

FRAILTYPACK: An R package for the analysis of correlated survival data with a general class of frailty models

V Rondeau¹ Y Mazroui¹ A Mauguen¹ A Diakité¹
JR Gonzalez Ruiz²

¹INSERM CR 897, Université Bordeaux2 - ²CREAL, Barcelona, SPAIN

1 ères rencontres R, Bordeaux, 2-3 Juillet 2012

Motivation

- Survival data : times to event
- **Recurrent events** : naturally ordered failure time data
Ex : cancer relapses, rehospitalizations
- **Clustered observations** into groups
Ex : meta-analysis, multicentric cohort

⇒ **Correlated data** in survival analysis

Statistical approach

- **Frailty models** : extensions of the Cox proportional hazards model
- 4 types of frailty models : **shared, nested, joint, additive**
- **Right-censored** or **left-truncated** data are considered.
- **Stratification** (on 2 modalities)
- **Gap** times or **calendar** times
- Time-independent or Time-dependent covariates
- Estimation : maximization of a **penalized likelihood**
Smooth hazard or survival functions (approximation with splines) $pl(\phi) = l(\phi) - \kappa \int_0^\infty \lambda_0''^2(t) dt$
- or estimation with simple maximisation of the likelihood (**parametrical approach**)

FRAILTYPACK

FRAILTYPACK : R library

<http://cran.r-project.org/web/packages/frailtypack/>

1st version : 2005

Cox proportional hazard model or shared frailty model

2nd version : 2009

Addition of Nested, Additive or Joint frailty models

3rd version : 2012

Addition of parametrical estimation, and concordance index

Free access : Windows, Unix and MAC OS X

Cox proportional hazard model

$$\lambda_i(t) = \lambda_0(t) \exp(\beta' \mathbf{X}_i)$$

```
fit.COX <- frailtyPenal(Surv(time,event) ~ var1 + var2 +  
cluster(id), n.knots=12, kappa1=10000, data=database,  
Frailty=FALSE)
```

Shared frailty model

- **Recurrent event times** ($j=1, \dots, n_i$) shared by each subject ($i=1, \dots, G$)
- Observations ($j=1, \dots, n_i$) **clustered** into groups ($i=1, \dots, G$)

$$\lambda_{ij}(t|z_i) = \lambda_0(t) z_i \exp(\beta' \mathbf{X}_{ij})$$

$$Z \sim \Gamma(1/\theta; 1/\theta)$$

with $E(Z) = 1$ and $\text{var}(Z) = \theta$

Shared frailty model : R application

```
frailtyPenal(formula = Surv(time, event) cluster(id) +  
as.factor(dukes) + as.factor(charlson) + sex + chemo,  
data = readmission2, Frailty = TRUE, cross.validation =  
TRUE, n.knots = 10, kappa1 = 1)
```

Shared Gamma Frailty model parameter estimates
using a Penalized Likelihood on the hazard function

	coef	exp(coef)	SE	coef (H)	SE	coef (HIH)	z	p
dukes.C	0.297572	1.346586		0.160814		0.160773	1.85041	6.4254e-02
dukes.D	1.056007	2.874867		0.194686		0.194668	5.42414	5.8234e-08
charlson.1-2	0.451561	1.570763		0.258891		0.258890	1.74422	8.1122e-02
charlson.3	0.410276	1.507234		0.137003		0.137002	2.99466	2.7475e-03
sexFemale	-0.537752	0.584060		0.138793		0.138701	-3.87449	1.0685e-04
chemoTreated	-0.206521	0.813409		0.143000		0.142862	-1.44420	1.4868e-01

	chisq	df	global	p
dukes	30.16313	2	2.82e-07	
charlson	3.04229	1	8.11e-02	

Frailty parameter, Theta: 0.672346 (SE (H): 0.142286) (SE (HIH): 0.142268)

```
penalized marginal log-likelihood = -3236.42  
LCV = the approximate likelihood cross-validation criterion  
in the semi parametrical case = 3.78066  
n= 861  
n events= 458 n groups= 403  
number of iterations: 9  
Exact number of knots used: 10  
Best smoothing parameter estimated by  
an approximated Cross validation: 208159031, DoF: 11.00
```


Shared frailty model : C measures

Compute **concordance probability estimation** for Cox or shared frailty models with clustered data
= Among 2 patients randomly chosen, the one who has the shorter predicted survival time will be the one with the shorter observed survival time

Cmeasures(fit.sharedFM)

```
----- Model -----  
Surv(time, event) ~ cluster(id) + as.factor(dukes) + as.factor(charlson) +  
sex + chemo  
  
----- Frequencies -----  
Number.patients Number.events Number.groups  
861 458 403  
  
----- Concordance Probability Estimation (conditional) -----  
  
Concordance            Between    Within    Overall  
Number pairs 368919.000 1311.000 370230.000
```

Nested frailty model

Two-level hierarchical structure

Examples of studies design :

- Subjects \in Families \in Cities
- Recurrent observations \in Subjects \in Cities

$$\lambda_{ijk}(t|v_i, w_{ij}) = \lambda_0(t) v_i w_{ij} \exp(\beta' \mathbf{X}_{ijk}) \quad (1)$$

v_i cluster random effects

w_{ij} sub-cluster random effects

Nested frailty model : R application

```
frailtyPenal(Surv(t1,t2,event)~cluster(gp)+subcluster(subgp)  
+cov1+cov2,data=dataNested,n.knots=8,kappa1=10000)
```

```
left truncated structure used
```

```
Nested Frailty model parameter estimates using  
a Penalized Likelihood on the hazard functions
```

	coef	exp(coef)	SE coef	(H)	SE coef	(HIH)	z	p
cov1	-0.558	0.572	0.133		0.133		-4.19	2.8e-05
cov2	1.274	3.575	0.149		0.148		8.56	0.0e+00

```
Frailty parameters:
```

```
alpha (group effect): 0.373 (SE(H):0.107) (SE(HIH):0.107)  
eta (subgroup effect): 0.110 (SE(H):0.0727) (SE(HIH):0.0726)
```

```
penalized marginal log-likelihood = -1020.88  
n= 400  
n events= 287 n groups= 20  
number of iterations: 22  
Exact number of knots used: 8  
Value of the smoothing parameter: 10000
```

Joint frailty model

- **Dependence** between **recurrent events** and a **terminating event**
- **Informative censoring** of recurrent events by death (ex : Censoring by death may be associated to worsening of symptoms)

$$\begin{cases} r_{ij}(t|\omega_i) = \omega_i r_0(t) \exp(\beta_1' \mathbf{Z}_{ijr}) & \text{(recurrent)} \\ \lambda_i(t|\omega_i) = \omega_i^\alpha \lambda_0(t) \exp(\beta_2' \mathbf{Z}_{id}) & \text{(death)} \end{cases} \quad (2)$$

Joint frailty model : R application

```
frailtyPenal(Surv(t.start,t.stop,event)~cluster(id)  
+as.factor(dukes)+sex+chemo  
+terminal(death),formula.terminalEvent= sex+chemo,  
data=readmission,hazard="Piecewise-per",  
nb.int1=7,nb.int2=6,  
Frailty=TRUE,joint=TRUE,recurrentAG=TRUE)
```

Joint frailty model : R application

left truncated structure used

Joint gamma frailty model for recurrent and a terminal event processes
using a Parametrical approach for the hazard function

Recurrences:

-----	coef	exp(coef)	SE	coef (H)	z	p
dukes.C	0.206467	1.229327		0.195331	1.05701	0.29051000
dukes.D	0.847687	2.334241		0.229447	3.69447	0.00022034
sexFemale	-0.582572	0.558460		0.167779	-3.47226	0.00051610
chemoTreated	-0.238974	0.787436		0.186810	-1.27923	0.20082000

	chisq	df	global	p
dukes	14.7664	2	0.000622	

Terminal event:

-----	coef	exp(coef)	SE	coef (H)	z	p
sexFemale	-0.291431	0.747194		0.266475	-1.09365	0.27411
chemoTreated	0.285226	1.330063		0.283601	1.00573	0.31455

Frailty parameters:

theta (variance of Frailties, Z): 1.12568 (SE (H): 0.110214))
alpha (Z^{alpha} for terminal event): 1.09677 (SE (H): 0.207369)

marginal log-likelihood = -4297.75

AIC = Aikaike information Criterion = 5.01597

The expression of the Aikaike Criterion is:

'AIC = (1/n)[np - l(.)]'

n= 861

n recurrent events= 458 n groups= 403

n terminal events= 109

number of iterations: 14

Exact number of time intervals nb.int1 used: 7

Exact number of time intervals nb.int2 used: 6

Additive frailty model

Two main sources of intertrial heterogeneity in survival data :

- **heterogeneity of the baseline risk**
due to differences in trial design, in treatment protocols, medical practices or in patient populations
- **heterogeneity of treatment effects across trials**
reflects differences in patient characteristics and in implementation of the protocol

$$\lambda_{ij}(t|u_i, v_i) = \lambda_0(t) \exp(u_i + v_i X_{ij1} + \sum_{k=1}^p \beta_k X_{ijk}) \quad (3)$$

$$u_i \sim \mathcal{N}(0, \sigma^2) \quad v_i \sim \mathcal{N}(0, \tau^2) \quad \text{cov}(u_i, v_i) = \rho\sigma\tau$$

Additive frailty model : R application

Another function in the R package FRAILTYPACK :
`additivePenal(Surv(t1,t2,event)~cluster(group)
+var1+var2+slope(var1),cross.validation=TRUE,
data=dataAdditive,correlation=TRUE,n.knots=10,kappa1=1)`

Additive gaussian frailty model parameter estimates
using a Penalized Likelihood on the hazard function

	coef	exp(coef)	SE coef	(H)	SE coef	(HIH)	z	p
var1	-0.1901	0.827	0.0618		0.0618		-3.076	0.0021
var2	-0.0212	0.979	0.0246		0.0246		-0.862	0.3900

Covariance (between the two frailty terms,

the intercept and the slope): -0.227 (SE: 0.047)

Corresponding correlation between the two frailty terms : -0.677

Variance for random intercept: 0.358 (SE (H): 0.0567) (SE (HIH): 0.0567)

Variance for random slope: 0.314 (SE (H): 0.0546) (SE (HIH): 0.0546)

penalized marginal log-likelihood = -30125.31

n= 10000

n events= 6887 n groups= 100

number of iterations: 7

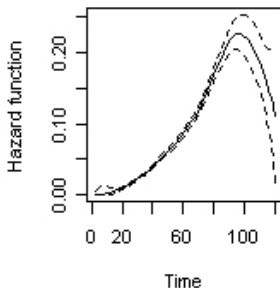
Exact number of knots used: 10

Smoothing parameter estimated by Cross validation: 16735

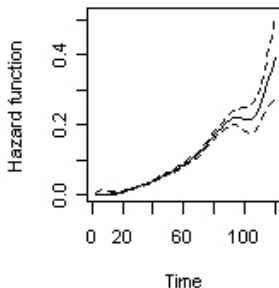
Additive frailty model : R application

```
plot(modAdd2cov_withoutCorr,type.plot="hazard",main="Correlation=False",conf.bands=TRUE)  
plot(modAdd2cov_withCorr,type.plot="hazard",main="Correlation=True",conf.bands=TRUE)
```

Correlation = False



Correlation = True



Perspectives for FRAILTYPACK

- Regular improvement of the R package
- More **general joint frailty** models (with 2 frailties, with time-dependent effects, for clustered data)
- propose a **dynamic prediction** tool
- include **interval-censoring** option
- include **log-normal** option for random effects

Ongoing Collaboration : the US-FDA has agreed that FRAILTYPACK can be used for primary analysis of a phase 3 trial on recurrent hospitalizations for severe heart failure

Comments / bugs more than welcome !

References

- Rondeau, V., Mazroui, Y. and Gonzalez, J.R. FRAILTYPACK : An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation. *J Stat Software*, 2012
- Duchateau, L. and Janssen, P. The frailty model. *Springer* 2008
- Rondeau, V., Michiels, S., Liqueur, B. and Pignon, J.P. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Stat Med* 2008 ;27 :1894–1910.
- Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. and Soubeyran, P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation : application on cancer events. *Biostatistics* 2007 ;8 :708–721.
- Cook, R.J. and Lawless, J.F. The statistical analysis of recurrent events. *Springer* 2007
- Rondeau, V., Filleul, L. and Joly, P. Nested frailty models using maximum penalized likelihood estimation *Stat Med* 2006 ;25 :4036–4052.
- Rondeau V, and Gonzalez, JR. Frailtypack : A computer program for the analysis of correlated failure time data using penalized likelihood estimation. *Comput meth prog biomed* 2005 ;80 :154–164.

<http://cran.r-project.org/web/packages/frailtypack/>