Target Trial Simulation

Susan Hoffman, MPH

Simulation Parameters

Exposure:

- The primary causal question of interest focuses on antibiotic initiation at or after 24 weeks gestation

Outcome

- Preterm delivery

Causal contrast:

- Intention to treat effect

Exchangeability considerations:

- Age (<18 - 4%; 18-35 - 79%; >=35 - 17%)
- Parity (0 - 41%, >=1 - 59%)
- HIV status (Negative - 78%; Living with HIV - 22%)

Data generation

This synthetic data set was simulated based on Caniglia, et al. (2023) and is not representative of a real data set or real causal effects. This data is meant to be used for this sandbox example only and will not match what was seen in the manuscript.

Analysis

```
# Getting number of exposed participants per week
exposure_counts <- final_data %>%
  group_by(week, exposure_info) %>%
  summarise(count = n()) %>%
  pivot_wider(names_from = exposure_info,
              values_from = count,
              names_prefix = "exposure_") %>%
  mutate(exposure_1 = coalesce(exposure_1, 0),
         exposure_0 = coalesce(exposure_0, 0))

# Getting the number of outcome events per week by exposure status
outcome_counts <- final_data %>%
  group_by(week, exposure_info) %>%
  summarise(outcome_events = sum(outcome, na.rm = TRUE))

# Data cleaning - creating factor variables

factor_names <- c('age', 'hiv', 'parity')

final_data[, factor_names] <- lapply(final_data[, factor_names], factor)

head(final_data)
```

```
  participant_id   age parity hiv  week exposure_info outcome
1              1 18-35      1   0 wk_24             0       0
2              1 18-35      1   0 wk_25             0       1
3              1 18-35      1   0 wk_26             0      NA
4              1 18-35      1   0 wk_27             0      NA
5              1 18-35      1   0 wk_28             0      NA
6              1 18-35      1   0 wk_29             0      NA
```

```
# Data cleaning - creating numeric week variable

final_data$week_num <- as.numeric(gsub("wk_", "", final_data$week))

# adjusted models by week

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

```r
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
}


# 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(adj_model_results[[wk]]$coefficients)
  ci <- exp(confint(adj_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 %"]
  ))
}

# Pooled results

library(splines)
```

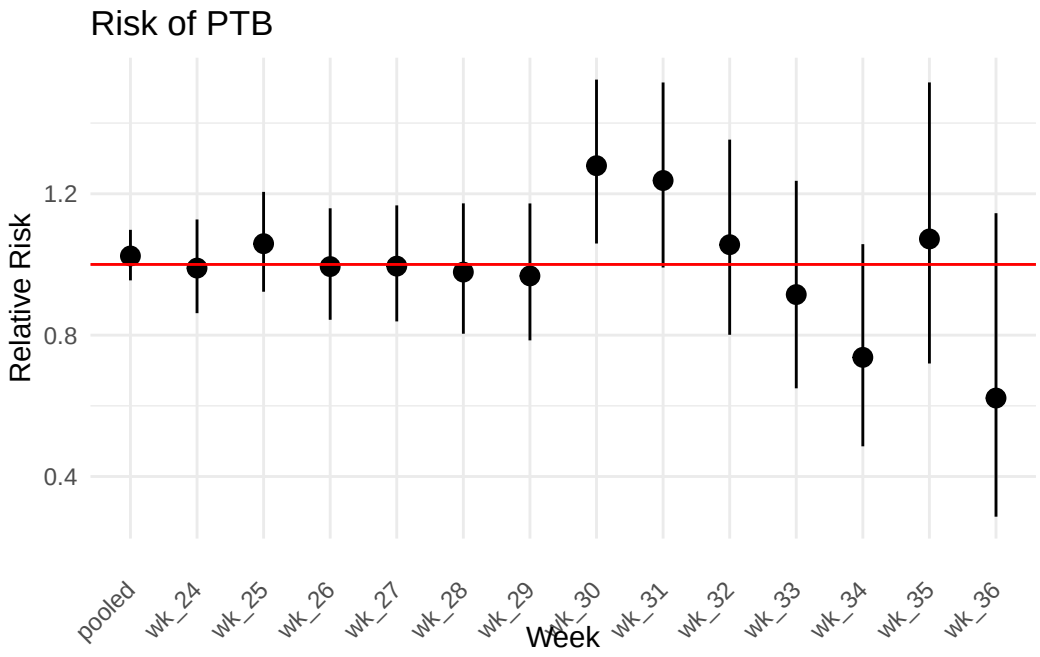
```r
model_pooled <- glm(outcome ~ exposure_info +
                      hiv +
                      parity +
                      age +
                      ns(week_num, df = 2), # adding flexibility to the model for GA
                    data = final_data,
                    family = binomial(link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

co_pooled <- exp(model_pooled$coefficients)

# need to bootstrap confidence intervals
replicates <- 5 # 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 unique 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 GA
                    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))
  }
```

```
}

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)
```

Tables and Figures

RR summary figure

```
library(ggplot2)

ggplot(results_adj, aes(x = Week,
                        y = Coefficient)) +
  geom_pointrange(aes(ymin = Lower_CI,
                      ymax = Upper_CI),
                  color = "black") +
  geom_point(color = "black",
             size = 3) +
  geom_hline(yintercept = 1,
             linetype = 'solid',
             color = 'red') +
  labs(x = 'Week',
       y = "Relative Risk",
       title = "Risk of PTB") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5, hjust = 1))
```



Table

To generate Latex tables online: https://www.tablesgenerator.com/latex_tables#

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)