COVID

Ryan Wei UNI: rw2844

1 Study design

1.1 Randomization Procedure

Participants will be randomly assigned to receive injections of either100 μg of mRNA-1273 vaccine or a placebo control in a 1:1 randomization ratio. Randomization will be stratified based on age and, if they are < 65 years of age, based on the presence or absence of risk factors for severe illness from COVID-19 based on CDC recommendation as of Mar 2020. There will be 3 strata for randomization: ≥ 65 years, < 65 years and categorized to be at increased risk ("at risk") for the complications of COVID-19, and < 65 years "not at risk".

Risk will be defined based on the study participants' relevant past and current medical history. At least 25% of enrolled participants, but not more than 40%, will be either ≥ 65 years of age or < 65 years of age and "at risk" at Screening.

Participants who are less than 65 years old will be categorized as at risk for severe COVID-19 illness if they have at least 1 of the following risk factors at Screening:

- Chronic lung disease (eg, emphysema and chronic bronchitis, idiopathic pulmonary fibrosis, and cystic fibrosis) or moderate to severe asthma
- Significant cardiac disease (eg, heart failure, coronary artery disease, congenital heart disease, cardiomyopathies, and pulmonary hypertension)
- Severe obesity (body mass index ≥ 40 kg/m2)
- Diabetes (Type 1, Type 2 or gestational)
- Liver disease

- Human Immunodeficiency Virus (HIV) infection

All participants will be assessed for efficacy and safety endpoints and provide a nasopharyngeal (NP) swab sample and blood sample before the first and second dose of IP in addition to a series of post-dose blood samples for immunogenicity through 24 months after the second dose of IP. Efficacy assessments will include surveillance for COVID-19 with RT-PCR confirmation of SARS-CoV-2 infection after the first and second dose of IP.

1.2 Statistical Analysis

Statistical Hypotheses: For the primary efficacy objective, the null hypothesis of this study is that the VE of mRNA-1273 to prevent first occurrence of COVID-19 is $\leq 30\%$ (ie, H_0 : efficacy: VE ≤ 0.3).

The study will be considered to meet the primary efficacy objective if the corresponding CI of VE rules out 30% at the primary analysis. In the primary analysis of VE of COVID-19, cases will be counted starting 14 days after the second dose of IP.

Vaccine efficacy is defined as the percent reduction in the hazard of the primary endpoint (mRNA-1273 vs placebo). Equivalently, the null hypothesis is: - H_0 : efficacy: hazard ratio (HR) ≥ 0.7 (equivalently, proportional hazards VE ≤ 0.3).

A stratified Cox proportional hazard model will be used to assess the magnitude of the treatment group difference (ie, HR) between mRNA-1273 and placebo at a 1-sided 0.025 significance level

1.3 Sample Size Calculation

- incident rate in the vaccine group : p_T
- incident rate in the control group : $p_C = 0.0075$
- vaccine efficacy: $\pi = 1 - \frac{p_T}{p_C} = 1 - R$
- null hypothesis: $H_0 : \pi \leq \pi_0$
- $u = n_T/n_C$

1.4 Method 1

- number of cases in the vaccine group follows a Poisson distribution with rate $\lambda_T = n_T p_T$
- number of cases in the controlled group follows a Poisson distribution with rate $\lambda_C = n_C p_C$
- number of cases in the vaccine group given the total number of cases S: Binomial(S, θ), where $\theta = \frac{\lambda_T}{\lambda_T + \lambda_C} = \frac{1-\pi}{1-\pi+u}$
- rewrite null hypothesis $H_0 : \theta \geq \theta_0$
- target risk reduction: $\pi_1 = 0.6 = 1 - \frac{p_T}{p_C}$, $p_T = 0.003$, $\theta_1 = 0.2857143$
- rejection margin risk reduction: $\pi_0 = 0.3$, $\theta_0 = 0.4117647$

$$n = \frac{\left[z_\alpha \sqrt{\theta_0 \left(1 - \theta_0\right)} + z_\beta \sqrt{\theta(1-\theta)}\right]^2}{\left(p_T + p_C\right)\left(\theta - \theta_0\right)^2}$$

```
pC = 0.0075
u = 1
pi0 = 0.3
theta0 = (1 - pi0)/(1 - pi0 + u)
pi1 = 0.6
pT = (1 - pi1) * pC
theta1 = (1 - pi1)/(1 - pi1 + u)

alpha = 0.025
beta = 0.80

d = 0.02

sample.size.VE.low <- function(pT, pC, theta0, theta1, alpha, beta){
```

```
  z.alpha = qnorm(1-alpha)
  z.beta = qnorm(beta)
  nom = (z.alpha * sqrt(theta0 * (1 - theta0)) + z.beta * sqrt(theta1 * (1 - theta1)))^2
  denom = (pT + pC) * (theta1 - theta0)^2
  return(nom/denom)
}

sample.size.VE.low(pT,pC,theta0,theta1,alpha,beta)/(1-d)
```

```
## [1] 11061.56
```

2 SAE

```
long.dat <- readxl::read_excel("./Q2b.xlsx")
baseline.dat <- readxl::read_excel("./Q2b_BL.xlsx")

long.dat <- left_join(long.dat, baseline.dat, by = "ID") %>%
  mutate(OBS = if_else(is.na(SAE), 0, 1)) %>%
  mutate(TIME = factor(TIME),
         GROUP = factor(GROUP, labels = c("Control", "Vaccine"), levels = c(0,1)),
         SITE = factor(SITE),
         SEX = factor(SEX, labels = c("Female", "Male"), levels = c(0,1)),
         SAE = factor(SAE, labels = c("No", "Yes"), levels = c(0,1)))

glmm.fit <- glmer(SAE ~ TIME * GROUP + SEX + AGE + (1|ID) , data = long.dat, family = bi
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
glmm.fit.1 <- glmer(SAE ~ TIME + GROUP + SEX + AGE + (1|ID) , data = long.dat, family =
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
glmm.fit.null <- glmer(SAE ~ TIME + SEX + AGE + (1|ID) , data = long.dat, family = binom
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
anova(glmm.fit.1, glmm.fit.null)
```

```
## Data: long.dat
## Models:
## glmm.fit.null: SAE ~ TIME + SEX + AGE + (1 | ID)
## glmm.fit.1: SAE ~ TIME + GROUP + SEX + AGE + (1 | ID)
##                npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## glmm.fit.null    6 2616.6 2674.3 -1302.3   2604.6
## glmm.fit.1       7 2615.2 2682.5 -1300.6   2601.2 3.4022  1    0.06511 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(glmm.fit.1, glmm.fit)
```

```
## Data: long.dat
## Models:
## glmm.fit.1: SAE ~ TIME + GROUP + SEX + AGE + (1 | ID)
## glmm.fit: SAE ~ TIME * GROUP + SEX + AGE + (1 | ID)
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## glmm.fit.1    7 2615.2 2682.5 -1300.6   2601.2
## glmm.fit      9 2617.8 2704.4 -1299.9   2599.8 1.3303  2     0.5142
```

```r
anova(glmm.fit, glmm.fit.null)
```

```
## Data: long.dat
## Models:
## glmm.fit.null: SAE ~ TIME + SEX + AGE + (1 | ID)
## glmm.fit: SAE ~ TIME * GROUP + SEX + AGE + (1 | ID)
##               npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## glmm.fit.null    6 2616.6 2674.3 -1302.3   2604.6
## glmm.fit         9 2617.8 2704.4 -1299.9   2599.8 4.7326  3     0.1925
```

```r
summary(glmm.fit)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##  Family: binomial  ( logit )
## Formula: SAE ~ TIME * GROUP + SEX + AGE + (1 | ID)
##    Data: long.dat
##
##      AIC      BIC   logLik deviance df.resid
##   2617.8   2704.4  -1299.9   2599.8   111443
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -0.064 -0.042 -0.039 -0.036 34.522
##
## Random effects:
##  Groups Name        Variance  Std.Dev.
##  ID     (Intercept) 5.417e-12 2.327e-06
## Number of obs: 111452, groups:  ID, 41194
##
```

```
## Fixed effects:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -7.28684    0.37810 -19.273   <2e-16 ***
## TIME2               0.05952    0.28029   0.212   0.8318
## TIME3               0.19198    0.27311   0.703   0.4821
## GROUPVaccine        0.46506    0.25913   1.795   0.0727 .
## SEXMale            -0.09763    0.15144  -0.645   0.5192
## AGE                 0.01435    0.00680   2.110   0.0349 *
## TIME2:GROUPVaccine -0.15088    0.37007  -0.408   0.6835
## TIME3:GROUPVaccine -0.42629    0.37442  -1.139   0.2549
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TIME2  TIME3  GROUPV SEXMal AGE    TIME2:
## TIME2       -0.378
## TIME3       -0.387  0.523
## GROUPVaccin -0.397  0.551  0.566
## SEXMale     -0.183  0.000  0.000 -0.005
## AGE         -0.827  0.000 -0.001 -0.013 -0.009
## TIME2:GROUP  0.286 -0.757 -0.396 -0.700 -0.001  0.001
## TIME3:GROUP  0.283 -0.382 -0.729 -0.692  0.000  0.001  0.484
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
summary(glmm.fit.1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##  Family: binomial  ( logit )
```

```
## Formula: SAE ~ TIME + GROUP + SEX + AGE + (1 | ID)
##    Data: long.dat
##
##      AIC      BIC   logLik deviance df.resid
##   2615.2   2682.5  -1300.6   2601.2   111445
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -0.061 -0.042 -0.039 -0.036 32.734
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ID     (Intercept) 6.88e-11 8.294e-06
## Number of obs: 111452, groups:  ID, 41194
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.18053    0.35513 -20.220   <2e-16 ***
## TIME2        -0.02803    0.18283  -0.153   0.8782
## TIME3        -0.03556    0.18541  -0.192   0.8479
## GROUPVaccine  0.27957    0.15193   1.840   0.0657 .
## SEXMale      -0.09763    0.15144  -0.645   0.5192
## AGE           0.01435    0.00680   2.110   0.0348 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TIME2  TIME3  GROUPV SEXMal
## TIME2       -0.250
## TIME3       -0.249  0.477
```

```
## GROUPVaccin -0.215  0.007  0.015
## SEXMale     -0.195 -0.001  0.000 -0.009
## AGE         -0.880  0.000  0.000 -0.022 -0.009
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
summary(glmm.fit.null)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 0) [glmerMod]
##  Family: binomial  ( logit )
## Formula: SAE ~ TIME + SEX + AGE + (1 | ID)
##    Data: long.dat
##
##      AIC      BIC   logLik deviance df.resid
##   2616.6   2674.3  -1302.3   2604.6   111446
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.0577 -0.0417 -0.0394 -0.0372 30.5411
##
## Random effects:
##  Groups Name        Variance  Std.Dev.
##  ID     (Intercept) 3.435e-11 5.861e-06
## Number of obs: 111452, groups:  ID, 41194
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.050199   0.350291 -20.127   <2e-16 ***
## TIME2       -0.030380   0.182824  -0.166   0.8680
```

```
## TIME3       -0.040833   0.185381  -0.220   0.8257
## SEXMale     -0.095150   0.151432  -0.628   0.5298
## AGE          0.014629   0.006879   2.127   0.0334 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) TIME2  TIME3  SEXMal
## TIME2   -0.252
## TIME3   -0.249  0.477
## SEXMale -0.201  0.000  0.000
## AGE     -0.908  0.000  0.000 -0.007
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
glmm.tbl <-
  tbl_regression(glmm.fit, exponentiate = T,
                 label = list(
                   TIME ~ "Time points",
                   GROUP ~ "Treatment group",
                   SEX ~ "Gender",
                   AGE ~ "Age"
      )
      )
```

```
library(table1)
```

```
##
## Attaching package: 'table1'

## The following objects are masked from 'package:base':
```

```
##
##     units, units<-
```

```
long.dat %>%
  group_by(TIME, GROUP) %>%
  mutate(SAE_TOTAL = sum(SAE == "Yes", na.rm = T),
         MISSING_TOTAL = sum(OBS == 0)) %>%
  select(TIME, GROUP, SAE_TOTAL, MISSING_TOTAL) %>%
  unique() %>%
  ggplot(aes(x=TIME, y=SAE_TOTAL, fill=GROUP)) +
  geom_bar(stat="identity", position=position_dodge()) +
  #labs(title="SAE_TOTAL vs TIME by GROUPS") +
  ylab("Number of subjects who experienced SAE") +
  xlab("Time (Months)") +
  labs(fill = "Treatment Group") +
  theme_bw() +
  theme(legend.position = "bottom")
```

```
long.dat %>%
  group_by(TIME, GROUP) %>%
  mutate(SAE_TOTAL = sum(SAE == "Yes", na.rm = T),
         MISSING_TOTAL = sum(OBS == 0)) %>%
  select(TIME, GROUP, SAE_TOTAL, MISSING_TOTAL) %>%
  unique() %>%
  ggplot(aes(x=TIME, y=MISSING_TOTAL, fill=GROUP)) +
  geom_bar(stat="identity", position=position_dodge()) +
  #labs(title="SAE_TOTAL vs TIME by GROUPS") +
  ylab("Number of missing observations") +
  xlab("Time (Months)") +
  labs(fill = "Treatment Group") +
```

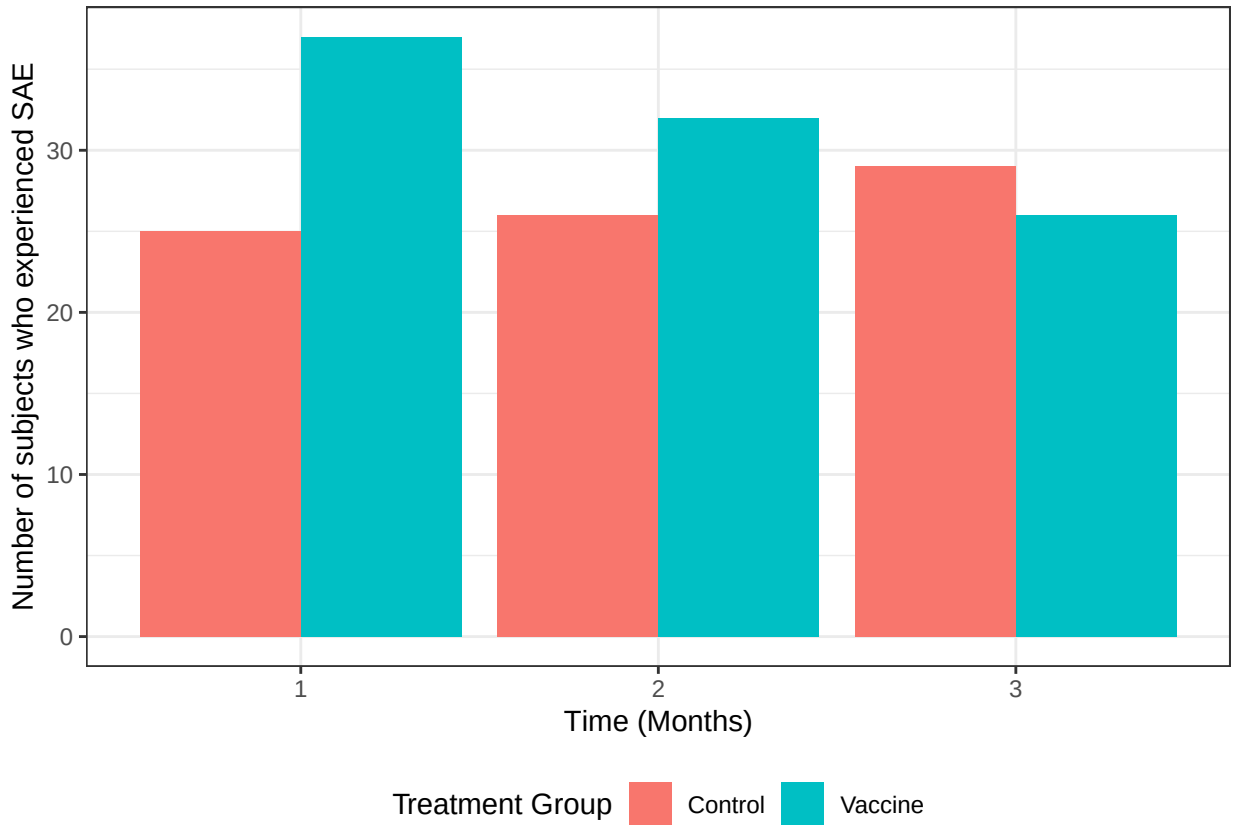


Figure 1: Number of subjects experienced serious adverse events (SAE) at each time point, stratified by treatment group.

```
  theme_bw() +
  theme(legend.position = "bottom")
```

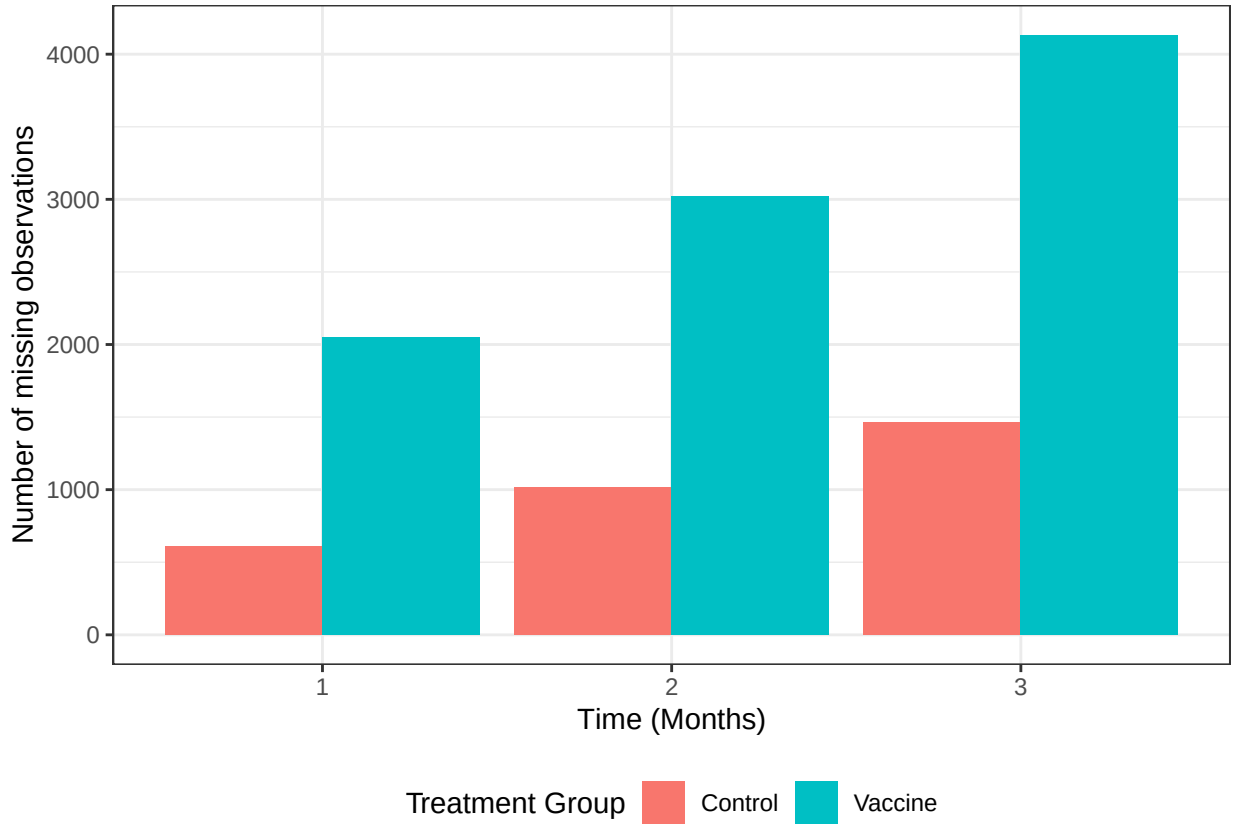


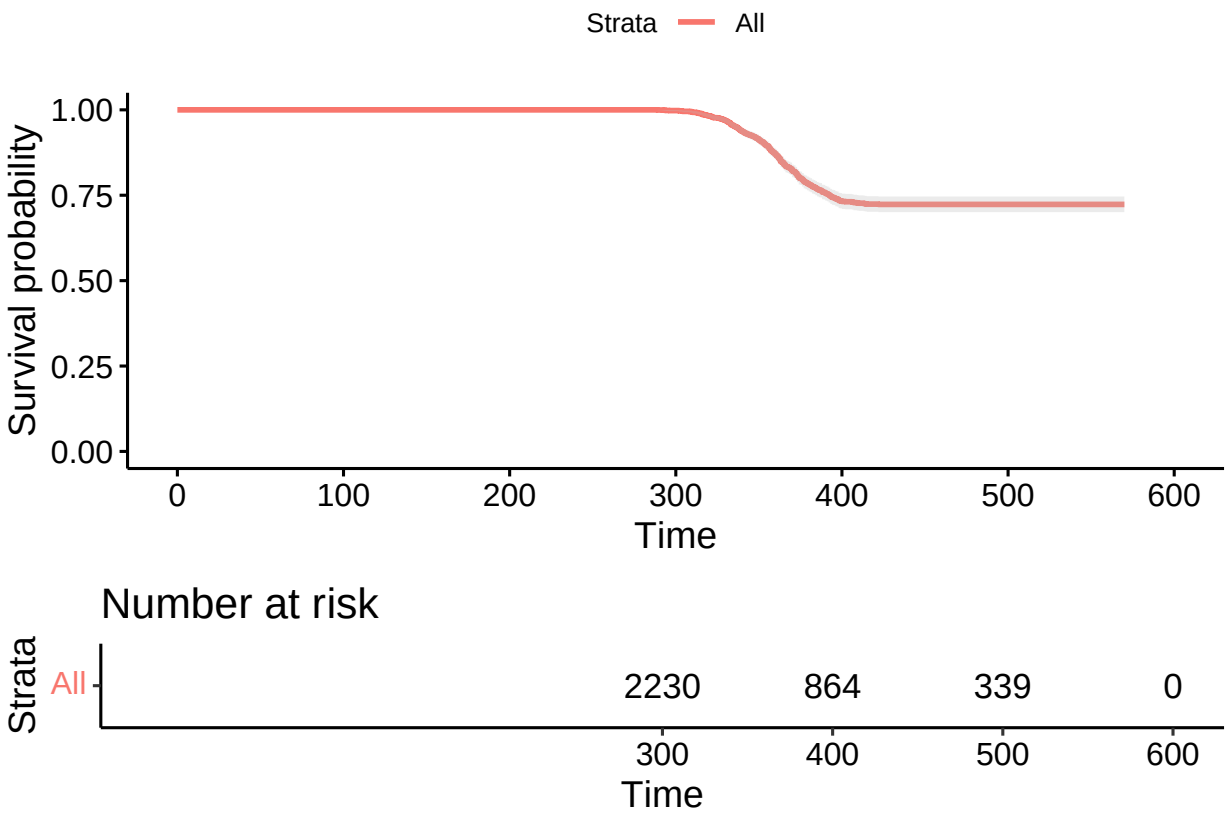
Figure 2: Number of missing observations at each time point, stratified by treatment group.

3 Survival

```
survival.dat <- readxl::read_excel("./Q2c.xlsx",col_types = rep("numeric", 5)) %>%
  mutate(InfectionTime = if_else(is.na(InfectionTime), LastFUTime, InfectionTime))

km.fit <- survfit(Surv(EnrollmentTime, InfectionTime, Infection) ~ 1, data = survival.da

ggsurvplot(km.fit, data = survival.dat, pval.method = TRUE, conf.int = TRUE, censor = F,
```

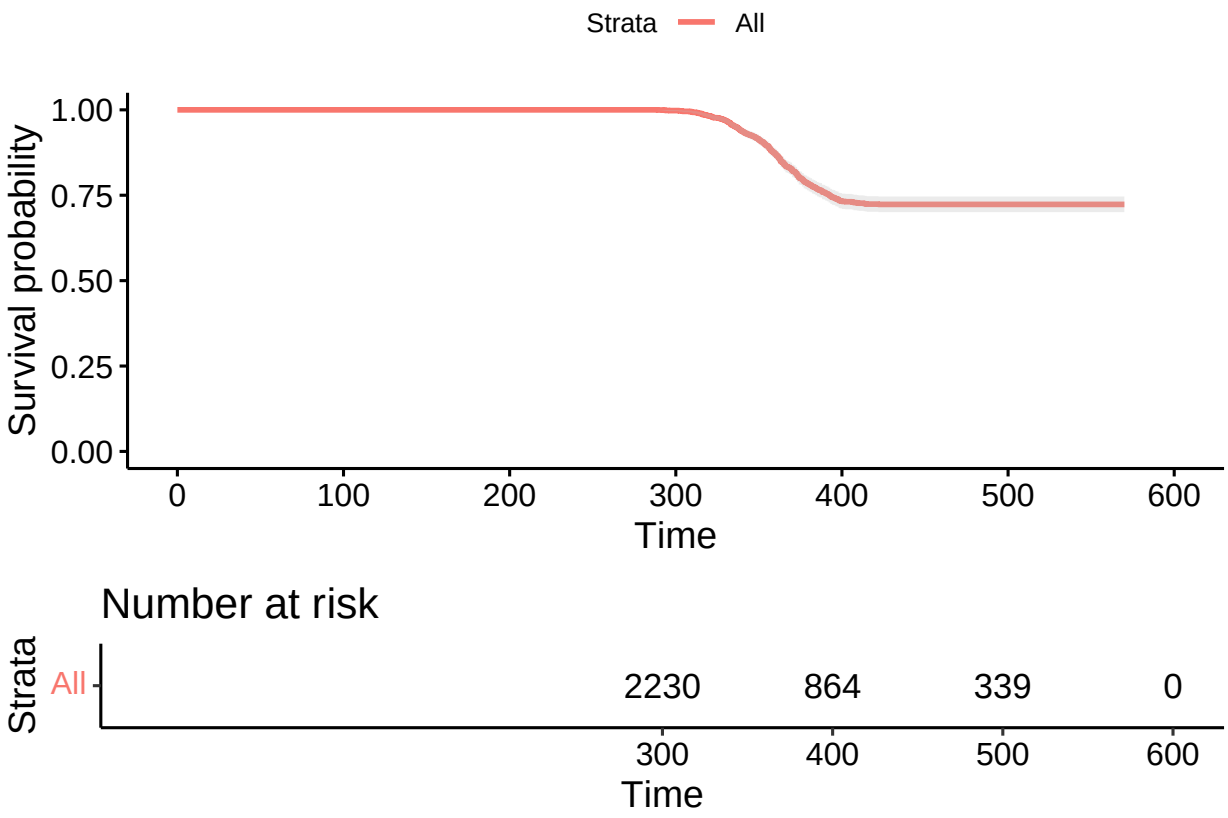



```
print(km.fit)
```

```
## Call: survfit(formula = Surv(EnrollmentTime, InfectionTime, Infection) ~
##     1, data = survival.dat)
##
##         n events median 0.95LCL 0.95UCL
## [1,] 2299    433     NA      NA      NA
```

```
km.fit2 <- survfit(Surv(InfectionTime, Infection) ~ 1, data = survival.dat)

ggsurvplot(km.fit2, data = survival.dat, pval.method = TRUE, conf.int = TRUE, censor = F
```




```
print(km.fit2)
```

```
## Call: survfit(formula = Surv(InfectionTime, Infection) ~ 1, data = survival.dat)
##
##         n events median 0.95LCL 0.95UCL
## [1,] 2299    433     NA      NA      NA
```