



Sciences Economiques et Sociales de la Santé
& Traitement de l'Information Médicale

sesstim.univ-amu.fr

Darlin Robert MBA

Doctorant au sein de l'équipe
QuanTIM - UMR 1252 - SESSTIM

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

Mai 2019



Cliquez ici pour voir l'intégralité des ressources associées à ce document

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

Darlin Mba¹, Roch Giorgi^{*2}

¹Aix Marseille Univ, INSERM, IRD, SESSTIM, Marseille, France

²Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Hop Timone, BioSTIC, Marseille, France

*Directeur de thèse

email : darlin.mba@univ-amu.fr

Survie nette (\neq survie globale ou brute)

- ☞ 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 pays et périodes)
- ☞ Son estimation repose sur l'utilisation des tables de mortalité (approche paramétrique ou non paramétrique)

Table de mortalité

Hommes	
Age	Prob. de décès
⋮	⋮
30-31	0.0020520
31-32	0.0021522
32-33	0.0022424
33-34	0.0023426
34-35	0.0024529
⋮	⋮
85-86	0.1321354
86-87	0.1441446
87-88	0.1573526
88-89	0.1716155
89-90	0.1870248
⋮	⋮

Femmes	
Age	Prob. de décès
⋮	⋮
30-31	0.0007502
31-32	0.0008003
32-33	0.0008503
33-34	0.0009003
34-35	0.0009504
⋮	⋮
85-86	0.0882394
86-87	0.0984490
87-88	0.1095782
88-89	0.1215099
89-90	0.1345577
⋮	⋮

Table de mortalité

Hommes	
Age	Prob. de décès
⋮	⋮
30-31	0.0020520
31-32	0.0021522
32-33	0.0022424
33-34	0.0023426
34-35	0.0024529
⋮	⋮
85-86	0.1321354
86-87	0.1441446
87-88	0.1573526
88-89	0.1716155
89-90	0.1870248
⋮	⋮

Femmes	
Age	Prob. de décès
⋮	⋮
30-31	0.0007502
31-32	0.0008003
32-33	0.0008503
33-34	0.0009003
34-35	0.0009504
⋮	⋮
85-86	0.0882394
86-87	0.0984490
87-88	0.1095782
88-89	0.1215099
89-90	0.1345577
⋮	⋮

Table de mortalité

Hommes	
Age	Prob. de décès
⋮	⋮
30-31	0.0020520
31-32	0.0021522
32-33	0.0022424
33-34	0.0023426
34-35	0.0024529
⋮	⋮
85-86	0.1321354
86-87	0.1441446
87-88	0.1573526
88-89	0.1716155
89-90	0.1870248
⋮	⋮

Femmes	
Age	Prob. de décès
⋮	⋮
30-31	0.0007502
31-32	0.0008003
32-33	0.0008503
33-34	0.0009003
34-35	0.0009504
⋮	⋮
85-86	0.0882394
86-87	0.0984490
87-88	0.1095782
88-89	0.1215099
89-90	0.1345577
⋮	⋮

Table de mortalité

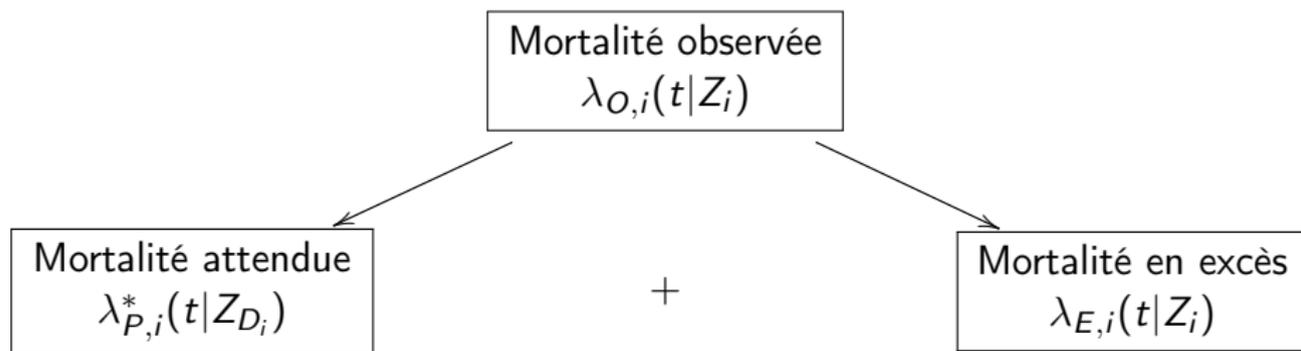
Hommes		Femmes	
Age	Prob. de décès	Age	Prob. de décès
⋮	⋮	⋮	⋮
30-31	0.0020520	30-31	0.0007502
31-32	0.0021522	31-32	0.0008003
32-33	0.0022424	32-33	0.0008503
33-34	0.0023426	33-34	0.0009003
34-35	0.0024529	34-35	0.0009504
⋮	⋮	⋮	⋮
85-86	0.1321354	85-86	0.0882394
86-87	0.1441446	86-87	0.0984490
87-88	0.1573526	87-88	0.1095782
88-89	0.1716155	88-89	0.1215099
89-90	0.1870248	89-90	0.1345577
⋮	⋮	⋮	⋮

👉 En France, Table[année, sexe, âge, département]

👉 Aux Etats-Unis, Table[année, sexe, âge, État, [race](#)]

👉 En Angleterre, Table[année, sexe, âge, comté, [classe sociale](#)]

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~~

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~~

☞ $gradient \in Z$, $gradient \notin Z_D$ (table de mortalité non stratifiée sur $gradient$)

☞ Effets de $gradient$ et des autres covariables sur la mortalité en excès $\lambda_E(t|Z)$ biaisés

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

$$\lambda_{O,i}(t|age_i, x_i) = \lambda_{P,i}^*(t|age_i) + \lambda_{E,i}(t|age_i, x_i)$$

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

$$\lambda_{O,i}(t|age_i, x_i) = \lambda_{P,i}^*(t|age_i) + \lambda_{E,i}(t|age_i, x_i)$$

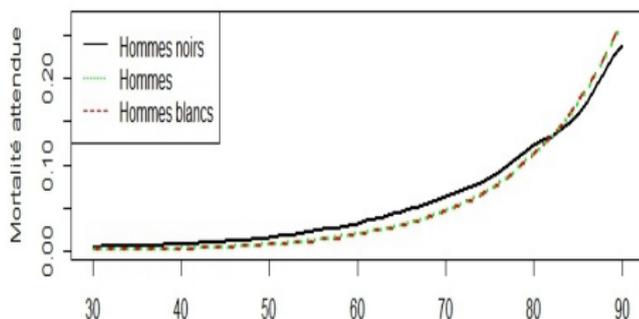
☞ Modèle de régression multivarié (Touraine et al. (2019))

$$\lambda_{O,i}(t|age_i, x_i) = \lambda_{P,i}(t|age_i, x_i = j) + \lambda_{E,i}(t|age_i, x_i)$$

$$\lambda_{O,i}(t|age_i, x_i) = \alpha_j \lambda_{P,i}^*(t|age_i) + \lambda_{E,i}(t|age_i, x_i)$$

$j = 1, \dots, J$: 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 est fonction de l'âge en introduisant des points de rupture

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

1 point de rupture (ϵ)

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

Forme généralisée (K points de rupture)

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

☞ $\{\text{age}_i, x_i\} \in Z$ et $x_i \notin Z_D$

☞ $\alpha_{jk} : (j * k)$ paramètres multiplicatifs corrigeant la mortalité attendue des patients de la modalité j et de la classe d'âge k

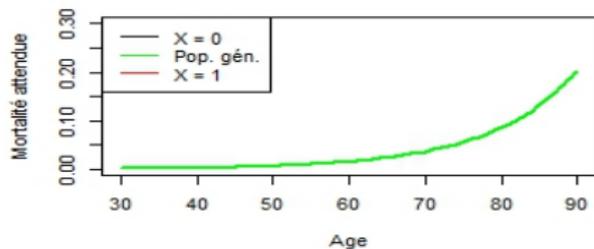
$$\lambda_{E,i}(t|\text{age}_i, x_i) = \lambda_{E,0,i}(t) e^{\beta_1 \text{age}_i + \beta_2 x_i}$$

$$\lambda_{E,0,i}(t) = \sum_{m=1}^M e^{\tau_m} I_m(t)$$

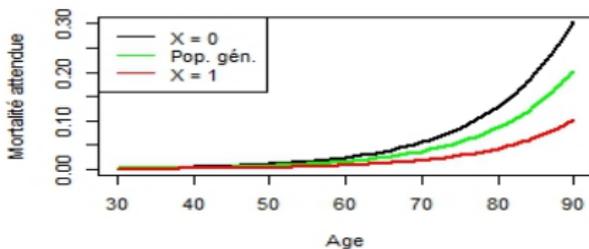
☞ Estimation des paramètres par Maximum de Vraisemblance

☞ Point de rupture déterminé de manière heuristique

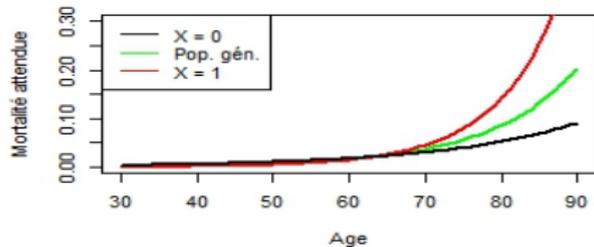
Scénario A : Table non stratifiée



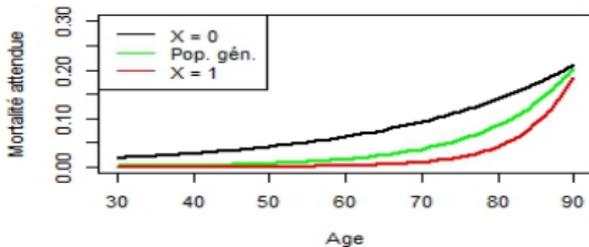
Scénario B : Proportionnalité



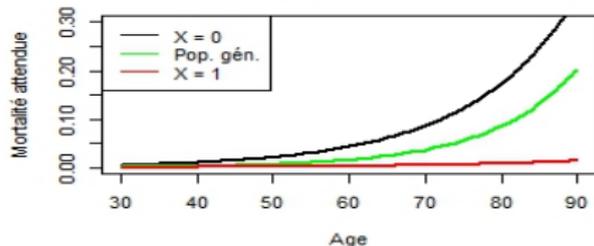
Scénario C : Non proportionnalité - Croisement



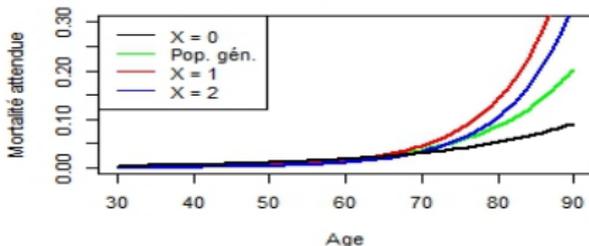
Scénario D : Non proportionnalité - Convergence



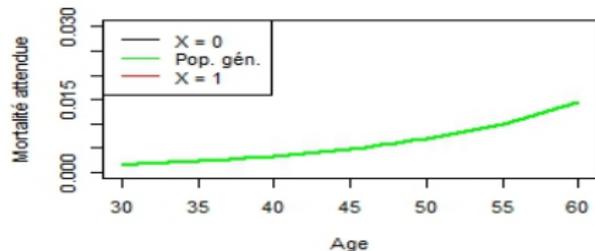
Scénario E : Non proportionnalité - Divergence



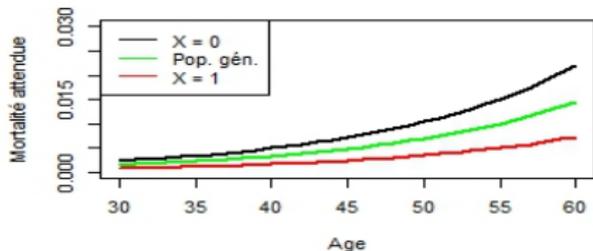
Scénario F : Non proportionnalité - 3 niveaux



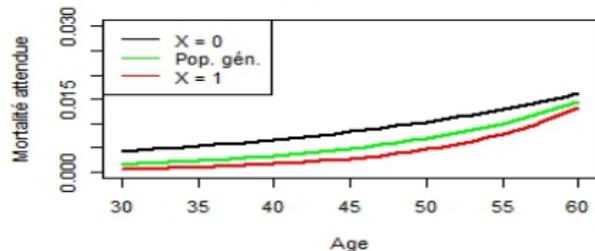
Scénario A : Table non stratifiée



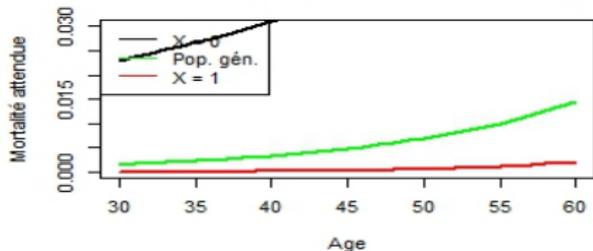
Scénario B : Proportionnalité



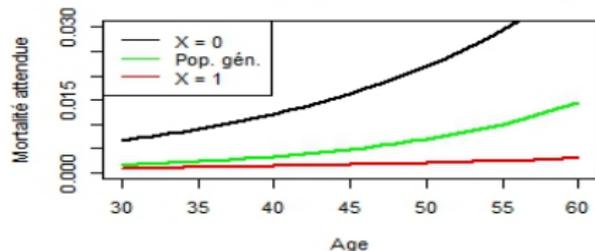
Scénario C : Non proportionnalité - Croisement



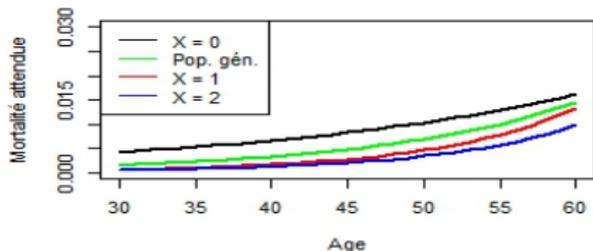
Scénario D : Non proportionnalité - Convergence



Scénario E : Non proportionnalité - Divergence



Scénario F : Non proportionnalité - 3 niveaux



Indicateurs de performance

Biais

Différence entre l'espérance de l'estimateur et la valeur théorique

RMSE : Root Mean Squared Error

Précision de l'estimateur, il est fonction du biais et de la variabilité de l'estimateur

ECR : Empirical Coverage Rate

Proportion d'occurrences où l'intervalle de confiance comprend la valeur théorique du paramètre

%AIC : Akaike Information Criterion

Proportion d'occurrences où un modèle a été sélectionné

Scénario A : Table non stratifiée

Modèles	β	$\hat{\beta}$	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	β	$\hat{\beta}$	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	β	$\hat{\beta}$	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	β	$\hat{\beta}$	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	β	$\hat{\beta}$	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	β	$\hat{\beta}$	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

Limite

- ➡ Modèle proportionnel par morceaux

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 !

