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.

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

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

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`

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

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.

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

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

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`

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

2.1 Interpret the effect of age at diagnosis on the excess mortality hazard by writing 1 sentence suitable to an epidemiological journal.

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)

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

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?

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).

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

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

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

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)

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?

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

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:

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

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?

Comparison

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

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.

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

6.2 Interpret the EDI's effect.

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.

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.

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

6.6 Plot the values of the shrinkage estimates

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 ?

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`).

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

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