DOVE3 – <u>D</u>urability <u>O</u>f <u>V</u>accine <u>E</u>fficacy 3

September 21, 2022

Introduction

DOVE3 is an R package for estimating the potentially waning long-term effectiveness of vaccination and prior infection against COVID-19 outcomes in observational studies (Lin et al., 2022). Effects of all exposures (i.e., vaccination and prior infection) are estimated simultaneously under a single Cox regression model, allowing the outcome of interest to be a recurrent event.

DOVE3 inputs a rectangular data set with the following information:

- **Subject ID**: Number that identifies subjects.

- **Entry time**: Calendar time when the subject enters the risk set.

- **Event Time**: Calendar time when the subject experiences the clinical outcome of interest (e.g., SARS-CoV-2 Infection, hospitalization, or death caused by infection).

- **Censoring Time**: Calendar time when the subject moves out of the risk set.

- **Vaccination Time**: Calendar time when vaccination takes place.

- **Vaccination Type**: Categorical variable indicating which vaccine type the subject receives.

- **Infection Time**: Calendar time when the subject is infected.

- **Infection Type**: Categorical variable indicating which dominant variant the infection is associated with.

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

Of note, both numerical and categorical covariates can be included, and all of the time variables are measured from the same time origin and are specified in units of whole days.

The primary analysis tool of the package is *dove3()*, which returns the estimated hazard/rate ratio for each baseline covariate and the estimated effectiveness of each exposure in reducing the instantaneous risk of the clinical outcome of interest. The standard errors and 95% confidence intervals are also provided.

In addition, the package includes three convenience functions: *outcome()*, which is used to specify the subject id, entry time, event time, and censoring time in the model statement of *dove3()*; and *exposure()*, which is used to specify the vaccination time, vaccine type, infection time, and infection type in the model statement of *dove3()*.

Functions

outcome()

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: subject id, entry time, event time, and censoring 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

```
outcome(subject_id, entry_time, event_time, censor_time)
```

where `subject.id` is the number that identifies subjects; `entry.time` is the calendar time when the subject enters the risk set; `event.time` is the calendar time when the subject experiences the clinical outcome of interest (NA, Inf, or an arbitrary value greater than the censoring time if the subject does not experience an event); `censor.time` is the calendar time when the subject moves out of the risk set. Note that all times must be provided in units of whole days.

exposure()

This convenience function is used as the right-hand side of a formula object for the sole purpose of simplifying the specification of required input variables: vaccination time, vaccine type, infection time, and infection type. 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

```
exposure(vaccine_time, vaccine_type, infection_time, infection_type)
```

where `vaccine.time` is the calendar time when vaccination takes place (NA, Inf, or an arbitrary value greater than the study end time if the subject is not vaccinated during the study period); `vaccine.type` is the categorical variable indicating which vaccine type the subject receives (an integer between 1 and the total number of vaccine types, with NA or an arbitrary value within this range if the subject is not vaccinated during the study period); `infection.time` is the calendar time when the subject is infected (NA, Inf, or an arbitrary value that is larger than the study end time if the subject is not infected during the study period); `infection.type` is the categorical variable indicating which dominant variant the infection is associated with (an integer between 1 and the total number of variants under investigation, with NA or an arbitrary value within this range if the subject is not infected during the study period). Note that all times must be provided in units of whole days.

dove3()

This function estimates effectiveness of vaccination and prior infection under the assumption that the log rate/hazard ratio for the effect of vaccine and prior infection is a piecewise linear function of time. The value object returned contains the estimated hazard/rate ratio for each baseline covariate and the estimated effectiveness of vaccine and prior infection in reducing the hazard rate, $EE_h(t)$, where t is the time elapsed since vaccination or infection. The 95% confidence intervals for the estimated effectiveness are provided.

The function call takes the following form:

```
dove3(formula,
      data,
      vaccine_infection_interaction = FALSE,
      vaccine_knots = NULL,
      vaccine_uninfected_knots = NULL,
      vaccine_infected_knots = NULL,
      prior_infection_knots = NULL,
      related_vaccine_types = NULL,
      last_piece_constant = FALSE,
      reinfection_cutoff = 14,
      plots = TRUE
)
```

where

- `formula` is a model statement. See below for further details.

- `data` is a data.frame object containing all required data as previously described.

- `vaccine_infection_interaction` is a logical object specifying the interaction between vaccination and prior infection. If TRUE, vaccine effects are allowed to be different between previously uninfected subjects and previously infected subjects; otherwise, average vaccine effects among all subjects are estimated, regardless of the prior infection status.

- `vaccine_knots` is a list object of which the kth element specifies the knots of the piecewise linear function for the log rate/hazard ratio of the kth vaccine type.

- `vaccine_uninfected_knots` is a list object of which the kth element specifies the knots of the piecewise linear function for the log rate/hazard ratio of the kth vaccine type given no prior infection before vaccination.

- `vaccine_infected_knots` is a list object of which the kth element specifies the knots of the piecewise linear function for the log rate/hazard ratio of the kth vaccine type given at least one prior infection before vaccination.

- `prior_infection_knots` is a list object of which the kth element specifies the knots of the piecewise linear function for the log rate/hazard ratio of the kth infection type. The first knot should be placed at reinfection_cutoff.

- `related_vaccine_types` is a list object of which the element takes the form c(i,j) and imposes a constraint that the slope of the first piece of the piecewise linear function is the same between the ith and jth vaccine types.

- `last_piece_constant` is a logical object specifying the effectiveness trend after the last change point. If TRUE, effectiveness is assumed to be constant after the last change point; otherwise, effectiveness is allowed to vary after the last change point.

- `reinfection_cutoff` is a positive scalar object specifying the gap time (in days) that defines distinct infections. If the gap time of two infections is larger than reinfection_cutoff, they considered as two different infections.

- `plots` is a logical object indicating whether graphical forms of the estimated effectiveness curves are to be generated.

If vaccine_knots or vaccine_uninfected_knots = NULL, knots will be placed at every month by default. If vaccine_infected_knots = NULL, knots will be placed at every other month by default. The input of vaccine_knots will be ignored when vaccine_infection_interaction = TRUE. The input of vaccine_uninfected_knots and vaccine_infected_knots will be ignored if vaccine_infection_interaction = FALSE. If prior_infection_knots = NULL, a default set of knots will be used, with the first knot placed at reinfection_cutoff and the subsequent knots placed at every three months.

The model statement is a formula object. The left side is an object returned by the *outcome()* function. The right-hand side is a combination of baseline covariates and the previously described *exposure()* function. Categorical baseline covariates can be specified, and all other categories are compared to the first category.

The `formula` input takes the following general structure

```
outcome(subject_id, entry_time, event_time, censor_time) ~ covariates +
    exposure(vaccine_time, vaccine_type, infection_time, infection_type)
```

where 'subject_id', 'entry_time', 'event_time', 'censor_time', 'covariates', 'vaccine_time', 'vaccine_type', 'infection_time', and 'infection_type' 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.

Examples

To illustrate the call structure and results of *dove3()*, we use the dataset provided with the package, exampleData. This dataset was simulated to mimics the surveillance data of vaccination and infection and contains the following observations for each of the 8,000 records:

- **subject.id**: The subject ID of each record.

- **event.time**: The event (infection, hospitalization or death) time in days. Infection is treated as recurrent event.

- **censor.time**: The censoring time for each subject in days.

- **entry.time**: The entry time in days.

- **Vtime**: The time of vaccination/booster in days.

- **Vtype**: The type of vaccination/booster.

- **infection.time**: The time of infection in days. If the outcome is infection, infection.time is the same as event.time.

- **infection.type**: The type of infection.

- **age**: A categorical variable of age (<18, 18-34, 35-49, 50-64, >=65).

- **gender**: A binary indicator of gender (1 for male and 0 for female).

- **priority**: A composite baseline risk score taking values 1-3.

The data can be loaded in the usual way

```
data(exampleData)
```

```
head(exampleData)
```

```
##   subject.id event.time censor.time entry.time Vtime Vtype infection.time
## 1          1        180         272        122   273     1            180
## 2          1        229         272        122   273     1            229
## 3          2        224         272        131   273     1            224
## 4          2        273         272        131   273     1            273
## 5          3        169         272        154   273     1            169
## 6          3        210         272        154   273     1            210
##   infection.type age gender priority
## 1              2 <18      1        1
## 2              2 <18      1        1
## 3              2 <18      1        2
## 4              2 <18      1        2
## 5              1 <18      1        1
## 6              1 <18      1        1
```

Consider the summary statistics

```
summary(exampleData)
```

```
##    subject.id      event.time     censor.time      entry.time
##  Min.   :   1.0   Min.   :123.0   Min.   :202.0   Min.   :119.0
##  1st Qu.: 997.8   1st Qu.:224.0   1st Qu.:272.0   1st Qu.:127.0
##  Median :1995.5   Median :255.0   Median :272.0   Median :132.0
##  Mean   :1995.7   Mean   :244.6   Mean   :271.9   Mean   :147.2
##  3rd Qu.:2994.0   3rd Qu.:273.0   3rd Qu.:272.0   3rd Qu.:159.0
```

```
##  Max.   :3992.0   Max.   :273.0   Max.   :272.0   Max.   :271.0
##     Vtime            Vtype        infection.time  infection.type
##  Min.   :141.0   Min.   :1.000   Min.   :123.0   Min.   :1.000
##  1st Qu.:273.0   1st Qu.:1.000   1st Qu.:224.0   1st Qu.:1.000
##  Median :273.0   Median :1.000   Median :255.0   Median :1.000
##  Mean   :259.6   Mean   :1.177   Mean   :244.6   Mean   :1.495
##  3rd Qu.:273.0   3rd Qu.:1.000   3rd Qu.:273.0   3rd Qu.:2.000
##  Max.   :273.0   Max.   :3.000   Max.   :273.0   Max.   :2.000
##      age              gender          priority
##  Length:8000      Min.   :0.0000   Min.   :1.00
##  Class :character 1st Qu.:0.0000   1st Qu.:2.00
##  Mode  :character Median :0.0000   Median :2.00
##                   Mean   :0.4349   Mean   :1.87
##                   3rd Qu.:1.0000   3rd Qu.:2.00
##                   Max.   :1.0000   Max.   :3.00
```

We can see that participants were enrolled in the study over a 5-month period ($119 \leq$ entry.time ≤ 271 days), the follow-up time ended on day 272 (censor.time ≤ 272 days). In this analysis, we will include in our model statement all three baseline covariates, age, gender and priority.

In the first example, we estimated the effectiveness of the 3 types of vaccination and the 2 types of the prior infection without interaction between them. The knots are placed as follows:

```
vaccine.knots = list("vac.type1" = c(30),
                     "vac.type2" = c(30,60),
                     "vac.type3" = c(30,60))
prior.infection.knots = list("inf.type1" = c(14),
                             "inf.type2" = c(14))
```

The function call takes the following form

```
formula = outcome(subject.id, entry.time, event.time, censor.time) ~
  age + gender + priority +
  exposure(Vtime, Vtype, infection.time, infection.type)
data = exampleData

result1 = dove3(formula = formula,
                data = data,
                vaccine_knots = vaccine.knots,
                prior_infection_knots = prior.infection.knots)
```

The function returns a list object with the following information.

Covariate Effects: The estimated (log) hazard/rate 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
## `age>=65`   -0.16351695 0.11596842 -1.410013 0.15853594 0.8491521 0.6765081
## `age18-34`  -0.12319040 0.08155531 -1.510513 0.13091246 0.8840953 0.7534928
## `age35-49`  -0.18112448 0.08285777 -2.185968 0.02881791 0.8343315 0.7092674
## `age50-64`  -0.27729728 0.08464601 -3.275964 0.00105302 0.7578292 0.6419785
## gender      -0.04074722 0.03003144 -1.356819 0.17483872 0.9600718 0.9051924
## priority     0.01306333 0.02524958  0.517368 0.60489928 1.0131490 0.9642305
```

```
##            upper .95
## `age>=65`  1.0658547
## `age18-34` 1.0373352
## `age35-49` 0.9814480
## `age50-64` 0.8945861
## gender     1.0182784
## priority   1.0645494
```

Estimated Effectiveness: Element **$effectiveness** contains the estimated effectiveness for different types of vaccination or prior infection in reducing the attack rate at each observed event time, together with its standard error and the 95% confidence interval. The ith element of result1$effectiveness corresponds to the ith type of vaccination or prior infection.

```
head(result1$effectiveness[[1]])
```

```
##       time         VE         se  lower .95  upper .95
## [1,]    0 0.00000000 0.000000000 0.000000000 0.00000000
## [2,]    1 0.01259900 0.003550150 0.005616134 0.01953283
## [3,]    2 0.02503927 0.007010842 0.011200726 0.03868414
## [4,]    3 0.03732280 0.010383769 0.016753955 0.05746136
## [5,]    4 0.04945157 0.013670592 0.022275996 0.07587181
## [6,]    5 0.06142754 0.016872945 0.027767025 0.09392266
```

Plots: Element **$plots** contains the graphical depictions of the estimated effectiveness that generated by default by *dove3()*. The ith element of result1$plots corresponds to the ith type of vaccination or prior infection, which can be regenerated as follows:
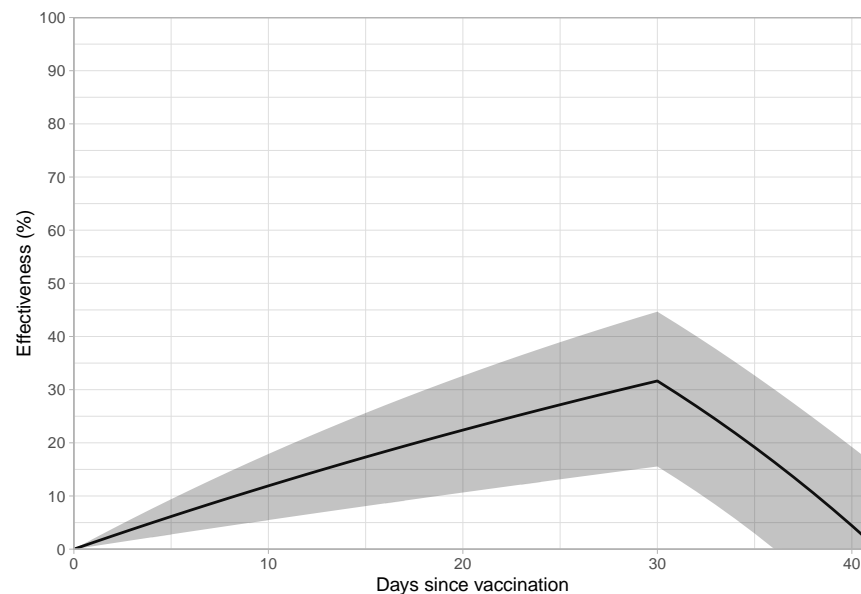
```
result1$plots[[1]]
```


Figure 1: Plots auto-generated by *dove3()*. The estimated effectiveness curve (black line) without interaction assumption and its 95% confidence intervals (grey shadow) are shown as a function of the time since vaccination.

In the second example, we estimated the effectiveness of the 3 types of vaccination and the 2 types of the prior infection with interaction between them. In addition, we assumed the first piece of the piecewise linear function is the same between the first and second vaccine types. The knots are placed as follows:

```
vaccine.uninfected.knots = list("vac.noinf.type1" = c(30),
                                "vac.noinf.type2" = c(30,60),
                                "vac.noinf.type3" = c(30,60))

vaccine.infected.knots = list("vac.noinf.type1" = c(30),
                              "vac.noinf.type2" = c(60),
                              "vac.noinf.type3" = c(60))
```

The function call takes the following form

```
result2 = dove3(formula = formula,
                data = data,
                vaccine_infection_interaction = TRUE,
                vaccine_uninfected_knots = vaccine.uninfected.knots,
                vaccine_infected_knots = vaccine.infected.knots,
                prior_infection_knots = prior.infection.knots,
                related_vaccine_types = list(c(1,2)))
```

The function returns a list object containing the following items.

Covariate Effects: The estimated (log) hazard/rate 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
## `age>=65`   -0.17114078 0.11594639 -1.4760337 0.1399348513 0.8427029 0.6713991
## `age18-34`  -0.12942273 0.08156914 -1.5866629 0.1125890167 0.8786025 0.7487910
## `age35-49`  -0.18808049 0.08288543 -2.2691623 0.0232584578 0.8285480 0.7043126
## `age50-64`  -0.28405610 0.08465038 -3.3556388 0.0007918191 0.7527244 0.6376487
## gender      -0.04078698 0.03001825 -1.3587396 0.1742291242 0.9600336 0.9051798
## priority     0.01236031 0.02526020  0.4893197 0.6246153721 1.0124370 0.9635328
##             upper .95
## `age>=65`   1.0577140
## `age18-34`  1.0309182
## `age35-49`  0.9746976
## `age50-64`  0.8885677
## gender      1.0182115
## priority    1.0638234
```

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

```
head(result2$effectiveness[[1]])
```

```
##      time         VE          se   lower .95  upper .95
## [1,]    0 0.00000000 0.000000000 0.00000000 0.00000000
## [2,]    1 0.01766819 0.003094380 0.01158444 0.02371449
## [3,]    2 0.03502421 0.006079416 0.02303469 0.04686660
## [4,]    3 0.05207359 0.008958006 0.03435228 0.06946967
## [5,]    4 0.06882173 0.011732979 0.04553877 0.09153673
## [6,]    5 0.08527396 0.014407098 0.05659568 0.11308047
```

Plots: Element **$plots** contains the graphical depictions of the estimated effectiveness:
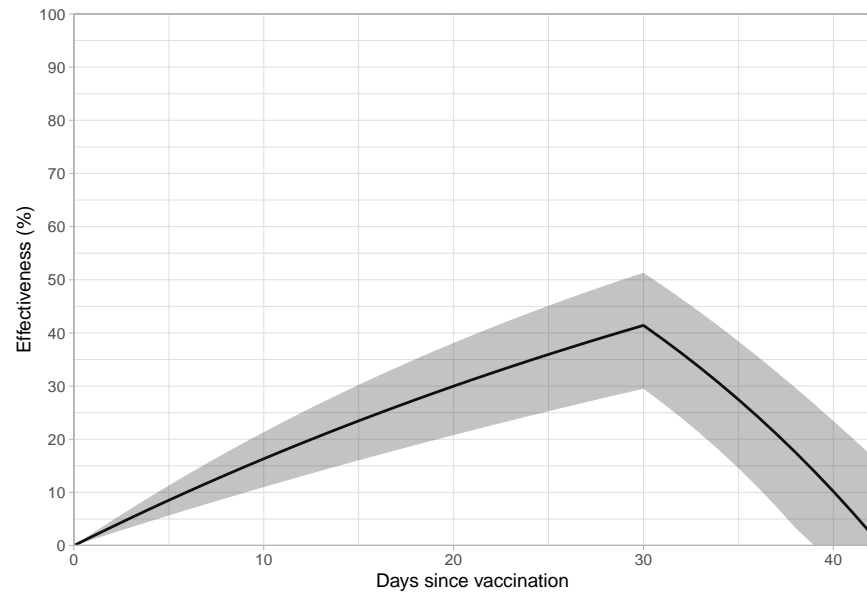
`result2$plots[[1]]`

Figure 2: Plots auto-generated by *dove3()*. The estimated effectiveness curve (black line) under interaction assumption and its 95% confidence intervals (grey shadow) are shown as a function of the time since vaccination.

References

Lin, D-Y, Gu, Y., Xu, Y., et al. Association of Primary and Booster Vaccination and Prior Infection With SARS-CoV-2 Infection and Severe COVID-19 Outcomes. JAMA. Published online September 26, 2022. doi:10.1001/jama.2022.17876