DOVE2 – Durability of Vaccine Efficacy

June 18, 2021

Methods

Kaplan-Meier Estimator

In the Kaplan-Meier (KM) method, the event time is measured from the participant's entry time. Let $\widehat{G}_0(\cdot)$ and $\widehat{G}_1(\cdot)$ denote the KM estimates of the survival functions for the placebo and vaccine groups, respectively. Then the KM estimates of the cumulative incidence at time t after randomization in the placebo and vaccine groups are $1-\widehat{G}_0(t)$ and $1-\widehat{G}_1(t)$, respectively. The KM estimate of the vaccine efficacy in reducing the cumulative incidence over the time period $(t_0, t]$ is

$$\widehat{VE}_{CI}\left(t_0, t\right) = 1 - \frac{\widehat{G}_1\left(t_0\right) - \widehat{G}_1(t)}{\widehat{G}_0\left(t_0\right) - \widehat{G}_0(t)},$$

where t_0 is set to 4 weeks by default to reflect the ramping vaccine effect after the 1st dose.

Standard Cox Model

In the standard Cox model, the event time is measured from the participant's entry time, and the events that occur during the first 4 weeks are excluded from the analysis. Let X denote the baseline covariates, and Z indicate, by the values 1 versus 0, whether the participant is on vaccine or placebo. The hazard function conditional on X and Z takes the form

$$\widetilde{\lambda}(t|X, Z) = \widetilde{\lambda}_0(t) e^{\beta^T X + \gamma Z},$$

where $\widetilde{\lambda}_0(\cdot)$ is an arbitrary baseline hazard function, and β and γ are log hazard ratios. We estimate the vaccine efficacy on the hazard rate by $1 - e^{\hat{\gamma}}$, where $\hat{\gamma}$ is the maximum partial likelihood estimate of γ.

Standard Poisson Model

Under the standard Poisson model, we estimate vaccine efficacy by $1 - E_1/E_0$, where E_1 is the number of events that occur after week 4 in the vaccine group, and E_0 is the corresponding number in the placebo group.

New Cox Model

In the new Cox model, the event time is measured from the start of the clinical trial. We fit the Cox model with a time-varying hazard ratio

$$\lambda(t|X, S) = \lambda_0(t) e^{\beta^T X + \eta(t-S) I(S<t)},$$

where S is the calendar time when the 1st dose occurs, $\lambda_0(\cdot)$ is an arbitrary baseline hazard function, and $\eta(\cdot)$ is the log hazard ratio. We center each component of X at its sample medium, such that $\lambda_0(\cdot)$ pertains to the medium of each covariate. We assume a piece-wise linear form for the log hazard ratio:

$$\eta(t) = \gamma_1 t + \gamma_2 (t - t_1)_+ + \gamma_3 (t - t_2)_+ + \cdots + \gamma_K (t - t_{K-1})_+,$$

where $t_1, \ldots, t_{K-1}$ are the $(K-1)$ pre-specified knots, and $\gamma_1, \ldots, \gamma_K$ are the K parameters to be estimated from the data. By default, two knots are placed at 30 and 60 days, so we only need to estimate three parameters in order to estimate $\eta(\cdot)$. We denote the maximum partial likelihood estimate of $\eta(\cdot)$ by $\hat{\eta}(\cdot)$, and denote the Breslow estimator of the cumulative baseline hazard function by $\widehat{\Lambda}_0(\cdot)$. Then the estimated cumulative incidence at time t after the time of vaccination s is given by

$$1 - \exp\left\{-\int_s^{s+t} e^{\hat{\eta}(u-s)} d\widehat{\Lambda}_0(u)\right\}$$

for the vaccine group and

$$1 - \exp\left\{\widehat{\Lambda}_0(s) - \widehat{\Lambda}_0(s+t)\right\}$$

for the placebo group. The vaccine efficacy in reducing the cumulative incidence over the time period $(t_0, t]$ for the individuals who received the 1st dose at time s can be estimated by

$$\widehat{VE}_{CI}\,(t_0, t; s) = 1 - \frac{\exp\left\{-\int_s^{s+t_0} e^{\hat{\eta}(u-s)} d\widehat{\Lambda}_0(u)\right\} - \exp\left\{-\int_s^{s+t} e^{\hat{\eta}(u-s)} d\widehat{\Lambda}_0(u)\right\}}{\exp\left\{\widehat{\Lambda}_0(s) - \widehat{\Lambda}_0(s+t_0)\right\} - \exp\left\{\widehat{\Lambda}_0(s) - \widehat{\Lambda}_0(s+t)\right\}}.$$

In addition, we can estimate the vaccine efficacy in reducing the hazard rate at time t after the 1st dose by

$$\widehat{VE}_{HR}(t) = 1 - e^{\hat{\eta}(t)},$$

which is independent of the vaccination time. We further construct the 95% confidence interval for $VE_{HR}(t)$ based on the 95% confidence interval for $\eta(t)$.

Software

The **DOVE2** package takes as input a rectangular data set with the following information:

- **Entry time**: Calendar time when the participant entered the trial (in days).
- **Event time**: Calendar time when the participant experienced symptomatic COVID-19 or when the follow-up ended, whichever occurred first (in days).
- **Event status**: Binary indicator on whether symptomatic COVID-19 occurred before the end of follow-up.
- **Vaccination status**: Binary indicator taking value 1 if vaccination occurred before the end of follow-up and 0 otherwise.
- **Vaccination time**: Calendar time when the 1st dose of vaccination took place (in days, with an arbitrary value if the participant was not vaccinated).
- **Covariates**: Baseline covariates (e.g., priority group, age, sex, ethnicity).

The primary analysis tool of the package is *dove2()*, which implements all four aforementioned methods and generates a figure that depicts the estimation results of the cumulative incidence, VE_{CI} and VE_{HR} based on the Kaplan-Meier method or the new Cox model.

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 *dove2()*, similar in spirit to the *cluster()* function of the **survival** package.

Finally, a simulated dataset is provided to illustrate the use of the software.

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 entered the trial; `vaccination_status` is the binary indicator taking value 1 if vaccination occurred before the end of follow-up and 0 otherwise; `vaccination_time` is the time when the 1st dose of vaccination took place (with an arbitrary value if the participant was not vaccinated).

Note that all the time variables are measured from the beginning of the clinical trial and are specified in units of days. For each individual, the `entry_time`, `event_time` and `vaccination_time` must satisfy `entry_time` $\leq$ `event_time` and `entry_time` $\leq$ `vaccination_time`. If `entry_time` $>$ `event_time` or `entry_time` $>$ `vaccination_time`, the case will be removed from the analysis and a message will be generated. The software automatically classify all participants into the placebo and vaccine groups according to whether they were vaccinated at study entry (i.e., `vaccination_time` = `entry_time`).

dove2()

This function is the primary tool of **DOVE2**. The value object returned contains the estimation results of vaccine efficacy based on each of the four methods. Graphical depictions of the estimates by the Kaplan-Meier method and the new Cox model can be generated upon request.

The function call takes the following form:

```
dove2(formula, data, changePts = c(30L, 60L), plots = TRUE)
```

where

- `formula` is a model statement. See below for further details.
- `data` is the data.frame object containing all required data as previously described.
- `changePts` is a numerical vector object to specify the change points (in days) of the piece-wise log-linear hazard ratio. The default change points are 30 and 60 days.
- `plots` is a logical object indicating whether graphical forms of the estimated vaccine efficacy in reducing the cumulative incidence, VE_{CI}, and in reducing the hazard rate, VE_{HR}, are to be generated. If TRUE (default), the plots will be generated and saved to a PDF file named "figure.pdf" in the current working directory.

The model statement is a formula object. The left-hand-side is an object returned by the *vaccine()* function and specifies all time variables. The right-hand-side contains all baseline covariates; a model without baseline covariates is allowed. The `formula` input takes the following general structure

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

where `entry_time`, `event_time`, `event_status`, `vaccination_status`, `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.

The value object returned by *dove2()* is a list containing four elements, one for each method. Specifically,

- **Kaplan-Meier method**: A list containing two elements:
  - The jump points and survival probabilities of the KM estimates for both the placebo and the vaccine groups.
  - The estimated $VE_{CI}(t_0, t)$ at all jump points, excluding the first 4 weeks of events after the 1st dose.
- **Standard Cox model**: A vector of the estimated vaccine efficacy and its standard error.
- **Standard Poisson model**: A scalar of the estimated vaccine efficacy.
- **New Cox model**: A list containing ten data frames:
  - The estimated cumulative incidence for individuals who were vaccinated on the following four dates: (1) first entry time; (2) medium entry time; (3) last entry time; (4) last entry time + 2 months.
  - The estimated $VE_{CI}(t_0, t; s)$ for individuals who were vaccinated at the above four dates.
  - The estimated $VE_{HR}(t)$ and its 95% confidence interval.
  - The estimation results on the hazard ratio of each covariate.

Illustration

We use the dataset 'dove2Data' provided with the package for illustration. This dataset was simulated under an unblinded, 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 time to event or censoring in days, whichever occurred first.
- **event.status**: The binary indicator on whether the event occurred before censoring.
- **vaccine.time**: The time of the 1st dose in days.
- **vaccine.status**: The binary indicator on whether the vaccination occurred before censoring.
- **priority**: A composite baseline priority score taking values 1-5.
- **sex**: A binary indicator of sex (male/female).

The data can be loaded in the usual way

```
data(dove2Data)
```

```
head(dove2Data)
```

```
##   entry.time event.time event.status vaccine.time vaccine.status priority sex
## 1        111        276            0          276              1        4   1
## 2         24        262            0          262              1        3   1
## 3        107        228            0          228              1        1   1
## 4         16        230            0          230              1        2   1
## 5         26        317            0          317              1        5   0
```

```
## 6          12          225              1            241                1          2    1
```

In this analysis, we will include in our model statement both baseline covariates. We will use the default values for the other arguments of the *dove2()* function. The function call takes only one line of code

```
result <- dove2(formula = Surv(event.time, event.status) ~ priority + sex +
                  vaccine(entry.time, vaccine.status, vaccine.time),
                data = dove2Data)
```

```
## tau set to Day 322
## Number of subjects: 40000
## Number of subjects in the placebo group: 20000
## Number of subjects in the vaccine group: 20000
## Warning: Removed 162 row(s) containing missing values (geom_path).
```

By default, a figure (as is shown in Figure 1) will be generated and saved to "figure.pdf" in the current working directory.

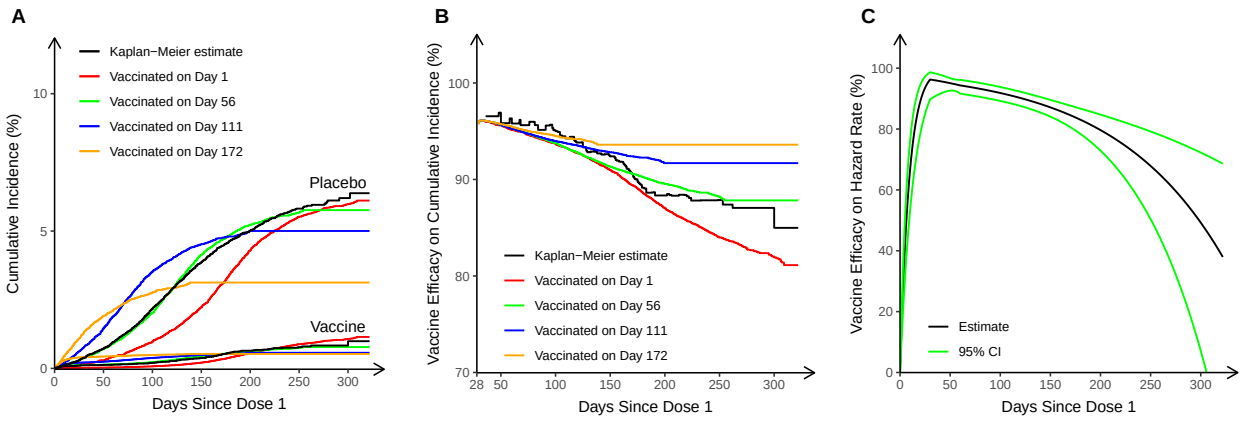

Figure 1: Estimation of vaccine efficacy in a clinical trial: A. Kaplan-Meier estimates of the cumulative incidence of disease for the vaccine and placebo groups and the cumulative incidence curves for participants who were vaccinated at various dates under the new Cox model; B. Estimates of VE_{CI} based on the Kaplan-Meier method and the new Cox model for participants who were vaccinated at various dates; C. Estimates and 95% confidence intervals for VE_{HR} under the new Cox model.

In addition to the figure, the function returns a list of four elements, one for each method. Some of them are used internally to generate the figure.

```
str(result, max=2)
```

```
## List of 4
##  $ KM     :List of 2
##   ..$ Sur_km  :List of 2
##   ..$ VE_CI_km:'data.frame': 77 obs. of  2 variables:
##  $ stdCox : num [1:2] 0.8919 0.0114
##  $ Poisson: num 0.889
```

```
##  $ newCox :List of 10
##   ..$ CI_1      :'data.frame':   247 obs. of  3 variables:
##   ..$ VE_CI_1   :'data.frame':   239 obs. of  2 variables:
##   ..$ CI_2      :'data.frame':   226 obs. of  3 variables:
##   ..$ VE_CI_2   :'data.frame':   201 obs. of  2 variables:
##   ..$ CI_3      :'data.frame':   176 obs. of  3 variables:
##   ..$ VE_CI_3   :'data.frame':   148 obs. of  2 variables:
##   ..$ CI_4      :'data.frame':   116 obs. of  3 variables:
##   ..$ VE_CI_4   :'data.frame':   87 obs. of  2 variables:
##   ..$ VE_HR     :'data.frame':   3221 obs. of  5 variables:
##   ..$ covariates: num [1:2, 1:7] -0.0109 0.0352 0.0227 0.0603 -0.4804 ...
##   .. ..- attr(*, "dimnames")=List of 2
```

In particular, under the new Cox model, 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 can be obtained by

```
result$newCox$covariates
```

```
##                 coef   se(coef)          z  Pr(>|z|) exp(coef) lower .95
## priority -0.01089040 0.02266749 -0.4804415 0.6309135 0.9891687 0.9461836
## sex       0.03520544 0.06031170  0.5837248 0.5594054 1.0358325 0.9203461
##          upper .95
## priority  1.034107
## sex       1.165810
```

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