Supplementary Materials of Impact of IPTp Regimen on Pregnancy Outcomes

# Overview

The Supplementary Appendix begins with comprehensive methodological details, including variable‐by‐variable missingness (Table S1), analysis of deviance comparing models with and without the malaria×SP interaction (Table S2), and variance inflation factors for the final interaction model (Table S3). It also presents machine‐learning tuning results with the top five elastic-net hyperparameter combinations ranked by mean cross-validated AUC (Table S4) alongside a heatmap illustrating AUC across the penalty–mixture grid (Figure S1). Full code excerpts document our data-cleaning steps, rsample splits, recipe definitions, and tune\_grid workflows. The appendix then moves on to additional results: a detailed stratification of outcome measures and malaria-exposure variables by IPTp arm (Table S5), bar graphs of total malaria episodes during pregnancy (Figure S2), gravidity (Figure S3), and parity (Figure S4) distributions by treatment arm, and a bootstrap calibration curve for the interaction model (Figure S5). Finally, it compares model discrimination with overlaid ROC curves for logistic regression, random forest, and XGBoost (Figure S6), presents a precision-recall curve for the top‐performing machine-learning model (Figure S7), and reports the test-set AUC for the gravidity-only model in women under 25 years (Table S6).

# Code and file information

Explain here what each code/file is and does, and in which order (if any) users need to run thing to reproduce everything. Essentially, give a full set of instructions to re-generate everything.

# Additional Method Details

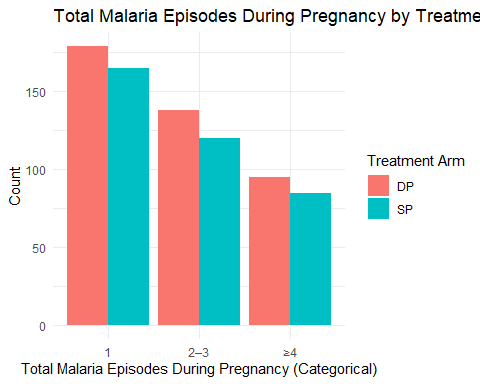
Often, the main manuscript only allows for an overview description of the methods. Use the supplement to describe all your methods, models and approaches in a lot of detail. Reference specific parts of your code as needed.

# Additional results

Show additional results here. Those can be some useful exploratory/descriptive figures or tables, or results from additional analyses that didn’t make it into the main text.

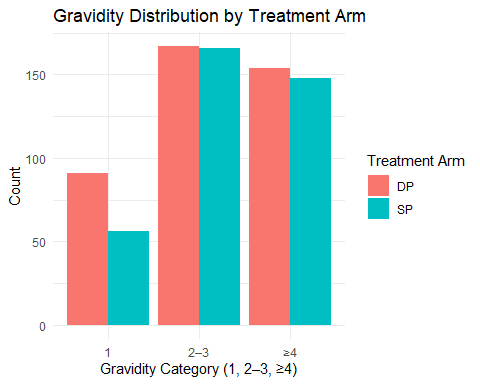
*Bar Graph of Total Malaria Episodes During Pregnancy by Treatment Arm*

# Load necessary libraries  
library(dplyr)  
library(ggplot2)  
library(here)  
  
# Read in the cleaned data  
promo\_data\_clean <- read.csv(here("data", "clean", "PROMO\_Data\_clean.csv"))  
  
# Recode 'total\_malaria\_episodes\_during\_pregnancy' into a categorical variable  
promo\_data\_clean <- promo\_data\_clean %>%  
 mutate(  
 MalariaEpisodesPreg\_cat = case\_when(  
 total\_malaria\_episodes\_during\_pregnancy %in% c(0, 1) ~ "1",  
 total\_malaria\_episodes\_during\_pregnancy %in% c(2, 3) ~ "2–3",  
 total\_malaria\_episodes\_during\_pregnancy >= 4 ~ "≥4"  
 ),  
 MalariaEpisodesPreg\_cat = factor(MalariaEpisodesPreg\_cat, levels = c("1", "2–3", "≥4"))  
 )  
  
# Create the bar graph  
ggplot(promo\_data\_clean, aes(x = MalariaEpisodesPreg\_cat, fill = study\_arm)) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Total Malaria Episodes During Pregnancy by Treatment Arm",  
 x = "Total Malaria Episodes During Pregnancy (Categorical)",  
 y = "Count",  
 fill = "Treatment Arm") +  
 theme\_minimal()



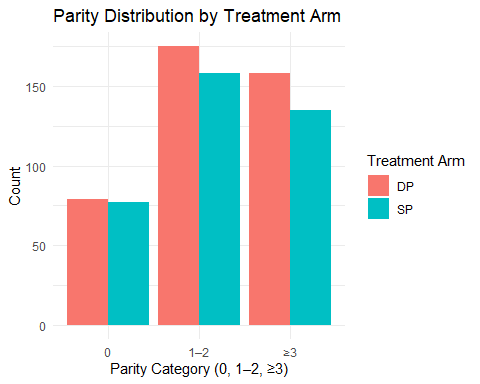
*Bar Graphs for Gravidity and Parity*

# Load necessary libraries  
  
# Read in the cleaned data  
promo\_data\_clean <- read.csv(here("data", "clean", "PROMO\_Data\_clean.csv"))  
  
# Recode Gravidity into categories: "1", "2–3", "≥4"  
promo\_data\_clean <- promo\_data\_clean %>%  
 mutate(  
 Gravidity\_cat = case\_when(  
 gravidity == 1 ~ "1",  
 gravidity %in% c(2, 3) ~ "2–3",  
 gravidity >= 4 ~ "≥4"  
 ),  
 Gravidity\_cat = factor(Gravidity\_cat, levels = c("1", "2–3", "≥4"))  
 )  
  
# Create the bar graph for Gravidity by Treatment Arm  
ggplot(promo\_data\_clean, aes(x = Gravidity\_cat, fill = study\_arm)) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Gravidity Distribution by Treatment Arm",  
 x = "Gravidity Category (1, 2–3, ≥4)",  
 y = "Count",  
 fill = "Treatment Arm") +  
 theme\_minimal()



*Parity Distribution by Treatment Arm*

# Load necessary libraries  
library(dplyr)  
library(ggplot2)  
library(here)  
  
# Read in the cleaned data  
promo\_data\_clean <- read.csv(here("data", "clean", "PROMO\_Data\_clean.csv"))  
  
# Recode Parity into categories: "0", "1–2", "≥3"  
promo\_data\_clean <- promo\_data\_clean %>%  
 mutate(  
 Parity\_cat = case\_when(  
 parity == 0 ~ "0",  
 parity %in% c(1, 2) ~ "1–2",  
 parity >= 3 ~ "≥3"  
 ),  
 Parity\_cat = factor(Parity\_cat, levels = c("0", "1–2", "≥3"))  
 )  
  
# Create the bar graph for Parity by Treatment Arm  
ggplot(promo\_data\_clean, aes(x = Parity\_cat, fill = study\_arm)) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Parity Distribution by Treatment Arm",  
 x = "Parity Category (0, 1–2, ≥3)",  
 y = "Count",  
 fill = "Treatment Arm") +  
 theme\_minimal()



## Example additional result

[Table 1](#tbl-resulttable1) shows an additional table summarizing a model fit.

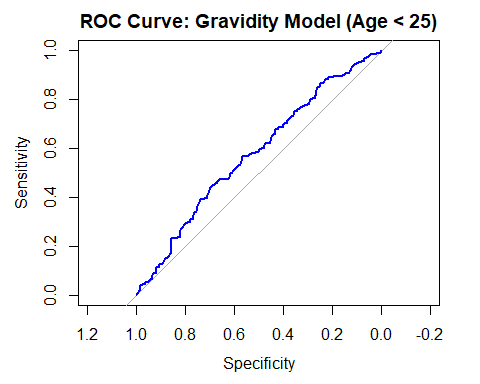
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Another fit table.   | term | estimate | std.error | statistic | p.value | | --- | --- | --- | --- | --- | | (Intercept) | 149.6997661 | 19.7518528 | 7.5790240 | 0.0001285 | | Weight | 0.2277371 | 0.2708841 | 0.8407177 | 0.4282860 | |

[Figure 1](#fig-result2) shows a scatterplot figure produced by one of the R scripts.

|  |
| --- |
| Figure 1: Height and weight. |

# Evaluating Discrimination (ROC Curve and AUC)

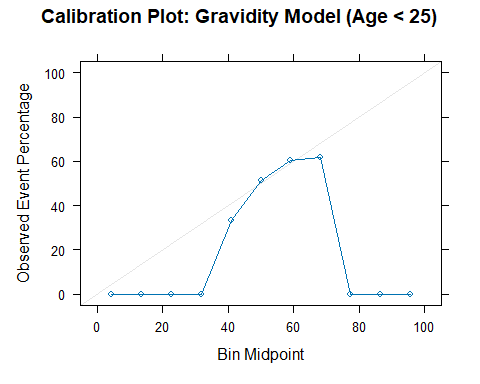
# 1) Load required libraries  
library(dplyr)  
library(pROC)  
library(caret)  
library(here)  
  
# 2) Read cleaned data and subset to women < 25  
promo\_data\_clean <- read.csv(here("data/clean/PROMO\_Data\_clean.csv"))  
promo\_data\_young <- promo\_data\_clean %>%  
 filter(age\_at\_enrollment\_years < 25) %>%  
 mutate(  
 # Composite adverse outcome: 1 if preterm, stillbirth, or low birth weight  
 low\_birth\_weight = as.integer(birth\_weight < 2.5),  
 adverse\_birth\_outcome = as.integer(  
 preterm\_births\_count > 0 |  
 stillbirth\_bin == 1 |  
 low\_birth\_weight == 1  
 )  
 )  
  
# 3) Fit logistic regression model  
model\_gravidity <- glm(  
 adverse\_birth\_outcome ~ gravidity + total\_malaria\_episodes +  
 study\_arm + education\_level,  
 family = binomial(link = "logit"),  
 data = promo\_data\_young  
)  
  
# 4) Generate predicted probabilities  
promo\_data\_young$predicted\_prob <- predict(  
 model\_gravidity,  
 newdata = promo\_data\_young,  
 type = "response"  
)  
  
# 5) ROC curve and AUC  
roc\_obj <- roc(  
 promo\_data\_young$adverse\_birth\_outcome,  
 promo\_data\_young$predicted\_prob  
)  
plot(  
 roc\_obj,  
 col = "blue",  
 lwd = 2,  
 main = "ROC Curve: Gravidity Model (Age < 25)"  
)



auc\_val <- auc(roc\_obj)  
cat("AUC:", round(auc\_val, 2), "\n")

AUC: 0.58

# 6) Calibration plot  
data\_cal <- data.frame(  
 observed = factor(promo\_data\_young$adverse\_birth\_outcome, levels = c(0,1)),  
 predicted = promo\_data\_young$predicted\_prob  
)  
cal\_plot <- calibration(  
 observed ~ predicted,  
 data = data\_cal,  
 class = "1"  
)  
plot(cal\_plot, main = "Calibration Plot: Gravidity Model (Age < 25)")



I see that the ROC curve for my model is close to the diagonal, with an AUC only slightly above 0.5. This tells me that the model doesn’t have strong discriminative ability for predicting adverse outcomes in women under 25. Additionally, the calibration plot shows that my predicted probabilities often stray from the ideal diagonal—especially in the mid-range—indicating that my model’s risk estimates don’t consistently match the observed rates. Overall, while gravidity is statistically significant, my model as a whole isn’t very effective at distinguishing between those who experience adverse outcomes and those who don’t, and its probability estimates need improvement.

During 5‑fold cross‑validation, the elastic‐net model achieved the highest mean AUC (0.55±0.01), followed by random forest (0.52±0.01) and XGBoost (0.50±0.01).

# Discussion

# References