

# iDOVE – Durability Of Vaccine Efficacy Against SARS-CoV-2 Infection

Yu Gu, Shannon T. Holloway, and Dan-Yu Lin

September 9, 2021

## Introduction

**iDOVE** is an R package for assessing potentially time-varying vaccine efficacy (VE) against SARS-CoV-2 infection under staggered enrollment of participants and time-varying community transmission, allowing crossover of placebo volunteers to the vaccine arm before the end of the study. Infection is not directly observed, but is rather known to occur between two examinations. **iDOVE** implements the methodology of Lin et al. (2021) for estimating time-varying VE against such interval-censored events, representing the log hazard ratio for the vaccine effect by a piece-wise linear function of time since vaccination. The special case of right-censored events is implemented by `dove2()` in the **DOVE** package.

**iDOVE** inputs a rectangular data set with the following information:

- **Entry time:** The time when the participant enters the trial.
- **Left interval time:** The last examination time when the test is negative.
- **Right interval time:** The first examination time when the test is positive.
- **Vaccination time:** The time when vaccination takes place.
- **Covariates:** Baseline covariates (e.g., priority group, age, sex, ethnicity).

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

The primary analysis tool of the package is `idove()`, which returns 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 rates over successive time periods. The standard errors and 95% confidence intervals are also provided.

In addition, the package includes three convenience functions: `intCens()`, which is used to wrap all of the input time variables together and is part of the model statement of `idove()`; `print()`, which displays the primary results of the analysis; and `plot()`, which generates plots of the estimated VE curves. Finally, a simulated dataset is provided to illustrate the use of the software.

## Functions

### `intCens()`

This convenience function is used as the left-hand side of a formula object for the sole purpose of simplifying the specification of required input variables: entry time, left interval time, right interval time, and vaccination time. This function is not intended to be used as a stand-alone feature. 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

```
intCens(entry_time, left_time, right_time, vaccination_time)
```

where `entry_time` is the time when the participant enters the trial; `left_time` is the last examination time when the test is negative; `right_time` is the first examination time when the test is positive (NA or Inf if the participant is never tested positive during the clinical trial); `vaccination_time` is the time when vaccination takes place (NA or Inf if the participant is not vaccinated during the trial). Note that all times must be provided in units of whole days.

### *idove()*

This function is the primary tool of **iDOVE**. The value object returned is an S3 object of class `iDOVE` that contains the estimated hazard ratio for each baseline covariate, the estimated VE 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 VE in reducing the attack rates over  $m$  successive time periods,  $VE_a(0, t_1), VE_a(t_1, t_2), \dots, VE_a(t_{m-1}, t_m)$ . By definition,  $VE_a(0, t) = VE_a(t)$ .

The function call takes the following form:

```
idove(formula, data, constantVE = FALSE, plots = TRUE,
      changePts = NULL, timePts = NULL, tol = 0.0001, maxit = 2000)
```

where

- `formula` is a model statement. See below for further details.
- `data` is a `data.frame` object containing all required data as previously described.
- `constantVE` is a logical object specifying the VE trend after the last change point. If `TRUE`, VE is assumed to be constant after the last change point; otherwise, VE is allowed to vary after the last change point.
- `plots` is a logical object indicating whether graphical forms of the estimated  $VE_a(t)$  and  $VE_h(t)$  curves are to be generated.
- `changePts` is an optional integer vector to specify the change points for the vaccine effect on the hazard ratio. 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) to capture the ramping vaccine effect after the initial shot.
- `timePts` is an optional integer vector to specify the time points  $(t_1, t_2, \dots, t_m)$  for partitioning the study period in the estimation of VE on the attack rates over successive time periods. If not provided, a default sequence  $t_1, 2t_1, 3t_1, \dots$  will be used, where  $t_1$  is the first change point. The sequence ends at the maximum of the finite left and right interval times from all participants.
- `tol` is the convergence threshold for the EM algorithm.
- `maxit` is the maximum number of iterations for the EM algorithm.

The model statement is a formula object. The left side is an object returned by the `intCens()` function and specifies all time variables. The right side contains all baseline covariates; a model without baseline covariates is allowed. Categorical baseline covariates can be specified, and all other categories are compared to the first category.

The `formula` input takes the following general structure

```
intCens(entry_time, left_time, right_time, vaccination_time) ~ covariates
```

where ‘event\_time’, ‘left\_time’, ‘right\_time’, ‘vaccination\_time’, and ‘covariates’ are place holders indicating the data that are to be provided; they should be replaced by the appropriate 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 maximum of all finite left and right ends of the time intervals. To ensure stable estimates, we suggest placing change points at times (since vaccination) when there are relatively large numbers of events and not placing change points at the right tail.

### *plot()*

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

### *print()*

When provided the value object returned by *idove()*, the tabular results are displayed.

## Examples

To illustrate the call structure and results of *idove()*, we use the dataset provided with the package, *idoveData*. This dataset was simulated under a blinded, 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.
- **left.time**: The left end of the time interval in days.
- **right.time**: The right end of the time interval in days.
- **vaccine.time**: The time of vaccination in days.
- **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(idoveData)
```

```
head(idoveData)
```

```
##   entry.time left.time right.time vaccine.time priority sex
## 1      113      186         Inf         186         5    0
## 2       59      286         Inf         286         2    0
## 3      101      308         Inf         101         1    0
## 4       12       68        212         212         4    1
## 5       83      105        136         221         5    0
## 6       93      116        151         222         4    1
```

Consider the summary statistics

```
summary(idoveData)
```

```
##   entry.time    left.time    right.time  vaccine.time    priority
## Min.   : 0.00  Min.   : 2.0  Min.   : 4  Min.   : 0.0  Min.   :1.000
## 1st Qu.: 30.00  1st Qu.:244.0  1st Qu.:Inf  1st Qu.: 60.0  1st Qu.:2.000
## Median : 60.00  Median :274.0  Median :Inf  Median :150.0  Median :3.000
```

```
## Mean : 60.15 Mean :261.8 Mean :Inf Mean :156.7 Mean :3.007
## 3rd Qu.: 90.00 3rd Qu.:298.0 3rd Qu.:Inf 3rd Qu.:252.0 3rd Qu.:4.000
## Max. :120.00 Max. :315.0 Max. :Inf Max. :315.0 Max. :5.000
## sex
## Min. :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean :0.4951
## 3rd Qu.:1.0000
## Max. :1.0000
```

```
summary(idoveData$right.time[is.finite(idoveData$right.time)])
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 4.0 174.2 225.0 208.7 261.0 315.0
```

We can see that participants were enrolled in the study over a 4-month period ( $0 \leq \text{entry.time} \leq 120$  days), the follow-up time ended on day 315 (left.time and finite right.time  $\leq 315$  days), and more than 75% of the participants were never tested positive during the follow-up (right.time = Inf indicates that a participant did not test positive during the course of the trial). 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 both baseline covariates, priority and sex.

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

```
model <- intCens(entry.time, left.time, right.time, vaccine.time) ~ priority + sex
result1 <- idove(formula = model,
                 data = idoveData,
                 changePts = 4*7,
                 timePts = c(4, 16, 28, 40)*7)
```

```
## changePts: {28}
## performing nonparametric maximum likelihood
## Iteration 100 : difference = 0.000140
## EM algorithm converged after 118 iterations
## Number of subjects: 40000
## Number of unique time points: 313
## Log-likelihood at final estimates: -10880.11
## PL converged after 2 iterations
## PL converged after 2 iterations
## PL converged after 2 iterations
## PL converged after 2 iterations
```

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

**call:** The unevaluated call.

```
result1$call
```

```
## idove(formula = model, data = idoveData, changePts = 4 * 7, timePts = c(4,
## 16, 28, 40) * 7)
```

**changePts:** The changePts of the analysis.

```
result1$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.

```
result1$covariates
```

```
##           coef   se(coef)      z    Pr(>|z|) exp(coef) lower .95
## priority 0.2055465 0.01537842 13.36591 9.565724e-41  1.228196  1.191729
## sex      0.2844556 0.04174790  6.81365 9.515337e-12  1.329038  1.224619
##           upper .95
## priority  1.265780
## sex       1.442361
```

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(result1$vaccine$VE_a)
```

```
##      time      VE_a      se lower .95 upper .95
## [1,]  0 0.00000000 0.00000000 0.00000000 0.00000000
## [2,]  1 0.02962129 0.001682040 0.02631888 0.03291249
## [3,]  2 0.05807850 0.003232211 0.05172201 0.06439238
## [4,]  3 0.08542300 0.004659225 0.07624518 0.09450964
## [5,]  4 0.11170374 0.005971253 0.09992265 0.12333064
## [6,]  5 0.13696737 0.007175949 0.12278727 0.15091824
```

```
tail(result1$vaccine$VE_a)
```

```
##      time      VE_a      se lower .95 upper .95
## [311,] 310 0.6004585 0.02828805 0.5409826 0.6522280
## [312,] 311 0.5994380 0.02853782 0.5394102 0.6516424
## [313,] 312 0.5984129 0.02878989 0.5378278 0.6510559
## [314,] 313 0.5973832 0.02904426 0.5362354 0.6504686
## [315,] 314 0.5963490 0.02930095 0.5346328 0.6498804
## [316,] 315 0.5953102 0.02955993 0.5330202 0.6492913
```

```
head(result1$vaccine$VE_h)
```

```
##      time      VE_h      se lower .95 upper .95
## [1,]  0 0.00000000 0.00000000 0.00000000 0.00000000
## [2,]  1 0.05865177 0.003296646 0.05216812 0.06509107
## [3,]  2 0.11386351 0.006206584 0.10161472 0.12594529
## [4,]  3 0.16583698 0.008763835 0.14848179 0.18283845
## [5,]  4 0.21476212 0.010999761 0.19290389 0.23602837
## [6,]  5 0.26081771 0.012943257 0.23500857 0.28575610
```

```
tail(result1$vaccine$VE_h)
```

```
##      time      VE_h      se lower .95 upper .95
## [311,] 310 0.2847861 0.1159957 0.017145854 0.4795455
## [312,] 311 0.2813357 0.1171253 0.010867601 0.4778470
```

```
## [313,] 312 0.2778685 0.1182635 0.004548372 0.4761435
## [314,] 313 0.2743847 0.1194103 -0.001812091 0.4744348
## [315,] 314 0.2708841 0.1205657 -0.008214052 0.4727210
## [316,] 315 0.2673665 0.1217297 -0.014657773 0.4710021
```

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

```
result1$vaccine$VE_period
```

```
##      left right      VE_a      se lower .95 upper .95
## [1,]   0    28 0.5178866 0.01726815 0.4828247 0.5505715
## [2,]  28   112 0.7731445 0.01572050 0.7401418 0.8019558
## [3,] 112   196 0.6601218 0.01976331 0.6190920 0.6967320
## [4,] 196   280 0.4907895 0.05547956 0.3695665 0.5887031
```

The graphical depictions of  $VE_a$  and  $VE_h$  estimates are generated by default by `idove()` and are shown in Figure 1. This figure can be regenerated using `plot()` as follows:

```
plot(x = result1)
```

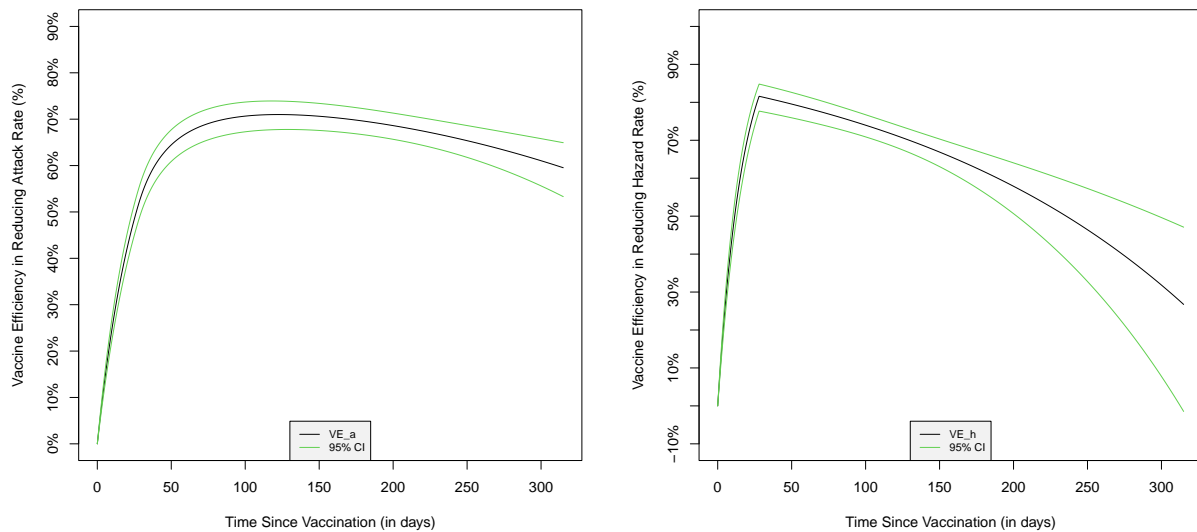


Figure 1: Plots auto-generated by `idove()`. On the left, the estimated  $VE_a(t)$  curve (black) and its 95% confidence intervals (green) are shown as a function of the time since vaccination. On the right, the estimated  $VE_h(t)$  curve (black) and its 95% confidence intervals (green) are shown as a function of the time since vaccination.

In the second example, 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

```
result2 <- idove(formula = model, data = idoveData, constantVE = TRUE)
```

```
## constantVE selected
## changePts not given; using AIC to select from {28,35,42,49,56}
## performing nonparametric maximum likelihood
```

```

## evaluating change point: 28
## EM algorithm converged after 28 iterations
## evaluating change point: 35
## EM algorithm converged after 29 iterations
## evaluating change point: 42
## EM algorithm converged after 29 iterations
## evaluating change point: 49
## EM algorithm converged after 30 iterations
## evaluating change point: 56
## EM algorithm converged after 29 iterations
## Day 28 (week 4) was selected as the change point by AIC
## Number of subjects: 40000
## Number of unique time points: 313
## Partial log-likelihood at final estimates: -10909.76
## PL converged after 2 iterations
## PL converged after 2 iterations
## PL converged after 1 iterations

```

The function returns a list object containing the following items.

**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 0.2022670 0.01522801 13.282564 2.922128e-40 1.224175 1.188177
## sex      0.2850122 0.04142350  6.880447 5.966491e-12 1.329778 1.226080
##           upper .95
## priority 1.261263
## sex      1.442247

```

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

```
result2$vaccine$VE
```

```

##           VE           se lower .95 upper .95
## 0.71086150 0.01471139 0.68054042 0.73830470

```

## References

Lin, D-Y, Gu, Y., Zeng, D., Janes, H. E., and Gilbert, P. B. (2021). Evaluating vaccine efficacy against SARS-CoV-2 infection. *Clinical Infectious Diseases*, ciab630, <https://doi.org/10.1093/cid/ciab630>