HAL open science

Prédiction de la pneumonie nosocomiale à l'aide d'un modèle multi-états.

Molière Nguile Makao, Jean-François Coeurjolly, Benoit Liquet, Jean-François Timsit

To cite this version:

Molière Nguile Makao, Jean-François Coeurjolly, Benoit Liquet, Jean-François Timsit. Prédiction de la pneumonie nosocomiale à l'aide d'un modèle multi-états.. 41èmes Journées de Statistique, Société Française de Statistique, May 2009, Bordeaux, France. inria-00386790

HAL Id: inria-00386790

https://inria.hal.science/inria-00386790v1

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.

Prédiction de la pneumonie nosocomiale à l'aide d'un modèle multi-états.

M. Nguile Makao[1], J.F. Coeurjolly[2], B. Liquet[3], J.F. Timsit [1,4].

[1] *Equipe 11 Epidémiologie pronostique des cancers et affections graves de l'U823 INSERM - Université Joseph Fourier.*
[4] *Réanimation médicale CHU Grenoble.*
[3] *Equipe Biostatistique de l'U897 INSERM ISPED - Université Victor Segalen Bordeaux 2.*
[2] *SAGAG, Dép. de statistiques, laboratoire Jean Kuntzman -Université Pierre-Mendès France.*

Résumé

Les modèles multi-états connaissent une utilisation croissante dans plusieurs domaines notamment en médecine, pour expliquer la survenue d'évènements au cours du temps pouvant être concurrents. Dans ce papier, nous utilisons un modèle multi-états pour construire la fonction de prédiction individualisée de la pneumonie nosocomiale en réanimation dans les trois jours qui suivent l'instant d'observation du patient. Le modèle multi-états est décrit par un processus semi-Markov non-homogène. Nous proposons ici une estimation paramétrique et non paramétrique de la prédiction qui tient compte de covariables fixes et dépendantes du temps. Une application est faite sur une base de données de 2873 patients de réanimation ventilés mécaniquement, extraite de la base OUTCOMEREA.

Mots-clés *: Modèle multi-états, Processus Semi-Markovien, Prédiction Pneumonie nosocomiale*

Abstract

The multistate models have become increasingly used in many fields notably in medicine in order to explain the occurrence of events (possibly competing) over time. In this paper, we use a multistate model to build an individualized prediction function of nosocomial pneumonia occurring in Intensive Care Units (ICU) over the three days that follow a patient observation. The multi state model is described by a non homogeneous semi-Markov process. We proposed a parametric and non parametric estimation of the prediction function which is taking into account fixed and time dependant covariates. An application is done on a database of 2873 patients (from the ICU) with mechanical ventilation, extracted from the OUTCOMEREA database.

Keywords*: Multistate model, Semi-Markov process, Prediction Nosocomial pneumonia*

[1] Institut Albert Bonniot, BP 170 Grenoble 38042 Cedex 9 France.
[2] B.S.H.M., 1251 Av. Centrale, BP 47 38040 Grenoble Cedex 9, France.
[3] 146 rue Léo Saignat 33076 Bordeaux Cedex.
[4] Hôpital Albert Michallon, BP 217, 38043 Grenoble Cedex 9, France.

1 Introduction

Les modèles multi-états sont considérés comme une généralisation des modèles de survie. Ils sont caractérisés par un processus stochastique à espace d'état fini et sont généralement utilisés pour décrire l'histoire d'un individu au cours du temps. Le but de ce travail est d'utiliser ce type de modèle pour construire une fonction de prédiction individualisée du changement d'état dans un intervalle de temps $[t+k, t+k+l]$ où t est l'instant de prédiction, en vue d'une application à la prédiction de la pneumonie nosocomiale en réanimation. Dans ce papier, nous proposons d'utiliser un processus Semi Markov non homogène pour prendre en compte la variabilité des intensités de transition qui sont habituellement estimées par des risques constants au cours du temps [2] et d'inclure des covariables fixes et dépendantes du temps pour mieux décrire l'hétérogénéité des patients.

2 Modèle multi-états pour la prédiction

2.1 Généralités

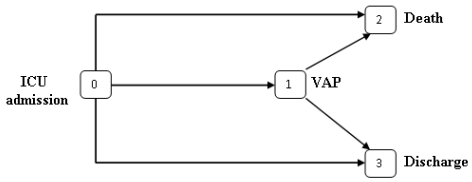

Figure 1: *Modèle multi-états pour les patients de réanimation*

Considérons le modèle multi-états (Figure 1), comportant 4 états $S = \{0, 1, 2, 3\}$ représentant respectivement être en réanimation sans pneumonie nosocomiale acquise sous ventilation mécanique (VAP), être en réanimation avec VAP, décéder en réanimation et sortir du service de réanimation. Soit un processus de Markov non homogène $(X(t))_{t\in[0,+\infty[}$ à 4 états dont chaque transition possible $h, j \in S$ est définie par la fonction d'intensité de transition :

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} \frac{\mathbf{Pr}(X(t+\Delta t) = j | X(t^-) = h)}{\Delta t} \quad \text{pour tout} \quad h \neq j. \tag{1}$$

Remarque : *Pour toute la modélisation, nous utilisons un processus semi-Markovien. Par conséquent t représente le temps passé dans un état et non dans l'étude* [3]

La fonction d'intensité cumulée est définie par :

$$A_{hj}(t) = \int_0^t \alpha_{hj}(u)du, \tag{2}$$

la fonction de survie dans l'état h par

$$S_h(t) = \prod_{j \neq h} S_{hj}(t) \tag{3}$$

où $S_{hj}(t) = \exp(-A_{hj}(t))$ et la fonction de sous-densité par,

$$f_{hj}(t) = S_h(t)\alpha_{hj}(t). \tag{4}$$

La propriété de semi-Markov assure que le processus est réinitialisé après chaque transition possible [2] [5]. Par conséquent, le modèle (Figure 1) peut être décomposé en 5 modèles de survie avec censure à droite correspondant aux transitions qui ne se sont pas réalisées. Pour toute transition possible $h, j \in S$ on note $Z_{hj}(.) = (\bar{Z}_{hj}, \tilde{Z}_{hj}(.))$ le vecteur des covariables, $\bar{Z}_{hj}$ le vecteur des covariables fixes, $\tilde{Z}_{hj}(.)$ le vecteur des covariables dépendantes du temps et $Z_h(.)$ le vecteur des covariables de toutes les transitions provenant de l'état h. On suppose que toutes les covariables respectent l'hypothèse de la proportionnalité. Pour chaque transition, on inclut les covariables par un modèle de Cox [6] de la manière suivante :

$$\alpha_{hj}(t|Z_{hj}(t)) = \alpha_{hj,0}(t)\exp(\beta_{hj}^t Z_{hj}(t)) \quad \text{pour tout } t \geq 0 \tag{5}$$

où $\alpha_{hj,0}(.)$ est l'intensité de transition de base, $\beta_{hj} = (\overline{\beta}_{hj}, \widetilde{\beta}_{hj})$ le vecteur des paramètres des effets des covariables avec $\widetilde{\beta}_{hj}$ pour les covariables dépendantes du temps et $\overline{\beta}_{hj}$ pour les covariables fixes.

2.2 Prédiction Individualisée

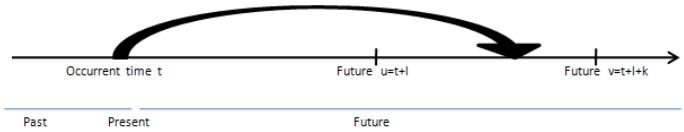

Figure 2: *Description de la prédiction*

Soit t l'instant de prédiction, on note $0 \leq \tau_1^i \leq \tau_2^i \leq \ldots \leq \tau_s^i \leq \ldots \leq t$ les instants d'observation passés de l'individu i. $\tilde{Z}_{hj}^i(.)$ est transformée en une fonction constante par morceaux égale à $\tilde{Z}_{hj}^i(\tau_s^i)$ pour tout $x \in [\tau_s^i, \tau_{s+1}^i[$. On définit la fonction covariable "prédite" pour tout instant t fixé par :

$$\tilde{Z}_{hj,t}^i(x) = \mathbf{1}_{[0,t]}(x)\tilde{Z}_{hj}^i(x) + \mathbf{1}_{]t,+\infty[}(x)\tilde{Z}_{hj}^i(t) \quad \text{pour tout } x \geq 0 \tag{6}$$

Par analogie avec les notations précédentes, on note $Z^i_{hj,t}(x) = (\bar{Z}^i_{hj}, \tilde{Z}^i_{hj,t}(x))$ et $Z^i_{h,t}(x) = (\bar{Z}^i_{hj}, \tilde{Z}^i_{hj,t}(x))$. La probabilité de l'individu i de passer de l'état h à l'état j entre $t+k$ et $t+k+l$ avec $k, l \in \mathbf{R}_+$ sachant qu'il se trouve dans l'état h à l'instant t (Figure 2) est définie par l'équation suivante :

$$\phi^i_{hj}(t,k,l|Z^i_{h,t}(x)) = \mathbf{Pr}(X(t+k+l) = j|X(t+k) = h, X(t) = h, Z^i_{h,t}(x), x \in [0,+\infty)) \quad (7)$$

En utilisant les fonctions de survie, d'intensité et d'intensité cumulée, on montre que l'équation (7) devient :

$$\phi^i_{hj}(t,k,l|Z^i_{h,t}(x)) = \int_{t+k}^{t+k+l} \prod_{r \neq h} \left(\frac{S_{hr}(u|Z^i_{hr,t}(u))}{S_{hr}(t+k|Z^i_{hr,t}(t))} \right) dA_{hj,0}(u) \exp(\beta^t_{hj} Z^i_{hj,t}(u)). \quad (8)$$

3 Estimation de la prédiction : Approche paramétrique et non paramétrique

Estimation paramétrique

Pour chaque transition possible, une distribution a été sélectionnée par le critère AIC [6] et estimée par la méthode du maximum de vraisemblance conditionnelle avec censure à droite. L'estimateur de prédiction s'écrit :

$$\widehat{\phi}^i_{hj}(t,k,l|Z^i_{h,t}(x)) = \int_{t+k}^{t+k+l} \prod_{r \neq h} \left(\frac{\hat{S}_{hr}(u|Z^i_{hr,t}(u))}{\hat{S}_{hr}(t+k|Z^i_{hr,t}(t))} \right) \hat{\alpha}_{hj,0}(u) \exp(\widehat{\beta}^t_{hj} Z^i_{hj,t}(u)) du \quad (9)$$

$\hat{\alpha}_{hj,0}(t) = \alpha_{hj,0}(t|\hat{\theta}_{hj})$ avec $\hat{\theta}_{hj}$ l'estimateur du vecteur des paramètres de la distribution.

Estimation non paramétrique

Les effets des covariables sont estimés par le maximum de vraisemblance partielle de Cox [6], l'intensité cumulée de base par l'estimateur de Breslow [1] $\hat{A}_{hj,0}(.)$ et les autres fonctions sont déduites des relations classiques d'analyse de survie. On note $t+k \leq t_1 < \ldots < t_p \leq \ldots \leq t+k+l$ les temps ordonnés de la transition h, j entre $t+k$ et $t+k+l$. L'estimateur non paramétrique de prédiction s'écrit :

$$\widehat{\phi}^i_{hj}(t,k,l|Z^i_{h,t}(x)) = \sum_{t_p} \prod_{r \neq h} \left(\frac{\hat{S}_{hr}(t_p|Z^i_{hr,t}(t_p))}{\hat{S}_{hr}(t+k|Z^i_{hr,t}(t))} \right) \exp(\widehat{\beta}^t_{hj} Z^i_{hj,t}(t_p)) \Delta \hat{A}_{hj,0}(t_p) \quad (10)$$

où $\Delta \hat{A}_{hj,0}(t_p) = \hat{A}_{hj,0}(t_p) - \hat{A}_{hj,0}(t_{p-1})$

Estimation non paramétrique lissée par la méthode des noyaux
Le risque cumulé de base a été calculé par l'estimateur Breslow. Nous avons estimé l'intensité de transition par l'estimateur à noyau (Gaussien) [7] et la fenêtre optimale de lissage b à été obtenue par validation croisée[4].
L'estimateur de prédiction s'écrit :

$$\widehat{\phi}^i_{hj}(t,k,l|Z^i_{h,t}(x)) = \sum_{t_p}\prod_{r\neq h}\left(\frac{\hat{S}_{hr}(t_i|Z^i_{hr,t}(t_p))}{\hat{S}_{hr}(t+k|Z^i_{hr,t}(t))}\right)\widetilde{\alpha}_{hj}(t_p|Z^i_{hj,t}(t_p))\Delta t_p \tag{11}$$

où

$$\widetilde{\alpha}_{hj}(t_p|Z^i_{hj,t}(t_p)) = \sum_{t_u}\frac{1}{b}K(\frac{t_u-t_p}{b})\exp(\widehat{\beta}^t_{hj}Z^i_{hj,t}(t_u))\Delta\hat{A}_{hj,0}(t_u)$$

$0 \leq t_1 \leq \ldots \leq t_u \leq \ldots \leq t_p$ sont les temps ordonnés de la transition h,j. b est la fenêtre de lissage et K le noyau. $\Delta t_p = t_p - t_{p-1}$

4 Application sur les données

La pneumonie acquise sous ventilation mécanique (VAP) est une pathologie fréquente associée à une lourde mortalité. La prédiction de sa survenue pourrait permettre un traitement plus précoce pour en diminuer son risque et ses conséquences. La pneumonie est plus fréquente chez l'homme et sa survenue est retardée par l'utilisation d'antibiotiques dans les premiers jours de suivi en réanimation.
Nous avons appliqué le modèle sur la base de données de 2873 patients de la réanimation ventilés mécaniquement pour évaluer la prédiction de quatre groupes. Cette prediction est faite dans les trois jours qui suivent l'observation des groupes ($k = 0, l = 3$). Les groupes 1 et 2 correspondent respectivement aux hommes ayant pris et n'ayant pas pris d'antibiotiques à l'admission. Les groupes 3 et 4 correspondent respectivement aux femmes ayant pris et n'ayant pas pris d'antibiotiques à l'admission. Les prédictions paramétriques et non paramétriques de ces quatre groupes sont représentées à la Figure 3.

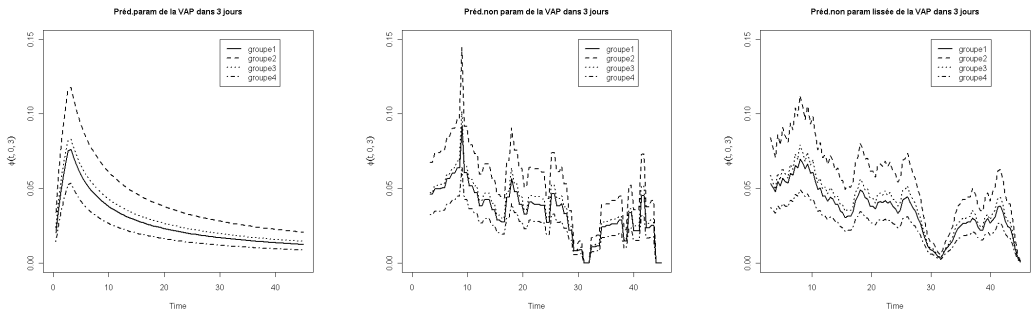

Figure 3: Prédiction paramétrique et non paramétrique

References

[1] P. K. Andersen and N. Keiding. Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11(2):91, 2002.

[2] S. Escolano, J. L. Golmard, A. M. Korinek, and A. Mallet. A multi-state model for evolution of intensive care unit patients: prediction of nosocomial infections and deaths. *Statistics in Medicine*, 19(24):3465–3482, 2000.

[3] P. Hougaard. Multi-state models: A review. *Lifetime Data Analysis*, 5(3):239–264, 1999.

[4] W. Härdle. *Nonparametric and Semiparametric Models*. Springer, 2004.

[5] A. F. Karr. *Point processes and their statistical inference*. Marcel Dekker New York, 1986.

[6] J. P. Klein and M. L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, 2003.

[7] K. B. Kulasekera, C. L. Williams, M. Coffin, and A. Manatunga. Smooth estimation of the reliability function. *Lifetime Data Analysis*, 7(4):415–433, 2001.