François LEFEBVRE

Aix Marseille Université, Inserm, IRD, SESSTIM, Marseille, France.
Groupe méthode en recherche clinique, service de santé publique,
Hôpitaux universitaires de Strasbourg, France

Utilisation des pseudo-observations pour la modélisation des covariables en analyse de survie

juin 2019

Utilisation des pseudo-résidus pour la modélisation des covariables en analyse de survie

François Lefebvre

Doctorant en biostatistique

SESSTIM (équipe QuanTIM)

Directeur de thèse

Pr Roch Giorgi

Webinar QuanTIM du 21 juin 2019

Contexte

- Étude de survie
  - estimation de la survie
  - étude de l’effet de covariables sur la survie
  - nécessité de modélisation dans le cadre de l’analyse multivariée :
    - modèle multiplicatif
    - modèle additif

Contexte
Modèles multiplicatifs

Modèle de Cox

$$\lambda(t|x_i) = \lambda_0(t) \times e^{\,(\beta^{*'} \times x_i(t))}$$

- $\lambda_0(t)$ est le risque instantané de base
- $\beta^* = (\beta_1^{\ *}, \ldots, \beta_p^{\ *})'$ est un vecteur de paramètres à estimer
- quand $\lambda_0(t)$ n'est pas spécifié, on obtient le modèle de Cox
- possibilité d'extension du modèle de Cox :
  - paramètres pouvant varier dans le temps

Contexte
Modèles multiplicatifs

Modèle de Cox

- il présente des avantages :
  - facilité d'utilisation
  - facilité d'interprétation : si β=0,693 alors HR=2
- mais aussi des limites :
  - proportionnalité des risques instantanés
  - pas nécessairement le plus approprié (effets en biologie ou en épidémiologie)

Contexte
Modèles additifs

Modèle proposé par Aalen (1989)

$$\lambda(t|x_i) = \beta_0(t) + \beta_1(t)x_{i1} + \cdots + \beta_p(t)x_{ip}$$

- $\beta_0(t)$ est le risque instantané de base
- les $\beta_j(t)$ sont les fonctions de régression décrivant les effets des covariables
- Lin a proposé un modèle avec des $\beta_j(t)$ constant dans le temps

Contexte
Modèles additifs
Modèle de Aalen

- Ce modèle présente de nombreux intérêts :
  - effet additif des covariables (modélisation plus adaptée pour les phénomènes biologiques ou épidémiologiques)
  - importance de l'effet ne dépend pas du risque de base (contrairement au modèle de Cox)
  - il n'y a pas d'hypothèses à vérifier
  - description de la variation de l'effet dans le temps
- mais aussi quelques inconvénients :
  - difficulté d'interprétation
  - possibilité de risque instantané négatif
  - peu utilisé

Problématique
comparaison de modèles

- Possibilité de visualiser et de tester :
  - l'adéquation des données à un modèle multiplicatif
  - l'adéquation des données à un modèle additif
- Pas de possibilité de comparer directement et simplement les deux types de modèles

Objectif

- Déterminer au vu des données si l'effet des covariables sur le risque instantané de base est multiplicatif ou additif, notamment quand les méthodes diagnostiques ne permettent pas de rejeter les deux types de modèles
  - proposition d'utiliser les pseudo-résidus
  - pour comparer différents types de modèles

Méthode
pseudo-observation

- Différence entre l'estimation de Kaplan-Meier calculée sur l'ensemble des patients et l'estimation de Kaplan-Meier calculée sur l'ensemble des patients excepté le patient i

$$\hat{S}_i(t) = n\hat{S}(t) - (n-1)\hat{S}^{-i}(t)$$

  - Permet de mesurer à chaque temps l'effet du sujet i sur la survie

Méthode
pseudo-résidu

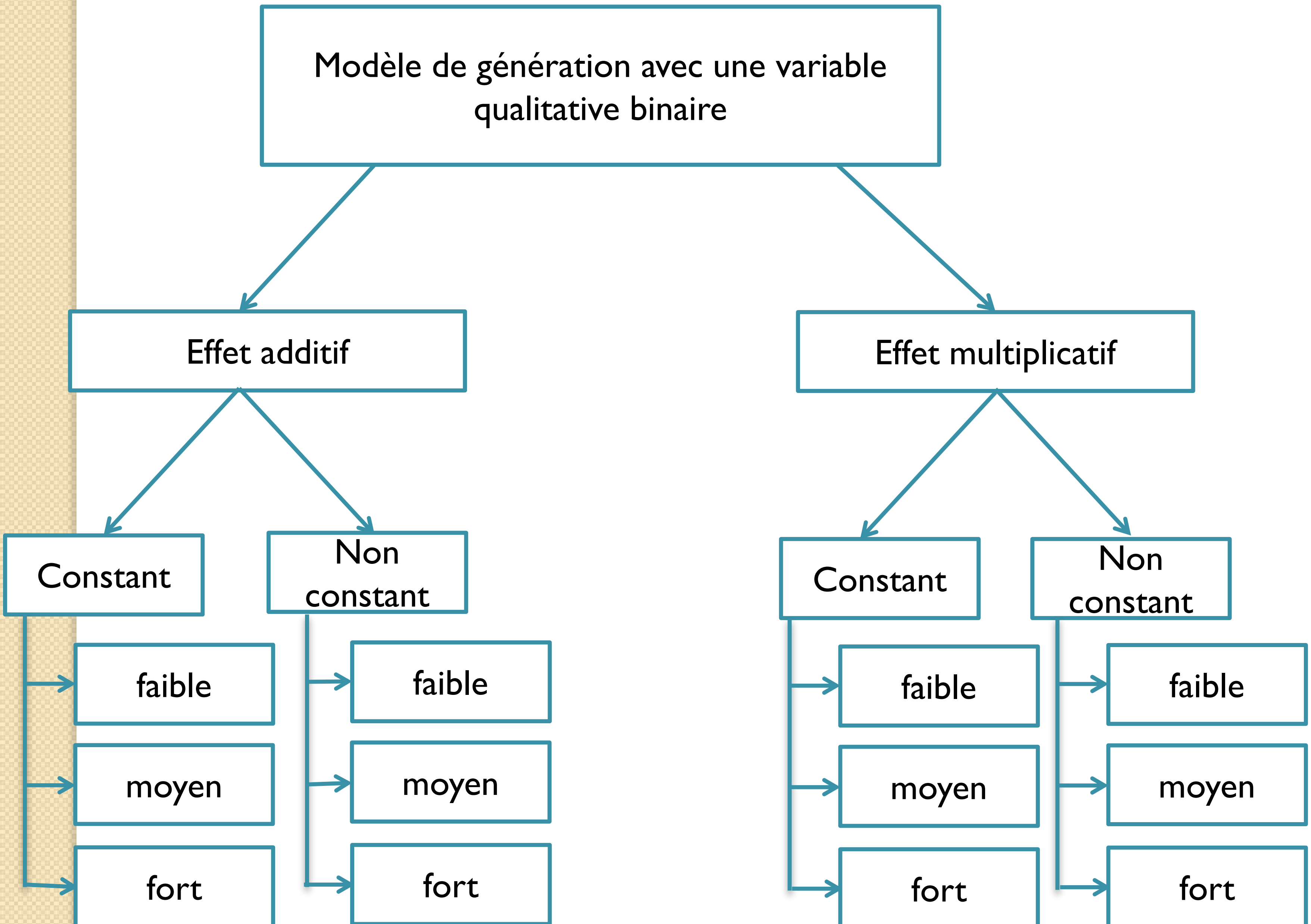
- Différence entre la pseudo-observation et la survie estimée pour le sujet i par un modèle de survie

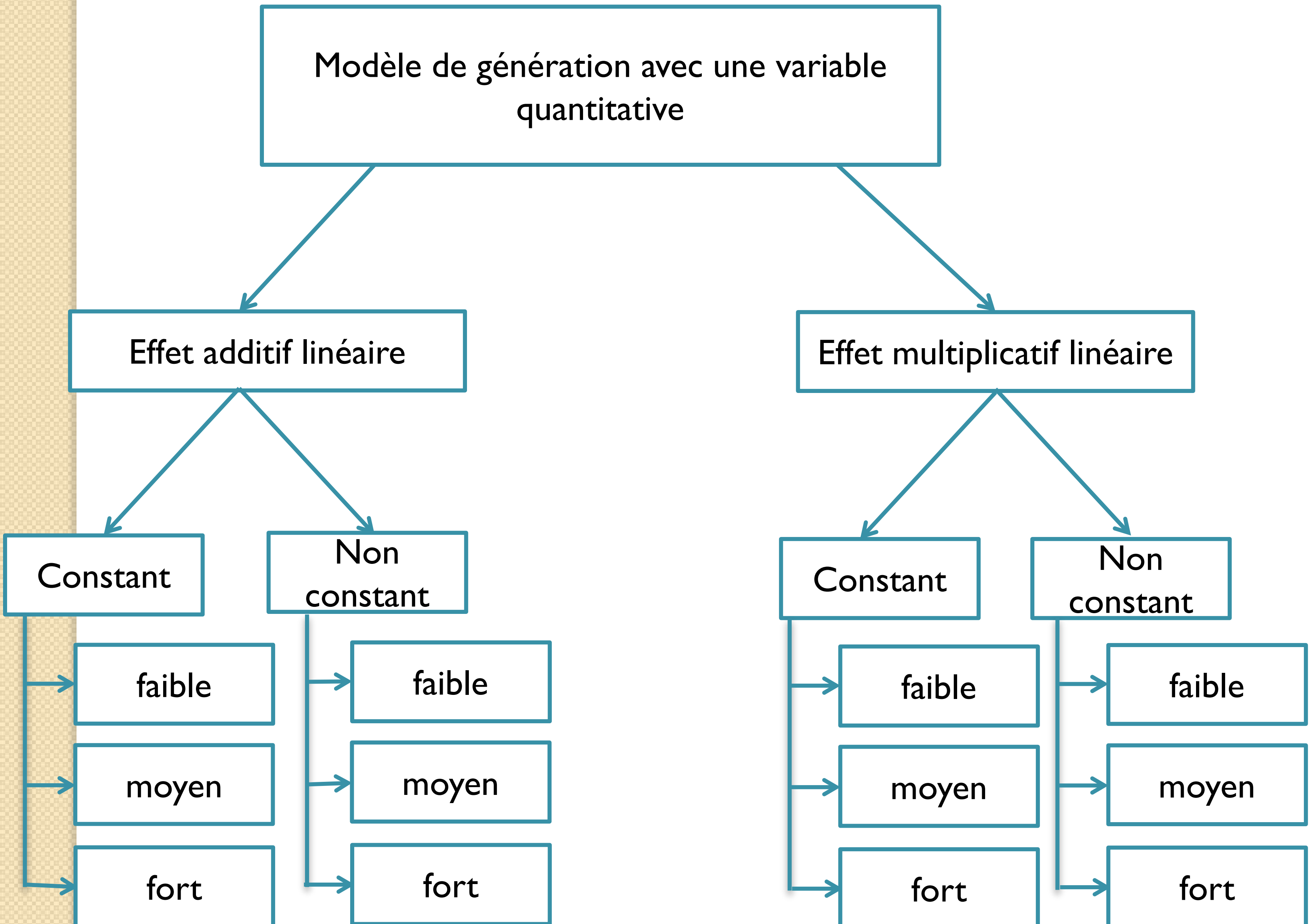
$$\hat{\varepsilon}_i(t) = \hat{S}_i(t) - \hat{S}(t|z_i)$$

  - $\hat{S}(t|z_i)$: survie estimée pour le sujet i avec un modèle
  - mesure une distance entre une estimation non-paramétrique des données et une estimation par un modèle
  - la somme des pseudo-résidus au carré représente un écart quadratique qui est d'autant plus faible que l'ajustement est bon
  - la comparaison de ces sommes de pseudo-résidus au carré permet de choisir le modèle le plus adapté aux données

Méthode
performances

- Étude de simulations
- Génération des données
  - risque de base suit une Weibull généralisée
  - covariable :
    - Effet multiplicatif et additif
    - Effet constant et non-constant
    - Effet linéaire et non linéaire
    - Effet plus ou moins important
    - Effectifs de 500, 1000 et 2000


Modèle de génération avec une variable qualitative binaire
Effet additif
Effet multiplicatif
Constant
Non constant
faible
moyen
fort
faible
moyen
fort
Constant
Non constant
faible
moyen
fort
faible
moyen
fort

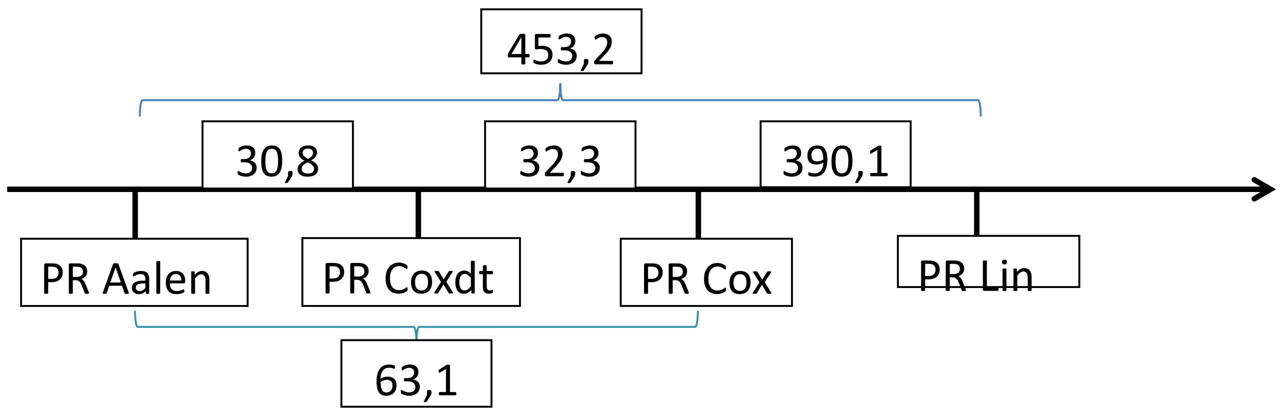

Modèle de génération avec une variable quantitative
Effet additif linéaire
Effet multiplicatif linéaire
Constant
Non constant
faible
moyen
fort
faible
moyen
fort
Constant
Non constant
faible
moyen
fort
faible
moyen
fort

Méthode

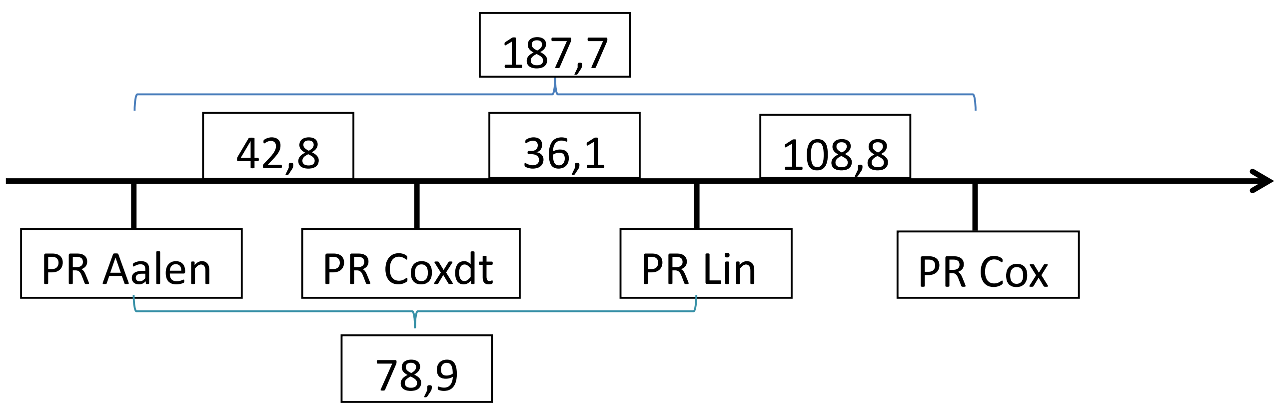
- Analyse des données
  - Modèle de Lin qui est un modèle additif correspondant au modèle de Cox (modèle additif semi paramétrique)
  - Modèle de Aalen
  - Modèle de Cox
  - Modèle de Cox avec effet dépendant du temps linéaire (pour assouplir l'hypothèse de proportionnalité)

Résultats variable qualitative binaire

Modèle de génération multiplicatif (effet constant moyen effectif 1000)



Modèle de génération additif (effet constant moyen effectif 1000)



Résultats variable qualitative binaire

Génération multiplicative

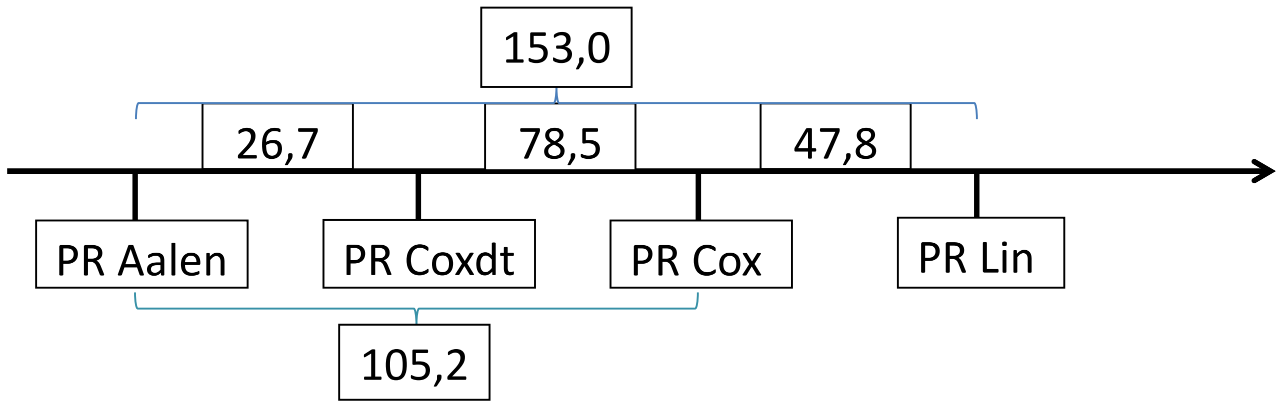
	Cox	Cox dt	Lin
Modèle retenu			
N=1000			
Cox dt	86,9 %		
Lin	5,0 %	0,8 %	
Aalen	100,0 %	100,0 %	100,0 %

Génération additive

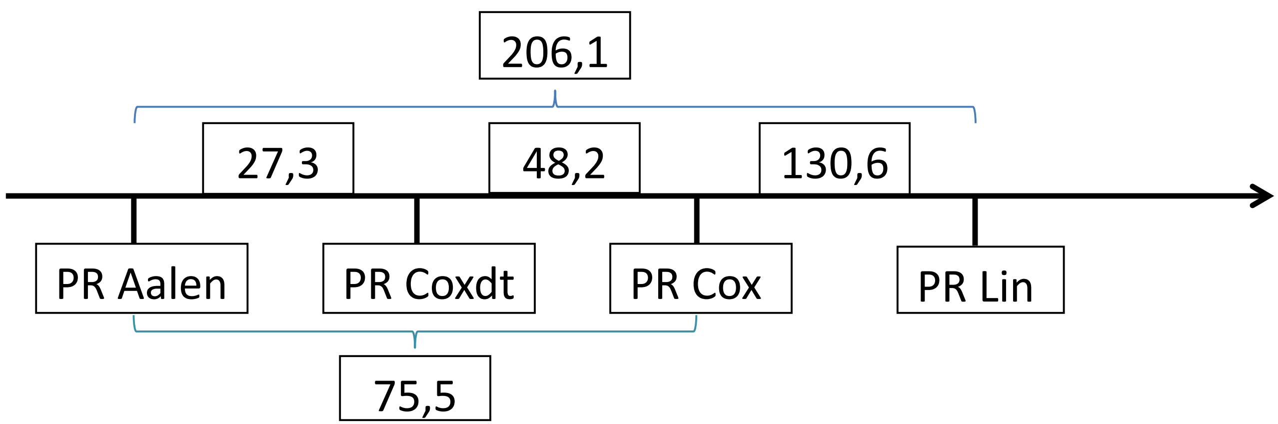
	Cox	Cox dt	Lin
Modèle retenu			
N=1000			
Cox dt	97,9 %		
Lin	83,1 %	26,8 %	
Aalen	100,0 %	100,0 %	100,0 %

Résultats variable qualitative binaire dépendant du temps

Modèle de génération multiplicatif (effet non constant moyen effectif 1000)



Modèle de génération additif (effet non constant moyen effectif 1000)



Résultats variable qualitative binaire dépendant du temps

Génération multiplicative

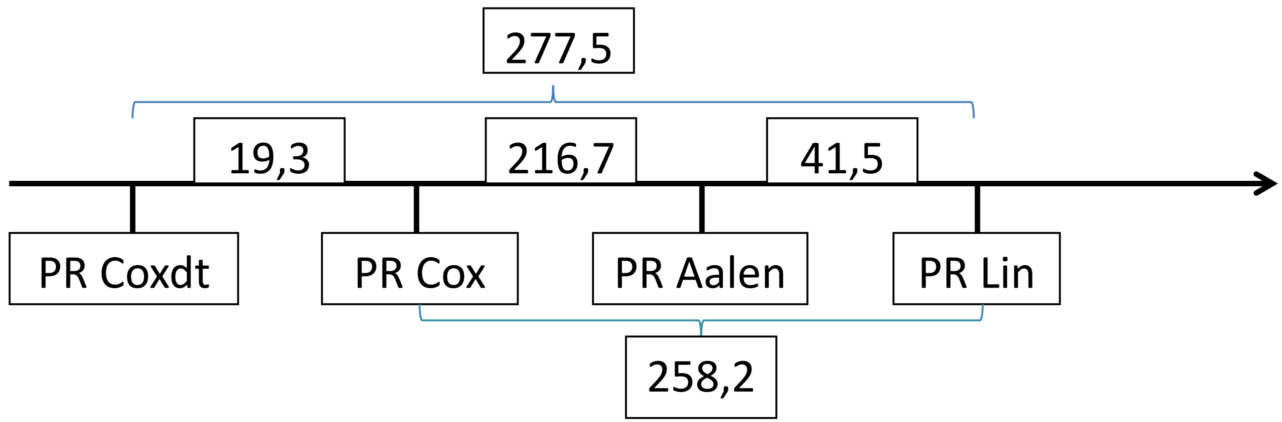
	Cox	Cox dt	Lin
Modèle retenu			
N=1000			
Cox dt	95,5 %		
Lin	18,5 %	2,0 %	
Aalen	100,0 %	100,0 %	100,0 %

Génération additive

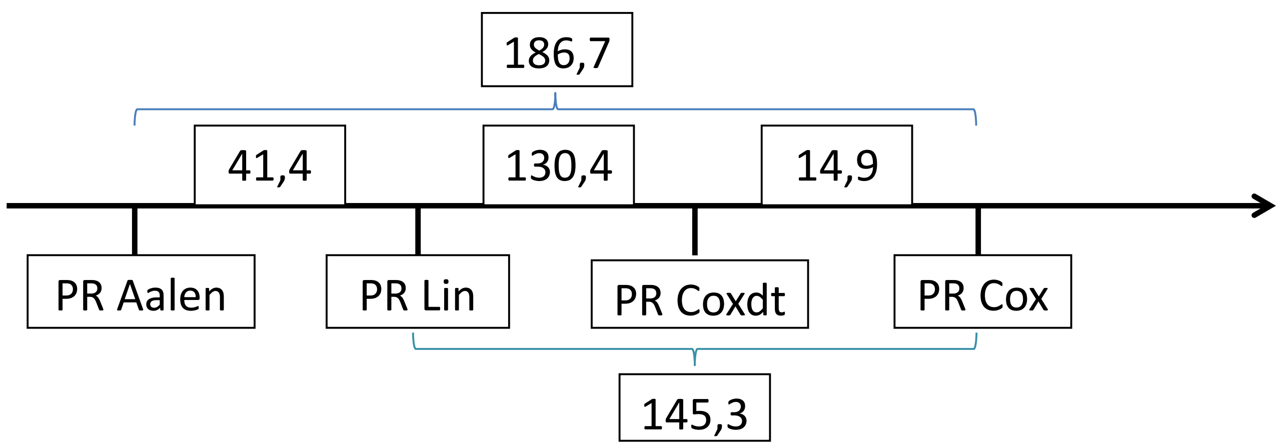
	Cox	Cox dt	Lin
Modèle retenu			
N=1000			
Cox dt	93,5 %		
Lin	8,5 %	4,0 %	
Aalen	100,0 %	100,0 %	100,0 %

Résultats variable quantitative linéaire

Modèle de génération multiplicatif (effet constant moyen effectif 1000)



Modèle de génération additif (effet constant moyen effectif 1000)



Résultats variable quantitative linéaire

Génération multiplicative

	Cox	Cox dt	Lin
Modèle retenu			
N=1000			
Cox dt	79,0 %		
Lin	29,5 %	28,0 %	
Aalen	33,0 %	32,5 %	92,5 %

Génération additive

	Cox	Cox dt	Lin
Modèle retenu			
N=1000			
Cox dt	74,0 %		
Lin	72,0 %	69,0 %	
Aalen	79,0 %	75,5 %	99,5 %

Conclusion
Variable qualitative binomiale

- Le modèle de Aalen s'ajuste toujours mieux : il est donc à privilégier
- Néanmoins, pour des raisons d'interprétation
  - si l'écart entre la somme des PR^2 avec Aalen et avec Cox est faible : on peut utiliser Cox
  - si l'écart entre la somme des PR^2 avec Aalen et Lin est faible : on peut utiliser Lin

Conclusion
Variable quantitative

- Le modèle qui a la plus petite somme des PR^2 s'ajuste le mieux
- Néanmoins, pour des raisons d'interprétation
  - si l'écart entre la somme des PR^2 avec le modèle qui s'ajuste le mieux et un autre est faible, on peut utiliser ce dernier

Perspectives

- Étendre l'étude à des effets non-linéaires et non-constants
- Étendre l'étude à des modèles avec des covariables qualitatives et quantitatives
- Étendre l'étude à la comparaison de modèles en survie nette

Références

- Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in medicine, 8* (8), 907-925.

- Aalen, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in medicine, 12* (17), 1569-1588.

- Lin, D. Y., & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika, 81* (1), 61-71.

Références

- Andersen, P. K., Klein, J. P., & Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, *90*(1), 15-27.

- Perme, M. P., & Andersen, P. K. (2008). Checking hazard regression models using pseudo-observations. *Statistics in medicine*, *27*(25), 5309-5328.

- Andersen, P. K., & Pohar Perme, M. (2010). Pseudo-observations in survival analysis. *Statistical methods in medical research*, *19*(1), 71-99.