

## frailtypack : Un package pour l'analyse de données de survie corrélées

A. Laurent<sup>a</sup> et Y. Mazroui<sup>b</sup> et A. Mauguen<sup>a</sup> et V. Rondeau<sup>a,c</sup>

<sup>a</sup>INSERM U897

ISPED

146 Rue Léo Saignat 33076 Bordeaux Cedex

Alexandre.Laurent@isped.u-bordeaux2.fr

Yassin.Mazroui@isped.u-bordeaux2.fr

Audrey.Mauguen@isped.u-bordeaux2.fr

Virginie.Rondeau@isped.u-bordeaux2.fr

<sup>c</sup>Université Bordeaux Segalen, ISPED

146 Rue Léo Saignat 33076 Bordeaux Cedex

**Mots clefs** : Survie, Données groupées, Modèles à fragilité, Modèles conjoints, Événements récurrents

Les modèles à fragilité sont une extension des modèles à risques proportionnels de Cox qui sont les plus populaires dans l'analyse de survie. Dans la plupart des applications cliniques, la population d'étude est considérée comme un échantillon hétérogène ou autrement dit un échantillon de plusieurs groupes homogènes pouvant représenter des familles ou des zones géographiques. Parfois pour des raisons économiques ou par omission, certaines variables étroitement liées à l'événement d'intérêt peuvent ne pas être mesurées. Les modèles à fragilité, ou modèles de survie à effet aléatoire permettent alors de tenir compte de l'éventuelle hétérogénéité de la population liée à des variables non mesurées.

`frailtypack` [1,2] est un package R qui proposent plusieurs types de modèles à fragilité pour données censurées à droite et tronquées à gauche. Le modèle à fragilité partagée (shared) qui est le modèle de base, est à utiliser lorsque nous savons que les données ne sont pas indépendantes. D'autre part, le risque de récurrences peut être interrompu par le décès, qui peut être décrit comme une censure informative. Ainsi, il est possible de modéliser conjointement les fonctions de risque associées aux événements récurrents et terminaux en considérant un effet aléatoire commun aux observations d'un même sujet [3]. Dans la continuité de ce même modèle, nous pouvons obtenir un modèle conjoint pour des données groupées [4] où l'effet aléatoire serait commun aux individus d'un même groupe. Toujours dans l'idée de la modélisation conjointe, un modèle multivarié [5] a été implémenté et permet l'étude cette fois de deux types d'événements récurrents en plus d'un événement terminal.

L'estimation des fonctions de risque de base se fait toujours de manière non paramétrique en approximant par des splines ou plus récemment en utilisant une paramétrisation Weibull ou une approximation en constant par morceaux. Les modèles shared et joint qui sont les plus utilisés sont les premiers à bénéficier d'améliorations notables. Jusqu'à maintenant le package ne proposait qu'une distribution de type Gamma pour les effets aléatoires, il est maintenant possible d'y considérer une distribution Log-Normale. De plus, les variables explicatives sont supposées avoir un effet constant sur le temps d'apparition de l'événement d'intérêt, il est désormais possible de considérer des variables qui ont un effet dépendant du temps en modélisant les coefficients de régression par une combinaison linéaire de B-splines. Le package propose

aussi l'inclusion de variables dépendantes du temps, de stratification mais aussi de résultats de prédiction.

Le package dont la première version date de 2005 et qui proposait uniquement de faire un simple modèle à fragilité partagée a bien été enrichi depuis. Initialement, le programme était écrit en Fortran 77 jusqu'à son adaptation au logiciel R dont l'utilisation dépend et se veut très proche du package `survival`. Cependant encore aujourd'hui, les corps de programme sont toujours implémentés en Fortran 90. L'outil étant en constante évolution, l'objectif de cette présentation sera de mettre en lumière les récentes améliorations faites sur les modèles à fragilité partagée et les modèles conjoints et montrer leur utilisation dans un cas concret d'épidémiologie.

## Références

- [1] Rondeau, V., Mazroui, Y. and Gonzalez, J. (2012). frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, **47**(4), 1-28
- [2] Rondeau, V. and Gonzalez, J. R. (2005). frailtypack: a computer program for the analysis of correlated failure time data using penalized likelihood estimation. *Comput Methods Programs Biomed*, **80**(2), 154-64.
- [3] Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. and Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*, **8**(4), 708-721.
- [4] Rondeau, V., Pignon, J.-P. and Michiels, S. (2011). A joint model for the dependence between clustered times to tumour progression and deaths: A meta-analysis of chemotherapy in head and neck cancer. *Statistical methods in medical research*
- [5] Mazroui, Y., Mathoulin-Pellissier, S., MacGrogan, G., Brouste, V. and Rondeau, V. (2013). Multivariate frailty models for two types of recurrent events with an informative terminal event: Application to breast cancer data. To appear in *Biometrical journal*