Advanced Methods in Epidemiology

Practical Session on Competing Risks

Prof. dr. Steven Abrams

Master of Epidemiology | Academic year: 2022-2023

Background

In survival analysis, one is interested in studying the time to a specific event, potentially in the presence of censoring (or truncation). Censored observations arise when the event of interest has not yet occurred prior to, for example, the end of the study period (i.e., referred to as administrative right censoring). Other reasons for (right) censoring are individuals that are lost to follow-up before the event time could have been recorded or the occurrence of another event that prevents the one of interest to occur. In the latter case, the other event can be a competing one if individuals experiencing the competing event are different from those that do not with regard to the occurrence of the event of interest. For example, in clinical oncology cancer-related mortality may be of primary interest, but other causes of death can prevent its occurrence and deaths caused by reasons other than cancer are typical examples of competing risks.

Survival analysis in the presence of competing risks poses additional challenges for the correct estimation of the hazard function (i.e., the rate or instantaneous probability of experiencing the event of interest) and the cumulative incidence function (i.e., the risk of experiencing the event up to time t). In standard survival analysis, there exists a direct one-to-one relationship between the cumulative incidence function (CIF) and the hazard function (see below). However, such a relationship does not necessarily exists in a competing risks setting. In this practical session, we demonstrate the use of different R functions to estimate the CIF and to study the effect of covariates on either the cause-specific hazard functions and/or CIF. Both the use of the Cox proportional hazard model for the estimation of covariate effects on the cause-specific hazards as well as the Fine and Gray model for the direct estimation of such effects on the CIF are considered hereunder in more detail.

Key concepts in survival analysis

Survival data can be characterized by means of the so-called hazard function $\lambda(t)$ which provides a dynamic description of the instantaneous risk of failing given survival until time t. Alternatively, one could consider the cumulative hazard function $\Lambda(t)$, which in discrete time, simplifies to the sum of $\lambda(t)$ from time zero up to time t (in a finite number of discrete time steps). The cumulative hazard function is defined as the integral of the hazard function on the interval $[0, t]$ in continuous time. In the absence of competing risks, the specification of the (cumulative) hazard function defines the underlying time-to-event distribution uniquely (for a continuous random variable T^*), implying a one-to-one relation with the CIF or survival function (i.e., 1 - CIF) $S(t)$ defined as:

$$S(t) = \exp[\Lambda(t)] = \exp(-\int_0^t \lambda(u)du). \quad (1)$$

The survival function $S(t)$ can be estimated non-parametrically using the Product-Limit estimator, also and more commonly referred to as the *Kaplan-Meier* estimator. Alternatively, one can rely on the Nelson-Aalen estimator to estimate the cumulative hazard function $\Lambda(t)$.

The cumulative distribution function (or cumulative incidence function in a non-competing risk setting), characterizing the risk of experiencing the event of interest, is defined as

$$\text{CIF}(t) = 1 - S(t) = 1 - \exp[\Lambda(t)] = 1 - \exp(-\int_0^t \lambda(u)du). \tag{2}$$

In the presence of competing risks, however, the CIF cannot be directly linked to the hazard function $\lambda(t)$ anymore. Moreover, the marginal cumulative incidence estimated using the Kaplan-Meier estimator is always larger than the sum of the cause-specific cumulative incidences for the different competing risks. The estimated marginal cumulative incidence, by means of the Kaplan-Meier estimator, represents the probability of having experienced the event of interest at time t in a hypothetical world in which other competing events are absent, at least if censoring induced by competing events is independent from the event time process of interest. The latter assumption, however, is untestable based on the event time data at hand.

Data application

The data we consider to demonstrate competing risks analysis is contained in the *riskRegression* package in R. A total of 205 patients with melanoma were followed after a surgical operation until the end of 1977. The data can be retrieved as follows:

```
# Import the data
#-----------------
#install.packages("riskRegression")
library(riskRegression)
data(Melanoma)
```

1. Study the variables in the dataset.

```
# Study the data
#----------------
str(Melanoma)
```

```
## 'data.frame':    205 obs. of  11 variables:
##  $ time    : int  10 30 35 99 185 204 210 232 232 279 ...
##  $ status  : num  2 2 0 2 1 1 1 1 2 1 ...
##  $ event   : Factor w/ 3 levels "censored","death.malignant.melanoma",..: 3 3 1 3 2 2 2 2 3 2 ..
##  $ invasion: Factor w/ 3 levels "level.0","level.1",..: 2 1 2 1 3 3 3 3 2 1 ...
##  $ ici     : Factor w/ 4 levels "0","1","2","3": 3 1 3 3 3 3 3 3 4 3 ...
##  $ epicel  : Factor w/ 2 levels "not present",..: 2 1 1 1 2 1 2 1 1 1 ...
##  $ ulcer   : Factor w/ 2 levels "not present",..: 2 1 1 1 2 2 2 2 2 2 ...
##  $ thick   : num  6.76 0.65 1.34 2.9 12.08 ...
##  $ sex     : Factor w/ 2 levels "Female","Male": 2 2 2 1 2 2 2 2 1 1 ...
##  $ age     : int  76 56 41 71 52 28 77 49 60 68 ...
##  $ logthick: num  1.911 -0.431 0.293 1.065 2.492 ...
```

```
head(Melanoma)
```

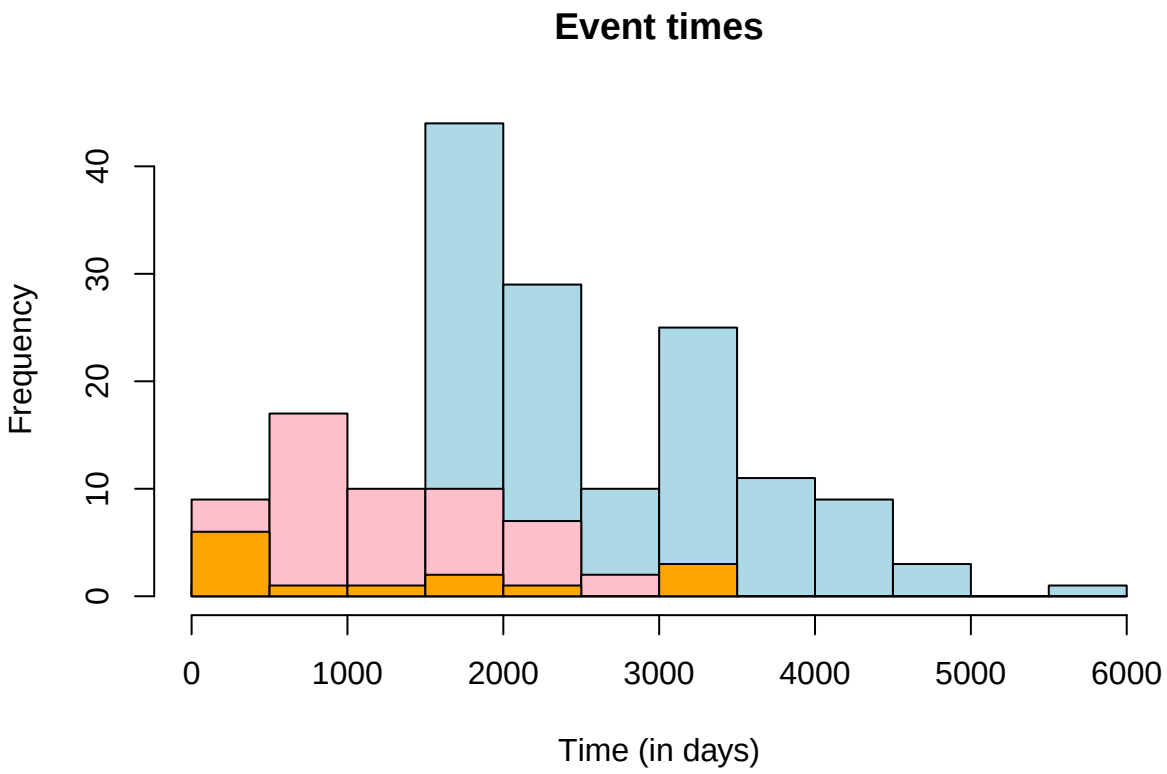
```
##   time status                    event invasion ici      epicel       ulcer
## 1   10      2       death.other.causes  level.1   2     present     present
## 2   30      2       death.other.causes  level.0   0 not present not present
## 3   35      0                 censored  level.1   2 not present not present
## 4   99      2       death.other.causes  level.0   2 not present not present
## 5  185      1 death.malignant.melanoma  level.2   2     present     present
## 6  204      1 death.malignant.melanoma  level.2   2 not present     present
##   thick    sex age   logthick
## 1  6.76   Male  76  1.9110229
```

```
## 2  0.65   Male  56 -0.4307829
## 3  1.34   Male  41  0.2926696
## 4  2.90 Female  71  1.0647107
## 5 12.08   Male  52  2.4915512
## 6  4.84   Male  28  1.5769147
```

```
## There are 11 variables included in the dataset:
## Event time (time) was measured in days from operation (time 0);
## Status represents the censoring indicator with 0 = censored, 1 = died from melanoma
##           and 2 = died from other causes;
## Event is defined in accordance with status but is converted to a factor variable
##           with three levels.
##
## Other variables are risk factors under study including invasion, inflammatory
## cell infiltration (ici), ulcer, thick, sex, age and tumor thickness on the
## natural log scale (logthick).
```

2. Graphically depict the event time distribution for the event of interest (i.e., death as a result of melanoma) and distinguish between censored and uncensored observations.

```
# Histogram of the observed event times
#---------------------------------------
A = Melanoma$time[Melanoma$status == 0]
B = Melanoma$time[Melanoma$status == 1]
C = Melanoma$time[Melanoma$status == 2]
c1 = "lightblue"; c2 = "pink"; c3 = "orange";

b <- min(c(A,B,C)) - 0.001       # Set the minimum for the breakpoints
e <- max(c(A,B,C))               # Set the maximum for the breakpoints
ax <- pretty(b:e, n = 12)        # Make a neat vector for the breakpoints

hgA <- hist(A, breaks = ax, plot = FALSE)  # Save first histogram data
hgB <- hist(B, breaks = ax, plot = FALSE)  # Save second histogram data
hgC <- hist(C, breaks = ax, plot = FALSE)  # Save second histogram data

par(mfrow = c(1,1))
plot(hgA, col = c1, main = "Event times", xlab = "Time (in days)")
plot(hgB, col = c2, add = TRUE)
plot(hgC, col = c3, add = TRUE)
```

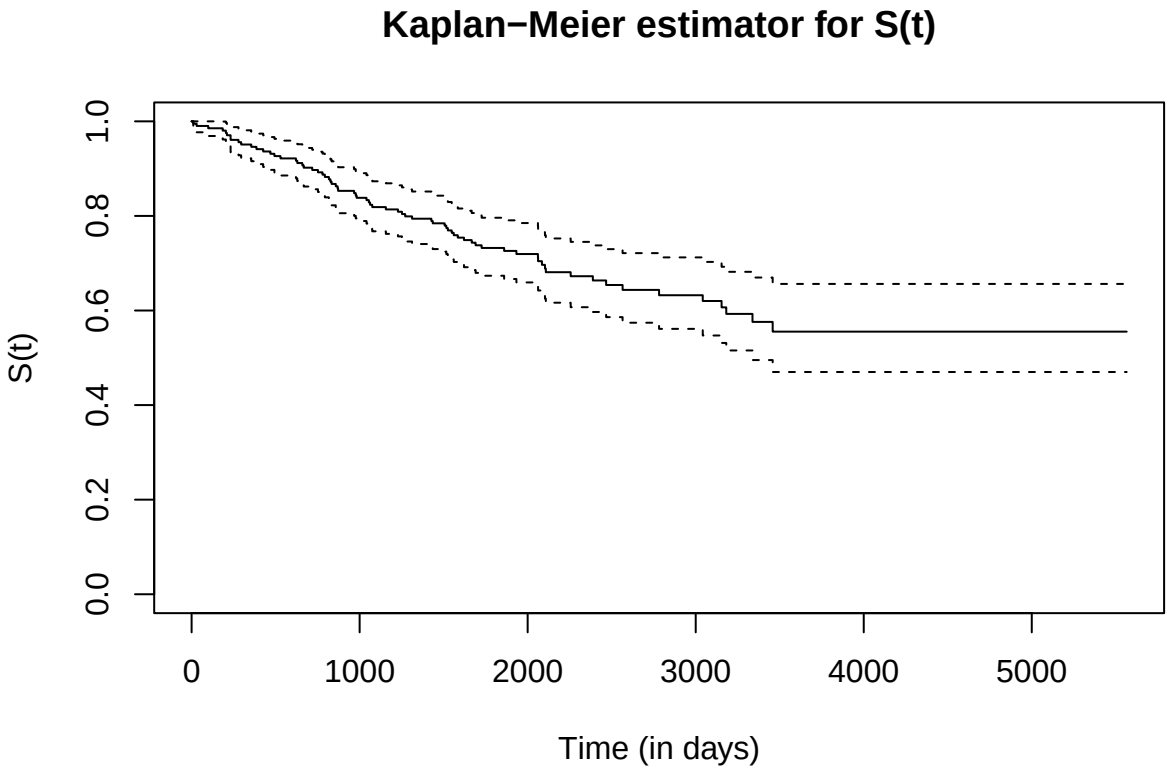


Competing risks analysis - Combined analysis

A first approach towards analysing competing risks survival data is to combine all events. This combined analysis provides insights into the overall hazard function.

1. Estimate the overall cumulative incidence function using the Kaplan-Meier estimator for the survival function.

```
# Estimation of the overall cumulative incidence function
#-------------------------------------------------------
library(survival)
surv_object <- Surv(time = Melanoma$time,
                    event = as.numeric(Melanoma$status != 0))

plot(surv_object, main = "Kaplan-Meier estimator for S(t)",
     xlab = "Time (in days)", ylab = "S(t)")
```

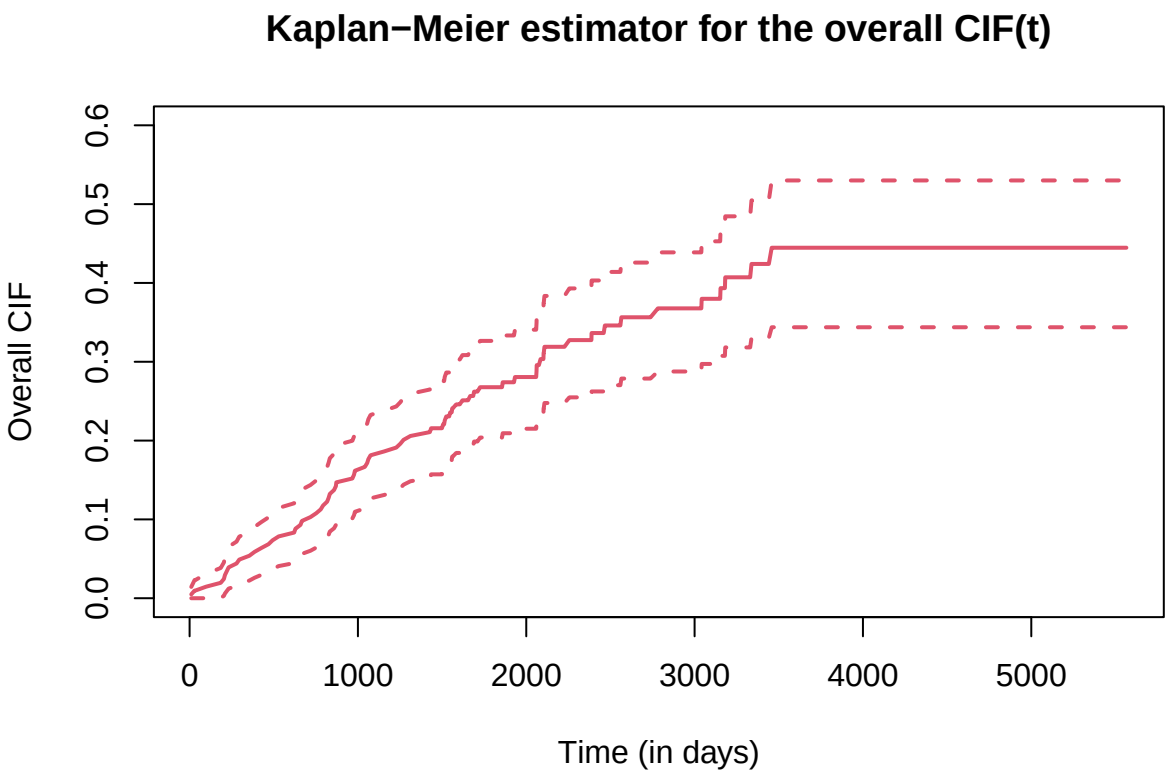



```
fit_combined <- survfit(surv_object ~ 1, data = Melanoma)
head(summary(fit_combined))
```

```
## $n
## [1] 205
##
## $time
##  [1]   10   30   99  185  204  210  232  279  295  355  386  426  469  493  529
## [16]  621  629  659  667  718  752  779  793  817  826  833  858  869  872  967
## [31]  977  982 1041 1055 1062 1075 1156 1228 1252 1271 1312 1427 1435 1506 1516
## [46] 1525 1548 1560 1584 1621 1667 1690 1726 1860 1933 2061 2062 2085 2103 2108
## [61] 2256 2388 2467 2565 2782 3042 3154 3182 3338 3458
##
## $n.risk
##  [1] 205 204 202 201 200 199 198 196 195 194 193 192 191 190 189 188 187 186 185
## [20] 184 183 182 181 180 179 178 177 176 175 174 173 172 171 170 169 168 167 166
## [39] 165 164 163 162 161 159 155 154 152 150 148 146 137 134 131 117 110  95  94
## [58]  92  90  88  80  75  69  63  57  52  46  44  35  28
##
## $n.event
##  [1] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##
## $n.censor
##  [1]  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [26]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  3  0  1  1  1  1
## [51]  8  2  2 13  6 14  0  1  1  1  7  4  5  5  5  4  5  1  8  6
##
## $surv
##  [1] 0.9951220 0.9902439 0.9853417 0.9804395 0.9755373 0.9706351 0.9608307
```

```
##  [8] 0.9559285 0.9510263 0.9461241 0.9412219 0.9363197 0.9314175 0.9265153
## [15] 0.9216131 0.9167109 0.9118087 0.9069065 0.9020043 0.8971021 0.8922000
## [22] 0.8872978 0.8823956 0.8774934 0.8725912 0.8676890 0.8627868 0.8578846
## [29] 0.8529824 0.8480802 0.8431780 0.8382758 0.8333736 0.8284714 0.8235692
## [36] 0.8186670 0.8137648 0.8088626 0.8039604 0.7990582 0.7941560 0.7892538
## [43] 0.7843516 0.7794186 0.7743901 0.7693616 0.7643000 0.7592046 0.7540749
## [50] 0.7489100 0.7434435 0.7378954 0.7322626 0.7260040 0.7194039 0.7118312
## [57] 0.7042586 0.6966036 0.6888636 0.6810356 0.6725226 0.6635556 0.6539389
## [64] 0.6435589 0.6322684 0.6201094 0.6066288 0.5928417 0.5759034 0.5553354
```

```r
plot(fit_combined$time, 1-fit_combined$surv, lwd = 2, col = 2, type = "l",
     main = "Kaplan-Meier estimator for the overall CIF(t)",
     xlab = "Time (in days)", ylab = "Overall CIF", ylim = c(0, 0.6))
lines(fit_combined$time, 1-fit_combined$lower, lwd = 2, col = 2, lty = 2)
lines(fit_combined$time, 1-fit_combined$upper, lwd = 2, col = 2, lty = 2)
```



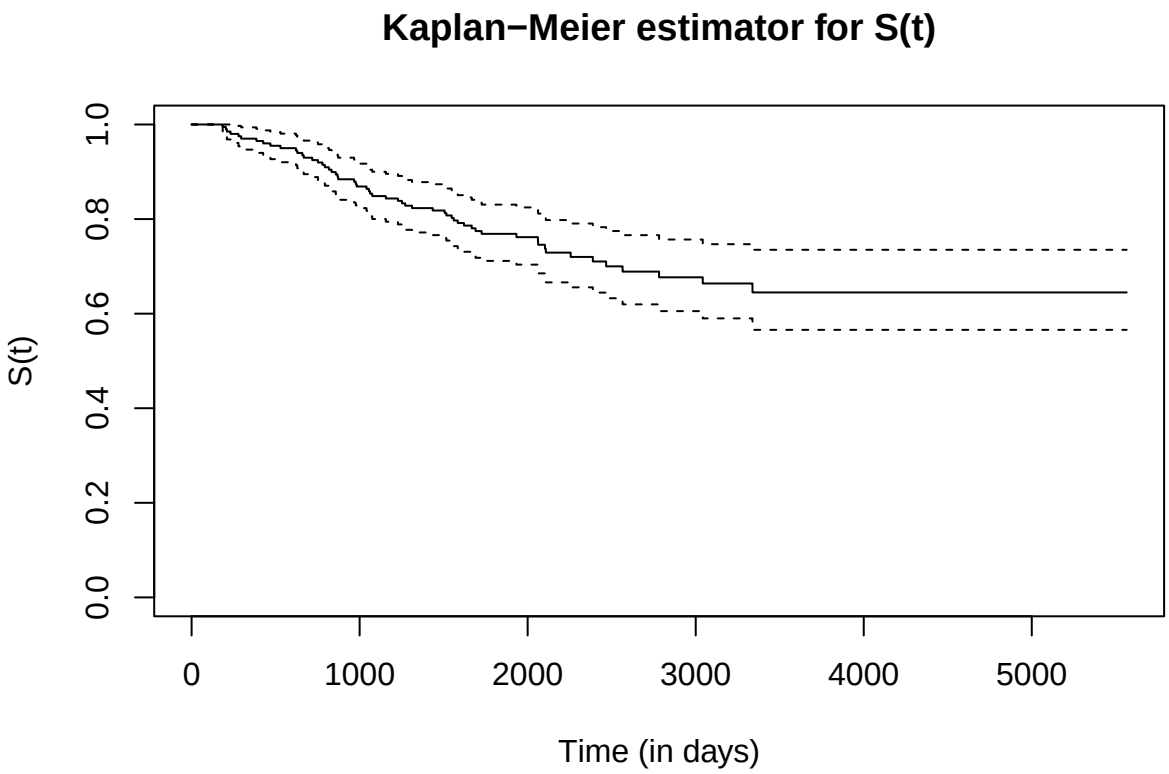
2. Interpret this result. What is the disadvantage of this approach?

```r
## The overall cumulative incidence function provides an estimate of the combined
## risk. However, the disadvantage of this approach is that the event-specific
## risks can not be derived. Furthermore, the effect of covariates with respect
## to the overall hazard function can not be used to assess the cause-specific
## effects.
```

Non-parametric estimation of the (cause-specific) hazard function

1. Estimate the marginal cumulative distribution function (CIF). More specifically, ignore the competing risk and consider such observations as being (right-)censored for the event of interest.

```r
# Kaplan-Meier estimator for the marginal CIF
#--------------------------------------------
library(survival)
surv_object <- Surv(time = Melanoma$time,
                    event = as.numeric(Melanoma$status == 1))

plot(surv_object, main = "Kaplan-Meier estimator for S(t)",
     xlab = "Time (in days)", ylab = "S(t)")
```

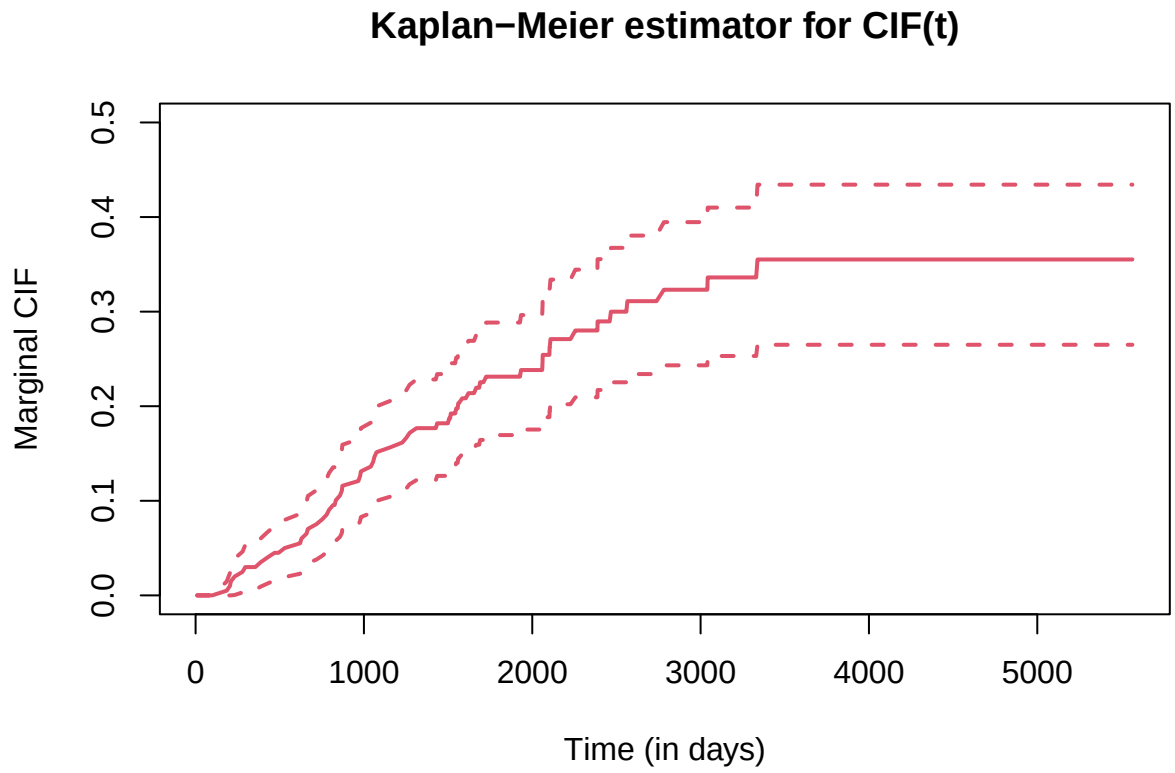



```r
fit1 <- survfit(surv_object ~ 1, data = Melanoma)
head(summary(fit1))
```

```
## $n
## [1] 205
##
## $time
##  [1]  185  204  210  232  279  295  386  426  469  529  621  629  659  667  718
## [16]  752  779  793  817  833  858  869  872  967  977  982 1041 1055 1062 1075
## [31] 1156 1228 1252 1271 1312 1435 1506 1516 1548 1560 1584 1621 1667 1690 1726
## [46] 1933 2061 2062 2103 2108 2256 2388 2467 2565 2782 3042 3338
##
## $n.risk
##  [1] 201 200 199 198 196 195 193 192 191 189 188 187 186 185 184 183 182 181 180
## [20] 178 177 176 175 174 173 172 171 170 169 168 167 166 165 164 163 161 159 155
## [39] 152 150 148 146 137 134 131 110  95  94  90  88  80  75  69  63  57  52  35
##
## $n.event
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
##
## $n.censor
##  [1]  4  0  0  1  0  0  1  0  0  1  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0
## [26]  0  0  0  0  0  0  0  0  0  0  1  1  3  2  1  1  1  8  2  2 20 14  0  3  1
## [51]  7  4  5  5  5  4 16
##
## $surv
##  [1] 0.9950249 0.9900498 0.9850746 0.9800995 0.9750990 0.9700985 0.9650721
##  [8] 0.9600457 0.9550192 0.9499662 0.9449132 0.9398602 0.9348072 0.9297542
## [15] 0.9247012 0.9196482 0.9145951 0.9095421 0.9044891 0.8994077 0.8943263
## [22] 0.8892449 0.8841635 0.8790821 0.8740007 0.8689193 0.8638379 0.8587565
## [29] 0.8536751 0.8485937 0.8435123 0.8384309 0.8333495 0.8282681 0.8231867
## [36] 0.8180738 0.8129286 0.8076839 0.8023702 0.7970211 0.7916358 0.7862137
## [43] 0.7804749 0.7746504 0.7687371 0.7617486 0.7537301 0.7457117 0.7374261
## [50] 0.7290462 0.7199331 0.7103340 0.7000393 0.6889276 0.6768411 0.6638250
## [57] 0.6448585
```

```r
plot(fit1$time, 1-fit1$surv, lwd = 2, col = 2, type = "l",
     main = "Kaplan-Meier estimator for CIF(t)",
     xlab = "Time (in days)", ylab = "Marginal CIF", ylim = c(0,0.5))
lines(fit1$time, 1-fit1$lower, lwd = 2, col = 2, lty = 2)
lines(fit1$time, 1-fit1$upper, lwd = 2, col = 2, lty = 2)
```



2. How do you interpret the aforementioned result?

```r
## The marginal CIF provides an unbiased estimate for the risk of dying from
## melanoma (prior to time t), at least under the assumption of independent
## censoring as a result of the competing risks, in the hypothetical world in
## which the competing risk does not exist. However, in case the occurrence
## of the competing risks is associated with the occurrence of the event of
## interest, the aforementioned assumption is violated. For example, if the
```

```
## (instantaneous) risk of experiencing the event of interest is different for
## individuals experiencing a competing event as compared to those that do not,
## the independence assumption is untenable. More specifically, if 'frail'
## individuals that are at high risk of dying from melanoma, are also at increased
## risk of experiencing the competing event (i.e., dying from another cause), the
## marginal CIF is unreliable and provides a biased estimate of the risk of
## dying from melanoma.
```

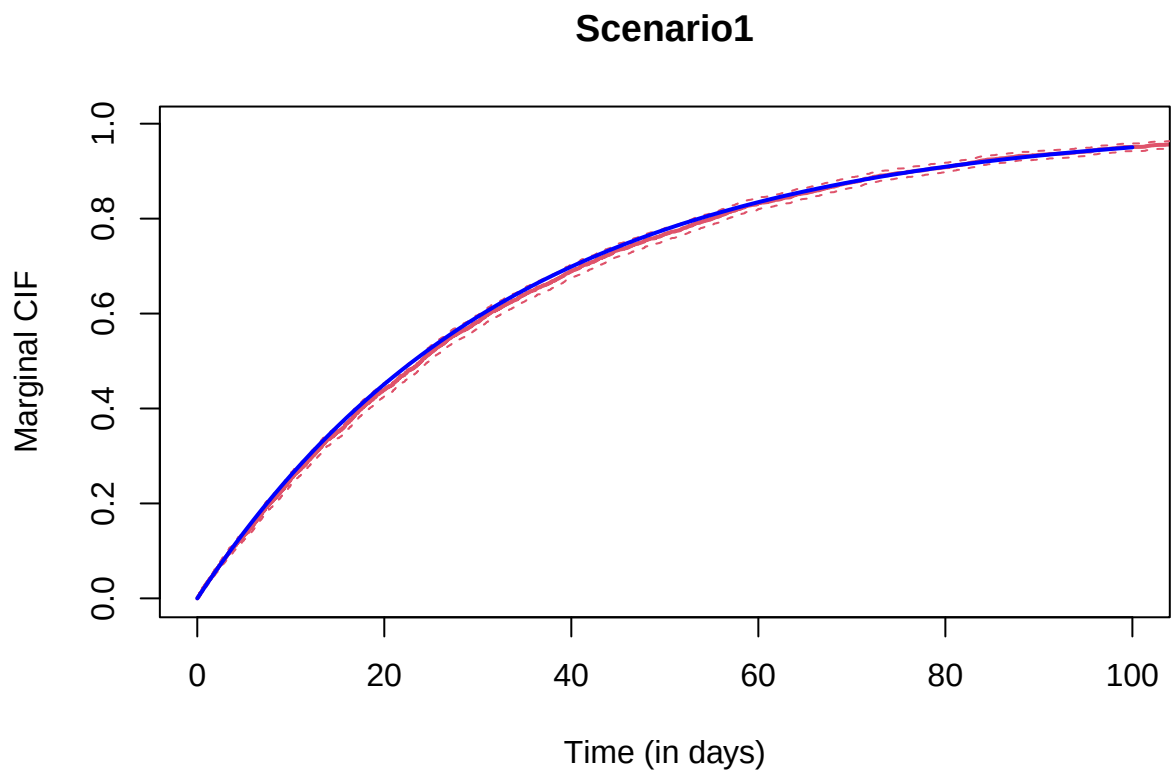
ADDITIONAL INFORMATION: Hereunder we demonstrate that bias is introduced in the estimation of the marginal CIF in case of dependence between event times T_1 and T_2. In the simulation approach considered below, we impose association using a copula approach (i.e., a copula is a function that can be used to join two marginal distributions into a joint bivariate distribution for two random variables) in which a single copula parameter measures the strength of association (theta) between the two random variables. The marginal distributions for T_1 and T_2 are assumed to be exponential with exponential rates $lambda_1$ and $lambda_2$. Independent censoring (exponential process with parameter cparam) is generated, for example, induced by administrative censoring. We consider different scenarios and investigate the impact on the estimated marginal CIF.

- In *scenario 1*, we assume that the rate of occurrence of event 1 is larger than the rate of occurrence of event 2 (the competing risk).
- In *scenario 2*, we assume that the rate of occurrence of event 1 is smaller than the rate of occurrence of event 2 (the competing risk).

```
library(copula)

# Generating dependent competing risks (T1, T2) and independent censoring time C
#-------------------------------------------------------------------------------
clayton.cmprsk.simul <- function(sample_size, hazard = "EXP",
                                          param1, param2, theta_parm,
                                          cens_hazard = "EXP", cparam, seednr){
  set.seed(seednr);
  random_samples <- rCopula(copula=claytonCopula(theta_parm, dim = 2), n=sample_size)
  u = random_samples[,1]; v = random_samples[,2];

  if (hazard == "EXP"){
    # message("Data generation: param-vector should include rate parameter");
    t1 = -log(1-u)/param1[1];
    t2 = -log(1-v)/param2[1];
  }

  if (hazard == "WEIBULL"){
    # message("Data generation: param-vector should include shape and decay parameters,
    # resp.");
    t1 = (-log(1-u)/(param1[2]))**(1/param1[1]);
    t2 = (-log(1-v)/(param2[2]))**(1/param2[1]);
  }

  if (cens_hazard == "EXP"){
    w = runif(sample_size, 0, 1);
    ct = -log(w)/cparam[1];
    status = rep(0, sample_size)
    for (i in 1:sample_size){
      if (t1[i] <= t2[i]){
        if (t1[i] <= ct[i]){
          status[i] = 1
        }
      }
      if (t1[i] > t2[i]){
        if (t2[i] <= ct[i]){
          status[i] = 2
        }
      }
```

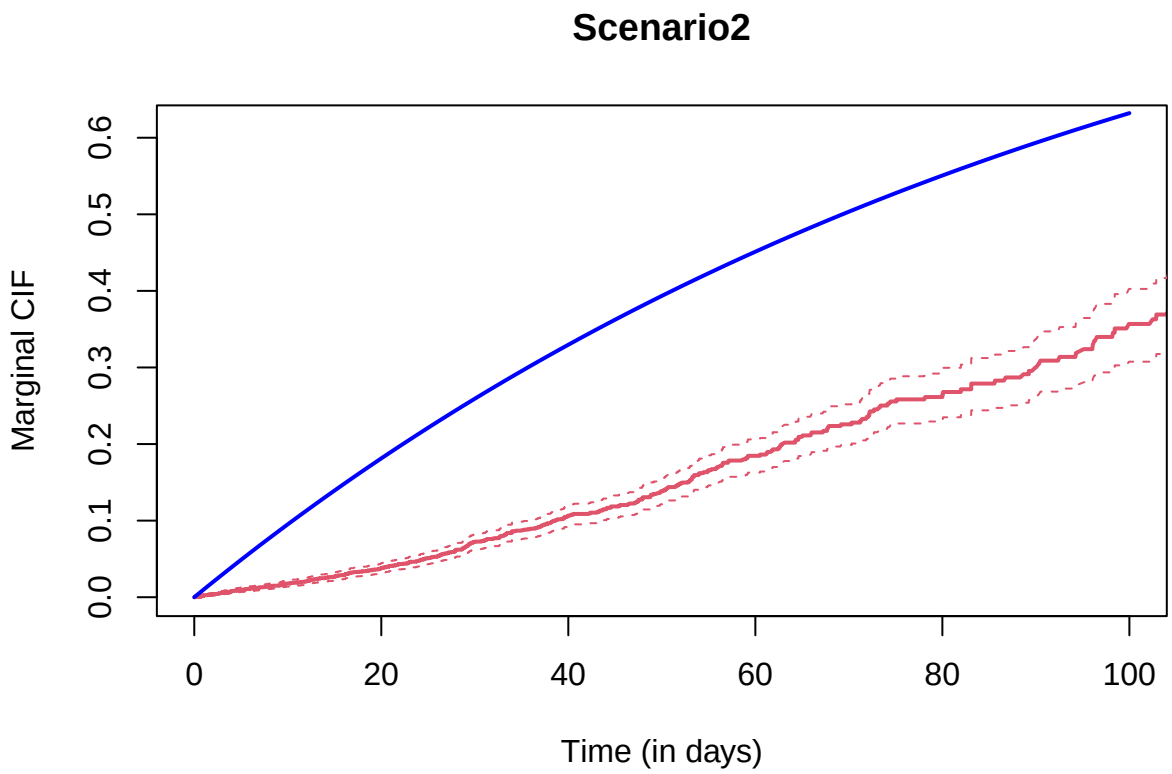
```
    }
    t = pmin(pmin(t1, t2), ct);
  }

  if (cens_hazard == "WEIBULL"){
    w = runif(sample_size, 0, 1)
    ct = (-log(w)/(cparam[2]))**(1/cparam[1]);
    status = rep(0, sample_size)
    for (i in 1:sample_size){
      if (t1[i] <= t2[i]){
        if (t1[i] <= ct[i]){
          status[i] = 1
        }
      }
      if (t1[i] > t2[i]){
        if (t2[i] <= ct[i]){
          status[i] = 2
        }
      }
    }
    t = pmin(pmin(t1, t2), ct);
  }

  simul.data = data.frame(t1 = t1, t2 = t2, ct = ct, t = t, status)
  return(simul.data)
}

# Scenario 1: T1 preceeds T2 most frequently
#-------------------------------------------
dat = clayton.cmprsk.simul(sample_size = 5000, hazard = "EXP",
                           param1 = 0.03, param2 = 0.01,
                           theta_parm = 2,
                           cens_hazard = "EXP",
                           cparam = 0.005,
                           seednr = 12345)
# head(dat)

# True marginal CIF
#------------------
lambda = 0.03
grid = seq(0,100,1)
true_cif = 1 - exp(-lambda*grid)

# Estimate the marginal CIF
#--------------------------
# library(survival)
surv_object <- Surv(time = dat$t,
                    event = as.numeric(dat$status == 1))

# plot(surv_object, main = "Kaplan-Meier estimator for S(t)",
#      xlab = "Time (in days)", ylab = "S(t)")

fit_sim1 <- survfit(surv_object ~ 1)
```

```r
par(mfrow = c(1,1))
plot(fit_sim1$time, 1-fit_sim1$surv, lwd = 2, col = 2, type = "l",
     main = "Scenario1",
     xlab = "Time (in days)", ylab = "Marginal CIF",
     xlim = c(0, 100))
lines(fit_sim1$time, 1-fit_sim1$lower, lty = 2, col = 2)
lines(fit_sim1$time, 1-fit_sim1$upper, lty = 2, col = 2)
lines(grid, true_cif, lwd = 2, col = "blue")
```




```r
## The estimated marginal CIF is close to the true marginal CIF (blue) due to
## the fact that P(T2 < T1) is small. Consequently, the event of interest T1
## is almost always directly observed and not preceded by the competing risk.

# Scenario 2: T2 preceeds T1 most frequently
#-------------------------------------------
dat = clayton.cmprsk.simul(sample_size = 5000, hazard = "EXP",
                           param1 = 0.01, param2 = 0.03,
                           theta_parm = 2,
                           cens_hazard = "EXP",
                           cparam = 0.005,
                           seednr = 12345)
# head(dat)

# True marginal CIF
#------------------
lambda = 0.01
grid = seq(0,100,1)
true_cif = 1 - exp(-lambda*grid)
```

```r
# Estimate the marginal CIF
#--------------------------
# library(survival)
surv_object <- Surv(time = dat$t,
                    event = as.numeric(dat$status == 1))

# plot(surv_object, main = "Kaplan-Meier estimator for S(t)",
#      xlab = "Time (in days)", ylab = "S(t)")

fit_sim2 <- survfit(surv_object ~ 1)

plot(fit_sim2$time, 1-fit_sim2$surv, lwd = 2, col = 2, type = "l",
     main = "Scenario2",
     xlab = "Time (in days)", ylab = "Marginal CIF",
     xlim = c(0, 100))
lines(fit_sim2$time, 1-fit_sim2$lower, lty = 2, col = 2)
lines(fit_sim2$time, 1-fit_sim2$upper, lty = 2, col = 2)
lines(grid, true_cif, lwd = 2, col = "blue")
```

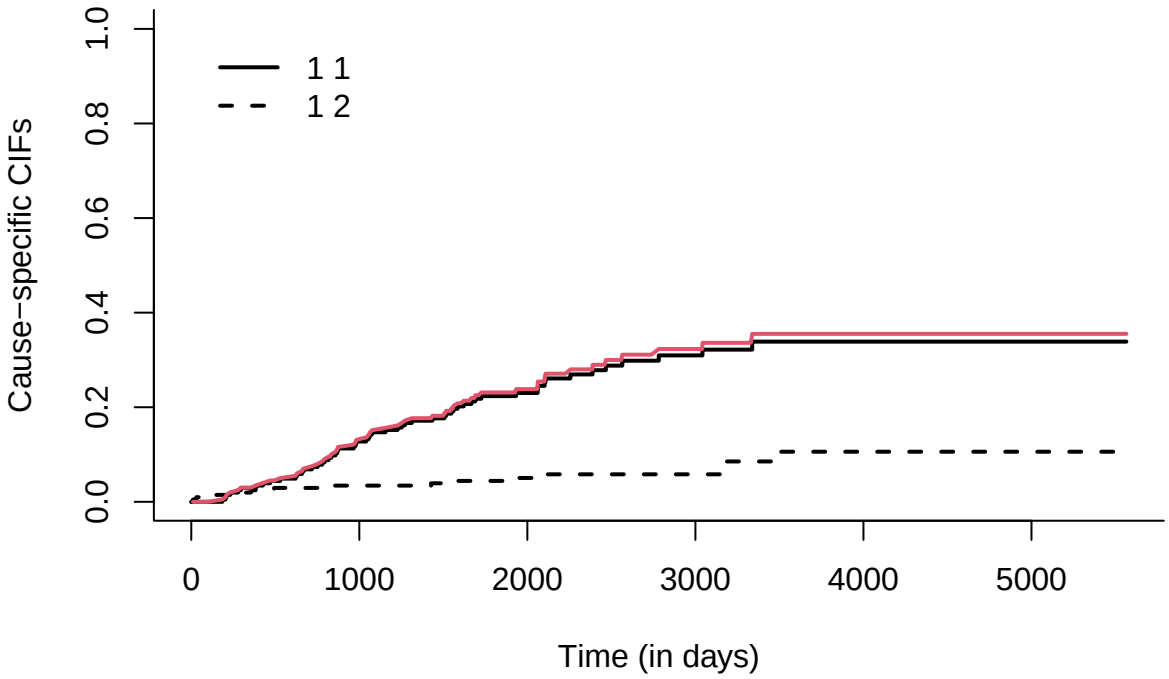



```r
## The estimated marginal CIF provides a biased estimate of the true underlying
## marginal CIF (blue curve) due to the presence of the dependent competing risk.

# Scenario 3: Independence between T1 and T2
#-------------------------------------------

## How would the figure look like in case of independence between T1 and T2?
```

3. The cumulative incidence functions (CIFs) for different causes, i.e., cause-specific CIFs, are used for the statistical description of survival data with competing risks. In order to do so, the Kaplan-Meier (KM) estimator can be used in case there are no competing risks OR when competing risks are assumed to be

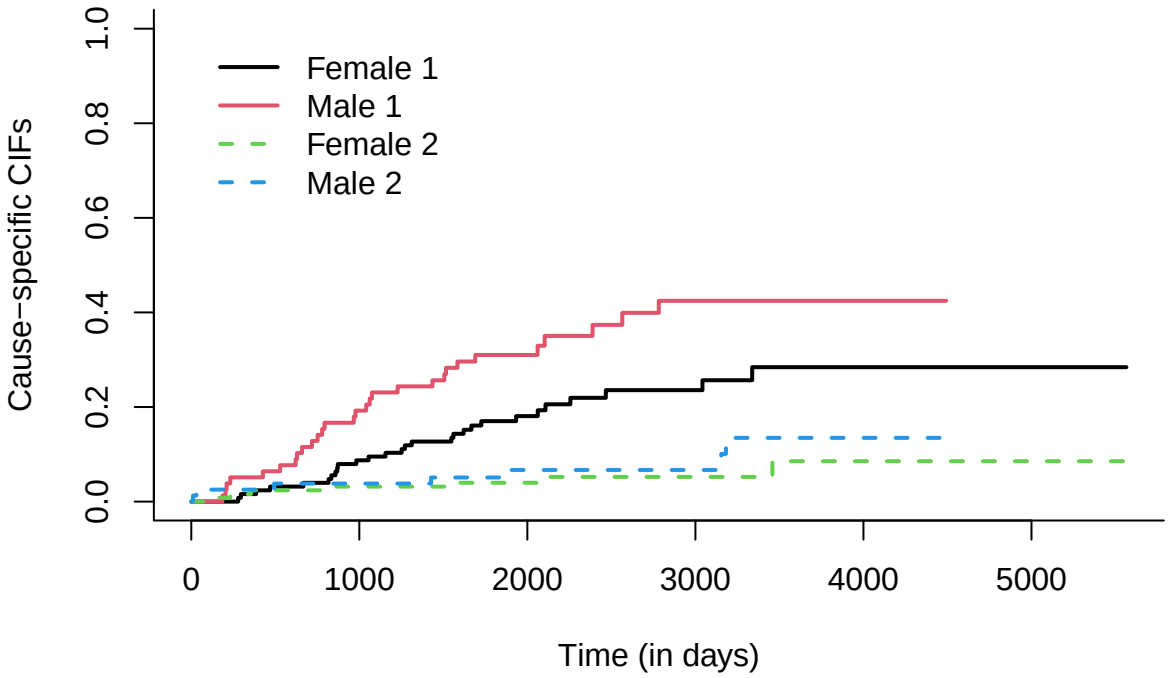
independent (see above). However, the KM method provides biased estimates in case the independence assumption is violated (i.e., the event time distributions for individuals that do experience or do not experience the competing event are different), hence, one should estimate cause-specific CIFs to gain additional insights into the nature of the survival data under study. The *cuminc()* function is the R package *cmprsk* can be used to estimate the cause-specific CIFs. Estimate the cause-specific CIF for death related to melanoma and for death related to another cause.

```
# Estimation of the cause-specific CIF
#--------------------------------------
library(cmprsk)
cif <- cuminc(ftime = Melanoma$time, fstatus = Melanoma$status, cencode = 0)

plot(cif, col=1, xlab="Time (in days)", ylab = "Cause-specific CIFs",
     lty = c(1,2), lwd = 2)
lines(fit1$time, 1-fit1$surv, lwd = 2, col = 2)
```



4. The *cuminc()* function also allows for the estimation of the cause-specific CIFs for different groups of patients, for example, males and females. Look in the documentation of the function to estimate the cause-specific CIFs for females and males separately.

```
# Estimation of the cause-specific CIFs for males and females
#------------------------------------------------------------
library(cmprsk)
cif_gender <- cuminc(ftime = Melanoma$time, fstatus = Melanoma$status,
                     group=Melanoma$sex)

plot(cif_gender, col=1:4, xlab="Time (in days)", ylab = "Cause-specific CIFs",
     lty = c(1,1,2,2), lwd = 2)
```



NOTE: The following arguments need to be specified in cuminc(): (1) a failure time variable, (2) a variable with distinct code for different causes of failure and optionally (3) a group argument that takes a variable specifying distinct groups. In this example, patients are divided in two groups by sex (males and females).

5. Interpret the results in question Q4.

```
## The cause-specific CIFs for males and females express the risk of experiencing
## each of the competing events for males vs. females. Apparently, there exists
## a large(r) difference between the risk of dying from melanoma between males and
## females as compared to the difference in risk of dying from another cause. More
## specifically, males tend to have a higher risk than females to die after the
## surgical operation as a result of melanoma. Differences between the survival
## probabilities for the other cause are small between males and females (dashed
## lines).
```

6. In order to perform a formal statistical test to check whether the difference between males and females is different, a modified χ^2-statistic can be considered (Gray, 1988). Perform this test (see documentation *cuminc()* function). Interpret the result.

```
cif_gender$Test
```

```
##        stat        pv df
## 1 5.8140209 0.0158989  1
## 2 0.8543656 0.3553203  1
```

```
## The first column of the output shows the $\chi^2$-statistic for the between-
## group test, and the second column presents the corresponding p-values.
##
## Based on the results, there exists a significant difference in risk of dying
## from melanoma between males and females at a 5% significance level
## (p-value = 0.016). There is no significant difference in mortality risk from
## other causes for male versus female patients (p-value = 0.360).
```

Cause-specific hazard regression

A cause-specific hazard regression model is used to investigate the effect of different covariates directly on the cause-specific hazard function (i.e., the rate of occurrence of a specific cause/event). A cause-specific hazard regression model can be fitted using the standard Cox proportional hazards regression model while treating failures from the cause of interest as events and failure from other causes as censored observations. The effect of covariates on the cause-specific hazard can be estimated using the partial likelihood method as proposed by Cox. Such a model is fit using the *coxph()* function in the *survival* package in R.

1. Assess the impact of gender, age and invasion on the cause-specific hazard function for dying as a result of the melanoma.

```
# Cox PH model for cause-specific hazards regression
#-------------------------------------------------------
csh <- coxph(Surv(time, status==1) ~ sex+age+invasion,data=Melanoma)
summary(csh)
```

```
## Call:
## coxph(formula = Surv(time, status == 1) ~ sex + age + invasion,
##     data = Melanoma)
##
##   n= 205, number of events= 57
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## sexMale          0.663383  1.941349 0.266320 2.491 0.012741 *
## age              0.009823  1.009871 0.008339 1.178 0.238840
## invasionlevel.1 1.037168  2.821217 0.328241 3.160 0.001579 **
## invasionlevel.2 1.403225  4.068300 0.380744 3.685 0.000228 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## sexMale             1.941     0.5151    1.1519     3.272
## age                 1.010     0.9902    0.9935     1.027
## invasionlevel.1     2.821     0.3545    1.4826     5.368
## invasionlevel.2     4.068     0.2458    1.9290     8.580
##
## Concordance= 0.7  (se = 0.035 )
## Likelihood ratio test= 26.7  on 4 df,   p=2e-05
## Wald test            = 24.39  on 4 df,   p=7e-05
## Score (logrank) test = 26.85  on 4 df,   p=2e-05
```

NOTE: The first argument of the coxph() function takes an object of class Surv, where the "status==1" indicates that only status value of 1 is considered as event and other values are considered as censored. The ***summary output*** *shows the coefficients and corresponding hazard ratios (HRs). The last five lines display statistics for the fitness of the model.*

2. Interpret the output of the Cox proportional hazards model fitted in question Q1 above.

```
## Based on the model output and the estimated HRs, we can conclude that there
## exists a significant difference in instantaneous probability of dying from
## melanoma between males and females. More specifically, the hazard of dying
## from melanoma is estimated to be approximately 2 times higher for males vs.
## females, accounting for death due to another reason. Furthermore, there is
## no significant effect of age on the cause-specific hazard (p-value = 0.239).
## The invasion level alters the rate of melanoma-related mortality significantly
## with the highest rate observed for invasion level 2, with the hazard being
```

```
## approximately four times higher as compared to the hazard with baseline
## invasion level (level 0).
```

NOTE: Alternatively, cause-specific regression can also be performed using the CSC() function contained in the riskRegression package in R.

```r
library(prodlim)
library(riskRegression)
csh_ext<-CSC(Hist(time,status)~sex+age+invasion,data=Melanoma)
print(csh_ext)
```

```
## CSC(formula = Hist(time, status) ~ sex + age + invasion, data = Melanoma)
##
## Right-censored response of a competing.risks model
##
## No.Observations: 205
##
## Pattern:
##
## Cause     event right.censored
##   1          57              0
##   2          14              0
##   unknown     0            134
##
##
## ----------> Cause:  1
##
## Call:
## coxph(formula = survival::Surv(time, status) ~ sex + age + invasion,
##     x = TRUE, y = TRUE)
##
##   n= 205, number of events= 57
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## sexMale          0.663383  1.941349 0.266320 2.491 0.012741 *
## age              0.009823  1.009871 0.008339 1.178 0.238840
## invasionlevel.1 1.037168  2.821217 0.328241 3.160 0.001579 **
## invasionlevel.2 1.403225  4.068300 0.380744 3.685 0.000228 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## sexMale             1.941     0.5151    1.1519     3.272
## age                 1.010     0.9902    0.9935     1.027
## invasionlevel.1     2.821     0.3545    1.4826     5.368
## invasionlevel.2     4.068     0.2458    1.9290     8.580
##
## Concordance= 0.7  (se = 0.035 )
## Likelihood ratio test= 26.7  on 4 df,   p=2e-05
## Wald test            = 24.39  on 4 df,   p=7e-05
## Score (logrank) test = 26.85  on 4 df,   p=2e-05
```

```
##
##
##
## ----------> Cause:  2
##
## Call:
## coxph(formula = survival::Surv(time, status) ~ sex + age + invasion,
##     x = TRUE, y = TRUE)
##
##   n= 205, number of events= 14
##
##                       coef exp(coef) se(coef)      z Pr(>|z|)
## sexMale           0.29893   1.34842  0.54620  0.547 0.584180
## age               0.09271   1.09715  0.02592  3.576 0.000349 ***
## invasionlevel.1 -0.88527   0.41260  0.64069 -1.382 0.167051
## invasionlevel.2 -1.34958   0.25935  1.12351 -1.201 0.229666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## sexMale            1.3484     0.7416   0.46227     3.933
## age                1.0971     0.9115   1.04279     1.154
## invasionlevel.1    0.4126     2.4236   0.11754     1.448
## invasionlevel.2    0.2593     3.8558   0.02868     2.345
##
## Concordance= 0.839  (se = 0.042 )
## Likelihood ratio test= 18.96  on 4 df,   p=8e-04
## Wald test            = 14.21  on 4 df,   p=0.007
## Score (logrank) test = 15.53  on 4 df,   p=0.004
```

```r
## The summary output is quite similar to that produced by coxph() function,
## except that the CSC() function automatically produces cause-specific hazard
## models for both types of events (cause 1 and 2). Using the fitted regression
## model, one can predict individual risk with given covariate values.
```

3. Interpret the results of the competing risk analysis obtained in Q2 above.

```r
## Interpretation in terms of the cause-specific hazard ratios. For example, a
## significant difference in hazard of dying due to melanoma exists between males
## and females at a 5% significant level. More specifically, the (cause-specific)
## hazard of dying due to melanoma is 1.94 (95% CI: 1.15 - 3.27) times higher for
## males as compared to females.
```

Subdistribution hazard regression

A subdistribution hazard regression model is also referred to as the Fine and Gray model. The motivation for the use of the Fine and Gray model is that the effect of a covariate on the cause-specific hazard function may be quite different from that on the cumulative incidence. In other words, a covariate may have strong influence on the cause-specific hazard, but no effect on the CIF. The difference between cause-specific hazard regression and subdistribution hazard regression is that both hazard functions rely on a different treatment of the competing risk events. The former considers competing risk events as non-informative censoring, whereas

the latter takes into account the informative censoring nature of the competing risk events.

The Fine and Gray model can be fitted using the *FGR() function* in the *riskRegression* package in R.

1. Consider the Fine and Gray model for the example at hand. More specifically, fit a subdistribution hazards regression model with gender, age and invasion as covariates.

```
# Fine and Gray model for subdistribution hazards regression
#-------------------------------------------------------------
shm <- FGR(Hist(time,status)~sex+age+invasion,data=Melanoma)
shm
```

```
##
## Right-censored response of a competing.risks model
##
## No.Observations: 205
##
## Pattern:
##
## Cause     event right.censored
##   1          57              0
##   2          14              0
##   unknown     0            134
##
##
## Fine-Gray model: analysis of cause 1
##
## Competing Risks Regression
##
## Call:
## FGR(formula = Hist(time, status) ~ sex + age + invasion, data = Melanoma,
##     cause = "1")
##
##                     coef exp(coef) se(coef)     z p-value
## sexMale          0.62762      1.87  0.27170 2.310  0.0210
## age              0.00565      1.01  0.00913 0.619  0.5400
## invasionlevel.1 1.04909      2.86  0.34040 3.082  0.0021
## invasionlevel.2 1.37802      3.97  0.40137 3.433  0.0006
##
##                 exp(coef) exp(-coef)  2.5% 97.5%
## sexMale              1.87      0.534 1.100  3.19
## age                  1.01      0.994 0.988  1.02
## invasionlevel.1      2.86      0.350 1.465  5.56
## invasionlevel.2      3.97      0.252 1.806  8.71
##
## Num. cases = 205
## Pseudo Log-likelihood = -274
## Pseudo likelihood ratio test = 24.6  on 4 df,
##
## Convergence: TRUE
```

2. Interpret the findings in question Q1 above.

```
## The estimated coefficient for cause 1 (i.e., mortality as a result of melanoma)
## deviates slightly from the one obtained in the cause-specific hazard model
## (HR: 1.87 vs. 1.94), as a result of different assumptions for the competing
## risks. The numerical values derived from the Fine and Gray model have no
```

```
## simple interpretation. These estimates reflect the ordering of the cumulative
## incidence curves.
##
## The cause-specific hazard function represents the time-dependent rate of
## failure due to cause per unit of time for patients who are still alive. The
## subdistribution hazard, however, is the rate of failure due to cause one per
## unit of time for patients who are either alive or have already failed from
## cause 2. In other words, patients who fail from other causes remain in the
## risk set.
```

3. Discuss the differences between the cause-specific hazard regression and subdistribution hazard regression models in terms of interpretation and formulation.

References

- Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. Annals of Statistics, 16:1141-54.
- Zhang, Z. (2017). Survival analysis in the presence of competing risks. Annals of Translational Medicine, 5(3):47.