Mathematical and Computer Modelling 52 (2010) 949–955

Contents lists available at ScienceDirect

Mathematical and Computer Modelling

journal homepage: www.elsevier.com/locate/mcm

# Statistical models for recurrent events and death: Application to cancer events

V. Rondeau *

*INSERM, CR897 (Biostatistic), Bordeaux, F-33076, France*
*Université Victor Segalen Bordeaux 2, Bordeaux, F-33076, France*

ARTICLE INFO

*Article history:*
Received 28 September 2009
Received in revised form 3 February 2010
Accepted 4 February 2010

*Keywords:*
Correlated survival times
Frailty models
Cure frailty model
Nested frailty model
Joint frailty model
Penalized likelihood

ABSTRACT

In many biomedical studies, subjects may experience the outcome of interest more than once over a period of observation; outcomes of this sort have been termed recurrent events. A model that is becoming increasingly popular for modeling association between recurrent survival times is the use of a frailty model. In recent years a number of papers appeared, extending the survival models to models that are suitable to handle more complex survival data as recurrent events. We present here frailty model extensions to analyze recurrent events data: cure frailty models for a mixture of susceptible and insusceptible subjects for the event of interest; nested frailty models when data are clustered at several hierarchical levels and joint frailty models for the joint analysis of recurrent events and death. We performed a semi-parametric penalized likelihood approach to estimate the different parameters. Those different models can be fitted using the R package "frailtypack".



## 1. Introduction

Owing to the natural evolution of a disease, several events often arise after a first treatment for the same subject. For example, patients with a primary invasive breast cancer and treated with breast conserving surgery may experience breast cancer recurrences, metastases or death. Patients may also experience repeated epileptic seizures or cancer patients may experience recurrent superficial tumors or repeated episodes of hospitalization. The structure of recurrent events is that of naturally ordered failure time data and the different events 'within' an individual are correlated. Statistical frailty models have been proposed to analyze these recurrent event data [1].

The specific feature that makes survival analysis different from classical statistical analysis is data censoring. Typically, the survival time is unknown for some of the subjects; the only information available being that the subject has survived up to a certain time. Thereafter the subject is no longer followed up. This type of censoring is called right censoring. To analyze recurrent event data, the focus can be placed on time-between-event (i.e., gap times) or time-to-event models (i.e., calendar times) (see [2]). These two timescales, which reflect two important aspects of the data, can be linked to the semi-Markov models in which the transition probability between two states depends only on the waiting times, whereas in non-homogeneous Markov models this transition depends only on the time since inclusion in the study (see Fig. 1). Analyses based on waiting times are often useful when events are relatively infrequent, when some type of individual renewal occurs after an event, or when prediction of the time to the next event is of interest.

In Section 2 we recall the shared frailty model with the construction of the full penalized log-likelihood and an application on breast cancer recurrences. In Section 3 we describe the cure frailty models for a mixture of susceptible and insusceptible subjects for the event of interest. The nested frailty models are explained in Section 4 when data are clustered at several hierarchical levels. The joint frailty models for the joint analysis of recurrent events and death is described in Section 5.

* Corresponding address: INSERM, CR897 (Biostatistic), Bordeaux, F-33076, France. Tel.: +33 557574531; fax: +33 556240081.
*E-mail address:* Virginie.Rondeau@isped.u-bordeaux2.fr.

0895-7177/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mcm.2010.02.002

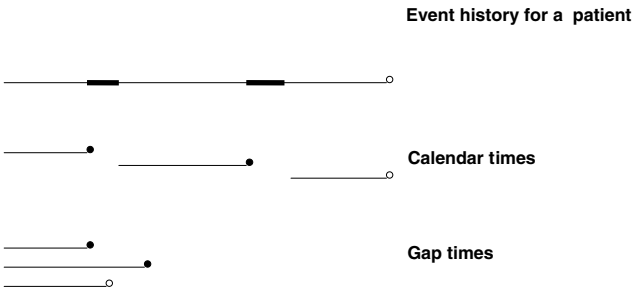


**Fig. 1.** Event history for a patient with recurrent events together with the calendar times and the gap times; • represents an event, ○ a censoring time and ■■ the non at risk periods.

## 2. Shared frailty model

### 2.1. The shared frailty model

The standard frailty model assumes all individuals eventually experience the event of interest with varying risks that are greater than zero [3]. The frailty models extend the Cox proportional hazards model [4] to account for unobservable heterogeneity among individuals.

We denote for subject $i$ ($i = 1, \ldots, N$), $X_{ij}$ the $j$th recurrent times ($j = 1, \ldots, n_i$), $C_i$ the censoring times. $T_{ij} = \min(X_{ij}, C_i)$ corresponds to each follow-up time and $\delta_{ij}$ is a binary indicator for recurrent events which is 0 if the observation is censored, and 1 if $X_{ij}$ is observed ($\delta_{ij} = I_{(T_{ij}=X_{ij})}$ where $I_{()}$ denotes indicator function). Let $Z_{ij} = (Z_{1ij}, \ldots, Z_{pij})$ be a vector of covariates (fixed or time-dependent) for individual $i$. We let $\omega_i$ random effects which follow a normal distribution with zero mean and variance $\theta^2$. The hazard at time $t_{ij}$ for an individual (with shared frailty $= \omega_i$) is of the form:

$$\lambda_{ij}(t_{ij} \mid \omega_i) = \lambda_0(t_{ij})\omega_i \exp(\beta' Z_{ij}), \tag{1}$$

where $\lambda_0$ is the baseline hazard function. The common random effect parameter $\omega_i$ takes into account the dependence between the successive events within a patient and better reflects the true clinical course of the disease in this heterogeneous population. Non-measured factors such as individual genetic predispositions or carcinogen exposure may lead to a heterogeneity among individuals that is a correlation between survival data of a particular patient. This model (1) corresponds to the survival function $S(t_{ij} \mid \omega_i) = \exp(-\Lambda_0(t_{ij})\omega_i \exp(\beta' Z_{ij}))$ where $\Lambda_0(.)$ is the cumulative baseline hazard function.

We introduce a semi-parametric penalized likelihood approach to estimate the different parameters: the regression coefficients, the variance of the random effects and the baseline hazard function $\lambda_0(t)$ [5]. In most situation it is reasonable to expect a smooth baseline hazard function, the piecewise constant modeling for the hazard function being often unrealistic. To introduce such *a priori* knowledge, we penalized the likelihood by a term which has large values for rough functions [6,7].

$$pl(\lambda_0(.), \beta, \theta) = l(\lambda_0(.), \beta, \theta) - \kappa \int_0^\infty \lambda_0^{''2}(t)\mathrm{d}t. \tag{2}$$

The roughness penalty function is represented by the squared norm of the second derivative of the hazard function [6]. The estimator $\hat{\lambda}(.)$ cannot be calculated explicitly but can be approximated on a basis of splines. Splines are piecewise polynomial functions that are combined linearly to approximate a function on an interval. We used cubic $M$-splines, which are a variant of cubic $B$-splines (for more details see [8]).

The estimated parameters are obtained by the robust Marquardt algorithm [9] which is a combination between a Newton–Raphson algorithm and a steepest descent algorithm. This algorithm has the advantage of being more stable than the Newton–Raphson algorithm while preserving its fast convergence property near the maximum.

We implemented in the R package "frailtypack", shared frailty models using a Penalized Likelihood estimation on the hazard function (see Appendix for an example) [10]. "Frailtypack" was first written in Fortran 90 and have been ported to the statistical software R.

### 2.2. Application: Breast cancer recurrences

We worked on a study of female patients with a primary invasive breast cancer and treated with breast-conserving surgery in one institute (Institut Bergonié, a regional comprehensive cancer center in South-Western France) from 1990 to 1994 and followed up to January 2006 ($n = 1006$) [11]. All the patients were prospectively included by one research assistant in a clinical, histological and biological database. Patients came to the hospital for a routine-visit every 6 months for 2 years, once a year for 3 years, then every 2 years for 5 years. Some other spontaneous inter-visits could take place.

There were 919 patients included in the analyses, 570 events were observed during the follow-up and 708 (77.0%) patients were still alive at the end of the follow-up period. The follow-up period varied between 150 days and 16 years

(mean = 11.8 years). The duration between two successive events decreased with time. By the time of study termination, 76 patients (8.3%) had died from other causes and 2 patients (0.2%) were lost to follow-up. The number of events ranged from 0 to 4 (average = 0.62 per patient). For each subject, events were categorized into locoregional, contralateral, metastases, second primary cancer, or death. First failures were mostly loco-regional recurrences (134 women over 919 at risk, 14.6%).

We analyzed the successive failures for each subject and their dependence using a random effects analysis. The rate of failure increased with grade (for grade 2 vs. grade 1 RelativeRisk = 1.74, 95% Confidence Interval = 1.29–2.34, for grade 3 vs. grade 1 Relative Risk = 2.02, 95% Confidence Interval = 1.43–2.85 in model 2), was higher for women with nodal invasion and with radiotherapy alone. The histological tumor size was associated with the rate of failure. The number of previous events (as a quantitative variable) was not associated with an increased risk of event given the frailty. The tendency was towards a higher risk of second failure compared to the risk of first failure. No major heterogeneity was observed in models after adjustment for the individual variables. The type of event influence the risk of developing a new failure. For instance the risk of failure was greatly increased (Relative Risk = 4.45) after developing a locoregional recurrence.

## 3. Cure frailty model

If a significant number of patients are "cured" and thus risk-free of recurrences after a first treatment, the population is then a mixture of susceptible and insusceptible subjects. In these cases, the standard survival analysis techniques, for example the Cox proportional hazards model [4] may not be appropriate, because they inherently assume that all the subjects have the same susceptibility to the disease and will eventually experience the event over a sufficiently long period of follow-up.

Cure models assume the existence of a cured fraction in the population. More generally they are used when a strong assumption of a subgroup non-susceptible to the event of interest is part of the marginal population. For example, a Kaplan–Meier survival curve with a long and stable plateau with heavy censoring at the tail may be taken as empirical evidence of such data [12]. Indeed, for such data, even if the follow-up were to be extended, a fraction of individuals would not experience the event of interest; that is, these individuals are no longer at risk. Cure models are becoming increasingly important and popular in clinical trials and medical research, especially in various types of oncology studies, such as breast cancer, leukemia and melanoma. To analyse recurrent events, it is first necessary to define the cured proportion to be modeled. Price and Manatunga [13] considered the use of cure frailty models to analyze a leukaemia recurrence with a cured fraction.

We proposed frailty cure models for recurrent event data with a flexible cured fraction that also depends on covariates.

### 3.1. Cure frailty model and inference

An alternative approach, that we propose, would be to account for the possibility of being "cured" after each event. Thus, probability of cure may evolve with time and is here defined as the probability to develop no further event after each event. The statistical power is increased when all available recurrences are considered together for the cured fraction instead of using only the first recurrence. Furthermore, this model can also be used to compare and evaluate time-dependent medical treatments on the times to recurrence among the uncured patients and on the cure probability.

Random effects are thus needed to take into account the intra-subject correlation. In this frailty cure model $\pi(.)$ is defined as the probability to develop a new event after each already developed event, so that $\pi(.)$ can be modeled by $\pi_{ij}(Z^*_{ij} \mid \omega_i)$. We therefore obtain the following model:

$$\begin{cases} S(t_{ij} \mid \omega_i) = 1 - \pi_{ij}(Z^*_{ij} \mid \omega_i) + \pi_{ij}(Z^*_{ij} \mid \omega_i) \exp(-\Lambda_0(t_{ij} \mid U_{ij} = 1) \exp(\beta' Z_{ij} + \omega_i)) \\ \pi_{ij}(Z^*_{ij} \mid \omega_i) = P(U_{ij} = 1 \mid Z^*_{ij}) = \dfrac{\exp(\mathbf{b}' Z^*_{ij} + \alpha\omega_i)}{(1 + \exp(\mathbf{b}' Z^*_{ij} + \alpha\omega_i))} \\ \omega_i \sim \mathcal{N}(0; \theta^2), \quad \text{with } \omega_i \perp \omega_{i'} \end{cases} \tag{3}$$

where $\alpha$ describes the association between the cured fraction and the recurrent rate due to the frailty. When $\alpha = 0$, the cured fraction and the recurrent events are independent, conditional on covariates and the different cured fractions for the same patient are independent of each other. When $\alpha = 1$, the effect of the frailty is identical for the recurrent rate and for the probability to develop a new event. When $\alpha > 0$ the recurrent rate and cured fraction are positively associated. Higher frailty will result in a higher risk of recurrence and in a higher probability to develop a new event. Inversely, when $\alpha < 0$ higher frailty will result in a higher survival among at risk patients but in a lower probability to develop a new event. The interpretation of the value of $\alpha$ makes sense only when heterogeneity is present, i.e., when the variance of the random effects is significantly different from zero.

In order to estimate parameters in the models, the method of maximum likelihood can be utilized using a piecewise constant baseline hazard function. Since the baseline hazard function $\lambda_0(t^* \mid U_{ij} = 1)$ is specified, then maximum likelihood estimates are obtained using the Newton–Raphson iterative procedure and numerical Gaussian quadrature integration techniques. Asymptotic standard errors are obtained by inverting the observed information matrix.

The resulting likelihood can then be maximized conveniently by Gaussian quadrature tools in standard statistical packages such as PROC NLMIXED in SAS 9.1 [14], a user-friendly procedure for estimation in nonlinear mixed models.

This model can also be applied to the analysis of breast cancer recurrences. In breast cancer studies the patients are never actually considered cured. However, in studies with long follow-up, one can argue that cases that have survived for a very long time without evidence of disease could be considered as having very low likelihood of dying from cancer. Our study on breast cancer relapses points to the existence of a large cured fraction.

## 4. Nested frailty model

The frailty model is a random effect survival model which allows for unobserved heterogeneity or for statistical dependence between observed survival data. The nested frailty model accounts for the hierarchical clustering of the data by including two nested random effects. Nested frailty models are particularly appropriate when data are clustered at several hierarchical levels naturally or by design. In such cases it is important to estimate the parameters of interest as accurately as possible by taking into account the hierarchical structure of the data.

### *4.1. Nested frailty model and inference*

We consider a multilevel proportional hazards model with two sets of nested random effects that act multiplicatively on the hazard, so that a large value of these variables increases the hazard [15]. Proportional hazards models are semi-parametric and fairly flexible and their covariates can be time-dependent. We treat the case of a cohort with $G$ independent clusters ($i = 1, \ldots, G$). Within the $i$th cluster, there are $J_i$ correlated sub-clusters ($j = 1, \ldots, J_i$). We treat the case of right-censored and left-truncated data. $T_{ijk}$ denotes the survival times under study for subject $k$ ($k = 1, \ldots, K_{ij}$) from subgroup $j$, and group $i$, and $C_{ijk}$ is the corresponding right-censoring time. The observations are $Y_{ijk} = \min(T_{ijk}, C_{ijk})$ and the censoring indicators $\delta_{ijk} = I_{\{T_{ijk} \leq C_{ijk}\}}$. The survival times may be left-truncated. We assume that the times of left truncation $\mathcal{L}_{ijk}$ are independent of the survival times $T_{ijk}$. Only subjects with $T_{ijk} > \mathcal{L}_{ijk}$ are observed. We define two random effects $v_i$ and $w_{ij}$ and assume that the cluster-level random effects $v_i$ and the sub-cluster random effects $w_{ij}$ are independent and gamma-distributed random effects ($\Gamma(1/\alpha;\, 1/\alpha)$ and $\Gamma(1/\eta;\, 1/\eta)$) with $E(v_i) = 1$, $\text{var}(v_i) = \alpha$ and $E(w_{ij}) = 1$, $\text{var}(w_{ij}) = \eta$. For identifiability, they have a mean equal to 1 at birth (i.e., at duration $t_{ijk} = 0$). If the variance is null, then observations from the same group are independent. A larger variance implies greater heterogeneity in frailty across groups and a greater correlation of the survival times for individuals belonging to the same group [16]. Mainly for reasons of mathematical convenience, the frailty terms are often assumed to follow a gamma distribution. This choice and other possibilities such as lognormal, positive stable distributions are discussed in several papers [17,18]. To construct the likelihood function, apart from the usual assumption of independent censoring, the censoring must be non-informative for $v_i$ and $w_{ij}$.

The hazard function conditional on the two frailties $v_i$ and $w_{ij}$, for individual $(i, j, k)$ is:

$$\lambda_{ijk}(t|v_i, w_{ij}) = v_i w_{ij} \lambda_0(t) \exp(\beta' \mathbf{X_{ijk}}) \tag{4}$$

where $\lambda_0(t)$ is the baseline hazard function; $\mathbf{X_{ijk}} = (X_{1ijk}, \ldots, X_{pijk})'$ denotes the covariate vector for the $k$th individual, with $p$ the number of covariates, and $\beta$ is the corresponding vector of regression parameters.

We also introduced a semi-parametric penalized likelihood approach to estimate the different parameters. The integrations in the full log-likelihood expression were evaluated using Gaussian quadrature. Laguerre polynomials were used to treat the integration $[0, \infty)$.

We implemented in the R package "frailtypack", nested frailty models using a Penalized Likelihood estimation on the hazard function (see Appendix for an example).

The simulation study demonstrates that this semi-parametric approach yields satisfactory results in this complex setting. In order to illustrate the MPnLE method and the nested frailty model, we analyzed a set of data on chronic granulomatous disease (CGD) [19]. CGD is a heterogeneous group of uncommon inherited disorders characterized by recurrent pyogenic infections that usually begin early in life and may lead to death in childhood. The CGD data set consists of a placebo-controlled randomized trial of gamma-interferon ($\gamma$-IFN) in CGD. The aim of the trial was to investigate the effectiveness of $\gamma$-IFN on serious infections in CGD patients. We illustrate that using a shared frailty model instead of a nested frailty model with two levels of regrouping leads to inaccurate estimates, with an overestimation of the variance of the random effects. We show that even when the frailty effects are fairly small in magnitude, they are important since they alter the results in a systematic pattern.

## 5. Joint frailty model

The time frame for an individual's repeated event process may depend on other "terminating" events, such as death. Often the recurrence of serious events, such as tumors and opportunistic infections, is associated with an elevated risk of death. In this context, the usual assumption of non-informative censoring of the recurrent event process by death, required by most statistical analyses, can be violated. This dependence should be accounted for in the joint modeling of recurrent events and deaths. Liu et al. [20] proposed a joint semi-parametric model for the intensity functions of both recurrent events and death by a shared gamma frailty model. In these models the frailty effect on recurrent events and death rates is not the same. In those approaches estimation is carried out through a Monte Carlo EM algorithm, which could

be time-consuming. Furthermore, these methods can not be used to correctly estimate hazard functions, which often have a meaningful interpretation in epidemiological studies. Most of the time, the baseline intensity estimate is based on Breslow's estimate leading to a piecewise-constant baseline hazard function or unspecified baseline hazard function.

We have proposed a non-parametric penalized likelihood method for estimating hazard functions in a general joint frailty model for recurrent events and terminal events, with both right censored survival data and delayed entries [21]. This approach is of interest for several reasons. Firstly, it makes it possible to deal with informative censoring for recurrent event data, in addition it also allows joint treatment of two processes which evolve with time leading to more accurate estimates. This work extends previous work by giving smoothed estimates of the two hazard functions which represent incidence and mortality rates in epidemiology. It is natural in epidemiology to impose a continuous hazard function with small local variations.

### 5.1. The joint frailty model

We denote for subject $i$ ($i = 1, \ldots, N$), $X_{ij}$ the $j$th recurrent times ($j = 1, \ldots, n_i$), $C_i$ the censoring times (not by death) and $D_i$ the death times. We first consider $X_{ij}$ as a time to event. $T_{ij} = \min(X_{ij}, C_i, D_i)$ corresponds to each follow-up time and $\delta_{ij}$ is a binary indicator for recurrent events which is 0 if the observation is censored or if the subject died, and 1 if $X_{ij}$ is observed ($\delta_{ij} = I_{(T_{ij}=X_{ij})}$ where $I_{()}$ denotes indicator function). Similarly, we note $T_i^*$ the last follow-up time for subject $i$, which is either a time of censoring or a time of death ($T_i^* = \min(C_i, D_i)$) and $\delta_i^* = I_{(T_i^*=D_i)}$. What we actually observe is $(T_{ij}, \delta_{ij}, \delta_i^*)$.

Following the model of Liu and Wolfe [20], the joint model for the hazard functions for recurrent event ($r_i(.)$) and death ($\lambda_i(.)$) is:

$$\begin{cases} r_{ij}(t|\omega_i) = \omega_i r_0(t) \exp(\beta_1' Z_{ij}(t)) = \omega_i r_{ij}(t) \\ \lambda_i(t|\omega_i) = \omega_i^\alpha \lambda_0(t) \exp(\beta_2' Z_i(t)) = \omega_i^\alpha \lambda_i(t). \end{cases} \tag{5}$$

The effect of the explanatory variables is assumed to be different for recurrent and for death times. The parameters $\beta_1$ and $\beta_2$ are interpretable in terms of the instantaneous probability of occurrence of the recurrent events and the terminal event, respectively, conditional on the subject's past event history and on being alive. The model and the estimation can deal with external time-dependent covariates in the sense of Kalbfleisch and Prentice [22] (page 197). The previous number of recurrent events can also be considered as an internal time-dependent covariate, that requires the survival of the individual for its existence, and its path thus carries direct information on the time to failure.

The random effects $\omega_i$ (frailties) are assumed independent. The gamma frailty density is adopted here with unit mean and variance $\theta$. The dependence between $T_i^*$ and $T_{ij}$ conditional on $Z_i(t)$ is solely due to the fact that the unobserved $\omega_i$ affects both the recurrent times and the death times. The common frailty parameter $\omega_i$ will take into account the heterogeneity in the data, associated with unobserved covariates.

In the traditional model, the assumption is that $\alpha = 0$ in (5), that is $\lambda_i(t)$ does not depend on $\omega_i$ and thus death (or the terminal event process) is not informative for the recurrent event rate $r_i(t)$, i.e., the two rates $\lambda_i(t)$ and $r_i(t)$ are not associated, conditional on covariates. When $\alpha = 1$, the effect of the frailty is identical for the recurrent events and for the terminating event. When $\alpha > 1$, the recurrent rate and the death rate are positively associated; higher frailty will result in higher risk of recurrence and higher risk of death.

We showed how maximum penalized likelihood estimation can be applied to non-parametric estimation of the continuous hazard functions in a general joint frailty model with right censoring and delayed entry. The integrations in the full log-likelihood expression were evaluated using Gaussian quadrature and Laguerre polynomials were used. We implemented in the R package "frailtypack", joint frailty models (see Appendix for an example).

### 5.2. Application: Follicular lymphomas

The approach we developed was motivated by a study of patients with follicular lymphoma (FL) undergoing episodic relapses of FL. Follicular lymphomas (FL) account for one third of non-Hodgkin lymphomas in adults. The prognosis of FL is heterogeneous and numerous treatments may be proposed [23]. The course of this disease is usually characterized by a response to initial treatment, followed by relapses, sometimes associated with high-grade non-Hodgkin lymphomas. After the initial treatment, each patient was monitored regularly for routine-visits and presence of FL relapses was notified at each visit. Estimation of the risk of recurrence allows for better planning of follow-up schedules after diagnosis or first treatment, and permits clinicians to determine therapeutic approaches based on the patient's risk of relapse. Furthermore, FL relapses may increase the risk of death. As a result there is an association between FL relapses process and the survival process, which precludes the use of standard analyses of recurrent events. Specifically, those subjects experiencing FL relapses at the highest rate are typically observed for shorter periods of observation due to mortality.

We have shown by simulation that using a reduced shared frailty model instead of a joint frailty model when there is a significant dependence between the two processes leads to unreliable estimates, with regression factors falsely non-significant or with an underestimation of the recurrence risk. This implies that the non-informative censoring of the recurrent event process by death needs to be taken into account in survival analysis to obtain accurate inferences.

## 6. Conclusion

We focussed on the different extensions of the simple shared frailty model to analyze recurrent events. All those frailty model techniques are available using SAS or R ("Frailtypack", a freeware package) to a wider audience. Other approaches allowing for additional correlation structures on the random effects may provide valuable insight for future research.

## Appendix. Frailtypack: Examples of frailty models using this R package

http://cran.r-project.org/.

### *A.1. Shared frailty models*

>frailtyPenal(Surv(start, stop, status)∼
**cluster**(patkey)
+rank+histosize+grade2+EIC+grade3+node13+node4
+emboll+excision+PRnegative
+radchemo+radhor+radchemohor,
data=breastcancer, Frailty=TRUE, n.knots=8,
kappa1=150, cross.validation=TRUE, recurrentAG=TRUE)

- "*data*": indicates the name of the data file. The database structure is different according to the model.
- "*recurrentAG*": Logical value (TRUE or FALSE). If recurrentAG=TRUE, it indicates that recurrent event times with a counting process approach is used with a calendar timescale. Calendar timescale means that the times correspond to the time since the entry/inclusion in the study. The gap timescale means that the times correspond to the time between two recurrent events. Default is FALSE, in particular for recurrent events or clustered data with gap-time as the timescale.
- "*Frailty*": Logical value (TRUE or FALSE). If Frailty= TRUE, the model includes a frailty term (a shared frailty model is fitted) and the variance of frailty parameter is estimated. If Frailty=FALSE, a Cox proportional hazards model is estimated using Penalized likelihood on the hazard function. Default is FALSE.
- "*n.knots*": is the number of knots to use. It corresponds to the n.knots+2 splines functions for the approximation of the baseline hazard function or the survival functions. The number of knots must be between 4 and 20. So it is recommended to start with a small number of knots (for instance: n.knots=7) and increase the number of knots until the graph of the baseline hazard function remains unchanged.
- "*kappa*1": is the smoothing parameter of the penalized likelihood.
- "*kappa*2": is the second smoothing parameter, it is required if the analysis is stratified (only for Cox, shared, nested and additive frailty models). This parameter will correspond to the smoothing parameter for the second baseline hazard function. If a joint frailty model is fitted, this parameter will correspond to the smoothing parameter for the death baseline hazard function (kappa1 being for the recurrent events baseline hazard function).
- "*cross.validation*": if=TRUE, indicates that a cross validation procedure is used for estimating the best smoothing parameter. kappa1 is used as the seed for the estimation of the smoothing parameter. Be careful, if a joint frailty model is fitted or a stratified analysis for Cox, shared, nested frailty model, the cross-validation method is not allowed, i.e., cross.validation=FALSE. Default is FALSE.

### *A.2. Nested frailty models*

>frailtyPenal(Surv(start, stop, status)∼
**cluster**(group)+**subcluster**(subgroup)
+cov1+cov2, data=dataNested, n.knots=8, kappa1=50 000)

- "*cluster*": identifies groups levels.
- "*subcluster*": identifies subgroups levels.

### *A.3. Joint frailty models*

>frailtyPenal(Surv(start, stop, status)∼ cluster(id)+var1+var2
+**terminal(status.terminal)**,
**formula.terminalEvent**=∼var1,
data=dataJoint, n.knots=7,
kappa1=1, kappa2=1, **joint**=TRUE, recurrentAG=TRUE)

- "*terminal*": is a special function used in the context of recurrent event models with terminal event (e.g., censoring variable related to recurrent events). It contains "status.terminal" a death indicator.
- "*start*": is the entry time
- "*stop*": is the last follow-up time

- "*joint*": Logical value (TRUE or FALSE). The default is FALSE. If TRUE a joint frailty model is fitted and if so, 'formula.terminalEvent' argument is required.
- "*formula.terminalEvent*": indicates which covariates the terminal event rate is adjusted for.

## References


[1] R. Cook, J. Lawless, The Statistical Analysis of Recurrent Events, Springer Publishing Company, New York, NY, 2007.
[2] L. Duchateau, P. Janssen, I. Kezic, C. Fortpied, Evolution of recurrent asthma event rate over time in frailty models, Applied Statistics 52 (3) (2003) 355–363.
[3] J. Vaupel, K. Manton, E. Stallard, The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography 16 (1979) 439–454.
[4] D. Cox, Regression models and life-tables, Journal of the Royal Statistical Society. Series B. Methodological 34 (1972) 187–220.
[5] V. Rondeau, D. Commenges, P. Joly, Maximum penalized likelihood estimation in a gamma-frailty model, Lifetime Data Analysis 9 (2) (2003) 139–153.
[6] F.O. Sullivan, Fast computation of fully automated log-density and log-hazard estimators, SIAM Journal on Scientific and Statistical Computing 9 (1988) 363–379.
[7] P. Joly, D. Commenges, L. Letenneur, A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia, Biometrics 54 (1) (1998) 185–194.
[8] J. Ramsay, Monotone regression splines in action, Statistical Science 3 (4) (1988) 425–461.
[9] D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, Journal of the Society for Industrial and Applied Mathematics 11 (2) (1963) 431–441.
[10] V. Rondeau, J. Gonzalez, Frailtypack: A computer program for the analysis of correlated failure time data using penalized likelihood estimation, Computer Methods and Programs in Biomedicine 80 (2) (2005) 154–164.
[11] I. de Mascarel, G. MacGrogan, S. Mathoulin-Pélissier, I. Soubeyran, V. Picot, J. Coindre, Breast ductal carcinoma in situ with microinvasion: A definition supported by a long-term study of 1248 serially sectioned ductal carcinomas, Cancer 94 (8) (2002) 2134–2142.
[12] J. Sy, J. Taylor, Estimation in a Cox proportional hazards cure model, Biometrics 56 (1) (2000) 227–236.
[13] D. Price, A. Manatunga, Modelling survival data with a cured fraction using frailty models, Statistics in Medicine 20 (9–10) (2001) 1515–1527.
[14] R. Littell, G. Milliken, W. Stroup, R. Wolfinger, SAS for Mixed Models, SAS Publishing, 2006.
[15] V. Rondeau, L. Filleul, P. Joly, Nested frailty models using maximum penalized likelihood estimation, Statistics in Medicine 25 (23) (2006) 4036–4052.
[16] T. Therneau, P. Grambsch, Modeling Survival Data: Extending the Cox Model, Springer Verlag, 2000.
[17] P. Hougaard, Frailty models for survival data, Lifetime Data Analysis 1 (3) (1995) 255–273.
[18] A. Pickles, R. Crouchley, A comparison of frailty models for multivariate survival data, Statistics in Medicine 14 (13) (1995) 1447–1461.
[19] T. Fleming, D. Harrington, Counting Processes and Survival Analysis, in: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section, 1991.
[20] L. Liu, R. Wolfe, X. Huang, Shared frailty models for recurrent events and a terminal event, Biometrics 60 (3) (2004) 747–756.
[21] V. Rondeau, S. Mathoulin-Pélissier, H. Jacqmin-Gadda, V. Brouste, P. Soubeyran, Joint frailty models for recurring events and death using maximum penalized likelihood estimation: Application on cancer events, Biostatistics 8 (4) (2007) 708–721.
[22] J. Kalbfleisch, R. Prentice, The Statistical Analysis of Failure Time Data, Wiley-Interscience, NY.
[23] P. Solal-Celigny, P. Roy, P. Colombat, J. White, J. Armitage, R. Arranz-Saez, W. Au, M. Bellei, P. Brice, D. Caballero, et al., Follicular lymphoma international prognostic index, Blood 104 (5) (2004) 1258–1265.