

TESTS DU LOG-RANK AJUSTÉ : ÉTUDE DE SIMULATIONS

Florent Le Borgne ^{1,2} & Yann Foucher ¹

¹ *Université de Nantes - EA4275-SPHERE - ITUN
bioStatistics, Pharmacoepidemiology and Human sciEnces REsearch team
1 rue Gaston Veil BP 53508 44035 Nantes Cedex 1
florent.le-borgne@etu.univ-nantes.fr / yohann.foucher@univ-nantes.fr*

² *IDBC/A2com
Espace Antrium Parc de la Teillais 35740 PACE*

Résumé. Dans les études observationnelles, la présence de facteurs de confusion est fréquente et la comparaison de différents groupes de sujets nécessite alors un ajustement. En présence de données de survie, cet ajustement peut être réalisé à l'aide d'un modèle multivarié (modèle de Cox le plus souvent) permettant de valider une différence observée des courbes de survie brutes obtenues par l'estimateur de Kaplan-Meier. Cependant, à partir de ce type de régression, l'effet du facteur explicatif d'intérêt est souvent résumé par le rapport des risques instantanés (RR). Cette perte d'information est tellement dommageable que la plupart des travaux de recherche en biologie ou en médecine présentent à la fois les courbes de survie brutes (biaisées mais représentant précisément les différences de survie) et les RR ajustés (non biaisés mais trop synthétiques). Une solution récente est l'utilisation de courbes de survie ajustées par une méthode de pondération basée sur les scores de propension accompagnées d'un test du Log-rank ajusté (IPTW, Inverse Probability of Treatment Weighted). Cependant, trois variantes sont retrouvées dans la littérature sans aucune comparaison des performances de ces tests en termes de risques de première et seconde espèces. Nous proposons donc dans ce travail une étude de simulations afin (i) d'évaluer si les performances de ces tests du Log-rank ajusté sont acceptables par rapport à un modèle de Cox classiquement utilisé, et (ii) choisir le plus performant des trois tests.

Mots-clés. Données de survie, estimateur de Kaplan-Meier ajusté, Log-rank ajusté, étude de simulations.

Abstract. In observational studies, the presence of confounding factors is common and the comparison of different groups of subjects requires adjustment. In the presence of survival data, this adjustment can be achieved with a multivariate model (usually a Cox model) allowing to validate a difference observed on the crude survival curves obtained by the Kaplan-Meier estimator. However, from this type of regression, the effect of the interesting factor is often summarized by the hazard ratio (HR). This loss of information is so damaging that most research projects in biology or medicine present both crude survival curves (biased but illustrating precisely the differences in survival) and the adjusted HR

(not biased but too synthetic). A recent solution is the use of adjusted survival curves by a weighting method based on propensity scores with an adjusted log-rank test (IPTW, Inverse Probability of Treatment Weighted). However, three alternatives are found in the literature without any comparison of the performance of these tests in terms of first and second kind error types. We propose in this work a simulations study in order to (i) evaluate if the performances of these adjusted log-rank tests are acceptable compared to a Cox model used classically, and (ii) choose the most powerful of the three tests.

Keywords. Survival data, adjusted Kaplan-Meier estimator, adjusted log-rank, simulations study.

1 Introduction

Les courbes de survie basées sur l'estimateur de Kaplan-Meier sont utilisées dans de nombreuses études cliniques. Dans les études randomisées où les patients sont répartis aléatoirement dans les différents groupes, l'estimateur de Kaplan-Meier et le test du Log-rank sont généralement appropriés. A l'inverse, dans les études observationnelles, les caractéristiques des sujets peuvent être déséquilibrées entre les groupes, ces facteurs de confusion rendant l'estimateur de Kaplan-Meier inadéquat. L'estimation de la causalité entre l'exposition et l'événement étudié requiert alors un ajustement sur les facteurs de confusion. Cet ajustement peut être réalisé à l'aide d'un modèle multivarié, cependant ce type de régression résume souvent l'effet du facteur explicatif d'intérêt en un rapport des risques instantanés (RR) entraînant une perte d'information très dommageable. Une solution est l'utilisation de courbes de survie ajustées [1,2] utilisant la méthode IPTW (Inverse Probability of Treatment Weighted) basée sur les scores de propension [3].

Le test du Log-rank est le test standard de comparaison de deux courbes de survie. Trois versions de ce test ont été adaptées à l'estimateur de Kaplan-Meier ajusté [2,4,5]. L'objectif de ce travail est double. Premièrement, nous souhaitons évaluer les performances des tests du Log-rank ajusté en les comparant au modèle de Cox à risques proportionnels, régression la plus utilisée en recherche biomédicale en présence de données censurées. Deuxièmement, nous souhaitons les comparer entre eux pour choisir le plus performant des trois.

Pour répondre à ces deux objectifs, nous réalisons une étude de simulations. Dans un premier temps, nous définissons l'estimateur de Kaplan-Meier ajusté par la méthode IPTW. Ensuite, nous exposons les trois différents tests du Log-rank ajusté retrouvés dans la littérature ainsi que le plan de simulations. Dans une dernière partie nous présentons les résultats obtenus.

2 L'estimateur de Kaplan-Meier ajusté

Soient $(T_i, \delta_i, X_i, Z_i)$, les variables aléatoires associées à l'individu i d'un échantillon de taille N ($i = 1, \dots, n$). T_i est le délai d'apparition de l'événement d'intérêt, δ_i est l'indicateur de censure ($\delta_i = 0$ si T_i est censuré à droite et $\delta_i = 1$ si T_i correspond à un évènement), X_i la variable explicative représentant le facteur d'exposition d'intérêt composé de K groupes ($k = 1, \dots, K$), et Z_i est le vecteur de covariables d'ajustement (possibles facteurs de confusion). Soit $p_{ik} = P(X_i = k | Z_i)$, la probabilité que l'individu i a d'être dans le groupe k sachant le vecteur de covariables Z_i . Soit $S_k(t)$ la fonction de survie du groupe k au temps t .

Supposons D_k le nombre de temps différents pour lesquels des évènements sont observés dans le groupe k . Au temps t_j ($j = 1, \dots, D_k$), le nombre d'évènements d_{jk} et le nombre d'individus à risque Y_{jk} dans le groupe k peuvent s'écrire respectivement $d_{jk} = \sum_{i:t_i=t_j} \delta_i I(X_i = k)$ et $Y_{jk} = \sum_{i:t_i \geq t_j} I(X_i = k)$. La méthode IPTW permet de réduire les biais de confusion entre les groupes en corrigeant la contribution de chaque individu i par un poids $w_{ik} = 1/p_{ik}$. Alors, le nombre pondéré d'évènements d_{jk}^w et le nombre pondéré d'individus à risque Y_{jk}^w au temps t_j sont définis comme :

$$d_{jk}^w = \sum_{i:t_i=t_j} w_{ik} \delta_i I(X_i = k)$$

$$Y_{jk}^w = \sum_{i:t_i \geq t_j} w_{ik} I(X_i = k)$$

Le nombre d'évènements et le nombre de sujets à risques dans l'ensemble de l'échantillon peut être noté $d_j^w = \sum_{k=1}^K d_{jk}^w$ et $Y_j^w = \sum_{k=1}^K Y_{jk}^w$. L'estimateur de Kaplan-Meier ajusté obtenu est le suivant :

$$\hat{S}_k(t) = \prod_{t_j \leq t} [1 - d_{jk}^w / Y_{jk}^w]$$

3 Test du Log-rank ajusté

Dans ce travail, considérons seulement deux groupes, notés $X = 0$ et $X = 1$.

Xu et al. proposent un test du Log-rank pour comparer deux courbes de survie ajustées comme définies précédemment [5]. Il s'agit d'un test équivalent au test du Log-rank standard en remplaçant (i) les nombres d'évènements observés par les nombres pondérés d'évènements et (ii) les nombres d'individus à risque observés par les nombres pondérés d'individus à risque. La statistique de test obtenue est $Z_0 = G^w / \sqrt{\text{Var}(G^w)}$. Elle suit asymptotiquement, sous H_0 (hypothèse de l'égalité des fonctions de survie), une loi nor-

male centrée réduite avec :

$$G^w = \sum_{j=1}^D d_{j1}^w - Y_{j1}^w \left(\frac{d_j^w}{Y_j^w} \right)$$

$$\text{et } \text{Var}(G^w) = \sum_{j=1}^D \left(\frac{Y_{j0}^w Y_{j1}^w d_j^w (n_j^w - d_j^w)}{(n_j^w)^2 (n_j^w - 1)} \right)$$

Un test alternatif est proposé par Xie et Liu [2]. Les auteurs précisent qu'il faut commencer par ajuster les poids de chaque individu. Ainsi, au temps t_j ($j = 1, \dots, D_k$), le poids pour un individu i dans le groupe k est réaffecté tel que $w'_{ijk} = w_{ik} \cdot Y_{jk} / Y_{jk}^w$. Le nombre d'événements pondéré devient $d_{jk}^{w'} = \sum_{i:t_i=t_j} w'_{ijk} \delta_i I(X_i = k)$ et le nombre d'individus à risque pondéré devient $Y_{jk}^{w'} = \sum_{i:t_i \geq t_j} w'_{ijk} I(X_i = k)$. Le test est basé sur la statistique :

$$G^w = \sum_{j=1}^D d_{j1}^{w'} - Y_{j1}^{w'} \left(\frac{d_j^{w'}}{Y_j^{w'}} \right)$$

de variance égale à :

$$\text{Var}(G^w) = \sum_{j=1}^D \left\{ \frac{d_j(Y_j - d_j)}{Y_j(Y_j - 1)} \sum_{i=1}^{Y_j} \left[\left(\frac{Y_{j0}^{w'}}{Y_j^{w'}} \right)^2 w_i'^2 X_i + \left(\frac{Y_{j1}^{w'}}{Y_j^{w'}} \right)^2 w_i'^2 (1 - X_i) \right] \right\}$$

La statistique de test du Log-rank ajusté proposée est $Z_0 = G^w / \sqrt{\text{Var}(G^w)}$ équivalente à la précédente.

Une dernière variante du test du Log-rank adapté à l'estimateur de Kaplan-Meier ajusté a été trouvée dans la littérature. Proposée par Sugihara [4], elle reprend les formules utilisées par Xie et Liu [2] à l'exception près de la ré-assignation des poids qui n'est pas réalisée.

4 Etude de simulations

Dans une première étape nous cherchons à nous placer dans un cas idéal pour le Cox multivarié et nous regardons si les résultats obtenus par les tests du Log-rank ajusté s'en rapprochent. Nous simulons donc des variables respectant l'hypothèse de proportionnalité des risques. Dans un second temps nous nous plaçons dans une situation où les conditions pour la réalisation du modèle de Cox ne s'appliquent plus en simulant un facteur de confusion ne respectant pas l'hypothèse de proportionnalité des risques.

4.1 Simulations des temps de survie à partir d'une fonction à risques proportionnels

Le modèle de Cox exprime une relation entre la fonction de risque associée à la survenue d'un événement et le vecteur des p variables explicatives $z = (z_1, \dots, z_p)$ tel que : $\lambda(t, y) = \lambda_0(t) \exp(\beta z)$ où $\beta = (\beta_1, \dots, \beta_p)$ est le vecteur des coefficients de régression et $\lambda_0(t)$ est la fonction de risque de base. Sa fonction de répartition est égale à :

$$F(t) = 1 - \exp(-\Lambda_0(t) * \exp(\beta z))$$

Où : $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. La fonction de répartition $F(t)$ est continue et strictement croissante sur \mathbb{R} et a pour limite 0 en $-\infty$ et 1 en $+\infty$. Dans ce cas particulier, $F(t)$ réalise une bijection de \mathbb{R} sur $]0, 1[$ et admet donc une fonction inverse $F^{-1}(t) :]0, 1[\rightarrow \mathbb{R}$. Si U est une variable aléatoire de loi uniforme sur $]0, 1[$, alors $F^{-1}(U)$ a la même que T . Pour un modèle à risques proportionnels avec une fonction de risque de base suivant une distribution de Weibull, les temps d'événements sont obtenus par :

$$t = \left(-\frac{1}{\lambda} \log(u) \exp(-\beta z) \right)^{1/\nu}$$

4.2 Simulations des temps pour un modèle à risques non proportionnels

Dans un deuxième temps, nous simulons des données de sorte que la variable explicative z_p ne respecte pas l'hypothèse de proportionnalité des risques. Pour cela on ajoute le terme $\beta_{p+1} \log(t) z_p$ à la fonction de risque du modèle tel que :

$$\lambda(t) = \lambda_0(t) \exp(\beta z + \beta_{p+1} \log(t) z_p)$$

Avec une fonction de risque de base, distribuée selon une loi de Weibull sous la contrainte $\beta_{p+1} z_p + \nu > 0$, la fonction de répartition $F(t)$ peut alors s'écrire :

$$F(t) = 1 - \exp \left(-\lambda \nu \exp(\beta z) \frac{t^{\beta_{p+1} z_p + \nu}}{\beta_{p+1} z_p + \nu} \right)$$

Les temps d'événements sont alors obtenus à partir des observations u d'une loi uniforme $]0, 1[$:

$$t = \left[-\frac{\ln(u)(\beta_{p+1} z_p + \nu) \exp(-\beta z)}{\lambda \times \nu} \right]^{1/(\beta_{p+1} z_p + \nu)}$$

4.3 Plan des simulations

Pour que les données simulées se rapprochent de données réelles, elles sont simulées selon des paramètres estimés à partir de données extraites de notre cohorte DIVAT (Données Informatisées et VALidées en Transplantation) de patients transplantés rénaux. Afin de ne pas trop complexifier le modèle, nous nous limitons à l'étude d'une variable d'intérêt et de 4 facteurs de confusion. Nous comparons les performances des différents modèles pour différentes tailles d'échantillons $N = 100, 250, 500, 1000$ et 1837 (taille de l'échantillon extrait de DIVAT), pour différents taux de censure égaux à 0.30 et 0.68 (taux de censure observé dans DIVAT) et un coefficient de régression associé à la variable d'intérêt égal à $0, 0,25, 0,365$ (valeur observée dans l'échantillon extrait de DIVAT) et $0,5$. Lorsque $\beta = 0$, nous calculons le risque de première espèce alors que la puissance des modèles est évaluée lorsque $\beta \neq 0$.

5 Résultats

Lorsque les variables sont simulées comme respectant l'hypothèse de proportionnalité des risques c'est à dire la situation où le modèle de Cox est optimal. Nous observons un risque de première espèce très proche de 5% pour le modèle de Cox ainsi que pour les tests du Log-rank proposés par Xie et Liu et par Sugihara. Une faible augmentation de ce risque lorsque les effectifs sont faibles ($N = 100$) respectivement $6,5\%, 7,7\%$ et $8,6\%$. Le test du Log-rank proposé par Xu et al. est en revanche associé à de très mauvais résultats en terme de risque de première espèce ($>40\%$) quelque soit la taille de l'échantillon.

En ce qui concerne le risque de seconde espèce, il diminue logiquement avec le taux de censure, lorsque la taille de l'échantillon augmente, lorsque la valeur du coefficient de régression augmente quelque soit la méthode. Les mêmes risques de seconde espèce sont obtenus avec les tests proposés par Xie et Liu et par Sugihara, également équivalents au risque de seconde espèce obtenu par un modèle de Cox lorsque les effectifs sont faibles (≤ 250). En revanche, les risques de seconde espèce sont plus importants que ceux obtenus avec un modèle de Cox lorsque les effectifs sont grands. Le test proposé par Xu et al. obtient les meilleurs résultats en terme de risque de seconde espèce.

Lorsque l'une des variables d'ajustement est simulée comme ne respectant l'hypothèse de proportionnalité des risques, situation où le modèle de Cox n'est plus adapté, ce dernier reste le modèle le plus performant.

En conclusion, le modèle de Cox montre de très bonnes performances lorsqu'il est dans sa situation optimale ce qui est conforme à nos attentes. Il montre également de bons résultats lorsque ses conditions d'applications ne sont pas respectées. Parmi les trois versions du test du Log-rank ajusté, celle proposée par Xu et al. ne semble pas adaptée. Parmi les deux autres, il semble que celle proposée par Xie et Liu soit la plus performante

pour son risque de première espèce légèrement meilleur. Comparativement au modèle de Cox, cette version du Log-rank ajusté montre un risque de première et de seconde espèce un peu plus grand. Cependant, elle reste néanmoins tout à fait correcte voir équivalente lorsque les effectifs sont grands. Une courbe de survie ajustée et un test du Log-rank ajusté, faciles à calculer et ne résumant pas les différences de survie entre deux groupes par un indicateur trop synthétique, semblent donc une bonne première approche descriptive préalable à la réalisation d'un modèle multivarié.

Bibliographie

- [1] Cole SR., Hernán MA. (2004). Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*, 75, 45-49.
- [2] Xie, J. and Lui, C. (2005). Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine*, 24, 3089-3110.
- [3] Rosenbaum PR., Rubin DB. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- [4] Sugihara M. (2010). Survival analysis using inverse probability of treatment weighted methods based on the generalized propensity score. *Pharmaceutical Statistics*, 9, 21-34.
- [5] Xu S., Shetterly S., Powers D., Raebel A., Tsai T., Ho M., Magid D. (2012). Extension of Kaplan-Meier methods in observational studies with time-varying treatment. *Value Health*, 15(1), 167-174.