

# MODÈLE DE SURVIE RELATIVE ADDITIVE SEMI-MARKOVIEEN

Florence Gillaizeau<sup>\*,†,§,1</sup> & Etienne Dantan<sup>\*,2</sup> & Magali Giral<sup>†,§,3</sup> & Yohann Foucher<sup>\*,§,4</sup>

<sup>\*</sup> EA 4275 - SPHERE - Biostatistique, Pharmacépidémiologie et Mesures Subjectives en Santé, Université de Nantes, 1 rue Gaston Veil, 44035 Nantes, France.

<sup>†</sup> INSERM UMR1064, Institut Transplantation-Urologie-Néphrologie, 30 Boulevard Jean Monnet 44035, Nantes, France.

<sup>§</sup> Centre Hospitalier Universitaire de Nantes, place Ricordeau 44093, Nantes, France.

E-mail: <sup>1</sup>[florence.gillaizeau@univ-nantes.fr](mailto:florence.gillaizeau@univ-nantes.fr); <sup>2</sup>[etienne.dantan@univ-nantes.fr](mailto:etienne.dantan@univ-nantes.fr);

<sup>3</sup>[magali.giral@chu-nantes.fr](mailto:magali.giral@chu-nantes.fr); <sup>4</sup>[yohann.foucher@univ-nantes.fr](mailto:yohann.foucher@univ-nantes.fr)

**Résumé.** Les délais d'apparition d'événements cliniques comme la progression de la maladie ou la mort peuvent être étudiés en utilisant des modèles multi-états. Cependant, lorsque la cause du décès du patient n'est pas disponible ou difficilement attribuable à la maladie, il n'est pas possible d'estimer si certains effets sont associés à une surmortalité liée à la maladie. L'analyse de survie relative permet d'estimer la survie nette (situation où la maladie serait la seule cause possible de décès) en prenant en compte le risque attendu dans la population générale. Dans le contexte des maladies chroniques, il peut être pertinent de considérer des modèles multi-états semi-Markoviens, les intensités de transition entre états dépendant davantage du temps déjà passé dans l'état en cours que du temps chronologique. Dans ce travail, nous proposons un nouveau modèle de survie relative additive semi-Markovien (SRSM). À partir de données simulées, nous montrons l'efficacité du modèle SRSM par rapport à deux modèles: une approche classique qui ne tient pas compte du risque attendu de décès dans la population générale, et un modèle où les différentes causes de décès seraient connues. L'utilité du modèle SRSM est aussi illustrée dans une application sur des patients transplantés rénaux.

**Mots-clés.** Données de survie, Survie relative, Modèles multi-états, Processus semi-Markovien, Progression de la maladie, Transplantation Rénale

**Abstract.** The times until clinical events such as disease progression or death can be studied using multistate models. However, when the cause of death for a patient is unavailable or hardly attributable to the disease, it is not possible to estimate if some effects are associated to an excess mortality related to the disease. Relative survival analysis allows to estimate the net survival (situation where the disease would be the only possible cause of death) taking into account of the expected risk of death in general population. In chronic disease context, it may be relevant to consider semi-Markov multistate models, the transition intensities between states depending more likely on the time already spent

in the current state than on the chronological time. In this paper, we propose a new semi-Markov additive relative survival (SMRS) model. From simulated data, we highlight the effectiveness of the SMRS model compared to two models: a classical approach that do not take account of the expected risk of death in general population, and a model where the different causes of death would be known. The usefulness of SMRS model is also illustrated in an application on kidney transplant recipients.

**Keywords.** Survival data, Relative survival, Multistate models, Semi-Markov Process, Disease Progression, Kidney Transplantation

## 1 Introduction

Les maladies chroniques sont caractérisées par leur longue durée et généralement leur lente progression. Pour étudier le délai de progression vers différents stades, les analyses de survie traditionnelles (Kaplan-Meier, Cox) ne sont pas adaptées et l'utilisation de modèles multi-états est nécessaire [1,2]. Parmi ceux-ci, le modèle semi-Markovien (SM) est intéressant car il considère que la probabilité qu'un patient passe d'un état à l'autre dépend du temps déjà passé dans cet état [3].

Le décès du patient étant souvent le stade terminal de l'évolution de la maladie, il semble important de s'intéresser plus spécifiquement aux décès liés à la maladie étudiée pour personnaliser une stratégie thérapeutique. Cependant, la cause de la mort est rarement disponible ou difficilement attribuable à une cause unique. Les modèles de survie relative à risques additifs ont été introduits comme une réponse méthodologique à cette question: la surmortalité due à une maladie est estimée en comparant les taux de mortalité observés avec ceux attendus dans les tables de mortalité de la population générale [4,5]. Cette décomposition permet d'estimer la survie nette, c'est-à-dire la survie dans la situation hypothétique où la maladie étudiée serait la seule cause possible de décès.

Dans ce travail, nous proposons un nouveau modèle de survie relative additive semi-Markovien (SRSM) qui combine les approches multi-états et survie relative. Dans une première partie, nous présentons le modèle SRSM. Dans une deuxième partie, nous montrons, par simulation, l'efficacité du modèle SRSM dans l'estimation des effets. L'intérêt du modèle SRSM est illustré sur des données de patients transplantés rénaux.

## 2 Le modèle de survie relative semi-Markovien

On note  $T$  le temps chronologique depuis l'origine de l'étude et  $S$  la durée (ou temps de séjour) dans un état. Soit  $\mathcal{X}$  l'espace fini des états cliniques possibles. Le processus stochastique étudié est  $\{X_m, T_m, m \in \mathbb{N}\}$ , où  $X_m$  est l'état du patient après la  $m$ -ième transition se produisant au temps  $T_m$  avec  $T_0 < T_1 < \dots < T_m$  ( $T_0 = 0$  et  $X_0 = 1$  par convention). Soit  $\epsilon$  l'espace des transitions  $ij$  possibles avec  $(i, j) \in (\mathcal{X}, \mathcal{X})$ , où  $i$  représente

un état transitoire avec  $j$  distinct de  $i$ . Nous notons  $Z$  le vecteur des caractéristiques du patient et  $Z_{ij}$  le sous-vecteur des caractéristiques associées à la transition  $ij$ .

Le modèle semi-Markovien (SM) considère que les intensités de transition entre deux états dépendent du temps déjà passé dans l'état en cours. Dans ce travail, nous nous intéressons à un modèle SM à temps homogène, c'est-à-dire qui considère que les intensités de transition entre deux états ne dépendent pas du temps depuis l'origine. La fonction de risque instantané spécifique de l'état  $X_m = i$  à l'état  $X_{m+1} = j$  après une durée  $s$ , sachant les caractéristiques du patient  $Z_{ij} = z_{ij}$  est alors définie par:

$$\lambda_{ij}(s|z_{ij}).ds = P(T_{m+1} - T_m \in ds, X_{m+1} = j | T_{m+1} - T_m > s, X_m = i, z_{ij}) \quad (1)$$

où nous notons  $ds$  la longueur de l'intervalle de temps  $[s, s + ds)$  et l'intervalle lui-même.

En analyse de survie relative additive, le risque de mortalité est supposé être la somme de deux composantes: la mortalité attendue dans la population et la mortalité en excès due à la maladie étudiée [6]. Le modèle de survie relative additive semi-Markovien (SRSM) considère donc un état de plus que le modèle SM en distinguant les deux causes possibles de décès, même si en réalité celles-ci ne sont pas distinguables.

En utilisant les notations  $iO$  et  $iE$  pour les transitions de l'état  $i$  vers le décès quelle que soit sa cause (Observé) et vers le décès lié à la maladie (en Excès), la fonction de risque instantané spécifique au décès lié à la maladie après une durée  $s$  dans l'état  $i$  s'écrit:

$$\lambda_{iE}(s|z_{iE}).ds = P(T_{m+1} - T_m \in ds, X_{m+1} = E | T_{m+1} - T_m > s, X_m = i, z_{iE}) \quad (2)$$

Le risque de mortalité attendu en population  $\lambda^P(\cdot)$  peut être obtenu à partir des taux de décès donnés par les tables de mortalité des instituts nationaux de statistiques. Ces tables indiquent la proportion de décès dans une année calendaire stratifiées sur l'année de naissance et le genre. Soient  $A$  la variable aléatoire de l'âge de décès du patient,  $a_i$  l'âge du patient observé à l'entrée dans l'état  $i$ ,  $y$  l'année de naissance du patient, et  $g$  son genre. La fonction de risque instantané pour la mortalité attendue (non liée à la maladie) après une durée  $s$  dans l'état  $i$ , i.e. à l'âge  $s + a_i$ , est alors:

$$\lambda^P(s + a_i|y, g).ds = P(A \in a_i + ds, X = P | A > s + a_i, y, g) \quad (3)$$

La fonction de risque instantané pour la mortalité toute cause peut alors être réécrite:

$$\lambda_{iO}(s|z_{iE}, a_i, y, g) = \lambda_{iE}(s|z_{iE}) + \lambda^P(s + a_i|y, g) \quad (4)$$

La fonction de risque cumulé correspondante est alors  $\Lambda_{iO}(s|z_{iE}, a_i, y, g) = \Lambda_{iE}(s|z_{iE}) + \Lambda^P(s + a_i|y, g) - \Lambda^P(a_i|y, g)$ , où  $\Lambda_{iE}(s|z_{iE})$  est la fonction de risque cumulé de décès lié à la maladie après une durée  $s$  dans l'état  $i$ , et  $\Lambda^P(s + a_i|y, g) - \Lambda^P(a_i|y, g)$  représente le risque cumulé de décès entre l'âge  $a_i$  et  $a_i + s$  dans la population générale.

La fonction de survie globale après une durée  $s$  dans l'état  $i$ , c'est-à-dire la probabilité de rester dans l'état  $i$  à la durée  $s$ , s'écrit:

$$S_i.(s|z, a_i, y, g) = \exp \left[ - \sum_{\substack{j:ij \in \epsilon \\ j \neq \text{décès}}} \Lambda_{ij}(s|z_{ij}) - \Lambda_{iE}(s|z_{iE}) - \Lambda^P(s + a_i|y, g) + \Lambda^P(a_i|y, g) \right] \quad (5)$$

A partir des équations (1), (4), et (5), la fonction de densité de probabilité spécifique à la transition  $ij$  après une durée  $s$  est:

$$f_{ij}(s|z, a_i, y, g) = \left( \mathbb{1}_{\{j \neq \text{death}\}} \lambda_{ij}(s|z_{ij}) + \mathbb{1}_{\{j = \text{death}\}} \lambda_{iO}(s|z_{iE}, a_i, y, g) \right) S_i.(s|z, a_i, y, g) \quad (6)$$

La contribution d'un patient à la vraisemblance du modèle s'écrit différemment selon le dernier état observé. Soit  $s_{ij}$  la durée dans l'état  $i$  avant la transition vers l'état  $j$ . Pour un patient dans un état terminal à son dernier temps de suivi, la contribution est

$\mathcal{L} = \prod_{ij \in \epsilon} \left\{ f_{ij}(s_{ij}|z, a_i, y, g) \right\}^{\delta_{ij}}$  alors que pour un patient censuré dans un état transitoire  $k$  (pour une durée  $s_k$ ) à son dernier temps de suivi, la contribution est  $\mathcal{L} S_k.(s_k|z, a_k, y, g)$ , avec  $\delta_{ij} = 1$  si la transition  $ij$  est observée, et  $\delta_{ij} = 0$  sinon.

Les fonctions de risque instantané (y compris le risque de décès en excès  $\lambda_{iE}$ ) sont modélisées avec des distributions paramétriques. Nous utilisons les algorithmes de Nelder et Mead [7] pour maximiser la fonction de vraisemblance et calculer la matrice d'Information correspondante.

## 3 Performances du modèle SRSM

### 3.1 Schéma des simulations

Nos choix ont été guidés par les données de patients transplantés rénaux issus de la cohorte DIVAT ([www.divat.fr](http://www.divat.fr)). Nous nous plaçons dans le cas particulier d'un modèle multi-états avec deux états transitoires (l'état d'origine noté  $X = 1$ , et un état  $X = 2$ ) et deux états terminaux observés ( $X = 3$  et le décès qui peut être lié ou non à la maladie ( $X = E$  et  $X = P$ )). Nous supposons qu'aucune transition de retour n'est cliniquement possible. La Figure 1 illustre la structure du modèle.

Nous avons effectué des simulations pour évaluer les performances du modèle SRSM proposé par rapport à deux autres modèles: un modèle SM où les causes de décès seraient connues (5 états) et le modèle SM classique sans distinction entre les décès (4 états). Pour chaque échantillon simulé de taille  $N$ , l'année d'entrée dans l'étude suit une loi uniforme (1998-2010), le genre  $g$  suit une loi binomiale avec 61% d'hommes et une autre variable explicative binaire  $z$  suit une loi binomiale avec une proportion de 30%. L'âge du patient  $a$  à l'entrée dans l'étude a été simulé avec une loi normale tronquée (de 18 à 80 ans) dont les paramètres variaient selon  $g$  et  $z$ . Pour chacune des 5 transitions à modéliser dans l'ensemble  $\{12, 13, 1E, 23, 2E\}$ , les temps de séjour ont été simulés selon une loi de Weibull en fonction de trois variables ( $A, g, z$ ). L'âge de décès attendu a été simulé selon une distribution de Weibull en fonction de  $a$  et  $z$  et sachant que le sujet était vivant au début de l'étude. Les temps de censure ont été simulés selon une distribution de Weibull. Nous avons estimé les paramètres des trois modèles pour 100 jeux de données générés

indépendamment selon six scénarios combinant trois tailles d'échantillon ( $N=500$ ,  $N=1000$ ,  $N=3000$  sujets) et trois taux de censure (15%, 30%, 60%).

## 3.2 Résultats

Nous présentons ici les performances du modèle SRSM pour un taux de censure 60% et  $N=3000$  (Table 1). Celles-ci sont presque aussi bonnes que celle du modèle SM à 5 états. Les biais absolus sont faibles ( $<0,04$ ). Les taux de couverture sont toujours supérieurs à 92%. Les erreurs standards empiriques et asymptotiques sont proches, ce qui reflète une estimation asymptotique correcte de la variance à partir de la matrice d'Information. Des résultats similaires sont observés pour les autres scénarios de simulation.

## 4 Conclusion

Nos simulations montrent de bonnes propriétés du modèle SRSM avec de faibles biais dans les estimations, quels que soient la taille d'échantillon et le taux de censure. Le modèle SRSM a un assez bon ajustement aux données de patients transplantés rénaux. Cette application illustre l'utilité du modèle SRSM pour distinguer l'effet des caractéristiques du patient sur le risque de décès lié ou non à la maladie. Nous envisageons de développer un package dans le logiciel R pour faciliter l'utilisation de tels modèles dans la communauté biostatistique.

## Bibliographie

- [1] Commenges, D. Multi-state models in epidemiology (1999), *Lifetime Data Analysis*, 5(4), 315–327.
- [2] Putter, H. et Fiocco, M. et Geskus, R.B., Tutorial in biostatistics: competing risks and multi-state models (2007), *Statistics in Medicine*, 26(11), 2389–2430.
- [3] Foucher, Y. et Mathieu, E. et Saint-Pierre, P. et Durand, J.F. et Daurès, J.P. (2005), A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease, *Biometrical Journal*, 47(6), 825–833.
- [4] Hakulinen, T. et Tenkanen, L. (1987), Regression analysis of relative survival rates, *Journal of the Royal Statistical Society - Series C*, 36, 309–317.
- [5] Estève, J. et Benhamou, E. et Croasdale, M. et Raymond, L. (1990), Relative survival and the estimation of net survival: elements for further discussion, *Statistics in Medicine*, 9(5), 529–538.
- [6] Perme, M. P. et Stare, J. et Estève, J. (2012), On estimation in relative survival, *Biometrics*, 68(1), 113–120.
- [7] Nelder, J.A. et Mead, R. (1965), A simplex-method for function minimization, *Computer journal*, 7(4), 308–313.

Figure 1: Représentation du modèle SRSM pour un modèle multi-états incluant le décès comme état observé. Le modèle SRSM considère 5 états avec deux types de décès ( $X = P$  et  $X = E$ ). Les flèches pour les transitions vers  $X = E$  et  $X = P$  sont représentées en pointillés car les deux états ne peuvent pas être distingués dans la réalité.

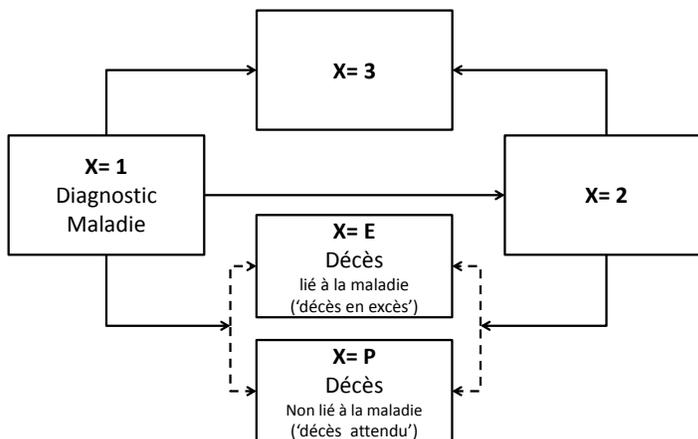


Table 1: Estimations des effets avec le modèle SRSM (100 jeux de données simulées, N=3000 patients, taux de censure=60%)

Coefficient	Valeur théorique <sup>1</sup>	Moyenne estimée	Biais absolu	EQM	ES	ES	Taux de couverture (%)
					Empirique	Asymptotique	
$\beta_{12}$ Homme	0.160	0.165	0.005	0.08	0.08	0.07	94
$\beta_{12}$ Age	-0.012	-0.012	0.000	0.00	0.00	0.00	97
$\beta_{12}$ z	0.210	0.211	0.001	0.08	0.08	0.08	94
$\beta_{13}$ Homme	-0.160	-0.193	-0.033	0.15	0.14	0.15	95
$\beta_{13}$ Age	0.014	0.014	0.000	0.01	0.01	0.01	95
$\beta_{13}$ z	0.910	0.902	-0.008	0.15	0.15	0.15	98
$\beta_{1E}$ Homme	0.180	0.205	0.025	0.23	0.23	0.23	95
$\beta_{1E}$ Age	-0.050	-0.049	0.001	0.01	0.01	0.01	95
$\beta_{1E}$ z	0.600	0.609	0.009	0.26	0.26	0.23	96
$\beta_{23}$ Homme	-0.420	-0.407	0.013	0.24	0.24	0.22	93
$\beta_{23}$ Age	-0.008	-0.007	0.001	0.01	0.01	0.01	97
$\beta_{23}$ z	0.400	0.420	0.020	0.24	0.24	0.23	95
$\beta_{2E}$ Homme	-0.150	-0.108	0.042	0.28	0.28	0.28	97
$\beta_{2E}$ Age	-0.035	-0.035	0.000	0.01	0.01	0.01	94
$\beta_{2E}$ z	0.740	0.770	0.030	0.32	0.32	0.28	92

EQM: Erreur Quadratique Moyenne, ES: Erreur Standard

<sup>1</sup> Valeurs théoriques pour le décès attendu avec la loi de Weibull (log(échelle), log(forme)): transition 12 (2.5,1.2), transition 13 (5.0,-0.3), transition 1E (1.1,0.2), transition 23 (2.8,-0.4), transition 2E (1.3,-0.1), censure à droite (0.6,-0.2).

Valeurs théoriques pour les effets liés au décès attendu:  $\beta_P$  Homme=0.4,  $\beta_P$  Age=0.02.