Computer Practical Day V
Exercises

Competing risks

In the SPSS file cod.sav you will find data of the cause of death (cod) of 3582 chronic myeloid leukemia (CML) patients from the EBMT, the European Group of Blood and Marrow Transplantation. Time is measured in months from transplant and the cause of death is recorded. The columns in the data are

id
: Patient identification number

survmon
: Time in months from transplantation to death or last follow-up

status
: Survival status; 0 = censored; 1,...,4 = death due to the following causes: Relapse (1), GvHD (2), Infection (3), Other causes (4)

cod
: Cause of death as factor with levels "Alive", "Relapse", "GvHD", "Infection", "Other"

match
: Donor-recipient gender match; factor with levels "No gender mismatch", "Gender mismatch"

tcd
: T-cell depletion; factor with levels "No TCD", "TCD", "Unknown"

year
: Year of transplantation; factor with levels "1985-1989", "1990-1994", "1995-1998"

age
: Patient age at transplant; factor with levels "<=20", "20-40", ">40"

1. Make an overview of the causes of death and of the patient characteristics.

2. Compute the Kaplan-Meier estimate of overall survival and make a survival plot. Note: status=1,…,4 all imply death, and only status=0 implies censoring. Select “status” as event indicator, and after having clicked “Define Event”, specify a range of values (1 through 4) as defining an event.

3. Repeat the procedure and make Kaplan-Meier survival curves for the three age categories. Look up the 5-yrs and 10-yrs overall survival probabilities for each of the three age categories. Use the log-rank test to test for differences in survival between the three age categories.

4. Use the Cox model to estimate the hazard ratios and 95% confidence intervals of the age groups 20-40 and >40 with respect to <=20. What is your conclusion with respect to overall survival?

5. Use the naive Kaplan-Meier (one-minus survival) to calculate 5-yrs and 10-yrs probabilities of death for the oldest age group, for each of the four causes of death. Add up the four probabilities. What is the total? Does it correspond to the overall (one minus) survival probability calculated in 3?

We are going to use the competing risks macro to calculate cumulative incidence probabilities for each of the three age groups. Save the macro and the example SPSS syntax file calling the macro to your working directory that also contains the data. The example SPSS syntax file contains already the code needed to call the macro for this problem; only the full path to your working directory should replace the path provided in the example syntax file.

After running the macro, you will get a new SPSS data file containing the results. The first column contains the group used. To see the results of the oldest age group you should scroll down. The column survmon contains the time, risk and totevent contain the number of patients at risk and the total number of events (for all causes of failure). Then, for each cause of failure, four columns follow, containing the cumulative incidence, its standard error, and lower and upper bound of the 95% confidence interval for the cumulative incidence. Go the variable view and select three decimals for the cumulative incidences.

6. Read off the 5-yrs and 10-yrs cumulative incidence probabilities of the four causes of death. Add them up. What is the total? Does it correspond to the overall survival probability calculated in 3?

Poisson regression

We will apply Poisson regression and look at the relation with Cox regression, using the PBC data.

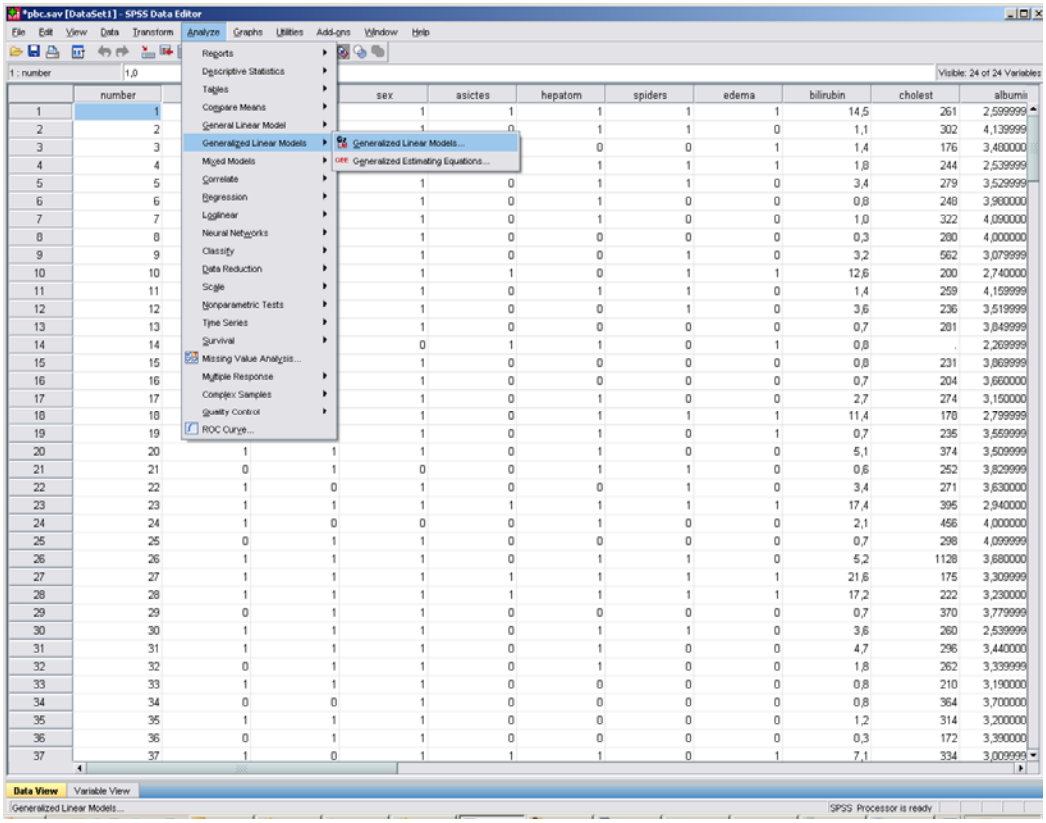
7. Open the PBC data, and make a Kaplan-Meier plot and a plot of the cumulative hazard. Write down the 3-yrs, 5-yrs, and 10-yrs survival probabilities (for later use). Do these plots agree with the assumption of a constant hazard? If you draw a straight line through the first 8 years (after that the cumulative hazard estimate may be too uncertain), what would the slope be?

8. Another way of estimating the constant hazard rate would be using the formula d/T, where d is the number of deaths, and T the total follow-up time in years. Calculate d and T, and the ratio.

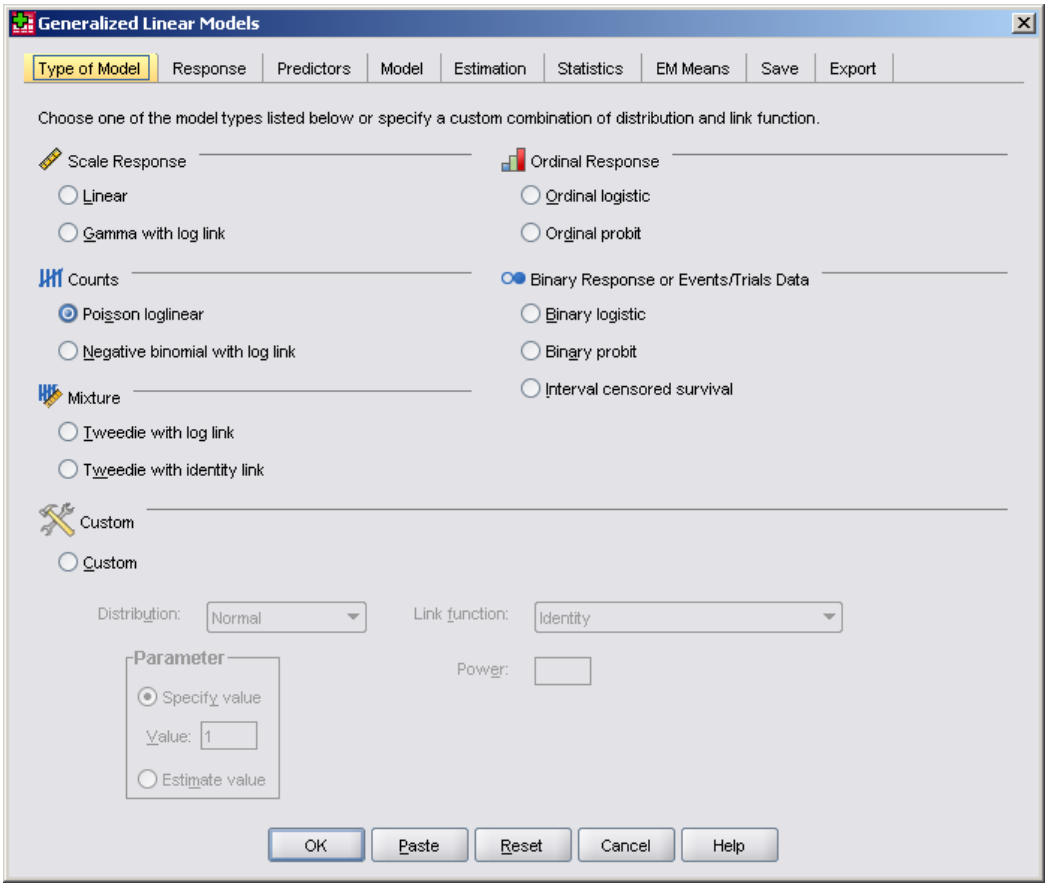
   Hint: You can use Analyze -> Descriptive Statistics -> Descriptives to calculate the total follow-up time in years. Select years as variables and under Options, select Sum.

Yet another way of estimating the constant hazard rate is using Poisson regression.

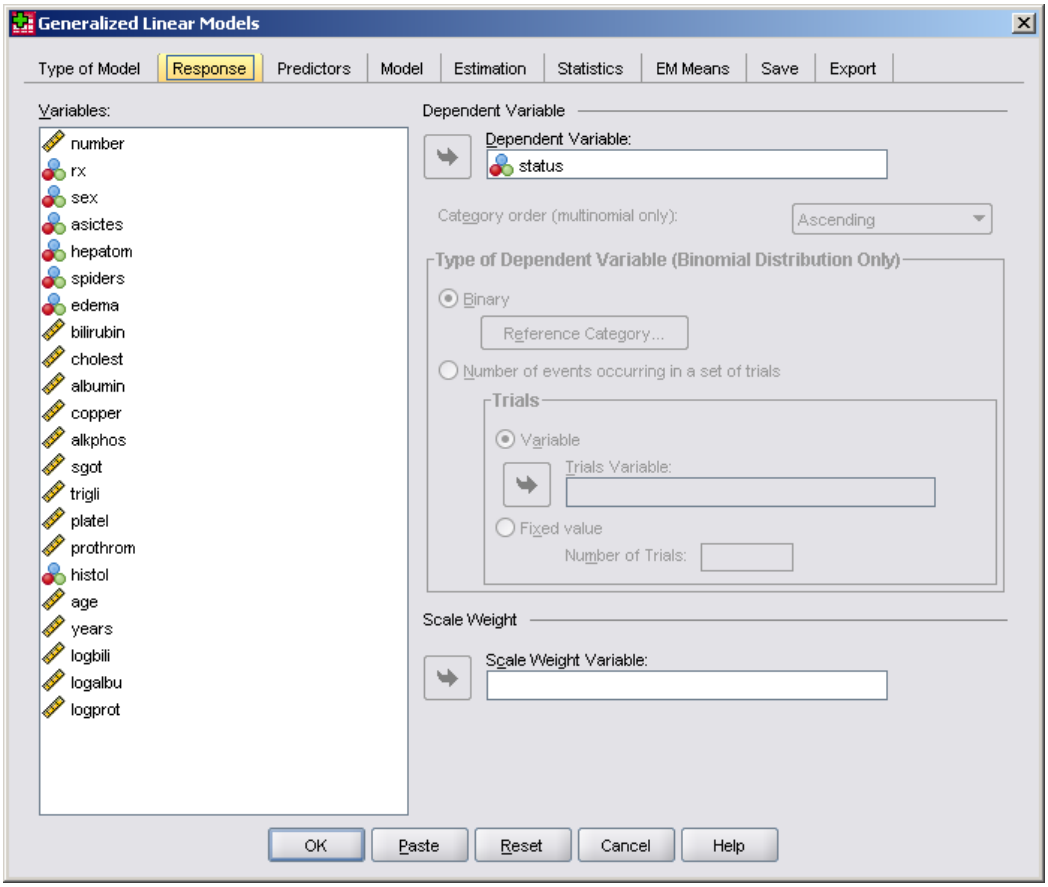
Choose Analyze -> Generalized Linear Models -> Generalized Linear Models.



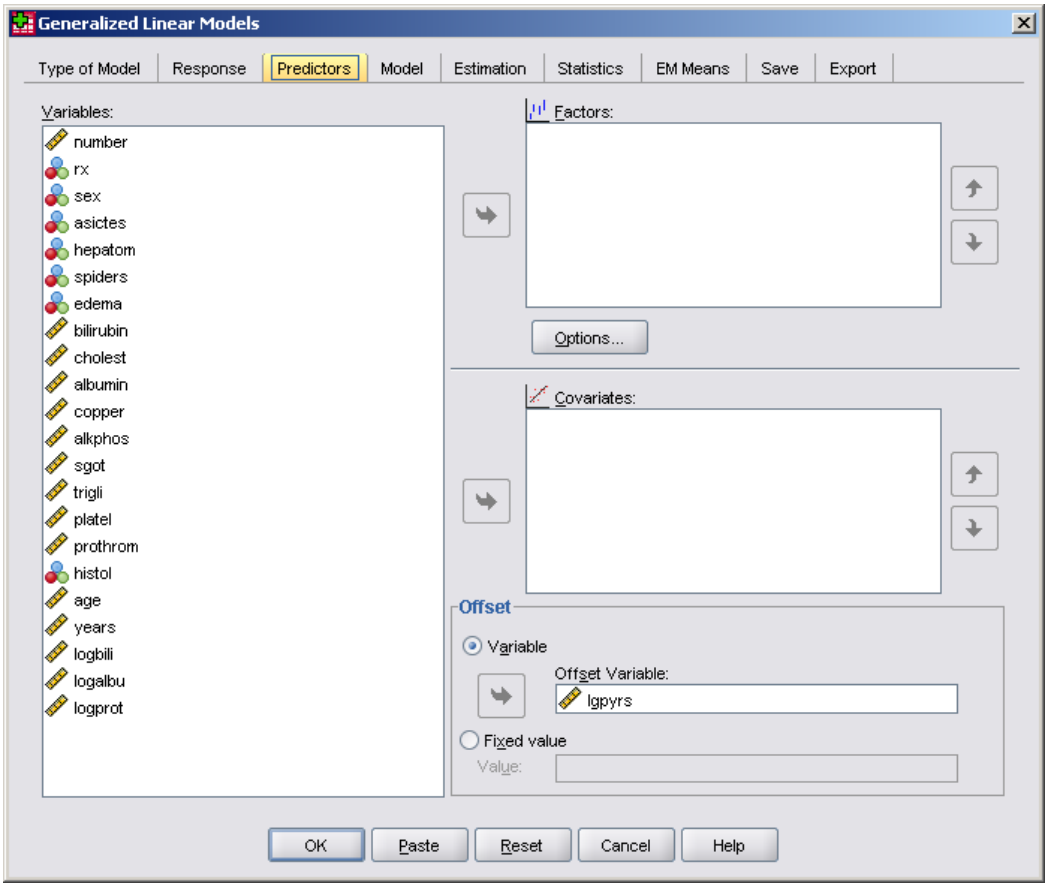
For type of model, choose Poisson loglinear. Click OK.



For Poisson regression, the number of events (in this case, simply 0 or 1, but in general this could also be 2 or more) is the outcome. On the Response tab, click status.



We will not include any covariates yet, but estimate only a single constant rate. Make a new variable lgpyrs, which is the natural log (LN) of the follow-up time (years); use Transform -> Compute for this. Use this new variable as offset.



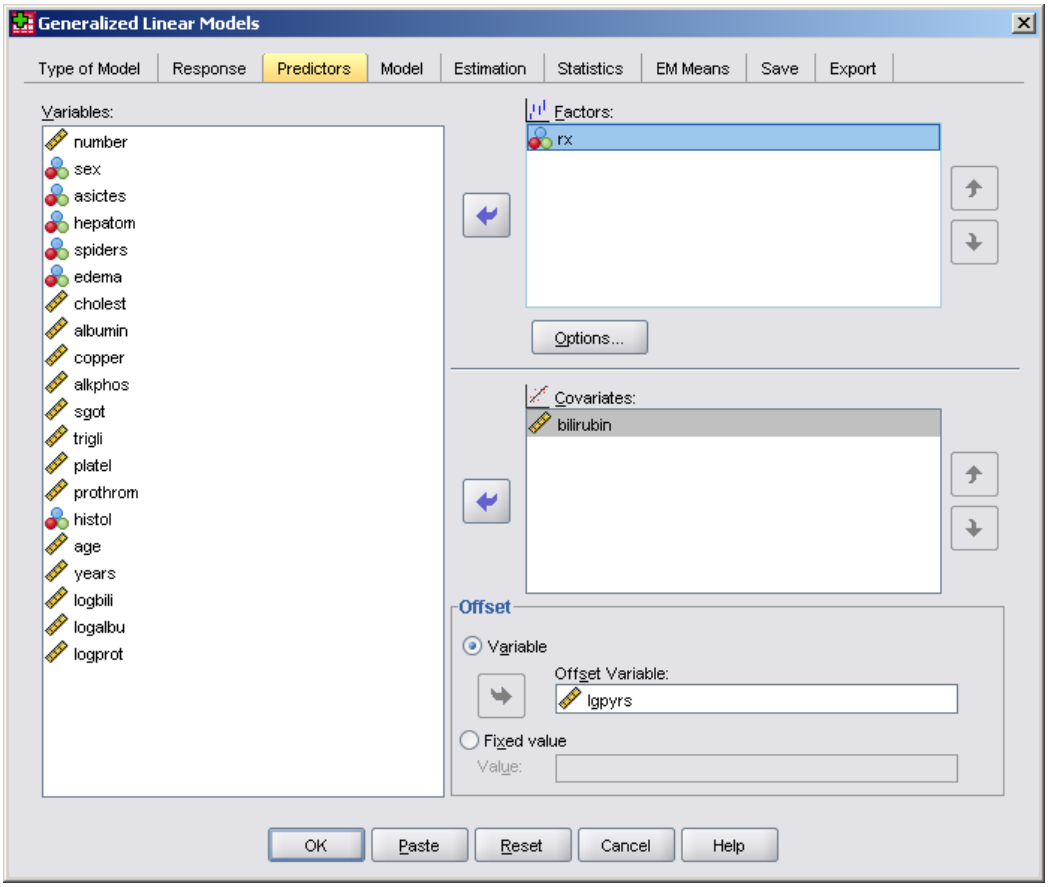
For now that is it; we'll keep the default values of everything else. Simply click OK. (You get a warning that you may ignore.)

9. Study the result. Look at the table with parameter estimates. The intercept regression coefficient is the (natural) log of the hazard rate. What is the estimate of the hazard rate? Construct a 95% confidence interval of the hazard rate. Compare the value with the previous exercise.

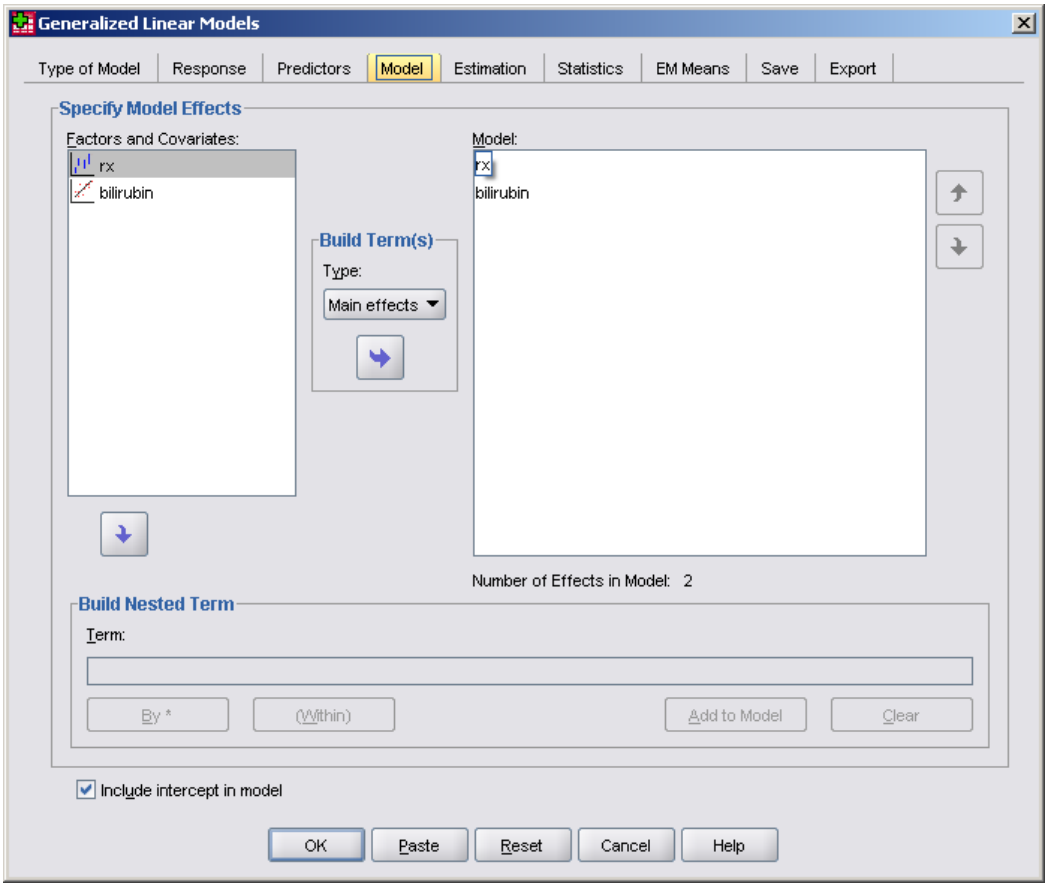
10. Based on the estimated hazard rate of the previous exercise, calculate the survival probabilities at 3, 5 and 10 years

Poisson regression is of course not really necessary for calculating a constant hazard rate. It starts to become useful when other covariates are also considered. We will look at one of the models considered yesterday, with treatment and bilirubin.

11. Repeat the Poisson regression, but now also include treatment as factor and bilirubin as covariate in the Predictors tab. The reason for including bilirubin as covariate rather than factor is that bilirubin is a continuous covariate.



In the Model tab, add treatment and bilirubin as main effects to the model.



What is your conclusion based on the output?

12. What is the hazard rate for an individual with a bilirubin value of 15, treated with DPCA? What is the predicted probability of surviving 5 years for such an individual? How large is this probability for an individual with a bilirubin value of 15, without treatment with DPCA?

13. Compare your results with the Cox regression results obtained yesterday. What is the difference of the Poisson regression model with the Cox model in terms of underlying assumptions? Compare estimated regression coefficients and standard errors. What do you notice?