



**HAL**  
open science

## Modèles de mélange pour données récurrentes : application aux récurrences des cancer

Virginie Rondeau, Emmanuel Schaffner, Fabien Corbière, Simone  
Mathoulin-Pélissier

► **To cite this version:**

Virginie Rondeau, Emmanuel Schaffner, Fabien Corbière, Simone Mathoulin-Pélissier. Modèles de mélange pour données récurrentes : application aux récurrences des cancer. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386615

**HAL Id: inria-00386615**

**<https://inria.hal.science/inria-00386615>**

Submitted on 22 May 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# MODÈLES DE MÉLANGE POUR DONNÉES RÉCURRENTES : APPLICATION AUX RÉCIDIVES DES CANCER

Virginie Rondeau<sup>1,2</sup>, Emmanuel Schaffner<sup>1</sup>, Fabien Corbière<sup>1,3</sup>, and  
Simone Mathoulin-Pélissier<sup>4</sup>

<sup>1</sup>*INSERM, U897 (Biostatistic), Bordeaux, F-33076, FRANCE;*

<sup>2</sup>*Université Victor Segalen Bordeaux 2, Bordeaux, F-33076, FRANCE;*

<sup>3</sup>*INRA-UMR 1225, Toulouse, F-31076, FRANCE;*

<sup>4</sup>*Institut Bergonié - Centre Régional de Lutte Contre le Cancer du Sud-Ouest (Clinical Research Unit), Bordeaux, F-33076, FRANCE.*

## RESUME

Dans les analyses de survie traditionnelles, en ayant un suivi suffisamment long on suppose que chaque individu est à risque de développer un événement. Cependant, si l'événement considéré est par exemple une récurrence de cancer, une fraction de la population dite " guérie ", ne va jamais rencontrer cet événement même après un très long suivi. Les modèles de mélange pourront d'une part permettre d'estimer cette fraction mais aussi la fonction de survie pour les patients susceptibles de développer une récurrence. Ces modèles de mélange avec taux de guérison généralisent le modèle à risques proportionnels de Cox en tenant compte de l'existence d'une sous population qui n'est pas à risque de développer l'événement d'intérêt. Nous avons proposé des modèles de mélange à effets aléatoires. Les estimations ont été obtenues par la procédure NLMIXED du logiciel SAS pour des modèles de survie avec fonctions de risque constantes par morceaux. Contrairement au simple modèle à fragilité, les méthodes proposées semblent prometteuses pour modéliser des données de survie multivariées avec des survivants à long-termes. Les données de récurrences de cancer du sein ont été utilisées.

## ABSTRACT

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. A certain proportion of subjects in the population who are not expected to experience the events of interest are considered to be "cured" or non-susceptible. To model correlated failure time data incorporating a surviving fraction, we compare two forms of cure rate frailty models. We illustrate the cure frailty models with the analysis of prognostic factors associated with breast cancer recurrences, metastases, new primary malignancy and death. Estimates were obtained by maximization of likelihood using SAS proc NLMIXED for a piecewise constant hazards model. As opposed to the simple frailty model, the proposed methods demonstrate great potential in modelling multivariate survival data with long-term survivors ("cured" individuals).

Dans les analyses de survie traditionnelles, en ayant un suivi suffisamment long, on suppose que chaque individu est à risque de développer un événement. Cependant, une fraction de la population, dite "guérie", ne va jamais rencontrer cet événement même après un très long suivi. Ainsi la fonction de survie reste à un niveau non nul après un certain temps. Cette problématique aborde la notion de modèles de mélange en analyse de survie pour fraction non à risque, une thématique sur laquelle j'ai récemment travaillé avec Pierre Joly et Fabien Corbière (son précédent étudiant en thèse) dans l'équipe de Biostatistique. La motivation initiale était de travailler sur des modèles de mélange pour données groupées ("Cure frailty models") en s'appuyant sur l'étude de la tremblante du mouton pour des données regroupées en élevages. Ces modèles ont été essentiellement proposés pour analyser des données de survie groupées (Yin, 2005) en introduisant des effets aléatoires pour tenir compte de la corrélation intra-groupe.

Nous avons récemment travaillé sur ces modèles de mélange pour analyser les données récurrentes de récurrences de cancer du sein. Dans l'analyse de données de cancer du sein, cette approche par modèles de guérison (ou modèles de mélange) est tout à fait réaliste, puisqu'il a été montré par exemple que certaines récurrences loco-régionales étaient potentiellement curables, n'étant pas suivies d'une étendue à distance de la maladie. A nouveau, dans cette analyse d'événements de survie répétés dans le temps, la population d'étude est un mélange de deux sous-populations : une fraction non à risque de récurrences (ou "guérie"), et une fraction à risque de récurrences. Nous avons proposé deux formes de modèles à fragilités avec fraction non à risque. Dans le premier modèle, les patientes non-susceptibles sont celles qui ne connaîtront jamais de récurrences. La probabilité d'être "guérie" est donc fixe pour chaque sujet. Le deuxième modèle tient compte de la possibilité de guérison après chaque nouvel événement. Ainsi, après chaque nouvel événement, une patiente a une nouvelle probabilité de "guérison". On notera  $U$ , indiquant si un individu développera ( $U=1$ ) ou ne développera pas une récurrence. Pour des événements observés, nous savons que  $U = 1$ , et que la patiente est non guérie. Lorsqu'une patiente est censurée, alors nous pouvons avoir  $U = 1$  ou  $U = 0$ . On introduit  $\pi(X^*) = P(U = 1 | X^*)$  la probabilité d'être non guérie qui dépend d'un vecteur de variables explicatives  $X^* = (X_1^*, \dots, X_q^*)'$ . Ce modèle de mélange à fragilités s'écrit :

$$\begin{cases} S(t_{ij} | X_{ij}, X_{ij}^*, \omega_i) = 1 - \pi_{ij}(X_{ij}^* | \omega_i) + \pi_{ij}(X_{ij}^* | \omega_i) \exp(-\Lambda_0(t_{ij} | U = 1) \exp(\beta' X_{ij} + \omega_i)) \\ \pi_{ij}(X_{ij}^* | \omega_i) = P(U = 1 | X_{ij}^*) = \frac{\exp(\mathbf{b}' X_{ij}^* + \alpha \omega_i)}{(1 + \exp(\mathbf{b}' X_{ij}^* + \alpha \omega_i))} \\ \omega_i \sim \mathcal{N}(0; \theta^2), \omega_i \perp \omega_{i'} \end{cases} \quad (1)$$

Les effets aléatoires  $\omega_i$  vont tenir compte de la corrélation intra-sujets des données, et le paramètre  $\alpha$  permettra de décrire l'association entre la probabilité d'être à risque et l'intensité du risque de récurrences de cancer liée à la variable de fragilité. Pour une valeur positive de  $\alpha$ , une fragilité élevée résultera en une probabilité plus élevée d'appartenir à la fraction à risque et en une survie réduite chez les sujets à risque. On s'intéresse à l'effet de facteurs sur le fait d'appartenir ou non à la fraction à risque, généralement par un

modèle logistique, mais aussi à l'effet de facteurs (pouvant être différents) sur la durée de survie des sujets à risques par un modèle à risques proportionnels.

Les estimations ont été obtenues par la procédure NLMIXED du logiciel SAS pour des modèles de survie avec fonctions de risque constantes par morceaux et des méthodes d'intégrations numériques par quadrature gaussiennes adaptatives. Contrairement au simple modèle à fragilités, les modèles de mélange à fragilités proposés semblent prometteurs pour modéliser des données de survie multivariées avec des patients guéris à long-terme.

## Bibliographie

- [1] Sy, JP., Taylor, JM. (2000) Estimation in a Cox proportional hazards cure model. *Biometrics*, 56, 227–236.
- [2] Peng, Y., Dear, KB. (2000) A nonparametric mixture model for cure rate estimation. *Biometrics*, 56, 237–43.
- [3] Yu, B., Tiwari, RC., Cronin, KA., Feuer, EJ. (2004) Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*, 23, 1733–1747.
- [4] Yin, G. (2005) Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics*, 61, 552–8.
- [5] Yu, B. (2008) A frailty mixture cure model with application to hospital readmission data. *Biometrical Journal*, 50, 386–394.