The emergence in December 2019 of a novel coronavirus, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has had devastating consequences globally. Control measures such as the use of masks, physical distancing, testing of exposed or symptomatic persons, contact tracing, and isolation have helped limit the transmission where they have been rigorously applied; however, these actions have been variably implemented and have proved insufficient in impeding the spread of coronavirus disease 2019 (Covid-19), the disease caused by SARS-CoV-2. Vaccines are urgently needed to reduce the morbidity and mortality associated with Covid-19. A pharmaceutical company therefore would like to conduct a phase III randomized (1-to-1 ratio), stratified, observer-blinded, placebo-controlled trial at 100 U.S. sites to demonstrate the efficacy for their developing vaccine.

The primary endpoint is the efficacy of the vaccine in preventing the first occurrence of symptomatic Covid-19 with onset at least 14 days after the second injection among participants who are seronegative at baseline. Vaccine efficacy is defined as (1 – “risk ratio”), where the “risk ratio” is the ratio of the “risk” of COVID-19 illness in the vaccine group to the corresponding risk in the placebo group. The “risk” can be measured by cumulative incidences, incidence rates, odds, or hazards whichever you think is appropriate.The trial will be designed to test the null hypothesis that the vaccine efficacy is 30% or less and provide 80% power to detect a 60% vaccine efficacy without planned interim analyses.

**Question**

1. As the study statistician, you are in charge of writing a comprehensive statistical analysis plan which includes the (1) randomization procedure, (2) analytic approach for primary analysis, and (3) sample size calculation. For a(1) and a(2), please make sure that your answer covers all the details that allow other statisticians to implement your plan and deal with potential issues that may arise when conducting the randomization and analysis. For a(3), if the information you may need for sample size calculation is not available (e.g., loss to follow-up rate), you can make your own assumptions and justify them. As for all statistical questions, there usually exists more than one appropriate analytic approach. So please make sure the method you use to conduct the sample size calculation is clearly specified (i.e., can be reproduced by other statisticians) and consistent with the analytic approach you propose in the statistical analysis plan.
2. After the company had your statistical analysis plan in place, the trial began, and data were collected. In addition to the primary efficacy outcome, the company also monitored the serious adverse events (SAE) closely by contacting study participants on a monthly basis for 3 months. SAE was defined as any untoward medical occurrence that was life-threatening or required inpatient hospitalization. Please use the provided data (Q2b\_BL.xlsx and Q2b.xlsx) to determine whether the vaccine group has greater odds of having SAE at any of the three assessment time points.

Listed below is the variable labels and value labels for the data sets.

Q2b\_BL.xlsx:

|  |  |  |
| --- | --- | --- |
| Variable Name | Variable Label | Value Label |
| ID | Study ID | NA |
| SITE | Site number | NA |
| SEX | Sex of the study participant | 0: Female; 1:Male |
| AGE (in years) | Age of the study participant | NA |

Q2b.xlsx:

|  |  |  |
| --- | --- | --- |
| Variable Name | Variable Label | Value Label |
| ID | Study ID | NA |
| TIME | Follow up time point | 1: 1-month follow-up  2: 2-month follow-up  3: 3-month follow-up |
| SAE | If any serious adverse event within a month prior to the follow-up time | 0: No; 1:Yes |
| GROUP | Randomization assignment | 0: Control group  1: Vaccine group |

1. After the trial was completed and the vaccine was proven to be efficacious, people are then curious to know what is the probability of contracting Covid-19 within 12 months, the median time to infection, and the mean time to infection after their second shot. To answer the above questions, the company recruited 2299 study participants who did get two shots of the vaccine in the original trial and did not get infected by COVID prior to the recruitment, followed them for up to 12 months, and collected the data as shown in Q2c.xlsx.
2. Please use the provided data set (i.e., Q2c.xlsx) to obtain the probability of getting infected within 12 months as well as the corresponding 95% confidence interval.
3. If someone asks you: “Based on the above data, what is the median time to infection, and the mean of the time to infection after the second shot?” How will you respond?

Listed below are the variable labels and value labels for the dataset Q2c.xlsx.

|  |  |  |
| --- | --- | --- |
| Variable Name | Variable Label | Value Label |
| ID | Study ID | NA |
| EnrollmentTime  (in days) | Number of days between the enrollment time and the time of the second shot | NA |
| LastFUTime | Number of days between the last FU time and the time of the second shot | NA |
| InfectionTime | Number of days between the infection time and the time of the second shot | NA |
| Infection | If infected | 0: No; 1: Yes |