# Introduction to target trial emulation: Worked example

**Susan Hoffman, MPH**

Department of Epidemiology
Emory University

SPER Methods Workshop

June 17, 2024

# Follow along!

https://github.com/ssb214/SPER24

# Study background

+ Caniglia, et al. (2023) – on GitHub

+ Exposure: Antibiotic initiation at or after 24 weeks gestation through 36 weeks gestation

+ Outcome: Preterm delivery

# Causal contrast

+ Intention to treat effect

+ The effect of being assigned to treatment, not the effect of the treatment itself

+ No confounders

Confounders

RCT Treatment Assignment → Treatment adherence → Outcome
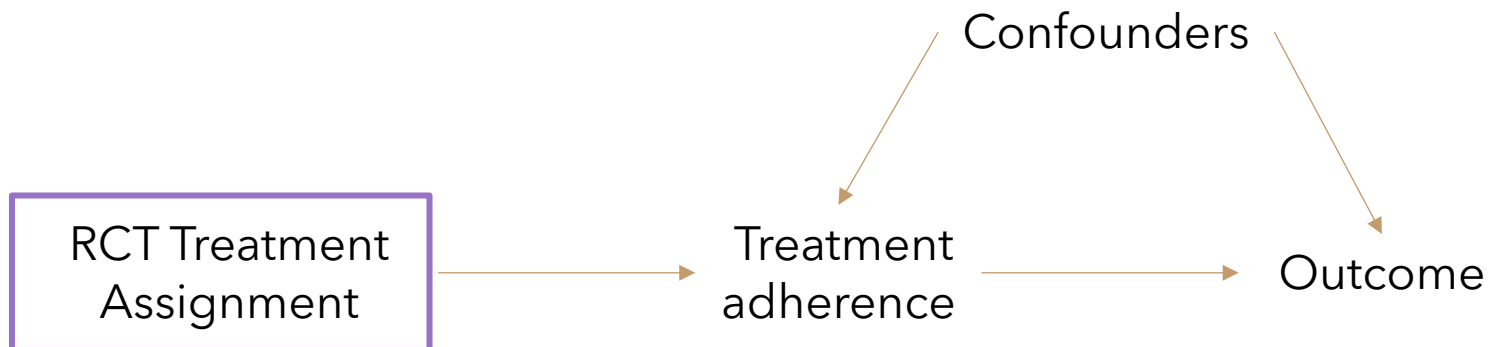
# Causal contrast

+ Observational analogue of the intention to treat effect
+ Now, we need to control for confounders (in this example, maternal age, parity, and HIV status)
+ The risk ratio is the estimate of interest

Confounders

Observational Treatment Ascertainment → Treatment adherence → Outcome
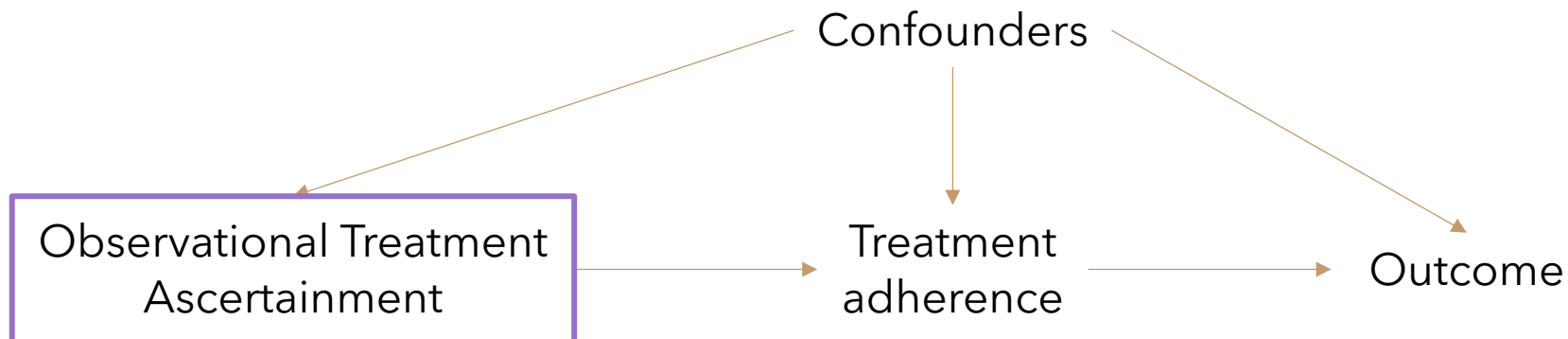
# "Data" structure

+ Define eligibility criteria

+ Specify time zero

+ Determine treatment ascertainment

+ Confounders

+ Follow-up and outcome

| participant_id | age | parity | hiv | week | exposure_info | outcome |
|---|---|---|---|---|---|---|
| 1 | 18-35 | 1 | 0 | wk_24 | 0 | 0 |
| 1 | 18-35 | 1 | 0 | wk_25 | 0 | 1 |
| 1 | 18-35 | 1 | 0 | wk_26 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_27 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_28 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_29 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_30 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_31 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_32 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_33 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_34 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_35 | 0 | NA |
| 1 | 18-35 | 1 | 0 | wk_36 | 0 | NA |
| 2 | 18-35 | 0 | 0 | wk_24 | 0 | 0 |
| 2 | 18-35 | 0 | 0 | wk_25 | 0 | 0 |
| 2 | 18-35 | 0 | 0 | wk_26 | 0 | 0 |
| 2 | 18-35 | 0 | 0 | wk_27 | 0 | 0 |
| 2 | 18-35 | 0 | 0 | wk_28 | 0 | 0 |

# Code overview

https://github.com/ssb214/SPER24

**Step 1: estimate a RR for each target trial within week 24-36**

```r
adj_model_results[[wk]] <- glm(outcome ~ exposure_info +
                hiv +
                parity +
                age,
            data = data_wk,
            family = binomial(link = "log")) # log link for RR
}
```

# General model for weeks 24-36

# Quick note: Families and links

| Odds Ratio | Risk Ratio | Risk Difference |
|---|---|---|
| Family = Binomial | Family = Binomial | Family = Binomial |
| Link = Logistic | Link = Log | Link = Identity |
| SE = model-based | SE = model-based | SE = model-based |
| | | |
| | Family = Poisson | Family = Gaussian |
| | Link = Log | Link = Identity |
| | SE = Bootstrap | SE = Bootstrap |

# Sequential target trials within each week

```r
# Create a list to store results for each week
adj_model_results <- list()

for (wk in paste0("wk_", 24:36)) {
  # Subset data for the current week
  data_wk <- final_data %>% filter(week == wk)

  # Fit log-binomial regression model and store results
  adj_model_results[[wk]] <- glm(outcome ~ exposure_info +
                                  hiv +
                                  parity +
                                  age,
                              data = data_wk,
                              family = binomial(link = "log")) # log link for RR
}
```

# Pulling results

```r
# Pull RR (exp. beta values) and 95% confidence intervals into a dataframe
## Create a null dataframe
results_adj <- data.frame(
  Week = character(),
  Coefficient = numeric(),
  Lower_CI = numeric(),
  Upper_CI = numeric(),
  stringsAsFactors = FALSE
)

for (wk in paste0("wk_", 24:36)) {
  # Extract coefficients and confidence intervals
  coefficients <- exp(crude_model_results[[wk]]$coefficients)
  ci <- exp(confint(crude_model_results[[wk]]))

  # Store results in the data frame
  results_adj <- rbind(results_adj, data.frame(
    Week = wk,
    Coefficient = coefficients["exposure_info"],
    Lower_CI = ci["exposure_info", "2.5 %"],
    Upper_CI = ci["exposure_info", "97.5 %"]
  ))
}
```

**Step 2: estimate RR for pooled results across weeks 24-36**

# Pooled model

```r
library(splines)

model_pooled <- glm(outcome ~ exposure_info +
                    hiv +
                    parity +
                    age +
                    ns(week_num, df = 2),  # adding flexibility to the model for GA
                    data = final_data,
                    family = binomial(link = "log"))  # log link for RR
```
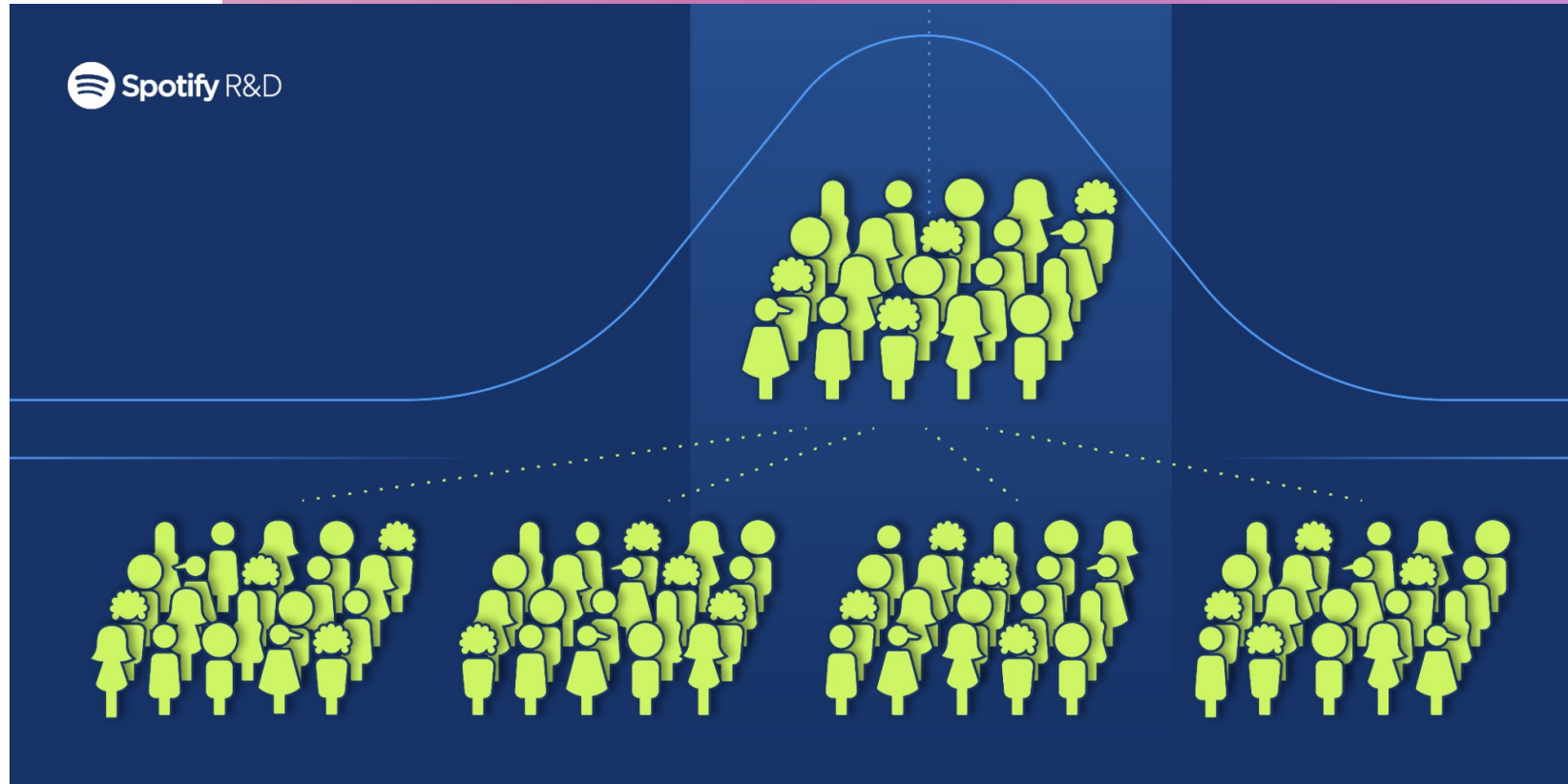
# Pooled results

+ Now we have the Beta estimates for the model, but because people can be present in more than one trial, we need to use the bootstrap for SE



```
Coefficients:
                           Estimate Std. Error   z value Pr(>|z|)
(Intercept)               -1.969904   0.017202   114.515   2e-16 ***
exposure_info              0.023687   0.027931     0.848   0.396
hiv1                      -0.009860   0.007708    -1.    0.201
parity1                    0.008752   0.006526     1.     0.180
age>35                     0.001125   0.017756     0.063   0.949
age18-35                   0.001545   0.016367     0.094    .925
ns(week_num, df = 2)1     -0.449437   0.0158      -28.415  <2e-16 ***
ns(week_num, df = 2)2     -0.696855   0.0    062   -40.842  <2e-1  ***
```

# The 👢 bootstrap

https://engineering.atspotify.com/2022/03/comparing-quantiles-at-scale-in-online-a-b-testing/

# The 👢 bootstrap

```
replicates <- 25 # this needs to be increased for a real analysis
bootstrap_estimates <- NULL
```

# The 👢 bootstrap

```r
replicates <- 25 # this needs to be increased for a real analysis
bootstrap_estimates <- NULL

for (i in 1:replicates) {
  # set seed to i so we get a uniquw sample each time
  set.seed(i)

  # sample the data
  index <- sample(1:nrow(final_data), nrow(final_data), replace = T)

  final_data_sample <- final_data[index,]
```

```
> index[1:10]
 [1]  452737  124413  856018   25173 1343338 1441262  640775  538191 1318949  549593
```

# The 👢 bootstrap

```r
replicates <- 25 # this needs to be increased for a real analysis
bootstrap_estimates <- NULL

for (i in 1:replicates) {
  # set seed to i so we get a uniquw sample each time
  set.seed(i)

  # sample the data
  index <- sample(1:nrow(final_data), nrow(final_data), replace = T)

  final_data_sample <- final_data[index,]

  # estimate the association in the resample
  model_boot <- glm(outcome ~ exposure_info +
                    hiv +
                    parity +
                    age +
                    ns(week_num, df = 2), # adding flexibility to the model for wee
                  data = final_data_sample,
                  family = binomial(link = "log")) # log link for RR

  bootstrap_estimates <- rbind(
    bootstrap_estimates,
    coef(model_boot)[2]
  )

  # Print iteration number when i is divisible by 10
  if (i %% 10 == 0) {
    print(paste("Iteration:", i))
  }
}
```

# Pooled results

```r
se_estimate <- sd(bootstrap_estimates)

lcl_pooled <- exp(coef(model_pooled)[2] - 1.96 * se_estimate)
ucl_pooled <- exp(coef(model_pooled)[2] + 1.96 * se_estimate)

pooled_dat <- data.frame(
  Week = "pooled",
  Coefficient = co_pooled[2],
  Lower_CI = lcl_pooled,
  Upper_CI = ucl_pooled
)

results_adj <- rbind(results_adj, pooled_dat)
```

| Antibiotic Initiation at: | Initiation | # of individuals | Events, # | RR (95% CI) |
|---|---|---|---|---|
| 24 weeks | No | 110082 | 15659 | 0.99 |
| | Yes | 1321 | 186 | (0.86, 1.12) |
| 25 weeks | No | 108628 | 13186 | 1.06 |
| | Yes | 1454 | 186 | (0.92, 1.21) |
| 26 weeks | No | 107316 | 10975 | 0.99 |
| | Yes | 1312 | 133 | (0.84, 1.16) |
| 27 weeks | No | 105876 | 8963 | 0.99 |
| | Yes | 1440 | 124 | (0.84, 1.17) |
| 28 weeks | No | 104555 | 7679 | 0.98 |
| | Yes | 1321 | 95 | (0.80, 1.17) |
| 29 weeks | No | 103095 | 6289 | 0.96 |
| | Yes | 1460 | 85 | (0.79, 1.17) |
| 30 weeks | No | 101656 | 5265 | 1.28 |
| | Yes | 1439 | 100 | (1.06, 1.52) |
| 31 weeks | No | 100249 | 4446 | 1.24 |
| | Yes | 1407 | 74 | (0.99, 1.51) |
| 32 weeks | No | 98962 | 3636 | 1.06 |
| | Yes | 1287 | 50 | (0.80, 1.35) |
| 33 weeks | No | 97722 | 2972 | 0.91 |
| | Yes | 1240 | 34 | (0.65, 1.24) |
| 34 weeks | No | 1317 | 2327 | 0.74 |
| | Yes | 96405 | 24 | (0.49, 1.06) |
| 35 weeks | No | 1303 | 1691 | 1.07 |
| | Yes | 95102 | 26 | (0.72, 1.52) |
| 36 weeks | No | 93841 | 1037 | 0.62 |
| | Yes | 1261 | 8 | (0.29, 1.15) |
| Pooled (24-36 weeks) | No | 1323489 | 84125 | 1.02 |
| | Yes | 17562 | 1125 | (0.98, 1.07) |



Risk of PTB
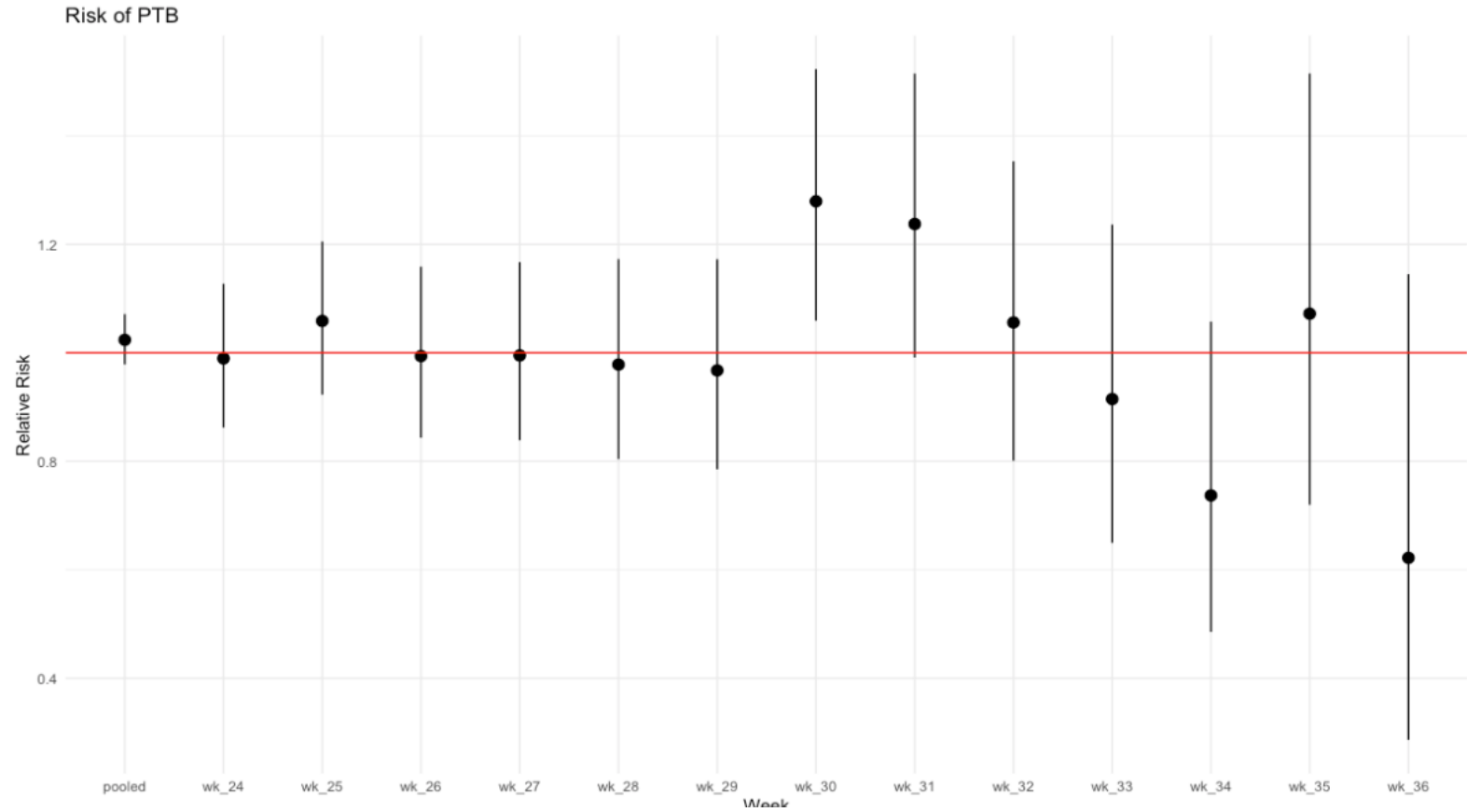
# Questions?

Email:
ssbucke@emory.edu