

UNIVERSITE DE LIMOGES
ECOLE DOCTORALE Biologie – Santé
Faculté de Médecine
INSERM UMR-S850

Thèse N° []

Thèse

pour obtenir le grade de

Docteur de l'Université de Limoges

Discipline/S spécialité : Pharmacologie

Présentée et soutenue publiquement par

Zeinab DAHER ABDI

Le 17 Décembre 2013

Modélisation pharmacocinétique et pharmacodynamique de population. Analyse des facteurs prédictifs de réponse aux médicaments immunosuppresseurs administrés en transplantation rénale

Thèse dirigée par le Professeur Annick ROUSSEAU et le Docteur Aurélie PRÉMAUD

JURY

Monsieur le Professeur Etienne CHATELUT : Rapporteur

Monsieur le Professeur Luc FRIMAT : Rapporteur

Madame le Professeur Marie ESSIG : Examineur

Monsieur le Professeur Eric BELLISSANT : Examineur

Madame le Professeur Annick ROUSSEAU : Directeur de thèse

Madame le Docteur Aurélie PRÉMAUD : Co-directeur de thèse

Monsieur le Professeur Pierre MARQUET : Membre invité

Madame le Docteur Mireille DROUET : Membre invité

Remerciements

A Madame le Professeur Annick Rousseau

Un proverbe djiboutien dit « *je dois une vie entière à celui qui m'a enseigné un seul mot* ». Merci pour toutes les notions statistiques que vous m'avez enseignées au cours de ces 3 ans de recherche. J'ai beaucoup appris auprès de vous et vos conseils et surtout vos qualités humaines m'ont permis d'effectuer ce travail dans un cadre idéal. Merci de m'avoir fait confiance et encouragée pendant les moments durs. Merci de m'avoir permis de participer à de nombreux congrès internationaux. Ce fut un grand plaisir d'avoir travaillé sous votre direction. Merci pour tout.

A Madame le Docteur Aurélie Prémaud

Un grand merci pour avoir accepté de co-diriger ce travail. Ton esprit bienveillant, ta rigueur scientifique et ton professionnalisme m'ont permis de mener à bien ce projet. Je te suis profondément reconnaissante pour tes commentaires pertinents lors de la relecture attentive des différents manuscrits, et pour ton soutien et ta patience au cours des derniers mois de thèse qui étaient particulièrement dures.

A Monsieur le Professeur Etienne Chatelut

Je suis très honorée que vous ayez accepté d'être rapporteur de mon travail de thèse et je vous en remercie. Je profite aussi de cette occasion pour vous témoigner ma gratitude et ma profonde reconnaissance pour votre soutien et vos conseils au cours de mon cursus universitaire à la faculté de pharmacie de Toulouse.

A Monsieur le Professeur Luc Frimat

Je vous remercie de l'intérêt que vous aviez porté à ce travail en acceptant d'être rapporteur. C'est un très grand honneur de vous compter parmi les membres du jury.

A Monsieur le Professeur Eric Bellissant

Je vous suis très reconnaissante et vous remercie d'avoir accepté d'examiner mon travail de thèse. Je suis très sensible à votre présence dans ce jury de thèse.

A Madame le Professeur Marie Essig

Merci pour l'aide et les précieux conseils que vous nous avez apporté tout au long de ce projet de thèse. Vous étiez la pierre angulaire de ce travail et votre point de vue clinique d'expert nous a permis de mieux interpréter les résultats obtenus avec les modèles statistiques. Merci aussi d'avoir accepté d'être examinateur de cette thèse.

A Monsieur le Professeur Pierre Marquet

Merci pour la confiance que vous m'avez accordée en m'accueillant dans votre équipe de recherche. Merci aussi pour l'honneur que vous me faites en participant à mon jury de thèse. Soyez assuré de ma profonde reconnaissance et de tout mon respect.

A Madame le Docteur Mireille Drouet et Monsieur Matthieu Filloux

Un grand merci pour votre collaboration qui a été constructive et essentielle pour faire avancer ce travail. Les données immunologiques étaient le maillon fort de ce projet, et vos suggestions et vos conseils pertinents sur le sujet nous ont permis de mener à bien ce travail. Merci aussi Mireille d'avoir accepté d'être membre invité dans mon jury de thèse, j'en suis très honorée.

A Madame le Professeur Sophie Alain et Madame Françoise Garnier

Merci pour votre contribution efficace et dynamique dans certaines parties des travaux de cette thèse. C'était un réel plaisir d'avoir travaillé avec vous.

A Madame le Professeur Nicole Tubiana-Mathieu, Madame le Docteur Sandrine Lavau-Denes et Madame Sophie Leobon

Merci de m'avoir permis d'élargir mon sujet de thèse sur un autre domaine très enrichissant qu'est l'oncogériatrie. Ce fut un réel plaisir d'avoir collaboré avec vous.

A tout le personnel de la Recherche clinique en Néphrologie et en Pharmacologie

Un merci particulier à Eliza, merci pour ta disponibilité et ta grande sympathie à chaque fois que je venais te solliciter pour les bases de données. Merci également à Hélène et Claire, vous étiez toujours d'une grande volonté pour m'aider à chercher des anciennes bases de données. Merci aussi à Fabrice, Alexandre et Séverine.

A l'ensemble des membres de l'unité UMR Inserm S-850

Merci à tous les étudiants pour l'ambiance et la bonne humeur qui rend si agréable le travail au labo. Un grand merci à Lynn et Karen pour les corrections d'anglais des manuscrits. Merci aussi à Patricia, François, Jean Baptiste, Nicolas, Julien, Fred, et Frank pour l'aide que vous m'avez apportée de près ou de loin au cours de ma thèse.

Un grand merci venant du cœur à Jean Hervé, toujours bienveillant à l'égard de tout le monde. Merci pour la chaleur humaine que vous dégagez et votre sincérité, des qualités rares de nos jours. Soyez assuré de toute mon estime.

A ma chère famille

Merci de m'avoir soutenue et encouragée toutes ces longues années d'étude. J'arrive enfin au bout du tunnel et je suis ravie d'être parmi vous très bientôt !

A mes chers parents, mes frères et sœurs, mes amis

Résumé

Ce travail porte sur des études de modélisation menées en transplantation rénale et en oncologie. L'effet de l'exposition longitudinale de l'acide mycophénolique (AMP) sur la survie sans rejet aigu du greffon rénal au cours de la 1^{ère} année post-greffe a été démontré dans un modèle joint développé par la méthode des modèles linéaires à effets mixtes. Ce premier modèle a permis de proposer des cibles d'exposition variant avec le délai post-greffe, pour le suivi thérapeutique du mycophénolate mofétil, prodrogue de l'AMP. Par la méthode des modèles non-linéaires à effets mixtes, un modèle joint plus complexe, prenant en compte l'exposition longitudinale aux médicaments immunosuppresseurs co-administrés a confirmé l'effet de l'exposition à l'AMP sur le devenir du patient greffé. Le rôle des infections et maladies à cytomégalovirus sur le devenir du patient greffé a aussi pu être démontré dans ce modèle. Une analyse de survie du greffon par arbre de régression et de classification a révélé l'existence de sept profils de patients dont la survie du greffon à 10 ans variait de 20 à 95%. La valeur de créatinine sérique et la protéinurie 12 mois après la greffe, le profil d'évolution de la créatinine sérique dans la 1^{ère} année post-greffe, le développement d'anticorps anti-HLA spécifiques du donneur, le rejet aigu, la présence d'anticorps anti-HLA avant la greffe chez les patients retransplantés étaient les principaux facteurs associés à une réduction de la survie du greffon. Un modèle pharmacocinétique de population de la capécitabine a été développé afin de quantifier l'influence du grand âge (≥ 75 ans) sur l'exposition à la capécitabine et à ses métabolites.

Mots clés : transplantation rénale, analyse de survie, immunosuppresseurs, rejet aigu, modèles joints, exposition longitudinale

Table des matières

CHAPITRE 1 : Introduction.....	8
I. <i>Modèles prédictifs d'évènements uniques ou répétés.....</i>	8
I.1. Analyses de survie	8
I.1.a. Méthode semi-paramétrique : modèle de Cox.....	9
I.1.b. Méthodes paramétriques	12
I.2. Régressions logistiques.....	15
I.2.a. Cas d'évènements uniques	15
I.1.a. Cas d'évènements répétés : modèles utilisant les équations d'estimation généralisées et modèles à effets mixtes.....	16
I.3. Modèles multi-états	19
II. <i>Méthodes pour prendre en compte une variable quantitative dépendante du temps.....</i>	20
II.1. Méthode non paramétrique k-means pour données longitudinales	20
II.2. Modèles longitudinaux paramétriques.....	22
II.2.a. Modèles utilisant les équations d'estimation généralisées.....	23
II.2.b. Modèles à effets mixtes	24
III. <i>Modèles conjoints</i>	24
IV. <i>Objectifs de la thèse.....</i>	27
CHAPITRE 2 : Analyse de l'influence de l'exposition aux immunosuppresseurs sur la survenue de différents évènements en greffe rénale.....	29
I. <i>Analyse de l'influence de l'exposition longitudinale à l'acide mycophénolique sur la survenue du rejet aigu.....</i>	29
II. <i>Analyse de l'influence de l'exposition longitudinale à l'acide mycophénolique et aux anticalcineurines sur un critère composite d'efficacité et sur les infections et maladies à cytomégalovirus.....</i>	41
III. <i>Analyse de l'influence de l'exposition longitudinale à l'acide mycophénolique sur la survenue d'anémie</i>	69
III.1. Contexte.....	69
III.2. Patients et méthodes	69
III.3. Résultats.....	71
III.4. Discussion.....	72
CHAPITRE 3 : Analyse de la survie du greffon rénal : étude sur une large cohorte de patients transplantés rénaux suivis à Limoges	73
CHAPITRE 4 : Etude de pharmacocinétique et de pharmacodynamie de la capécitabine chez des patients âgés de plus de 75 ans	98
CHAPITRE 5 : Discussion et conclusion	120
Références bibliographiques	131
Annexes	136

Glossaire

ACN:	anticalcineurine
AMP:	acide mycophénolique
AUC:	area under the curve
CMV:	cytomégalovirus
CsA:	cyclosporine A
5'-DFCR:	5'-deoxy-5-fluorocytidine
5'-DFUR:	5'-deoxy-5-fluorouridine
DSA	donor specific antibodies
FBAL:	alpha-fluoro-beta-alanine
5-FU:	5-fluorouracile
GEE:	generalized estimating equations
GLMM:	generalized linear mixed-effects models
HLA:	human leucocyte antigen
IS:	immunossuppresseurs
KmL:	k-means for longitudinal data
OR:	odds ratio
MMF:	mycophénolate mofétil
PK:	pharmacocinétique
RSF:	random survival forest
TAC:	tacrolimus

CHAPITRE 1 : Introduction

Cette première partie d'introduction a pour objectif de présenter brièvement le principe des différentes méthodes statistiques et de modélisations utilisées dans cette thèse.

I. Modèles prédictifs d'évènements uniques ou répétés

En épidémiologie, différents types de modèles permettent d'analyser la survenue des évènements étudiés. Parmi eux, on distingue les modèles de survie, les modèles de régressions logistiques et les modèles multi-états.

I.1. Analyse de survie

L'analyse de survie est une branche des statistiques qui s'intéresse à l'étude du délai de survenue d'un évènement donné. Initialement, cette notion a été employée dans l'analyse de la fiabilité en ingénierie. Puis, dans la moitié du XX^{ème} siècle, elle connaît son utilisation en médecine, d'abord en oncologie, où l'évènement étudié était le décès, d'où l'emploi du terme « survie ». En épidémiologie, cette méthode est particulièrement adaptée dans le contexte des études longitudinales comme les enquêtes de cohorte (suivi des patients dans le temps) ou les essais thérapeutiques (évaluation de l'efficacité d'un médicament).

Dans le domaine médical, l'évènement étudié peut être de nature diverse : au-delà du décès, il peut s'agir, par exemple, de l'apparition d'une maladie, de rechute, de complications, d'un évènement indésirable, ou encore de guérison.

Une des principales propriétés caractérisant les données de survie est le fait qu'elles sont partiellement observées : données censurées. Il existe différents types de censure : censure à gauche, censure à droite et censure par intervalles. La censure à gauche concerne les sujets dont l'évènement s'est déjà produit avant la première observation ; ces derniers sont en général exclus de l'analyse. La censure à droite concerne les sujets dont l'évènement n'est pas encore survenu pendant la durée d'observation ou les perdus de vue. Enfin, la censure par intervalles intervient quand les temps exacts de survenue de l'évènement sont inconnus ; la seule information disponible est que les évènements se sont produits pendant un intervalle de temps t_i et t_{i+1} (COMMENGES, 2002). Les censures doivent être aléatoires, cela signifie que

les raisons de ces censures doivent être indépendantes de l'évènement étudié. Cette condition est fondamentale pour que les censures ne biaisent pas les résultats.

Les données de survie peuvent être décrites par différentes fonctions mathématiques dont les 3 principales sont la fonction de survie, la fonction de risque instantané et la fonction de risque cumulé.

La fonction de survie désigne la probabilité de survivre au moins jusqu'à la date t , et est définie par l'équation suivante :

$$S(t) = P(T > t). \quad [\text{Eq.1}]$$

La fonction de risque instantané (appelé « *hazard function* » en anglais) désigne la probabilité que l'évènement survienne entre t et $t+dt$, avec $dt \rightarrow 0$, sachant qu'il n'a pas eu lieu jusqu'à l'instant t . Elle est définie par l'équation suivante :

$$h(t) = \lim_{dt \rightarrow 0} 1/dt P(t < T \leq t + \frac{dt}{T} > t), t \geq 0. \quad [\text{Eq.2}]$$

La fonction de risque cumulé correspond à l'intégrale du risque instantané $h(t)$. Elle est définie par l'équation suivante :

$$H(t) = \int_0^t h(s) ds, t \geq 0. \quad [\text{Eq.3}]$$

Les méthodes d'estimation de la survie peuvent être non paramétriques, semi-paramétriques ou paramétriques. Les méthodes non paramétriques ne font aucune hypothèse sur la nature de la distribution des données observées. Les 2 principales méthodes non paramétriques utilisées sont la méthode actuarielle (BENJAMIN, 1963) et la méthode de Kaplan-Meier (KAPLAN et MEIER, 1958). A l'inverse, les méthodes semi-paramétriques et paramétriques font des hypothèses qui sont respectivement partielles et totales sur la distribution des données observées (COLLETT, 2003). Nous allons détailler plus amplement ces 2 dernières méthodes dans les paragraphes suivants.

1.1.a. Méthode semi-paramétrique : modèle de Cox.

En 1972, David Cox propose un modèle à « risques proportionnels » qui permet d'exprimer le risque instantané de survenue de l'évènement en fonction de variables explicatives exogènes (COX, 1972). Ainsi, le risque instantané (appelé « *hazard* » en anglais) s'écrit de la façon suivante :

$$h(t) = h_0(t) \exp (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k), \quad [\text{Eq.4}]$$

où h_0 représente le risque de base (appelé « *baseline hazard* » en anglais) en l'absence de variables explicatives, et $\beta = (\beta_1 + \beta_2 + \dots + \beta_k)'$ représente le vecteur des coefficients de

régression quantifiant l'effet des variables explicatives x_k sur le risque de survenue de l'évènement.

Le principe de la méthode semi-paramétrique de Cox est d'estimer uniquement les coefficients de régression (partie du modèle paramétrée) et de ne spécifier aucune hypothèse sur la forme de distribution de la fonction de hasard de base « h_0 » (partie du modèle non paramétrée). Une hypothèse importante à vérifier dans le modèle de Cox est l'hypothèse de proportionnalité des risques qui sous-entend que le rapport des risques instantanés (ou « hazard ratio (HR) » en anglais) est constant au cours du temps. Soient deux individus i et j dont les variables explicatives sont x_i et x_j respectivement, le HR associé à la variable explicative x pour ces individus s'écrit de la façon suivante :

$$HR = h_{0i}(t).exp(\beta.x_i)/h_{0j}(t).exp(\beta.x_j) = exp(\beta x_i)/exp(\beta x_j) = exp(\beta(x_i - x_j)), \quad [Eq.5]$$

où h_{0i} et h_{0j} représentent les risques de base chez les individus i et j respectivement, et β le coefficient de régression associé à la variable explicative x . Différentes interprétations du HR sont possibles en fonction de la nature de la variable explicative x :

- pour une variable binaire codée en a et b , $HR = exp(\beta(a-b))$, ce qui se traduit par une augmentation du risque de $exp(\beta(a-b))$ chez le sujet i par rapport au sujet j ;
- pour une variable quantitative, le rapport des risques est estimé pour une augmentation d'une unité de la variable explicative x : $HR = exp(\beta)$; cela signifie que le risque augmente de $exp(\beta)$ quand la variable x augmente d'une unité.

Si la prévalence de l'évènement est très faible, le « hazard ratio » pourrait être interprété comme un risque relatif (RR) dont il est numériquement proche : $RR \sim HR$ (TIMSIT et al., 2005). Plusieurs méthodes existent pour vérifier l'hypothèse de proportionnalité des risques. Ces méthodes sont basées soit sur de la validation graphique (KALBFLEISCH et PRENTICE, 2002), soit sur des tests statistiques (analyse des résidus de Schoenfeld) (THERNEAU et GRAMBSCH, 2000).

Le modèle de Cox a conduit par la suite à de nombreuses extensions: prise en compte de l'interaction entre les différentes variables explicatives, stratification de l'effet des variables explicatives, introduction d'une évolution temporelle, etc.

L'interaction de 2 variables explicatives dans le modèle de Cox s'exprime sous forme de produits des 2 variables concernées. Cette interaction permet d'exprimer l'effet d'une variable donnée sur le risque de survenue de l'évènement en fonction de l'autre variable explicative.

Prenons un exemple pour illustrer ce phénomène. Soient x_1 et x_2 , 2 variables explicatives de la survenue d'un évènement, l'introduction d'une interaction de ces 2 variables dans le modèle de Cox peut s'écrire de la façon suivante :

$$h(t)=h_0.\exp (\beta_1x_1+\beta_2x_2+\beta_3x_1x_2), \quad [\text{Eq.6}]$$

où h_0 est le risque de base, β_1 et β_2 sont les coefficients de régression associés aux variables x_1 et x_2 respectivement, β_3 estime l'interaction entre les 2 variables x_1 et x_2 . L'impact d'une variable explicative sur le risque devient ainsi fonction du niveau de l'autre variable explicative étudiée.

La stratification dans le modèle de Cox concerne les variables définissant des sous-catégories de la population étudiée. Une analyse stratifiée est préférée dans le cas où la mesure directe de l'impact d'une variable sur le risque n'est pas importante ou dans le cas où l'hypothèse de proportionnalité des risques n'est pas vérifiée. Dans ces cas, la variable concernée ne peut pas être incluse directement dans le modèle, un modèle de Cox est par conséquent estimé dans chacune des strates constituant cette variable catégorielle. Dans le modèle de Cox stratifié, les coefficients de régression sont contraints à être identiques entre les différentes strates ou catégories, mais les risques de base diffèrent. Prenons l'exemple d'une variable binaire codée en a et b :

$$\text{Strate a : } h(t)=h_{0a}.\exp (\beta_1x_1+\beta_2x_2+\dots+\beta_kx_k), \quad [\text{Eq.7}]$$

$$\text{Strate b: } h(t)=h_{0b}.\exp (\beta_1x_1+\beta_2x_2+\dots+\beta_kx_k), \quad [\text{Eq.8}]$$

où h_{0a} et h_{0b} représentent les risques de base dans les strates a et b respectivement, et $\beta=(\beta_1+\beta_2+\dots+\beta_k)$ représente le vecteur des coefficients de régression associés aux variables explicatives x_k . L'effet des variables explicatives sur le risque instantané de survenue de l'évènement $h(t)$ est ainsi le même entre les 2 strates (le vecteur des coefficients de régression est identique dans les 2 strates), mais le risque de base h_0 est différent.

Une autre variante intéressante du modèle de Cox consiste à la prise en compte de variables dépendantes du temps (exposition au traitement, marqueur biologique,...). Ceci nécessite la connaissance de la valeur des covariables pour chaque temps t . Ainsi si, par exemple, le suivi inclut n visites, la valeur de chaque covariable doit être connue à chaque visite. Plusieurs logiciels statistiques, tels que le logiciel R, proposent aujourd'hui des méthodes adaptées à la prise en compte de l'évolution dans le temps de certaines variables dans le modèle de Cox. Ces méthodes sont notamment basées sur le découpage des temps de suivi de chaque individu en différents intervalles, chaque intervalle étant défini par la date de survenue d'un changement de la valeur de la variable explicative. Dans le cas des variables dépendantes du temps, l'hypothèse de proportionnalité des risques n'est pas toujours vérifiée, dans la mesure où le rapport des risques dépend du temps :

$$h_i(t) / h_j(t) = \exp\{ \beta(x_i(t) - x_j(t)) \}, \quad [\text{Eq.9}]$$

où h_i et h_j représentent les risques instantanés chez les individus i et j respectivement, $x_i(t)$ et $x_j(t)$ les variables explicatives dépendantes du temps et β le coefficient de régression associé à la variable explicative x .

1.1.b. Méthodes paramétriques

Les modèles paramétriques permettent de spécifier une hypothèse concernant la distribution des durées de survie. Ainsi, le choix du modèle paramétrique sera fonction de la distribution attribuée à la fonction de risque instantané de base $h_0(t)$. On distingue notamment 3 grands types de distributions attribuées à la fonction de risque de base :

- distribution exponentielle : $h_0(t)$ constante,
- distribution de Gamma, Weibull, ou Gompertz : $h_0(t)$ est monotone croissante ou décroissante,
- et distribution en forme de « cloche » ou de « U ».

Un grand intérêt des méthodes paramétriques, par rapport aux méthodes non-paramétriques et semi-paramétriques, consiste dans le fait qu'elles permettent la prise en compte des évènements répétés dans le temps. En effet, les analyses de survie s'intéressent classiquement à la survenue d'un évènement unique. Or les évènements biomédicaux peuvent être de nature répétée comme par exemple les épisodes d'épilepsie, de rejet aigu du greffon ou d'effets indésirables liés aux médicaments. Il est donc intéressant de prendre en compte dans le processus de modélisation cette évolution des évènements dans le temps.

Dans le présent manuscrit, nous nous limiterons à décrire le principe des modèles de survie avec une distribution exponentielle, de Weibull ou de Gompertz qui sont les modèles paramétriques les plus couramment utilisés en analyse de survie. Dans les paragraphes suivants, les fonctions de risque instantané seront présentées sous leur forme de base, c'est-à-dire en l'absence de variables explicatives ($h(t)=h_0(t)$).

I.1.b.(1) Distribution exponentielle

La distribution la plus simple des données de survie correspond à la distribution exponentielle qui implique une fonction de risque instantané $h(t)$ constante au cours du temps et une fonction de survie $S(t)$ qui décroît exponentiellement vers zéro, qui sont définies respectivement par les équations [Eq.10] et [Eq.11] :

$$h(t) = k , \quad \text{[Eq.10]}$$

$$S(t) = \exp(-kt). \quad \text{[Eq.11]}$$

Une propriété importante du modèle de survie avec une distribution exponentielle est qu'il est dit « sans mémoire », c'est-à-dire que la probabilité de survenue de l'évènement pour un individu donné reste la même au cours du temps. Dans certaines situations, le risque instantané peut être constant sur une période donnée mais varier d'une période à une autre au cours du temps. Un modèle différent peut ainsi être considéré dans chacun des intervalles de temps définis.

Le modèle de survie avec une distribution exponentielle permet une plus grande flexibilité des calculs et une facilité d'interprétation des résultats, ce qui constitue son avantage par rapport aux autres modèles paramétriques.

I.1.b.(2) Distribution de Weibull

En 1951, W. Weibull propose une nouvelle loi de distribution pour les données de survie qui est une extension de la distribution exponentielle vue précédemment (WEIBULL et al., 1951). Selon la loi de Weibull, la fonction de risque instantané et la fonction de survie sont définies respectivement par les équations [Eq.12] et [Eq.13] :

$$h(t) = \lambda \cdot \alpha \cdot t^{(\alpha-1)} , t > 0, \quad \text{[Eq.12]}$$

$$S(t) = \exp(-\lambda t^\alpha). \quad \text{[Eq.13]}$$

Les 2 paramètres caractérisant la distribution de Weibull sont λ , un paramètre d'échelle, et α , un paramètre de forme. En fonction des valeurs prises par α ($\alpha > 0$), on obtient différentes formes de la fonction de hasard (figure 1).

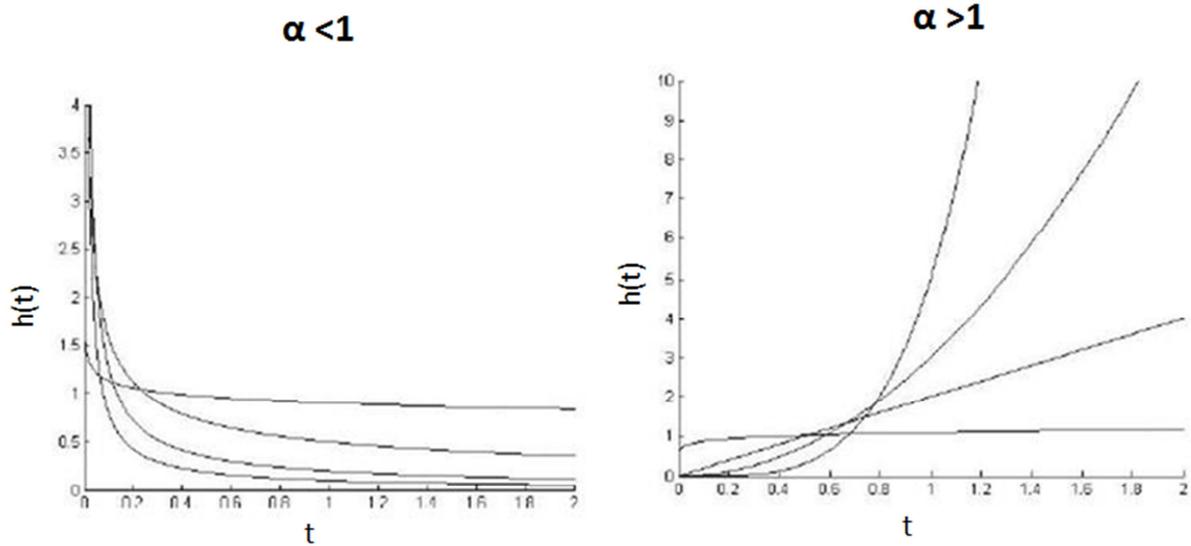


Figure 1: Formes de la fonction de hazard $h(t)$ selon différentes valeurs de α (adapté de PLANCHET, 2013).

- si $\alpha > 1$, $h(t)$ est monotone croissante
- si $\alpha < 1$, $h(t)$ est monotone décroissante
- si $\alpha = 1$, on retrouve la distribution exponentielle

La fonction de risque instantané $h(t)$ selon Weibull peut parfois s'écrire avec une paramétrisation différente :

$$h(t) = \beta_1 \cdot \exp(\beta_2 \ln(t)), \text{ avec } \beta_1 = \alpha \cdot \lambda \text{ et } \beta_2 = \alpha - 1. \quad [\text{Eq.14}]$$

Cette nouvelle paramétrisation permet d'obtenir une expression plus proche de celle du modèle de survie avec une distribution exponentielle.

I.1.b.(3) Distribution de Gompertz

Comme la distribution de Weibull, la distribution de Gompertz suppose aussi un risque instantané monotone. Ce modèle a été proposé en 1825 par Benjamin Gompertz, l'un des premiers mathématiciens à s'être intéressé aux analyses de survie pour estimer la probabilité de décéder à l'âge t (GOMPERTZ, 1825). Le modèle de survie avec une distribution de Gompertz est défini par la fonction de risque instantané $h(t)$ suivante :

$$h(t) = \beta_0 \cdot \exp(\beta_1 \cdot t). \quad [\text{Eq.15}]$$

Selon les valeurs attribuées à β_1 , on obtient différentes formes de $h(t)$:

- si $\beta_1 > 0$, $h(t)$ est monotone croissante

- si $\beta_1 < 0$, $h(t)$ est monotone décroissante
- si $\beta_1 = 0$, on retrouve la distribution exponentielle.

I.2. Régressions logistiques

On distingue deux grands types de régression logistiques appropriés respectivement à la survie d'un évènement unique ou répété dans le temps.

I.2.a. Cas d'évènements uniques

Pour étudier l'association entre la survenue d'un évènement donné et des variables explicatives, le modèle logistique prédit la probabilité P de survenue de l'évènement grâce à l'équation suivante :

$$\text{Log}(P/1-P) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \quad [\text{Eq.16}]$$

où P est la probabilité de survenue de l'évènement, α est l'intercept, et $\beta = (\beta_1 + \beta_2 + \dots + \beta_k)$ est le vecteur des coefficients de régression associés aux variables explicatives x_k .

La forme de l'équation logistique ci-dessus (Eq.16) correspond à une transformation logit. Cette transformation mathématique appliquée à la probabilité P permet d'assurer que celle-ci se trouve toujours dans un intervalle compris entre 0 et 1.

L'interprétation des résultats d'une régression porte principalement sur les coefficients β traduisant l'association entre la survenue de l'évènement et les variables explicatives : un coefficient $\beta > 0$ traduit une association positive, un coefficient $\beta < 0$ traduit une association négative et enfin un coefficient nul traduit une absence d'association. En pratique clinique, on raisonne plutôt en odds-ratio (OR) ou rapport de cotes ($\text{OR} = \exp(\beta)$) de par sa signification clinique intuitive. Lorsque la fréquence de l'évènement étudié est rare dans la population, l'odds-ratio est égal approximativement au risque relatif (LETRILLIART, 1998). En général, l'intercept α n'a pas de signification mais dans le cas des études longitudinales, il permet d'estimer la fréquence de base de l'évènement en l'absence des variables explicatives ($X_1 = X_2 = \dots = X_k = 0$).

I.2.b. Cas d'évènements répétés : modèles utilisant les équations d'estimation généralisées et modèles à effets mixtes

Les études longitudinales peuvent conduire à l'analyse d'évènements qui sont répétés au cours du temps. Dans ce cas, il est nécessaire de prendre en considération la corrélation entre les mesures répétées et ceci peut se faire grâce à 2 approches : approche utilisant les équations d'estimation généralisées (GEE) et approche des modèles à effets mixtes.

I.2.b (1) Modèles utilisant les GEE

Les GEE sont des équations dérivant des modèles linéaires généralisés. Ce système d'équations a été proposé par Liang et Zeger en 1986 (LIANG et ZEGER, 1986). Ces équations sont souvent associées aux régressions linéaire, logistique ou de Poisson pour permettre de tenir compte de la corrélation entre mesures répétées.

La résolution de ces équations GEE se fait grâce à un algorithme modifié du score de Fisher qui consiste en plusieurs étapes, que nous allons décrire ici dans le cas d'une régression logistique.

- 1^{re} étape : estimation des paramètres β avec un modèle de régression logistique sans dépendance entre les observations qui est défini par,

$$\text{Log}(P_i/1-P_i)=\alpha + \beta_1x_{i1} + \beta_2x_{i2} + \dots + \beta_kx_{ik} , \quad [\text{Eq.17}]$$

où P_i est la probabilité de survenue de l'évènement chez l'individu i , α est l'intercept, et $\beta=(\beta_1+\beta_2+\dots+\beta_k)$ 'est le vecteur des coefficients de régression associés aux covariables x_{ik} . Ce modèle logistique est dit moyenné sur la population, il suppose ainsi que la relation entre la survenue de l'évènement et la matrice des variables explicatives (ou covariables) X_i est la même pour tous les sujets.

- 2^e étape : estimation de la matrice de corrélation à l'aide des résidus des valeurs estimées de β en tenant compte de la structure de cette matrice.

- 3^e étape : re-estimation des paramètres β en tenant compte de la matrice de corrélation estimée à l'itération précédente.

Selon les hypothèses sur la nature des liens entre les mesures répétées, plusieurs choix sont possibles pour la structure de la matrice de corrélation nommée ici $R(\alpha)$.

- *Matrice identité*

Dans cette matrice les mesures répétées chez un individu i sont considérées comme indépendantes : la corrélation entre Y_{ij} et $Y_{ij'}$ est nulle pour $j \neq j'$, l'indice j dénotant l'ordre

chronologique dans lequel les mesures ont été réalisées ($R(\alpha) = I_{n_i \times n_i}$). Aucun paramètre n'est à estimer dans cette matrice.

$$R(\alpha) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

- *Matrice échangeable*

On considère que toutes les mesures répétées chez un individu ont une corrélation commune : la corrélation entre Y_{ij} et $Y_{ij'}$ est égale à α . On a ainsi un seul paramètre α à estimer dans cette matrice. Les mesures répétées chez les différents individus peuvent être effectuées à des temps différents et inégalement espacés.

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{pmatrix}$$

- *Matrice autorégressive*

Dans ce type de structure, on suppose que la corrélation entre Y_{ij} et Y_{ij0} est $\alpha^{|j'-j|}$ pour $j \neq j'$. Cette structure de matrice est utilisée quand on considère que les corrélations entre les mesures diminuent avec le temps : plus deux mesures sont éloignées dans le temps, plus leur corrélation est faible. Les mesures répétées chez les différents individus doivent être effectuées à des temps identiques et également espacés.

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & \dots & \alpha^{n_i-1} \\ \alpha & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha \\ \alpha^{n_i-1} & \dots & \alpha & 1 \end{pmatrix}.$$

- *Matrice non-structurée*

Aucune structure particulière n'est imposée ici à la matrice de corrélation. Chaque paire de mesures a sa propre corrélation, ce qui conduit à l'estimation d'un grand nombre de paramètres : un total de $J(J-1)/2$ paramètres à estimer dans la matrice. Les mesures répétées chez les différents individus doivent être effectuées à des temps identiques mais qui peuvent être inégalement espacés.

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1J} \\ \alpha_{12} & 1 & \cdots & \alpha_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1J} & \alpha_{2J} & \cdots & 1 \end{pmatrix}$$

I.2.b (2) Modèles à effets mixtes

Les modèles à effets mixtes permettent aussi de prendre en compte les données corrélées. La prise en compte de la corrélation se fait ici en ajoutant un effet aléatoire au modèle logistique. Le modèle est ainsi composé d'une partie fixe, identique pour chaque individu, constituant la réponse moyenne dans la population, et d'une partie aléatoire traduisant la variabilité liée à chaque sujet :

$$\text{Log}(P_i/1-P_i) = (\alpha + \eta_j) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}, \quad [\text{Eq.18}]$$

où P_i est la probabilité de survenue de l'évènement chez l'individu i , α est l'intercept, η_j est l'effet aléatoire associé à α ($\eta_j \sim N(0, \mu^2)$), et $\beta = (\beta_1 + \beta_2 + \dots + \beta_k)$ est le vecteur des coefficients de régression associés aux variables explicatives (ou covariables) x_{ik} .

Par opposition au modèle utilisant les GEE, le modèle à effets mixtes est un modèle dit « spécifique du sujet » : le modèle suppose ainsi que les relations entre la survenue de l'évènement et la matrice des covariables X_i diffèrent d'un sujet à l'autre. Grâce à l'effet aléatoire associé à l'intercept, le modèle permet d'estimer ainsi une probabilité P_i de survenue de l'évènement spécifique de chaque sujet.

I.3. Modèles multi-états

Les modèles multi-états sont adaptés pour étudier l'évolution d'un processus dans le temps qui est décrit comme une séquence d'évènements ou états. Les transitions entre les différents états peuvent être réversibles, ou non, conduisant ainsi respectivement à des états dits « récurrents » ou « transitoires ». Il existe aussi des états dits « absorbants », c'est-à-dire que le processus ne peut plus sortir de cet état quand il y entre. Les modèles multi-états sont très utiles en épidémiologie clinique où ils permettent d'étudier la progression d'une maladie donnée à travers ses différents stades ou encore la réponse du patient au traitement en considérant différents stades de la maladie. Le cas le plus simple des modèles multi-états est le modèle de survie qui est constitué uniquement de deux états : « vivant » et « mort ». Quand le processus étudié comprend plus de 2 états, les modèles de type Markovien sont employés, permettant d'étudier les transitions possibles qui unissent les différents états définis (ANDERSEN et KEIDING, 2002; ANDERSEN et al., 1991). Les modèles markoviens sont basés sur différentes hypothèses dont la principale est l'hypothèse markovienne. Celle-ci suppose que l'information sur les états antérieurs est résumée par l'état présent au temps t (le passé n'est pas pris en compte). Les modèles markoviens étudient les forces de passage, nommées intensités de transition, entre les différents états. Par exemple, l'intensité de transition λ_{ij} correspond à la force de passage de l'état i à l'état j . λ_{ij} est défini comme le risque instantané de transition et est calculé par le modèle. Un autre paramètre important dans les modèles markoviens est la probabilité de transition $P_{t,j | i}$ qui définit la probabilité d'être dans l'état j dans le futur sachant que l'état présent est i . Celle-ci est modélisée pour une durée donnée.

Les intensités de transitions λ_{ij} peuvent être définies en fonction de deux échelles de temps : t qui est la durée totale du suivi et d qui correspond à la durée passée dans l'état présent. Selon la dépendance ou non de λ_{ij} de t et/ou de d , plusieurs modèles de type Markovien peuvent être définis :

- si $\lambda_{ij}(t,d) = \lambda_{ij}$, le modèle est dit homogène par rapport au temps, c'est-à-dire que les intensités de transition sont constantes au cours du temps et ne dépendent ni de t ni de d ;
- si $\lambda_{ij}(t,d) = \lambda_{ij}(t)$, le modèle est dit non-homogène par rapport au temps ;
- si $\lambda_{ij}(t,d) = \lambda_{ij}(d)$, le modèle est dit semi-Markovien homogène par rapport au temps ;
- si $\lambda_{ij}(t,d) = \lambda_{ij}(t,d)$, le modèle est dit semi-Markovien non homogène par rapport au temps.

Le temps exact d'entrée dans un état n'est pas toujours connu, et il peut être utile de considérer des intervalles de temps plutôt qu'une échelle continue de temps.

II. Méthodes pour prendre en compte une variable quantitative dépendante du temps

En sciences appliquées, il est fréquent d'effectuer des mesures chez un individu i à plusieurs moments j dans le temps. C'est le cas par exemple des études longitudinales, dans lesquelles une réponse biologique ou clinique peut être mesurée chez différents individus à plusieurs temps successifs. La variable longitudinale mesurée peut être quantitative (ex : concentration sanguine d'un médicament, niveau d'un marqueur biologique dans les fluides biologiques, etc.) ou qualitative binaire (ex : survenue d'un évènement clinique de façon répétée dans le temps). Dans cette partie, nous allons nous intéresser aux variables quantitatives, le cas des variables qualitatives binaires ayant été traité dans la section I.

Le traitement statistique des données longitudinales pose un certain nombre de problématiques auxquelles il faudrait apporter des solutions.

- 1^{re} problématique : le nombre de mesures disponibles par individu peut varier en raison d'une durée variable de suivi des patients ou à cause de données manquantes. Une question importante est donc de savoir quelle mesure prendre en compte pour refléter au mieux la réponse globale de chaque patient : moyenne, médiane, minimum, maximum, etc ?
- 2^e problématique : les temps auxquels les mesures sont effectuées peuvent varier d'un individu à un autre. Les questions portent alors sur les temps à prendre en compte : ne faut-il prendre que les temps similaires entre les différents patients au risque d'ignorer certaines mesures disponibles pour certains patients uniquement et de perdre ainsi de l'information ?
- 3^e problématique : l'évolution de la variable à expliquer au cours du temps n'est pas toujours linéaire, alors comment intégrer l'allure non linéaire de la courbe d'évolution ?

L'objectif de cette partie est de présenter les différentes méthodes et modèles statistiques adaptés à l'analyse de données longitudinales dans le cas d'une variable quantitative, qui permettent d'apporter des réponses aux différentes problématiques soulevées.

II.1. Méthode non paramétrique k-means pour données longitudinales

La méthode « k-means » pour données longitudinales (KmL) est une méthode statistique permettant la partition de variables longitudinales en différents groupes homogènes appelés « cluster ». L'organisation des données en groupes a une longue et riche histoire dans une variété de domaines scientifiques (JAIN, 2008). L'un des premiers algorithmes proposés pour partitionner des données était la méthode des k-means, publiée en 1955. Cet algorithme

permettait de partitionner des données en différents clusters ayant des propriétés similaires. Chaque cluster doit ainsi présenter deux propriétés essentielles : une faible similarité inter-cluster et une grande similitude intra-cluster. Chaque donnée ou observation appartient au cluster avec la moyenne la plus proche. L'algorithme KmL, publié en 2010 par C. Genolini (GENOLINI et FALISSARD, 2010), est une variante de k-means adaptée aux données longitudinales, c'est-à-dire faisant l'objet de mesures répétées dans le temps. Il permet ainsi d'obtenir des clusters correspondant à différents profils d'évolution (ou trajectoires) de la variable étudiée.

Cet algorithme a été développé dans le package « KmL » qui est implémenté dans le logiciel R. Les principales propriétés caractérisant l'algorithme KmL sont les suivantes :

- la méthode « KmL » est dite « modèle-indépendante », c'est-à-dire qu'elle ne fait pas d'hypothèse sur la forme des différents profils qui peut être propre à chaque cluster,
- la méthode KmL peut supporter des données manquantes (si l'absence des données n'est pas liée à la réponse étudiée),
- le nombre optimum de clusters est sélectionné sur la base du critère statistique de Calinski et Harabasz (CALINSKI et HARABASZ, 1974). Les différents nombres de clusters (ou partitions) trouvés par KmL sont triés selon le critère de Calinski et Harabatz en ordre décroissant dans le sens du nombre de clusters le plus optimum au moins optimum (Figure 2). Le nombre de clusters sélectionné est indiqué par un point noir sur la figure 2, et les profils moyens correspondants sont représentés à droite de la figure 2.

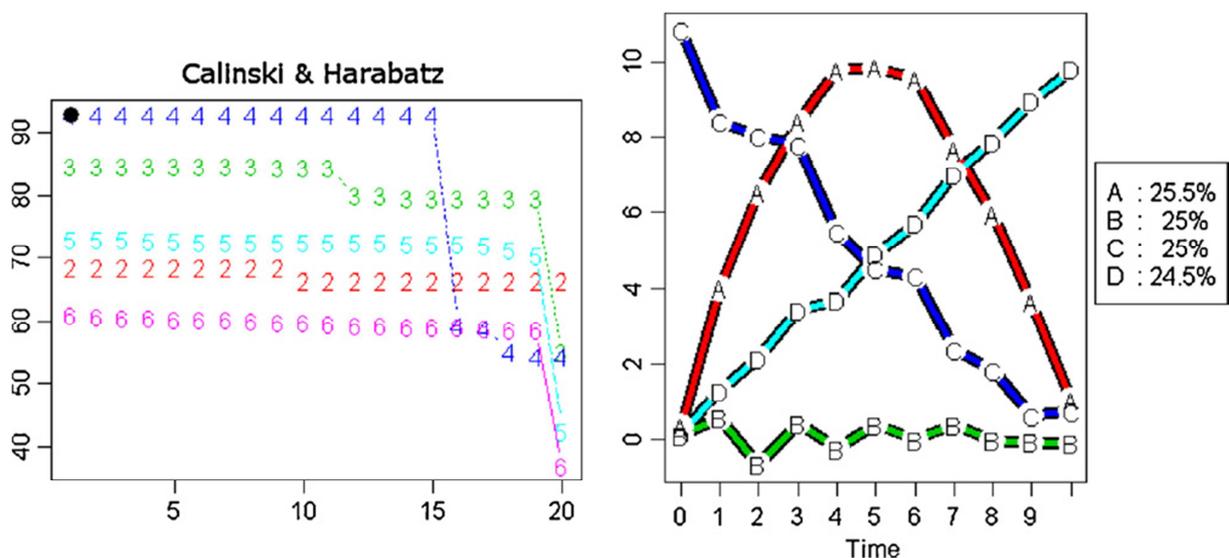


Figure 2: Illustration graphique des clusters obtenus avec le package KmL (adapté de GENOLINI et FALISSARD, 2010).

La méthode de classification selon KmL a montré son grand intérêt en épidémiologie clinique pour la classification de marqueurs quantitatifs longitudinaux (durée de sommeil, durée d'hospitalisation, etc) permettant ainsi l'identification de groupes de patients présentant des caractéristiques similaires.

Le premier exemple d'application de la méthode KmL concerne une étude réalisée par l'Institut de la statistique du Québec chez 1492 enfants suivis jusqu'à l'âge de 6 ans (TOUCHETTE et al., 2007). Cette étude avait pour objectif d'étudier, chez ces enfants, l'association entre des profils de durée de sommeil journalière et le développement comportemental et cognitif à leur entrée à l'école. Quatre clusters de durée de sommeil ont ainsi été identifiés par la méthode KmL.

Le deuxième exemple d'application de la méthode KmL porte sur une étude réalisée par l'Institut Mutualiste Montsouris, consistant à analyser des durées d'hospitalisation annuelles pour anorexie mentale chez 331 adolescents suivis de 0 à 26 ans (GENOLINI et FALISSARD, 2010). L'algorithme KmL a permis d'identifier ainsi 3 clusters de durée d'hospitalisation.

La méthode non-paramétrique de KmL présente un grand avantage par rapport aux méthodes basées sur des modèles : plus de flexibilité sur l'analyse des données car pas d'hypothèse a priori sur la distribution des données. Toutefois, cette méthode possède certaines limites :

- la corrélation entre les mesures chez un même individu n'est pas prise en compte,
- les mesures doivent être réalisées à des moments identiques au cours du temps chez les différents individus (mêmes temps de visites pour les différents individus),
- l'influence de facteurs explicatifs potentiels sur l'évolution de la variable longitudinale ne peut être étudiée.

Les méthodes paramétriques permettent notamment de s'affranchir de ces limites.

II.2. Modèles longitudinaux paramétriques

Une des caractéristiques des études longitudinales est que les mesures répétées chez un même individu peuvent être potentiellement corrélées. Par exemple, dans les études portant sur la greffe rénale, les créatinines sériques mesurées à deux temps différents chez un même individu sont probablement corrélées. Par conséquent, il est raisonnable de prendre en considération la corrélation des mesures pour un individu donné dans les analyses statistiques. Les paragraphes suivants présentent deux méthodes statistiques basées sur des modèles (approche paramétrique) et permettant de prendre en compte la corrélation entre les mesures

répétées chez un même individu : les modèles utilisant les équations d'estimation généralisées et les modèles à effets mixtes.

A la différence de la méthode KmL, ces 2 méthodes peuvent permettre l'analyse de mesures réalisées à des temps variables d'un individu à un autre, et l'étude de l'influence de potentiels facteurs explicatifs sur l'évolution temporelle de la variable à expliquer.

II.2.a. Modèles utilisant les équations d'estimation généralisées

Nous avons vu précédemment dans la section *I.2.b(1)*, l'utilisation des équations d'estimation généralisées (GEE) dans le cadre d'une régression logistique (variable binaire). Les modèles GEE présentés ici sont adaptés aux variables quantitatives et sont associés à une régression linéaire. De manière similaire aux GEE utilisées pour la régression logistique, les GEE permettent ici d'estimer les paramètres moyens de la régression linéaire en tenant compte de la corrélation entre les mesures répétées. Ils peuvent être définis par l'équation suivante :

$$y_{ij} = \beta_0 + \beta_1 t_j + \varepsilon_{ij}, \quad [\text{Eq.19}]$$

où y_{ij} est la variable mesurée chez l'individu i au temps j , β_0 est l'intercept, β_1 est la pente de régression en fonction du temps, et ε_{ij} est un terme correspondant à l'erreur résiduelle non expliquée par le modèle ($\varepsilon_{ij} \sim N(0, \sigma^2)$).

L'approche par GEE n'est pas appropriée pour étudier directement l'effet du temps sur la variable à expliquer y , mais envisage plutôt l'étude de la moyenne mesurée dans la population et une spécification de la matrice de corrélation entre les mesures répétées. Les mêmes matrices de corrélation que pour la régression logistique peuvent être utilisées ici : identité, échangeable, autorégressive, échangeable et non structurée. Comme il a été mentionné plus haut (*section I.2.b(1)*), les matrices échangeable et non-structurée permettent d'analyser des mesures répétées effectuées à des temps variables d'un individu à un autre. Par contre, dans le cas de la matrice autorégressive, les mesures répétées chez les différents individus doivent être réalisées à des temps identiques.

D'autres facteurs explicatifs potentiels de la variable y peuvent être introduits dans l'équation du modèle :

$$y_{ij} = \beta_0 + \beta_1 t_j + \beta_k x_{ik} + \varepsilon_{ij}, \quad [\text{Eq.20}]$$

où β_k est le coefficient de régression associé à la variable explicative x_k .

II.2.b. Modèles à effets mixtes

De la même manière que les modèles à effets mixtes pour variables binaires (cf. section I.2.b(2)), les modèles à effets mixtes pour variables quantitatives utilisent un effet aléatoire pour prendre en compte la corrélation entre les mesures répétées. Cet effet aléatoire correspond à la variabilité liée à chaque sujet. La forme de la liaison entre la variable à expliquer et le temps est bien définie avec ces modèles à effets mixtes. L'évolution de la variable en fonction du temps peut ainsi être décrite par des fonctions linéaires ou non linéaires (polynomiale, exponentielle, Emax, Weibull, Gompertz, etc.). Pour illustrer ce type de modèles à effets mixtes, nous allons présenter le cas du modèle linéaire qu'est le cas le plus simple :

$$y_{ij} = (\beta_0 + \eta_{j0}) + (\beta_1 + \eta_{j1})t_j + \varepsilon_{ij}, \quad [\text{Eq.21}]$$

où y_{ij} est la variable mesurée chez l'individu i au temps j , β_0 et β_1 sont respectivement l'intercept et la pente de régression en fonction du temps et correspondent aux effets fixes du modèle, η_{j0} et η_{j1} sont les effets aléatoires associés respectivement à β_0 et β_1 ($\eta_j \sim N(0, \mu^2)$), et ε_{ij} un terme correspondant à l'erreur résiduelle non expliquée par le modèle ($\varepsilon_{ij} \sim N(0, \sigma^2)$). Les effets aléatoires peuvent porter soit sur l'un des deux paramètres (l'intercept ou la pente), soit sur les deux en même temps.

Les modèles à effets mixtes permettent de prendre en compte des mesures répétées prises à des temps variables d'un individu à un autre, présentant ainsi un plus grand intérêt par rapport à la méthode non-paramétrique de KmL.

D'autres facteurs explicatifs potentiels de la variable y peuvent être introduits dans l'équation du modèle :

$$y_{ij} = (\beta_0 + \eta_{j0}) + (\beta_1 + \eta_{j1})t_j + (\beta_k + \eta_{jk})x_{ik} + \varepsilon_{ij}, \quad [\text{Eq.22}]$$

où β_k est le coefficient de régression associé à la variable explicative x_k , et η_{jk} l'effet aléatoire associé à β_k ($\eta_j \sim N(0, \mu^2)$).

III. Modèles conjoints

Deux types de données sont fréquemment recueillis au cours des études longitudinales ou dans les essais cliniques :

- des données quantitatives répétées (niveau d'un marqueur biologique dans le sang, exposition à un médicament, etc.),
- des données binaires décrivant la survenue ou non d'un événement d'intérêt (décès, sortie d'étude, développement d'une pathologie, etc.).

Généralement, ces deux types de données sont séparément analysés dans les études, mais l'étude de leur association est particulièrement nécessaire dans 2 situations de données (JACQMIN-GADDA et al., 2004). La première situation correspond au cas où l'objectif principal est d'étudier la survenue de l'évènement ; c'est le cas par exemple des analyses de survie. La variable quantitative répétée est considérée dans ce cas comme une variable explicative dépendante du temps. Parmi les évènements les plus étudiés, il y a le décès. On peut citer l'exemple le plus courant des études de VIH où l'évolution du nombre de lymphocytes CD4 ajusté sur le type de traitement anti-VIH a été étudiée dans la survenue du décès des patients (RIZOPOULOS, 2010). Il y a également les études de cancérologie telles que les études du cancer de la prostate où l'évolution du niveau d'anticorps PSA (Prostate specific antigen) a été étudiée dans l'apparition de métastases ou la survenue du décès (JACQMIN-GADDA et al., 2010). Un autre évènement intéressant est aussi le rejet dans les études de greffe rénale. Différentes études ont été publiées récemment s'intéressant à l'association entre l'évolution de l'exposition au traitement immunosuppresseur et l'apparition du premier épisode de rejet aigu (FROBEL et al., 2013; ZHOU et al., 2012). La deuxième situation correspond au cas où l'objectif principal est d'étudier l'évolution de la variable quantitative répétée. Un des problèmes les plus fréquents lors de l'étude de l'évolution d'une variable répétée concerne les données manquantes qui peuvent être dues, entre autres, à des sorties d'étude. L'absence des données peut être aléatoire, c'est-à-dire non liée à la variable étudiée. Dans ce cas, les données manquantes n'induisent pas de biais dans le modèle d'évolution de la variable quantitative répétée. Par contre, dans d'autres cas, les données manquantes sont dites informatives (LITTLE et RUBIN, 2002), et peuvent ainsi entraîner un biais de sélection affectant les estimateurs des paramètres du modèle d'évolution de la variable quantitative répétée. Pour corriger ce biais potentiel, une solution est de considérer les absences de données comme des évènements que l'on pourra appeler sortie d'études, et de modéliser ensemble l'évolution de la variable quantitative répétée et le délai des sorties d'étude.

Les méthodes classiquement utilisées pour étudier l'association entre la survenue de l'évènement et la variable quantitative répétée sont par exemple les modèles de survie tels que le modèle de Cox dans lequel la variable répétée est considérée comme une variable explicative dépendante du temps. Cependant cette méthode présente deux limites principales : elle nécessite que les valeurs de la variable quantitative répétée soient connues à tous les temps d'évènements et elle ne permet pas de tenir compte de l'erreur de mesure associée à cette variable. Pour contourner la première limite, une méthode souvent utilisée est l'imputation avec la dernière valeur observée (Last Observation Carried Forward : LOCF).

Cependant, celle-ci induit une estimation biaisée des paramètres du modèle surtout quand les temps entre les mesures sont longs ou quand l'erreur de mesure est grande.

Les modèles conjoints (appelés aussi modèles joints) constituent une nouvelle approche statistique de plus en plus utilisée dans le domaine biostatistique et permettent de s'affranchir des limites posées par les méthodes classiques. Cette approche des modèles conjoints permet d'analyser ensemble l'évolution de la variable quantitative répétée et la survenue de l'évènement grâce à une combinaison de deux sous-modèles : un sous-modèle pour décrire l'évolution de la variable quantitative répétée et un sous-modèle pour décrire la survenue de l'évènement. Il existe deux grands types de modèles conjoints : les modèles conjoints dits « à effets aléatoires partagés » (TSIATIS et WULFSON, 1997 ; TSIATIS et DAVIDIAN, 2004) et les modèles conjoints dits « à classes latentes » (LIN et al., 2002 ; PROUST-LIMA et al., 2007 ; JACQMIN-GADDA et al., 2010).

Les modèles conjoints « à effets aléatoires partagés » permettent d'estimer un profil moyen d'évolution de la variable quantitative répétée et un risque d'évènement fonction des caractéristiques continues de cette variable. Pour décrire l'évolution de la variable quantitative répétée, un modèle à effets mixtes est utilisé. Le modèle à effets mixtes, dont le principe a été décrit dans la section *II.2.b*, permet de tenir compte de l'erreur de mesure et d'avoir des estimations des valeurs individuelles de la variable répétée à tous les temps. Pour décrire la survenue de l'évènement, deux types de modèles peuvent être utilisés : un modèle de survie (semi-paramétrique ou paramétrique) et un modèle de régression logistique. L'estimation des modèles conjoints peut se faire soit par une approche en deux étapes, soit par une approche simultanée. L'approche en deux étapes consiste à modéliser la variable quantitative répétée grâce au modèle à effets mixtes puis d'utiliser les prédictions individuelles obtenues avec le modèle comme variable explicative dépendante du temps dans le modèle pour décrire l'évènement (modèle de survie ou modèle logistique). L'approche simultanée consiste à estimer simultanément les paramètres des deux sous-modèles par maximisation de la vraisemblance conjointe (TSIATIS et DAVIDIAN, 2004). Bien que l'estimation en deux étapes soit plus flexible et rapide que l'approche simultanée, son utilisation n'est pas appropriée dans le cas où l'objectif de l'étude est centré sur l'évolution de la variable quantitative répétée. Dans ce cas, la maximisation de la vraisemblance conjointe est nécessaire permettant d'obtenir des estimateurs non biaisés.

Les modèles conjoints « à classes latentes » supposent que la population est divisée en k classes latentes avec un profil moyen d'évolution de la variable quantitative répétée et un risque d'évènement associé propre à chaque classe. Le profil moyen d'évolution de la variable

quantitative est décrit dans chaque classe à l'aide d'un modèle à effet mixtes. Ce type de modèles conjoints est ainsi adapté pour étudier l'association entre une variable quantitative répétée et un risque d'évènement dans une population hétérogène.

IV. Objectifs de la thèse

Le challenge du traitement immunosuppresseur (IS) en greffe d'organes, consiste à trouver un équilibre subtil qui permet d'éviter les épisodes de rejet aigu du greffon tout en limitant les effets secondaires et le risque d'infections associé à ces médicaments. Le traitement IS d'entretien est constitué le plus souvent d'une trithérapie associant un antimétabolite (azathioprine (AZA) ou plus fréquemment aujourd'hui mycophénolate mofétil (MMF)), un anticalcineurine (cyclosporine (CsA) ou tacrolimus (TAC)) et un corticoïde. Les médicaments IS possèdent une fenêtre thérapeutique étroite et sont caractérisés par une variabilité interindividuelle importante de leur pharmacocinétique, ce qui justifie le recours au suivi thérapeutique pharmacologique (STP) pour certains de ces médicaments. Le STP est obligatoire pour les anticalcineurines (ACN) et le rôle bénéfique de ce dernier est reconnu par les cliniciens. A l'inverse, l'intérêt du STP pour le MMF, prodrogue de l'acide mycophénolique (AMP), reste débattu à l'heure actuelle (BYRNE et al., 2011; VAN GELDER, 2011).

La première partie de cette thèse a pour objectif d'analyser les facteurs prédictifs de réponse aux traitements IS administrés en transplantation rénale. Pour répondre à cet objectif, nous avons réalisé 3 études. Dans la première étude présentée au Chapitre 2 section I, la réponse étudiée est le rejet aigu du greffon dans la première année post-greffe. L'effet de l'exposition longitudinale à l'AMP sur la survenue du rejet aigu a été testé en utilisant l'approche des modèles conjoints pour données longitudinales et données de survie. Dans la seconde analyse présentée au Chapitre 2 section II, la réponse étudiée est constituée d'un critère composite d'efficacité incluant le rejet aigu, le retour en dialyse et le décès, et d'un deuxième critère considéré comme marqueur potentiel de sur-immunosuppression et incluant les infections et maladies à cytomégalovirus (CMV). L'effet de l'exposition longitudinale à l'AMP et aux ACN, ainsi que les effets d'autres covariables caractéristiques du donneur et du receveur ont été étudiés sur ces deux critères de réponse aux traitements IS, en utilisant respectivement un modèle de survie paramétrique et un modèle de régression logistique. Enfin, dans la troisième étude présentée au Chapitre 2 section III, la réponse étudiée correspond à l'anémie qui est un des principaux effets indésirables attribuables au MMF. L'effet de l'exposition à l'AMP sur la survenue de l'anémie a été étudié en utilisant parallèlement 3 méthodes adaptées à la prise en

compte de données longitudinales dans des modèles de régression logistique avec effets mixtes, régression logistique utilisant les GEE et modèle de Markov.

La deuxième partie de cette thèse, présentée au Chapitre 3, est consacrée à présenter les résultats que nous avons obtenus dans un travail collaboratif réalisé dans une large cohorte de patients transplantés rénaux suivis à Limoges et dans lequel une analyse de survie non paramétrique basée sur des arbres de régression et forêts aléatoires a été menée pour établir une classification des facteurs prédictifs de survie du greffon rénal, 10 ans après la transplantation.

Enfin, la troisième partie de cette thèse, présentée au Chapitre 4, porte sur une étude réalisée en oncogériatrie dont l'objectif était d'analyser la pharmacocinétique (PK) de la capécitabine et de ses métabolites chez les sujets âgés de plus de 75 ans, et évaluer la relation entre l'exposition à ces produits et leur toxicité dans cette même population. Ce travail s'appuie sur l'essai CAPAGEC mené à Limoges chez les sujets de plus de 75 ans, et nous avons utilisé le modèle PK développé précédemment par Urien et al. (2005)

CHAPITRE 2 :

Analyse de l'influence de l'exposition aux immunosuppresseurs sur la survenue de différents évènements en greffe rénale

I. Analyse de l'influence de l'exposition longitudinale à l'acide mycophénolique sur la survenue du rejet aigu

Le rejet aigu est un critère d'échec du traitement IS classiquement retenu comme critère d'efficacité dans les essais cliniques. L'association entre l'exposition à l'AMP et le rejet aigu en greffe rénale reste très débattue à l'heure actuelle (Byrne et al., 2011; van Gelder, 2011). La présente étude évalue cette association en prenant en compte l'exposition longitudinale à l'AMP et le premier épisode de rejet aigu survenu au cours de la première année post-greffe. La base de données analysée inclue 490 patients transplantés rénaux traités avec du MMF associé à un ACN (CsA ou TAC). L'exposition longitudinale à l'AMP était appréciée au moyen des différentes valeurs d'AUC mesurées chez les patients au cours de la première année post-greffe ($AMP\ AUC = f(t)$). Le grand intérêt de cette étude reposait sur l'approche de modélisation mise en œuvre pour étudier l'association exposition/rejet. En effet, une approche de modèles conjoints permettait d'étudier l'influence de l'évolution au cours du temps de l'exposition à l'AMP sur le risque de rejet aigu. Les variabilités inter et intra-individuelles de l'exposition à l'AMP étaient ainsi prises en compte dans cette approche. L'exposition longitudinale à l'AMP a été décrite par un modèle polynomial à effets mixtes et un modèle de survie paramétrique a été utilisé pour la prédiction du risque de rejet aigu. Le modèle conjoint développé dans le package JM implémenté dans le logiciel R (RIZOPOULOS, 2010) a permis de montrer une association significative entre l'exposition longitudinale à l'AMP et le risque de rejet aigu ($p=0.0081$). Le modèle, validé par bootstrap a ensuite permis de déterminer des valeurs seuils d'exposition à l'AMP dépendants du temps en utilisant des courbes ROC adaptées aux modèles conjoints. Des valeurs seuils d'exposition à

l'AMP variant de 35 à 41 mg.h/L au cours de la première année ont ainsi été proposées à l'issue de cette analyse.

Cette étude a fait l'objet d'une publication dans le journal *Pharmacological Research* en 2013.

ARTICLE 1:

Impact of longitudinal exposure to mycophenolic acid on acute rejection in renal-transplant recipients using a joint modeling approach

Z. Daher Abdi, M. Essig, D. Rizopoulos, Y. Le Meur, A. Prémaud,
J.B. Woillard, J.P. Rérolle, P. Marquet, A. Rousseau

Pharmacological Research 72 (2013) 52– 60

II. Analyse de l'influence de l'exposition longitudinale à l'acide mycophénolique et aux anticalcineurines sur un critère composite d'efficacité et sur les infections et maladies à cytomégalovirus

La précédente étude a montré une association significative entre l'exposition longitudinale à l'AMP et le risque de rejet aigu mais la principale limite de cette étude était qu'elle ne tenait pas compte de l'exposition longitudinale à l'ACN co-administré avec le MMF. L'analyse de la relation entre expositions longitudinales aux IS co-administrés (AMP et ACN) et efficacité a fait l'objet d'une seconde étude qui est la première à proposer un modèle de survie intégrant l'effet de deux variables longitudinales.

Un critère composite d'efficacité associant le rejet aigu, le retour en dialyse et le décès a été considéré. Les données issues de 222 patients transplantés rénaux ayant reçu du MMF associé à un ACN (CsA ou TAC) ont été analysées. Sur la base de mesures répétées d'AUC de l'AMP, de C2 de CsA et de C0 de Tacro, obtenues au cours de la première année post-greffe, les expositions longitudinales aux médicaments immunosuppresseurs ont été décrites à l'aide de modèles non linéaires à effets mixtes. Une approche séquentielle était conduite. Ainsi les prédictions individuelles obtenues avec ces modèles longitudinaux ont été introduites comme des covariables dépendantes du temps dans un modèle de survie paramétrique. Les temps exacts de survenue des événements n'étant pas forcément connus une censure par intervalle a été appliquée. Les modèles ont été développés dans le logiciel Nonmem® Version 7. Cette analyse a montré une association significative entre l'exposition longitudinale à l'AMP et le risque de rejet aigu, de dialyse ou de décès, dans un modèle ajusté sur l'exposition longitudinale à l'ACN co-administré. Ce modèle a de plus confirmé l'association entre infections et maladies à CMV, et survie sans échec (rejet aigu, dialyse, décès). Ces résultats sont en accord avec le rôle négatif du CMV dans la survie du greffon et du patient, précédemment rapporté (ERIKSSON et al., 1996; FRYD et al., 1980; SAGEDAL et al., 2004).

Un 2^e critère de réponse aux IS correspondant aux infections et maladies à CMV a également été étudié afin de tester l'hypothèse que les infections et maladies à CMV seraient un marqueur de sur-immunosuppression. Cependant, l'analyse par régression logistique, qui tenait compte des événements répétés, n'a montré aucune influence de l'exposition longitudinale aux IS sur la survenue des infections et maladies à CMV. Par ailleurs, cette étude a permis de quantifier le risque associé aux statuts donneur/receveur du CMV sur la

survenue des infections et maladies à CMV. En prenant comme référence le sous-groupe D-/R-, le risque le plus élevé était obtenu dans le sous-groupe D+/R-(D+/R- vs. D-/R-: OR=10.8, CI90%: 4.7-48.5), un risque intermédiaire était obtenu avec le sous-groupe D+/R+ (D+/R+ vs. D-/R-: OR=7.9, CI90%: 3.5-36.3), et le risque le plus faible était obtenu avec le sous-groupe D-/R+ (D-/R+ vs. D-/R-: OR=3.9, CI90%: 1.3-19.2).

Cette étude fait l'objet d'un manuscrit soumis pour publication dans le journal *American Journal of Transplantation*

III. Analyse de l'influence de l'exposition longitudinale à l'acide mycophénolique sur la survenue d'anémie

III.1. Contexte

L'anémie est un évènement fréquent chez les patients transplantés. Elle est d'origine multifactorielle mais peut être attribuée dans certains cas au MMF, dont elle constitue l'un des principaux effets indésirables (VANRENTERGHEM et al., 2003). Le but de cette analyse était d'étudier l'influence de l'exposition à l'AMP sur le risque de survenue de l'anémie au cours de la 1^{re} année après la transplantation en utilisant parallèlement 3 approches de modélisation adaptées aux données longitudinales : régression logistique utilisant les GEE (modèle GEE), régression logistique avec effets mixtes (modèle GLMM), et modèle de Markov. Cette étude a fait l'objet d'un poster publié au congrès de PAGE à Glasgow en Juin 2013 (<http://www.page-meeting.org/default.asp?abstract=2870>).

III.2. Patients et méthodes

Nous avons utilisé pour cette analyse les données issues de patients transplantés rénaux inclus dans l'essai randomisé d'APOMYGRE (NCT0019967). Ces patients avaient reçu soit une dose fixe, soit une dose ajustée de MMF, en association avec la cyclosporine (LE MEUR et al., 2007). Les valeurs d'AUC_{0-12h} de l'AMP ont été obtenues à plusieurs temps au cours de la 1^{re} année post-greffe (jours 7 et 14, mois 1, 3, 6 et 12). Elles ont été estimées à l'aide d'estimateurs bayésiens précédemment validés (PREMAUD et al., 2005). L'anémie a été définie ici comme une variable binaire: 1 pour présence d'anémie (hémoglobine <10 g/dl) et 0 pour absence d'anémie. La présence ou l'absence d'anémie était renseignée aux différentes visites du patient. Les anémies récurrentes chez un même patient au cours de la période de suivi étaient prises en compte dans l'analyse.

L'association exposition à l'AMP/anémie a été étudiée en considérant l'AUC de l'AMP obtenue à la visite antérieure au diagnostic de l'anémie. Nous faisons ainsi l'hypothèse qu'il existait un certain délai entre une exposition donnée à l'AMP et la survenue de l'anémie. L'effet d'autres covariables correspondant aux caractéristiques du donneur et du receveur (âge du donneur, âge du receveur, genre, stratégie de posologie du MMF (bras fixe vs bras ajusté)) a été également pris en compte dans le modèle. La valeur d'hémoglobine du patient mesurée à

1 mois après la greffe a été également testée comme covariable dans le modèle. Les caractéristiques des patients inclus dans l'analyse ont été résumées dans le tableau 1.

Nombre de patients	130
Temps de visites après la transplantation	J7, J14, M1, M3, M6, M12
Anémie, évènements/ patients (n)	42 / 32
Age du receveur (ans)	49,8 ± 13,7
Age du donneur (ans)	47,4 ± 14,3
Hémoglobine à 1 mois (g/dl)	11,5 ± 1,4
Genre, H/F (n)	84/ 46
Bras, Dose fixe/ Dose ajustée (n)	65/ 65

Tableau 1: Caractéristiques des patients de l'étude

J: jour, M: mois, H: homme, F: femme. Les variables quantitatives sont exprimées en moyenne ± écart-type.

Les modèles GEE et GLMM ont été développés dans le package lme4 implémenté dans le logiciel R (BATES et al., 2013; HØJSGAARD et al., 2006), et le modèle de Markov a été développé dans le logiciel NONMEM® (BEAL et al., 2009) en utilisant la méthode d'estimation de Laplace. L'équation *logit* suivante a été utilisée pour les 3 types de modèles:

$$\text{Log} (P_i/1-P_i) = \alpha + \eta_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik},$$

où P_i correspondait à la probabilité d'avoir une anémie chez l'individu i , α correspondait à l'intercept et η_j l'effet aléatoire associé à cet intercept ($\eta_j \sim N(0, \mu^2)$), et enfin β_k représentait le coefficient de régression associé à la covariable x_{ik} . L'effet aléatoire η_j était utilisé que dans les modèles GLMM et de Markov permettant d'avoir ainsi des estimations de paramètres spécifiques du sujet. A l'inverse l'effet aléatoire n'était pas introduit dans le modèle GEE et des estimations de paramètres moyennées sur la population étaient obtenues avec ce modèle. Différents types de matrices de corrélation ont été testées pour le modèle GEE : échangeable, autorégressive et non-structurée.

A la différence des deux autres modèles, le modèle de Markov permettait de prendre en compte 4 types de transitions correspondant à des changements d'état du patient en fonction de sa valeur d'hémoglobine mesurée :

- transition 1 : valeur normale d'hémoglobine → anémie
- transition 2 : anémie → valeur normale d'hémoglobine
- transition 3 : valeur normale d'hémoglobine → valeur normale d'hémoglobine
- transition 4 : anémie → anémie

Le modèle de Markov a été évalué sur la base de graphes diagnostiques appelés « *Posterior Predictive check : PPC* » obtenus à partir d'un jeu de simulations des données et permettant de vérifier les propriétés prédictives du modèle.

III.3. Résultats

Une matrice autorégressive a été sélectionnée pour le modèle GEE. Cette matrice était particulièrement adaptée à nos données qui correspondaient à des mesures répétées prises à des temps identiques chez les individus mais inégalement espacées. Les 2 facteurs prédictifs de la survenue d'anémie sélectionnés dans les 3 modèles étaient la valeur d'hémoglobine à 1 mois et l'AUC de l'AMP. Les odds-ratio [95%IC] associés à l'exposition à l'AMP étaient du même ordre de grandeur dans les 3 modèles : 1,027 [1,001-1,053], $p=0,044$ pour le modèle GLMM; 1,016 [1,002-1,031], $p=0,029$ pour le modèle GEE et 1,020 [1,003-1,444] pour le modèle de Markov. Ainsi, une augmentation d'une unité d'AUC de l'AMP (mg.h/L) était associée à une augmentation du risque d'anémie de 2.7%, 1.6% ou 2.0% respectivement avec les modèles GLMM, GEE et Markov. Les graphes de PPC obtenus avec le modèle de Markov ont montré que le modèle prédisait bien la transition1 (valeur normale d'hémoglobine → anémie) : le nombre de transitions 1 observé était compris dans l'intervalle de prédiction à 95%. Cependant, la transition dans le sens inverse (anémie → valeur normale d'hémoglobine) n'était pas bien prédite par le modèle. (Figure 3).

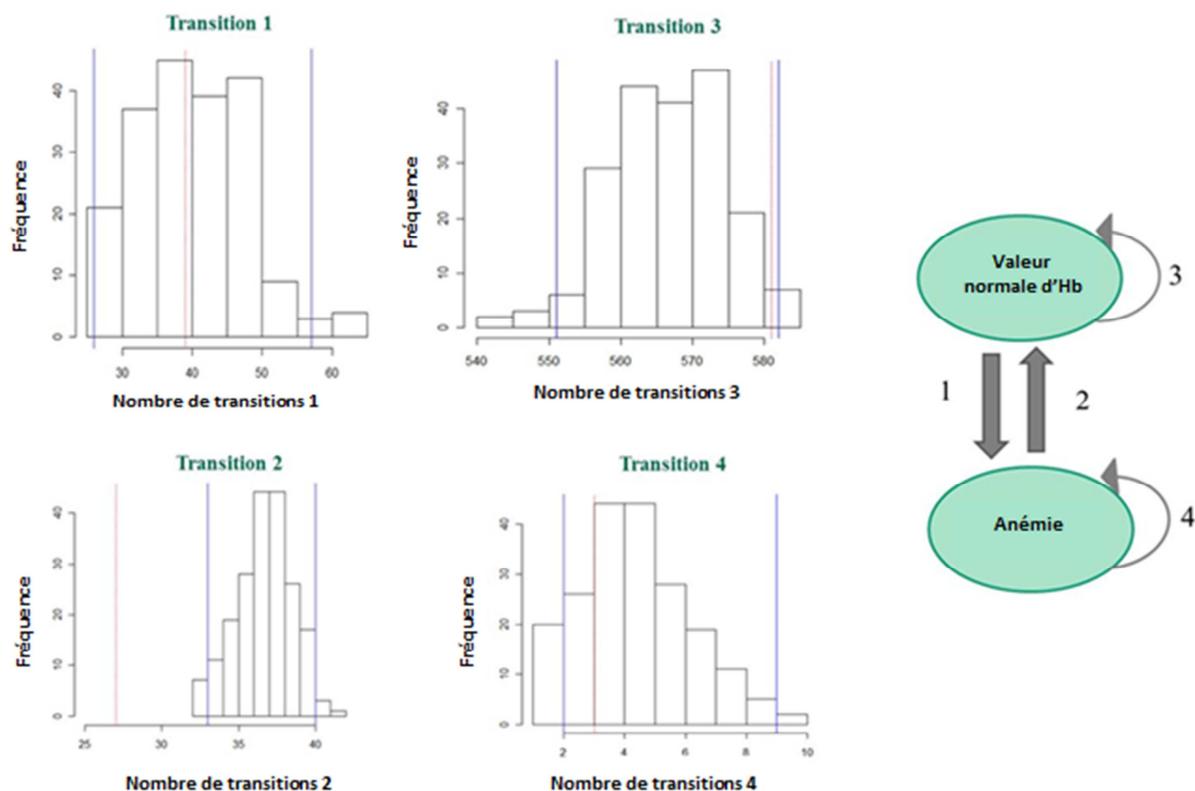


Figure 3:Graphes de « PPC » obtenus avec le modèle de Markov

Les histogrammes représentent la distribution des nombres de transitions simulés à partir du modèle. Les traits rouges représentent les nombres de transitions observés, et les traits bleus correspondent au 2.5^e et 97.5^e percentiles des nombres de transitions simulés.

III.4. Discussion

Une association significative entre l'exposition à l'AMP et la survenue de l'anémie a été montrée chez les patients transplantés rénaux en utilisant parallèlement 3 types de modèles. Les OR obtenus étaient estimés avec une bonne précision. L'intérêt des 3 modèles utilisés dans cette analyse était la prise en compte des événements répétés au cours du temps. Ces modèles étaient ainsi particulièrement adaptés au cas de l'anémie qui pourrait survenir plusieurs fois chez un même patient au cours du temps. Le modèle de Markov permettait en plus de prendre en compte l'aspect dynamique du processus étudié, en étudiant les différentes transitions au cours du temps correspondant au passage du patient d'un état à un autre. L'AUC de l'AMP n'était pas une covariable significative dans la transition correspondant au passage d'un état anémique à une valeur normale d'hémoglobine. En effet, l'érythropoïétine était une covariable potentielle pour cette transition, mais celle-ci n'a pas pu être testée dans cette analyse en raison d'un grand nombre de données manquantes (>50%).

CHAPITRE 3 :

Analyse de la survie du greffon rénal : étude sur une large cohorte de patients transplantés rénaux suivis à Limoges

Le travail présenté dans cette partie concerne une analyse de survie du greffon à long terme dans la cohorte des patients ayant bénéficié d'une greffe rénale entre 1984 et 2011 et suivis à Limoges. Ce projet avait pour objectif d'identifier et de hiérarchiser les facteurs prédictifs de la survie du greffon rénal à 10 ans de façon à établir une classification conditionnelle des facteurs de risque prédictifs de l'échec de greffe.

La première partie du travail a consisté à créer une base de données commune au service de Néphrologie, Dialyse, Transplantations du CHU de Limoges et au Laboratoire HLA, Service d'Immunologie et Immunogénétique du CHU de Limoges. Cette base de données a inclus les données archivées dans le registre national Cristal tenu par l'agence de Biomédecine (données du donneur, bilan pré-greffe du receveur et suivi annuel clinico-biologique du receveur), des données de suivi clinico-biologique du receveur issues d'une base de données clinique tenue au sein du service de Néphrologie, Dialyse, Transplantations (base MEDWARE déclarée à la CNIL). Cette dernière base pouvait fournir les données de suivi des différents bilans réalisés durant la première année post-greffe pour les patients greffés ces cinq dernières années. Pour les patients pour lesquels ces données n'étaient pas disponibles dans MEDWARE, un recueil rétrospectif dans les dossiers cliniques papier a été réalisé. Cette base a été complétée par les bilans de suivi immunologique qui prévoyaient une recherche d'anticorps anti HLA une fois par semaine pendant les deux premiers mois, puis tous les ans et lorsqu'il existait une indication clinique (évènement immunisant, rejet ...)

Un point fort de cette cohorte était que les anticorps anti-HLA avaient été identifiés et quantifiés par la technique Luminex® (Luminex Corporation, Austin, USA) qui est la technique la plus sensible actuellement. Cette technique est utilisée en routine depuis 2003 au CHU de Limoges. Les sérums collectés avant 2003 et analysés initialement par la technique de micro-lymphocytotoxicité dépendante du complément ont été contrôlés par la technique Luminex®, ces échantillons étant stockés dans la collection biologique de l'établissement.

Enfin, les nombres de mismatch HLA A-B-DR ont été calculés en considérant les sous-spécificités des antigènes HLA (ou « *splits* ») du donneur et du receveur.

Un atout de cette base de données était de disposer d'un suivi à long terme des patients; 203 patients étaient suivis plus de dix ans post greffe. Les analyses de survie en greffe rénale précédemment décrites dans la littérature avaient une période de suivi post-greffe inférieure, huit ans pour la plus longue (FOUCHER et al., 2010).

Des méthodes d'analyse de survie non paramétriques basées sur des arbres de régression ont été mises en œuvre dans ce projet pour identifier les facteurs prédictifs de survie du greffon rénal à 10 ans. En analyse de la survie, différentes stratégies de modélisation peuvent être envisagées pour prédire le risque d'événements futurs. Le choix par défaut est classiquement un modèle de régression de Cox. Le modèle de Cox est « semi-paramétrique » ; il suppose une indépendance entre le processus de censure et le temps de survie étudié. Cette hypothèse peut être difficilement acceptable dans l'analyse de survie des greffons rénaux puisque la mesure de la créatinine sérique et le débit de filtration glomérulaire dépendent de facteurs multiples et sont liés à de nombreuses comorbidités. Une approche alternative pour l'analyse de survie est basée sur les arbres de décision : arbres de classification et de régression et forêts d'arbres.

Les arbres de décisions sont largement utilisés pour leur lisibilité, ils peuvent fournir des règles logiques de classification des individus à partir de traits descriptifs. L'arbre décrit une suite de questions relatives aux caractéristiques de chaque sujet. Sa représentation graphique est une structure en arborescence constituée de nœuds. L'idée de base est de mettre ensemble des entités qui se ressemblent et de séparer celles qui diffèrent les unes des autres. Généralement, la base de données est partagée en deux sous-ensembles : un sous ensemble d'apprentissage contenant 65 à 80% des individus et un sous ensemble d'évaluation. La procédure consiste à diviser récursivement (souvent de façon binaire) l'ensemble d'apprentissage sur la base de tests statistiques jusqu'à ce que l'on obtienne des sous-ensembles de données ne contenant (presque) que des données appartenant à une même classe (et ne pouvant plus être partagés).

Les « forêts aléatoires » adaptées aux données de survie ont été proposées par Léo Breiman en 1996 (BREIMAN, 1996). Un nombre préalablement fixé d'arbres de décision est alors généré, à l'aide de processus d'apprentissage aléatoires. En 2008, Ishwaran *et al* ont développé un programme dans R (package RandomSurvivalForest-RSF-) dédié aux forêts aléatoires pour données censurées à droite (ISHWARAN et al., 2008). Dans l'algorithme utilisé la

randomisation intervient à deux étapes : 1) chaque arbre est construit sur un sous-échantillon tiré aléatoirement et 2) à chaque nœud, une sélection de variables explicatives choisies aléatoirement est testées. L'algorithme RSF fonctionne en 5 étapes ainsi décrites par Ishwaran *et al* :

- 1- Constituer « *narbres* » échantillons, par bootstrap à partir d'une fraction de la base de données originale. Chaque échantillon « bootstrap » exclut en moyenne 37% des données de la base originale (37% est la valeur par défaut) qui constituent un groupe d'évaluation nommé par les anglo-saxons « out-of-bag ».
- 2- Construire un arbre de survie pour chaque échantillon « bootstrappé ». A chaque nœud, plusieurs variables explicatives possibles sont choisies aléatoirement. Retenir la covariable qui maximise la différence de survie entre les nœuds enfants.
- 3- Arrêter la croissance de l'arbre quand chaque nœud terminal contient un nombre minimum prédéfini de décès (ou d'évènements non censurés)
- 4- Calculer la moyenne des fonctions de risque cumulé pour chaque arbre (à l'aide de l'estimateur de Nelson-Aalen), et combiner les informations de tous les arbres constituant la forêt. Chaque arbre fournit un estimateur du risque cumulé pour chaque nœud terminal.
- 5- Calculer le taux d'erreur sur l'ensemble des groupes d'évaluation (*narbres* groupes).

Cette méthode des forêts aléatoires appliquée aux données de survie, fournit ainsi une liste validée de prédicteurs, classés par ordre d'importance. Le recours à la procédure de ré-échantillonnage par « bootstrap » (par défaut 1000 échantillons « bootstrappés ») et le calcul du taux d'erreur sur groupe d'évaluation indépendant sont des points forts de cette méthode. Ainsi, selon Breiman et Cutler (2004), dans les forêts aléatoires, la validation croisée ou un test dans un groupe indépendant n'est pas nécessaire pour obtenir une estimation non biaisée du taux d'erreur. La limite de cette méthode est liée à la difficulté d'interprétation des résultats pour une utilisation en clinique, car cette méthode ne fournit pas d'arbre « final ».

Pour cette raison, nous avons utilisé, en complément, le package R-party (disponible dans le logiciel R) qui produit un arbre conditionnel par une série de divisions binaires (partitionnement récursif) des données. Cette dernière analyse a été conduite sur une base de données réduite incluant les variables considérées comme les plus importantes à l'issue de l'analyse par forêts aléatoires de survie. Schématiquement, l'algorithme de cette construction d'arbre fonctionne comme suit: 1) tester l'hypothèse nulle globale de l'indépendance entre chaque variable d'entrée et la réponse c'est-à-dire la survie. Si l'hypothèse d'indépendance ne peut être écartée avec aucune covariable, le développement de l'arbre s'arrête à ce nœud.

Sinon, la variable d'entrée avec la plus forte association à la réponse est sélectionnée. Cette association est mesurée par une valeur de p correspondant à l'hypothèse nulle partielle testée. 2) de façon similaire, une division binaire de la variable d'entrée qui vient d'être sélectionnée est envisagée. 3) les étapes 1 et 2 sont répétées dans un processus récursif (TORSTEN et al, 2006).

A l'issue de ces analyses, une classification conditionnelle des différents facteurs de risque impliqués dans la survie du greffon rénal à 10 ans a été obtenue. Chez les patients ayant une créatinine sérique à 12 mois supérieure à 150 $\mu\text{mol/L}$, la survie du greffon dépendait principalement des variables caractéristiques de la dysfonction chronique du greffon (profils d'évolution de la créatininémie au cours de la première année et protéinurie à 12 mois post-greffe). Près de 40% des patients classés dans ces sous-groupes avaient reçu un greffon issu d'un donneur âgé de plus de 60 ans. Lorsque la créatinine sérique à 12 mois était inférieure à 150 $\mu\text{mol/L}$, la survie du greffon dépendait principalement de l'apparition d'anticorps anti-HLA spécifiques du donneur (donor specific antibodies : DSA), celle-ci étant plus précoce chez les patients ayant des anticorps anti-HLA avant la greffe. La survie des patients ayant développé des DSA était fortement associée au développement de rejet aigu. Enfin, la survie du greffon chez les patients qui présentaient une créatinine sérique inférieure à 150 $\mu\text{mol/L}$ à l'issue de la première année post-greffe et qui n'avaient pas développé des DSA dépendait de la présence d'anticorps anti-HLA avant greffe associée à une re-transplantation. Ce dernier sous-groupe, sans aucun des facteurs de risques retenus, avait une survie du greffon à 10 ans voisine de 95%.

Une première version du manuscrit présentant les résultats obtenus est proposée ci-après.

CHAPITRE 4 :

Etude de pharmacocinétique et de pharmacodynamie de la capécitabine chez des patients âgés de plus de 75 ans

Les objectifs de cette étude étaient d'évaluer l'impact de l'âge sur la pharmacocinétique de la capécitabine et de ses métabolites (5'-DFCR : 5'-deoxy-5-fluorocytidine, 5'-DFUR : 5'-deoxy-5-fluorouridine, 5-FU : 5-fluorouracile et FBAL : alpha-fluoro-beta-alanine) et la relation exposition/effet de ces molécules chez les sujets âgés.

Ce travail s'appuyait sur l'essai clinique CAPAGEC (NCT00812864) mené à Limoges qui incluait des sujets âgés (≥ 75 ans) atteints de cancers du sein ou colorectaux et traités avec de la capécitabine en monothérapie. Cette base initiale a été combinée avec des données pharmacocinétiques issues de 2 études de phase I incluant des patients plus jeunes (<75 ans) atteints de cancers métastatiques et qui avaient reçu de la capécitabine associée à de l'irinotecan ou de l'irofulven (URIEN et al., 2005).

L'analyse de la pharmacocinétique a été réalisée par une approche de population par la méthode non linéaire à effets mixtes implémentée dans le logiciel NONMEM®. La pharmacocinétique de la capécitabine et de ses 3 métabolites (5'-DFCR, 5'-DFUR et 5-FU) a été décrite par le modèle structural à 4 compartiments développé par Urien et al. (2005). L'analyse des covariables a montré un effet significatif du sexe sur la constante d'élimination du 5-FU. Cependant, aucun effet de l'âge n'a été montré sur la pharmacocinétique de la capécitabine et de ses métabolites. En outre, aucune différence majeure n'a été observée dans la tolérance au traitement et le taux de réponse chez les patients âgés par rapport aux patients plus jeunes.

L'analyse de la relation exposition-toxicité a mis en évidence des expositions médianes à la capécitabine et à 3 de ses métabolites, supérieures chez les patients ayant développé un syndrome main-pied par rapport à ceux qui ne l'ont pas développé. Par ailleurs, aucune différence significative des expositions médianes aux différentes molécules n'a été mise en évidence entre les patients ayant présenté une réponse partielle ou une stabilisation de la maladie, et les non-répondeurs.

Cette étude fait l'objet d'un manuscrit soumis pour publication dans le journal *Cancer Chemotherapy and Pharmacology*.

CHAPITRE 5 : Discussion et conclusion

Ce travail de thèse s'articule autour de 3 parties. Les 2 premières parties (*Chapitres 2 et 3*) étaient consacrées à des études menées en transplantation rénale s'intéressant aux facteurs prédictifs du devenir du greffon. La 3^e partie (*Chapitre 4*) concernait une étude de pharmacocinétique et de pharmacodynamie menée en oncologie, sur la capécitabine administrée au sujet âgé (≥ 75 ans). Ces 3 parties seront discutées séparément au cours de cet exposé.

Discussion de la 1^{re} partie

La 1^{re} partie de cette thèse (*Chapitre 2*) était consacrée à analyser les facteurs prédictifs de réponse aux traitements immunosuppresseurs (IS) de maintenance administrés en transplantation rénale (MMF et ACN) en s'intéressant notamment à 3 critères de réponse :

- ❖ critère d'efficacité ou d'échec du traitement : rejet aigu, retour en dialyse, décès
- ❖ marqueur potentiel de sur-immunosuppression : infection ou maladies à CMV
- ❖ critère de toxicité : anémie

Le facteur prédictif principal de réponse aux IS considéré au cours de ces analyses est l'exposition à ces IS qui est une variable quantitative faisant l'objet de mesures répétées au cours du temps. Différents marqueurs d'exposition sont utilisés selon le type de médicament IS :

- ❖ aire sous la courbe des concentrations entre deux administrations du médicament (exemple : AUC_{0-12h}) pour l'acide mycophénolique (AMP), métabolite actif du MMF ;
- ❖ concentrations sanguines 2h après administration du médicament (C2h) pour la ciclosporine (CsA) ;
- ❖ concentrations sanguines résiduelles (C0) pour le tacrolimus (TAC).

L'exposition aux IS varie au cours du temps. L'exposition à l'AMP est soumise à une grande variabilité inter et intra-individuelle. Il y a une augmentation des AUCs, ainsi que des AUC/dose de l'AMP, dans les 1^{ers} mois suivant la transplantation. Cette augmentation de l'exposition à l'AMP a déjà été évoquée par différents auteurs qui l'ont expliquée en partie par une diminution de la clairance apparente de l'AMP (VAN HEST et al., 2007). Par ailleurs,

les concentrations des ACN diminuent au cours du temps du fait de la diminution des cibles d'exposition recommandées pour ces traitements (EKBERG et al., 2007).

Les travaux menés dans cette partie de thèse illustrent l'intérêt de prendre en compte l'évolution au cours du temps de l'exposition aux IS dans les analyses de l'association exposition-réponse, et ce en utilisant des méthodes statistiques novatrices adaptées aux données longitudinales. L'approche méthodologique employée dans les deux premières études (*Chapitre 2, sections I et II*) est basée sur les modèles conjoints, récemment introduits dans le domaine biostatistique. Dans ces travaux, nous avons mis en œuvre les modèles conjoints dits « à effets aléatoires partagés ». Le principe de ces modèles consiste à modéliser l'évolution au cours du temps de l'exposition au médicament IS à l'aide d'un modèle à effets mixtes et de l'introduire dans un modèle de survie ou logistique utilisé pour prédire le critère de réponse. Les critères de réponse sont analysés ici sous forme de variable binaire (oui/non). L'estimation des deux sous modèles peut être simultanée par maximisation de la vraisemblance conjointe (*Chapitre 2, section I*). Une estimation en deux étapes est également possible, consistant à utiliser les prédictions individuelles obtenues avec le modèle à effets mixtes comme covariable dépendante du temps dans le modèle de prédiction de la réponse (*Chapitre 2, section II*). Les modèles conjoints « à effets aléatoires partagés » utilisés dans ces études, présentent certains intérêts par rapport aux modèles conjoints « à classes latentes ». En effet, ces modèles peuvent être combinés à des courbes ROC spécialement adaptées aux données de survie, ce qui n'a pas été à ce jour réalisé dans les modèles conjoints à classes latentes. De plus, les modèles conjoints à classes latentes publiés à ce jour ont recours à un modèle linéaire pour décrire l'évolution du marqueur longitudinal. Le choix des modèles conjoints à effets aléatoires partagés facilite l'implémentation de modèles non linéaires complexes pour décrire l'évolution du marqueur longitudinal.

Dans la troisième étude (*Chapitre 2, section III*), différents types de modèles adaptés à l'analyse d'évènements répétés ont été utilisés pour la prédiction de la réponse au traitement IS. Les valeurs observées de l'exposition au médicament IS ont été introduites dans ces modèles comme covariable dépendante du temps.

Grâce à ces nouvelles approches méthodologiques, les différentes études menées dans cette partie de thèse nous ont permis d'obtenir un certain nombre de résultats nouveaux concernant la relation exposition-réponse aux traitements IS, et dont certains pouvant avoir un impact clinique.

- L'exposition longitudinale à l'AMP a une influence sur le risque de rejet aigu (*Chapitre 2, sections I et II*).

Le rejet aigu représente un des principaux obstacles au succès de la greffe. Il existe deux grands types de rejet aigu : le rejet aigu médié par les anticorps (rejet humoral) et le rejet aigu médié par les lymphocytes T (rejet cellulaire). L'association entre l'exposition à l'AMP et le rejet aigu en greffe rénale reste très débattue à l'heure actuelle (BYRNE et al., 2011; VAN GELDER, 2011). Par conséquent le STP de l'AMP n'est réalisé que dans certains centres de transplantation (SAINT-MARCOUX et al., 2011). Les principales limites des études précédemment décrites dans la littérature concernaient la méthodologie statistique utilisée qui consistait à comparer entre les groupes avec et sans rejet les moyennes d'exposition obtenues à un temps donné au cours de la période post-transplantation. En utilisant un modèle à effets mixtes pour l'exposition longitudinale à l'AMP et un modèle de survie paramétrique pour la prédiction du rejet, nous avons pu montrer dans 2 études séparées une association significative entre l'exposition longitudinale à l'AMP et le rejet aigu chez des sujets transplantés rénaux, en s'intéressant principalement au premier épisode de rejet aigu au cours de la 1^{re} année post-transplantation. Dans la 1^{re} étude (*Chapitre 2, section I*), le rejet aigu était analysé comme seul critère d'efficacité et dans la 2^e étude (*Chapitre 2, section II*), il était inclus dans un critère composite d'efficacité comprenant le rejet aigu, le retour en dialyse et le décès. Bien que le principe de l'approche méthodologique utilisée dans les 2 études restait la même, la 2^e étude nous a permis d'aller plus loin dans la méthodologie par rapport à la 1^{re} étude grâce aux options disponibles dans le logiciel utilisé NONMEM® (BEAL et al., 2009) : utilisation d'équations plus sophistiquées et flexibles pour décrire l'évolution de l'exposition longitudinale, prise en compte de l'exposition longitudinale de l'ACN co-administré avec l'AMP et utilisation de la censure par intervalles dans le modèle de survie. En effet, dans le package JM implémenté dans R (RIZOPOULOS, 2010) qui a été utilisé pour la 1^{re} étude, les équations pour décrire l'évolution de l'exposition longitudinale ne pouvaient être que linéaires ou polynomiales. A l'inverse, le logiciel NONMEM® nous a permis de tester différentes équations pour l'exposition longitudinale : polynomiale, exponentielle, Emax, Weibull, Gompertz, etc. L'avantage de ceci était de pouvoir sélectionner le meilleur modèle pour décrire l'exposition longitudinale, et de réduire ainsi le biais qui pourrait être induit dans l'analyse de l'association exposition-réponse. Par ailleurs, le logiciel NONMEM® nous a permis de prendre en compte simultanément les expositions à l'AMP et l'ACN co-administré dans le

modèle de prédiction de la réponse. A notre connaissance, cette analyse était la 1^{re} permettant de prendre en compte l'effet de la co-exposition longitudinale aux IS sur le critère de réponse à ces traitements. D'autres auteurs ont déjà étudié l'effet de l'exposition longitudinale des IS sur le risque de survenue du rejet aigu dans des modèles de survie qui ont également été développés dans NONMEM® (FROBEL et al., 2013; ZHOU et al., 2012). Cependant aucune de ces études ne prenait en compte la co-exposition des IS co-administrés. Dans notre analyse, l'association entre l'exposition longitudinale aux ACN et le critère composite d'efficacité (rejet aigu, retour en dialyse, décès) n'était pas significative. Ce résultat ne contredit pas le rôle majeur et bénéfique du STP des ACN, mais pourrait s'expliquer par le fait que les échecs observés dans cette étude ne seraient pas imputables à l'exposition longitudinale aux ACN. Enfin, un autre avantage du logiciel NONMEM® par rapport au logiciel R était de pouvoir utiliser une méthode de censures par intervalles dans le modèle de survie prédisant le critère d'efficacité. La censure par intervalles est une méthode utilisée dans les analyses de survie quand les temps exacts de survenue de l'évènement sont inconnus (COMMENGES, 2002). Nous avons considéré que cette méthode était la plus adaptée aux différents évènements étudiés ici (rejet aigu, retour en dialyse, décès). En effet, il est bien connu que les réactions immunologiques impliquées dans le rejet aigu précèdent le diagnostic de celui-ci. D'autre part, on pourrait supposer que le greffon commence à devenir non-fonctionnel avant le diagnostic de décès ou la date de retour en dialyse. L'intervalle considéré pour les différents évènements était ainsi compris entre la visite précédant le diagnostic de l'évènement et la visite correspondant au diagnostic de l'évènement. Le travail réalisé dans la 2^e étude a permis également de confirmer l'association entre infection et maladie à CMV et survenue du rejet, décès et perte du greffon, dans le modèle tenant compte de l'exposition longitudinale des IS co-administrés (AMP et ACN). En effet, plusieurs auteurs avaient déjà rapporté que le CMV contribuait à une augmentation de l'incidence du rejet d'allogreffe (REISCHIG, 2010; SAGEDAL et al., 2002) et à une diminution de la survie du patient et du greffon rénal (ERIKSSON et al., 1996; FRYD et al., 1980; SAGEDAL et al., 2004). En ajustant sur l'effet de l'exposition longitudinale des 2 IS co-administrés, notre analyse a permis ainsi de confirmer le rôle négatif des infection et maladie à CMV sur le rejet aigu et l'échec de greffe.

A l'issue de ces travaux qui nous ont permis de montrer une association significative entre l'exposition longitudinale à l'AMP et le rejet aigu, nous avons pu proposer des seuils d'exposition dépendants du temps en utilisant des courbes ROC spécifiques aux modèles

conjoints (*Chapitre 2, section I*). Des seuils variant légèrement au cours de la première année entre 35 mg.h/L et 41 mg.h/L ont été proposés. Ces seuils étaient en accord avec ceux déterminés dans l'essai randomisé RCTT (VAN GELDER et al., 1999) et utilisé dans les essais FDCC et APOMYGRE (VAN GELDER et al., 2008; LE MEUR et al., 2007). Ces seuils pourraient être considérés comme des valeurs minimums d'exposition pour avoir une diminution du risque de rejet. Il a été montré au cours de cette étude que les patients ayant développé un rejet ont été sous-exposés par rapport à ces seuils à un moment donné pendant la période du suivi : soit à un temps proche du rejet, soit à un temps antérieur au rejet. Ces résultats illustraient ainsi l'importance d'une exposition optimale à l'AMP chez les patients transplantés tout au long de la 1^{re} année, d'où l'intérêt de réaliser le STP pour cet IS.

- L'exposition aux IS (AMP et ACN) n'est pas associée au risque de survenue des infections et maladies à CMV (*Chapitre 2, section II*)

Les infections à CMV représentent les infections virales opportunistes les plus courantes en transplantation (KANJ et al., 1996; STRATTA et al., 1992). Notre hypothèse de départ était que les infections et maladies à CMV seraient un marqueur de sur-immunosuppression. Dans une étude de sondage récente, la plupart des cliniciens ont indiqué qu'ils diminuaient les doses d'IS lors de maladies à CMV sévères et récidivantes (LE PAGE et al., 2013). Cependant, jusqu'à présent, aucune relation entre les infections et maladies à CMV et l'exposition longitudinale aux IS n'avaient été rapportée dans la littérature. Grâce à un modèle logistique à effets mixtes permettant de tenir compte des événements répétés, nous avons montré que l'exposition aux IS n'influe pas sur le risque de survenue des infections et maladies à CMV. Nos résultats permettraient ainsi d'éviter, en pratique clinique, une diminution inutile des doses d'IS chez les patients. Par ailleurs, cette étude nous a permis de quantifier le risque associé aux statuts donneur-receveur (D/R) du CMV dans la survenue des infections et maladies à CMV. En effet, bien que l'association entre les statuts D/R du CMV et les infections et maladies à CMV était connue et déjà rapportée dans des études précédentes (KOTTON et al., 2013), celle-ci n'avait pas été quantifiée jusqu'à présent. En prenant comme référence le sous-groupe D-/R-, nous avons calculé les OR associés aux différents sous-groupes grâce au modèle de régression logistique. Le risque le plus élevé était ainsi obtenu dans le sous-groupe D+/R- (D+/R- vs. D-/R-: OR=10.8, CI90%: 4.7-48.5), un risque intermédiaire était obtenu avec le sous-groupe D+/R+ (D+/R+ vs. D-/R-: OR=7.9, CI90%: 3.5-36.3), et le risque le plus

faible était obtenu avec le sous-groupe D-/R+ (D-/R+ vs. D-/R-: OR=3.9, CI90%: 1.3-19.2).

- L'exposition à l'AMP pourrait être associée au risque de survenue de l'anémie (*Chapitre 2, section III*).

Les causes d'anémie en transplantation sont multifactorielles. L'anémie peut être attribuée dans certains cas au MMF dont elle constitue un des principaux effets indésirables (VANRENTERGHEM et al., 2003). Par 3 méthodes différentes adaptées aux données longitudinales, nous avons montré que l'exposition à l'AMP pourrait être associée à la survenue de l'anémie (codée en oui/non) au cours de la 1^{re} année suivant la transplantation. L'intérêt d'utiliser ces 3 méthodes en parallèle était de pouvoir étudier la relation entre l'exposition à l'AMP et la survenue d'anémie en considérant différentes hypothèses pour la prise en charge des mesures répétées. En effet, bien que ces 3 méthodes soient toutes adaptées aux données longitudinales, les hypothèses sous-jacentes à chacune d'entre elles sont différentes (*cf. Chapitre Introduction*). Les OR associés à l'exposition à l'AMP étaient du même ordre de grandeur avec les 3 méthodes utilisées : régression logistique avec effets mixtes (modèle GLMM), régression logistique utilisant les GEE (modèle GEE), et modèle de Markov. Une augmentation d'une unité de l'AUC_{0-12h} de l'AMP (en mg.h/L) était associée à une augmentation du risque d'anémie de 2.7%, 1.6% ou 2.0% respectivement avec les modèles GLMM, GEE et Markov.

Cependant, une limite de cette analyse pourrait être le traitement de l'anémie comme une variable binaire. Une approche plus informative serait de prendre en compte les différents grades de l'anémie. En raison de l'indisponibilité de certaines données, cette approche n'a pas pu être testée dans notre analyse.

Dans les différentes études réalisées dans cette partie de thèse, les facteurs relatifs aux caractéristiques du donneur et du receveur (âge du donneur, âge du receveur, genre, nombre de HLA mismatches A-B-DR, temps d'ischémie froide et retransplantation) ont également été testés comme facteurs prédictifs de la réponse aux IS dans les modèles tenant compte des expositions longitudinales aux IS. Mais aucun de ces facteurs n'étaient significatifs.

Les principales limites de ces études étaient cependant la non prise en compte des doses de corticoïdes administrés dans les différents modèles développés. Les corticoïdes sont associés au MMF et aux ACN dans le traitement d'immunosuppression de maintenance utilisé en greffe rénale. Bien que ces médicaments ne bénéficient pas de STP comme le MMF et les ACN, il était cependant intéressant de tenir compte des doses de corticoïdes administrés dans

les différentes périodes du suivi des patients. Ces données n'étant pas disponibles pour tous les patients, nous n'avons pas pu les inclure dans notre analyse.

Une autre limite concernant l'analyse de l'association exposition-rejet aigu était de ne pas prendre en compte le type de rejet aigu dans l'analyse : rejet aigu médié par les anticorps et rejet aigu médié par les lymphocytes T. Cette information de classification n'était pas renseignée pour les différents rejets inclus dans l'analyse.

Discussion de la 2^e partie

La deuxième partie de cette thèse (*Chapitre 3*) était consacrée aux résultats issus d'un travail collaboratif réalisé dans une large cohorte de patients transplantés rénaux suivis à Limoges. Ce travail concernait l'étude des facteurs prédictifs de la survie à 10 ans du greffon rénal. Les études précédemment rapportées dans la littérature concernant les facteurs prédictifs de la survie à long terme du greffon rénal étaient principalement basées sur des analyses univariées de comparaison de groupe ou sur des modèles de Cox. D'autres approches alternatives déjà rapportées avaient pour objectifs la construction de scores additifs à partir des caractéristiques du donneur et du receveur, obtenues en pré-greffe ou collectées dans la première année post-greffe (FOUCHER et al., 2010; NYBERG et al., 2003, 2005).

Une classification conditionnelle des différents facteurs de risque impliqués dans la survie du greffon rénal à 10 ans a été obtenue. Contrairement à un score additif, cette classification conditionnelle permettait d'avoir des estimations de survie dans chaque sous-groupe de patients à partir de facteurs prédictifs qui n'étaient pas les mêmes dans l'ensemble de la population et qui variaient conditionnellement aux valeurs d'autres facteurs prédictifs. Sept sous-groupes ont été identifiés avec des survies du greffon à 10 ans variant entre 20% et 95%. Les facteurs prédictifs identifiés dans cette étude étaient connus comme facteurs de risque de perte du greffon. Cependant, l'utilisation d'arbres conditionnels pour la classification de ces facteurs était nouvelle, ce qui constituait un point fort de cette analyse.

Le premier critère discriminant retenu était la créatinine sérique mesurée à 12 mois post-greffe. Chez les patients ayant une créatinine sérique à 12 mois supérieure à 150 $\mu\text{mol/L}$, la survie du greffon dépendait principalement des variables caractéristiques de la dysfonction chronique d'allogreffe (profils d'évolution de la créatininémie au cours de la première année et protéinurie à 12 mois). Une très forte association était trouvée entre l'âge du donneur et la créatinine sérique à 12 mois. Près de 40% des patients classés dans ce sous-groupe avaient reçu un greffon issu d'un donneur âgé (>60 ans). Chez les patients avec une créatinine sérique

à 12 mois inférieure à 150 $\mu\text{mol/L}$, le principal critère discriminant était l'apparition d'anticorps anti-HLA spécifiques du donneur (DSA), celle-ci étant plus précoce chez les patients ayant des anticorps anti-HLA avant la greffe. La survie des patients ayant développé des DSA était fortement associée à la présence de rejet aigu. A 10 ans post-greffe, le taux d'échec de greffe était, de 60% pour ceux qui avaient développé un rejet aigu versus 10% pour ceux qui n'en n'avaient pas développé ($p=0.012$). Cooper *et al* (2011) avaient précédemment rapporté que l'impact clinique des DSA était lié à leur association avec des épisodes de rejet aigu concomitants ou précédant leur formation plutôt qu'à leur apparition. Dans la présente étude, la survie post-DSA du greffon était fortement impactée par la survenue du rejet aigu et elle était plus faible chez les patients ayant présenté un premier épisode de rejet aigu avant l'apparition des DSA. Nous avons montré dans les travaux précédents (*Chapitre 2, sections I et II*) que le risque de rejet aigu était influencé par l'exposition longitudinale aux IS, d'où l'intérêt du suivi thérapeutique pour ces traitements afin d'améliorer la survie du greffon.

Enfin, la survie du greffon chez les patients avec une créatinine sérique à 12 mois inférieure à 150 $\mu\text{mol/L}$ et n'ayant pas développé de DSA était associée à la présence d'anticorps anti-HLA pré-greffe et aux antécédents de greffe rénale antérieure.

La classification des facteurs de risque proposée dans cette étude permet d'apporter une nouvelle vision sur les facteurs de risque de survie du greffon rénal en étudiant leur interdépendance avec une nouvelle approche méthodologique. Cette étude est la première à proposer une classification conditionnelle des facteurs de risque d'échec de greffe rénale, avec un recul à 10 ans post-greffe. Elle fournit ainsi des éléments qui pourraient être utiles pour la prise de décision en clinique.

Discussion de la 3^e partie

La troisième et dernière partie de cette thèse (*Chapitre 4*) concernait une étude de la PK de la capécitabine et de ses métabolites (5'DFCR, 5'DFUR, 5FU et FBAL) chez les sujets âgés (≥ 75 ans), et l'évaluation de la relation exposition/effet de ces produits dans cette même population. Très peu d'essais cliniques en cancérologie sont réalisés chez le sujet âgé et les relations dose-concentrations et concentrations-effet sont mal connus dans cette population (HUTCHINS *et al.*, 1999; LEWIS *et al.*, 2003). Le travail réalisé dans cette partie s'appuyait sur l'essai clinique CAPAGEC (NCT00812864) mené au CHU de Limoges. Cet essai incluait

des sujets âgés de plus de 75 ans atteints de cancers du sein ou colorectaux et qui étaient traités avec de la capécitabine en monothérapie.

L'analyse des covariables a montré un effet significatif du sexe sur la constante d'élimination du 5-FU (k_{40}). k_{40} était significativement plus faible chez les femmes par rapport aux hommes. Des résultats similaires ont été obtenus dans de précédentes études réalisées chez des patients ayant reçu une chimiothérapie incluant du 5-FU (ETIENNE et al., 1998; TERRET et al., 2000). Cet effet du sexe sur la PK de la capécitabine et de ses métabolites reste toutefois modéré et ne nécessite pas un ajustement de dose. Aucune influence de l'âge n'a été montrée sur la PK de la capécitabine et de ses métabolites. En outre, aucune différence majeure n'a été observée dans la tolérance au traitement et le taux de réponse chez les patients âgés par rapport aux patients plus jeunes. La capécitabine et ses métabolites étant éliminés majoritairement par voie rénale, la clairance de la créatinine (estimée selon la formule de Cockcroft-Gault) a été également testée comme covariable dans l'analyse PK. Aucun effet de la clairance de la créatinine sur les différents paramètres PK n'a été montré. Ces résultats étaient en accord avec ceux rapportés par Poole et al. (2002) et Gieschke et al. (2003) qui n'ont montré aucun effet significatif de la clairance de la créatinine sur l'exposition systémique à la capécitabine ou au 5-FU.

L'analyse de la relation exposition-toxicité a montré des expositions médianes de la capécitabine et 3 de ses métabolites (5'-DFCR, 5'-DFUR et 5-FU), supérieures chez les patients ayant développé un syndrome main-pied par rapport à ceux qui ne l'ont pas développé. Par ailleurs, les expositions médianes des différents produits n'étaient pas significativement différentes entre les patients chez qui la maladie avait été stabilisée ou chez qui une réponse partielle était observée et les non-répondeurs. Certains auteurs avaient déjà rapporté l'absence de relation entre l'exposition systémique à la capécitabine et à ses métabolites et la réponse à la chimiothérapie (GIESCHKE et al., 2003).

Conclusion

Ce travail de thèse a été consacré à des études de modélisation menées en oncologie et en transplantation rénale.

L'étude de pharmacocinétique et de pharmacodynamie menée en oncologie, sur la capécitabine administrée au sujet âgé (de plus de 75 ans) n'a montré aucun effet significatif de l'âge sur la pharmacocinétique de la capécitabine et de ses métabolites. Cette étude apportait ainsi de nouveaux arguments dans la prise en charge thérapeutique des patients âgés qui pourraient bénéficier d'une chimiothérapie par capécitabine avec des niveaux de toxicité similaires à ceux observés chez les patients plus jeunes

En transplantation rénale, ce travail a concerné l'étude de facteurs prédictifs du devenir du greffon, en utilisant de nouvelles approches méthodologiques. Dans un modèle joint pour données longitudinales et données de survie, nous avons montré une relation entre l'exposition longitudinale à l'AMP et la survie sans rejet aigu au cours de la 1^{re} année post-greffe. Cette étude apportait ainsi de nouveaux arguments en faveur d'un suivi thérapeutique du MMF, pro-drogue de l'AMP, au cours de cette 1^{re} année post-greffe. Des cibles d'AUC de l'AMP ont été déterminées en utilisant des courbes de ROC spécifiquement adaptées aux modèles joints, ces valeurs seuils confirmaient les cibles proposées dans la littérature. Par la méthode des modèles non-linéaires à effets mixtes, nous avons ensuite développé un modèle plus complexe pour étudier l'effet de chacun des médicaments immunosuppresseurs co-administrés sur un critère composite d'efficacité incluant : rejet aigu, retour en dialyse et décès survenue principalement pendant la première année post-greffe (3 évènements sur 38 étaient survenus au-delà de la première année). Ce modèle a confirmé l'effet de l'exposition longitudinale à l'AMP sur le devenir du patient greffé et a montré l'impact des infections et maladies à CMV sur ce critère d'efficacité. Une hypothèse étudiée dans notre travail était que les infections et maladies à CMV seraient liées à une sur-immunosuppression. Aucune relation entre exposition aux médicaments immunosuppresseurs et apparition d'infection ou de maladie à CMV n'a été mise en évidence quel que soit la variable d'exposition prise en compte. La dernière partie du travail mené en transplantation, s'est appuyée sur une large cohorte de patients transplantés rénaux qui avaient bénéficié d'un suivi annuel des anticorps anti HLA, en plus du suivi classique clinico-biologique. Cette analyse de cohorte a conduit à

une classification conditionnelle des facteurs prédictifs de l'échec de greffe et à l'identification de sept sous-groupes de patients avec des survies du greffon à 10 ans variant entre 20% et 95%. Les facteurs discriminants retenus étaient la créatinine sérique mesurée à 12 mois post-greffe, le profil d'évolution de la créatinine sérique dans la première année post-greffe (ces deux facteurs étant fortement associés à l'âge du donneur), la protéinurie à 12 mois post-greffe, le développement d'anticorps anti-HLA spécifiques du donneur (DSA), le rejet aigu, la présence d'anticorps anti-HLA avant la greffe chez les patients retransplantés. Cette classification fournit des données scientifiques nouvelles qui pourraient guider le choix de stratégie clinique chez un patient en fonction de ses facteurs de risque.

Références bibliographiques

- ANDERSEN P. K., HANSEN L. S., KEIDING N. Assessing the influence of reversible disease indicators on survival. *Stat Med* 1991; **10**:1061–1067.
- ANDERSEN P. K., KEIDING N. Multi-state models for event history analysis. *Stat Methods Med Res* 2002; **11**:91–115.
- BATES D., MAECHLER M., BOLKER B. *lme4: Linear mixed-effects models using Eigen and syntax* - R package version 2.15.0 (2012). Available at: [Url:http://CRAN.R-project.org/package=lme4](http://CRAN.R-project.org/package=lme4) (accessed 11/03/2013).
- BEAL S.L., SHEINER L.B., BOECKMANN A., BAUER R.J. *NONMEM User's Guides (1989-2009)*. Icon Development Solutions, Ellicott City, USA, 2009.
- BENJAMIN B. Actuarial methods of mortality analysis; adaptation to changes in the age and cause pattern. *Proc R Soc Lond B Biol Sci* 1963; **159**:38–65.
- BREIMAN L. Bagging Predictors . *Machine Learning* 1996; **24**:123-140.
- BREIMAN L., CUTLER A. *Random Forests - version 5.1* (2004). Available at: [Url: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm](http://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm) (accessed 05/04/2011).
- BYRNE R., YOST S. E., KAPLAN B. Mycophenolate mofetil monitoring: is there evidence that it can improve outcomes? *Clin Pharmacol Ther* 2011; **90**:204–206.
- CALINSKI T., HARABASZ J. A dendrite method for cluster analysis. *Commun Stat - Theory Methods* 1974; **3**:1–27 .
- COLLETT D. *Modelling survival data in medical research - 2nd edition*. Chapman & Hall/CRC, London, UK, 2003.
- COMMENGES D. Inference for multi-state models from interval-censored data. *Stat Methods Med Res* 2002; **11**:167–182.
- COOPER J.E., GRALLA J., CAGLE L., GOLDBERG R., CHAN L., WISEMAN A.C. Inferior kidney allograft outcomes in patients with de novo donor-specific antibodies are due to acute rejection episodes. *Transplantation* 2011; **91**:1103–1109.
- COX D.R. Regression models and life tables (with discussion). *J Royal Statistical Soc B* 1972; **34**:187-220.
- EKBERG H., TEDESCO-SILVA H., DEMIRBAS A., et al. Reduced exposure to calcineurin inhibitors in renal transplantation. *N Engl J Med* 2007; **357**:2562–2575.

ERIKSSON B.M., WIRGART B.Z., CLAEISSON K., et al. A prospective study of rapid methods of detecting cytomegalovirus in the blood of renal transplant recipients in relation to patient and graft survival. *Clin Transplant* 1996; **10**:494–502.

ETIENNE M.C., CHATELUT E., PIVOT X., LAVIT M., PUJOL A., CANAL P., et al. Co-variables influencing 5-fluorouracil clearance during continuous venous infusion. A NONMEM analysis. *Eur J Cancer* 1998 ; **34**:92-97.

FOUCHER Y., DAGUIN P., AKL A., et al. A clinical scoring system highly predictive of long-term kidney graft survival. *Kidney Int* 2010; **78**:1288–1294.

FROBEL A.K., KARLSSON M.O., BACKMAN J.T., et al. A Time-to-Event Model for Acute Rejections in Paediatric Renal Transplant Recipients Treated with Ciclosporin A. *Br J Clin Pharmacol* 2013; **76**:603-615.

FRYD D.S., PETERSON P.K., FERGUSON R.M., SIMMONS R.L., BALFOUR H.H., NAJARIAN J.S. Cytomegalovirus as a risk factor in renal transplantation. *Transplantation* 1980; **30**:436–439.

GENOLINI C., FALISSARD B. KmL: k-means for longitudinal data. *Comput Stat* 2010; **25**:317–328.

GIESCHKE R., BURGER H.U., REIGNER B., BLESCH K. S., STEIMER J.L. Population pharmacokinetics and concentration-effect relationships of capecitabine metabolites in colorectal cancer patients. *Br J Clin Pharmacol* 2003; **55**:252–263.

GOMPERTZ B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philos Trans R Soc Lond* 1825; **115**:513–583.

HØJSGAARD S., HALEKOH U., YAN J. The R Package geepack for Generalized Estimating Equations. *J Stat Softw* 2006; **15**:1—11.

HUTCHINS L. F., UNGER J. M., CROWLEY J. J., COLTMAN C. A., ALBAIN K. S. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med* 1999; **341**: 2061–2067.

ISHWARAN H., KOGALUR U. B., BLACKSTONE E. H., LAUER M. S. Random survival forests. *Ann Appl Stat* 2008; **2**:841–860.

JACQMIN-GADDA H., THIÉBAUT R., DARTIGUES J.F. Joint modeling of quantitative longitudinal data and censored survival time. *Rev Epidemiol Santé Publique* 2004; **52**:502–510.

JACQMIN-GADDA H., PROUST-LIMA C., TAYLOR J. M. G., COMMENGES D. Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics* 2010; **66**:11–19.

JAIN A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 2010; **31**:651-666.

KALBFLEISCH J. D., PRENTICE R. L. *The statistical analysis of failure time data - 2nd edition*. John Wiley & sons, New Jersey, USA, 2002.

KANJ S. S., SHARARA A. I., CLAVIEN P. A., HAMILTON J. D. Cytomegalovirus infection following liver transplantation: review of the literature. *Clin. Infect Dis* 1996; **22**:537–549.

KAPLAN E. L., MEIER P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* 1958; **53**:457–481.

LE MEUR Y., BÜCHLER M., THIERRY A., et al. Individualized mycophenolate mofetil dosing based on drug exposure significantly improves patient outcomes after renal transplantation. *Am J Transplant* 2007; **7**:2496–2503.

LE PAGE A. K., JAGER M. M., KOTTON C. N., SIMOONS-SMIT A., RAWLINSON W. D. International survey of cytomegalovirus management in solid organ transplantation after the publication of consensus guidelines. *Transplantation* 2013; **95**:1455–1460.

LETRILLIART L. Les enquêtes cas-temoins : quand, comment? *Sang Thrombose Vaisseaux* 1998; **10**:116-122.

LEWIS J.H., KILGORE M.L., GOLDMAN D.P., et al. Participation of patients 65 years of age or older in cancer clinical trials. *J Clin Oncol* 2003; **21**:1383–1389.

LIANG K.Y., ZEGER S. L. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22 .

LIN H., TURNBULL B.W., MCCULLOCH C.E., SLATE E.H. Latent class models for joint analysis of longitudinal biomarker and event process data : application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* 2002; **97**:53-65.

LITTLE R. J. A., RUBIN D. B. *Statistical analysis with missing data - 2nd edition*. John Wiley & sons, New York, USA, 2002.

NYBERG S.L., MATAS A.J., KREMERS W.K., et al. Improved scoring system to assess adult donors for cadaver renal transplantation. *Am J Transplant* 2003; **3**:715–721.

NYBERG S.L., BASKIN-BEY E.S., KREMERS W., PRIETO M., HENRY M.L., STEGALL M.D. Improving the prediction of donor kidney quality: deceased donor score and resistive indices. *Transplantation* 2005; **80**:925–929.

PLANCHET F. *Modèles de durée - Introduction - Support de cours 2013-2014 - Version 3.4* (2013). Available at: Url: [http://www.ressources-actuarielles.net/EXT/ISFA/fp-isfa.nsf/0/1430AD6748CE3AFFC1256F130067B88E/\\$FILE/Seance1.pdf?OpenElement](http://www.ressources-actuarielles.net/EXT/ISFA/fp-isfa.nsf/0/1430AD6748CE3AFFC1256F130067B88E/$FILE/Seance1.pdf?OpenElement) (accessed 15/09/2013).

POOLE C., GARDINER J., TWELVES C., JOHNSTON P., HARPER P., CASSIDY J., et al. Effect of renal impairment on the pharmacokinetics and tolerability of capecitabine (Xeloda) in cancer patients. *Cancer Chemother Pharmacol* 2002; **49**:225-234.

PRÉMAUD A., LE MEUR Y., DEBORD J., et al. Maximum a posteriori bayesian estimation of mycophenolic acid pharmacokinetics in renal transplant recipients at different postgrafting periods. *Ther. Drug Monit* 2005; **27**: 354–361.

PROUST-LIMA C., AMIEVA H., DARTIGUES J.F., JACQMIN-GADDA H. Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. *American Journal of Epidemiology* 2007, **165**:344-350.

REISCHIG T. Cytomegalovirus-associated renal allograft rejection: new challenges for antiviral preventive strategies. *Expert Rev Anti Infect Ther* 2010; **8**:903–910.

RIZOPOULOS J. JM: an R package for the joint modeling of longitudinal and time-to-event data. *J Stat Softw* 2010; **35**:1-33 .

SAGEDAL S., NORDAL K.P., HARTMANN A., et al. The impact of cytomegalovirus infection and disease on rejection episodes in renal allograft recipients. *Am J Transplant* 2002; **2**:850–856.

SAGEDAL S., HARTMANN A., NORDAL K.P., et al. Impact of early cytomegalovirus infection and disease on long-term recipient and kidney graft survival. *Kidney Int* 2004; **66**:329–337.

SAINT-MARCOUX F., VANDIERDONCK S., PRÉMAUD A., DEBORD J., ROUSSEAU A., MARQUET P. Large scale analysis of routine dose adjustments of mycophenolate mofetil based on global exposure in renal transplant patients. *Ther Drug Monit* 2011; **33**:285–294.

STRATTA R.J., SHAEFFER M.S., MARKIN R.S., et al. Cytomegalovirus infection and disease after liver transplantation. An overview. *Dig Dis Sci* 1992; **37**:673–688.

TERRET C., ERDOCIAN E., GUIMBAUD R., BOISDRON-CELLE M., MCLEOD H.L., FÉTY-DEPORTE R. et al. Dose and time dependencies of 5-fluorouracil pharmacokinetics. *Clin Pharmacol Ther* 2000; **68**:270-279.

THERNEAU T. M., GRAMBSCH P. M. *Modeling survival data: extending the Cox model*. Statistics for biology and health - Springer, 2000.

TIMSIT J. F., ALBERTI C., CHEVRET S. Cox proportional hazards regression analysis. *Rev Mal Respir* 2005; **22**:1058–1064.

TORSTEN H., KURT H., ACHIM Z. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat* 2006; **15**:651-674.

TOUCHETTE E., DOMINIQUE P., SEGUIN J.R., BOIVIN M., TREMBLAY R.E., MONTPLAISIR J.Y. Associations Between Sleep Duration Patterns and Behavioral/Cognitive Functioning at School Entry. *Sleep* 2007 ;**30**:1213–1219.

TSIATIS A.A., DAVIDIAN M. Joint modeling of longitudinal and time-to-event data: an overview. *Clin Statistica Sinica* 2004;**14**:809–834.

URIEN S., REZAÍ K., LOKIEC F. Pharmacokinetic modelling of 5-FU production from capecitabine--a population study in 40 adult patients with metastatic cancer. *J Pharmacokinet Pharmacodyn* 2005; **32**:817–833.

VAN GELDER T., HILBRANDS L.B., VANRENTERGHEM Y., et al. A randomized double-blind, multicenter plasma concentration controlled study of the safety and efficacy of oral mycophenolate mofetil for the prevention of acute rejection after kidney transplantation. *Transplantation* 1999; **68**:261–266.

VAN GELDER T., SILVA H.T., DE FIJTER J.W., et al. Comparing mycophenolate mofetil regimens for de novo renal transplant recipients: the fixed-dose concentration-controlled trial. *Transplantation* 2008; **86**:1043–1051.

VAN GELDER T. Therapeutic Drug Monitoring for Mycophenolic Acid Is Value for (Little) Money. *Clin Pharmacol Ther* 2011; **90**:203–204 .

VAN HEST R. M., VAN GELDER T., BOUW R., et al. Time-dependent clearance of mycophenolic acid in renal transplant recipients. *Br J Clin Pharmacol* 2007; **63**:741–752 .

VANRENTERGHEM Y., PONTICELLI C., MORALES J.M., et al. Prevalence and management of anemia in renal transplant recipients: a European survey. *Am J Transplant* 2003; **3**:835–845.

WEIBULL W. A statistical distribution function of wide applicability. *Journal of Applied Mechanics* 1951; **18**:292-297.

WULFSOHN M.S., TSIATIS A.A. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.

ZHOU Z., SHEN J., HONG Y., KAUL S., PFISTER M., ROY A. Time-varying belatacept exposure and its relationship to efficacy/safety responses in kidney-transplant recipients. *Clin Pharmacol Ther* 2012;**92**:251–257.

Annexes

Annexe 1 :

Code Nonmem du modèle de survie paramétrique utilisé pour prédire le critère d'efficacité dans le manuscrit 1(Chapitre 2, section II).

```
$PROBLEM TTE MODEL
$INPUT ID CODE TIME DVID MDV DV E01 EMAX EC50 GAM E02 E2 K2
        E03 E3 K3 CMV
$DATA Base-totale.csv IGNORE @
$SUBROUTINE ADVAN=6 TOL=6
$MODEL COMP=(CUMHAZ)
$OMEGA 0 FIX ; ETA1
$PK
```

```
IF(NEWIND.LE.1) CHLAST=0
BASHAZ=THETA(1)*EXP(ETA(1))
BETA1=THETA(2)
BETA2=THETA(3)
BETA3=THETA(4)
BETA4=THETA(5)
```

```
$DES
```

```
DPRG1=E01+(EMAX*(T**GAM))/((EC50**GAM)+(T**GAM))
DPRG2=E2+E02*EXP(-K2*T)
DPRG3=E3+E03*EXP(-K3*T)
IF (CODE.LE.348) THEN
    CNI=0
ELSE
    CNI=1
ENDIF
```

```
DADT(1)=BASHAZ*EXP((BETA1*DPRG1)+(BETA2*DPRG2*(1-
CNI)))+(BETA3*DPRG3*CNI)+(BETA4*CMV))
```

```
$ERROR
```

```
CUMHAZ=A(1)
DISPRG1=E01+(EMAX*(TIME**GAM))/((EC50**GAM)+(TIME**GAM))
DISPRG2=E2+E02*EXP(-K2*TIME)
DISPRG3=E3+E03*EXP(-K3*TIME)
IF (CODE.LE.348) THEN
    CNIS=0
ELSE
    CNIS=1
ENDIF
```

```
HAZARD=BASHAZ*EXP((BETA1*DISPRG1)+(BETA2*DISPRG2*(1-
CNIS)))+(BETA3*DISPRG3*CNIS)+(BETA4*CMV))
```

```
IF (DVID.EQ.2.AND.DV.EQ.0) THEN
  Y=EXP(-CUMHAZ)
  CHLAST=CUMHAZ
ELSE
  CHLAST=CHLAST
ENDIF
```

```
IF (DVID.EQ.2.AND.DV.EQ.2) THEN
  Y=1-EXP(-(CUMHAZ-CHLAST))
ENDIF
```

```
$THETA (0,0.01,10) ; BASHAZ
```

```
-0.01 ; BETA1
```

```
0.0001 ; BETA2
```

```
0.01 ; BETA3
```

```
2.39 ; BETA4
```

```
$ESTIMATION NSIG=4 PRINT=1 NOABORT MAX=9999 METHOD=0 LIKE
$TABLE ID TIME DV DVID CUMHAZ HAZARD ONEHEADER NOPRINT
FILE=table.fit
```

Annexe 2 :

Code Nonmem du modèle de régression logistique utilisé pour prédire le risque des infections et maladies à CMV dans le manuscrit 1 (Chapitre II, section II)

```
$PROBLEM LOGISTIC MODEL
$INPUT ID CODE TIME DV STAT
$DATA Data-cmv.csv IGNORE @
$OMEGA 0.01 ; ETA(1)
$PRED

BASE=THETA(1)
BETA2=0
BETA3=0
BETA4=0
IF(STAT.EQ.2) BETA2=THETA(2)
IF(STAT.EQ.3) BETA3=THETA(3)
IF(STAT.EQ.4) BETA4=THETA(4)

A=EXP(BASE+BETA2+BETA3+BETA4+ETA(1))
P1=A/(1+A)

IF (DV.EQ.1) THEN; probability to have CMV infection or disease
  Y=P1
ENDIF

IF (DV.EQ.0) THEN; probability to not have CMV infection or disease
  Y=1-P1
ENDIF

$THETA -0.1 ; BASE
0.01 ; BETA2
0.01 ; BETA3
0.01 ; BETA4
$ESTIMATION NSIG=4 PRINT=1 MAX=9900 METHOD=COND LAPLACE LIKE
MSFO=msfb1

$TABLE ID TIME DV NOPRINT ONEHEADER FILE=sdtab1
$TABLE ID MPA CSA TAC PROP STAT NOPRINT ONEHEADER FILE=cotab1
$TABLE ID BASE BETA2 BETA3 BETA4 NOPRINT ONEHEADER FILE=patab1
```

Annexe 3 :

Code Nonmem du modèle de Markov utilisé pour prédire la survenue d'anémie.

```
$PROBLEM MARKOV MODEL
$INPUT ID TIME PDV DV MPA BRAS AGER AGED SEX HBMI
$DATA Data-anemie.csv IGNORE @
$OMEGA 0.1 ; ETA(1)
0.1 ; ETA(2)
$PRED

BASE0=THETA(1)
BETA10=THETA(2)
BETA20=THETA(3)
BASE1=THETA(4)
BETA11=THETA(5)
BETA21=THETA(6)

A0=EXP(BASE0+(BETA10*MPA)+(BETA20*HBMI)+ETA(1))
A1=EXP(BASE1+(BETA11*MPA)+(BETA21*HBMI)+ETA(2))
P0=A0/(1+A0)
P1=A1/(1+A1)

IF (PDV.EQ.0.AND.DV.EQ.1) Y=P0
IF (PDV.EQ.0.AND.DV.EQ.0) Y=1-P0
IF (PDV.EQ.1.AND.DV.EQ.0) Y=P1
IF (PDV.EQ.1.AND.DV.EQ.1) Y=1-P1

$THETA 0.1 ; BASE0
(0,0.09) ; BETA10
(-1,-0.001) ; BETA20
0.1 ; BASE1
(0,0.09) ; BETA11
(-1,-0.001) ; BETA21
$ESTIMATION NSIG=4 PRINT=1 MAX=9900 METHOD=COND LAPLACE LIKE
MSFO=msfb3
$COVARIANCE
$TABLE ID TIME PDV NOPRINT ONEHEADER FILE=sdtab3
$TABLE ID PDV NOPRINT ONEHEADER FILE=cotab3
$TABLE ID PDV NOPRINT ONEHEADER FILE=patab3
```

Annexe 4 :

Code Nonmem du modèle développé par Urien et al (2005) et utilisé pour prédire la pharmacocinétique de la capécitabine et ses métabolites.

```
$PROBLEM CAPECITABINE N40 TLAG
$ABBREVIATED DERIV2=NO
$INPUT ID IDO DATE TIME DV AMT IDV MDV STDY AGE X BSA BWT BILT
$DATA base-finale.csv IGNORE=#
$OMEGA 0 FIX ; OKA
$OMEGA 0.234 ; OV1
$OMEGA 0.126 ; Ocl
$OMEGA 0 FIX ; Ocl12
$OMEGA .449 ; ok23
$OMEGA .159 ; oK34
$OMEGA .165 ; ok40
$OMEGA 0.462 ; tlag
$OMEGA BLOCK(1)
1.52 ; EOC1
$OMEGA BLOCK(1) SAME

$PRED

IF (MDV.EQ.1) THEN
  DS =AMT
  T0 =TIME
ENDIF
IF (DATE.EQ.1) EOC1=ETA(9)
IF (DATE.GT.25) EOC1=ETA(10)

KA=THETA(1)*EXP(ETA(1)+EOC1)

V1 =THETA(2)*EXP(ETA(2))

CL10=THETA(3)*EXP(ETA(3))

CL12= THETA(4)*EXP(ETA(4))

K23=THETA(5)*EXP(ETA(5))

K34 =THETA(6)*EXP(ETA(6))

TVK40 =THETA(7)
K40 =(TVK40*((AGE/65.5)**THETA(9)))*EXP(ETA(7))

TLAG =THETA(8)*EXP(ETA(8))
T1 = TIME - T0 - TLAG
V2=1;;180
V3=1;;90
V4=1;;20
K10 = CL10/V1
K12 = CL12/V1
E11 =K10+K12
; S11=1 SA1=1 S11=1 S20=1 S30=1 S40=1
IF (T1.GT.0) THEN
  A01 =DS*KA
  DN1 =KA - E11
  F11 =EXP(-E11*T1)
  FAA =EXP(-KA *T1)
```

```

ENDIF

IF (T1.GT.0.AND.IDV.EQ.1) F=A01*(F11 - FAA)/DN1/V1
IF (T1.GT.0.AND.IDV.EQ.2) THEN
  FM0=A01*K12
  CMA=FAA/(E11-KA)/(K23-KA)
  C11=F11/(KA-E11)/(K23-E11)
  C22=EXP(-K23*T1)/(KA-K23)/(E11-K23)
  F =FM0*(CMA+C11+C22)/V2
ENDIF
IF (T1.GT.0.AND.IDV.EQ.3) THEN
  FM0=A01*K12*K23
  CMA=FAA/(E11-KA)/(K23-KA)/(K34-KA)
  C11=F11/(KA-E11)/(K23-E11)/(K34-E11)
  C22=EXP(-K23*T1)/(KA-K23)/(E11-K23)/(K34-K23)
  C33=EXP(-K34*T1)/(KA-K34)/(E11-K34)/(K23-K34)
  F =FM0*(CMA+C11+C22+C33)/V3
ENDIF
IF (T1.GT.0.AND.IDV.EQ.4) THEN
  FM0=A01*K12*K23*K34
  CMA=FAA/(E11-KA)/(K23-KA)/(K34-KA)/(K40-KA)
  C11=F11/(KA-E11)/(K23-E11)/(K34-E11)/(K40-E11)
  C22=EXP(-K23*T1)/(KA-K23)/(E11-K23)/(K34-K23)/(K40-K23)
  C33=EXP(-K34*T1)/(KA-K34)/(E11-K34)/(K23-K34)/(K40-K34)
  C44=EXP(-K40*T1)/(KA-K40)/(E11-K40)/(K23-K40)/(K34-K40)
  F =FM0*(CMA+C11+C22+C33+C44)/V4
ENDIF
IPRED = F
IRES=DV-IPRED
IF (IPRED.GT.0) IWRES=IRES/IPRED
IF (IDV.EQ.1) Y = F+EPS(1)
IF (IDV.EQ.2) Y = F+EPS(2)
IF (IDV.EQ.3) Y = F+EPS(3)
IF (IDV.EQ.4) Y = F+EPS(4)

$THETA (0,1.06) ; KA
$THETA (0,246) ; V1
$THETA (0,242) ; cl10
$THETA (0,11.4) ; cl12
$THETA (0,9.93) ; k23
$THETA (0,4.77) ; k34
$THETA (0,61.2) ; k40
$THETA (0,0.29) ; tlag
$THETA (0,0.1) ; AGE sur K40
$$SIGMA BLOCK(2)
14.67 ; P1
3.1 13.83 ; P2
$$SIGMA BLOCK(2)
33.75 ; P3
2.6 .41 ; P4
$ESTIMATION METHOD=0 MAXEVAL=9999 POSTHOC PRINT=10
$COVARIANCE MATRIX=S
$TABLE ID T1 IDV IPRED IRES IWRES STDY ONEHEADER NOPRINT FILE=r01
$TABLE ID T1 KA V1 CL10 CL12 K23 K34 K40 STDY ONEHEADER FIRSTONLY
NOPRINT FILE=c

```