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

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

June 17, 2021

Introduction

iDOVE is an R package for assessing potentially time-varying vaccine efficacy (VE) against SARS-CoV-2 infection under staggered enrollment and time-varying community transmission, allowing crossover of placebo volunteers to the vaccine arm. The infection time can be either interval- or right-censored, the latter being a special case of the former with very frequent monitoring. iDOVE implements both the method for interval-censored data proposed in Lin et al. (2021) and the standard Cox regression method for right-censored data. The iDOVE function takes as input 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, with an arbitrary value that is greater than the last examination time (e.g., the end time of the clinical trial) if the participant is not vaccinated during the trial.
- **Covariates**: Baseline covariates (e.g., priority group, age, sex, ethnicity).

Note that an arbitrary number of baseline covariates can be included. 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 accommodates both interval- and right-censored infection time data. The implementation uses linear splines to approximate the log hazard ratio with respect to vaccination and, thus, is capable of estimating either the constant or the potentially waning vaccine efficacy. Specifically, in K-piece linear splines, the log hazard ratio at time t is given by

$$\eta(t) = \gamma_1 t + \gamma_2 (t - x_1)_+ + \gamma_3 (t - x_2)_+ + \cdots + \gamma_K (t - x_{K-1})_+,$$

where $x_1, \ldots, x_{K-1}$ are the $K-1$ pre-specified change points, $\gamma_1, \ldots, \gamma_K$ are the K spline parameters, and $t_+ = t$ if $t > 0$ and 0 otherwise. The first $K-1$ spline parameters are always estimated from the data, whereas γ_K is either estimated or given by $\gamma_K = -\sum_{k=1}^{K-1} \gamma_k$, depending on the assumption of vaccine efficacy after the last change point x_{K-1}.

Function *idove()* 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 rate (instantaneous risk), and the estimated vaccine efficacy in reducing the attack rates over successive time periods.

In addition, the package includes three convenience functions *intCens()*, which is used to wrap all the input time variables together as the response required in the model statement of *idove()*; *print()*, which displays the primary results of the analysis; and *plot()*, which generates plots of the estimated vaccine efficacies. Finally, a simulated dataset is provided to illustrate the use of the software.

Functions

intCens()

This convenience function is used as the response 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. 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 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 rates over m successive time periods, $VE_a(0,t_1), VE_a(t_1,t_2), \ldots, VE_a(t_{m-1},t_m)$. By definition, $VE_a(0,t) = VE_a(t)$.

The function call takes the following form:

```
idove(formula, data, constantVE = FALSE, rightCens = 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 the data.frame object containing all required data as previously described.
- `constantVE` is a logical object. If TRUE, the vaccine efficacy is assumed to be constant in the period after the last change point. If FALSE, the VE is assumed to be potentially waning after the last change point.
- `rightCens` is a logical object. If TRUE, the standard Cox regression for right-censored infection time data is used estimate the vaccine efficacy. If FALSE, the nonparametric maximum likelihood method for interval-censored infection time data is used.
- `plots` is a logical object indicating whether graphical forms of the estimated vaccine efficacy in reducing the attack rate, $VE_a(t)$, and in reducing the hazard rate, $VE_h(t)$, are to be generated.

- `changePts` is an optional integer vector to specify the change points $(x_1, \ldots, x_{K-1})$ of the piece-wise log-linear hazard ratio. If no change points are provided, one change point will automatically be selected among Weeks 4, 5, 6, 7, 8 by AIC.
- `timePts` is an optional vector to specify the time points $(t_1, t_2, \ldots, t_m)$ for partitioning the study period in the estimation of the attack rates over successive time periods. If not provided, a default sequence $t_1, 2t_1, 3t_1, \ldots$ 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 or Newton-Raphson algorithm.
- `maxit` is the maximum number of iterations for the EM or Newton-Raphson algorithm.

The model statement is a formula object. The left-hand-side is an object returned by the *intCens()* function and specifies all time variables. The right-hand-side contains all baseline covariates; a model without baseline covariates is allowed. Note that 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

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

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

When `right_time - left_time` ≤ 2 for all individuals whose infection times are truly interval-censored (i.e., `right_time` is finite), the software assumes that the examinations are completed daily or every two days, and performs the standard Cox regression, regardless of the value provided through input `rightCens`. In general, we suggest placing change points at times (since vaccination) when there are sufficient probabilities of events. Also, we suggest not placing change points at the tail, otherwise the estimation on final pieces might be unstable. The two measures of vaccine efficacy, $VE_a(t)$ and $VE_h(t)$, are estimated up to the maximum of all finite left and right ends of the intervals. However, the estimates at the tail may not be reliable because there are very few participants under follow-up. 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 or every two months.

The value object returned by *idove()* is an S3 object of class iDOVE that has additional attributes that are used by package convenience functions:

- **knots**: The knots of the linear spline.
- **tau**: The length of the trial in days.
- **gamma**: The estimated spline parameters.
- **covgamma**: The covariance matrix of the spline parameters.

plot()

When provided the value object returned by *idove()*, this convenience function creates/recreates plots of the estimated vaccine efficacy 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 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
## 19471        113       186        Inf          186        5   0
## 6506          59       286        Inf          286        2   0
## 31984        101       308        Inf          101        1   0
## 11892         12        68        212          212        4   1
## 15134         83       105        136          221        5   0
## 9060          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 ($0 \leq$ entry.time ≤ 120 days); that the follow-up time ended on day 315 (left.time and finite right.time ≤ 315 days); and that 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 vaccine efficacy 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)
```

```
## 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: -10880.1052514139

## 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)
```

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 **$VE_a** contains the estimated vaccine efficacy in reducing the attack rate at the endpoint of each time interval, together with its standard error and the 95% confidence interval. Element **$VE_h** contains the estimated vaccine efficacy in reducing the hazard rate at the endpoint of each time interval, together with its standard error and the 95% confidence interval.

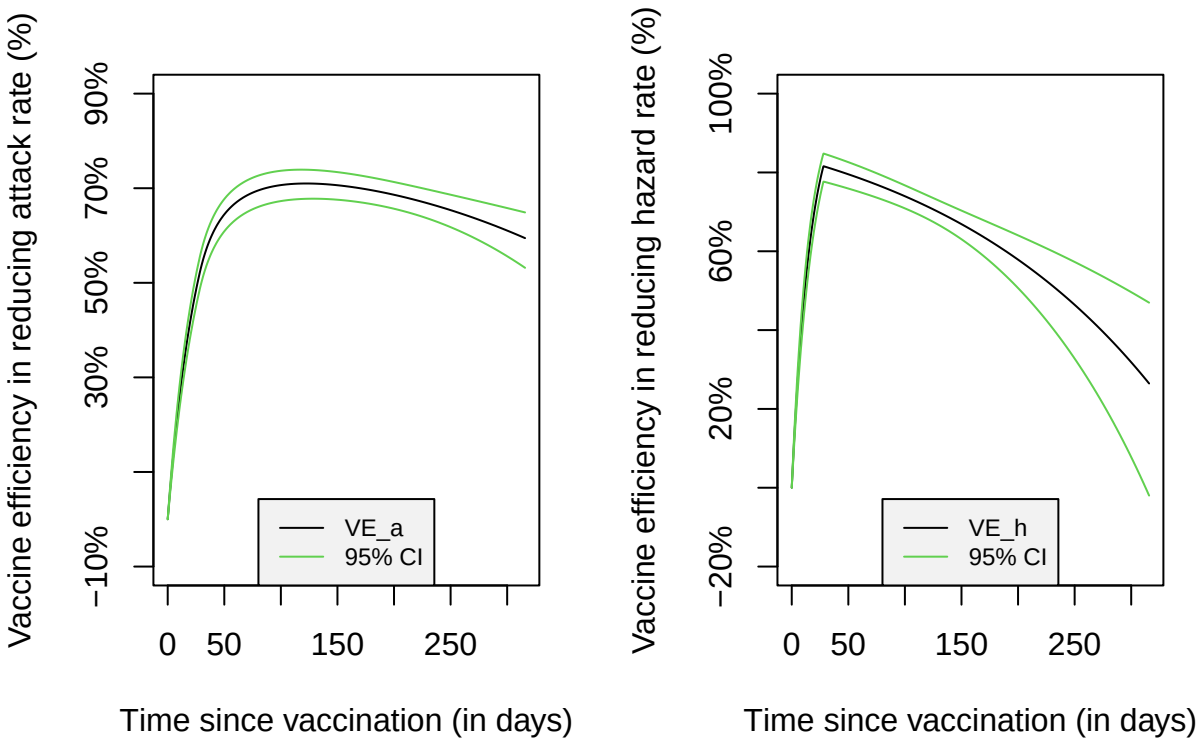

Figure 1: Plots auto-generated by *idove()*. On the left, the estimated curve of vaccine efficacy in reducing the attack rate, $VE_a(t)$ (black) and its 95% confidence intervals (green) are shown as a function of the time since vaccination. On the right, the estimated curve of vaccine efficacy in reducing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

```
result1$vaccine$VE_a
```

```
##      time      VE_a         se lower .95 upper .95
## [1,]   28 0.5178866 0.01726815 0.4828247 0.5505715
## [2,]  112 0.7093301 0.01597110 0.6762789 0.7390067
## [3,]  196 0.6882408 0.01434110 0.6588261 0.7151195
## [4,]  280 0.6290054 0.02185599 0.5835965 0.6694625
```

```
result1$vaccine$VE_h
```

```
##      time      VE_h         se lower .95 upper .95
## [1,]   28 0.8159191 0.01805048 0.7769117 0.8481060
## [2,]  112 0.7242073 0.01486782 0.6934711 0.7518615
## [3,]  196 0.5868033 0.03229145 0.5184075 0.6454856
## [4,]  280 0.3809426 0.08582152 0.1876609 0.5282364
```

Element **$VE_period** contains the estimated vaccine efficacy 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)
```

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 vaccine efficacy after the change point and thus only the constant vaccine efficacy is estimated. The function call takes the following form

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

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

## performing nonparametric maximum likelihood

## EM algorithm converged after 28 iterations
## EM algorithm converged after 29 iterations
## EM algorithm converged after 29 iterations
## EM algorithm converged after 30 iterations
## EM algorithm converged after 29 iterations

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

## Number of subjects: 40000

## Partial log-likelihood at final estimates: -10909.7584228153

## 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 vaccine efficacy, 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. Submitted.