

# DEUX STRATÉGIES ALTERNATIVES POUR COMPARER L'EFFET DES FACTEURS DE RISQUE ENTRE DEUX GROUPES: ILLUSTRATION PAR L'ANALYSE DES PREMIÈRES ET DES SECONDES TRANSPLANTATIONS RÉNALES

Katy Trébern-Launay <sup>1</sup> & Yann Foucher <sup>2</sup>

<sup>1</sup> *Institut de Transplantation, Urologie Néphrologie (ITUN), CHU de Nantes, Inserm U1064, 30 bd Jean Monnet, 44093 Nantes, France, katy.launay@univ-nantes.fr*

<sup>2</sup> *EA 4275 - SPHERE Biostatistique, Recherche Clinique et Mesures Subjectives en Santé, 1 rue Gaston Veil, 44035 Nantes, France, yohann.foucher@univ-nantes.fr*

**Résumé.** Le pronostic des receveurs d'une seconde transplantation rénale (STR) a fréquemment été comparé à celui des receveurs d'une première transplantation (PTR). Cependant, peu d'études se sont intéressées à comparer les différences d'effet des facteurs de risque sur l'échec de greffe entre les deux groupes. Pour étudier cette hétérogénéité des facteurs, nous proposons deux stratégies alternatives : (i) un modèle de survie relative à risques multiplicatifs (MSR) et (ii) un modèle de Cox stratifié (MCS) sur le rang de greffe et faisant l'hypothèse de sous-vecteurs de variables explicatives. Le critère de jugement analysé dans cette étude est l'échec de greffe (retour en dialyse ou décès du patient). L'estimation des paramètres est basée sur la maximisation de la vraisemblance partielle. Nous avons développé une procédure associant des simulations de Monte-Carlo à un ré-échantillonnage par bootstrap afin de prendre en compte la variabilité des paramètres attendus dans l'estimation de la partie relative du modèle MSR. Nous avons montré pour la première fois en transplantation rénale que : (i) le sexe du donneur était un facteur de risque spécifique des STR, (ii) l'effet lié à l'âge du receveur était potentialisé chez les STR et (iii) l'effet lié à l'âge du donneur était atténué chez les STR. Par rapport aux modèles de survie classiquement utilisés, les deux modèles proposés apportent des informations nouvelles et cliniquement pertinentes. Ces approches pourraient être développées dans d'autres champs d'application clinique pour comparer l'effet des facteurs entre deux groupes d'études. Nous avons proposé un package R pour implémenter l'approche MSR.

**Mots-clés.** Survie Relative, Modèle de Cox Stratifié, Seconde Transplantation Rénale

**Abstract.** Whereas the prognosis of second kidney transplant recipients (STR) compared to the first ones has been frequently analyzed, no study has addressed the issue of comparing the risk factor effects on graft failure between both groups. Here, we propose two alternative strategies to study the heterogeneity of risk factors between two groups of patients: (i) a multiplicative-regression model for relative survival (MRS) and (ii) a stratified Cox model (SCM) specifying the graft rank as strata and assuming subvectors

of the explicatives variables. These developments were motivated by the analysis of factors associated with time to graft failure (return-to-dialysis or patient death) in second kidney transplant recipients compared to the first ones. Estimation of the parameters was based on partial likelihood maximization. Monte-Carlo simulations associated with bootstrap re-sampling was performed to calculate the standard deviations for the MRS. We demonstrate, for the first time in renal transplantation, that: *(i)* male donor gender is a specific risk factor for STR, *(ii)* the adverse effect of recipient age is enhanced for STR and *(iii)* the graft failure risk related to donor age is attenuated for STR. While the traditional Cox model did not provide original results based on the renal transplantation literature, the proposed relative and stratified models revealed new findings that are useful for clinicians. These methodologies may be of interest in other medical fields when the principal objective is the comparison of risk factors between two populations.

**Keywords.** Relative Survival, Stratified Cox Model, Second Kidney Transplant

## 1 Introduction

Trébern-Launay et al. (2012) ont récemment montré que les receveurs d'une seconde transplantation rénale (STR) avaient un risque d'échec de greffe (retour en dialyse ou décès du patient) plus important que les receveurs d'une première transplantation rénale (PTR) après quelques années post-transplantation. Cet excès de risque d'échec des STR est cependant négligeable au regard du bénéfice de la retransplantation en termes de survie et de qualité de vie. En revanche, dans le contexte actuel de grave pénurie de greffons, la compréhension des différences d'effets des facteurs sur le devenir de la transplantation entre PTR et STR est essentielle afin d'identifier les patients susceptibles d'en tirer le meilleur bénéfice.

Les modèles de survie classiquement utilisés permettent difficilement d'étudier cette hétérogénéité des facteurs. Pour répondre à cet objectif, nous proposons deux stratégies appliquées à 2206 FTR et 566 STR, issus de la cohorte française multicentrique DIVAT (Données Informatisées et VALidées en Transplantation - <http://www.divat.fr>). La première stratégie est l'adaptation d'un modèle de survie relative à risques multiplicatifs (MSR), tel que celui proposé par Andersen et al. (1985) mais dont l'originalité est d'étudier l'échec de greffe des STR (groupe d'intérêt) en prenant comme référence un échantillon de PTR (groupe contrôle). En effet, les modèles de survie relative s'attachent toujours à l'étude de la mortalité observée par rapport à la mortalité en population générale issue de tables.

La seconde stratégie est l'adaptation originale d'un modèle de Cox stratifié (MCS) définissant le rang de greffe comme une strate et faisant l'hypothèse d'un vecteur de variables explicatives divisé en sous-vecteurs de variables entrant soit dans le risque attendu (PTR), soit dans le risque spécifique (STR), soit dans les deux groupes mais avec des

effets communs ou différents.

## 2 Modèle de survie relative à risques multiplicatifs (MSR)

Supposons les individus indexés par  $j$  ( $j = 1, \dots, n_a$ ) dans le groupe contrôle des PTR et par  $i$  ( $i = 1, \dots, n_r$ ) dans le groupe d'intérêt des STR.  $n_a$  and  $n_r$  représentent les tailles d'échantillons des groupes respectifs. Le risque d'échec instantané au temps  $t_i$  pour le  $i$ ème individu du groupe d'intérêt des STR est noté  $h^o(t_i|z_i)$ , avec  $z_i$  le vecteur des variables explicatives. Le modèle à risques multiplicatifs proposé par Andersen et al. (1985) suppose la décomposition du risque observé dans ce groupe d'intérêt en deux risques :

$$h^o(t_i|z_i) = h^a(t_i|z_i^a) h^r(t_i|z_i^r) \quad (1)$$

avec  $h^a(t_i|z_i^a)$  le risque instantané attendu pour un individu du groupe contrôle des PTR avec des caractéristiques similaires à celle du  $i$ ème individu et  $h^r(t_i|z_i^r)$  le risque instantané relatif (ou risque spécifique) pour le  $i$ ème individu du groupe observé.  $z_i^a$  et  $z_i^r$  sont des sous-ensembles de  $z_i$  qui représentent les vecteurs de variables liées respectivement au risque attendu et au risque spécifique.

Pour modéliser la fonction de risque instantané attendu, nous avons fait l'hypothèse d'un modèle semi-paramétrique à risques proportionnels tel que proposé par Cox (1972). Ainsi, le risque instantané attendu pour le  $j$ ème individu ( $j = 1, \dots, n_a$ ) du groupe contrôle peut être ré-écrit :

$$h^a(t_j|z_j^a) = h_0^a(t_j) \exp(\beta^a z_j^a) \quad (2)$$

avec  $h_0^a(t_j)$  la fonction de risque instantané attendu de base non estimée et  $\beta^a$  le vecteur des coefficients de régression associés aux variables  $z_j^a$ . L'estimation des paramètres  $\hat{\beta}^a$  du modèle est obtenue en maximisant le logarithme de la vraisemblance partielle parmi les  $n_a$  individus du groupe contrôle :

$$\log \mathcal{PL}_a(\beta^a) = \sum_{j=1}^{n_a} \delta_j \left\{ \beta^a z_j^a - \log \left( \sum_{k:t_k \geq t_j} \exp(\beta^a z_k^a) \right) \right\} \quad (3)$$

avec  $\delta_j$  l'indicatrice de l'événement, égale à 1 lorsque l'échec est observé pour le  $j$ ème individu et 0 sinon. L'estimation de la matrice de variance-covariance,  $\hat{V}(\hat{\beta}^a)$  s'appuie sur l'estimation de la matrice Hessienne correspondante.

Nous avons également fait l'hypothèse d'un modèle semi-paramétrique à risques proportionnels pour modéliser la fonction de risque instantané relatif. Ainsi le risque spécifique

pour le  $i$ ème individu du groupe d'intérêt ( $i = 1, \dots, n_r$ ) est défini par :

$$h^r(t_i|z_i) = h_0^r(t_i) \exp(\beta^r z_i^r) \quad (4)$$

avec  $h_0^r(t_i)$  la fonction de risque instantané relatif de base non estimée et  $\beta^r$  le vecteur des coefficients de régression associés aux variables  $z_i^r$ . Si on fait l'hypothèse que les paramètres de la partie attendue estimés juste avant sont non-aléatoires, les paramètres  $\beta^r$  peuvent être obtenus en maximisant :

$$\log \mathcal{P}\mathcal{L}_r(\beta^r) = \sum_{i=1}^{n_r} \delta_i \left\{ \hat{\beta}^a z_i^a + \beta^r z_i^r - \log \left( \sum_{k:t_k \geq t_i} \exp(\hat{\beta}^a z_k^a) \exp(\beta^r z_k^r) \right) \right\} \quad (5)$$

Pour une variable prise en compte seulement dans la partie relative dans le modèle (1),  $\exp(\beta^r)$  représente le rapport des risques (RR) dans le groupe contrôle, identique au RR estimé dans un modèle à risque proportionnel (4) à partir des individus du groupe d'intérêt. A l'inverse, pour les variable prises en compte dans la partie attendue et dans la partie relative dans le modèle (1),  $\exp(\beta^r)$  représente un facteur de pondération entre le RR attendu dans le groupe contrôle ( $\exp(\hat{\beta}^a)$ ) et le RR observé dans le groupe d'intérêt ( $\exp(\hat{\beta}^a) \times \exp(\beta^r)$ ). Ainsi, pour les variables impliquées dans les deux modèles (2) et (4),  $\beta^r = 0$  signifie que la variable a le même effet dans les deux groupes. Si  $\beta^r > 0$  ou  $\beta^r < 0$ , l'effet lié à la variable est respectivement augmenté ou diminué dans le groupe d'intérêt par rapport groupe contrôle.

Contrairement aux modèles de survie relative classiques basés sur les tables de mortalité, le risque attendu estimé ici ne peut pas raisonnablement être supposé constant étant donné que ses paramètres étaient estimés à partir d'un échantillon. Pour prendre en compte cette variabilité associée au modèle attendu (2) dans l'estimation de la partie relative (4), nous avons développé une procédure en 3 étapes basée sur des simulations de Monte-Carlo associée à un ré-échantillonnage par bootstrap ( $b = 1, \dots, 1000$ ) :

- (a) Génération de 1000 vecteurs de paramètres  $\hat{\beta}_b^{a*}$  simulés en s'appuyant sur une distribution normale multivariée  $\mathcal{N}(\hat{\beta}^a, \hat{V}(\hat{\beta}^a))$ , et obtenus par maximisation de la log-vraisemblance partielle (3). Cette première étape permettait de prendre en compte la variance du risque attendu.
- (b) Génération de 1000 échantillons de bootstrap de taille  $n_r$  à partir du groupe des STR. Cette seconde étape permettait de prendre en compte la fluctuation d'échantillonnage.
- (c) A partir de ces 1000 échantillons de bootstrap, estimation de 1000 vecteurs de paramètres  $\hat{\beta}_b^r$  de la partie relative, obtenus par maximisation de la log-vraisemblance partielle (5) dans laquelle les paramètres simulés  $\hat{\beta}_b^{a*}$  à chaque itération  $b$  étaient utilisés à la place des paramètres  $\hat{\beta}^a$ .

Le vecteur  $\hat{\beta}^r$  final a été obtenu par les moyennes des 1000 vecteurs de paramètres relatifs  $\hat{\beta}_b^r$  ainsi estimés et les intervalles de confiance ont été obtenus par les percentiles.

### 3 Le modèle de Cox stratifié (MCS)

Pour le  $i$ ème individu de la cohorte complète ( $i = 1, \dots, n$ ), c'est à dire incluant le groupe contrôle et le groupe d'intérêt ( $n = n_a + n_r$ ),  $z_i$  correspond au vecteur de variables explicatives entrant dans le modèle pour les premières et/ou les secondes transplantations. Nous avons développé un modèle de Cox stratifié spécifiant le rang de greffe comme une strate ( $k = a$  pour le groupe contrôle et  $k = r$  pour le groupe d'intérêt). Le modèle de Cox stratifié est alors défini par :

$$h_k(t_i|z_i) = h_{k,0}(t_i) \exp(\beta z_i); \quad \text{for } k = a, r. \quad (6)$$

où  $h_{k,0}(t_i)$  est la fonction de risque instantané de base dans la strate  $k$  au temps  $t_i$ .

Le vecteur  $z_i$  de variables explicatives peut être divisé en quatre sous-vecteurs : (i) un vecteur  $z_i^a$  de variables explicatives entrant uniquement dans le risque attendu (leurs valeurs valent 0 si  $k = r$ ); (ii) un vecteur  $z_i^r$  de variables explicatives entrant uniquement dans le risque spécifique (leurs valeurs valent 0 si  $k = a$ ); (iii) un vecteur  $z_i^c$  de variables explicatives communes et avec le même effet dans les deux groupes; et (iv) un vecteur  $z_i^d$  sous-vecteur des variables explicatives de  $z_i^c$  mais avec des effets différents entre les deux strates. Le modèle de Cox stratifié (6) peut être ré-écrit :

$$h_k(t_i|z_i) = h_{k,0}(t_i) \exp(\beta^a z_i^a + \beta^c z_i^c + \beta^d z_i^d \delta_{ir} + \beta^r z_i^r); \quad \text{for } k = a, r. \quad (7)$$

avec  $\beta = (\beta^a, \beta^c, \beta^d, \beta^r)$  le vecteur des coefficients de régression et  $\delta_{ir}$  égal à 1 pour l'individu  $i$  appartenant à la strate  $k = r$  et 0 sinon.

Alors,  $\exp(\beta^a)$  représente le RR associé aux variables spécifiques du groupe contrôle et  $\exp(\beta^r)$  représente le RR associé aux variables spécifiques du groupe d'intérêt. Pour les variables prises en compte uniquement dans le vecteur  $z^c$  (mais pas dans le vecteur  $z^d$ ),  $\exp(\beta^c)$  représente le RR associé aux  $z^c$ , commun aux deux groupes. Pour les variables prises en compte à la fois dans  $z^c$  et dans  $z^d$ , les RR associés aux  $z^c$  valent  $\exp(\beta^c)$  dans le groupe contrôle et  $\exp(\beta^c + \beta^d)$  dans le groupe d'intérêt. Ainsi,  $\exp(\beta^d)$  représente un facteur de pondération entre le RR attendu dans le groupe contrôle ( $\exp(\beta^c)$ ) et le RR observé dans le groupe d'intérêt ( $\exp(\beta^c) \times \exp(\beta^d)$ ). On retrouve ainsi le principe des modèles de survie relative à risques multiplicatifs qui ont pour objet de tester l'homogénéité de l'effet des facteurs de risque entre deux populations.

## 4 Résultats

Dans cette étude, nous avons utilisé les deux stratégies afin de comparer l'effet des facteurs de risque sur le risque d'échec de greffe entre les premières et les secondes transplantations rénales. Les résultats des deux stratégies étaient concordants. Ils ont permis de montrer que le sexe masculin du donneur et une longue attente en dialyse avant la retransplantation étaient des facteurs de risque spécifiques des STR. De plus, l'effet délétère lié à un âge élevé du receveur semblait potentialisé chez les STR par rapport aux PTR alors que l'effet délétère lié à un âge élevé du donneur semblait atténué chez les STR par rapport aux PTR.

## 5 Discussion

Nous avons illustré l'avantage de ces approches alternatives avec la problématique des secondes transplantations rénales mais elles pourraient être utiles dans bon nombre d'autres applications cliniques et épidémiologiques lorsque l'objectif de l'étude est de comparer l'effet des facteurs de risque entre deux populations et que l'ajustement sur des facteurs spécifiques d'une des populations est susceptible d'apporter une meilleure approche du risque.

## Bibliographie

- [1] Trébern-Launay, K. et Foucher, Y. et Giral, M. et Legendre, C. et Kreis, H. et Kessler, M. et Ladrière, M. et Kamar, N. et Rostaing, L. et Garrigue, V. et Mourad, G. et Morelon, E. et Soulillou, JP. et Dantal, J. (2012), Poor Long-Term Outcome in Second Kidney Transplantation: A Delayed Event, PLoS ONE, 7, e47915.
- [2] Andersen, PK. et Borch-Johnsen, K. et Deckert, T. et Green, A. et Hougaard, P. et Keiding, N. et Kreiner, S. (1985), A Cox regression model for the relative mortality and its application to diabetes mellitus survival data, Biometrics, 41-4, 921–932.
- [3] Cox, DR. (1972), Regression Models and Life-Tables, Journal of the Royal Statistical Society. Series B (Methodological), 34-2, 187–220.