Flexible hazard-based Mixed-effects Models for the Excess Mortality Hazard Practical

A. Belot and the CENSUR Working Survival Group

October 19, 2019

Abstract

In this practical, we use data of men diagnosed with a Lip-Oral Cavity-Pharynx cancer between 1997 and 2010 in a French region (Basse Normandie), and followed up to the 30^{th} of June 2013.

In the **first exercise** of the practical, the objective is to **describe the association between the age at diagnosis and the excess mortality hazard**, and we will focus on interpreting the results of the different models and their underlying assumption.

In the **second exercise**, you'll describe the association between an ecological deprivation index and the excess mortality hazard, adjusting on age at diagnosis and accounting for the hierarchical structure of the data.
The practical will be mainly using the package `mexhaz`

Introduction

The file is called `fakeLOCP.dat`, and is a tabulate delimited text file. It contains the following columns:

- sex: 1 for man, 2 for woman
- ydiag : year of diagnosis (continuous)
- agediag: age at diagnosis (continuous)
- EDI: European Deprivation Index (rounded with 2 decimals) associated to an IRIS (see below)
- quintile: Deprivation quintile of the French population
- timesurv : survival time in years
- status: Indicator of the event (0=alive, 1=dead)
- myclus: IRIS of residence at diagnosis (character)
- expectedrate: Population (Expected) mortality rate at the time of the last known vital status (and the corresponding age and year)

Data preparation

Import the data in R and explore the data

Start by loading the package `mexhaz` and by defining the path in `mypath` below to indicate where your datafile is located.

```
> # Load necessary package
> library(mexhaz)
> mypath <- "C:/Users/Aurel/Dropbox/WorkStat/Teaching/Corte2019/FPMM/Practicals/"
> mydat <- read.delim(paste0(mypath, "fakeLOCP.dat"))
```

Then we have to do some data management. To do so, copy and paste the R-code given below to

- Select only men
- Create a survival time variable by censoring patients at 10 years, with the corresponding vital status
- Create the dummy variables called `Iagecat**`, corresponding to the following age classes [15-45[, [45-55[, [55-65[, [65-75[and [75-++[. For example, `Iagecat1545` will equal 1 for patients aged between 15 and 45 years old, and 0 otherwise.
- Create the variables corresponding to the age at diagnosis (centred at 70 years old) expressed in a Truncated Power Basis spline of degree 3 with one knot at 0 (i.e. 70 years old as it has been already centred). The same is done but with the age centred at 70 years and rescaled (by dividing by 10).

- Summarize the available information (number of events/censored before 10 years) by deprivation quintiles
- Summarize the continuous variable agediag and EDI

```
> # select only men
> mydat <- subset(mydat, sex==1)
> summary(mydat)

      sex         ydiag         agediag           EDI             quintile
 Min.   :1   Min.   :1997   Min.   :16.49   Min.   :-7.830   Min.   :1.000
 1st Qu.:1   1st Qu.:2000   1st Qu.:52.51   1st Qu.:-1.280   1st Qu.:3.000
 Median :1   Median :2003   Median :60.14   Median : 0.410   Median :4.000
 Mean   :1   Mean   :2003   Mean   :61.06   Mean   : 1.078   Mean   :3.542
 3rd Qu.:1   3rd Qu.:2007   3rd Qu.:69.21   3rd Qu.: 2.390   3rd Qu.:5.000
 Max.   :1   Max.   :2010   Max.   :95.41   Max.   :17.400   Max.   :5.000
    timesurv            status           myclus       expectedrate
 Min.   : 0.001747   Min.   :0.0000   Min.   :  1.0   Min.   :0.000446
 1st Qu.: 0.727183   1st Qu.:1.0000   1st Qu.:135.0   1st Qu.:0.008523
 Median : 1.936618   Median :1.0000   Median :379.0   Median :0.013475
 Mean   : 3.445450   Mean   :0.8008   Mean   :398.7   Mean   :0.026127
 3rd Qu.: 5.101770   3rd Qu.:1.0000   3rd Qu.:632.0   3rd Qu.:0.027092
 Max.   :16.456786   Max.   :1.0000   Max.   :976.0   Max.   :0.364189

> # create survival time and corresponding vital status
> mydat$timesurv10y <- pmin(mydat$timesurv, 10)
> mydat$status10y <- ifelse(mydat$timesurv10y==mydat$timesurv,mydat$status,0)
> # Creation of useful variables
>
> mydat$Iagecat1545 <- ifelse(mydat$agediag>=15 & mydat$agediag<45,1,0)
> mydat$Iagecat4555 <- ifelse(mydat$agediag>=45 & mydat$agediag<55,1,0)
> mydat$Iagecat5565 <- ifelse(mydat$agediag>=55 & mydat$agediag<65,1,0)
> mydat$Iagecat6575 <- ifelse(mydat$agediag>=65 & mydat$agediag<75,1,0)
> mydat$Iagecat75pp <- ifelse(mydat$agediag>=75 ,1,0)
> # Alternative coding:
> # mydat$agecat <- cut(mydat$agediag, breaks=c(15, 45, 55, 65, 75, 150), right = FALSE)
> # table(mydat$agecat)
>
> # Creation of the variable for a spline of age (deg 3, knot 70) in a truncated power basis
> # not reduced
> mydat$agediagc=mydat$agediag-70
> mydat$agediagc2 <- mydat$agediagc^2
> mydat$agediagc3 <- mydat$agediagc^3
> mydat$agediagc3plus <- (mydat$agediagc-0)^3*(mydat$agediagc>0)
> # redcued
> mydat$agediagcr=(mydat$agediag-70)/10
> mydat$agediagcr2 <- mydat$agediagcr^2
> mydat$agediagcr3 <- mydat$agediagcr^3
> mydat$agediagcr3plus <- (mydat$agediagcr-0)^3*(mydat$agediagcr>0)
```

Briefly summarize the information available (R-code given below):

- Tabulate the number of events observed before 10 years by deprivation quintiles
- Summarize the continuous variable agediag and EDI

```
> # Number of events by deprivation quintiles
> with(mydat, table(quintile, status10y))

        status10y
quintile   0   1
       1 101 224
       2 129 338
       3 150 498
       4 208 809
       5 175 807

> # Distribution of EDI and agediag in the population
> summary(mydat$EDI)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -7.830  -1.280   0.410   1.078   2.390  17.400
```

```
> summary(mydat$agediag)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  16.49   52.51   60.14   61.06   69.21   95.41 
```

Exercise 1

Model 1: Excess mortality hazard regression model with time-fixed effects of age at diagnosis in categories

1. Using `mexhaz`, fit an excess mortality hazard regression model, assuming an exponentiel of a cubic B-spline with 2 knots located at 1 and 5 years for the baseline hazard, and time-fixed effects for the covariables age-groups (reference category 65-75). Save the model in an object called `FPM1`

$$\lambda_E(t,\mathbf{x}) = \lambda_0(t)\exp\Big(\sum_{i=1,i\neq 4}^{5}\beta_i Iagecat_i\Big)$$

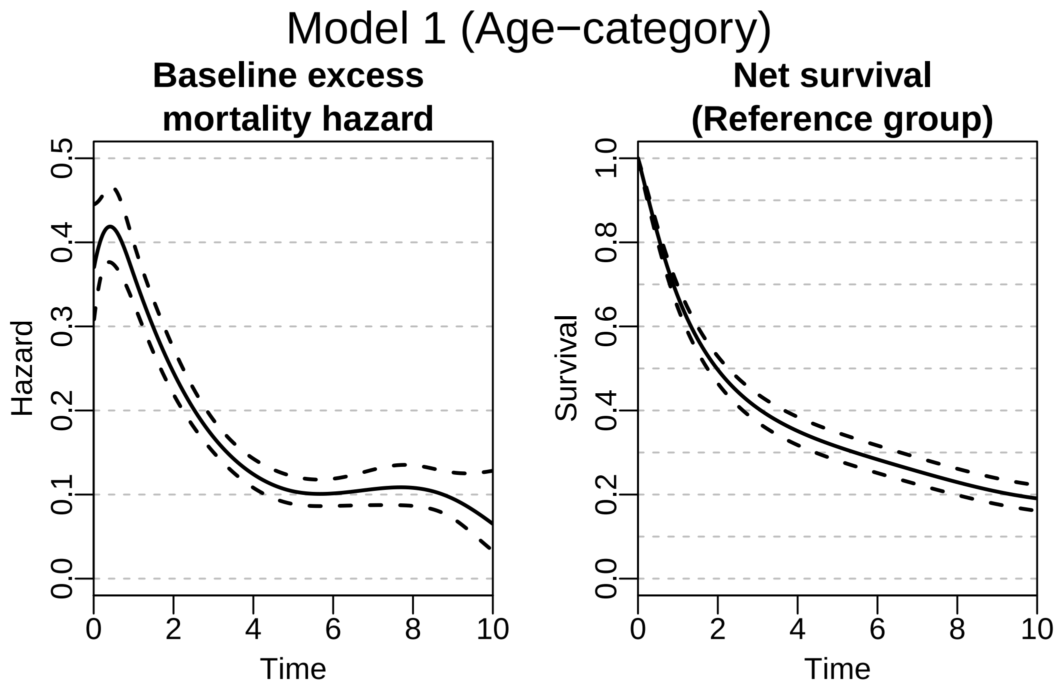
```
> FPM1 <- mexhaz(Surv(timesurv10y, status10y)~ Iagecat1545 + Iagecat4555 + Iagecat5565 + Iagecat75pp,
                expected="expectedrate", degree=3, knots = c(1,5), base="exp.bs", data=mydat, verbose = 0)

> summary(FPM1)

Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ Iagecat1545 + 
    Iagecat4555 + Iagecat5565 + Iagecat75pp, data = mydat, expected = "expectedrate", 
    base = "exp.bs", degree = 3, knots = c(1, 5), verbose = 0)

Coefficients:
             Estimate   StdErr  t.value   p.value    
Intercept   -0.999687 0.096820 -10.3252 < 2.2e-16 ***
BS3.1        0.228656 0.133900   1.7077  0.087790 .  
BS3.2       -0.367664 0.126097  -2.9157  0.003572 ** 
BS3.3       -1.903339 0.276768  -6.8770 7.232e-12 ***
BS3.4       -0.832278 0.292946  -2.8411  0.004523 ** 
BS3.5       -1.732011 0.359464  -4.8183 1.510e-06 ***
Iagecat1545 -0.325472 0.099345  -3.2762  0.001063 ** 
Iagecat4555 -0.097581 0.059862  -1.6301  0.103171    
Iagecat5565 -0.060875 0.058401  -1.0424  0.297320    
Iagecat75pp  0.127271 0.079306   1.6048  0.108629    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (for proportional effect variables):
               Coef     HR CI.lower CI.upper
Iagecat1545 -0.3255 0.7222   0.5944   0.8775
Iagecat4555 -0.0976 0.9070   0.8066   1.0200
Iagecat5565 -0.0609 0.9409   0.8391   1.0551
Iagecat75pp  0.1273 1.1357   0.9722   1.3268

log-likelihood: -5898.2098 (for 10 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

1.1 Quantify and interpret the effects of the covariable age-group on the excess mortality hazard by writing 2 sentences suitable to an epidemiological journal, based on the Excess Hazard Ratios.

Patients from the age-group [15,45[have an excess mortality hazard (EMH) which is about 30% lower than patients in the age-group [65,75[(EHR 0.72, 95% CI (0.59, 0.88)).
Patients aged 75 years old and more have an EMH which is 1.14 times the one for patients aged between 65 and 75 years.

1.2 Predict and plot the baseline excess mortality hazard and the corresponding net survival

```
> mytime <- seq(0.01,10,0.01)
> predFPM1 <- predict(FPM1, time.pts = mytime,
                      data.val = data.frame(Iagecat1545=0, Iagecat4555=0, Iagecat5565=0, Iagecat75pp=0))
> par(mfrow=c(1,2), oma = c(0, 0, 2, 0), mgp=c(1.5,0.4,0), mar=c(3, 3, 2.5, 1))
> plot(predFPM1, which="hazard", xlim=c(0,10), ylim=c(0,0.5),
    conf.int = T, main="Baseline excess \n mortality hazard", lwd=2,
    panel.first=abline(h=seq(0,0.5,by=0.1), tck=1, lty=8, col="grey"))
> plot(predFPM1, which="surv", xlim=c(0,10), ylim=c(0,1),
    conf.int = T, main="Net survival \n (Reference group)", lwd=2,
    panel.first=abline(h=seq(0,1,by=0.1), tck=1, lty=8, col="grey"))
> mtext("Model 1 (Age-category)", outer = TRUE, cex = 1.5)
```



1.3 Predict the net survival at 10 years for each age-group

```
> datatopred <- data.frame(rbind(c(1,0,0,0), c(0,1,0,0), c(0,0,1,0), c(0,0,0,0), c(0,0,0,1) ))
> names(datatopred) <- c("Iagecat1545", "Iagecat4555", "Iagecat5565", "Iagecat75pp")
> predFPM1ageall <- predict(FPM1, time.pts = 10, data.val = datatopred)
> predFPM1ageall$results$surv

[1] 0.3022862 0.2225523 0.2103943 0.1907846 0.1523685
```

Model 2 : Excess mortality hazard regression model with linear and time-fixed effects of age at diagnosis

2. Create a variable corresponding to the agediag centered, called `agediagc`, `agediagc=agediag-70`, fit the following model (the baseline being parametrised in the same way as model `FPM1`), and save it in an object called `FPM2`

$$\lambda_E(t, \mathbf{x}) = \lambda_0(t)\exp\left(\beta_1 agediagc\right)$$

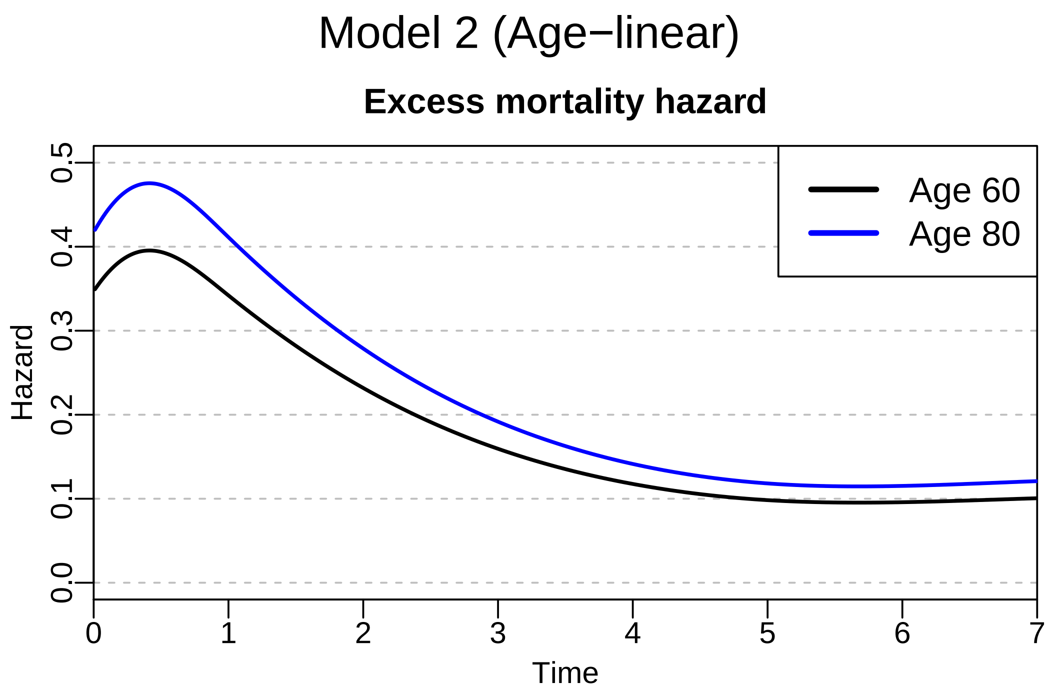
```
> FPM2 <- mexhaz(Surv(timesurv10y, status10y)~ agediagc , expected="expectedrate",
                degree=3, knots = c(1,5), base="exp.bs", data=mydat, verbose = 0)

> summary(FPM2)

Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ agediagc, data = mydat,
    expected = "expectedrate", base = "exp.bs", degree = 3, knots = c(1,
        5), verbose = 0)
```

```
Coefficients:
             Estimate    StdErr  t.value   p.value    
Intercept -0.9663932 0.0898646 -10.7539 < 2.2e-16 ***
BS3.1      0.2305179 0.1337747   1.7232  0.084946 .  
BS3.2     -0.3645436 0.1259955  -2.8933  0.003836 ** 
BS3.3     -1.8962882 0.2765186  -6.8577 8.263e-12 ***
BS3.4     -0.8338228 0.2926964  -2.8488  0.004415 ** 
BS3.5     -1.7264390 0.3588259  -4.8114 1.564e-06 ***
agediagc   0.0092191 0.0019062   4.8363 1.381e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (for proportional effect variables):
       Coef     HR CI.lower CI.upper
[1,] 0.0092 1.0093   1.0055    1.013

log-likelihood: -5896.7752 (for 7 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

2.1 Interpret the effect of age at diagnosis on the excess mortality hazard by writing 1 sentence suitable to an epidemiological journal.
For one unit increase of age at diagnosis, the EMH is increased by 1 % (the EMH is multiplied by 1.01 -EHR-)

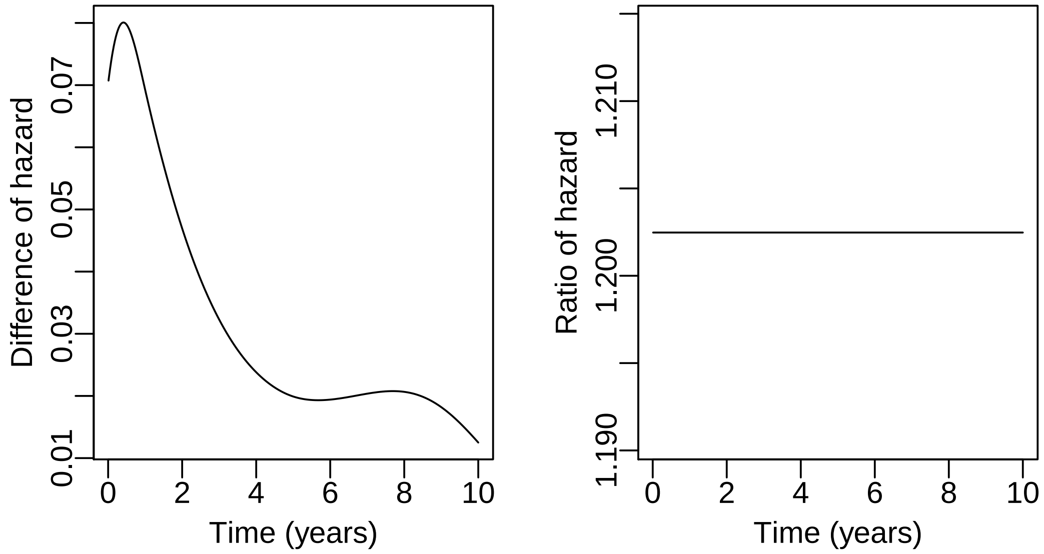
2.2 Predict and plot the excess mortality hazard for men aged 60 and 80 years old (Make sure you use the correct value for the agediag, as it was centred)

```
> myagediag <- 60; myagediagc <- myagediag-70;
> predFPM2age60 <- predict(FPM2, time.pts = mytime,
                    data.val = data.frame(agediagc=myagediagc))
> myagediag <- 80; myagediagc <- myagediag-70;
> predFPM2age80 <- predict(FPM2, time.pts = mytime,
                    data.val = data.frame(agediagc=myagediagc))
> par(oma = c(0, 0, 2, 0), mgp=c(1.5,0.4,0), mar=c(3, 3, 2.5, 1))
> plot(predFPM2age60, which="hazard", xlim=c(0,7), ylim=c(0,0.5),
      conf.int = F, main="Excess mortality hazard", lwd=2,
      panel.first=abline(h=seq(0,0.5,by=0.1), tck=1, lty=8, col="grey"))
> lines(predFPM2age80, which="hazard", conf.int = F, col="blue", lwd=2)
> legend("topright", c("Age 60", "Age 80"), col=c("black", "blue"), lwd=3, cex=1.2, bg="white")
> mtext("Model 2 (Age-linear)", outer = TRUE, cex = 1.5)
```



2.3 Is the difference between the 2 hazards constant over time? and what about their ratio? Check your answer using the objects `predFPM2age60` and `predFPM2age80`.

```
> diffhaz <- predFPM2age80$results$hazard-predFPM2age60$results$hazard
> ratiohaz <- predFPM2age80$results$hazard/predFPM2age60$results$hazard
> par(mfrow=c(1,2), oma = c(0, 0, 2, 0), mgp=c(1.5,0.4,0), mar=c(3, 3, 2.5, 1))
> plot(mytime, diffhaz, type="l", xlab="Time (years)", ylab="Difference of hazard")
> plot(mytime, ratiohaz, type="l", xlab="Time (years)", ylab="Ratio of hazard")
```



Check that you can obtain the value of the Hazard ratio from the results provided after fitting `FPM2`

```
> exp((80-60)*FPM2$coefficients["agediagc"])

agediagc
1.202474
```

2.4 Calculate the ratio between the excess mortality hazard at 5 years of a 56 years old men and the excess mortality hazard at 5 years of a 55 years old men. Repeat the calculation for two men of 83 and 82 years old. Was it expected, and why?

```
> (Ratio5y56vs55 <- predict(FPM2, time.pts = 5, data.val = data.frame(agediagc=56-70))$results$hazard/
    predict(FPM2, time.pts = 5, data.val = data.frame(agediagc=55-70))$results$hazard)

[1] 1.009262

> (Ratio5y83vs82 <- predict(FPM2, time.pts = 5, data.val = data.frame(agediagc=83-70))$results$hazard/
     predict(FPM2, time.pts = 5, data.val = data.frame(agediagc=82-70))$results$hazard)

[1] 1.009262
```

Model 3 : Excess mortality hazard regression model with non-linear and time-fixed effect of age at diagnosis

3. Create the 3 variables useful to model the agediagc effect with a cubic spline defined in a Truncated Power Basis (i.e. the quadratic, cubic and cubic plus basis term) with a knot at 0 (i.e. 70 years old).

$$\lambda_E(t, \mathbf{x}) = \lambda_0(t) \exp\left(f(agediagc)\right)$$

```
> FPM3w <- mexhaz(Surv(timesurv10y, status10y)~ agediagc + agediagc2 + agediagc3 + agediagc3plus,
                  expected="expectedrate", degree=3, knots = c(1,5), base="exp.bs", data=mydat, verbose = 0)

> summary(FPM3w)
```

```
Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ agediagc + agediagc2 +
    agediagc3 + agediagc3plus, data = mydat, expected = "expectedrate",
    base = "exp.bs", degree = 3, knots = c(1, 5), verbose = 0)

Coefficients:
                 Estimate      StdErr t.value   p.value
Intercept     -1.0000e+00  1.0524e-01 -9.5020 < 2.2e-16 ***
BS3.1         -1.0000e+00  1.7340e-01 -5.7671 8.778e-09 ***
BS3.2         -1.0000e+00  1.5376e-01 -6.5036 8.977e-11 ***
BS3.3         -1.0000e+00  2.6873e-01 -3.7212 0.0002015 ***
BS3.4         -1.0000e+00  2.5051e-01 -3.9919 6.691e-05 ***
BS3.5         -1.0000e+00  2.5798e-01 -3.8763 0.0001080 ***
agediagc      -3.9925e-08  4.3434e-03  0.0000 0.9999927
agediagc2      7.6171e-07  1.3980e-04  0.0054 0.9956529
agediagc3     -6.1377e-06          NA      NA        NA
agediagc3plus  1.3322e-06  1.7740e-05  0.0751 0.9401419
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (for proportional effect variables):
              Coef HR CI.lower CI.upper
agediagc         0  1   0.9915   1.0086
agediagc2        0  1   0.9997   1.0003
agediagc3        0  1      NaN      NaN
agediagc3plus    0  1   1.0000   1.0000

log-likelihood: -6296.2157 (for 10 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

3.1 What do you observe? Check `FPM3w$code`.

```
> FPM3w$code

[1] 2
```

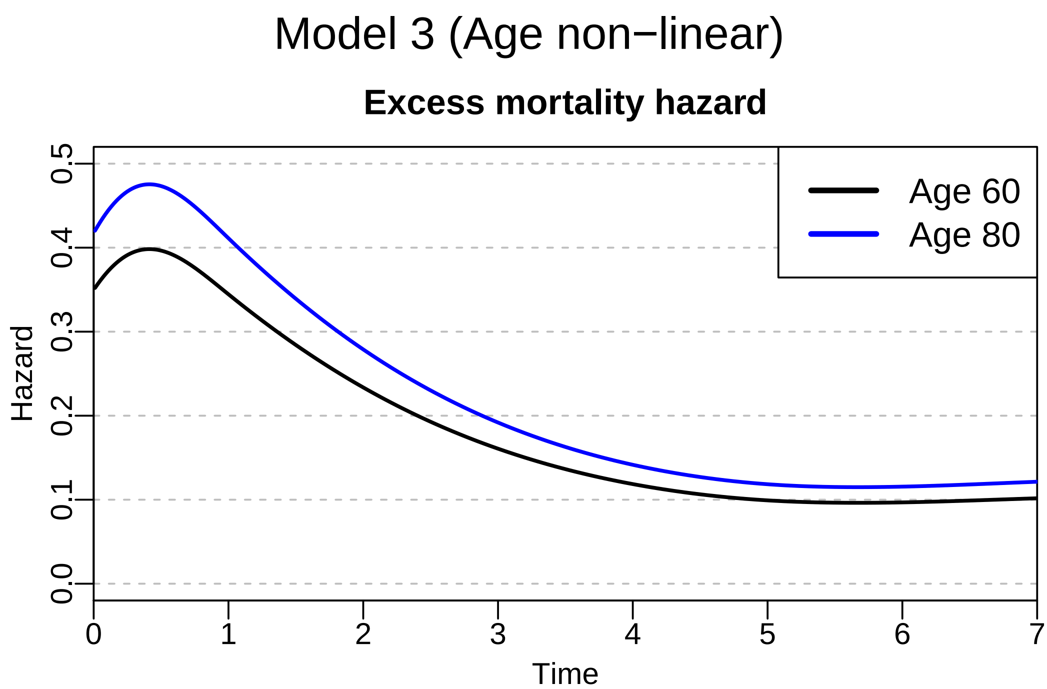
3.2 Try now to fit the same model, but using the variable agediag centered AND rescaled (divided by 10)

```
> FPM3 <- mexhaz(Surv(timesurv10y, status10y)~ agediagcr + agediagcr2 + agediagcr3 + agediagcr3plus,
                expected="expectedrate", degree=3, knots = c(1,5), base="exp.bs", data=mydat, verbose = 0)

> summary(FPM3)

Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ agediagcr + agediagcr2 +
    agediagcr3 + agediagcr3plus, data = mydat, expected = "expectedrate",
    base = "exp.bs", degree = 3, knots = c(1, 5), verbose = 0)

Coefficients:
                Estimate    StdErr  t.value   p.value
Intercept      -0.994164  0.092221 -10.7803 < 2.2e-16 ***
BS3.1           0.229640  0.133790   1.7164  0.086175 .
BS3.2          -0.364248  0.126024  -2.8903  0.003873 **
BS3.3          -1.896150  0.276564  -6.8561 8.358e-12 ***
BS3.4          -0.828284  0.292709  -2.8297  0.004686 **
BS3.5          -1.722550  0.359110  -4.7967 1.682e-06 ***
agediagcr       0.087667  0.051063   1.7168  0.086101 .
agediagcr2      0.049912  0.050142   0.9954  0.319606
agediagcr3      0.018932  0.012632   1.4987  0.134047
agediagcr3plus -0.036187  0.053532  -0.6760  0.499100
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (for proportional effect variables):
                  Coef     HR CI.lower CI.upper
agediagcr       0.0877 1.0916   0.9876   1.2066
```

```
agediagcr2       0.0499 1.0512   0.9528   1.1598
agediagcr3       0.0189 1.0191   0.9942   1.0447
agediagcr3plus -0.0362 0.9645   0.8684   1.0712

log-likelihood: -5894.5212 (for 10 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

3.3 Predict and plot the excess mortality hazard for men aged 60 and 80 years old (Make sure you use the correct value for the agediag, as it was centred AND reduced)

```
> myagediag <- 60
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10;
> myagediagcr2 <- myagediagcr^2; myagediagcr3 <- myagediagcr^3;
> myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> predFPM3age60 <- predict(FPM3, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus))
> myagediag <- 80
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10;
> myagediagcr2 <- myagediagcr^2; myagediagcr3 <- myagediagcr^3;
> myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> predFPM3age80 <- predict(FPM3, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus))
> par(oma = c(0, 0, 2, 0), mgp=c(1.5,0.4,0), mar=c(3, 3, 2.5, 1))
> plot(predFPM3age60, which="hazard", xlim=c(0,7), ylim=c(0,0.5),
    conf.int = F, main="Excess mortality hazard", lwd=2,
    panel.first=abline(h=seq(0,0.5,by=0.1), tck=1, lty=8, col="grey"))
> lines(predFPM3age80, which="hazard", conf.int = F, col="blue", lwd=2)
> legend("topright", c("Age 60", "Age 80"), col=c("black", "blue"), lwd=3, cex=1.2, bg="white")
> mtext("Model 3 (Age non-linear)", outer = TRUE, cex = 1.5)
```



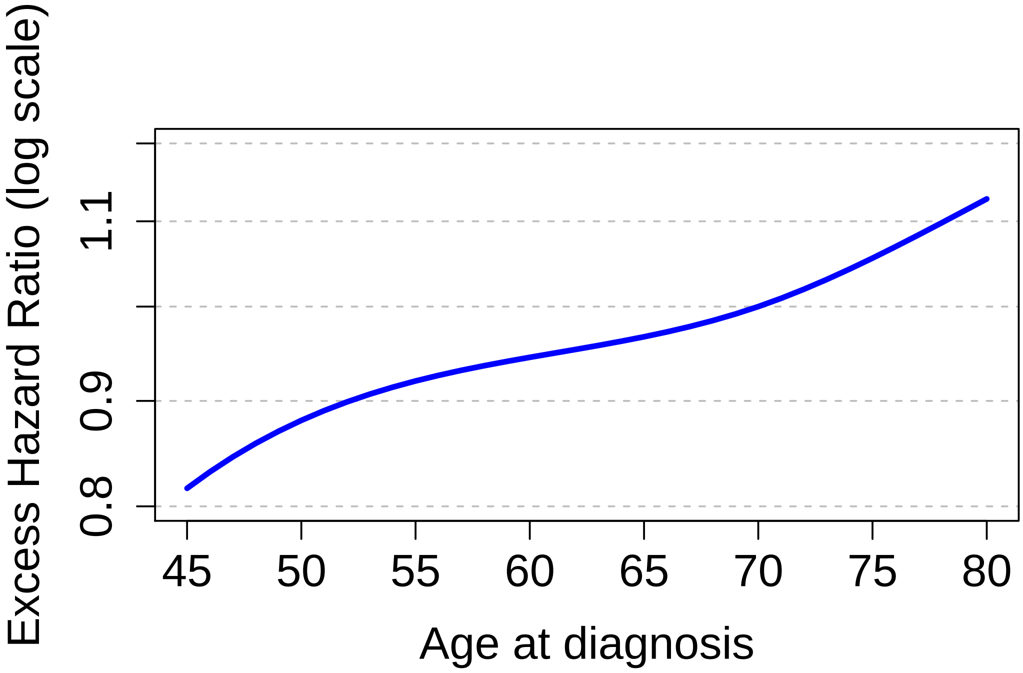
3.4 Calculate the ratio between the excess mortality hazard at 5 years of a 56 years old men and the excess mortality hazard at 5 years of a 55 years old men. Repeat the calculation for two men of 83 and 82 years old. Was it expected and why?

```
> myagediag <- 56
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10; myagediagcr2 <- myagediagcr^2
> myagediagcr3 <- myagediagcr^3; myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> haz5y56 <- predict(FPM3, time.pts = 5,
                    data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
```

```
                                 agediagcr3=myagediagcr3,
                                 agediagcr3plus=myagediagcr3plus))$results$hazard
> myagediag <- 55
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10; myagediagcr2 <- myagediagcr^2
> myagediagcr3 <- myagediagcr^3; myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> haz5y55 <- predict(FPM3, time.pts = 5,
                   data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                 agediagcr3=myagediagcr3,
                                 agediagcr3plus=myagediagcr3plus))$results$hazard
> haz5y56/haz5y55

[1] 1.006257

> myagediag <- 83
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10; myagediagcr2 <- myagediagcr^2
> myagediagcr3 <- myagediagcr^3; myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> haz5y83 <- predict(FPM3, time.pts = 5,
                   data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                 agediagcr3=myagediagcr3,
                                 agediagcr3plus=myagediagcr3plus))$results$hazard
> myagediag <- 82
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10; myagediagcr2 <- myagediagcr^2
> myagediagcr3 <- myagediagcr^3; myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> haz5y82 <- predict(FPM3, time.pts = 5,
                   data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                 agediagcr3=myagediagcr3,
                                 agediagcr3plus=myagediagcr3plus))$results$hazard
> haz5y83/haz5y82

[1] 1.013239
```

3.5 Plot the non-linear EHR according to age, limiting the range value of age to the 5^{th} and 95^{th} percentiles

```
> quantile(mydat$agediag, na.rm=T, probs = c(0, 0.025, 0.05, 0.25, 0.5, 0.75, 0.95, 0.975, 1))

      0%     2.5%       5%      25%      50%      75%      95%    97.5%
16.48645 41.65726 44.20541 52.50768 60.14140 69.21228 80.95031 84.60489
    100%
95.41242

> ageplot=(seq(45,80)-70)/10
> HRage <- exp(ageplot*FPM3$coefficients[c("agediagcr")] +
               ageplot^2*FPM3$coefficients[c("agediagcr2")] +
               ageplot^3*FPM3$coefficients[c("agediagcr3")] +
               ageplot^3*FPM3$coefficients[c("agediagcr3plus")]*(ageplot>0))
> plot(10*ageplot+70, HRage, log="y", type="l", xlim=c(45,80), ylim=c(0.8,1.2),
     xlab="Age at diagnosis", ylab="Excess Hazard Ratio (log scale)",
     lwd=3, cex.lab=1.5, cex.axis=1.5, col="blue",
     panel.first=abline(h=seq(0,2,by=0.1), tck=1, lty=8, col="grey"))
```



Model 4 : Excess mortality hazard regression model with non-linear and time-dependent effects of age at diagnosis

4. Fit the model including now non-linear effect of age, assuming non-proportional hazard for age at diagnosis:

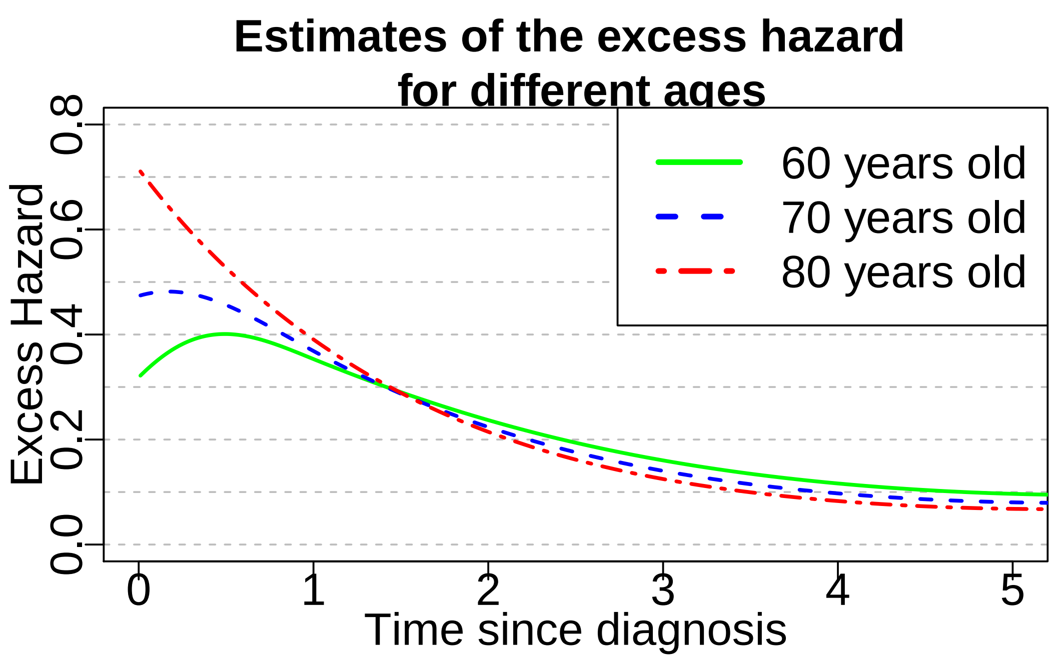
$$\lambda_E(t, \mathbf{x}) = \lambda_0(t) \exp\left(\beta(t) \cdot agediagc + f(agediagc)\right)$$

```
> FPM4 <- mexhaz(Surv(timesurv10y, status10y)~ agediagcr + agediagcr2 + agediagcr3
                + agediagcr3plus + nph(agediagcr),
                expected="expectedrate", degree=3, knots = c(1,5), base="exp.bs",
                data=mydat, verbose = 0)

> summary(FPM4)

Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ agediagcr + agediagcr2 +
    agediagcr3 + agediagcr3plus + nph(agediagcr), data = mydat,
    expected = "expectedrate", base = "exp.bs", degree = 3, knots = c(1,
        5), verbose = 0)

Coefficients:
                 Estimate    StdErr t.value    p.value
Intercept       -0.747659  0.100826 -7.4153 1.522e-13 ***
BS3.1            0.070822  0.158254  0.4475 0.6545282
BS3.2           -0.716351  0.175904 -4.0724 4.758e-05 ***
BS3.3           -2.609399  0.431472 -6.0477 1.628e-09 ***
BS3.4           -1.036828  0.470458 -2.2039 0.0275999 *
BS3.5           -1.958812  0.609818 -3.2121 0.0013298 **
agediagcr        0.410230  0.089717  4.5725 4.990e-06 ***
agediagcr2       0.032534  0.050985  0.6381 0.5234469
agediagcr3       0.019078  0.012721  1.4997 0.1337890
agediagcr3plus  -0.049602  0.055369 -0.8958 0.3703996
agediagcr*BS3.1 -0.278293  0.119032 -2.3380 0.0194464 *
agediagcr*BS3.2 -0.475218  0.122114 -3.8916 0.0001015 ***
agediagcr*BS3.3 -0.726026  0.274943 -2.6406 0.0083124 **
agediagcr*BS3.4 -0.339405  0.293767 -1.1554 0.2480264
agediagcr*BS3.5 -0.347759  0.376874 -0.9227 0.3562052
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (for proportional effect variables):
                   Coef     HR CI.lower CI.upper
```

```
agediagcr        0.4102 1.5072   1.2641   1.7970
agediagcr2       0.0325 1.0331   0.9348   1.1417
agediagcr3       0.0191 1.0193   0.9942   1.0450
agediagcr3plus -0.0496 0.9516   0.8537   1.0607

log-likelihood: -5868.0992 (for 15 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

4.1 Predict and plot the excess mortality hazard for patients aged 60, 70 and 80 years old. What would you conclude from this plot?

```
> myagediag <- 60
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10; myagediagcr2 <- myagediagcr^2
> myagediagcr3 <- myagediagcr^3; myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> predFPM4age60 <- predict(FPM4, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus))
> myagediag <- 70
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10; myagediagcr2 <- myagediagcr^2
> myagediagcr3 <- myagediagcr^3; myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> predFPM4age70 <- predict(FPM4, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus))
> myagediag <- 80
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10; myagediagcr2 <- myagediagcr^2
> myagediagcr3 <- myagediagcr^3; myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> predFPM4age80 <- predict(FPM4, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus))
> par(oma = c(0, 0, 2, 0), mgp=c(1.5,0.4,0), mar=c(3, 3, 2.5, 1))
> plot(0,1, xlim=c(0,5),ylim=c(0,0.8), type="n",
    xlab="Time since diagnosis",ylab="Excess Hazard", cex.lab=1.5, cex.axis=1.5, cex=1.5,
    panel.first=abline(h=seq(0,0.8,by=0.1), tck=1, lty=8, col="grey"))
> lines(predFPM4age60, which="hazard", lwd=2, col="green",lty.pe=1, conf.int = F)
> lines(predFPM4age70, which="hazard", lwd=2, col="blue",lty.pe=2, conf.int = F)
> lines(predFPM4age80, which="hazard", lwd=2, col="red",lty.pe=6, conf.int = F)
> title("Estimates of the excess hazard \n for different ages", cex.main=1.5)
> legend("topright",c("60 years old", "70 years old", "80 years old"),
      col=c("green", "blue", "red"), lty=c(1,2,6), lwd=3,cex=1.5, bg="white")
```



Comparison

5. Calculate the Akaike Information Criterion for the 4 models to assess their fit. Which model would you favor?

```
> res <- t(sapply(list(FPM1, FPM2, FPM3, FPM4),
                  function(x) {
                    round( rbind(x$loglik, x$n.par, -2 * x$loglik +2 * x$n.par), 3)} ))
> res <- as.data.frame(res)
> rownames(res) <- c("Age cat", "Age lin", "Age Nlin", "Age Nlin TD")
> colnames(res) <- c("log-lik", "N Param", "AIC")
> res

              log-lik N Param      AIC
Age cat     -5898.210      10 11816.42
Age lin     -5896.775       7 11807.55
Age Nlin    -5894.521      10 11809.04
Age Nlin TD -5868.099      15 11766.20
```

Exercise 2

In this part, we are now interested in quantifying the impact of the EDI (an ecological continuous measure of deprivation) on the excess mortality hazard in men with LOCP cancers. AS we have seen during the lecture, it's recommended to account for the hierarichal structure of the data. Here, patients are nested in cluster (denoted by the variable `myclus`). Our interest here is to describe how the EDI is associated to the EMH, so by default in all our models, we will assume a non-linear and time-dependent effect of age.

Mixed-effects Model for the Excess mortality hazard: example of an analysis studying the association between deprivation and cancer survival

6.1 Fit a mixed-effect excess hazard model with non-linear and time-dependent effect of agediag, and linear and time-fixed effect of the EDI, and including a random effect for the cluster level. For the log of the baseline excess hazard, assume a cubic B-spline with 2 knots located at 1 and 5 years. Parametrise the non linear effect of age using a spline defined in the Truncated Power Basis, with one knot at 70 years. Hint: use the centred and reduced version of age for avoiding convergence issues.

$$\lambda_E(t, \mathbf{x} \mid w_d) = \lambda_0(t) \exp\left(\beta_1(t) \cdot agediagc + f(agediagc) + \beta_2 \cdot EDI + w_d\right)$$

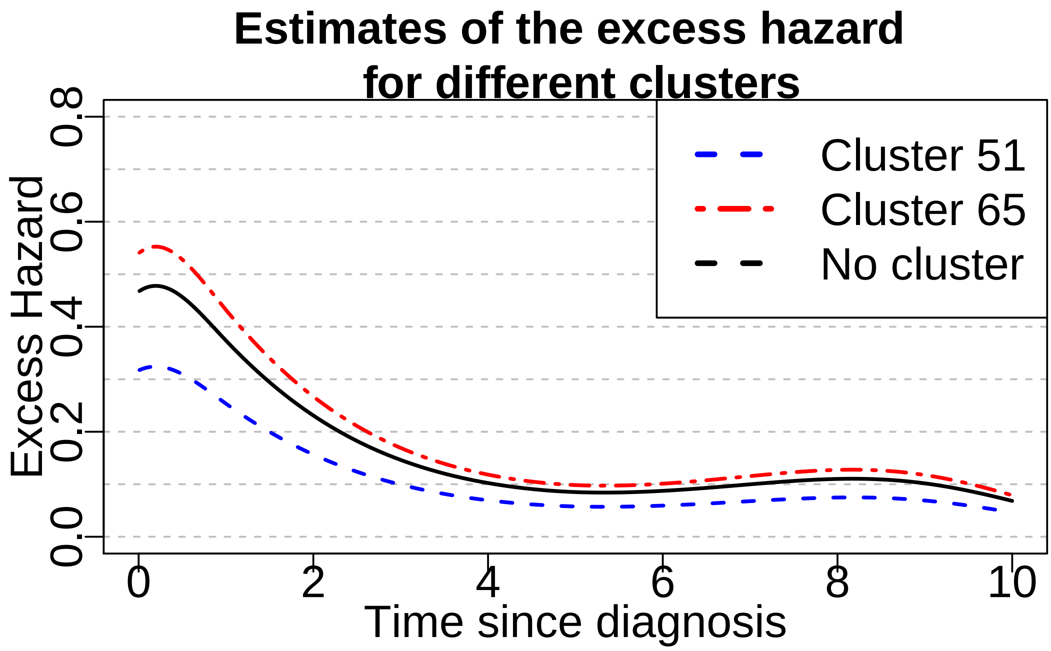
```
> FPMixM1 <- mexhaz(Surv(timesurv10y, status10y)~ agediagcr + agediagcr2 + agediagcr3 +
                        agediagcr3plus + EDI + nph(agediagcr), expected="expectedrate",
                      degree=3, knots = c(1,5), base="exp.bs", random="myclus", data=mydat, verbose = 0)

> summary(FPMixM1)

Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ agediagcr + agediagcr2 +
    agediagcr3 + agediagcr3plus + EDI + nph(agediagcr), data = mydat,
    expected = "expectedrate", base = "exp.bs", degree = 3, knots = c(1,
        5), random = "myclus", verbose = 0)

Coefficients:
                   Estimate     StdErr t.value   p.value
Intercept        -0.8016961  0.1025906 -7.8145 7.286e-15 ***
BS3.1             0.0816080  0.1583314  0.5154 0.6062894
BS3.2            -0.6583725  0.1771171 -3.7172 0.0002047 ***
BS3.3            -2.5511923  0.4329093 -5.8931 4.157e-09 ***
BS3.4            -0.9143514  0.4737267 -1.9301 0.0536740 .
BS3.5            -1.9222775  0.6275715 -3.0630 0.0022080 **
agediagcr         0.4145444  0.0904554  4.5829 4.750e-06 ***
agediagcr2        0.0336922  0.0519711  0.6483 0.5168433
agediagcr3        0.0197883  0.0128946  1.5346 0.1249687
agediagcr3plus   -0.0522282  0.0567544 -0.9202 0.3575071
EDI               0.0199766  0.0064031  3.1198 0.0018245 **
agediagcr*BS3.1  -0.2733748  0.1190096 -2.2971 0.0216742 *
agediagcr*BS3.2  -0.4726560  0.1224240 -3.8608 0.0001151 ***
agediagcr*BS3.3  -0.7327744  0.2751726 -2.6630 0.0077819 **
agediagcr*BS3.4  -0.3230596  0.2939296 -1.0991 0.2717994
agediagcr*BS3.5  -0.3776726  0.3814777 -0.9900 0.3222316
```

```
myclus [log(sd)] -1.3873554  0.1717392 -8.0783 9.016e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (for proportional effect variables):
                   Coef     HR CI.lower CI.upper
agediagcr        0.4145 1.5137   1.2677   1.8074
agediagcr2       0.0337 1.0343   0.9341   1.1452
agediagcr3       0.0198 1.0200   0.9945   1.0461
agediagcr3plus -0.0522 0.9491   0.8492   1.0608
EDI              0.0200 1.0202   1.0074   1.0331

log-likelihood: -5856.7251 (for 17 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

6.2 Interpret the EDI's effec
For one unit increase of the EDI, the EMH is increased by 2 % conditional on the random effect (i.e. the comparison is done in clusters with the same level of the Random effect w_d)

6.3 Predict and plot the excess mortality hazard for a 70 years old men, with EDI equals to 2 and from the cluster 51. Repeat the same operation for cluster 65, and for an individual not among the observed cluster.

```
> myagediag <- 70
> myagediagc <- myagediag-70; myagediagcr <- (myagediag-70)/10;
> myagediagcr2 <- myagediagcr^2; myagediagcr3 <- myagediagcr^3;
> myagediagcr3plus <- (myagediagcr-0)^3*(myagediagcr>0)
> predFPMixM1Clus51 <- predict(FPMixM1, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus,
                                            EDI=2), cluster="51")
> predFPMixM1Clus65 <- predict(FPMixM1, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus,
                                            EDI=2), cluster="65")
> predFPMixM1Noclus <- predict(FPMixM1, time.pts = mytime,
                      data.val = data.frame(agediagcr=myagediagcr, agediagcr2=myagediagcr2,
                                            agediagcr3=myagediagcr3, agediagcr3plus=myagediagcr3plus,
                                            EDI=2), cluster=NULL)
> par(oma = c(0, 0, 2, 0), mgp=c(1.5,0.4,0), mar=c(3, 3, 2.5, 1))
> plot(0,1, xlim=c(0,10),ylim=c(0,0.8), type="n",
     xlab="Time since diagnosis",ylab="Excess Hazard", cex.lab=1.5, cex.axis=1.5, cex=1.5,
     panel.first=abline(h=seq(0,0.8,by=0.1), tck=1, lty=8, col="grey"))
> lines(predFPMixM1Clus51, which="hazard", lwd=2, col="blue", lty.pe=2, conf.int = F)
> lines(predFPMixM1Clus65, which="hazard", lwd=2, col="red", lty.pe=6, conf.int = F)
> lines(predFPMixM1Noclus, which="hazard", lwd=2, col="black",  conf.int = F)
> title("Estimates of the excess hazard \n for different clusters", cex.main=1.5)
> legend("topright",c("Cluster 51", "Cluster 65", "No cluster"),
       col=c("blue", "red", "black"), lty=c(2,6), lwd=3,cex=1.5, bg="white")
```



6.4 Extract the values of the shrinkage estimates for these 2 clusters using the `ranef` function. Are those values coherent with the previous graph? Explain why.

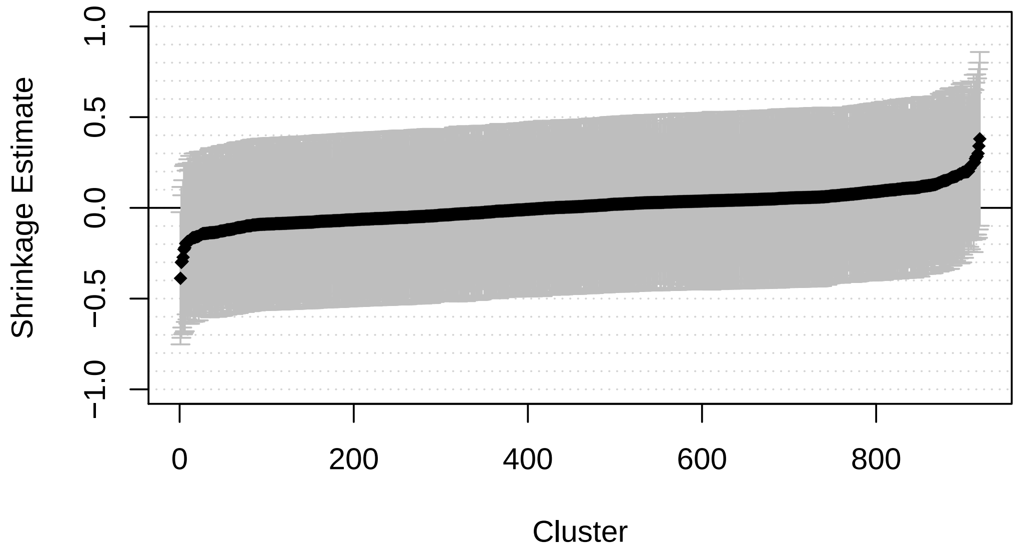
```
> ranef(FPMixM1)[ranef(FPMixM1)$cluster %in% c(51,65),]

   cluster     mu.hat se.mu.hat
51      51 -0.3880077 0.1858356
65      65  0.1452619 0.2503849
```

6.5 Do you think that a fixed-effect model would be suitable? and a stratified model? Why? Check the numbers of patients/events in clusters 51 and 65 to justify your answer

```
> with(mydat[mydat$myclus %in% c(51,65),], table(status10y, myclus))

         myclus
status10y 51 65
        0  9  0
        1 18  3
```

6.6 Plot the values of the shrinkage estimates

```
> datatoplot <- ranef(FPMixM1)
> datatoplot <- datatoplot[order(datatoplot$mu.hat),]
> datatoplot$LBmuhat <- datatoplot$mu.hat-qnorm(0.975)*datatoplot$se.mu.hat
> datatoplot$UBmuhat <- datatoplot$mu.hat+qnorm(0.975)*datatoplot$se.mu.hat
> plot(seq(1,nrow(datatoplot)), datatoplot$mu.hat, type="n",  ylim=c(-1,1),
      xlab="Cluster", ylab="Shrinkage Estimate",
      panel.first=abline(h=seq(-1,1,0.1), tck=1, lty=3, col="lightgray"))
> abline(h=0)
> for (i in seq(1,nrow(datatoplot))){
   arrows(i,datatoplot$LBmuhat[i],i,datatoplot$UBmuhat[i],code=3,
          length=0.05,angle=90, col="grey")}
> points(seq(1,nrow(datatoplot)), datatoplot$mu.hat, type="p", pch=18,
      xlab="Cluster", ylab="Shrinkage Estimate")
```



6.7 Complexify the model by adding a non-linear effect of the EDI, and compare the estimated standard deviation of the random effect. How (intuitively) would you explain this ?

```
> mydat$EDI2 <- mydat$EDI^2
> mydat$EDI2plus <- (mydat$EDI-0)^2*(mydat$EDI>0)
> FPMixM2 <- mexhaz(Surv(timesurv10y, status10y)~ agediagcr + agediagcr2 + agediagcr3 +
                  agediagcr3plus + EDI + EDI2 + EDI2plus+ nph(agediagcr),
                expected="expectedrate", degree=3, knots = c(1,5), base="exp.bs",
                random="myclus", data=mydat, verbose = 0)

> summary(FPMixM2)

Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ agediagcr + agediagcr2 +
    agediagcr3 + agediagcr3plus + EDI + EDI2 + EDI2plus + nph(agediagcr),
    data = mydat, expected = "expectedrate", base = "exp.bs",
    degree = 3, knots = c(1, 5), random = "myclus", verbose = 0)

Coefficients:
                   Estimate    StdErr t.value   p.value
Intercept        -0.7807354 0.1036866 -7.5298 6.466e-14 ***
BS3.1             0.0802849 0.1584113  0.5068 0.6123186
BS3.2            -0.6696140 0.1771173 -3.7806 0.0001591 ***
BS3.3            -2.5547264 0.4327148 -5.9039 3.896e-09 ***
BS3.4            -0.9366500 0.4741604 -1.9754 0.0483046 *
BS3.5            -1.9308659 0.6264792 -3.0821 0.0020719 **
agediagcr         0.4089900 0.0903621  4.5261 6.212e-06 ***
agediagcr2        0.0254311 0.0516362  0.4925 0.6223941
agediagcr3        0.0177126 0.0127818  1.3858 0.1659080
agediagcr3plus   -0.0465922 0.0563966 -0.8262 0.4087753
EDI               0.0538287 0.0160961  3.3442 0.0008341 ***
EDI2              0.0013163 0.0063655  0.2068 0.8361942
EDI2plus         -0.0052176 0.0072452 -0.7201 0.4714830
agediagcr*BS3.1  -0.2742676 0.1190873 -2.3031 0.0213341 *
agediagcr*BS3.2  -0.4761123 0.1224081 -3.8896 0.0001023 ***
agediagcr*BS3.3  -0.7323613 0.2752552 -2.6607 0.0078351 **
agediagcr*BS3.4  -0.3312902 0.2945010 -1.1249 0.2607016
agediagcr*BS3.5  -0.3784073 0.3819032 -0.9908 0.3218308
myclus [log(sd)] -1.5227607 0.2163002 -7.0400 2.312e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Hazard ratios (for proportional effect variables):
                  Coef     HR CI.lower CI.upper
agediagcr       0.4090 1.5053   1.2609   1.7971
agediagcr2      0.0254 1.0258   0.9270   1.1350
agediagcr3      0.0177 1.0179   0.9927   1.0437
agediagcr3plus -0.0466 0.9545   0.8546   1.0661
EDI             0.0538 1.0553   1.0225   1.0891
EDI2            0.0013 1.0013   0.9889   1.0139
EDI2plus       -0.0052 0.9948   0.9808   1.0090

log-likelihood: -5850.6971 (for 19 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

6.8 Complexify the model by adding a time-dependent effect of the EDI in addition to the non-linear effect of the EDI (save it as `FPMixM3`).

```
> FPMixM3 <- mexhaz(Surv(timesurv10y, status10y)~ agediagcr + agediagcr2 + agediagcr3 +
                     agediagcr3plus + EDI + EDI2 + EDI2plus + nph(agediagcr + EDI),
                   expected="expectedrate", degree=3, knots = c(1,5), base="exp.bs",
                   random="myclus", data=mydat, verbose = 0)

iteration = 0
Step:
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Parameter:
 [1] -1.0 -1.0 -1.0 -1.0 -1.0 -1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
[16]  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.1
Function Value
[1] 6204.889
Gradient:
 [1]  -126.33192  -231.37816   -80.34857    60.45167    51.92336    20.83247
 [7]  -107.12074   285.41492 -1740.38304  -104.63506  -184.73075 -1445.09006
[13] -1349.99072   119.47591    77.53900   -59.59098   -63.53493   -30.68888
[19]  -300.80588  -132.56937    73.38056    48.65408    20.77480   165.43047

iteration = 254
Parameter:
 [1] -0.750835626  0.059557037 -0.728821437 -2.540743074 -1.062822107
 [6] -1.919444794  0.403940983  0.025319259  0.017513734 -0.044794130
[11]  0.024880258  0.001692035 -0.005803558 -0.272795469 -0.460439679
[16] -0.747674389 -0.289089546 -0.380474844  0.019698454  0.062719557
[21] -0.024669884  0.137629081 -0.008244258 -1.520737327
Function Value
[1] 5847.647
Gradient:
 [1]  5.293259e-04 -2.592060e-04 -1.396984e-03  1.544257e-03 -3.867925e-04
 [6] -3.932807e-04  6.402843e-04  2.228262e-04  1.382432e-04 -2.182787e-05
[11] -2.819434e-05  1.195986e-03  7.721610e-04 -8.176357e-04 -9.895302e-04
[16]  7.003109e-05 -4.620233e-04 -2.743946e-03  1.164153e-04  5.911716e-05
[21] -7.821654e-05 -1.446097e-04 -2.964953e-04 -2.182925e-04

Gradient relatif proche de z.
L'ittion courante est probablement la solution.

Computation of the Hessian
Computation of the covariance matrix of the shrinkage estimators

Data
  Name N.Obs.Tot N.Obs N.Events N.Clust
 mydat      3439  3439     2676     919

Details
 Iter Eval   Base Nb.Leg Nb.Aghq Optim Method Code    LogLik Total.Time
  254 4390 exp.bs     20      10   nlm    ---    1 -5847.647     398.34

> summary(FPMixM3)
```

```
Call:
mexhaz(formula = Surv(timesurv10y, status10y) ~ agediagcr + agediagcr2 +
    agediagcr3 + agediagcr3plus + EDI + EDI2 + EDI2plus + nph(agediagcr +
    EDI), data = mydat, expected = "expectedrate", base = "exp.bs",
    degree = 3, knots = c(1, 5), random = "myclus", verbose = 0)

Coefficients:
                   Estimate    StdErr t.value   p.value
Intercept        -0.7508356 0.1057101 -7.1028 1.481e-12 ***
BS3.1             0.0595570 0.1621036  0.3674 0.7133426
BS3.2            -0.7288214 0.1801255 -4.0462 5.321e-05 ***
BS3.3            -2.5407431 0.4376720 -5.8051 7.019e-09 ***
BS3.4            -1.0628221 0.4772506 -2.2270 0.0260143 *
BS3.5            -1.9194448 0.6239013 -3.0765 0.0021109 **
agediagcr         0.4039410 0.0904939  4.4637 8.315e-06 ***
agediagcr2        0.0253193 0.0516158  0.4905 0.6237880
agediagcr3        0.0175137 0.0127948  1.3688 0.1711456
agediagcr3plus   -0.0447941 0.0561501 -0.7978 0.4250668
EDI               0.0248803 0.0295249  0.8427 0.3994623
EDI2              0.0016920 0.0063888  0.2648 0.7911455
EDI2plus         -0.0058036 0.0072739 -0.7979 0.4250080
agediagcr*BS3.1  -0.2727955 0.1193887 -2.2849 0.0223777 *
agediagcr*BS3.2  -0.4604397 0.1225260 -3.7579 0.0001742 ***
agediagcr*BS3.3  -0.7476744 0.2758594 -2.7103 0.0067549 **
agediagcr*BS3.4  -0.2890895 0.2946554 -0.9811 0.3266076
agediagcr*BS3.5  -0.3804748 0.3790808 -1.0037 0.3156052
EDI*BS3.1         0.0196985 0.0379489  0.5191 0.6037395
EDI*BS3.2         0.0627196 0.0343327  1.8268 0.0678145 .
EDI*BS3.3        -0.0246699 0.0727633 -0.3390 0.7345984
EDI*BS3.4         0.1376291 0.0722048  1.9061 0.0567219 .
EDI*BS3.5        -0.0082443 0.0943828 -0.0873 0.9303991
myclus [log(sd)] -1.5207373 0.2128578 -7.1444 1.100e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (for proportional effect variables):
                  Coef     HR CI.lower CI.upper
agediagcr       0.4039 1.4977   1.2542   1.7885
agediagcr2      0.0253 1.0256   0.9269   1.1349
agediagcr3      0.0175 1.0177   0.9925   1.0435
agediagcr3plus -0.0448 0.9562   0.8565   1.0675
EDI             0.0249 1.0252   0.9675   1.0863
EDI2            0.0017 1.0017   0.9892   1.0143
EDI2plus       -0.0058 0.9942   0.9801   1.0085

log-likelihood: -5847.647 (for 24 degree(s) of freedom)

number of observations: 3439, number of events: 2676
```

6.9 Compare the fit of the 4 models (with and without a non-linear and a time-dependent effect of the EDI) using Akaike criterion

```
> res <- t(sapply(list(FPMixM1, FPMixM2, FPMixM3),
                 function(x) {
                   round( rbind(x$loglik, x$n.par, -2 * x$loglik +2 * x$n.par), 3)} ))
> res <- as.data.frame(res)
> rownames(res) <- c("EDI lin", "EDI Nlin", "EDI Nlin-TD")
> colnames(res) <- c("log-lik", "N Param", "AIC")
> res

              log-lik N Param      AIC
EDI lin     -5856.725      17 11747.45
EDI Nlin    -5850.697      19 11739.39
EDI Nlin-TD -5847.647      24 11743.29
```

6.10 Plot the non linear effect of the EDI using the model with the better fit according to the Akaike criterion

```
> q05EDI=quantile(mydat$EDI, probs=c(0.05))
> q95EDI=quantile(mydat$EDI, probs=c(0.95))
> ediplot <-round(seq(round(q05EDI,1),round(q95EDI,1),0.1),1)
> NlinEHR.EDI <- exp(ediplot*FPMixM2$coefficients[c("EDI")] +
                ediplot^2*FPMixM2$coefficients[c("EDI2")] +
                ediplot^2*FPMixM2$coefficients[c("EDI2plus")]*(ediplot>0))
> plot(ediplot, NlinEHR.EDI, type="l",xlab="EDI", ylab="Excess Hazard Ratio (log scale)",
     xlim=c(q05EDI,q95EDI), ylim=c(0.8, 1.25), lwd=2, cex.lab=1.5, cex.axis=1.5, col="blue",
     panel.first=abline(h=seq(0,2,by=0.1), tck=1, lty=8, col="grey"))
```