DOVE – <u>D</u>urability <u>O</u>f <u>V</u>accine <u>E</u>fficacy

August 31, 2021

Introduction

DOVE is an R package for evaluating the durability of vaccine efficacy (VE) in a randomized, placebo-controlled clinical trial with staggered enrollment of participants and potential crossover of placebo recipients (Lin et al., 2021a; 2021b). It inputs a rectangular dataset 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 non-negative value if the participant is not vaccinated.

- **Covariates**: Baseline covariates (e.g., priority group, age, sex, ethnicity).

Of note, an arbitrary number of baseline covariates can be included, and all time variables are measured from the start of the trial and are specified in units of days.

The two primary analysis tools of the package are *dove()* and *dove2()*, the formal argument structures of which are similar and were chosen to resemble that of the *coxph()* function of the **survival** package. The underlying methodologies for *dove()* and *dove2()* are detailed in Lin et al. (2021a) and Lin et al. (2021b), respectively. Function *dove()* allows the vaccine effect to be an arbitrary function of time, whereas function *dove2()* assumes that the log hazard ratio for the vaccine effect is a piecewise linear function of time. Both functions return the estimated hazard ratio for each baseline covariate, the estimated VE in reducing the attack rate (cumulative incidence), the estimated VE in reducing the hazard rate (instantaneous risk), and the estimated VE in reducing the attack rate over successive time periods.

We recommend *dove()* for exploratory analyses and *dove2()* for formal analyses. The latter yields more stable estimates, together with proper confidence intervals, for VE in reducing the hazard rate. Both functions handle potentially right-censored events (e.g., symptomatic COVID-19, severe COVID-19, death). For interval-censored infection endpoints, iDOVE should be used instead.

The DOVE package includes convenience functions *print()*, *plot()*, and *vaccine()*. Function *vaccine()* is used to simplify the specification of input variables required in the model statements of *dove()* and *dove2()*, similar in spirit to the *cluster()* function of the **survival** package. A simulated dataset is provided to illustrate the use of the software.

Functions

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; although 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.

dove()

This function estimates VE as a nonparametric function of time. The value object returned contains the estimated hazard ratio for each baseline covariate, estimated VE in reducing the attack rate, $VE_a(t)$, and in reducing the hazard rate, $VE_h(t)$, where t is the time elapsed since vaccination, as well as the estimated VE 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. See below for further details.

- `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 object specifying the time points $(t_1, t_2, \ldots, t_m)$ for partitioning the study period.

- `bandwidth` is an optional numeric object specifying the 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, the default time periods are every 60 days. (If $\tau < 60$ days, timePts must be provided.) 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 0.3. This input is ignored if `plots` is FALSE.

The model statement is a formula object. The left-hand side is a survival analysis 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. Categorical baseline covariates can be specified, and if provided, all other categories are compared to the first category. 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 place holders indicating the data that are to be provided; they will be replaced by the variable names in the header of the input data.

The two VE measures, $VE_a(t)$ 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. We also constrain VE_h to be 0 at day 0 and non-decreasing at the right tail. 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.

dove2()

This function estimates VE under the assumption that the log hazard ratio for the vaccine effect is a piecewise linear function of time. The value object returned is similar to that returned by *dove()* and contains the estimated hazard ratio for each baseline covariate, estimated VE in reducing the attack rate, $VE_a(t)$, and in reducing the hazard rate, $VE_h(t)$, where t is the time elapsed since vaccination, as well as the estimated VE in reducing the attack rates over m successive time periods, $VE_a(0, t_1), VE_a(t_1, t_2), \ldots, VE_a(t_{m-1}, t_m)$. The 95% confidence intervals for all three measures of VE are provided.

The function call takes the following form:
```
dove2(formula, data, plots = TRUE, changePts = NULL,
        constantVE = FALSE, timePts = NULL)
```

where `formula`, `data`, `plots`, and `timePts` are as described above for *dove()*. Input `changePts` is an optional numerical vector to specify the change points, in units of days, of the piece-wise log-linear hazard ratio for the vaccine effect. If no change points are provided, one change point will automatically be selected among Weeks 4, 5, 6, 7, 8 by the Akaike information criterion (AIC). Input `constantVE` is a logical object indicating the VE trend after the last change point. If specified as TRUE, VE is assumed to be constant in the period after the last change point; otherwise it is allowed to vary after the last change point. If `timePts` is not specified, the default sequence of the multiples of the first change point will be used.

The model statement is as previously described for *dove()*. Specifically, the `formula` input takes the following general structure
```
Surv(event_time, event_status) ~ covariates +
        vaccine(entry_time, vaccination_statue, vaccination_time)
```

To ensure stability, we suggest placing change points at the time points where the events are relatively frequent and not placing change points near the right tail. We estimate $VE_a(t)$ and $VE_h(t)$ up to the maximum of all observed event times.

plot()

When provided a value object of class DOVE (the object returned by *dove()* and *dove2()*), this convenience function creates/recreates plots of the estimated VE in reducing the attack rate, $VE_a(t)$, and in reducing the hazard rate, $VE_h(t)$.

print()

When provided a value object of class DOVE (the object returned by *dove()* and *dove2()*), the tabular results are displayed.

Examples

We use the dataset provided with the package, doveData, to illustrate the analyses. This dataset was simulated under a priority-tier dependent crossover design with a ramping vaccine effect between dose 1 and dose 2 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=observed; 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 priority sex
## 1         69        212            1           NA              0        4   1
## 2         92        320            0          250              1        3   0
## 3         21        137            1           NA              0        3   0
## 4         31        320            0          186              1        5   1
## 5         32        320            0          251              1        3   0
## 6         72        320            0          188              1        5   1
```

The summary statistics are shown below

```
summary(doveData)
```

```
##    entry.time        event.time      event.status       vaccine.time
##  Min.   :  0.00   Min.   :  9.0   Min.   :0.00000   Min.   :  0.0
##  1st Qu.: 31.00   1st Qu.:320.0   1st Qu.:0.00000   1st Qu.: 57.0
##  Median : 60.00   Median :320.0   Median :0.00000   Median :112.0
##  Mean   : 60.25   Mean   :310.8   Mean   :0.06657   Mean   :148.1
##  3rd Qu.: 90.00   3rd Qu.:320.0   3rd Qu.:0.00000   3rd Qu.:247.0
##  Max.   :120.00   Max.   :320.0   Max.   :1.00000   Max.   :319.0
##                                                     NA's   :3234
##  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 :1.0000
##  Mean   :0.9192   Mean   :3.005   Mean   :0.5035
##  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 ≤ 120 days), the follow-up ended on day 320 (event.time ≤ 320 days) with a $\sim 6.7\%$ event rate, and $\sim 92\%$ of the participants were vaccinated over the course of the study period. In addition, the priority (risk) score is evenly distributed across participants, who are equally distributed between the two sex groups.

dove()

First, we illustrate a *dove()* analysis. Here, 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

```r
result1 <- dove(formula = Surv(event.time, event.status) ~ priority + sex +
                        vaccine(entry.time, vaccine.status, vaccine.time),
                data = doveData)
```

```
## tau = 320
```

```
## timePts not given; default values will be used
```

```
## method converged after 13 iterations
```

The function returns an object of class DOVE containing a list with the following elements. For brevity, we show only a snapshot of the large tabular results.

call: The unevaluated call.

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.

```r
result1$covariates
```

```
##                coef    se(coef)         z Pr(>|z|) exp(coef) lower .95 upper .95
## priority 0.2139999 0.01444336 14.816485        0  1.238622  1.204050  1.274188
## sex      0.3645546 0.03940875  9.250601        0  1.439873  1.332842  1.555498
```

Vaccine Efficacy: Element **$efficacy** contains the estimated VE 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.

```r
head(result1$vaccine$efficacy)
```

```
##      time      VE_a         se   lower .95 upper .95 hazardRatio
## [1,]    0 0.0000000 0.0000000  0.00000000 0.0000000   1.0000000
## [2,]    1 0.3529937 0.2056809 -0.20644423 0.6530157   0.6470063
## [3,]    2 0.2547967 0.1569215 -0.12595279 0.5067929   0.8434002
## [4,]    3 0.1559060 0.1377541 -0.16227101 0.3869806   1.0418753
## [5,]    4 0.2198001 0.1153130 -0.04235380 0.4160218   0.5885176
## [6,]    5 0.2577284 0.1011185  0.03055578 0.4316670   0.5905584
```

```r
tail(result1$vaccine$efficacy)
```

```
##          time      VE_a         se lower .95 upper .95 hazardRatio
## [284,]   298 0.5614314 0.02587078 0.5076771 0.6093166   0.9504455
## [285,]   299 0.5595833 0.02620395 0.5051090 0.6080614   0.9911702
## [286,]   301 0.5588241 0.02650251 0.5036975 0.6078275   1.1093593
```

```
## [287,]  304 0.5588767 0.02682445 0.5030393 0.6084403     1.3075374
## [288,]  306 0.5569388 0.02734465 0.4999669 0.6074196     1.4752206
## [289,]  307 0.5532190 0.02807463 0.4946606 0.6049916     1.5850495
```

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

```
result1$vaccine$period_efficacy
```

```
##      left right       VE_a          se  lower .95 upper .95
## [1,]    0    60 0.6821729 0.02126667 0.63763337 0.7212379
## [2,]   60   120 0.7239177 0.02133139 0.67877631 0.7627153
## [3,]  120   180 0.6755593 0.02529073 0.62200229 0.7215281
## [4,]  180   240 0.4436504 0.04417056 0.34997670 0.5238249
## [5,]  240   300 0.2799564 0.09087456 0.07787798 0.4377503
```

The graphical depictions of estimates returned in **vaccine$efficacy** are generated by default by *dove()* and are shown in Figure 1. This figure can be regenerated using *plot()* as follows:
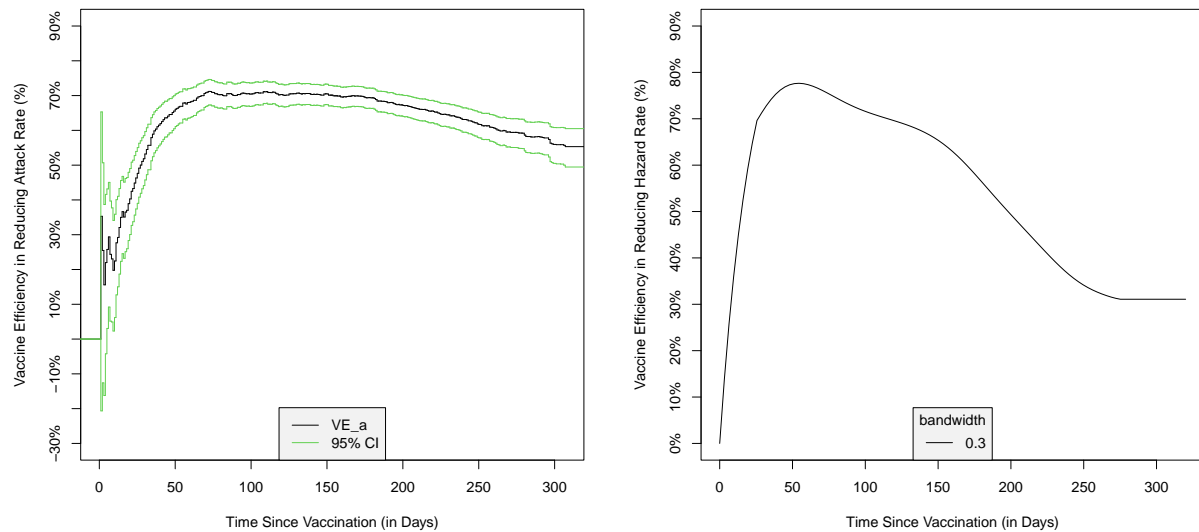
```
plot(x = result1)
```


Figure 1: Plots auto-generated by *dove()*. On the left, the estimated VE curve 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 VE curve in reducing the hazard ratio, $VE_h(t)$, is shown as a function of the time since vaccination.

dove2()

In the first analysis illustrating *dove2()*, we set Week 4 as the change point and assume a potentially waning VE after 4 weeks. We estimate VE_a over 0-4, 4-16, 16-28, 28-40 weeks. Note that all times must be provided in the unit of days. The function call takes the following form

```
result2 <- dove2(formula = Surv(event.time, event.status) ~ priority + sex +
                         vaccine(entry.time, vaccine.status, vaccine.time),
                data = doveData,
```

```
                changePts = 4*7,
                timePts = c(4, 16, 28, 40)*7)
```

```
## tau = 320
```

```
## Number of subjects: 40000
```

```
## log partial-likelihood at final estimates: -27351.81
```

The function returns an S3 object of class DOVE, which contains a list object with the following information.

call: The unevaluated call.

```
result2$call
```

```
## dove2(formula = Surv(event.time, event.status) ~ priority + sex +
##     vaccine(entry.time, vaccine.status, vaccine.time), data = doveData,
##     changePts = 4 * 7, timePts = c(4, 16, 28, 40) * 7)
```

changePts: The changePts of the analysis.

```
result2$changePts
```

```
## [1] 28
```

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

```
result2$covariates
```

```
##                 coef    se(coef)          z     Pr(>|z|) exp(coef) lower .95
## priority 0.2145486 0.01438123 14.918649 2.492693e-50  1.239302  1.204858
## sex      0.3646241 0.03940478  9.253296 2.176748e-20  1.439973  1.332945
##          upper .95
## priority  1.274732
## sex       1.555594
```

When no baseline covariates are provided, this element will be NA.

Vaccine Efficacy: Element **$VE_a** contains the daily VE estimate in reducing the attack rate, together with its standard error and the 95% confidence interval. Element **$VE_h** contains the daily VE estimate in reducing the hazard rate, together with its standard error and the 95% confidence interval.

```
head(result2$vaccine$VE_a)
```

```
##      time      VE_a          se  lower .95  upper .95
## [1,]    0 0.00000000 0.000000000 0.00000000 0.00000000
## [2,]    1 0.03024399 0.001222094 0.02784573 0.03263634
## [3,]    2 0.05927458 0.002346379 0.05466442 0.06386226
## [4,]    3 0.08714643 0.003379452 0.08049861 0.09374618
## [5,]    4 0.11391156 0.004327491 0.10538895 0.12235298
## [6,]    5 0.13961952 0.005196276 0.12937430 0.14974418
```

```
tail(result2$vaccine$VE_a)
```

```
##         time      VE_a         se lower .95 upper .95
## [316,]   315 0.5385651 0.02499808 0.4868731 0.5850498
## [317,]   316 0.5370153 0.02518384 0.4849277 0.5838355
```

```
## [318,]   317 0.5354577 0.02537152 0.4829703 0.5826167
## [319,]   318 0.5338922 0.02556114 0.4810008 0.5813935
## [320,]   319 0.5323189 0.02575270 0.4790191 0.5801658
## [321,]   320 0.5307377 0.02594622 0.4770253 0.5789336
```

```r
head(result2$vaccine$VE_h)
```

```
##      time       VE_h          se  lower .95  upper .95
## [1,]    0 0.00000000 0.000000000 0.00000000 0.00000000
## [2,]    1 0.05987195 0.002394147 0.05516769 0.06455278
## [3,]    2 0.11615925 0.004501610 0.10729190 0.12493851
## [4,]    3 0.16907651 0.006348134 0.15654055 0.18142616
## [5,]    4 0.21882552 0.007957412 0.20307225 0.23426738
## [6,]    5 0.26559596 0.009351233 0.24703692 0.28369756
```

```r
tail(result2$vaccine$VE_h)
```

```
##          time       VE_h          se  lower .95 upper .95
## [316,]   315 0.05159091 0.09347856 -0.1505214 0.2181981
## [317,]   316 0.04603717 0.09443096 -0.1582220 0.2142741
## [318,]   317 0.04045090 0.09539185 -0.1659755 0.2103312
## [319,]   318 0.03483192 0.09636127 -0.1737821 0.2063694
## [320,]   319 0.02918004 0.09733931 -0.1816423 0.2023886
## [321,]   320 0.02349506 0.09832601 -0.1895563 0.1983886
```

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

```r
result2$vaccine$VE_period
```

```
##      left right       VE_a         se lower .95 upper .95
## [1,]    0    28 0.5242172 0.01230293 0.4994819 0.5477300
## [2,]   28   112 0.7708700 0.01303526 0.7438421 0.7950462
## [3,]  112   196 0.6258165 0.01823842 0.5883060 0.6599094
## [4,]  196   280 0.3889351 0.04248634 0.2997210 0.4667834
```

The graphical depictions of VE_a and VE_h estimates are generated by default by *dove2()* and are shown in Figure 2. This figure can be regenerated using *plot()* as follows:

```r
plot(x = result2)
```

In the final analysis, we have the software use AIC to choose a change point among Weeks 4, 5, 6, 7, 8. We assume a constant VE after the change point, and thus only the constant VE is estimated. The function call takes the following form

```r
result3 <- dove2(formula = Surv(event.time, event.status) ~ priority + sex +
                          vaccine(entry.time, vaccine.status, vaccine.time),
                 data = doveData,
                 constantVE = TRUE)
```

```
## tau = 320
```

```
## changePts not given; using AIC to select from {28,35,42,49,56}
```

```
## Day 28 (week 4) was selected as the change point by AIC
```

```
## Number of subjects: 40000
```

```
## log partial-likelihood at final estimates: -27422.25
```
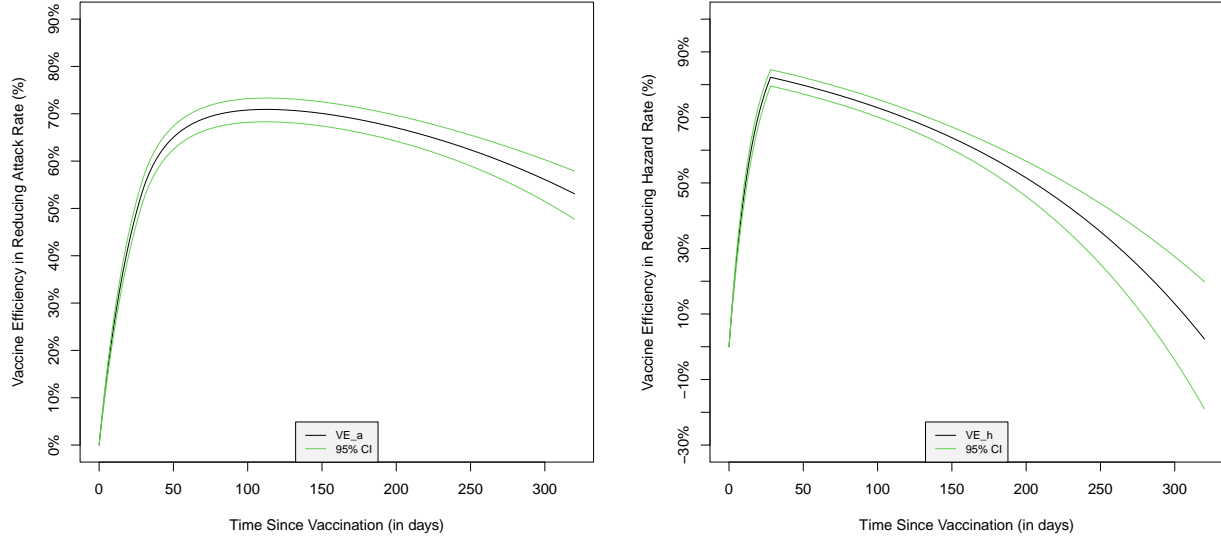
Figure 2: Plots auto-generated by *dove2()*. On the left, the estimated VE curve in reducing the attack rate, $VE_a(t)$ (black) and its 95% confidence intervals (green) are shown as a function of the time since vaccination. On the right, the estimated VE curve in reducing the hazard ratio, $VE_h(t)$ (black) and its 95% confidence intervals (green) are shown as a function of the time since vaccination.

The function returns a list object containing the following items.

changePts: The change point selected.

```
result3$changePts
```

```
## [1] 28
```

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

```
result3$covariates
```

```
##                coef   se(coef)         z     Pr(>|z|) exp(coef) lower .95
## priority 0.2127205 0.01450545 14.664868 1.082278e-48  1.237039  1.202364
## sex      0.3652731 0.03940404  9.269939 1.862523e-20  1.440907  1.333812
##          upper .95
## priority  1.272713
## sex       1.556602
```

Vaccine Efficacy: Element **$VE** contains the estimated constant VE, together with its standard error and the 95% confidence interval.

```
result3$vaccine$VE
```

```
##         VE         se  lower .95  upper .95
## 0.68233226 0.01443303 0.65274570 0.70939801
```

Plots cannot be generated when `constantVE` = TRUE.

9

```
plot(x = result3)
```

plot() is not available for the provided analysis (constantVE = TRUE)

References

Lin DY, Zeng D, Gilbert PB (2021a). Evaluating the long-term efficacy of COVID-19 vaccines. Clinical Infectious Diseases, ciab226, https://doi.org/10.1093/cid/ciab226

Lin, D-Y, Gu, Y., Zeng, D., Janes, H. E., and Gilbert, P. B. (2021b). Evaluating Vaccine Efficacy Against SARS-CoV-2 Infection. Clinical Infectious Diseases, ciab630, https://doi.org/10.1093/cid/ciab630