

Rapport de stage M2 statistique

Yann Labrune

10 avril 2008

Table des matières

1	Introduction	3
2	Présentation de la structure d'accueil	4
3	Théorie de l'analyse de survie	5
3.1	Définitions importantes	5
3.1.1	La censure	5
3.1.2	la troncature	6
3.1.3	Les différents modèles d'analyse de survie	6
3.2	Fonctions associées aux distributions de survie	6
3.3	Estimation de la survie : la méthode de Kaplan-Meier	7
3.4	Le test du Log-Rank	7
3.5	Le modèle de Cox	8
3.5.1	Principe du modèle	8
3.5.2	Calcul des coefficients de régression β	8
3.5.3	Interprétation des coefficients de régression β	9
3.5.4	Hypothèses à vérifier	9
3.5.5	Test d'hypothèse des coefficients de régression β	10
3.5.6	Le modèle stratifié	11
3.5.7	Les données temps-dépendantes	11
4	L'influence du poids du rein dans la survie des patients transplantés	14
4.1	Problématique	14
4.2	Descriptif de l'échantillon	15
4.2.1	Descriptif des covariables	15
4.3	méthodes	17
4.3.1	Caractérisation du poids du rein	17
4.3.2	Choix des valeurs seuil	17
4.3.3	Validation des modèles	17
4.4	statistiques descriptives	17
4.5	Résultats	19
4.5.1	Tests du Log-Rank	19
4.5.2	L'analyse multivariée, le modèle de Cox	20
4.5.3	discussion des résultats	21
4.5.4	Modèle de Cox corrigé	21
4.5.5	discussion des résultats corrigé	22
5	L'influence du poids du rein sur l'apparition de protéinurie	23
5.1	Problématique	23
5.2	résultats	23
5.2.1	modèles univariés	23
5.2.2	Modèle de Cox : la protéinurie	24

5.2.3	discussion des résultats	25
A	Hypothèses de proportionnalité et de log-linéarité	26

Chapitre 1

Introduction

Chapitre 2

Présentation de la structure d'accueil

Chapitre 3

Théorie de l'analyse de survie

L'analyse de données de survie est l'étude du délai de survenue d'un évènement précis. Ce dernier peut être un rejet de greffon, l'apparition d'un cancer ou la mort réelle des individus étudiés. Grâce à cette analyse, nous sommes capables de décrire les durées de survie. Nous pouvons également expliquer ou prédire la durée de survie de sujets par rapport à certains facteurs étudiés, ces derniers étant appelés facteurs pronostiques.

L'analyse de survie est une branche particulière des statistiques avec ses méthodes et ses définitions propres. Nous allons donc ici présenter les différentes méthodes que l'on peut utiliser en analyse de données de survie ainsi que quelques notions nécessaires à la compréhension du modèle de Cox.

3.1 Définitions importantes

Tout d'abord, il est important de définir la date d'origine associée au délai de survenue de l'évènement que l'on souhaite étudier. Cette date correspond souvent à la date d'entrée des patients dans l'étude. La variable étudiée en analyse de survie est le délai de survenue d'un évènement. Cette variable aléatoire, continue et positive est appelée « **durée de survie** ». Une des caractéristiques principales des données de survie est l'existence de données incomplètes. Ces données peuvent être soit censurées soit tronquées.

3.1.1 La censure

La censure est le phénomène le plus souvent rencontré lorsque l'on recueille des données de survie.

La censure à droite Une durée de survie est censurée à droite si l'individu n'a pas subi l'évènement étudié avant sa dernière observation. Les individus qui se trouvent dans ce cas ne peuvent pas être supprimés étant donné que ces données nous apportent quand même une information : la date de survenue de l'évènement étudié est supérieure à la date de la dernière observation (en supposant que la date de début de l'étude est notée $t=0$). La censure à droite peut être rencontrée pour deux raisons :

- Le sujet n'a pas encore subi l'évènement à la fin de l'étude.
- Le sujet (qui n'avait pas encore subi l'évènement) a quitté l'étude en cours.

La censure à gauche Une durée de survie est censurée à gauche si la seule information en notre possession est que l'évènement étudié a eu lieu avant une date t .

La censure par intervalle Une durée de survie est censurée par intervalle si on sait que l'évènement étudié a eu lieu entre deux dates.

3.1.2 la troncature

Une observation est dite tronquée si elle est observée conditionnellement à un autre évènement.

La troncature à gauche Seul les individus qui n'ont pas présentés l'évènement étudié à la date d'inclusion dans l'étude sont retenus pour celle-ci. Ici, la date d'inclusion dans l'étude et le commencement de celle-ci sont différents. Pour éviter ce type de troncature il faudrait suivre les individus depuis leur naissance.

3.1.3 Les différents modèles d'analyse de survie

Différentes approches existent pour aborder les données d'analyse de survie. On distingue deux grands types de méthodes, les méthodes paramétriques et les méthodes non-paramétriques. Parmi les méthodes non-paramétriques, la plus connue est la méthode de Kaplan-Meier. Cette méthode ne cherche pas à modéliser la fonction de risque mais plutôt à la décrire. Les méthodes paramétriques reposent quant à elles sur des hypothèses sur la forme de la fonction de risque. Le modèle le plus fréquemment utilisé est le modèle de Weibull.

Le modèle de Cox ne fait partie d'aucun de ces deux groupes. En effet, c'est un modèle semi-paramétrique ce qui lui permet d'être plus efficace que les méthodes non-paramétriques et moins contraignant en terme d'hypothèse que les méthodes paramétriques.

3.2 Fonctions associées aux distributions de survie

Soit X une variable aléatoire de durée de vie. Plusieurs fonctions peuvent caractériser la loi de probabilité de X :

– **La fonction de densité** $f(x)$

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x}$$

$f(x)$ représente la probabilité que l'évènement se produise juste après x .

– **La fonction de répartition** $F(x)$

$$F(x) = P(X \leq x) = \int_0^x f(u) du$$

Cette fonction représente la probabilité de présenter l'évènement entre 0 et x , elle est liée à la fonction de densité de probabilité.

– **La fonction de survie** $S(x)$

$$S(x) = P(X > x) = 1 - F(x) = \int_x^{\infty} f(u) du$$

La fonction de survie représente la probabilité de survivre au-delà de x .

– **La fonction de risque instantané** $\lambda(x)$

$$\lambda(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x | X \geq x)}{\Delta x}$$

Cette fonction représente la probabilité que l'évènement se produise juste après x sachant qu'il n'a pas eu lieu avant x .

– **La fonction de risque cumulé** $\Lambda(x)$

$$\Lambda(x) = \int_0^x \lambda(t) dt$$

C'est l'intégrale de la fonction de risque.

3.3 Estimation de la survie : la méthode de Kaplan-Meier

La méthode de Kaplan-Meier est la méthode non paramétrique la plus utilisée aujourd'hui. Elle repose sur une idée simple : être encore en vie après un instant t c'est être en vie juste avant t et ne pas mourir à l'instant t . A partir de cette idée, nous allons être capable de construire un "tableau de survie" qui nous donnera la valeur de la fonction de survie à chaque temps de mort d'un individu. La fonction de survie par la méthode de Kaplan-Meier s'écrit comme suit :

$$S(t) = \prod_{t_i < t} \frac{(n_i - d_i)}{n_i}$$

où n_i est le nombre d'individus à risque au temps t_i et d_i est le nombre de personne décédé au temps t_i .

3.4 Le test du Log-Rank

Le test du Log-Rank est un test permettant de comparer deux ou plusieurs fonctions de survie à la condition que les courbes de survie étudiées ne se croisent pas. L'hypothèse H_0 testée est l'égalité des fonctions de survie contre H_1 les fonctions de survie sont différentes.

Son principe est le suivant : à chaque temps de décès, on calcule la proportion de sujet à risque dans chacun des groupes, on cumule les différences entre décès observés et décès attendus jusqu'à la date du dernier décès observé, on évalue la significativité de la différence cumulée en la divisant par sa variance.

Soit $n_{A,i}$ et $n_{B,i}$: le nombre de sujets à risque au temps t_i dans les groupes A et B et $n_i = n_{A,i} + n_{B,i}$ les nombres de sujets à risque total au temps t_i .

Soit $d_{A,i}$ et $d_{B,i}$: le nombre d'événement observé au temps t_i dans les groupes A et B et $d_i = d_{A,i} + d_{B,i}$ le nombre d'événement total observé au temps t_i .

La statistique de test est calculée comme suit :

$$U = \sum_{t_i} d_{B,i} - \frac{d_i n_{B,i}}{n_i}$$
$$Var(U) = \sum_{t_i} d_i \left(\frac{n_i - d_i}{n_i - 1} \right) \frac{n_{A,i} n_{B,i}}{n_i^2}$$

On obtient :

$$\frac{U^2}{Var(U)} \rightsquigarrow \chi_1^2$$

Si la valeur de la statistique est supérieure à un χ^2 à 1 degré de liberté avec un risque α fixé, on rejette l'hypothèse d'égalité des fonctions de survie.

3.5 Le modèle de Cox

3.5.1 Principe du modèle

Le modèle de Cox a été introduit en 1972 par David Cox. C'est le modèle le plus utilisé en analyse des données de survie avec censure à droite dans le domaine de la biostatistique. Historiquement, il a d'abord été utilisé pour analyser des durées de survie, c'est-à-dire le délai jusqu'à la survenue du décès. C'est pourquoi le vocabulaire employé classiquement utilise souvent les termes de "décédé" et "vivant" plutôt que ceux de "malade" et "non malade". Il peut s'appliquer cependant à tout autre type d'étude où l'on étudie le délai d'apparition d'un évènement.

Le modèle de Cox est devenu au fil des ans le modèle de référence pour l'analyse statistique dans le secteur médical. Il permet d'étudier les facteurs de risque de décès sans donner à la distribution des durées de survie une forme particulière. Il s'agit donc d'un modèle semi-paramétrique.

Il existe deux hypothèses pour la validité de ce modèle, l'hypothèse de log-linéarité des données et l'hypothèse de proportionnalité que nous verrons en détail dans la suite de ce rapport.

Le modèle de Cox exprime une relation entre la fonction de risque associée à la survenue d'un évènement et un vecteur de p variables explicatives $Z = (Z_1, \dots, Z_p)$. Soit $\beta = (\beta_1, \dots, \beta_p)$ un vecteur de p coefficients associés aux p variables explicatives.

On note la fonction de risque instantané du modèle de Cox comme suit :

$$\lambda(t, Z_1, \dots, Z_p) = \lambda_0(t) \underbrace{\exp\left(\sum_{i=1}^p \beta_i Z_i\right)}_{\text{partie paramétrique}}$$

On constate que cette fonction de risque est séparée en deux parties.

La première partie $\lambda_0(t)$ est la fonction de risque de base au temps t , c'est à dire la fonction de risque lorsque toutes les variables $Z = (Z_1, \dots, Z_p)$ sont égales à zéro.

La seconde partie $\exp(\sum_{i=1}^p \beta_i Z_i)$ représente les différences de risque apportées par le vecteur des variables Z . C'est seulement cette partie que l'on va tenter de modéliser grâce au modèle de Cox. On peut noter que cette partie ne dépend pas du temps, cela implique que l'association entre Z_i et la survenue du décès est constante au cours du temps. Cette hypothèse (appelée hypothèse des risques proportionnels) est une des hypothèses importantes du modèle de Cox sur laquelle nous reviendrons. Nous pouvons aussi écrire le modèle de Cox par rapport à la fonction de survie, ainsi on obtient :

$$\begin{aligned} S(t, Z_1, \dots, Z_p) &= \exp\left\{-\int_0^t \lambda(u, Z_1, \dots, Z_p) du\right\} \\ &= \exp\left\{-\int_0^t \exp\left(\sum_{i=1}^p \beta_i Z_i\right) \lambda_0(u) du\right\} \\ &= \exp\left\{-\exp\left(\sum_{i=1}^p \beta_i Z_i\right) \int_0^t \lambda_0(u) du\right\} \\ &= \exp\left\{-\int_0^t \lambda_0(u) du\right\}^{\exp(\sum_{i=1}^p \beta_i Z_i)} \\ &= S_0(t)^{\exp(\sum_{i=1}^p \beta_i Z_i)} \end{aligned}$$

Où $S_0(t)$ est la fonction de survie de base.

3.5.2 Calcul des coefficients de régression β

Le principal but de ce modèle est d'étudier l'incidence des variables $Z = (Z_1, \dots, Z_p)$ sur la survie (ou sur la fonction de risque, nous avons démontré que c'était équivalent). Pour cela, il nous faut calculer les coefficients β associés à chaque variable. Afin de trouver ces coefficients, on calcule la vraisemblance partielle du modèle. La vraisemblance partielle (ou vraisemblance de Cox) correspond

au produit des probabilités conditionnelles d'observer à l'instant t_i l'évènement D_i sachant que l'on a le groupe A_i de sujets à risque juste avant t_i . Afin d'estimer le vecteur des paramètres β , il faut maximiser cette vraisemblance. Soit $L^*(\beta)$ la vraisemblance partielle du modèle de Cox.

$$\begin{aligned} L^*(\beta) &= \prod_{i \in D} \frac{\lambda_0(t_i) \exp(\beta Z_i)}{\sum_{l \in A_i} \lambda_0(t_i) \exp(\beta Z_l)} \\ &= \prod_{i \in D} \frac{\exp(\beta Z_i)}{\sum_{l \in A_i} \exp(\beta Z_l)} \end{aligned}$$

On peut noter que en présence d'ex-aequos, cette vraisemblance n'est pas calculable. L'estimateur de β ainsi que sa variance sont obtenus par la méthode du maximum de vraisemblance (partielle). Lorsque les paramètres composant le vecteur β sont déterminés, on va alors vouloir interpréter ces coefficients et tester si la survie est significativement indépendante de la variable Z_j associée au paramètre β_j .

3.5.3 Interprétation des coefficients de régression β

Pour comprendre le sens des coefficients du modèle, nous allons considérer seulement une variable dichotomique Z codée (0,1). Le risque instantané chez les sujets du groupe correspondant à $Z = 0$ est égal à $\lambda_0(t)$; chez ceux du groupe correspondant à $Z = 1$, il vaut $\lambda_1(t) = \lambda_0(t) \exp(\beta)$. Considérons le risque relatif à l'instant t $RR(t)$, qui est par définition le rapport des fonctions de risque :

$$RR(t) = \frac{\lambda_1(t)}{\lambda_0(t)} = \exp(\beta)$$

L'interprétation du risque relatif est assez intuitive, les individus du groupe $Z = 1$ ont $RR(t)$ fois plus de chance de mourir au temps t que les individus du groupe correspondant à $Z = 0$. Un risque relatif inférieur à 1 indique quel groupe où $Z = 0$ est plus à risque que le groupe où $Z = 1$, il n'y a pas de différence de survie entre les groupes quand $RR(t) = 1$ et donc quand $\beta = 0$. Le risque relatif est constant au cours du temps dans ce modèle vu qu'il ne dépend que de β , ce qui implique une proportionnalité des fonctions de risque (c'est pourquoi on dit que ce modèle est à risque proportionnel).

3.5.4 Hypothèses à vérifier

hypothèse de proportionnalité

Cette hypothèse est à vérifier pour la validité du modèle. Pour la vérifier nous allons utiliser une méthode de validation graphique. Il nous faut estimer les courbes de survie dans chacun des groupes puis de tracer les courbes $\log(-\log(S(t)))$. Pour que cette hypothèse soit vérifiée, les courbes doivent avoir un écart constant au cours du temps :

$$\begin{aligned} \text{Log}(-\text{Log}(S(t, Z))) &= \text{Log}(-\text{Log}(S_0(t)^{\exp(\sum_{i=1}^p \beta_i Z_i)})) \\ &= \text{Log}(\exp(\sum_{i=1}^p \beta_i Z_i) \times -\text{Log}(S_0(t))) \\ &= \sum_{i=1}^p \beta_i Z_i + \text{Log}(-\text{Log}(S_0(t))) \end{aligned}$$

Pour les groupes $Z = 0$ et $Z = 1$ l'écart entre les deux courbes est de β . Si sur le graphique les courbes n'ont pas un écart constant, le β n'est donc pas constant, ce qui n'est pas pris en compte dans le modèle de Cox.

Hypothèse de log-linéarité

L'hypothèse sous jacente lorsque l'on utilise une variable explicative dans un modèle de Cox est l'hypothèse de log-linéarité. Elle suppose que le coefficient de régression β augmente linéairement pour une augmentation d'une unité de la variable associée. Afin de tester cette hypothèse, on peut discrétiser la variable en plusieurs groupes, puis tracer un nuage de point avec en abscisse la valeur centre du groupe et en ordonnée la valeur du coefficient β calculé au sein de ce groupe. L'hypothèse est vérifiée si les points du nuage sont alignés.

Comme nous pouvons nous en douter, cette hypothèse est rarement vérifiée dans la pratique. Ainsi, une solution consiste à mettre les variables n'ayant que peu de modalités sous une forme disjonctive complète, et de choisir une valeur seuil pour les autres variables (on sépare la variable en deux groupes distincts).

3.5.5 Test d'hypothèse des coefficients de régression β

Il existe plusieurs tests qui vont nous renseigner sur l'effet des différentes variables sur la survie. Cela revient à tester la nullité des coefficients $(\beta_1, \dots, \beta_p)$. Nous allons voir dans cette partie deux types de tests qui vont pouvoir répondre à cette problématique : le test de Wald et le test du rapport de vraisemblance.

Le test de Wald

Soit β_i ($i = 1, \dots, p$) le paramètre correspondant à la i^{e} variable. L'hypothèse à tester sera $H_0 : \beta_i = 0$ contre $H_1 : \beta_i \neq 0$. Sous H_0 , le rapport du coefficient à son écart type suit une loi normale centrée réduite (Théorème Central Limite) :

$$W = \frac{\beta_i}{\sqrt{\text{Var}(\beta_i)}} \rightsquigarrow N(0, 1)$$

Si $W \in [-1, 96; 1, 96]$ alors on ne rejettera pas l'hypothèse H_0 et on pourra dire que le coefficient β n'est significativement pas différent de 0 au seuil 0,05.

Ici, on ne teste qu'un seul coefficient, il serait possible de tester plusieurs coefficients à la fois par ce test mais cela complexifierait d'autant plus la statistique de test ; il est préférable dans ce cas d'utiliser le test du rapport de vraisemblance.

Le test du rapport de vraisemblance

Soit β le vecteur des paramètres que l'on veut tester.
Soit $L^*(0)$ la vraisemblance partielle du modèle de Cox sans les paramètres que l'on souhaite tester.
Soit $L^*(1)$ la vraisemblance partielle du modèle de Cox avec les paramètres que l'on souhaite tester.
La statistique de test est :

$$RV = -2 \times \log\left(\frac{L^*(0)}{L^*(1)}\right) \rightsquigarrow \chi_d^2$$

Sous l'hypothèse nulle, RV suit une loi du χ_d^2 où d est le nombre de degrés de liberté et correspond aux nombres de paramètres testés.

3.5.6 Le modèle stratifié

Le modèle de Cox à cet avantage de ne pas estimer la fonction de risque de base, on peut donc aisément stratifier ce modèle en définissant une fonction de risque de base pour chaque strate, mais le risque relatif de chaque variable reste identique dans chaque strate.

Le modèle s'écrit alors comme suit :

$$\lambda_k(t|z) = \lambda(t|z, \text{strate } k) = \lambda_{0k}(t) \exp(\beta z)$$

La vraisemblance partiel de Cox du modèle stratifié se calcul alors comme suit :

$$\begin{aligned} L^*(\beta) &= \prod_{i \in D} \frac{\lambda_{0k}(t_i) \exp(\beta Z_i)}{\sum_{l \in A_i} \lambda_{0k}(t_i) \exp(\beta Z_l)} \\ &= \prod_{k=1}^K \prod_{i=1}^{n_k} \frac{\lambda_{0k}(t_i) \exp(\beta Z_i)}{\sum_{l \in A_i} \lambda_{0k}(t_i) \exp(\beta Z_l)} \\ &= \prod_{k=1}^K L_k^*(\beta) \end{aligned}$$

où $L_k^*(\beta)$ représente la vraisemblance partielle de cox pour le modèle de la strate k .

On obtient alors un ajustement sur la variable de stratification (dans notre cas le centre hospitalier). La variable sur laquelle on stratifie ne repose plus nécessairement sur l'hypothèse de proportionnalité puisque l'on applique un modèle de Cox par strates. Cependant, aucune estimation directe de l'importance de l'effet de la strate sur la survie n'est fourni par le modèle et il faut garder à l'esprit que la précision dans l'estimation des coefficients et la puissance de l'analyse diminuent avec le nombre de strates.

3.5.7 Les données temps-dépendantes

Il reste relativement fréquent d'avoir une des variables explicatives qui est dépendante du temps. On entend par temps-dépendants soit l'effet de la variable variant au cours du temps, soit la valeur de la variable qui se modifie. Le modèle de Cox reste-il une solution à notre problème ? Nous pouvons toujours utiliser le modèle de Cox en y appelant quelques modifications. Il y a trois solutions :

1. Stratifier le modèle sur la variable temps-dépendante (valable uniquement pour les variables qualitatives)
2. Partitionner le temps (nous obtiendrons ainsi des risques relatifs différents par intervalles de temps)
3. Estimer les paramètres du modèle avec des variables dépendantes du temps.

Certaines de nos variables temps dépendantes sont par exemple des mesures biologique (créatinine, protéinurie...) issu d'échantillons sanguin ou urinaire. En pratique, les valeur de ces variables sont inconnues la majeure partie du temps. A la place, on utilise la valeur la plus récente qui à été mesurée. Puisqu'elles sont mesurées qu'une douzaine de fois pour chaque patient, on préférera le terme de "*mis à jour*" plutôt que celui de variables "*temps-dépendantes*".

Le terme variable temps dépendante va être réservé dans notre étude aux variables qualitatives dont on connaît la date d'apparition et donc la valeur exacte tout au long du suivi (le rejet aigu par exemple).

Pour comprendre comment analyser ce type de données, il est remarquable que le modèle de Cox ne fait pas la différence entre un individu qui rentre dans l'étude à un temps 0 et qui meurt à un temps t (C1) et deux individus dont l'un entre au temps 0 et est censuré à la date u (C2) et l'autre entre au temps u et meurt à la date t (C3) :

$$\begin{aligned}
 C1 &= \lim_{h \rightarrow 0} \frac{P(t < T < t + h)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{P(t < T < t + h, T > u)}{h} \text{ car par définition } P(t > u) \\
 &= \lim_{h \rightarrow 0} \frac{P(t < T < t + h | T > u)}{h} P(t > u) \\
 &= C2 \times C3
 \end{aligned}$$

Lorsqu'on a une variable temps dépendante, il est donc possible de remplacer un individu par plusieurs sujets "fictifs" correspondant à chaque changement de la variable temps dépendante. Par exemple :

<i>TempsDebut</i>	<i>TempsFin</i>	<i>RejetAigu</i>	<i>status</i>	<i>Var1</i>	<i>Var2</i>
0	6,6	0	0	1	0
6,6	111	1	1	1	0

Dans cet exemple, l'individu meurt à 111 mois mais fait un épisode de rejet aigu entre le 6^e et 7^e mois ; avant cette date il n'y a donc pas de rejet aigu et après cette date on note l'apparition du rejet. On peut noter que les autres variables restent les mêmes d'une ligne à l'autre.

Bien que le modèle de Cox ne fasse pas la différence, pour que celui-ci soit valide, il faut que tous les individus soient indépendants entre eux, ce qui n'est plus le cas. Néanmoins, il est couramment admis que les résultats du modèle ne changent pas énormément.

Pour une variable temps-dépendante dont c'est l'effet (et non plus la valeur de la variable) qui varie au cours du temps, la méthode n'est pas toujours la même, on va séparer plusieurs périodes dans lesquelles l'effet de la variable est constant. Pour ce faire, on va procéder comme précédemment à la différence que nous allons créer cette fois plusieurs variables. Dans notre exemple la variable x à un effet variant au cours du temps, cet effet est cependant constant dans les intervalles $[0; 24]$ et $]24; 144]$ où 144 est le temps de fin de l'étude.

$$\begin{aligned}
 \lambda(t, z, x) &= \lambda_0(t) \exp(\beta z + \beta(t)x) \\
 &= \lambda_0(t) \exp(\beta z + \mathbb{1}_{[0;24]} \gamma_1 x + \mathbb{1}_{]24;144]} \gamma_2 x) \\
 &= \lambda_0(t) \exp(\beta z + \gamma_1 x_1 + \gamma_2 x_2)
 \end{aligned}$$

x_1 correspond à la variable x sur l'intervalle de temps $[0; 24]$.

x_2 correspond à la variable x sur l'intervalle de temps $]24; 144]$.

Exemple pour un individu mort au temps 111 dont $x = 1$ et $z = 1$:

<i>Tempsdebut</i>	<i>Tempsfin</i>	<i>x1</i>	<i>x2</i>	<i>status</i>	<i>z</i>
0	24	1	0	0	1
24	111	0	1	1	1

Exemple pour un individu mort au temps 93 ayant fait un rejet au temps 0.25 dont $x = 1$ et $z = 1$:

<i>Tempsdebut</i>	<i>Tempsfin</i>	<i>x1</i>	<i>x2</i>	<i>Rejet</i>	<i>status</i>	<i>z</i>
0	0,25	1	0	0	0	1
0,25	24	1	0	1	0	1
24	93	0	1	1	1	1

Pour les individus ayant $x = 0$, les variables $x1$ et $x2$ restent toutes deux à zéro.

Chapitre 4

L'influence du poids du rein dans la survie des patients transplantés

4.1 Problématique

En 2006, parmi les 9 226 malades en attente de greffe rénale, 3 274 malades ont accédé à la liste d'attente au cours de l'année et 5 952 malades étaient en attente au 1er janvier. Le nombre estimé de porteurs de greffon fonctionnel en France est de 27 492 au 31 décembre 2006. Ces chiffres nous montrent que la greffe de rein et a fortiori de tout organes n'est pas un phénomène rare comme on pourrait le penser mais au contraire une procédure médicale de plus en plus courante. Malheureusement, à l'heure actuelle nous sommes face à une véritable pénurie d'organe ce qui se ressent sur le temps d'attente des malades. En 2005, il y avait 3,4 candidats pour un greffon utilisable dans l'année. De cela découle un temps d'attente de plus en plus long, en 2006 après 18,6 mois d'attente (médiane d'attente), les malades inscrits entre 2003 et 2006 ont 50 % de chance d'être greffés (4.1).

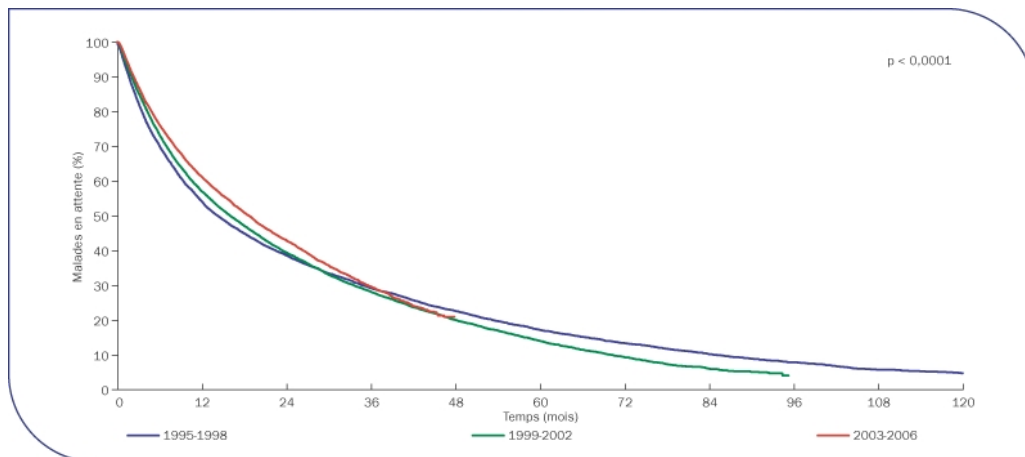


FIG. 4.1 – Durée d'attente avant greffe de rein par période d'inscription

Les patients les plus touchés sont dialysés plusieurs fois par semaine pendant cette attente. Malheureusement, la dialyse, seule alternative à la transplantation, ne permet pas de guérison et empêche le patient de vivre normalement ; la transplantation est la seule véritable solution. Les reins sont attribués selon un système assez complexe qui est basé sur plusieurs critères tels que l'état du patient, la compatibilité HLA et l'âge...

Au vu du nombre très supérieur de demande que d'offre, il est très important pour les médecins de déterminer les facteurs qui ont un effet sur la survie des greffons afin de pouvoir répondre au mieux aux attentes des patients et d'optimiser la durée de vie des greffons. Beaucoup de ces facteurs ont déjà été mis à jour par les connaissances sans cesse grandissantes des médecins sur le corps humain, il est

cependant souvent nécessaire de valider les connaissances ou intuitions des médecins par la statistique afin de pouvoir quantifier le gain apporté par la prise en compte de tel ou tel facteur. Notre étude consiste à tester l'effet du poids du rein sur la survie des patients transplantés rénaux.

4.2 Descriptif de l'échantillon

La base de données est constituée de 1189 greffes de rein ou de rein-pancréat. La plus ancienne greffe a été effectuée le 3 avril 1995 et la plus récente le 20 janvier 2006. Elle contient plusieurs dizaines de variables telles que des informations sur le donneur, le receveur, les antécédents, diverses infections, divers relevés immunologiques, etc.

Nous avons été dans l'obligation de supprimer certaines greffes de notre étude pour plusieurs raisons :

- Les individus censurés qui n'ont eu aucun suivi (n=39).
- Ceux dont on ne connaît pas le poids du rein greffé (variable centrale de notre étude) (n=5).
- Les individus qui ont été greffés du rein et du pancréas en même temps, ce qui implique une différence de protocole opératoire (n=64).
- Les individus greffés de moins de 18 ans qui posent un problème légal et qui ne sont pas comparable au adulte en terme de survie. (n=20).
- Les individus dont on ne connaît pas le poids (n=1).

Notre base finale ne contient donc plus que 1060 individus, les statistiques ont été faites sur ceux-ci.

4.2.1 Descriptif des covariables

Nous n'avons eu besoin que d'une petite partie des variables de cette base. Nous allons ici présenter celles qui sont importantes pour les experts médicaux :

- **Le poids du rein :**

Le nom de cette variable est explicite, elle représente le poids (en grammes) du rein que l'on greffe sur le patient.

- **Le poids du receveur :**

C'est le poids (en kilogramme) du receveur au moment de la greffe.

- **L'âge du receveur :**

Il s'agit de l'âge du receveur (en années) au moment de la greffe.

- **L'âge du donneur :**

Il s'agit de l'âge du donneur (en années) au moment de la greffe.

- **L'ischémie froide :**

Elle représente le temps (en minutes) entre le moment où le rein a été prélevé sur le donneur et le moment où il a été greffé sur le receveur.

- **DGF :**

DGF signifie "**Delayed Graft Function**" Elle représente le délai après la greffe que met le rein pour redémarrer. En effet, après une greffe le rein n'est que très rarement directement opérationnel et souvent il faut attendre plusieurs jours avant que le rein ne se remette en marche. On peut noter que cette variable est très fortement corrélée avec le nombre de dialyses du receveur après la greffe.

- **Le numéro de la greffe :**

Ici, on a cherché à savoir si la greffe étudiée est la première du receveur ou si il a déjà subi une transplantation qui a échoué. Etant donné le nombre important de première greffe dans notre base,

nous avons séparé les individus dont cette greffe était la première des autres individus qui, on peut le supposer, présentent plus de risque.

– **Les incompatibilités HLA :**

HLA est un sigle qui signifie “**H**uman **L**euco**C**yte **A**ntigen**S**”. Les HLA permettent la reconnaissance du soi et du non-soi.

Il est établi par les experts que plus le nombre d’incompatibilités HLA est important, plus la transplantation a des chances d’échouer. On peut noter que l’on ne greffe pas des personnes dont le nombre d’incompatibilités HLA est supérieur à 6 à cause du trop grand risque de rejet. Dans notre base, le nombre d’incompatibilités est compris entre 0 et 6.

– **PRA :**

Panel reactive antibody (anticorps réactifs contre les échantillons du panel) – Mesure du degré de sensibilisation d’une personne (exposition à des molécules de HLA étrangères et formation d’anticorps contre ces molécules) aux différentes molécules de HLA présentes dans la population générale. Plus le pourcentage de PRA est élevé, plus le degré de sensibilisation est important, ce qui réduit la probabilité qu’un organe de donneur décédé soit acceptable (résultat négatif à l’épreuve de compatibilité croisée).

– **Créatinine du donneur :**

C’est la concentration de la créatine du donneur dans le sang, celle-ci dépend de la capacité d’élimination du rein et de la masse musculaire. Son évaluation permet d’apprécier un dysfonctionnement de la filtration rénale.

– **Les centres de greffes :**

Elle représente le lieu où la greffe a été effectuée.

– **Le ratio poids du rein / poids du receveur :**

C’est le ratio entre les poids du rein greffé (en gramme) et le poids du receveur (en Kilogramme) au moment de la greffe.

– **Le rejet aigu :**

Le rejet aigu, qui a lieu le plus souvent au cours de la première année post transplantation, se traduit par une diminution soudaine de la fonction rénale. L’organisme du receveur n’accepte pas le greffon et le rejette donc. Il est parfaitement réversible lorsqu’il est détecté et traité à temps. Pour le traiter, on administre au receveur des immunosuppresseurs qui affaiblissent les défenses de l’organisme.

C’est une variable temps-dépendante, car la date du rejet aigu apporte de l’information à notre modèle.

4.3 méthodes

4.3.1 Caractérisation du poids du rein

Le but de notre étude est de chercher et de caractériser l'influence du poids du rein sur la survie des patients transplantés rénaux. Pour cela, différentes façon de prendre en compte le poids du rein se sont présentées à nous : le poids du rein tel-quel, le ratio poids du rein/poids du receveur, le ratio poids du rein/surface corporel du receveur. Pour choisir parmi ces possibilités nous avons effectué des tests du log-rank sur chacune de variables. Le ratio poids du rein/poids du receveur ayant la plus forte influence sur la survie en univarié, nous avons choisi de caractériser le poids du rein par cette variable.

4.3.2 Choix des valeurs seuil

Nous avons vu dans la partie théorique que si les variables continues ne sont pas log-linéaire, on ne peut pas les inclure dans le modèle tel-quel, il est nécessaire de les discrétiser en plusieurs groupes. Dans certain cas, le choix de la valeur seuil ne revient pas au statisticien mais à l'expert médical, en effet, plusieurs choses sont importante dans ce choix, l'interprétabilité du seuil, l'homogénéité des seuils dans la littérature nécessaire à une comparabilité des résultats, et enfin le bien fondé statistique de ce seuil.

Nous allons voir ici comment choisir la valeur de passage entre deux groupes lorsque qu'il n'y a pas de valeur communément admise.

Notre but est de trouver les groupes les plus différent du point de vue de la survie. Nous avons donc procédé à plusieurs tests du log-rank pour chaque variables en faisant varié la valeur seuil, puis nous avons choisi le seuil interprétable le plus proche de la valeur seuil qui maximise la statistique de test (et donc la vraisemblance). Bien entendu, il faut veiller à ne pas avoir un groupe avec trop peu d'effectif ce qui nuirait à l'analyse.

4.3.3 Validation des modèles

Nous avons choisi de ne retenir pour l'analyse multivariée que les variables ayant une p-value inférieur à 0,2 au test du log-rank. La meilleure solution consiterait à inclure toutes les covariables dans le modèle de Cox pour ensuite écarter celles qui n'ont aucune influence sur la survie ; mais la complexité du modèle multivarié serait trop importante pour pouvoir estimer tous les paramètres. On ne fait donc, grâce au test du log-rank, qu'un premier tri de nos covariables.

Nous avons ensuite enlevé une à une de l'analyse multivarié toutes les covariables n'ayant pas une p-value inférieure à 0,05. Nous avons aussi testé quelques interactions qui, de l'avis des medecins, peuvent apporter de l'information à notre modèle.

4.4 statistiques descriptives

L'échantillon se compose de 659 (62%) receveurs hommes et 401 (38%) receveurs femmes, les donneurs sont aussi en majorité des hommes : 719 (68%) contre 339 femmes (32%). La moyenne d'âge des receveurs est de $45,6 \pm 13,13$ ans ¹, la moyenne d'âge des donneurs est de $39,78 \pm 15,29$ ans. La moyenne du temps d'ischémie froide est de 24,4 heures \pm 9,5 heures. La majorité des patients transplantés de l'échantillon en sont à leur première greffe de rein (938, soit 88,4%). Le ratio poids du rein/poids du receveur a pour moyenne $3,19 \pm 1,13$ g.Kg⁻¹. Le rein met en moyenne $5,40 \pm 6,96$ jours pour redemarrer, la médiane quand à elle est de 2 jours. La grande majorité des personnes greffées n'ont pas connu d'épisode de rejet aigu (786, soit 74%). 270 patients de notre base ont subi l'évènement (retour en dialyse ou mort) parmi lesquels 188 ont eu un retour en dialyse et 82 sont morts.

¹moyenne \pm écarts-types

Avant de déterminer le type d'analyse univarié à faire, nous avons vérifié les hypothèses de log-linéarité et de proportionnalité. L'âge du receveur c'est avéré être la seule variable à la fois log-linéaire et proportionnelle, nous avons donc pu l'intégrer en tant que variable continue dans le modèle de Cox ce qui nous a contraint à utiliser un modèle de Cox univarié et non un test du log-rank pour tester son impact sur la survie. Les experts n'ayant pas de seuil défini pour la créatinine du donneur nous avons dû la calculer comme expliqué dans la section 4.3.2 page 17. Nous avons trouvé un seuil de $120 \mu\text{mol/l}$. En ce qui concerne les âges du donneur et du receveur, le seuil communément admis est de 55 ans, afin d'assurer une homogénéité et une comparabilité dans les articles nous avons repris ce seuil. Pour l'ischémie froide, nous avons trouvé un seuil statistique de 2160 minutes soit 36 heures. Pour la DGF, un seuil de 6 jours nous a été recommandé. Pour le nombre de greffes et le PRA, un seuil s'est imposé à nous au vu de la distribution des valeurs (une majorité de patients avec une première greffe ainsi qu'une PRA=0), nos seuils sont donc : première greffe/plus d'une greffe, PRA=0/PRA \neq 0. Pour la variable ratio, nous avons choisi de faire deux groupes : ratio inférieur à $2,3 \text{ g.Kg}^{-1}$ (groupe 0) et ratio supérieur ou égal à $2,3 \text{ g.Kg}^{-1}$ (groupe 1). Dans notre base, le nombre d'incompatibilités est compris entre 0 et 6. Nous avons choisi deux modalités pour cette variable : incompatibilité inférieure à 5 (groupe 0), incompatibilité supérieure ou égale à 5 (groupe 1). Le tableau 4.1 nous montre la distribution de l'échantillon par rapport aux deux modalités du ratio. Un test du χ^2 est effectué afin de voir si cette distribution est, ou non, similaire dans chaque groupe. On peut voir que les seules variables où la distribution n'est pas similaire au seuil $\alpha = 0,05$ sont les sexes du receveur et du donneur ainsi que le taux de créatinine du donneur.

	ratio supérieur à		ratio inférieur à		χ^2	P-value
	Modalité	Effectif	Effectif			
Sexe du receveur	Femmes	347	54	22,26	< 0,0001	
	Hommes	490	169			
Sexe du donneur	Femmes	227	112	42,90	< 0,0001	
	Hommes	608	111			
Age du receveur	< 55 ans	614	153	1,98	0,1590	
	\geq 55 ans	223	70			
Age du donneur	< 55 ans	680	181	0,02	0,8732	
	\geq 55 ans	153	42			
Créatine du donneur	<120	643	183	4,68	0,0304	
	\geq 120	181	33			
Ischémie froide	<36h	697	179	1,63	0,2019	
	\geq 36h	134	44			
DGF	\leq 6j	580	163	1,16	0,2815	
	>6j	240	56			
nb de greffes	1	736	202	1,21	0,2705	
	>1	101	21			
nb d'incompatibilités HLA	<5	721	189	0,58	0,4480	
	\geq 5	100	31			
PRA	0	632	177	1,21	0,2722	
	autre	197	45			
rejet aigu	0	615	167	0,18	0,6703	
	1 ou plus	222	56			

TAB. 4.1 – tableau des statistiques descriptives

4.5 Résultats

4.5.1 Tests du Log-Rank

Nous allons présenter dans cette partie les résultats du test du log-Rank de nos covariables sous forme d'un tableau unique 4.2. Vous pouvez voir les courbes de survie de Kaplan-Meier ainsi que les résultats des tests de proportionnalité et de log-linéarité en annexe.

	Modalité	Effectif	Nombre d'événements observés	χ^2	P-value
Sexe du receveur	Femmes	401	103	0,2	0,673
	Hommes	659	167		
Sexe du donneur	Femmes	339	94	1,2	0,265
	Hommes	719	175		
Créatinine du donneur	<120	826	201	6,0	0,0143
	≥120	214	65		
Ischémie froide	<36h	876	206	9,9	0,0017
	≥36h	178	64		
DGF	≤6j	743	151	22,1	<0,0001
	>6j	296	109		
nb de greffes	1	938	227	5,0	0,0256
	>1	122	43		
nb d'incompatibilités HLA	<5	910	224	1,6	0,201
	≥ 5	131	39		
PRA	0	809	191	3,7	0,054
	autre	242	79		
ratio	<2.3	223	70	3,8	0,0505
	≥ 2.3	837	200		

TAB. 4.2 – tableau récapitulatif des tests du Log Rank pour la survie

Les covariables en gras sur ce tableau sont celles retenues pour l'analyse multivariée. Nous pouvons constater sur ce tableau que seules deux covariables n'ont pas été retenues pour l'analyse multivariée sexe du receveur et sexe du donneur avec des p-values respectives de 0,673 et 0,265.

On constate aussi que la variable qui a la plus forte influence sur la survie dans ce modèle est la DGF avec une p-value inférieure à 0,0001. Le ratio poids du rein/du poids du receveur semble avoir une influence sur la survie, en effet, sa p-value est d'environ 0,05. La PRA à une p-value de 0,054 ce qui indique une influence presque significative au seuil 0,05. Nous avons choisi de retenir les incompatibilités HLA qui ont une p-value de 0,201 et donc peut être une influence sur la survie non négligeable en multivarié. L'ischémie froide à quand à elle une p-value de 0,017 et la créatinine du donneur une de 0,0143.

Comme vous pouvez le remarquer la variable "âge du receveur" ne fait pas partie de ce tableau ; ceci s'explique par le fait que nous avons choisi de l'inclure en tant que variable continue, un test du log rank n'est donc pas adapté. Nous avons ainsi construit pour cette variable un modèle de Cox univarié pour tester son influence sur la survie : tableau 4.3. Un modèle de Cox univarié est aussi nécessaire pour la variable rejet aigus celle-ci étant temps-dépendante les résultats se trouve dans le tableau 4.4.

	Coefficient	Risque Relatif	IC à 95 %	P-value
âge du receveur	0,189	1,21	[1,1 ;1,33]	< 0,0001

TAB. 4.3 – Modèle de Cox univarié sur l'âge du receveur

	Coefficient	Risque Relatif	IC à 95 %	P-value
rejet aigus	0,537	1,71	[1,33 ;2,21]	< 0,0001

TAB. 4.4 – Modèle de Cox univarié sur le rejet aigus

L'influence sur la survie de l'âge du receveur n'est, comme on peut le constater sur le tableau 4.3, pas négligeable. Pour deux groupes de patients avec 10 ans d'écart (les valeurs de la variables ont été divisé par 10) le groupe de patient le plus vieux aura 1,21 fois plus de chance de retourner en dialyse ou de mourir. Avec une p-value<0,0001, cette covariable est retenue pour l'analyse multivariée. Le tableau 4.4 nous informe quand à lui sur l'influence du rejet aigus sur la survie. On constate que pour deux groupes d'individus l'un ayant subit un rejet aigus et l'autre non, le premier groupe a 1,71 fois plus de chance de perdre son greffon ou de mourir que l'autre. Une p-value<0,0001 nous permet de retenir cette variable pour l'analyse multivarié.

4.5.2 L'analyse multivariée, le modèle de Cox

	Coefficient	Risque Relatif	IC à 95 %	P-value
ratio poids du rein/poids du receveur	-0,299	0,741	[0,555 ;0,99]	0,0430
rejet aigu	0,638	1,893	[1,422 ;2,52]	< 0,0001
âge du receveur	0,173	1,189	[1,071 ;1,32]	0,0012
ischémie froide	0,309	1,363	[1,001 ;1,86]	0,0500
creatinine du donneur	0,311	1,365	[1,015 ;1,83]	0,0390
DGF	0,614	1,85	[1,232 ;2,14]	0,0006
nombre de greffe	0,672	1,959	[1,360 ;2,82]	0,0003

TAB. 4.5 – Modèle de Cox multivarié avec le rejet aigus

Le tableau 4.5 montre les variables ayant une influence significative sur la survie au seuil $\alpha = 0,05$. On peut voir sur celui-ci que si le ratio poids du rein/poids du receveur est supérieur à $2,3g.Kg^{-1}$, alors le risque de mourir ou de retourner en dialyse est environ 1,35 (1/0,74) fois plus élevé que si le ratio été en dessous de $2,3g.Kg^{-1}$, sa p-value (0,0430) nous indique que son influence est significative. Le rejet aigus est comme l'on pouvait s'y attendre est une variable incontournable, en effet, le fait de faire au moins un rejet aigus au cours de la greffe multiplie par 1,93 le risque de retour en dialyse ou de mourir. Une fonction renal qui redémarre tardivement (après les 6 premiers jours de la greffe) est mauvais signe pour la tenu du greffon, on peut voir que cela multiplie par 1,85 le risque de perdre celui-ci. Un patient ayant déjà perdu un greffon lors d'un greffe précédente a presque 2 fois plus de risque d'échec lors de cette greffe qu'un patient qui n'a subit aucune autre transplantation de rein. L'âge du receveur est lui aussi important pour la survie du greffon, en effet,

le risque d'échec de la greffe est multiplié par environ 1,2 pour une augmentation de 10 de l'âge, par exemple, pour deux groupes de patients identiques ayant 30 ans de différence, le groupe le plus âgé aura : $1,2 \times 1,2 \times 1,2 = 1,728$ fois plus de chance de perdre le greffon que le groupe le plus jeune. La créatinine du donneur a elle aussi un influence sur la survie, en effet une créatinine trop forte (supérieur à $120\mu\text{ mol/l}$) réduit les chance de survie du greffon, c'est tout à fait normal étant données que la créatinine est un facteur de bonne santé du rein.

4.5.3 discussion des résultats

Nos résultats sont parfaitement en accord avec ce que l'on pouvait attendre d'une telle étude, les variables restante dans le modèle de Cox sont bien les variables les plus importantes d'un point de vue médical.

Néanmoins, en regardant de plus près les courbes de kaplan-Meier en annexe figure A.1 page 26 ainsi que le test de proportionnalité figure A.2 page 26, nous pouvons constater que l'effet du ratio n'est pas constant au cours du temps, ce qui n'est pas pris en compte dans le tableau 4.5.

En effet, il semble presque nul entre 0 et 24 mois puis un effet apparait clairement après 24 mois. On peut donc supposer que notre modèle a sous-estimé l'effet du ratio en "confondant" ces deux périodes. Pour pallier à cela, nous allons séparer ces deux périodes dans l'analyse. Pour ce faire, nous avons procédé comme nous l'avons vu section 3.5.7 page 12.

4.5.4 Modèle de Cox corrigé

	Coefficient	Risque Relatif	IC à 95 %	P-value
ratio1	-0,0234	0,977	[0,569 ;1,676]	0,9300
ratio2	-0,4213	0,656	[0,466 ;0,924]	0,0160
rejet aigu	0,6410	1,898	[1,426 ;2,528]	< 0,0001
âge du rece- veur	0,1743	1,190	[1,072 ;1,322]	0,0012
creatinine du donneur	0,3098	1,363	[1,014 ;1,832]	0,0400
DGF	0,4849	1,624	[1,231,2,143]	0,0006
nombre de greffes	0,6784	1,971	[1,368,2,840]	0,0003
ischémie froide	0,3077	1,360	[0,999 ;1,853]	0,0510

TAB. 4.6 – Modèle de Cox multivarié final

Ici, nous avons testé l'interaction entre le ratio et l'ischémie froide, le test de Wald ne pouvant nous donner des résultats (étant donné que nous testons en fait 2 interactions à la fois : ratio1 et ratio2) nous avons dû utiliser le test du rapport de vraisemblance. Nous en avons conclu que l'interaction n'est pas significative, c'est pourquoi elle n'est pas présente dans le modèle ci-dessus. On voit sur le tableau 4.6 que l'effet du ratio 1 et du ratio2 est conforme à ce que l'on pouvait attendre au vu des courbes de Kaplan-Meier : l'effet du ratio1 (c'est à dire l'effet du ratio entre 0 et 24 mois) est quasiment nul (RR=0.977) contrairement à l'effet du ratio2 (l'effet du ratio après 24 mois) qui lui est supérieur à l'effet du ratio dans le modèle précédent (RR=0,656), le groupe de patient avec un ratio inférieur à $2,3g.Kg^{-1}$ a donc $1/0,656=1,524$ fois plus chance de perdre son greffon que l'autre groupe. Par ailleurs, les autres variables ne sont pas affectées par le changement dans la façon de prendre en compte le poids du rein et l'effet reste donc quasiment le même.

4.5.5 discussion des résultats corrigé

Au vu des résultats, on peut supposer que quel que soit le ratio poids du rein/poids du receveur la greffe à autant de chance de prendre, mais à partir de deux 2 ans les petits reins commencent à peiner s'ils sont greffé sur un "gros" patient, le rein ne peut plus assuré sa fonction se qui entraine le retour en dialyse ou la mort du receveur. Les résultats confirmant un fort effet de poids du rein sur la survie, on peut espérer que la même étude effectué sur des cohortes de patients différents apporte des résultats similaire. Si tel est le cas, il n'est pas à exclure que le poids du rein (ou plus exactement le ratio poids du rein/poids du receveur) face un jour partie des facteurs utilisés pour l'attribution des reins.

Chapitre 5

L'influence du poids du rein sur l'apparition de protéinurie

5.1 Problématique

Une protéinurie est définie comme l'élimination pathologique dans les urines d'une quantité de protéines supérieure à 150 mg (soit 0,15 g) par 24 heures. Dans cette étude nous allons rechercher chez les patients ceux qui présente une forte protéinurie (0,50g/24h). L'événement étudié n'est plus ici la mort ou le retour en dialyse de l'individus mais le passage au dessus d'un certain seuil(0,50g/24h) de Protéinurie.

Ici, il faut bien avoir à l'esprit que nous sommes face à des données censurées par intervalle, en effet le taux de protéine éliminer dans les urines n'est relever qu'as intervalles régulier de temps. Lorsque nous voyons l'apparition d'un taux supérieur à 0,50g/24h, notre seul certitude est que le passage au-dessus de ce seuil est apparu entre le dernier relever et celui-ci.

Le modèle de Cox ne peut pas normalement pas faire face à ce genre de censure. Un biais apparaîtra alors dans nos résultats. Ce biais peut cependant être considéré comme négligeable.

Les résultats des test du Log-Rank sont notés dans le tableau 5.1 :

5.2 résultats

5.2.1 modèles univariés

Nous pouvons constater sur le tableau 5.1 que bien que une forte protéinurie soit un indice d'un disfonctionnement du greffon, les covariables qui influent sur la protéinurie ne sont pas les mêmes que les variables qui influent sur la survie du greffon. On voit sur le tableau ci-dessus que le sexe du receveur, qui n'avait pas d'influence sur la survie du greffon, a maintenant une influence non négligeable (p-value=0,0033) sur le taux de protéinurie. Le nombre de greffes du patient n'a pas d'effet sur le taux de protéinurie en univarié (p-value=0,274) alors que cette covariable était une des covariables les plus influentes pour la survie des greffons. La PRA avec une p-value de 0,0189 influt sur la protéinurie. La créatinine du donneur n'a vraisemblablement rien à voir avec le taux de protéinurie post greffe avec une p-value à 0,635 cette variable est la moins influente de toutes celles testées. Le nombre d'incompatibilités HLA et de greffe n'ont pas d'influence significative au seuil $\alpha=0,2$ sur le taux de protéinurie.

	Modalité	Effectif	Nombre d'événements observés	χ^2	P-value
Sexe du receveur	Femmes	401	166	8,6	0,0033
	Hommes	659	317		
Sexe du donneur	Femmes	339	164	1,2	0,281
	Hommes	719	319		
âge du donneur	<55 ans	861	383	7,2	0,00709
	≥ 55 ans	195	100		
âge du receveur	<55 ans	767	351	0,7	0,399
	≥ 55 ans	293	132		
Créatinine du donneur	<120	826	376	0,2	0,635
	≥ 120	214	99		
Ischémie froide	<36h	876	389	3,2	0,0716
	≥ 36 h	178	90		
DGF	≤ 6 j	743	301	32,7	<0,0001
	>6j	296	173		
nb de greffes	1	938	422	1,2	0,274
	>1	122	61		
nb d'incompatibilités HLA	<5	910	412	1	0,323
	≥ 5	131	65		
PRA	0	809	351	5,5	0,0189
	autre	242	126		
ratio	<2.3	223	122	12,2	0,0005
	≥ 2.3	837	361		

TAB. 5.1 – Tests du log-rank pour l'événement protéinurie

5.2.2 Modèle de Cox : la protéinurie

On peut constater sur le tableau 5.2 que le nombre de variables influant le taux de protéinurie est bien plus faible que le nombre de variable influant sur la survie. On peut voir aussi que le sexe fait maintenant partie des covariables significativement influante avec 1,23 fois plus de risque de dépasser le seuil de 0,5g/24h de protéinurie pour les hommes que pour les femmes. Si la fonction rénale ne reprend qu'après 6 jours, ce taux est 1,49 fois plus élevé que pour les patient dont la fonction rénale à repris avant cette date. L'âge du donneur (et donc l'âge du rein) tient une place importante dans l'apparition d'une protéinurie, en effet le risque d'apparition de celle-ci est multiplié par 1,34 par tous les 10 ans. On voit aussi que avoir fait un rejet aigu est un facteur aggravant de la protéinurie avec un risque multiplié par 1,51. Nous constatons qu'un ratio élevé est toujours un facteur protecteur d'une forte protéinurie avec $1/0,74=1,52$ fois plus de chance d'éviter celle-ci pour les gros ratios (supérieur à 2,3 g.Kg⁻¹).

	Coefficient	Risque Relatif	IC à 95 %	P-value
ratio poids du rein/poids du receveur	-0,300	0,74	[0,597 ;0,918]	0,0062
rejet aigu	0,409	1,51	[1,195 ;1,895]	0,0005
âge du donneur	0,292	1,34	[1,067 ;1,682]	0,012
DGF	0,402	1,49	[1,222 ;1,828]	< 0,0001
sexe du receveur	0,207	1,23	[1,013 ;1,494]	0,036

TAB. 5.2 – Modèle de Cox pour l'événement protéinurie

5.2.3 discussion des résultats

Nous allons ici rassembler les résultats des deux études précédentes. On a vu que le ratio poids du rein/poids du receveur à une influence sur la survie après 24 mois (p-value=0,0160), mais aussi une influence sur une forte protéinurie (p-value=0,0062). La cohérence de ces résultats nous incite à penser que le ratio influt sur la survie du greffon via la protéinurie.

Annexe A

Hypothèses de proportionnalité et de log-linéarité

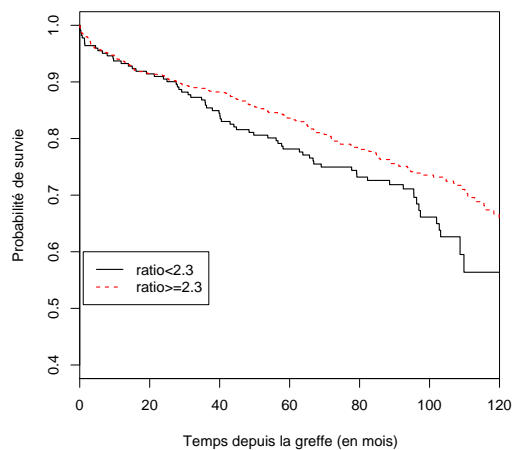


FIG. A.1 – Survie de notre échantillon (n=1060) selon le ratio

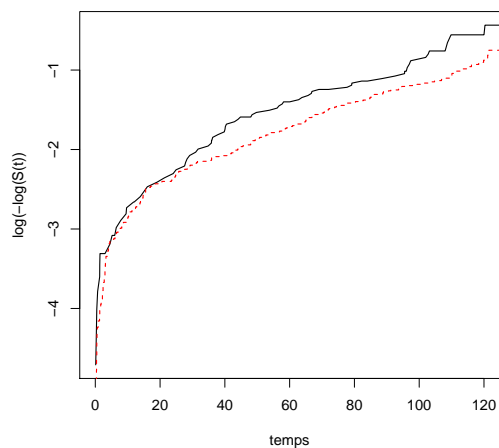


FIG. A.2 – Test de Proportionnalité pour la variable ratio