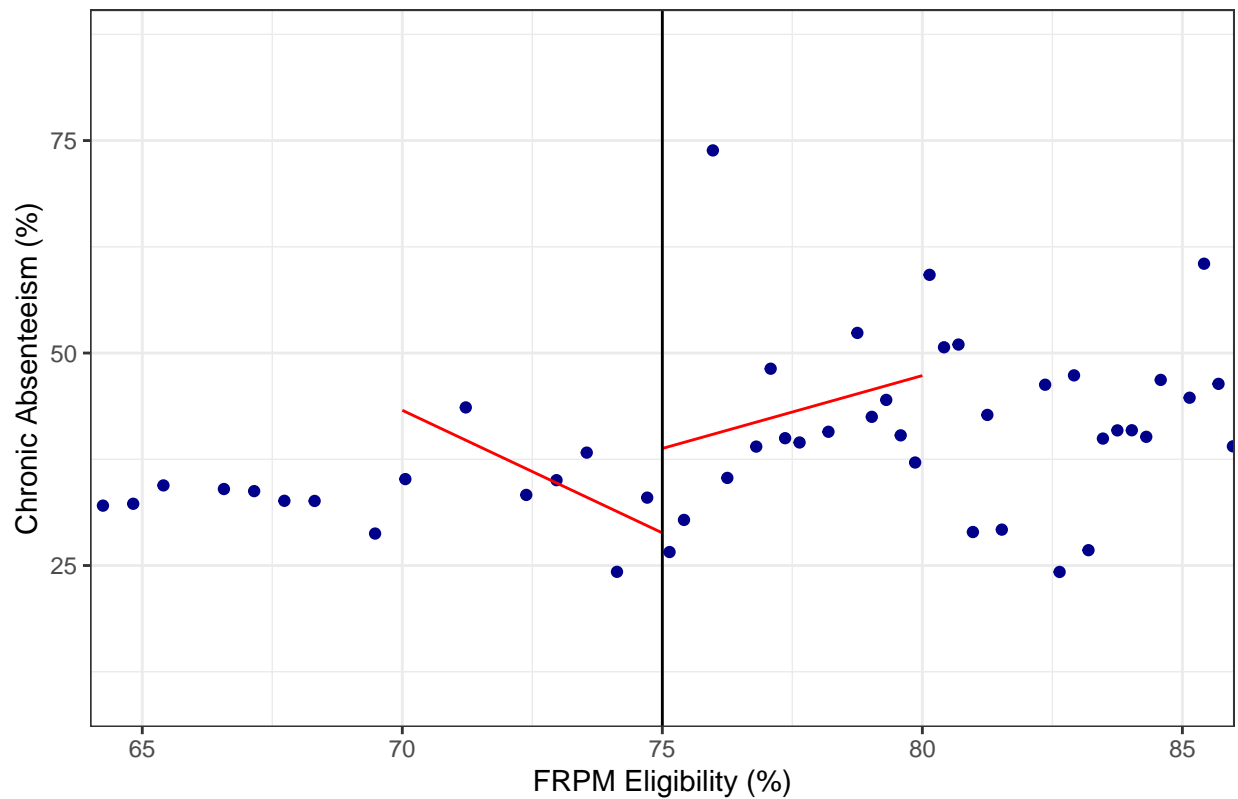
RDD Plot at h = 4.443 (Optimal Bandwidth)
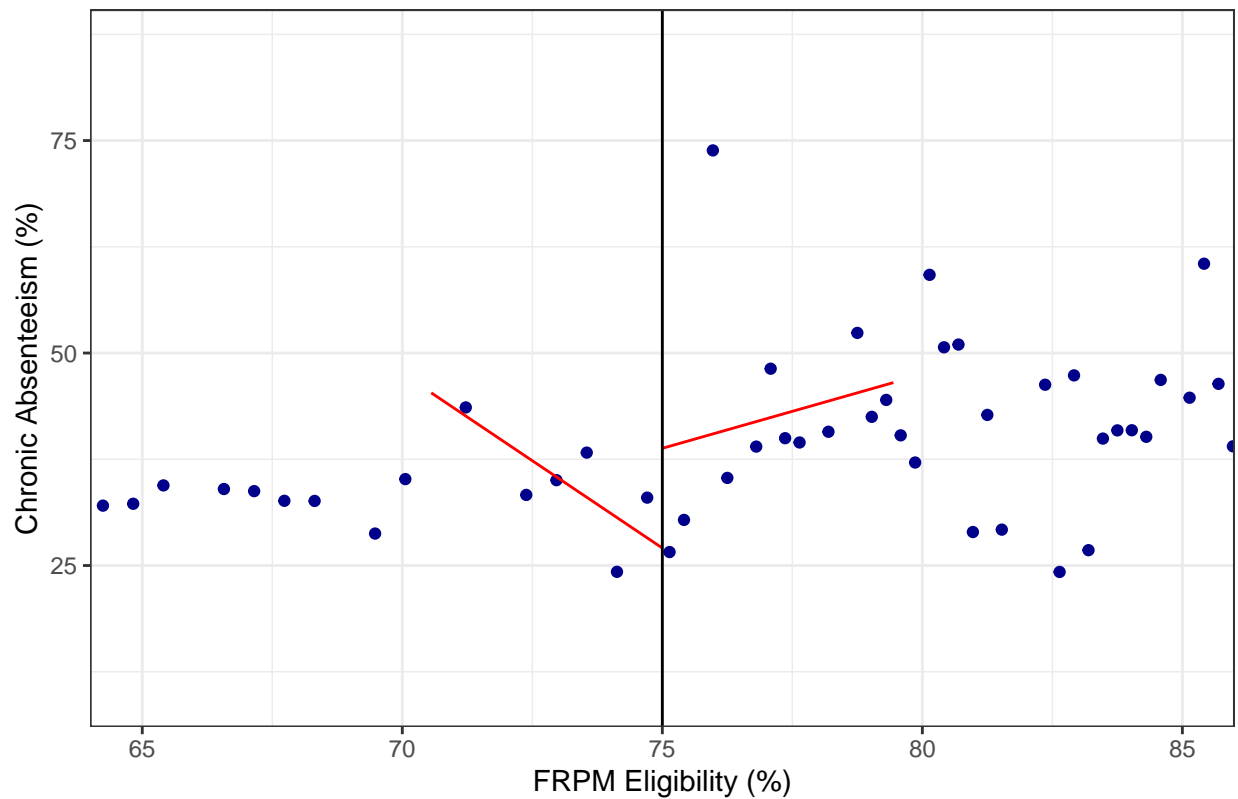

```
# Sensitivity: Narrower
rdplot(merged$chronic_absenteeism, merged$frpm_rate,
       c = 75, h = 3, p = 1,
       x.lim = c(65, 85),
       title = "RDD Plot at h = 3 (Narrow Bandwidth)",
       x.label = "FRPM Eligibility (%)",
       y.label = "Chronic Absenteeism (%)")
```

```
## [1] "Mass points detected in the running variable."
```

RDD Plot at h = 3 (Narrow Bandwidth)


```
# Sensitivity: Wider
rdplot(merged$chronic_absenteeism, merged$frpm_rate,
       c = 75, h = 5, p = 1,
       x.lim = c(65, 85),
       title = "RDD Plot at h = 5 (Wide Bandwidth)",
       x.label = "FRPM Eligibility (%)",
       y.label = "Chronic Absenteeism (%)")
```

```
## [1] "Mass points detected in the running variable."
```
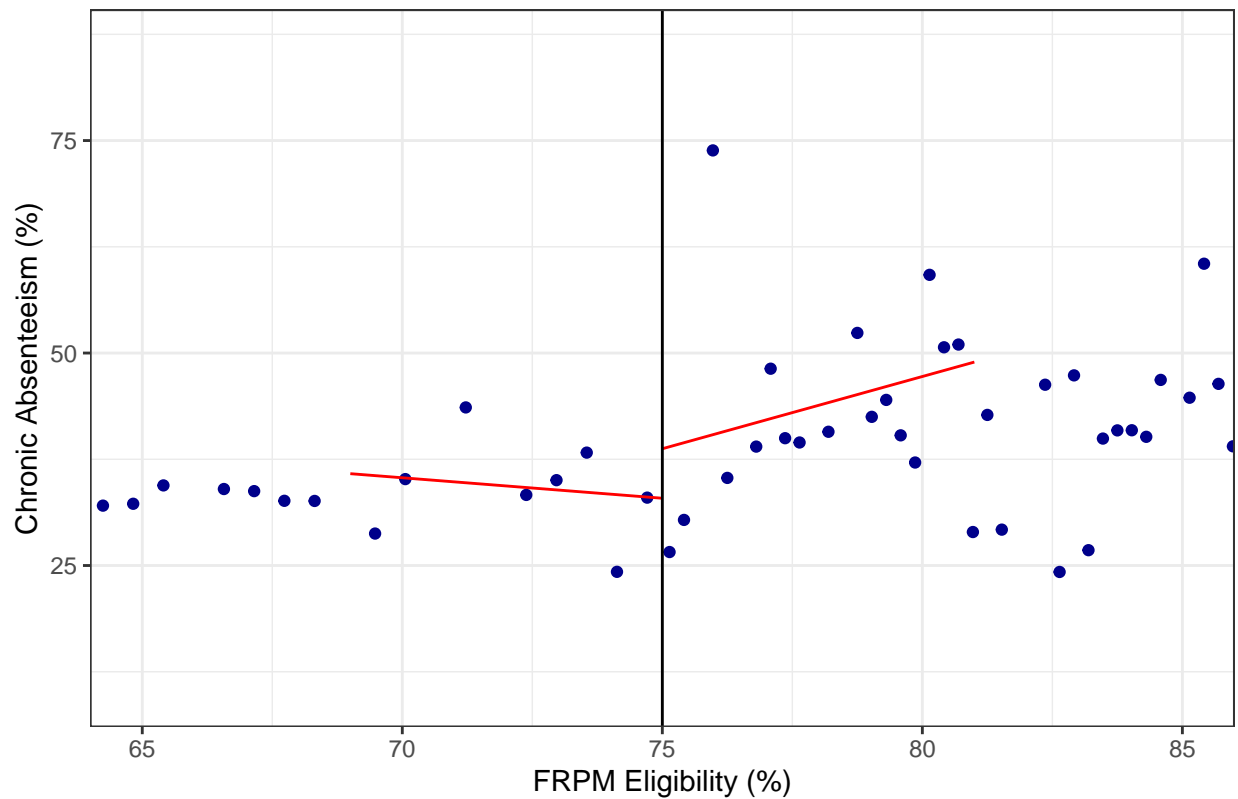
RDD Plot at h = 5 (Wide Bandwidth)


```
# Sensitivity: Even wider
rdplot(merged$chronic_absenteeism, merged$frpm_rate,
       c = 75, h = 6, p = 1,
       x.lim = c(65, 85),
       title = "RDD Plot at h = 6 (Extra Wide Bandwidth)",
       x.label = "FRPM Eligibility (%)",
       y.label = "Chronic Absenteeism (%)")
```

```
## [1] "Mass points detected in the running variable."
```

RDD Plot at h = 6 (Extra Wide Bandwidth)

VII. Run the McCrary Density Test

What: Check whether schools are manipulating FRPM % to qualify.

Why it's important: RDD assumes schools cannot manipulate assignment. Bunching invalidates causal claims.

How it supports your thesis: If there's no manipulation, the assignment can be seen as as-good-as-random, strengthening causal identification.

```r
library(rdrobust)
library(rddensity)

merged <- merged %>%
  mutate(
    frpm_rate = ifelse(frpm_rate <= 1, frpm_rate * 100, frpm_rate),
    running   = frpm_rate - 75,
    treated   = ifelse(frpm_rate >= 75, 1, 0)
  )

ggplot(merged, aes(frpm_rate, chronic_absenteeism)) +
  geom_point(alpha = .4) +
  geom_vline(xintercept = 75, linetype = "dashed") +
  labs(x = "FRPM eligibility rate (%)",
       y = "Chronic absenteeism rate (%)")
```

```
## Warning: Removed 4710 rows containing missing values or values outside the scale range
## (`geom_point()`).
```
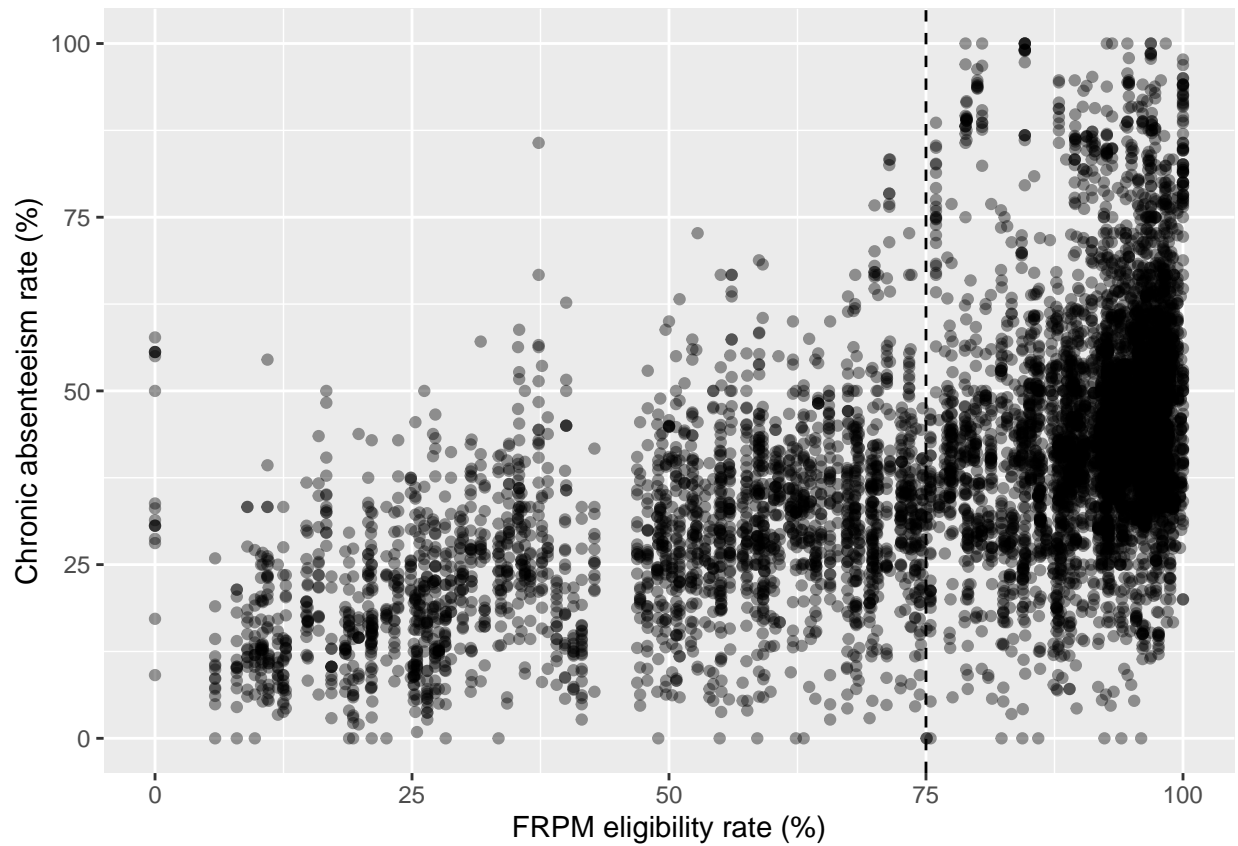


```
library(rddensity)

# 1. McCrary density test
dens75 <- rddensity(merged$frpm_rate, c = 75)
dens75$test              # prints t_jk and p_jk
```

```
## $t_asy
## [1] NA
##
## $t_jk
## [1] -1.545891
##
## $p_asy
## [1] NA
##
## $p_jk
## [1] 0.1221309
```

```
# 3. Density test statistics (use the right object)
dens75$test              # NOT rd75$test
```

```
## $t_asy
```

```
## [1] NA
##
## $t_jk
## [1] -1.545891
##
## $p_asy
## [1] NA
##
## $p_jk
## [1] 0.1221309
```

```r
# 4. McCrary density plot
# McCrary density test object
dens75 <- rddensity(merged$frpm_rate, c = 75)

# Density plot    rdd object first, X second
rdplotdensity(
  dens75,                   # ← rdd    (object returned by rddensity)
  merged$frpm_rate,         # ← X      (numeric running-variable vector)
  title  = "McCrary density plot at 75% FRPM",
  xlabel = "FRPM eligibility rate (%)",
  ylabel = "Estimated density"
)
```


McCrary density plot at 75% FRPM

```
## $Estl
```
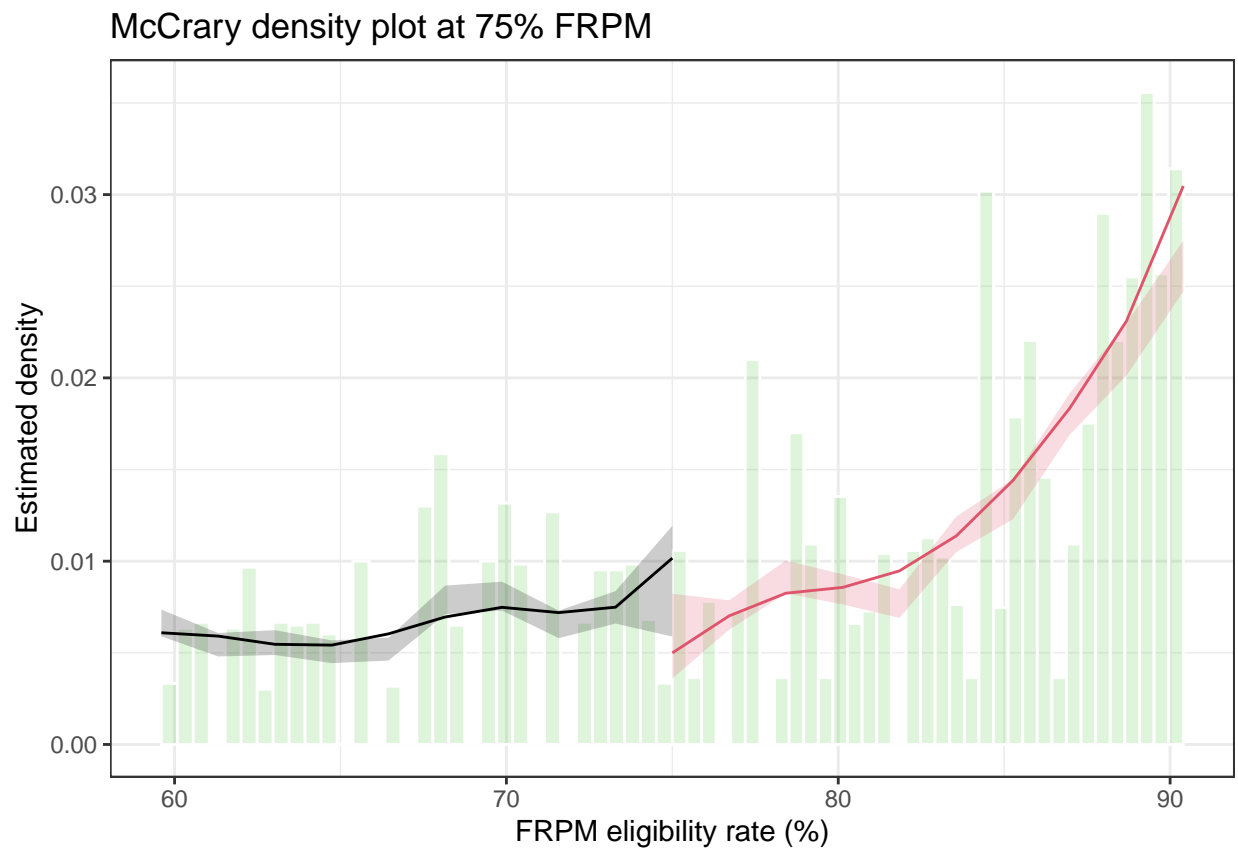
```
## Call: lpdensity
##
## Sample size                                      3583
## Polynomial order for point estimation    (p=)    2
## Order of derivative estimated            (v=)    1
## Polynomial order for confidence interval (q=)    3
## Kernel function                                  triangular
## Scaling factor                                   0.273511450381679
## Bandwidth method                                 user provided
##
## Use summary(...) to show estimates.
##
## $Estr
## Call: lpdensity
##
## Sample size                                      9518
## Polynomial order for point estimation    (p=)    2
## Order of derivative estimated            (v=)    1
## Polynomial order for confidence interval (q=)    3
## Kernel function                                  triangular
## Scaling factor                                   0.726564885496183
## Bandwidth method                                 user provided
##
## Use summary(...) to show estimates.
##
## $Estplot
```


McCrary density plot at 75% FRPM
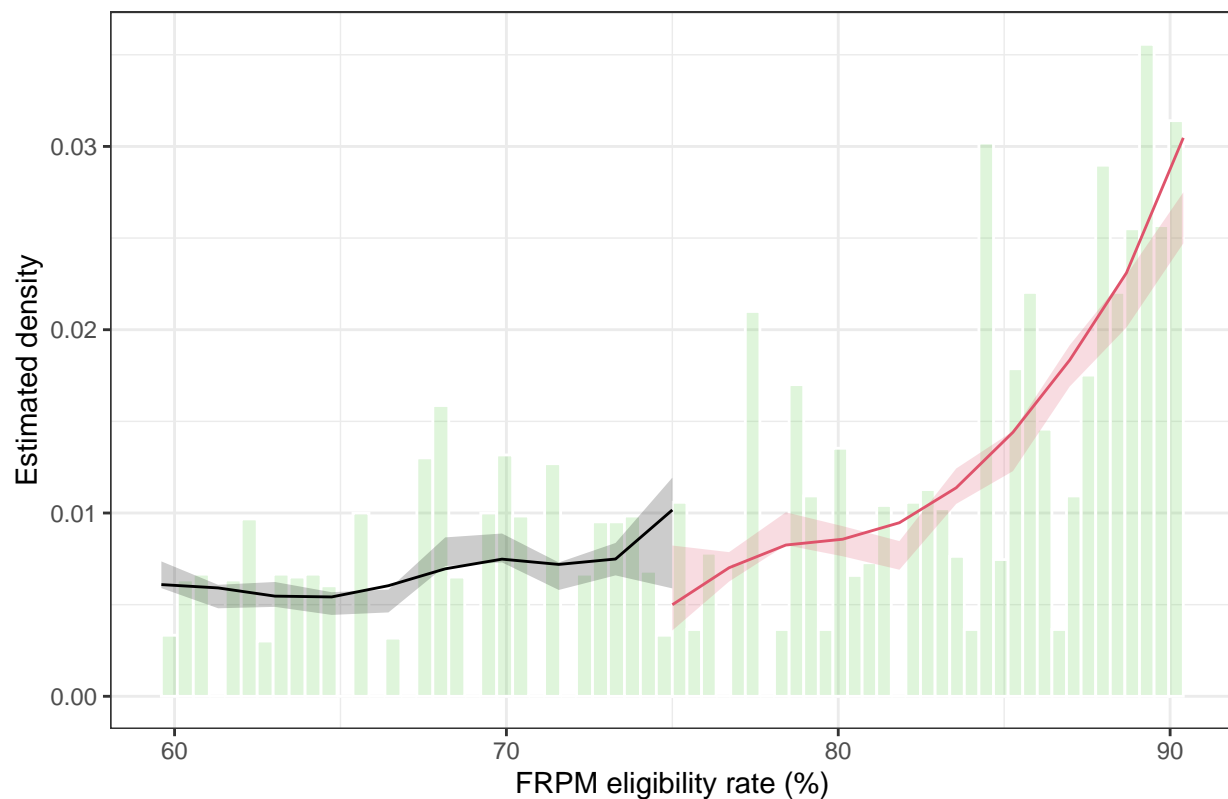
VIII. Covariate Balance Checks (In Progress)

IX. Robustness Checks

1. Bandwidth Sensitivity

Try different bandwidths (e.g., h = 2 to h = 6). Linear results are most credible. Avoid high-order polynomials due to overfitting.

```
# Fit models
bw_models <- list(
  h2 = rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 2, p = 1),
  h3 = rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 3, p = 1),
  h4 = rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 4.443, p = 1),
  h5 = rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 5, p = 1),
  h6 = rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 75, h = 6, p = 1)
)

# Create table
bw_results <- do.call(rbind, Map(function(model, h) {
  data.frame(
    Bandwidth = h,
    Coef = model$coef[1, 1],
    SE = model$se[1, 1],
    Pval = model$pv[1, 1],
    CI_Lower = model$ci[1, 1],
    CI_Upper = model$ci[1, 2]
  )
}, bw_models, c(2, 3, 4.443, 5, 6)))
knitr::kable(bw_results, digits = 3, caption = "Bandwidth Sensitivity: Local Linear RDD Estimates")
```

Table 1: Bandwidth Sensitivity: Local Linear RDD Estimates

	Bandwidth	Coef	SE	Pval	CI_Lower	CI_Upper
h2	2.000	-2.429	2.811	0.388	-7.939	3.081
h3	3.000	8.075	2.071	0.000	4.016	12.134
h4	4.443	10.368	1.734	0.000	6.969	13.767
h5	5.000	10.175	1.626	0.000	6.988	13.361
h6	6.000	8.157	1.441	0.000	5.332	10.981

2. Polynomial Order Sensitivity

Include results but de-emphasize cubic and quadratic fits. Note they are unstable and reverse sign — a common RDD issue.

```
# Models: m7, m8, m9 (h = 4, p = 1-3)
poly_models <- list(m7, m8, m9)
poly_info <- data.frame(Poly_Order = 1:3, Bandwidth = 4)

poly_table <- do.call(rbind, Map(function(model, spec) {
```

```
  data.frame(
    Poly_Order = spec$Poly_Order,
    Coef = model$coef[1,1],
    SE = model$se[1,1],
    Pval = model$pv[1,1],
    CI_Lower = model$ci[1,1],
    CI_Upper = model$ci[1,2]
  )
}, poly_models, split(poly_info, seq(nrow(poly_info))))))

knitr::kable(poly_table, digits = 3, caption = "Polynomial Order Sensitivity at h = 4")
```

Table 2: Polynomial Order Sensitivity at h = 4

Poly_Order	Coef	SE	Pval	CI_Lower	CI_Upper
1	9.984	1.862	0.000	6.334	13.634
2	1.781	2.766	0.520	-3.639	7.202
3	-13.603	4.368	0.002	-22.164	-5.042

3. Placebo Cutoffs (Next Step Suggestion)

Try cutoffs at 85% or 90% where no program change should occur. Null effects there will further validate your 75% results.

```
# Placebo cutoffs at 85% and 90%
p70 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 85, h = 4, p = 1, masspoints = "adjus
p80 <- rdrobust(merged$chronic_absenteeism, merged$frpm_rate, c = 90, h = 4, p = 1, masspoints = "adjus

placebo_results <- do.call(rbind, Map(function(model, cutoff) {
  data.frame(
    Placebo_Cutoff = cutoff,
    Coef = model$coef[1,1],
    SE = model$se[1,1],
    Pval = model$pv[1,1],
    CI_Lower = model$ci[1,1],
    CI_Upper = model$ci[1,2]
  )
}, list(p70, p80), c(70, 80)))

knitr::kable(placebo_results, digits = 3, caption = "Placebo Cutoffs (No Expected Treatment Effect)")
```

Table 3: Placebo Cutoffs (No Expected Treatment Effect)

Placebo_Cutoff	Coef	SE	Pval	CI_Lower	CI_Upper
70	0.102	1.245	0.935	-2.339	2.543
80	-0.358	1.046	0.732	-2.407	1.692