Estimating relative survival models with the S+ software *LM* function

Remontet L, Bossard N, Belot A, Esteve J

Dept. of Biostatistics, Hospices Civils de Lyon, Université Claude Bernard (CNRS UMR 55 58), Lyon, France

Background

Relative survival from a given disease has been defined as the ratio of the observed survival over the expected survival in the absence of disease. This definition is equivalent to an additive model for the corresponding mortality rates. Several methods of estimation exists and regression models have been proposed to study the influence of covariates on relative survival [1-2].Unfortunately, none of these methods have been implemented in standard softwares in the form of routine procedure, with the exception of recent Giorgi function [3].

Objective:

Estimating relative survival model of Esteve &al using a simple algorithm which can be readily implemented in any standard software equipped with a weighted linear least square regression procedure.

Method

The proposed algorithm takes benefit from the identity, up to a constant, of the survival and Poisson likelihood. When the baseline hazard is taken as a step function in Esteve model, Dickman and Bolard noted that the splitting of individual observation in sub intervals, where the baseline rate is constant enables simple modeling especially when time dependent covariate are needed [2;4]. Dickman showed that the maximum likelihood relative survival estimates can be obtained from a GLM with Poisson error and implemented this idea from a SAS and STATA GLM procedure. We have implemented this methodology using an iteratively reweighted least squares procedure.

Description of the algorithm

1.Ingredient: T_i, δ_i survival data for individual i; Z(t) vector of covariates including dummy's for the time axis partition; β vectors of parameters to be estimated for the survival model $\lambda(t,z,x)=e^{\beta z(t)}+\mu(x+t,z)$ where μ is the expected hazard function known from adequate lifes tables and x is age at diagnostic. Tik, δ_{ik}, τ_{ik} are the values computed for each interval k after the splitting of the data.

2.Algorithm

$$t_{ik} \leftarrow T_{ik} - T_{ik-1}$$

$$\beta \leftarrow \text{initial}\beta;\ y \leftarrow y(\text{initial}\beta);\ \text{lik} \leftarrow 0;\ \text{oldlik} \leftarrow 0;\ \Delta\text{lik} \leftarrow 1;\ \Delta\beta \leftarrow 1$$

$$\textit{WHILE}(\Delta\text{lik} > \varepsilon \ \textit{OR}\ \Delta\beta > \varepsilon)\ \textit{DO}$$

$$\{\tau_{ik} \leftarrow e^{\beta z_{ik}} + \mu_i$$

$$\text{lik} \leftarrow \sum_i \sum_k t_{ik}(\tau_{ik} - \mu_i) + \delta_{ik}\log(\tau_{ik})$$

$$w_{ik} \leftarrow \frac{(\tau_{ik} - \mu_i)^2 t_{ik}}{\tau_{ik}}$$

$$y \leftarrow y + \frac{\delta_{ik} - \tau_{ik} t_{ik}}{(\tau_{ik} - \mu_i) t_{ik}}$$

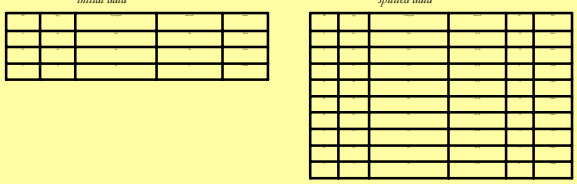
$$(y,\beta) \leftarrow \text{regress}(y \sim x, \text{weight} = w_{ik})$$

$$\Delta\text{lik} \leftarrow \text{lik} - \text{oldlik};\ \Delta\beta \leftarrow \|\beta - \text{old}\beta\|;\ \text{oldlik} \leftarrow \text{lik};\ \text{old}\beta \leftarrow \beta$$

$$\}$$

Example of splitting procedure

in a 5 years survival analysis



Example using S+

We used lm(y~x, weight=w_{ik}) as the regression procedure. We studied the prognostic value of age at diagnostic and of Cathepsin D, a tumour proteolytic factor, in a retrospective hospital cohort of 771 patients with breast cancer. The effects of these two continuous covariates were modelled using regression spline (piecewise polynomials of degree 3, with one knot corresponding to the median value). Figure 1 and figure 2 shows respectively the effects of age and Cathepsin on the log mortality hazard ratio (adjusted on tumour size and nodal status).

Figure 1: Log mortality hazard ratio according to age at diagnostic

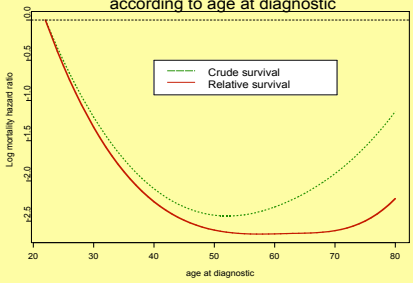
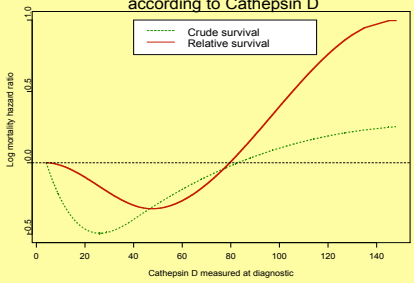


Figure 2: Log mortality hazard ratio according to Cathepsin D



Conclusion

When the baseline hazard is a step function, sophisticated modeling of relative survival can be carried out in any software equipped with a weighted linear least square procedure. Time dependent covariable analysis can be easily done by an adequate splitting of the date. Non proportional excess hazard can be evaluated by testing interaction between interval interval and the variable of interest.

References:

1)Esteve J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: element for further discussion. *Statistic in Medicine* 1990;**9**;529-538

2) Dickman PW, Slogett A, Hills M, Hakulinen T. Regression models for relative survival. *Statistics in Medicine* 2004;**23**:51-64

3) Giorgi R, Abrahamowicz M, Quantin C &al. A relative survival regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine* 2003;**22**:2767-2784

4) Bolard P, Quantin C, Esteve J &al Modelling time-dependent hazard ratios in relative survival. Application to colon cancer. *Journal of clinical Epidemiology* 2001;**54**:986-996