

Modèle de régression du taux de mortalité en excès par cancer corrigeant une mortalité attendue potentiellement inexacte : estimation avec points de rupture

Darlin R. Mba<sup>1</sup>, Juste A. Goungounga<sup>1</sup>,  
Nathalie Grafféo<sup>1,2</sup>, Roch Giorgi<sup>3</sup>

<sup>1</sup>Aix Marseille Univ, INSERM, IRD, SESSTIM, Marseille, France

<sup>2</sup>Institut Paoli-Calmettes, Marseille, France

<sup>3</sup>Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Hop Timone, BioSTIC, Marseille, France

13ème conférence EPICLIN & 26ème Journée des statisticiens des CLCC

email : [darlin.mba@univ-amu.fr](mailto:darlin.mba@univ-amu.fr)

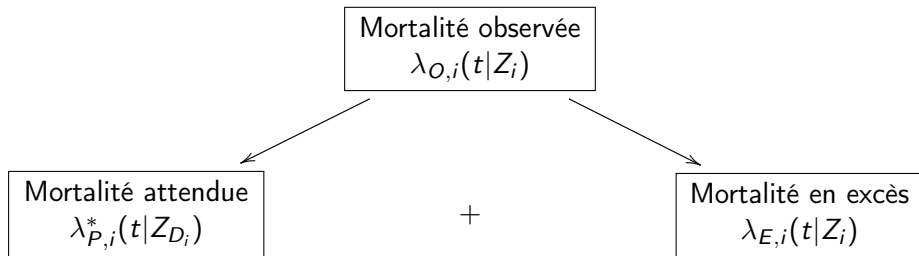
## Survie nette ( $\neq$ survie globale)

- Survie qui serait observée si la seule cause de décès possible était la pathologie d'intérêt (e.g. cancer étudié)
- Registres de cancer (cause de décès non disponible)
- Affranchissement des différences de mortalité dues aux autres causes (comparaisons notamment entre populations et périodes)<sup>1</sup>
- Son estimation repose sur l'utilisation des tables de mortalité (approche paramétrique ou non paramétrique)

---

1. Danieli et al. (2012)

# Modèle additif de mortalité



- ☞  $t$  : temps (délai) depuis le diagnostic
- ☞  $Z_i$  : vecteur des covariables d'intérêt du patient  $i$
- ☞  $Z_{D_i}$  : vecteur de variables démographiques du patient  $i$ , avec  $Z_{D_i} \subset Z_i$
- ☞  $\lambda_{O,i}$  : mortalité observée du patient  $i$
- ☞  $\lambda_{E,i}$  : mortalité en excès du patient  $i$  (imputable au cancer)
- ☞  $\lambda_{P,i}^*$  : mortalité attendue dans la population générale du patient  $i$

## Problème

Table de mortalité standard (appropriée) ?

☞ Dans certains registres de cancer spécialisés en France

Cancer étudié  
(registre de cancer,  
mortalité observée) :  $Z$

- sexe
- année
- âge
- gradient social
- ...

Table de mortalité  
(population générale,  
mortalité attendue) :  $Z_D$

- sexe
- année
- âge
- ~~gradient social~~

☞ Effets de *gradient* et des autres covariables sur la mortalité en excès  $\lambda_E(t|Z)$  biaisés<sup>2</sup>

👉 Modèle classique (Estève et al. (1990))

$$\lambda_{O,i}(t|Z_i) = \lambda_{P,i}^*(t|Z_{D_i}) + \lambda_{E,i}(t|Z_i)$$

☞ Modèle classique (Estève et al. (1990))

$$\lambda_{O,i}(t|Z_i) = \lambda_{P,i}^*(t|Z_{D_i}) + \lambda_{E,i}(t|Z_i)$$

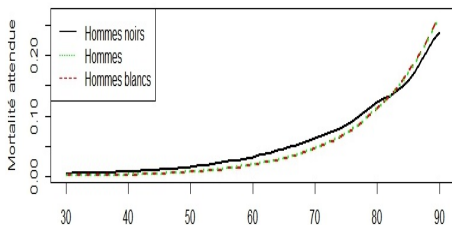
☞ Modèle de régression multivarié (Touraine et al. (2019))

$$\lambda_{O,i}(t|Z_i) = \lambda_{P,i}(t|Z_{D_i}, \mathbf{x}_i = j) + \lambda_{E,i}(t|Z_i)$$

$$\lambda_{O,i}(t|Z_i) = \alpha_j \lambda_{P,i}^*(t|Z_{D_i}) + \lambda_{E,i}(t|Z_i)$$

$j = 1, \dots, J$  : modalité de la variable supplémentaire

- Limite : hypothèse (forte) de proportionnalité entre les risques instantanés des différentes modalités de la variable supplémentaire
- Ce n'est pas toujours le cas en pratique



## Objectif

- Proposer un modèle de régression (flexible) dont la proportionnalité dépend du temps, en introduisant des points de rupture

## Modèle proposé (avec 1 point de rupture) et sa généralisation

### 1 point de rupture ( $\epsilon$ )

$$[\alpha_{j1}\mathbb{1}(\text{age}_i + t < \epsilon) + \alpha_{j2}\mathbb{1}(\text{age}_i + t > \epsilon)]\lambda_{P,i}^*(t|Z_{D_i}) + \lambda_{E,i}(t|Z_i)$$

### Forme généralisée (K points de rupture)

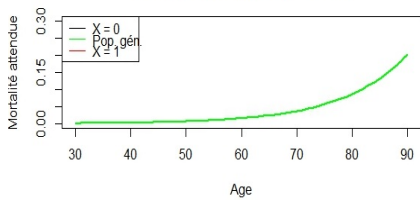
$$\sum_{k=1}^{K+1} \alpha_{jk} \mathbb{1}_k(\text{age}_i + t) \lambda_{P,i}^*(t|Z_{D_i}) + \lambda_{E,i}(t|Z_i)$$

- ☞ Estimation des paramètres par Maximum de Vraisemblance
- ☞ Si connaissance épidémiologique alors point de rupture fixé a priori
- ☞ Sinon, point de rupture déterminé de manière heuristique<sup>3</sup>
  - subdivision de la distribution de l'âge en **déciles**
  - pour k points de rupture,  $C_9^k$  candidats
  - sélection du candidat minimisant le critère BIC

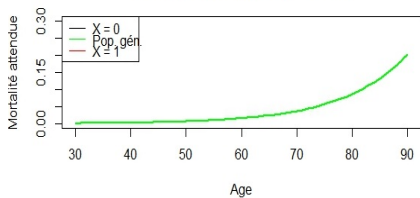
3. Molinari et al. (2001), Bessaoud et al. (2005)



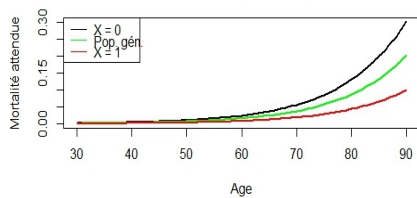
## Scénario A : Table non stratifiée



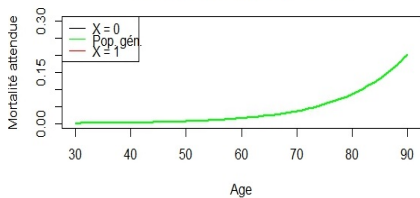
Scénario A : Table non stratifiée



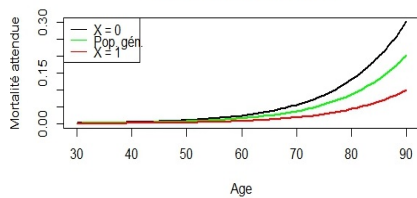
Scénario B : Proportionnalité



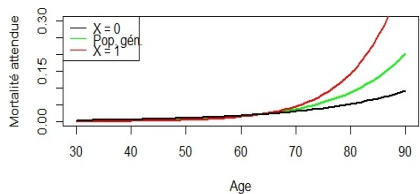
## Scénario A : Table non stratifiée



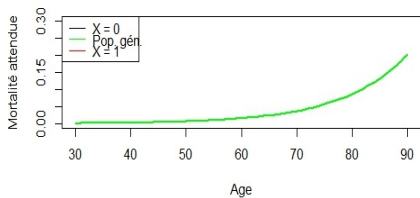
## Scénario B : Proportionnalité



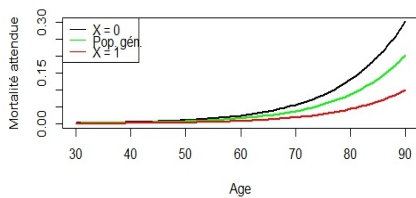
## Scénario C : Non proportionnalité - Croisement



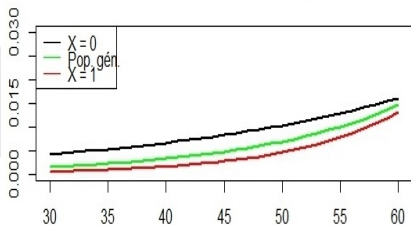
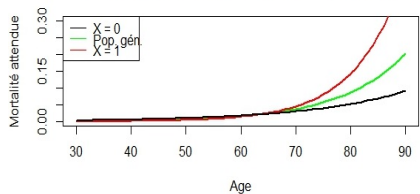
Scénario A : Table non stratifiée



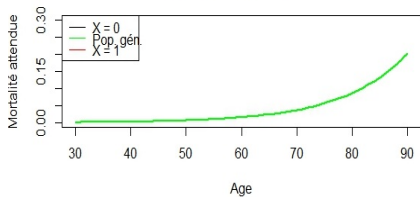
Scénario B : Proportionnalité



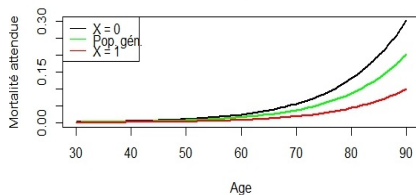
Scénario C : Non proportionnalité - Croisement



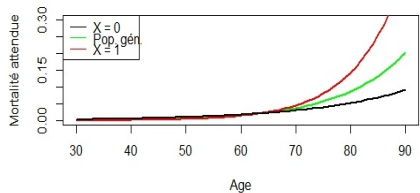
## Scénario A : Table non stratifiée



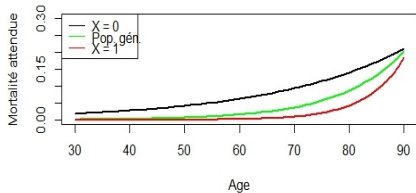
## Scénario B : Proportionnalité



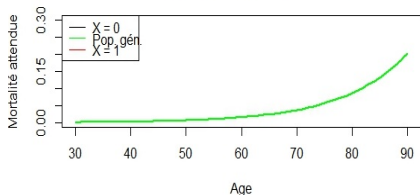
## Scénario C : Non proportionnalité - Croisement



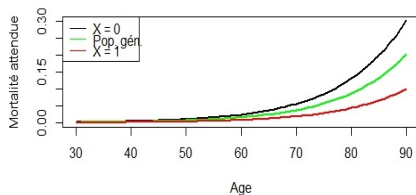
## Scénario D : Non proportionnalité - Convergence



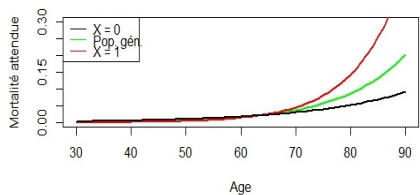
Scénario A : Table non stratifiée



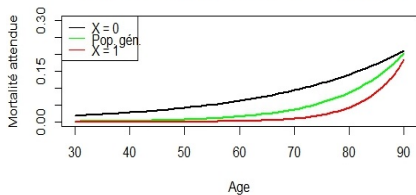
Scénario B : Proportionnalité



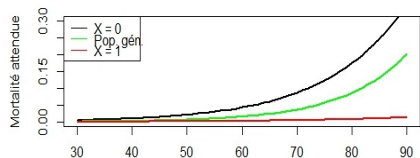
Scénario C : Non proportionnalité - Croisement



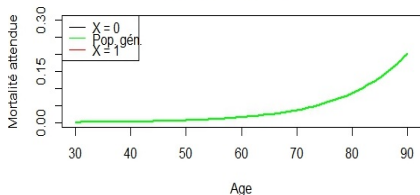
Scénario D : Non proportionnalité - Convergence



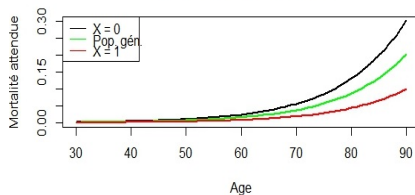
Scénario E : Non proportionnalité - Divergence



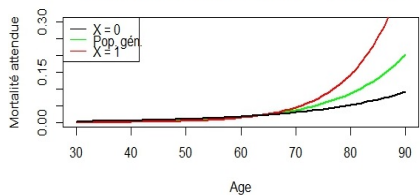
Scénario A : Table non stratifiée



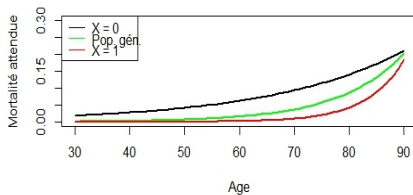
Scénario B : Proportionnalité



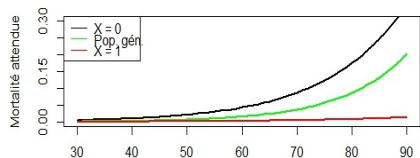
Scénario C : Non proportionnalité - Croisement



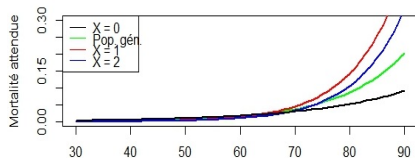
Scénario D : Non proportionnalité - Convergence



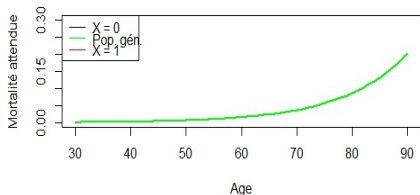
Scénario E : Non proportionnalité - Divergence



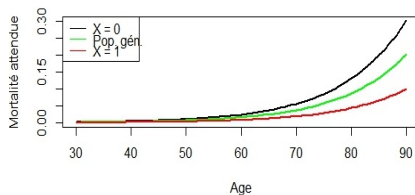
Scénario F : Non proportionnalité - 3 niveaux



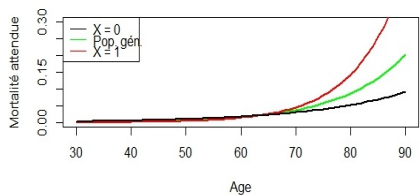
Scénario A : Table non stratifiée



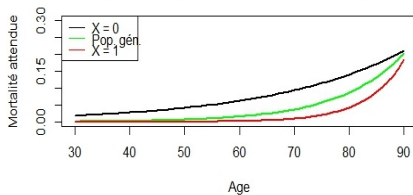
Scénario B : Proportionnalité



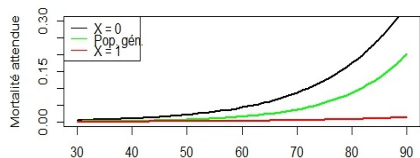
Scénario C : Non proportionnalité - Croisement



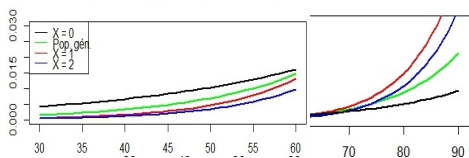
Scénario D : Non proportionnalité - Convergence



Scénario E : Non proportionnalité - Divergence



Scénario F : Non proportionnalité - 3 niveaux





## Scénario A : Table non stratifiée

Modèles	$\beta$	$\hat{\beta}$	Biais	RMSE	ECR	%AIC
Estève et al.	0.30	0.288	-0.012	0.05	95.1	66.7
	-0.20	-0.198	0.002	0.13	95.1	
Touraine et al.	0.30	0.306	0.006	0.08	94.0	08.4
	-0.20	-0.171	0.029	0.21	96.4	
Modèle proposé	0.30	0.300	0.000	0.09	91.4	24.9
	-0.20	-0.136	0.064	0.25	96.9	

## Scénario B : Proportionnalité

Modèles	$\beta$	$\hat{\beta}$	Biais	RMSE	ECR	%AIC
Estève et al.	0.30	0.189	-0.111	0.12	30.1	04.50
	-0.20	-0.566	-0.366	0.39	18.2	
Touraine et al.	0.30	0.308	0.008	0.08	90.5	44.94
	-0.20	-0.190	0.010	0.23	95.6	
Modèle proposé	0.30	0.301	0.001	0.08	90.6	50.56
	-0.20	-0.160	0.040	0.27	94.9	

## Scénario C : Non proportionnalité - Croisement

Modèles	$\beta$	$\hat{\beta}$	Biais	RMSE	ECR	%AIC
Estève et al.	0.30	0.435	0.135	0.14	14.6	00.60
	-0.20	0.188	0.388	0.41	17.7	
Touraine et al.	0.30	0.232	-0.068	0.11	88.5	25.03
	-0.20	-0.381	-0.181	0.27	82.0	
Modèle proposé	0.30	0.251	-0.049	0.10	90.9	74.37
	-0.20	-0.268	-0.068	0.25	93.3	

## Scénario D : Non proportionnalité - Convergence

Modèles	$\beta$	$\hat{\beta}$	Biais	RMSE	ECR	%AIC
Estève et al.	0.30	0.198	-0.102	0.11	40.1	24.33
	-0.20	-0.736	-0.536	0.55	00.6	
Touraine et al.	0.30	0.248	-0.052	0.09	89.2	12.31
	-0.20	-0.529	-0.329	0.39	57.9	
Modèle proposé	0.30	0.263	-0.037	0.09	91.7	63.36
	-0.20	-0.395	-0.195	0.32	76.7	

## Scénario E : Non proportionnalité - Divergence

Modèles	$\beta$	$\hat{\beta}$	Biais	RMSE	ECR	%AIC
Estève et al.	0.30	0.088	-0.212	0.22	00.7	00.00
	-0.20	-0.894	-0.694	0.70	00.1	
Touraine et al.	0.30	0.294	-0.006	0.05	94.7	54.96
	-0.20	-0.261	-0.061	0.24	92.3	
Modèle proposé	0.30	0.285	-0.015	0.06	94.6	45.04
	-0.20	-0.201	-0.001	0.29	93.7	

## Scénario F : Non proportionnalité - 3 niveaux

Modèles	$\beta$	$\hat{\beta}$	Biais	RMSE	ECR	%AIC
Estève et al.	0.30	0.361	0.061	0.08	73.5	13.94
	-0.10	0.257	0.357	0.40	47.1	
	-0.20	0.043	0.243	0.28	61.5	
Touraine et al.	0.30	0.246	-0.054	0.10	89.5	34.50
	-0.10	-0.281	-0.181	0.37	95.8	
	-0.20	-0.337	-0.137	0.25	84.8	
Modèle proposé	0.30	0.259	-0.041	0.09	92.8	51.56
	-0.10	-0.207	-0.107	0.39	96.7	
	-0.20	-0.241	-0.041	0.24	93.7	

- ➡ Modèle multivarié introduisant un point de rupture a des meilleures performances comparativement à un modèle sans point de rupture

## Limites

- ➡ Modèle proportionnel par morceaux
- ➡ Nombre de points de rupture pourrait entraîner un sur-paramétrage du modèle

## Perspectives

- ➡ Implémenter des fonctions plus souples

- Akaike H. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 1974 ; 19 : 716-723.
- Bessaoud F, Daures J-P and Molinari N. Free knot splines for logistic models and threshold selection. Computer Methods and Programs in Biomedicine 2005 ; 77 :1-9.
- Danieli C, Remontet L, Bossard N, Roche L, Belot A. Estimating net survival : the importance of allowing for informative censoring. Statistics in Medicine 2012 ; 31 :775 -786.
- Estève J, Benhamou E, C, Raymond L. Relative survival and estimation of net survival : elements for further discussion. Statistics in Medicine 1990 ; 9 : 529-38.
- Grafféo N, Jooste V, Giorgi R. The impact of additional life-table variables on excess mortality estimates. Statistics in Medicine 2012 ; 31 :4219-30.
- Molinari N, Daures J-P and Durand J-F. Regression splines for threshold selection in survival data analysis. Statistics in Medicine 2001 ; 20 :237–247.
- Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics 1978 ; 6 :461-464.
- Touraine C, Grafféo N, Giorgi R and the CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Statistical Methods in Medical Research 2019.  
<https://doi.org/10.1177/0962280218823234> (in press).



Merci pour votre attention !

