DOVE: An R Package for Evaluating the Durability Of Vaccine Efficacy

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

March 1, 2021

1 Introduction

DOVE is an R package for evaluating the durability of vaccine efficacy in a randomized, placebo-controlled clinical trial with staggered enrollment of participants and potential crossover of placebo recipients (Lin et al., 2021). It inputs a rectangular data set with the following information:

- **Entry time**: Calendar time when the participant enters the trial.
- **Event time**: Calendar time when the participant experiences the clinical event of interest (e.g., symptomatic COVID-19) or their follow-up ends, whichever occurs first.
- **Event status**: Binary indicator taking the value 1 if the clinical event of interest occurs before the end of follow-up and 0 otherwise.
- **Vaccination status**: Binary indicator taking the value 1 if vaccination occurs before the end of follow-up and 0 otherwise.
- **Vaccination time**: Calendar time when vaccination takes place, with an arbitrary value if the participant is not vaccinated.
- **Covariates**: Baseline covariates (e.g., priority group, age, sex, ethnicity).

Note that an arbitrary number of baseline covariates can be included and that all of the time variables are measured from the start of the trial.

The primary analysis tool of the package is `dove()`, the formal argument structure of which was chosen to resemble that of the `coxph()` function of the **survival** package. Function `dove()` returns the estimated hazard ratio for each baseline covariate, the estimated vaccine efficacy in reducing the attack rate (cumulative incidence), the estimated vaccine efficacy in reducing the hazard ratio (instantaneous risk), and the estimated vaccine efficacy in reducing the attack rate over successive time periods.

In addition, the package includes a convenience function `vaccine()`, which is used to simplify the specification of input variables required in the model statement of `dove()`, similar in spirit to the `cluster()` function of the **survival** package. Finally, a simulated dataset is provided to illustrate the use of the software.

2 Functions

2.1 `vaccine()`

This convenience function is used as a component of the right-hand-side of a formula object for the sole purpose of simplifying the specification of required input variables: entry time, vaccination status and vaccination time. This function is not intended to be used as a stand-alone feature; though for completeness, the function ensures that the input data obey basic constraints and returns the data in a predictable format for use in internal functions.

The usage is

```
vaccine(entry_time, vaccination_status, vaccination_time)
```

where `entry_time` is the time when the participant enters the trial; `vaccination_status` is the binary indicator of vaccination, and `vaccination_time` is the time when vaccination takes place.

2.2 `dove()`

This function is the primary tool of **DOVE**. The value object returned contains the estimated hazard ratio for each baseline covariate, estimated vaccine efficacy in reducing the attack rate, $VE_a(t)$, and in reducing the hazard rate, $VE_h(t)$, where t is time elapsed since vaccination, as well as the estimated vaccine efficacy in reducing the attack rate over m successive time periods, $VE_a(0,t_1), VE_a(t_1,t_2), \ldots, VE_a(t_{m-1},t_m)$. By definition, $VE_a(0,t) = VE_a(t)$.

The function call takes the following form:

```
dove(formula, data, plots = TRUE, timePts = NULL, bandwidth = NULL)
```

where `formula` is a model statement, `data` is the data.frame object containing all required data as previously described, `plots` is a logical object indicating whether graphical forms of the $VE_a(t)$ and $VE_h(t)$ results are to be generated, `timePts` is an optional vector to specify the time points $(t_1, t_2, \ldots, t_m)$ for partitioning the study period, and `bandwidth` is a tuning parameter for the bandwidth used in the kernel estimation of $VE_h(t)$. To obtain reliable estimates of VE_a (t_{j-1}, t_j) $(j = 1, \ldots, m)$, we suggest using broad time periods, such as every month, every two months, or every quarter. If `timePts` is not provided, then the default time periods are every 60 days. We suggest choosing `bandwidth` between 0.1 and 1.0: a smaller bandwidth yields a less biased estimate of $VE_h(t)$, whereas a larger bandwidth yields a smoother estimate of the VE_h curve. The default value of `bandwidth` is 0.5. This input is ignored if `plots` is FALSE.

The model statement is a formula object. The left-hand-side is a survival object as returned by the `Surv()` function of the **survival** package and specifies the event time and event status. The right-hand-side is a combination of baseline covariates and the previously described `vaccine()` function. The `formula` input takes the following general structure

```
Surv(event_time, event_status) ~ covariates +
    vaccine(entry_time, vaccination_status, vaccination_time)
```

where 'event_time', 'event_status', 'covariates', 'entry_time' 'vaccination_status' and 'vaccination_time' are used here as place holders indicating the data that are to be provided; they are to be replaced by the variable names in the header of the input data.

Of note, the two measures of vaccine efficacy, $VE_a(t)$ and $VE_h(t)$, are estimated up to the last observed event time. However, the estimates near the end of crossover where there are very few placebo participants under follow-up may not be reliable. For estimating VE_a over successive time periods, the last time period should not extend beyond the point that there are still a few placebo participants under follow-up.

3 Example

To illustrate the call structure and results of `dove()`, we use the dataset provided with the package, doveData. This dataset was simulated under a priority-tier dependent crossover design and contains the following observations for each of the 40,000 participants:

- **entry.time**: The entry time in days
- **event.time**: The event time in days
- **event.status**: The event indicator (1=event; 0=censored)
- **vaccine.time**: The time of vaccination in days; NA if not vaccinated
- **vaccine.status**: The indicator of vaccination (1=vaccinated; 0 = not vaccinated)
- **priority**: A composite baseline risk score taking values 1-5
- **sex**: A binary indicator of sex (male/female)

The data can be loaded in the usual way

```
data(doveData)

head(doveData)
  entry.time event.time event.status vaccine.time vaccine.status
1          5        320            0            5              1
2          1        320            0            1              1
3         19        320            0           19              1
4        102        320            0          102              1
5         21        320            0           21              1
6        113        320            0          113              1
  priority sex
1        4   0
2        2   1
3        4   1
4        5   0
5        4   1
6        2   1
```

Consider the summary statistics

```
summary(doveData)
  entry.time       event.time     event.status      vaccine.time
Min.   :  0.00   Min.   :  1.0   Min.   :0.00000   Min.   :  0.0
1st Qu.: 30.00   1st Qu.:320.0   1st Qu.:0.00000   1st Qu.: 52.0
Median : 61.00   Median :320.0   Median :0.00000   Median :103.0
Mean   : 60.98   Mean   :315.2   Mean   :0.03318   Mean   :139.3
3rd Qu.: 92.00   3rd Qu.:320.0   3rd Qu.:0.00000   3rd Qu.:239.0
Max.   :122.00   Max.   :320.0   Max.   :1.00000   Max.   :320.0
                                                   NA's   :6192
vaccine.status      priority          sex
Min.   :0.0000   Min.   :1.000   Min.   :0.0000
1st Qu.:1.0000   1st Qu.:2.000   1st Qu.:0.0000
Median :1.0000   Median :3.000   Median :0.0000
Mean   :0.8452   Mean   :2.994   Mean   :0.4985
3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:1.0000
Max.   :1.0000   Max.   :5.000   Max.   :1.0000
```

We see that participants were enrolled in the study over a 4-month period ($0 \leq$ entry.time ≤ 122 days); that the follow-up time ended on day 320 (event.time ≤ 320 days) with few events ($\sim 3.3\%$ of the participants); and that $\sim 85\%$ of the participants were vaccinated over the course of the study period (6192 'missing' data entries in vaccine.time). In addition, the priority (risk) score is evenly distributed across participants, who are equally distributed between the two sex groups.

In this analysis, we will include in our model statement baseline covariates, priority and sex. In addition, we will use the default partitioning of the study period and the default tuning parameter for the bandwidth. The function call takes the following form

```
result <- dove(formula = Surv(event.time, event.status) ~ priority + sex +
                  vaccine(entry.time, vaccine.status, vaccine.time),
                data = doveData)
```

The function returns a list object containing the following items. For brevity, we show only a snapshot of the large tabular results.

Covariate Effects: The estimated hazard ratio for each covariate, together with the (estimated) standard error, the 95% confidence interval, and the two-sided p-value for testing no covariate effect.

```
result$covariates
              coef   se(coef)        z    Pr(>|z|) exp(coef)
priority 0.1857095 0.01989881 9.332693 0.00000e+00  1.204072
sex      0.2799022 0.05534819 5.057116 4.25644e-07  1.323000
         lower .95 upper .95
priority  1.158016  1.251961
sex       1.186989  1.474597
```

Vaccine Efficacy: Element **$efficacy** contains the estimated vaccine efficacy in reducing the attack rate at each observed event time, together with its standard error and the 95% confidence interval. In addition, the raw estimate of the hazard ratio at each observed event time is provided.

```
head(result$vaccine$efficacy)
```

```
     time      VE_a         se lower .95 upper .95 hazardRatio
[1,]    0 1.0000000 0.00000000 1.0000000 1.0000000   0.0000000
[2,]    3 0.9613860 0.03859012 0.9945543 0.7261984   0.1158419
[3,]   10 0.9767919 0.01642067 0.9942008 0.9071225   0.1162388
[4,]   23 0.9794174 0.01030862 0.9922878 0.9450680   0.2413194
[5,]   24 0.9752265 0.01110811 0.9897124 0.9403429   0.1211639
[6,]   29 0.9753030 0.01011848 0.9889364 0.9448695   0.1216495
```

```
tail(result$vaccine$efficacy)
```

```
       time      VE_a         se lower .95 upper .95 hazardRatio
[137,]  302 0.7917180 0.02697112 0.8384062 0.7315406    1.861729
[138,]  305 0.7865112 0.02835943 0.8354487 0.7230195    2.212932
[139,]  306 0.7794720 0.03001049 0.8311014 0.7120604    2.367466
[140,]  307 0.7553499 0.03546448 0.8158579 0.6749592    7.626032
[141,]  309 0.7474560 0.03743709 0.8111344 0.6623075    2.928507
[142,]  318 0.7101707 0.06037742 0.8073291 0.5640182   14.129609
```

Element **$period_efficacy** contains the estimated vaccine efficacy in reducing the attack rate over each time period, its standard error, and the 95% confidence interval.

```
result$vaccine$period_efficacy
```

```
     left right      VE_a          se lower .95 upper .95
[1,]    0    60 0.9675288 0.008207847 0.9467078 0.9802151
[2,]   60   120 0.9487351 0.010818674 0.9224724 0.9661012
[3,]  120   180 0.9054394 0.016414666 0.8671162 0.9327104
[4,]  180   240 0.7113107 0.040308193 0.6204388 0.7804266
[5,]  240   300 0.4791499 0.104800674 0.2273428 0.6488937
```

The graphical depictions of estimates returned in **vaccine$efficacy** are generated by default by `dove()` and are shown in Figure 1. For the plots, we assume that the times are expressed in days in the data.

References

Lin DY, Zeng D, Gilbert PB (2021). Evaluating the long-term efficacy of COVID-19 vaccines. doi: https://doi.org/10.1101/2021.01.13.21249779.

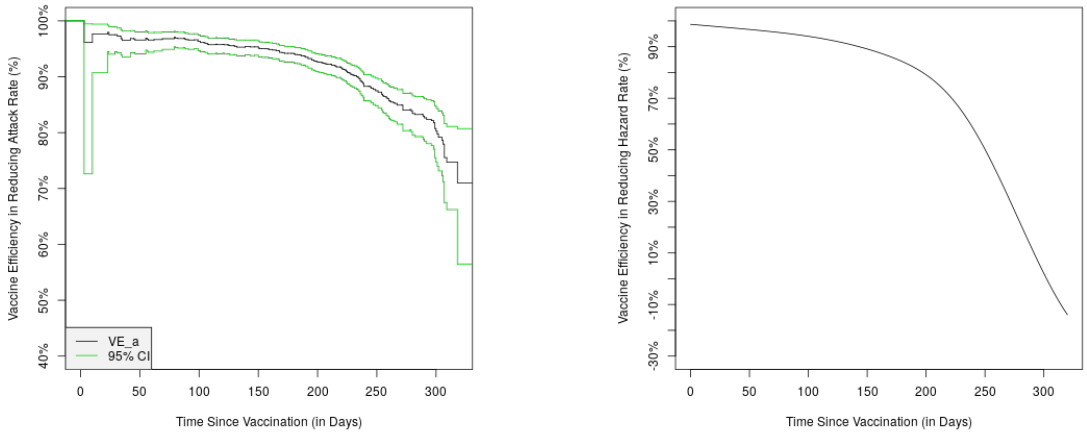

Figure 1: Plots auto-generated by `dove()`. On the left, the estimated curve of vaccine efficacy in reducing the attack rate, $VE_a(t)$ (black) and its 95% confidence intervals (green) are shown as a function of the time since vaccination. On the right, the estimated curve of vaccine efficacy in reducing the hazard ratio, $VE_h(t)$, is shown as a function of the time since vaccination.