RSURV: a function to perform relative survival analysis with S-PLUS or R

Roch Giorgi *, Julie Payan, Joanny Gouvernet

LERTIM, Faculté de Médecine, Université de la Méditerranée, 27 Bd Jean Moulin
13385 Marseille Cedex, France.
* E-mail address: roch.giorgi@ap-hm.fr

SUMMARY
Relative survival is a method used to estimate net survival using the expected mortality in the general population. This method is frequently used in cancer registries, more particularly with the Esteve et al. regressive proportional hazards model. Recently, extensions of this model have been developed to account for time-dependent covariate and for time-dependent hazards using B-spline functions. We propose a function, RSurv, to take into account these extensions. Written in the R/S language this function has the same structure of the standard Cox function coxph of R and S-PLUS software with the goal to homogenise survival functions and to take advantages of the power of R and S-PLUS software. We also propose a function, plot.RSurv, for plotting relative survival curves and time-dependent hazards ratio. The usage of these functions is exemplified by a study of a breast cancer hospital-based data set.

Key words: Relative survival; Proportional hazards models; Non-proportional hazards models; Time-dependent covariates; Survival analysis.

1 Introduction
Relative survival, a method used to estimate net survival, is commonly used for cancer patient survival analysis [e.g. Ref. 1-3]. Different relative survival regression models, based on the assumption of proportional hazards (PH) for disease-related mortality, have been developed [4-7]. This assumption constrains the hazards ratio (HR), which
RSURV : A function to perform relative survival analysis with S-PLUS or R

describes the effect of a prognostic factor on survival, to be constant over time. It is well
known that this assumption may be too strong and several extensions of the Esteve et al.
PH model [4] have been proposed to estimate time-dependent hazards ratio [8-10]. On
the other hand, others extensions of the Esteve et al. PH model are possible, such as
taking time-dependent covariates into account (i.e. recurrence of a disease, change in the
subject’s treatment.).

We propose a new function, RSurv, written in the R/S language, to analyze
proportional relative survival using the Esteve et al. model [4], with the possibility to
account also for both time-dependent covariate and non-proportional hazards.

2 Relative survival models implemented

In the Esteve et al. PH model [4], the observed hazard function for total mortality, \( \lambda_t \), at
time \( t \) after diagnosis of an individual age \( a \) at diagnosis and with a vector of covariate
\( z \), which could contain age, is expressed as:

\[
\dot{\lambda}_t(t, z, a) = \lambda_e(t + a, z) + \sum_{k=1}^{r} \tau_k I_k(t) \exp(\beta z)
\]

where \( \lambda_e \), which depends only on \( z \), a subvector of \( z \), is the expected mortality
hazard function for overall mortality in a general population and is obtained from
relevant mortality statistics using external sources, generally quantified based on
published age- and sex-specific mortality rates. Terms on the right side of the above
equation represent the disease-related mortality hazard function, where \( \tau_k \) is the
disease-related baseline mortality hazard in the \( k^{th} \) time-segment for patients with \( z = 0 \),
\( I_k \) is an indicator function of the time-segment (\( I_k = 1 \) if \( t_{k-1} < t < t_k \), and 0 otherwise),
and \( \beta \) is the log hazards ratio (HR) corresponding to the covariates.

Taking time-dependent covariate into account is easily achievable using the counting
process approach, which has proved, in the framework of the Cox model [11], to be a
powerful theoretical tool [12]. For each subjects having a modification of his time-
dependent covariate status, we just have to consider new intervals of risk determined by
the time of his status change. In this way, each subject \( i \) is represented by a set of
observations delimited by the time in which the time-dependent status has changed with
the corresponding death or censor and covariate status.

Published in: Computer Methods and Programs in Biomedicine 2005; 78:175-178.
Time-dependent HR is implemented using quadratic B-spline functions with two interiors knots, as proposed by Giorgi et al. and described in ref. [10]. This model makes it possible to adjust effects of both HR having time-dependent effects, using smooth functions of time, and HR having constant effects on disease-specific mortality.

3 Program description

The RSurv function is written in the R/S language. In order to homogenise survival function, its structure is similar to that of the standard coxph function for fitting Cox model [11] using R or S-PLUS software. Several simple options allow the user to specify the Esteve et al. PH model, with or without time-dependent covariate, or the Giorgi et al. model for which parameter estimates based on maximum likelihood should be obtained. The most important arguments of the function RSurv are the following:

- **formula**: a formula object of the Surv R and S-PLUS function containing the time and censoring indicator status on the left of a ‘~’ operator and the covariates to be fitted on the right (e.g. formula=Surv(time, status)~X1+X2+X3). In a time-dependent covariate framework, and as with the coxph function, we simply have to replace Surv(time, status) with Surv(start, stop, status), where (start, stop] is the interval of risk, open on the left and closed on the right, of each subject, delimited by the time of change in the time-dependent covariate status;

- **data**: a data frame in which to interpret the variables named in the formula;

- **ratedata**: a data frame of the hazards mortality in general population;

- **bsplines**: its presence implies the use of the Giorgi et al. model (by default, the Esteve et al. model is used). It corresponds to a vector of logical values indicating which covariates have a time-dependent HR (e.g. bsplines=c(T, F, T) indicates assumption of time-dependent HR for X1 and X3, and a PH effect for X2);

- **interval**: a vector indicating the location of time-segment for the Esteve et al. model, or the location of the knots for the Giorgi et al. B-splines model. Determination of the intervals might be user’s defined or automatically computed, using the “NA” notation, according to the quantile of the distribution of deaths (e.g. interval=c(0,NA,NA,60));
RSURV: A function to perform relative survival analysis with S-PLUS or R

- covtest: a vector of logical values indicating for which covariates a test is required (loglikelihood ratio, Wald, and score tests of no effect for the Esteve et al. model; loglikelihood ratio test of PH effect for the Giorgi et al. model).

The function RSurv return an object of class “esteve.ph” or “giorgi.tdph”, according to the choose model, representing the fit. Objects of this class have methods for the function plot.RSurv used to plot results of the function RSurv (i.e. relative survival curves for the Esteve et al. model or log HR functions for the Giorgi et al. model).

4 RSurv function usage example

Usage of the function RSurv is illustrated by data for 2,764 incident cases of breast cancer. We considered four prognostic factors, coded as binaries covariates: age (below 34 years, 35 – 49, referent category, 50 - 69 and 75 or older), tumour size (T: less than 2 cm, referent category, greater than 2 cm), node status (N: absence, referent category, or presence of nodes at diagnosis), metastasis status (M: absence, referent category, or presence of metastasis at diagnosis).

Table 1 Output of the function RSurv when all the covariates have a PH effect (Esteve et al. model): log hazards ratio (HR) of the covariates and the coefficients of the disease-related baseline mortality hazard (with 6 time-segments, in months), with theirs standard errors, their 95% confidence limits, the z-statistics and the correspondents p values.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Log(HR)</th>
<th>Standard Error</th>
<th>95% Log(HR) Confidence Limits</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age34</td>
<td>1.0703</td>
<td>0.2104</td>
<td>6.5e-01 1.482</td>
<td>5.087</td>
<td>3.6e-07</td>
</tr>
<tr>
<td>Age50.69</td>
<td>0.1210</td>
<td>0.1514</td>
<td>-1.7e-01 0.417</td>
<td>0.799</td>
<td>4.2e-01</td>
</tr>
<tr>
<td>Age70</td>
<td>-0.3959</td>
<td>0.1514</td>
<td>-1.1e-00 0.286</td>
<td>-1.136</td>
<td>2.6e-01</td>
</tr>
<tr>
<td>Tumour size</td>
<td>1.0321</td>
<td>0.1741</td>
<td>6.9e-01 1.373</td>
<td>5.926</td>
<td>3.1e-09</td>
</tr>
<tr>
<td>Node</td>
<td>1.0299</td>
<td>0.1773</td>
<td>6.8e-01 1.377</td>
<td>5.807</td>
<td>6.4e-09</td>
</tr>
<tr>
<td>Metastasis</td>
<td>2.0814</td>
<td>0.1864</td>
<td>1.7e-00 2.446</td>
<td>11.167</td>
<td>0.0e+00</td>
</tr>
<tr>
<td>[0 - 12]</td>
<td>0.0007</td>
<td>0.0003</td>
<td>-3.7e-06 0.001</td>
<td>1.950</td>
<td>5.1e-02</td>
</tr>
<tr>
<td>[12 - 24]</td>
<td>0.0036</td>
<td>0.0009</td>
<td>1.7e-03 0.005</td>
<td>3.704</td>
<td>2.1e-04</td>
</tr>
<tr>
<td>[24 - 36]</td>
<td>0.0041</td>
<td>0.0011</td>
<td>1.8e-03 0.006</td>
<td>3.561</td>
<td>3.7e-04</td>
</tr>
<tr>
<td>[36 - 48]</td>
<td>0.0057</td>
<td>0.0015</td>
<td>2.6e-03 0.008</td>
<td>3.660</td>
<td>2.5e-04</td>
</tr>
<tr>
<td>[48 - 60]</td>
<td>0.0067</td>
<td>0.0018</td>
<td>3.1e-03 0.010</td>
<td>3.658</td>
<td>2.5e-04</td>
</tr>
<tr>
<td>[60 - 120]</td>
<td>0.0053</td>
<td>0.0012</td>
<td>2.8e-03 0.007</td>
<td>4.257</td>
<td>2.1e-05</td>
</tr>
</tbody>
</table>

Published in: Computer Methods and Programs in Biomedicine 2005; 78:175-178.
RSURV : A function to perform relative survival analysis with S-PLUS or R

Statements for an analysis with the Esteve et al. model, using all the covariates and 6
time-segments for the disease-related baseline mortality hazard, is as follow:
RSurv(Surv(time,status)~age34+age50.69+age70+T+N+M,
    interval=c(0,12,24,36,48,60,120),
    data=breast,ratedata=Mortal)

Part of the printed output can be seen from Table 1. In this model, the effect of age
greater than 50 years is non-significant, whereas all the others covariates have a
significant effect. To test the effect of all the ages categories, we just have to add
covtest=c(T,T,T,F,F,F) in the previous command. Then, RSurv return an
object with components for the loglikelihood ratio, Wald, score tests, and the degree of
freedom.

The presence of a time-dependent covariates will be treated in the same way as with the
R/S-PLUS coxph function using the Surv(start,stop,status) notation. User
would just have to add one more variable in the data file, in order to represent each
subject by a set of observations over its intervals at risk, with time intervals delimited by
the time of change in the time-dependent covariate status.

Modelling time-dependent HR with this data set could be handling in the following
way:
RSurv(Surv(time,status)~age34+age50.69+age70+T+N+M,
    bsplines=c(T,T,T,T,T,T),covtest=c(T,T,T,F,F,F),
    interval=c(0,NA,NA,120),data=breast,ratedata=Mortal)

which impose a TD effect for all the covariates, and test a TD effect for age. The
corresponding likelihood ratio test of TD for age is non significant (p = 0.97 with 12
degrees of freedom), and thus, according to the principle of parsimony, it is preferable
to impose a PH effect for age in the model. Other fits show a PH effect for tumour size
and metastasis status, and a TD effect for node status. Therefore, the fitted model
becomes a mixture of covariate having PH and TD effects:
fitTDPH<-
    RSurv(Surv(time,status)~age34+age50.69+age70+T+N+M,
        bsplines=c(F,F,F,F,T,F),interval=c(0,NA,NA,120),
        data=breast,ratedata=Mortal)

Estimates were obtained with the function Rsurv and Fig. 1 gives the plot, obtains
RSURV : A function to perform relative survival analysis with S-PLUS or R

with the function plot.RSurv, of the adjusted estimate of the log HR, with the 95% confidence limits, for the presence of node at diagnosis. The statement for plotting is

\[
\text{Plot.RSurv(fitTDPH,c(’N’),conf.int=T)}.
\]

![Graph](image)

**Fig. 1** Adjusted estimates of log hazards ratio for node prognostic factor (solid curve) and its 95 per cent confidence limits (dashed lines). Plot obtained with the function plot.RSurv.

5 Conclusion

We proposed a R/S-PLUS function which allows to fit the Esteve *et al.* regressive relative survival proportional hazards model, its extensions in case of time-dependent covariate and time-dependent hazards ratio. The structure of this function is similar to that of the R/S-PLUS *coxph* function for the Cox model with the goal to homogenise survival functions, and to take advantages of the power of R and S-PLUS software. The RSurv function, its associated plot.RSurv function for plotting relative survival curves and time-dependent HR, are proposed in open sources and can be downloaded free of charge from the Web site of the Laboratory for Education and Research in Medical Information Processing (LERTIM) at the Faculty of Medicine of Marseille, France (http://cybertim.timone.univ-mrs.fr/CybErtim/LERTIM/Default.htm, in the MEDuS project of the Research directory).
Acknowledgements

We thank Pr Pascal Bonnier, Department of Gynecology and Obstetrics, Hospital of the Conception, Marseille, France, for providing data of its hospital-based registry for illustrative example, and Dr Terry Therneau for his kindness authorisation to use the basic structure of its \textit{coxph} function.

References


R. Giorgi et al.
RSURV: A function to perform relative survival analysis with S-PLUS or R

[12] Therneau T, Grambsch PM. Modeling survival data: Extending the Cox model.