

Extension de l'analyse de la survie nette au domaine de la recherche clinique

UMR 1252 SESSTIM

Thèse présentée par Juste GOUNGOUNGA

Sous la direction du Pr Roch GIORGI

pour obtenir le grade de

Docteur d'Aix-Marseille Université

03 Décembre 2018



Faculté
de Médecine
Aix-Marseille Université



A*Midex
Initiative d'excellence Aix-Marseille

Plan de présentation

Contexte de la recherche clinique : survie globale et survie nette

Partie 1 : limites des méthodes d'estimation de la survie nette en recherche clinique

Partie 2 : prise en compte de la non-comparabilité en termes de mortalité due aux autres causes (situation unicentrique)

Partie 3 : extension à la modélisation de l'hétérogénéité inter-centres en recherche clinique (situation multicentrique)

Conclusion

Survie globale et mortalité observée

- L'analyse des données de survie : étude du délai écoulé jusqu'à la survenue de l'évènement d'intérêt

Survie globale et mortalité observée

- L'analyse des données de survie : étude du délai écoulé jusqu'à la survenue de l'évènement d'intérêt
 - Survie globale $S_O(t)$: évènement « décès quelle que soit la cause »
 - ☞ Probabilité qu'un patient reste en vie jusqu'au temps t après le diagnostic
 - ☞ Relation entre survie globale $S_O(t)$ et mortalité observée $\lambda_O(t)$

Survie globale et mortalité observée

- L'analyse des données de survie : étude du délai écoulé jusqu'à la survenue de l'évènement d'intérêt
 - Survie globale $S_O(t)$: évènement « décès quelle que soit la cause »
 - ☞ Probabilité qu'un patient reste en vie jusqu'au temps t après le diagnostic
 - ☞ Relation entre survie globale $S_O(t)$ et mortalité observée $\lambda_O(t)$

$$S_O(t) = \exp \left[- \int_0^t \lambda_O(u) du \right]$$

Survie nette et mortalité en excès

- L'analyse des données de survie : étude du délai écoulé jusqu'à la survenue de l'évènement d'intérêt
 - Survie nette $S_E(t)$: évènement « décès dû au cancer étudié »

Survie nette et mortalité en excès

- L'analyse des données de survie : étude du délai écoulé jusqu'à la survenue de l'évènement d'intérêt
 - Survie nette $S_E(t)$: évènement « décès dû au cancer étudié »
 - ☞ Survie que l'on observerait si le cancer étudié était la seule cause possible de décès¹
 - ☞ Relation entre survie nette $S_E(t)$ et mortalité excès $\lambda_E(t)$

1. ESTEVE et al. 1990.

Survie nette et mortalité en excès

- L'analyse des données de survie : étude du délai écoulé jusqu'à la survenue de l'évènement d'intérêt
 - Survie nette $S_E(t)$: évènement « décès dû au cancer étudié »
 - ☞ Survie que l'on observerait si le cancer étudié était la seule cause possible de décès¹
 - ☞ Relation entre survie nette $S_E(t)$ et mortalité excès $\lambda_E(t)$

$$S_E(t) = \exp \left[- \int_0^t \lambda_E(u) du \right]$$

1. ESTEVE et al. 1990.

Estimation de la survie nette : cause de décès connue

- Estimation de la survie nette pour un cancer donné
 - Approche traditionnelle : survie **spécifique** ou méthode cause-spécifique²
 - ☞ Censurer les individus vivants
 - ☞ Censurer les décès dus aux autres causes
 - ☞ Utiliser les décès dus au cancer (Kaplan-Meier, Cox)

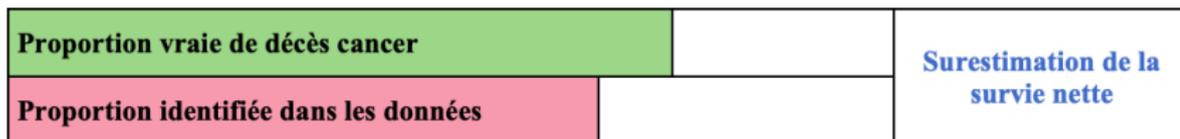
Méthode de survie spécifique : biais de classification

- Potentiels biais : erreurs de classification de la cause de décès³

3. SCHAFFAR et al. 2015.

Méthode de survie spécifique : biais de classification

- Potentiels biais : erreurs de classification de la cause de décès³



3. SCHAFFAR et al. 2015.

Études populationnelles : cause de décès non disponible

- Études populationnelles : méthodes « populationnelles »
- Registre de cancer : enregistrement continu et exhaustif des diagnostics de cancer sur un territoire défini

➔ Sous l'hypothèse d'indépendance

$$\underbrace{\lambda_{O,i}(t)}_{\text{Décès}} = \underbrace{\lambda_{E,i}(t)}_{\text{Cancer}} + \underbrace{\lambda_{P,i}(t)}_{\text{Autres causes}}$$

Exemple de table de mortalité

- Données sur la mortalité attendue d'une population donnée
- Taux de mortalité journaliers pour un individu donné :
 $-\log(1 - q)/365.25$
- q probabilité de décès d'un individu qui va avoir un âge donné dans l'année

Table de mortalité disponible sur R

```
> head(as.table(survival::survexp.us[, , "2000"]), 7)
  sex
age  male      female
0  2.096990e-05  1.719282e-05
1  1.506232e-06  1.259701e-06
2  1.040581e-06  7.667056e-07
3  7.940919e-07  6.297781e-07
4  6.297781e-07  5.202411e-07
5  5.750090e-07  4.654742e-07
6  5.202411e-07  4.107084e-07
```

$1 - q$: Probabilité pour un individu qui aura 5 ans en 2000 d'atteindre son 6^{eme} anniversaire

Partie 1

[Journal of Cancer Research and Clinical Oncology](#)

└ April 2018, Volume 144, [Issue 4](#), pp 791–793 | [Cite as](#)

Commentary on: Survival benefit of mantle cell lymphoma patients enrolled in clinical trials; a joint study from the LYSA group and French cancer registries

Authors

[Authors and affiliations](#)

Juste Aristide Goungounga, Roch Giorgi 

Letter to the Editor

First Online: 29 January 2018

126

Downloads 26 November 2018

To the Editors:

We read with great interest the article by Augustin et al. (2017) comparing the prognosis of patients diagnosed with mantle cell lymphoma (MCL) according to their source of recruitment, population-based or clinical trials. They have shown that patients recruited in clinical trials have a better prognosis, in terms of net survival, than the entire population of patients diagnosed with MCL, and provided conclusions in favor of reconsideration of the inclusion criteria to have samples more representative of MCL patients. Such a comparative study is important, because of its potential impact on clinical research, as the authors rightly argue, but we want to highlight the fact that the difference pointed out in this study may be biased.



Objectif général - Objectifs spécifiques

- Clarifier l'intérêt de la survie nette en recherche clinique

Objectif général - Objectifs spécifiques

- Clarifier l'intérêt de la survie nette en recherche clinique
 - ☞ Présenter quelques méthodes d'estimation de la survie nette
 - ☞ Identifier l'impact de l'utilisation des tables de mortalité en recherche clinique

- Méthode cause-spécifique (cause de décès connue)
 - ☞ Approche classique
 - ☞ Approche avec la pondération IPCW

- Méthode cause-spécifique (cause de décès connue)
 - ☞ Approche classique
 - ☞ Approche avec la pondération IPCW

- Méthodes « populationnelles » (cause de décès inconnue)
 - ☞ Approche du « ratio-estimate »⁴
 - ☞ Approche paramétrique d'estimation du taux en excès
 - ☞ Avec pondération IPCW

4. EDERER 1961 ; EDERER et HEISE 1959 ; HAKULINEN 1982.

Méthode cause-spécifique classique

$$\hat{S}_E(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \quad (\text{KAPLAN et MEIER 1958})$$

$$\hat{S}_E(t, X) = \exp \left(\overbrace{- \sum_{i=1}^n \frac{1(t_i \leq t) \delta_i}{\sum_{j \in R(t_i)} \exp \beta X_j}}^{\text{Breslow}} \right) \exp(\beta X) \quad (\text{COX 1975})$$

Méthode de survie spécifique avec pondération IPCW⁵

$$\hat{S}_E^w(t) = \exp\left(-\int_0^t \frac{dN_E^w(u)}{Y^w(u)}\right) \quad (\text{Nelson-Aalen pondéré})$$

- $N_{Ei}(t) = 1(t_{O_i} \leq t, T_{O_i} \leq C_i, t_{E_i} \leq t_{P_i})$
- dN_E représentation différentielle du processus N_E
- $Y_i(t)$ Indicateur du statut à risque pour chaque individu i
- $dN_{Ei}^w(t) = \frac{dN_{Ei}(t)}{S_{P_i}(t_i^-)}$ et $Y_i^w(t) = \frac{Y_i(t)}{S_{P_i}(t^-)}$
- $S_{P_i}(t_i^-)$: Survie attendue dans la population générale

5. PERME, STARE et ESTÈVE 2012 ; MORISOT 2015 ; SCHAFFAR et al. 2017.

Taux de base en excès constant par intervalles ou B-splines

$$\overbrace{\exp(\beta X) \sum_k^r \tau_k I_k(t)}^{\hat{\lambda}_E(t|X)} = \lambda_O(t | X) - \lambda_P(t | Z_D)$$

(ESTEVE et al. 1990)

Taux de base en excès constant par intervalles ou B-splines

$$\underbrace{\hat{\lambda}_E(t|X)}_{\exp(\beta X) \sum_k^r \tau_k I_k(t)} = \lambda_O(t | X) - \lambda_P(t | Z_D)$$

(ESTEVE et al. 1990)

$$\underbrace{\sum_{j=-2}^2 \nu_j \mathbb{B}_{j,3}(t) \exp \left[\sum_{i=1}^{p-l} \sum_{j=-2}^2 \nu_j^* \mathbb{B}_{j,3}(t) X_i + \left(\sum_{h=1-l+p}^p \beta X_h \right) \right]}_{\hat{\lambda}_E(t|X)} = \lambda_O(t | X) - \lambda_P(t | Z_D)$$

(GIORGI et al. 2003)

Méthode populationnelle avec pondération IPCW

- Estimateur non paramétrique

$$\hat{\Lambda}_E(t) = \underbrace{\int_0^t \frac{dN^w(u)}{Y^w(u)}}_{\text{Observé}} - \underbrace{\int_0^t \frac{\sum_{i=1}^n Y_i^w(u) d\Lambda_{P_i}(u)}{Y^w(u)}}_{\text{Population}}$$

(PERME, STARE et ESTÈVE 2012)

- ↪ $N_i(t) = 1$ ($t_{O_i} \leq t$, $T_{O_i} \leq C_i$)
- ↪ dN représentation différentielle du processus N
- ↪ $Y_i(t)$ Indicateur du statut à risque pour l'individu i
- ↪ $dN_i^w(t) = \frac{dN_i(t)}{S_{P_i}(t_i^-)}$ et $Y_i^w(t) = \frac{Y_i(t)}{S_{P_i}(t^-)}$
- ↪ $S_{P_i}(t_i^-)$: Survie attendue dans la population générale

Exemple d'introduction du concept de survie nette en recherche clinique (1/2)

- Objectif de l'article⁶ : comparaison de
 - ☞ Caractéristiques cliniques
 - ☞ **Survie nette** des patients (lymphome du manteau)
- Données de 2008-2012 : deux essais cliniques (phase II et III, 372 patients) vs six registres de cancers (312 patients)

6. AUGUSTIN et al. 2016.

7. NELSON et al. 2007.

Exemple d'introduction du concept de survie nette en recherche clinique (1/2)

- Objectif de l'article⁶ : comparaison de
 - ☞ Caractéristiques cliniques
 - ☞ **Survie nette** des patients (lymphome du manteau)
- Données de 2008-2012 : deux essais cliniques (phase II et III, 372 patients) vs six registres de cancers (312 patients)
- Méthodes utilisées
 - ☞ Régression logistique : variables associées à l'inclusion
 - ☞ Estimateur de Pohar-Perme : survie nette
 - ☞ Modèle de régression de Nelson et al.⁷ : survie nette et facteurs pronostiques

6. AUGUSTIN et al. 2016.

7. NELSON et al. 2007.

Exemple d'introduction du concept de survie nette en recherche clinique (2/2)

■ Résultats

- ☞ Meilleur pronostic dans les essais vs registres
 - Plus jeunes (âge médian 60 vs 74 ans)
 - Meilleure survie nette après 4 ans (79,9% vs 60,3%)
 - Faible mortalité en excès : plus de 65 ans

■ Conclusion

- ☞ Patients inclus dans les essais \neq pratique médicale quotidienne
- ☞ Élargir les critères de sélection \Rightarrow améliorer la représentativité de la population générale

Méthodes d'estimation de la survie nette dans l'article : potentiels biais en recherche clinique

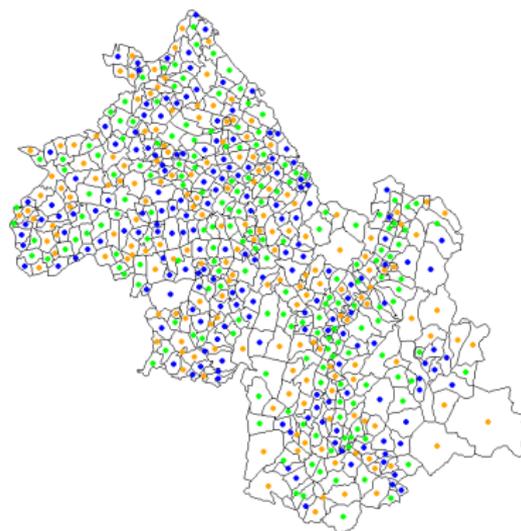
- Estimateur de Pohar-Perme et modèle de NELSON et al. 2007
- L'hypothèse de comparabilité est **invalidé** :
 - Si un facteur influençant la mortalité autres causes est distribuée différemment dans l'échantillon étudié et dans la population générale⁸
 - Essai clinique : patients sélectionnés selon des critères spécifiques

8. SARFATI, BLAKELY et PEARCE 2010.

Exemple de sélection des patients en recherche clinique

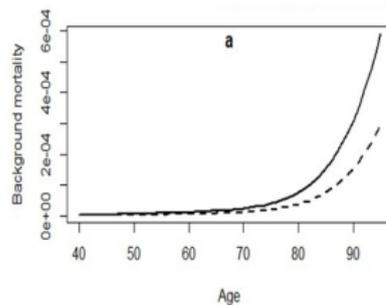
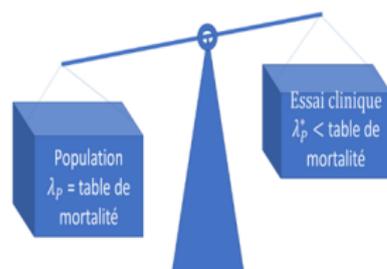


Essai clinique : mortalité due aux autres causes $<$ table de mortalité

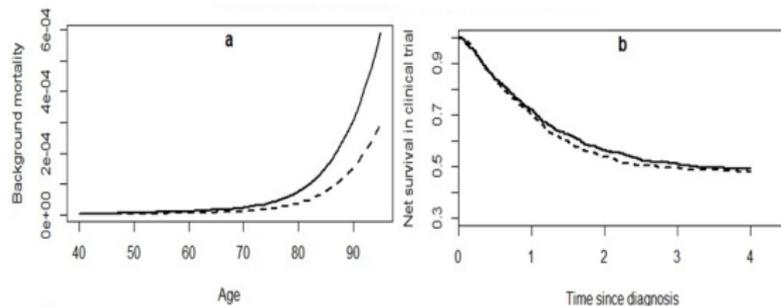
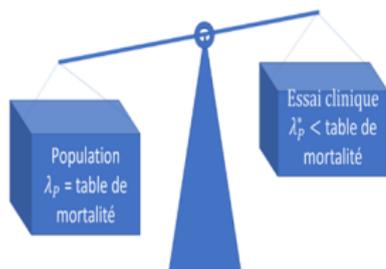


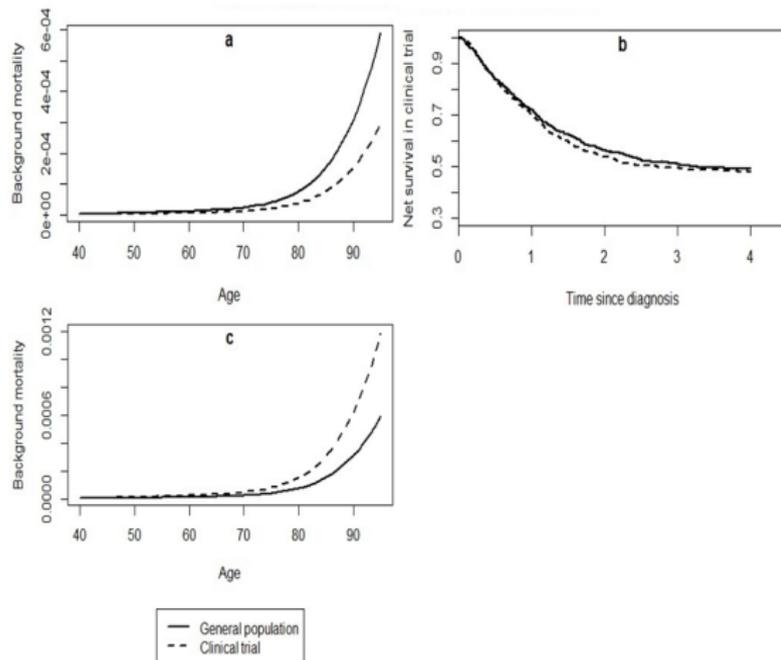
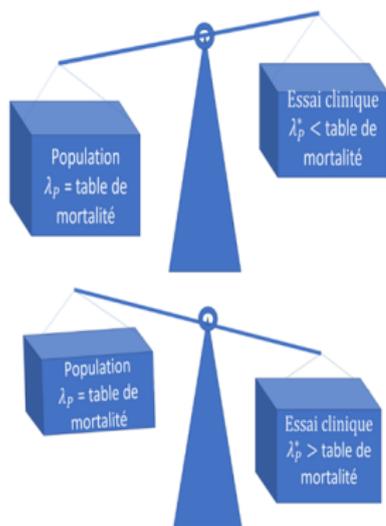
Registre de cancer : mortalité due aux autres causes \simeq table de mortalité

Cas 1 : λ_p^* dans l'essai $< \lambda_p$ (mélanome)

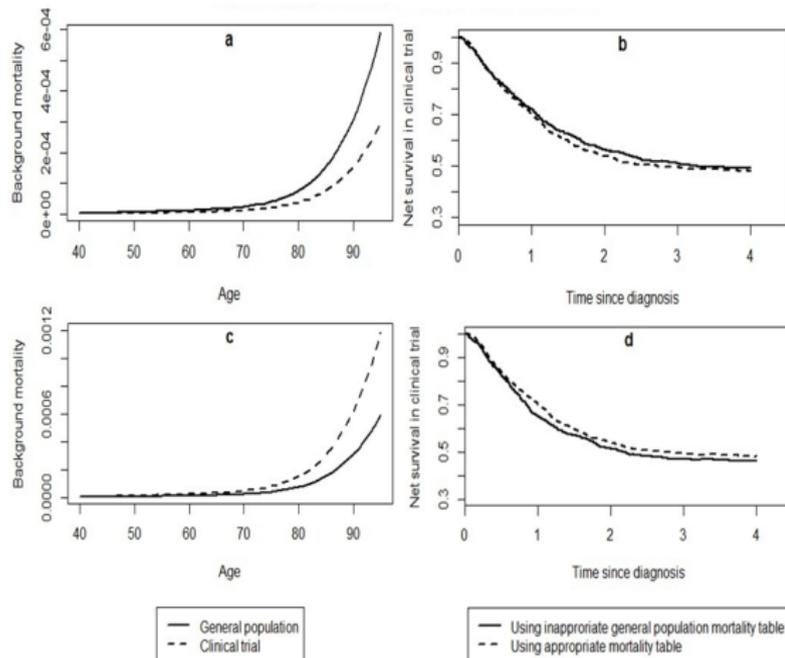
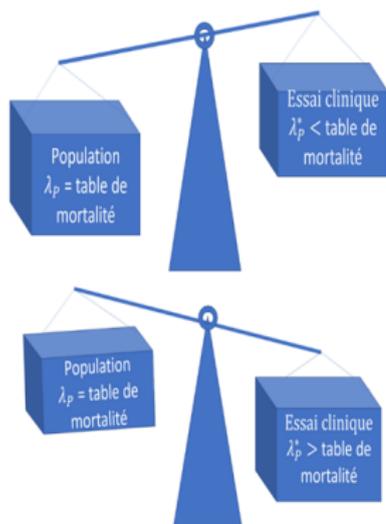


Cas 1 : λ_p^* dans l'essai $< \lambda_p$ (mélanome)



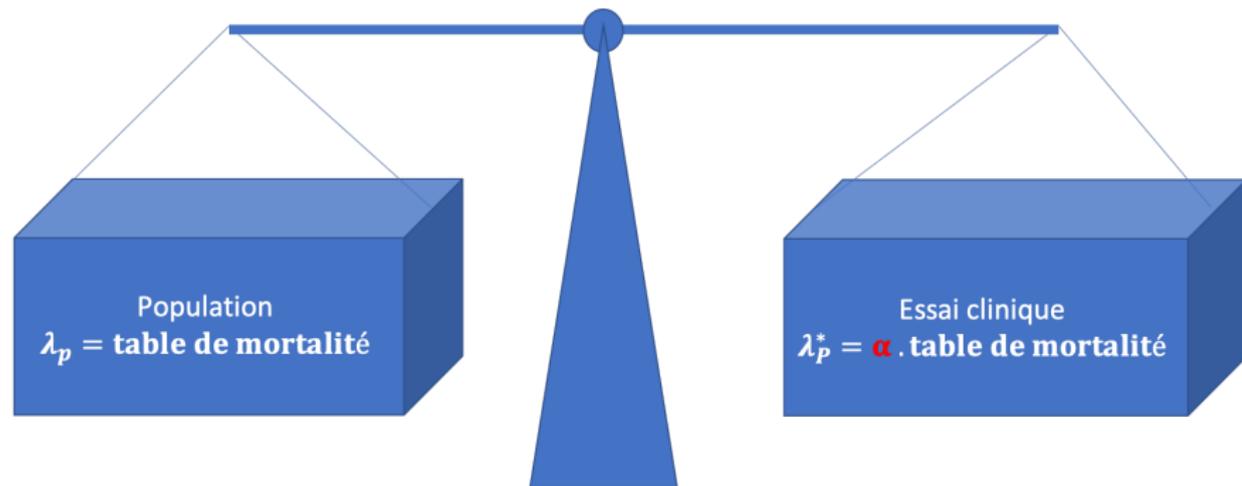
Cas 2 : λ_p^* dans l'essai $>$ λ_p (cancer du poumon)

Cas 2 : λ_p^* dans l'essai $>$ λ_p (cancer du poumon)



Correction de la mortalité attendue dans les essais cliniques

- Une solution à ce problème pourrait consister à :



$$\lambda_{Oi}(t) = \alpha \lambda_{Pi}(t) + \lambda_{Ei}(t)$$

(effet de sélection moyen)

Conclusion

- La survie nette \Rightarrow études de registre
- Critères de sélection : mortalité autres causes différente de celle de la population générale
- Estimation de la survie nette en recherche clinique \Rightarrow surestimation ou sous-estimation
- Solution
 - Construction d'un modèle de régression : correction de la mortalité attendue des patients

Partie 2

Correcting for misclassification and selection effects in estimating net survival in clinical trials.

Juste Aristide Goungounga¹, Célia Touraine^{1,2}, Nathalie Graffeo^{3,4}, Roch Giorgi⁵, and the CENSUR working survival group.

¹ Aix Marseille Univ, INSERM, IRD, SESSTIM Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Marseille, France.

² Unité de Biométrie, Institut du Cancer de Montpellier (ICM), Univ Montpellier, Montpellier, France.

³ INSERM U1153, Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité (CRESS), ECSTRA Team, Hôpital Saint Louis, Paris, France.

⁴ Université Paris Diderot, Paris, France.

⁵ Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM (Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale), Hop Timone, BioSTIC (Biostatistique et Technologies de l'Information et de la Communication), Marseille, France.

Corresponding author :

Prof Roch Giorgi
SESSTIM, Faculté de Médecine
27 Boulevard Jean Moulin
F-13385 Marseille Cedex 5
E-mail: roch.giorgi@ap-hm.fr



(Première relecture - non majeur)

Objectif général - Objectifs spécifiques

- Proposer un modèle de mortalité en excès : correction du biais de non comparabilité
- Estimer le biais lié à la mauvaise classification et intérêt des méthodes populationnelles
- Évaluer le modèle proposé par rapport à d'autres méthodes
 - En absence/présence de problème de classification
 - En absence/présence de biais de sélection

Méthodes comparés

- Méthode cause-spécifique
 - Estimateur de Kaplan-Meier (KM)
 - Modèle de Cox (Cox)
 - Estimateur de Nelson-Aalen pondéré (wNA)

Méthodes comparés

- Méthode cause-spécifique
 - Estimateur de Kaplan-Meier (KM)
 - Modèle de Cox (Cox)
 - Estimateur de Nelson-Aalen pondéré (wNA)
- Méthodes « populationnelles »
 - Estimateur de Pohar-Perme (PP)
 - Modèle « Rescaled B-spline (RBS) » proposé

Modèle proposé : « Rescaled B-spline »

- Modèle « Rescaled B-spline » :

$$\overbrace{\sum_{j=-2}^2 \nu_j \mathbb{B}_{j,3}(t)}^{\hat{\lambda}_E(t)} \exp[\beta X] = \lambda_O(t | X) - \alpha \lambda_P(t | Z)$$

Plan de simulation

- Simulation de 1000 jeux de données avec 2000 patients de même sexe
- Pour chaque patient, nous avons généré indépendamment :
 - Âge et traitement
 - $T_P \sim \alpha \lambda_P$ et $\alpha = (0.5, 1, 2, 4)$
 - $T_E \sim$ Weibull généralisée et $T_C \sim \mathcal{U}[0, b]$ avec $T_{X_C} = 50\%$
 - ☞ $T_O = \min(T_P, T_E, T_C)$;
 - Censure administrative : 15 ans

Plan de simulation

- Simulation de 1000 jeux de données avec 2000 patients de même sexe
- Pour chaque patient, nous avons généré indépendamment :
 - Âge et traitement
 - $T_P \sim \alpha \lambda_P$ et $\alpha = (0.5, 1, 2, 4)$
 - $T_E \sim$ Weibull généralisée et $T_C \sim \mathcal{U}[0, b]$ avec $T_{X_C} = 50\%$
 - $T_O = \min(T_P, T_E, T_C)$;
 - Censure administrative : 15 ans
 - $statut = 1$ si $T_O \leq T_{Cens.}$, 0 sinon
 - $cause = 1$ si $T_O = T_E$, 0 sinon (0%, 20%, 30% d'erreurs)
- Indicateurs statistiques de performance : biais, REQM, TRE

Absence de problème de classification et de biais de sélection en termes de taux attendus

Méthodes & Misc.	Survie nette (5 ans)			Survie nette (15 ans)		
	Biais x 100	REQM	TRE	Biais x 100	REQM	TRE
$\alpha=1$						
Gold standard	0.012	0.012	94.7	0.025	0.016	95.2
Cause-spécifique						
KM & 0%	0.126	0.012	94.5	0.778	0.018	92.8
Cox & 0%	0.036	0.011	95.7	0.046	0.016	97.7
wNA & 0%	0.027	0.012	94.4	0.143	0.016	94.3
Populationnelles						
PP	-0.359	0.014	94.5	-0.999	0.020	94.0
RBS	-0.102	0.015	94.2	-0.459	0.037	93.1

Biais liés à la mauvaise classification et intérêt des méthodes populationnelles

Méthodes & Misc.	Survie nette (5ans)			Survie nette (15ans)		
	Biais x 100	REQM	TRE	Biais x 100	REQM	TRE
<i>$\alpha=1$</i>						
Gold standard	0.012	0.012	94.7	0.025	0.016	95.2
Cause-spécifique						
KM & 30%	0.126	0.012	94.5	7.245	0.074	0.4
Cox & 30%	0.036	0.011	95.7	6.366	0.065	4.7
wNA & 30%	0.027	0.012	94.4	6.726	0.069	1.8
Populationnelles						
PP	-0.359	0.014	94.5	-0.999	0.020	94.0
RBS	-0.102	0.015	94.2	-0.459	0.037	93.1

Performances des méthodes populationnelles en fonction de l'importance du biais de sélection

Méthodes	Survie nette (5 ans)			Survie nette (15 ans)		
	Biais x 100	RMSE	TRE	Biais x 100	RMSE	TRE
$\alpha=1$						
PP	-0.359	0.014	94.5	-0.999	0.020	94.0
RBS	-0.102	0.015	94.2	-0.459	0.037	93.1
$\alpha=0.5$						
PP	0.897	0.015	88.0	2.618	0.031	67.4
RBS	-0.023	0.012	94.9	-0.251	0.028	90.1
$\alpha=2$						
PP	-2.784	0.031	50.7	-6.831	0.070	8.2
RBS	-0.284	0.021	93.4	-0.668	0.053	94.0
$\alpha=4$						
PP	-7.274	0.074	0.1	-15.629	0.157	0.00
RBS	-0.469	0.031	93.1	-0.934	0.076	93.8

Présence de biais de sélection et de mauvaise classification

- Mauvaises performances : KM, WNA, PP

Méthodes & Misc.	Survie nette (5 ans)			Survie nette (15 ans)		
	Biais x 100	REQM	TRE	Biais x 100	REQM	TRE
$\alpha=1$						
Cox & 30%	0.036	0.011	95.7	6.366	0.065	4.7
RBS	-0.102	0.015	94.2	-0.459	0.037	93.1
$\alpha=0.5$						
Cox & 30%	0.033	0.010	97.2	6.404	0.065	7.0
RBS	-0.023	0.012	94.9	-0.251	0.028	90.1
$\alpha=2$						
Cox & 30%	0.036	0.011	97.3	6.386	0.065	7.3
RBS	-0.284	0.021	93.4	-0.668	0.053	94.0
$\alpha=4$						
Cox & 30%	0.037	0.011	97.3	6.329	0.065	12.6
RBS	-0.469	0.031	93.1	-0.934	0.076	93.8

Conclusion et perspectives

- Première évaluation par simulation sur design unicentrique
- La survie nette spécifique : moins biaisée en théorie (données hypothétiques)
- Mauvaise classification de la cause de décès : recours aux approches utilisant les tables de mortalité
 - Le modèle de « Rescaled B-spline (RBS) » : alternative à recommander dans les essais à long terme

Conclusion et perspectives

- Première évaluation par simulation sur design unicentrique
- La survie nette spécifique : moins biaisée en théorie (données hypothétiques)
- Mauvaise classification de la cause de décès : recours aux approches utilisant les tables de mortalité
 - Le modèle de « Rescaled B-spline (RBS) » : alternative à recommander dans les essais à long terme
- Perspectives
 - Extension aux essais cliniques multicentriques
 - Investigations de l'impact de l'hétérogénéité inter-centres

Partie 3

Correcting for heterogeneity and non-comparability bias with a rescaled random-effect excess hazard model in multicenter clinical trials.

J. A. Goungounga¹, N. Grafféo^{1,2}, H. Charvat³, R. Giorgi⁴

Abstract

Net survival is a key indicator in cancer population-based studies to assess the burden of the studied cancer and to compare it between countries or periods. It is defined as the survival that would be observed, in a hypothetical world, where the disease of interest would be the only possible cause of death. In long-term clinical trial setting, where the cause of death is known to be unreliable, we proposed a rescaled relative survival model to correct for the selection bias occurring as individuals selected in a randomized clinical trial are not representative of the general population.

However, heterogeneity among patients from several medical centers exists in multicenter clinical trials. Their net survival depending on the center, the assumption of independence of net survival times does not hold. Thus, we propose to introduce a random effect in the rescaled excess hazard model accounting for heterogeneity in multicenter clinical trials and correcting selection bias. We studied the performance of our model in a simulation study and illustrate its use on data from a multicenter clinical trial. Thus, we recommend taking these two sources of bias into account when using the relative survival methods in clinical research.



(en préparation)

Hétérogénéité inter-centres

- Essais cliniques multicentriques : critères de sélection similaires
- Caractéristiques des centres : différences d'un centre à l'autre
→ Impact sur la mortalité due au cancer dans chaque centre
 - Accessibilité aux soins (privé/public)
 - Niveau de la prise en charge (grand centre hospitalier de référence/centre hospitalier standard)
- Prise en compte de ces différences : hétérogénéité inter-centres en survie globale⁹ et en survie nette¹⁰
- Problématique de la prise en compte de l'hétérogénéité inter-centres en même temps que le biais de sélection impactant sur les taux attendus

9. GLIDDEN et VITTINGHOFF 2004.

10. DUPONT et al. 2013 ; CHARVAT et al. 2016.

Objectif général - Objectifs spécifiques

- Proposer un modèle qui prend en compte la présence d'hétérogénéité en même temps que le biais de sélection

Objectif général - Objectifs spécifiques

- Proposer un modèle qui prend en compte la présence d'hétérogénéité en même temps que le biais de sélection
- Évaluer les performances de ce modèle suivant différents critères
 - ➔ Nombre de centres
 - ➔ La force de l'hétérogénéité inter-centres
 - ➔ Importance de l'effet de sélection

Objectif général - Objectifs spécifiques

- Proposer un modèle qui prend en compte la présence d'hétérogénéité en même temps que le biais de sélection
- Évaluer les performances de ce modèle suivant différents critères
 - ➔ Nombre de centres
 - ➔ La force de l'hétérogénéité inter-centres
 - ➔ Importance de l'effet de sélection
- Comparer ce modèle à ceux ne prenant pas en compte ces deux types de biais.

Modélisation de l'hétérogénéité et du biais de sélection

$$\overbrace{\lambda_{E,0}(t) \exp[\beta_1 X_1 + f(X_2) + g(t) \cdot X_3 + u_c]}^{\hat{\lambda}_E(t|X)}$$
$$= \lambda_O(t | X) - \lambda_P(t | Z), \text{ avec } u_c \sim \mathbb{N}(0, \sigma^2) \quad (\text{CHARVAT et al. 2016})$$

Modélisation de l'hétérogénéité et du biais de sélection

$$\overbrace{\lambda_{E,0}(t) \exp[\beta_1 X_1 + f(X_2) + g(t) \cdot X_3 + u_c]}^{\hat{\lambda}_E(t|X)}$$

$$= \lambda_O(t | X) - \lambda_P(t | Z), \text{ avec } u_c \sim \mathbb{N}(0, \sigma^2) \quad (\text{CHARVAT et al. 2016})$$

$$\overbrace{\lambda_{E,0}(t) \exp[\beta X + u_c]}^{\hat{\lambda}_E(t|X)}$$

$$= \lambda_O(t | X) - \alpha \lambda_P(t | Z), \text{ avec } u_c \sim \mathbb{N}(0, \sigma^2) \quad (\text{Modèle proposé})$$

Plan de simulation

- Simulation de 1000 jeux de données avec 2000 patients
- Pour chaque patient, nous avons généré indépendamment :
 - ☞ Âge : $\beta_{age} = 0.05$
 - ☞ Traitement : $P(\text{trt}) = 50\%$ $\beta_{trt} = -0.5$
 - ☞ Effet centre : $u_c \sim \mathcal{N}(0, \sigma^2)$ avec $\sigma = (0.25, 0.5, 1)$
 - ☞ Nombre de centres (équilibré) : (10, 20, 50, 100)
 - ☞ $T_P \sim \alpha \cdot \lambda_P$ et $\alpha = (0.5, 1, 2, 4)$
 - ☞ $T_E \sim \exp(u_c)$. Weibull généralisée
 - ☞ $T_C \sim \mathcal{U}[0, b]$ et $T_{X_C} = 50\%$
 - ☞ $T_O = \min(T_P, T_E, T_C)$;

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp(\beta X) + \lambda_{P,i}(t | Z), \quad (\text{M I})$$

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp(\beta X_i) + \alpha \lambda_{P,i}(t | Z), \quad (\text{M II})$$

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp[(\beta X_i) + u_c] + \lambda_{P,i}(t | Z), \quad (\text{M III})$$

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp[(\beta X) + u_c] + \alpha \lambda_{P,i}(t | Z), \quad (\text{M IV})$$

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp(\beta X) + \lambda_{P,i}(t | Z), \quad (\text{M I})$$

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp(\beta X_i) + \alpha \lambda_{P,i}(t | Z), \quad (\text{M II})$$

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp[(\beta X_i) + u_c] + \lambda_{P,i}(t | Z), \quad (\text{M III})$$

$$\lambda_{O,i}(t | X) = \lambda_{E,0}(t) \exp[(\beta X) + u_c] + \alpha \lambda_{P,i}(t | Z), \quad (\text{M IV})$$

- Critères de performance : Biais, REQm, TRE

Modèles	M I			M II		
	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$
$\sigma_{théorique}$						
Nclust=10	Mauvais	Mauvais	Bon	Mauvais	Mauvais	Bon
Nclust=20	Mauvais	Insatisfaisant	Insatisfaisant	Insatisfaisant	Insatisfaisant	Insatisfaisant
Nclust=50	Mauvais	Insatisfaisant	Bon	Insatisfaisant	Bon	Bon
Nclust=100	Mauvais	Insatisfaisant	Bon	Insatisfaisant	Mauvais	Bon
$\theta/\lambda_E/S_E$	Insatisfaisant			Insatisfaisant		

Modèles	M III			M IV		
	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$
$\sigma_{théorique}$						
Nclust=10	Bon	Bon	Bon	Bon	Bon	Bon
Nclust=20	Bon	Bon	Bon	Bon	Bon	Bon
Nclust=50	Bon	Bon	Bon	Bon	Bon	Bon
Nclust=100	Bon	Bon	Bon	Bon	Bon	Bon
$\theta/\lambda_E/S_E$	Satisfaisant			Satisfaisant		

Modèles $\sigma_{\text{théorique}}$	M I			M II		
	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$
Nclust=10	Mauvais	Insatisfaisant	Insatisfaisant	Insatisfaisant	Mauvais	Insatisfaisant
Nclust=20	Mauvais	Insatisfaisant	Bon	Insatisfaisant	Mauvais	Bon
Nclust=50	Bon	Bon	Bon	Bon	Bon	Bon
Nclust=100	Bon	Bon	Bon	Bon	Bon	Bon
$\theta/\lambda_E/S_E$	Insatisfaisant			Insatisfaisant		

Modèles $\sigma_{\text{théorique}}$	M III			M IV		
	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$
Nclust=10	Insatisfaisant	Bon	Bon	Bon	Bon	Bon
Nclust=20	Bon	Bon	Bon	Bon	Bon	Bon
Nclust=50	Bon	Bon	Bon	Bon	Bon	Bon
Nclust=100	Bon	Bon	Bon	Bon	Bon	Bon
$\theta/\lambda_E/S_E$	Insatisfaisant			Satisfaisant		

Modèles $\sigma_{théorique}$	M I			M II		
	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$
Nclust=10	Mauvais	Mauvais	Insatisfaisant	Insatisfaisant	Insatisfaisant	Bon
Nclust=20	Mauvais	Mauvais	Insatisfaisant	Insatisfaisant	Insatisfaisant	Insatisfaisant
Nclust=50	Mauvais	Mauvais	Insatisfaisant	Insatisfaisant	Insatisfaisant	Insatisfaisant
Nclust=100	Mauvais	Mauvais	Insatisfaisant	Insatisfaisant	Insatisfaisant	Insatisfaisant
$\theta/\lambda_E/S_E$	Insatisfaisant			Insatisfaisant		

Modèles $\sigma_{théorique}$	M III			M IV		
	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.25$
Nclust=10	Mauvais	Insatisfaisant	Insatisfaisant	Bon	Bon	Bon
Nclust=20	Insatisfaisant	Insatisfaisant	Insatisfaisant	Bon	Bon	Bon
Nclust=50	Insatisfaisant	Insatisfaisant	Insatisfaisant	Bon	Bon	Bon
Nclust=100	Insatisfaisant	Insatisfaisant	Insatisfaisant	Bon	Bon	Bon
$\theta/\lambda_E/S_E$	Insatisfaisant			Satisfaisant		

Conclusion et perspectives

- Conclusion
 - Modèle proposé complète le modèle RBS : essais multicentriques
 - Bonnes performances : taille des centres, force de sélection, taille d'hétérogénéité

Conclusion et perspectives

- Conclusion
 - Modèle proposé complète le modèle RBS : essais multicentriques
 - Bonnes performances : taille des centres, force de sélection, taille d'hétérogénéité
- Perspectives
 - Fragilité multivariée : individus différents par centre (plausibilité clinique)
 - Package \mathbb{R} intégrant ces nouveaux développement : intérêt pour l'utilisation en recherche clinique

Références

- [1] A AUGUSTIN et al. « Lymphome du manteau dans deux essais cliniques du groupe LYSA : comparaison avec les données de population ». In : *Revue d'Épidémiologie et de Santé Publique* 64 (2016), S20.
- [2] Hadrien CHARVAT et al. « A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates ». In : *Statistics in medicine* 35.18 (2016), p. 3066–3084.
- [3] David R COX. « Partial likelihood ». In : *Biometrika* 62.2 (1975), p. 269–276.
- [4] Cyrielle DUPONT et al. « Description of an approach based on maximum likelihood to adjust an excess hazard model with a random effect ». In : *Cancer epidemiology* 37.4 (2013), p. 449–456.
- [5] F EDERER et H HEISE. « The effect of eliminating deaths from cancer on general population survival rates, methodological note 11 : End results evaluation section ». In : *National Cancer Institute* (1959).
- [6] Fred EDERER. « The relative survival rate : a statistical methodology ». In : *NCI Monograph* 6 (1961), p. 101–121.
- [7] J ESTEVE et al. « Relative survival and the estimation of net survival : elements for further discussion ». In : *Statistics in medicine* 9.5 (1990), p. 529–538.
- [8] Roch GIORGI et al. « A relative survival regression model using B-spline functions to model non-proportional hazards ». In : *Statistics in medicine* 22.17 (2003), p. 2767–2784.
- [9] David V GLIDDEN et Eric VITTINGHOFF. « Modelling clustered survival data from multicentre clinical trials ». In : *Statistics in medicine* 23.3 (2004), p. 369–388.
- [10] Els GOETGHEBEUR et Louise RYAN. « A modified log rank test for competing risks with missing failure type ». In : *Biometrika* 77.1 (1990), p. 207–211.
- [11] Juste Aristide GOUNGOUNGA et Roch GIORGI. « Commentary on : Survival benefit of mantle cell lymphoma patients enrolled in clinical trials ; a joint study from the LYSA group and French cancer registries ». In : *Journal of cancer research and clinical oncology* 144.4 (2018), p. 791–793.
- [12] Timo HAKULINEN. « Cancer survival corrected for heterogeneity in patient withdrawal ». In : *Biometrics* (1982), p. 933–942.
- [13] Edward L KAPLAN et Paul MEIER. « Nonparametric estimation from incomplete observations ». In : *Journal of the American statistical association* 53.282 (1958), p. 457–481.
- [14] Qing LIU et Donald A PIERCE. « A note on Gauss—Hermite quadrature ». In : *Biometrika* 81.3 (1994), p. 624–629.
- [15] Adeline MORISOT. « Méthodes d'analyse de survie, valeurs manquantes et fractions attribuables temps dépendantes : application aux décès par cancer de la prostate ». Thèse de doct. Université Montpellier, 2015.
- [16] Christopher P NELSON et al. « Flexible parametric models for relative survival, with application in coronary heart disease ». In : *Statistics in medicine* 26.30 (2007), p. 5486–5498.
- [17] Maja Pohar PERME, Janez STARE et Jacques ESTÈVE. « On estimation in relative survival ». In : *Biometrics* 68.1 (2012), p. 113–120.
- [18] L REMONTET et al. « An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies ». In : *Statistics in medicine* 26.10 (2007), p. 2214–2228.
- [19] Diana SARFATI, Tony BLAKELY et Neil PEARCE. « Measuring cancer survival in populations : relative survival vs cancer-specific survival ». In : *International journal of epidemiology* 39.2 (2010), p. 598–610.
- [20] Robin SCHAFFAR et al. « Cause-specific or relative survival setting to estimate population-based net survival from cancer? An empirical evaluation using women diagnosed with breast cancer in Geneva between 1981 and 1991 and followed for 20 years after diagnosis ». In : *Cancer epidemiology* 39.3 (2015), p. 465–472.
- [21] Robin SCHAFFAR et al. « Estimation of net survival for cancer patients : Relative survival setting more robust to some assumption violations than cause-specific setting, a sensitivity analysis on empirical data ». In : *European Journal of Cancer* 72 (2017), p. 78–83.
- [22] Bart VAN ROMPAYE, Els GOETGHEBEUR et Shabbar JAFFAR. « Design and testing for clinical trials faced with misclassified causes of death ». In : *Biostatistics* 11.3 (2010), p. 546–558.

Remerciements



Aux membres du
groupe CENSUR

Extension de l'analyse de la survie nette au domaine de la recherche clinique

UMR 1252 SESSTIM

Thèse présentée par Juste GOUNGOUNGA

Sous la direction du Pr Roch GIORGI

pour obtenir le grade de

Docteur d'Aix-Marseille Université

03 Décembre 2018



Faculté
de Médecine
Aix-Marseille Université



A*Midex
Initiative d'excellence
Aix-Marseille