COVE: An R Package for Evaluating the COVID-19 Vaccine Efficacy

Dan-Yu Lin, Donglin Zeng, and Shannon T. Holloway

May 17, 2021

1 Introduction

COVE is an R package that implements a simple and rigorous framework for evaluating the efficacy of vaccines based on the dual or triple primary endpoints of infection, symptomatic disease, and severe disease (Lin et al., 2021). It inputs a rectangular data set with the following information:

- **Vaccination Status**: Binary indicator taking the value 1 if the participant is vaccinated and 0 otherwise.
- **Follow-up time**: Length of follow-up for the participant.
- **Infection status**: Binary indicator taking the value 1 if the participant is infected during the follow-up and 0 otherwise.
- **Symptomatic disease status**: Binary indicator taking the value 1 if the participant experiences symptomatic disease during the follow-up and 0 otherwise.
- **Severe disease status**: Binary indicator taking the value 1 if the participant develops severe disease during the follow-up and 0 otherwise.

The primary analysis tool of the package is `cove()`, which returns the Z-scores and estimated vaccine efficacy for single endpoints, the test statistics by combining the Z-scores or the score statistics for the dual and triple endpoints, as well as the critical values adjusted for multiple testing. A print function is also provided to access a verbose description of these results.

2 Methods

We consider the endpoints of SARS-CoV-2 infection, symptomatic COVID-19, and severe COVID-19, referring to them as infection, symptomatic disease, and severe disease, respectively. A large number of individuals are randomly assigned to vaccine or placebo, and the trial records whether or not each participant has developed each of the three endpoints by the end of follow-up, as well as their length of follow-up.

We formulate the effect of the vaccine on each of the three endpoints through a Poisson model. We define the vaccine efficacy in terms of the proportionate reduction in the event rate between vaccinated and un-vaccinated individuals. We estimate the vaccine efficacy by the maximum likelihood method and provide the corresponding 95% confidence interval.
For each of the three endpoints, we calculate the score statistic for testing the null hypothesis that the vaccine efficacy is less than some threshold VEnull against the alternative hypothesis that the vaccine efficacy is greater than VEnull. We divide the score statistic by its standard error to create a standard-normal test statistic, i.e., Z-score.
We propose to test the null hypotheses on two or three endpoints, adjusting the critical value for the test statistics, so as to control the overall type I error at the desired level α. Typically, α is set to 0.05, and one-sided tests based on $\alpha/2$ are performed. If the Z-score for any endpoint is less than the adjusted critical value, then we consider the vaccine to be successful. We refer to this method as multiple testing.
If the effects of a vaccine are expected to be similar among the two or three endpoints, then we can enhance statistical power by combining the evidence of the vaccine effects on those endpoints and performing a single test of overall vaccine efficacy. Specifically, we take the sum of the score statistics or the Z-scores (i.e., standardized score statistics) and divide the sum by its standard error to create a standard-normal test statistic, i.e., Z-score. We consider the vaccine to be successful if the Z-score is less than the lower $100\alpha/2$ percentile of the standard normal distribution.

3 Functions

3.1 `cove()`

This function is the primary tool of **COVE**. The function call takes the following form:

```
cove(vaccinated, time, infection, symptomatic, severe, nMC = 20000L,
      alpha = 0.05, VEnull = c(0.3,0.3,0.3))
```

where `vaccinated`, `time`, `infection`, `symptomatic`, and `severe` are as described above. Input parameters `nMC` and `alpha` are the number of Monte Carlo samples (default 20,000) and the two-sided family-wise type I error (default 0.05), respectively, to be used in determining the adjusted critical value. Input `VEnull` is a vector of the three null values for the vaccine efficacy on infection, symptomatic disease, and severe disease (in this order); the default is (0.3,0.3,0.3).

3.2 `print()`

This is a convenience function provided to improve the readability of the results returned by `cove()`.

4 Example

To illustrate the call structure and results of `cove()`, we use the dataset provided with the package, coveData. This dataset contains the following observations for each of the 27,000 participants:

- **vaccinated** The indicator of vaccination (1 = vaccinated; 0 = not vaccinated)
- **time** The follow-up time in days
- **infection** The indicator of infection (1 = infected; 0 = not infected)
- **symptomatic** The indicator of symptomatic disease (1 = symptomatic; 0 = not symptomatic)
- **severe** The indicator of severe disease(1 = severe; 0 = not severe)

The data can be loaded in the usual way

```
data(coveData)
```

```
head(coveData)
```

```
  vaccinated time infection symptomatic severe
1          0  341         1           1      1
2          1  349         1           0      0
3          0  347         1           1      0
4          0  337         0           0      0
5          1  337         0           0      0
6          0  329         0           0      0
```

Consider the summary statistics

```
summary(coveData)
```

```
  vaccinated       time       infection        symptomatic
Min.   :0.0   Min.   :300   Min.   :0.00000   Min.   :0.000000
1st Qu.:0.0   1st Qu.:315   1st Qu.:0.00000   1st Qu.:0.000000
Median :0.5   Median :330   Median :0.00000   Median :0.000000
Mean   :0.5   Mean   :330   Mean   :0.01359   Mean   :0.008407
3rd Qu.:1.0   3rd Qu.:345   3rd Qu.:0.00000   3rd Qu.:0.000000
Max.   :1.0   Max.   :360   Max.   :1.00000   Max.   :1.000000
    severe
Min.   :0.00000
1st Qu.:0.00000
Median :0.00000
Mean   :0.00137
3rd Qu.:0.00000
Max.   :1.00000
```

We see that 50% of the participants were vaccinated during the follow-up period, which ranges from 300 to 360 days. Approximately 1.35% of the participants experienced SARS-CoV-2 infection; 0.84% experienced symptomatic COVID-19; and 0.14% experienced severe COVID-19.

The analysis, assuming default values for `nMC`, `alpha`, and `VEnull`, is completed as follows

```
result <- cove(vaccinated = coveData$vaccinated,
               time = coveData$time,
               infection = coveData$infection,
               symptomatic = coveData$symptomatic,
               severe = coveData$severe)
```

The function returns a list object containing the following items.

Single Endpoints: Element **$singleEP** contains a matrix. The first column contains the the Z-scores for the three endpoints: infection (INF), symptomatic disease (SYMP), and severe disease (SVR); the subsequent columns provide the estimated vaccine efficacy for each endpoint, its standard error, and the 95% confidence interval.

```
result$singleEP
       ZScore        VE         se lower .95 upper .95
INF  -3.205959 0.5021912 0.05482289 0.3822600 0.5988384
SYMP -2.870171 0.5261172 0.06708859 0.3745688 0.6409439
SVR  -1.848661 0.6297415 0.13698983 0.2353864 0.8207051
```

Multiple Endpoints: Element **$multipleEP** contains a matrix. The first two columns are standard-normal test statistics (i.e., Z-scores) obtained by combining the individual Z-scores (ComZScore) and by combining the individual score statistics (ComScoreStats). The third (AdjCriticalValue) provides the adjusted critical values for multiple testing.

```
result$multipleEP
            ComZScores ComScoreStats AdjCriticalValue
INF-SYMP      -3.219400     -3.237862        -2.163443
SYMP-SVR      -2.826059     -2.972032        -2.223006
INF-SYMP-SVR  -3.244596     -3.310001        -2.349799
```

A verbose summary of the results of the analysis can be obtained using `print()`

```
print(result)
Single Endpoints

  Infection: Z-score = -3.205959
  Symptomatic: Z-score = -2.870171
  Severe: Z-score = -1.848661

  Infection: estimated vaccine efficacy = 50.22%, 95% CI = (38.23%, 59.88%)
  Symptomatic: estimated vaccine efficacy = 52.61%, 95% CI = (37.46%, 64.09%)
  Severe: estimated vaccine efficacy = 62.97%, 95% CI = (23.54%, 82.07%)
```

```
Multiple Endpoints

Combining Z-scores
  Infection-Symptomatic: Z-score = -3.2194
  Symptomatic-Severe: Z-score = -2.826059
  Infection-Symptomatic-Severe: Z-score = -3.244596

Combining score statistics
  Infection-Symptomatic: Z-score = -3.237862
  Symptomatic-Severe: Z-score = -2.972032
  Infection-Symptomatic-Severe: Z-score = -3.310001

Multiple testing
  Infection-Symptomatic: adjusted critical value = -2.163443
  Symptomatic-Severe: adjusted critical value = -2.223006
  Infection-Symptomatic-Severe: adjusted critical value = -2.349799
```

References

Lin, DY, Zeng, D, Mehrotra, DV, Corey, L, and Gilbert, PB (2021). Evaluating the efficacy of COVID-19 vaccines. Clinical Infectious Diseases, ciaa1863, https://doi.org/10.1093/cid/ciaa1863