proj4

Jiayi Shi

2025-04-10

# Load data

sae <- read\_excel("Q2b.xlsx") %>%   
 janitor::clean\_names()  
baseline <- read\_excel("Q2b\_BL.xlsx") %>%   
 janitor::clean\_names() %>%   
 mutate(sex = factor(sex, levels = c(0, 1), labels = c("female", "male")))  
  
df <- left\_join(sae, baseline, by="id") %>%   
 mutate(group = factor(group, levels = c(0,1), labels = c("Control", "Vaccine")),  
 time = factor(time, levels = c(1,2,3), labels = c("Month1", "Month2", "Month3")))  
  
infection\_df <- read\_excel("Q2c.xlsx") %>%   
 janitor::clean\_names()

### Part A: SAE Analysis

# GEE logistic regression   
#gee\_sae <- geeglm(sae ~ group \* time + sex + age, data = df, id = , family = binomial, corstr = "ar1")  
#summary(gee\_sae)  
  
glmm\_sae <- glmmTMB::glmmTMB(sae ~ group \* time + sex + age + (1|site:id), data = df, family = binomial())  
summary(glmm\_sae)

## Family: binomial ( logit )  
## Formula: sae ~ group \* time + sex + age + (1 | site:id)  
## Data: df  
##   
## AIC BIC logLik deviance df.resid   
## 1645.8 1732.3 -813.9 1627.8 111443   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## site:id (Intercept) 120.7 10.99   
## Number of obs: 111452, groups: site:id, 41194  
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -13.74729 1.09735 -12.528 <2e-16 \*\*\*  
## groupVaccine 0.70602 0.55309 1.277 0.2018   
## timeMonth2 -0.01612 0.39483 -0.041 0.9674   
## timeMonth3 0.46712 0.39681 1.177 0.2391   
## sexmale -0.09682 0.46328 -0.209 0.8345   
## age 0.01338 0.02088 0.641 0.5218   
## groupVaccine:timeMonth2 -0.28876 0.53495 -0.540 0.5893   
## groupVaccine:timeMonth3 -1.09788 0.54407 -2.018 0.0436 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#vcov(glmm\_sae)  
  
# compares groups within each time point  
library(emmeans)  
emmeans(glmm\_sae, pairwise ~ group | time)

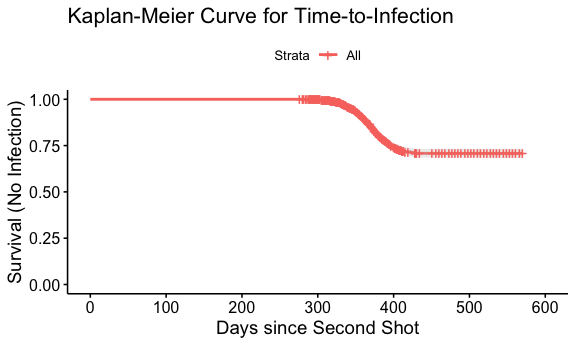
## $emmeans  
## time = Month1:  
## group emmean SE df asymp.LCL asymp.UCL  
## Control -13.2 0.489 Inf -14.2 -12.2  
## Vaccine -12.5 0.433 Inf -13.3 -11.6  
##   
## time = Month2:  
## group emmean SE df asymp.LCL asymp.UCL  
## Control -13.2 0.490 Inf -14.2 -12.3  
## Vaccine -12.8 0.458 Inf -13.7 -11.9  
##   
## time = Month3:  
## group emmean SE df asymp.LCL asymp.UCL  
## Control -12.7 0.460 Inf -13.6 -11.8  
## Vaccine -13.1 0.480 Inf -14.1 -12.2  
##   
## Results are averaged over the levels of: sex   
## Results are given on the logit (not the response) scale.   
## Confidence level used: 0.95   
##   
## $contrasts  
## time = Month1:  
## contrast estimate SE df z.ratio p.value  
## Control - Vaccine -0.706 0.553 Inf -1.277 0.2018  
##   
## time = Month2:  
## contrast estimate SE df z.ratio p.value  
## Control - Vaccine -0.417 0.569 Inf -0.733 0.4636  
##   
## time = Month3:  
## contrast estimate SE df z.ratio p.value  
## Control - Vaccine 0.392 0.565 Inf 0.693 0.4881  
##   
## Results are averaged over the levels of: sex   
## Results are given on the log odds ratio (not the response) scale.

### Part B: Time-to-Infection Analysis

sum(infection\_df$infection)

## [1] 433

# Fit a Kaplan-Meier survival curve.  
km\_fit <- survfit(Surv(enrollment\_time, last\_fu\_time, infection) ~ 1, data = infection\_df)  
  
# Plot the survival curve  
ggsurvplot(km\_fit, conf.int = TRUE,  
 xlab = "Days since Second Shot",   
 ylab = "Survival (No Infection)",  
 title = "Kaplan-Meier Curve for Time-to-Infection")



summary(km\_fit, times = 365)

## Call: survfit(formula = Surv(enrollment\_time, last\_fu\_time, infection) ~   
## 1, data = infection\_df)  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 365 1318 215 0.876 0.00799 0.86 0.892

median <- summary(km\_fit)$table["median"]  
cat("Median time-to-infection (days):", median, "\n")

## Median time-to-infection (days): NA

# the median time-to-infection might not be reached (i.e., the survival function does not drop below 50%)  
  
restricted\_mean <- summary(km\_fit)$table["rmean"]  
  
cat("Restricted mean time-to-infection (days):", restricted\_mean, "\n")

## Restricted mean time-to-infection (days): 511.5923

# Note:  
# In many vaccine studies, if a majority of subjects do not become infected,  
  
# and the mean time-to-infection (restricted mean) should be reported with the time horizon clearly defined.