

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE MONTPELLIER 1

FORMATION DOCTORALE : BIostatistique
ECOLE DOCTORALE : INFORMATION, STRUCTURES, SYSTÈMES

Thèse présentée et soutenue publiquement par

VANESSA ROUSSEAU

en vue d'obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER 1

MODÈLES FRAGILISÉS DE SURVIE RELATIVE APPLIQUÉS AUX DONNÉES CENSURÉES DANS LE CAS DE LA TRANSPLANTATION RÉNALE.

Thèse soutenue le Vendredi 10 Octobre 2008 devant le jury composé de :

M.	JEAN-PIERRE DAURÈS	PU-PH, Université de Montpellier 1	(Directeur)
M.	GILLES DUCHARME	PU, Université de Montpellier 2	(Examineur)
M.	ROCH GIORGI	MCU-PH, Université d'Aix-Marseille 2	(Rapporteur)
Mme.	MAGALI GIRAL	PH, CHU de Nantes	(Examineur)
Mme.	CATHERINE HUBER	PU, Université de Paris 5	(Examineur)
M.	PAUL LANDAIS	PU-PH, Université de Paris 5	(Rapporteur)

*“La véritable science enseigne, par dessus tout,
à douter et à être ignorant.”
Miguel de Unamuno (le sentiment tragique de la vie).*

REMERCIEMENTS

JE tiens tout d'abord à remercier, Mr Jean-Pierre Daurès, sans qui, cette thèse n'aurait pu avoir lieu. Son professionnalisme ainsi que son sens de l'écoute et de la compréhension font de lui un homme singulier. La confiance qu'il m'a accordée pendant ces trois années m'a encouragée à accomplir cette thèse et ses connaissances et conseils au quotidien m'ont été plus que bénéfiques. Je lui dois beaucoup.

Je remercie les professeurs Gilles Ducharme, Roch Giorgi, Magali Giral, Catherine Huber et Paul Landais d'avoir accepté de composer le jury et d'examiner ce travail.

Un grand merci à Christel, Christelle et Yohann, les relecteurs personnels de mon mémoire. Durant ces trois années, ils se sont toujours montrés disponibles et m'ont toujours apporté aides et précieux conseils, je les remercie profondément. Parmi eux, Christel, ce quotidien, ces longues discussions et ces nombreuses questions sans réponse... Il y a aussi Yohan et ses virées journalières dans le bureau, Christophe et sa conception particulière de la tarte aux pommes. Je salue aussi l'équipe que j'ai cotoyée pendant ces trois années, plus particulièrement Muriel pour sa bonne humeur innée. Je n'oublie pas le dévouement et l'efficacité de Sylvie, surtout lors des derniers instants.

Ces années à Montpellier ont commencé par la rencontre avec Mr K et la Carotte et je resterai persuadée que ce premier tournant fut le bon. En chemin et pour longtemps j'espère, Marie et Paul... Cette thèse m'aura permis de découvrir de nouveaux amis mais je n'oublie pas pour autant les anciens. Comment oublier les massiens ? Alex, Axelle, Isa, Juju, Lolo, Nico et ma Jaja bien sûr... Nul ne se doute mais le choix de ma carrière n'a pas été évident, seules Amélie et Sylvie vous le diront... Ma découverte de la Biostatistique fut accompagnée de Sophie et d'une Schtroumpfette porte-bonheur.

Mes derniers remerciements s'adressent à ma famille. Tout d'abord, je remercie ma p'tite maman pour la confiance qu'elle m'a toujours accordée. Ensuite, même avec l'éloignement, je n'oublie pas Stéphanie, ma soeur. Et puis, une autre soeur, un peu particulière, qui a été présente lors de mes veillées estudiantines, Danette. Je remercie aussi François, l'évolution de mon cursus scolaire dépend aussi de lui. Je tiens à faire un détour par la capitale pour remercier le soutien de ma famille "trop éloignée" : mon père, Zabeth, Jimmy, Jess, ainsi que les deux mamies. Par la même occasion, je ne serais jamais assez reconnaissante de Biloute et

Biquet pour leur accueil durant mon épopée parisienne et pour tout le reste...

Je dédie naturellement ces dernières lignes à mes grands-parents maternels...

TABLE DES MATIÈRES

TABLE DES MATIÈRES	vii
LISTE DES FIGURES	ix
INTRODUCTION	1
1 DONNÉES ET PROBLÉMATIQUES CLINIQUES	5
1.1 CONTEXTE CLINIQUE	7
1.2 DONNÉES	8
1.3 STATISTIQUES DESCRIPTIVES ET ÉLÉMENTAIRES	11
1.3.1 Statistiques générales	12
1.3.2 Statistiques par centre de transplantation	13
1.3.3 Statistiques par année de transplantation	15
1.3.4 Corrélation entre les covariables	15
1.4 PROBLÉMATIQUES CLINIQUES	16
DISCUSSION-CONCLUSION	18
2 POUR LA SURVIE...	19
2.1 PREMIÈRES DÉFINITIONS	21
2.1.1 Les fonctions de survie	21
2.1.2 Les mécanismes de censure	22
2.2 LES MODÈLES CLASSIQUES DE SURVIE	22
2.2.1 Modélisation non-paramétrique	22
2.2.2 Modélisation paramétrique	23
2.2.3 Modélisation semi-paramétrique	24
2.3 MODÈLE STRATIFIÉ	25
2.4 STRATÉGIE DE SÉLECTION DES COVARIABLES	25
2.5 APPLICATION AUX TRANSPLANTÉS RÉNAUX	26
2.5.1 Survie non paramétrique	26
2.5.2 Modèle paramétrique	27
2.5.3 Modèles de Cox	31
DISCUSSION-CONCLUSION	34
3 LA FRAGILITÉ	35
3.1 MODÈLES DE FRAGILITÉ "UNIVARIÉS"	39
3.2 MODÈLES DE FRAGILITÉ "MULTIVARIÉS"	41
3.2.1 Le modèle de fragilité partagée	42
3.2.2 Le modèle de fragilité corrélée	44
3.3 LE MODÈLE DE FRAGILITÉ PROPOSÉ	45
3.4 APPLICATION AUX TRANSPLANTÉS RÉNAUX	48
3.4.1 Modèle de Cox avec fragilité	48

3.4.2	Modèle de régression de Weibull avec fragilité	50
	DISCUSSION-CONCLUSION	55
4	LA SURVIE RELATIVE	57
4.1	DÉFINITION DES DIFFÉRENTES NOTIONS DE SURVIE	59
4.2	MÉTHODE D'ESTÈVE <i>et al.</i>	60
4.3	MODÈLE D'HAKULINEN ET TENKANEN	62
4.4	MODÈLE D'ANDERSEN <i>et al.</i>	63
4.5	TABLE DE MORTALITÉ	64
4.5.1	Définition et présentation	64
4.5.2	Analyse des tables vitales	65
4.6	CHOIX D'UN MODÈLE DE SURVIE RELATIVE	66
4.7	APPLICATION AUX TRANSPLANTÉS RÉNAUX	68
4.7.1	Résultats avec le modèle d'Andersen	68
4.7.2	Résultats avec le modèle paramétrique	71
	DISCUSSION-CONCLUSION	74
5	SURVIE RELATIVE MULTIPLICATIVE AVEC FRAGILITÉ	77
5.1	LE MODÈLE	79
5.2	APPLICATION AUX TRANSPLANTÉS RÉNAUX	80
5.2.1	Fragilité sur les centres de transplantation	81
5.2.2	Fragilité sur les années de transplantation	83
	DISCUSSION-CONCLUSION	86
6	PRISE EN COMPTE DE LA CENSURE INFORMATIVE	89
6.1	QU'EST-CE QU'UN MODÈLE SEMI-MARKOVIEN ?	93
6.2	LE MODÈLE PROPOSÉ	94
6.2.1	Notations et hypothèses	94
6.2.2	Introduction des covariables	96
6.2.3	Ecriture de la vraisemblance	96
6.3	TEST DE L'HYPOTHÈSE D'HOMOGENÉITÉ	98
6.3.1	La statistique de test	98
6.4	APPLICATION AUX TRANSPLANTÉS RÉNAUX	100
6.4.1	Stratégie de sélection des covariables	101
6.4.2	Estimation des paramètres du modèle	101
6.4.3	Test de l'hypothèse d'homogénéité	107
	DISCUSSION-CONCLUSION	113
	DISCUSSION-CONCLUSION-PERSPECTIVES	117
A	ANNEXES	123
A.1	DÉMONSTRATIONS AVEC LA FRAGILITÉ	125
A.2	TABLEAU D'EFFECTIFS SELON LES ANNÉES DE GREFFE	130
	BIBLIOGRAPHIE	133

LISTE DES FIGURES

2.1	Survie.	26
2.2	Survies Kaplan Meier par centre.	27
2.3	Survies Weibull et Exponentiel par centre.	28
2.4	Hypothèse de proportionnalité (modèle de Weibull).	29
2.5	Hypothèse de proportionnalité (modèle de Weibull suite).	30
2.6	Fonction de Risque selon les covariables (modèle de Weibull).	31
3.1	Risques attendu et naturel.	40
3.2	Risque selon les covariables (Weibull fragilisé sur les centres).	52
3.3	Risque selon les covariables (Weibull fragilisé sur les années).	54
4.1	Survie attendue par âge, sexe et année.	67
4.2	Ratio des λ_{Obs} d'Andersen Femme Vs Homme.	72
4.3	Graphique fonction de risque relative selon le DGF (Weibull).	74
5.1	Risque selon les covariables (modèle MSRF sur les centres).	82
5.2	Modèle MSRF sur les centres : RR selon le sexe.	83
5.3	Risque selon les covariables (modèle MSRF sur les années).	85
5.4	Modèle MSRF sur les années : RR selon le sexe.	86
6.1	Modèle à trois états.	95
6.2	Schéma des trois états possibles pour chaque patient.	100
6.3	Modèle semi-markovien : Risque selon les covariables.	104
6.4	Modèle semi-markovien avec survie relative (DGF).	106
6.5	Modèle semi-markovien avec survie relative (sexe).	107

INTRODUCTION

DANS le domaine médical, l'intérêt porté à la connaissance des pathologies par le biais des analyses statistiques est grandissant. De nombreuses méthodes existent et sont largement utilisées. Les domaines d'application sont très divers : essais thérapeutiques, registres de populations, épidémiologie descriptive ou analytique, ... Ainsi, le type de données étudiées ne se limite pas à l'événement terminal d'une pathologie létale mais peut s'étendre à tout autre événement faisant intervenir le temps. Il peut s'agir de la récurrence ou de la guérison d'une pathologie par exemple.

Les premières analyses s'intéressant à une durée de vie remonteraient au XVII^{ème} siècle. Depuis, les méthodes concernant les données de survie ne cessent d'évoluer afin de prendre en compte leurs différentes caractéristiques et les nombreuses problématiques qu'elles engendrent. Le modèle le plus connu et le plus utilisé en analyse de survie est le modèle de Cox (20) apparu au début des années 1970. Il possède la particularité de prendre en compte simultanément plusieurs variables explicatives pour expliquer la survie sans faire d'hypothèse sur la distribution du temps de survie. Sa première approche fut l'analyse d'un événement unique mais il s'est étendu à l'analyse de plusieurs événements par l'ajout de termes aléatoires. Par ailleurs, les effets aléatoires peuvent aussi être utilisés pour modéliser des facteurs cachés. En effet, une limite de la plupart des modèles de survie est qu'ils supposent que la population étudiée est homogène. Cela suppose que le risque de décès est le même pour tous les individus de cette population. Cependant ce n'est pas toujours le cas, les individus ont des propensions différentes sous l'effet d'un traitement ou l'influence de certaines covariables. Dans de nombreuses situations, il est impossible de mesurer toutes les covariables pertinentes liées à la pathologie d'intérêt et donc d'introduire ces covariables comme effets fixes dans le modèle. Les raisons sont souvent économiques ou liées à la méconnaissance de l'importance de certaines covariables sur la pathologie étudiée. Par ailleurs, lorsque l'étude porte, par exemple, sur des patients issus de même famille ou un événement à répétition sur un même patient, la présence d'un facteur caché liant les unités statistiques d'un même groupe peut être suspectée.

Ainsi, l'approche de la fragilité est un concept de modélisation statistique qui vise à représenter l'hétérogénéité dans une population étudiée provoquée par les covariables non mesurées. Fondamentalement, le concept de la fragilité est apparu au début des années 1920 dans les travaux de Greenwood et Yule (31) sur la "prédisposition à l'accident". Ce n'est qu'en 1979 que le terme de fragilité lui-même a été présenté par Vaupel *et al.*

(57) dans les modèles de survie. Cependant, en 1978, sans utiliser le terme de fragilité, Clayton (19) avait déjà largement promu son concept par son application à des données de familles. En termes statistiques, un modèle de fragilité est un modèle à effet aléatoire dans lequel celui-ci a un effet multiplicatif sur la fonction de risque de base. On peut distinguer deux grandes catégories de modèles de fragilité : les univariés et les multivariés. Dans le premier cas, l'effet aléatoire concerne l'individu et il est utilisé pour décrire l'influence des covariables non observées. L'idée principale est que les individus ont des fragilités différentes et que les individus les plus fragiles vont décéder plus tôt que les autres. Le second cas considère que l'effet aléatoire porte sur le groupe d'individus. Cet effet modélise la dépendance entre les temps d'évènements d'individus appartenant à un même groupe. Il peut s'agir, par exemple, des données de survie de patients issus de même famille, de jumeaux ou d'évènements récurrents sur un même patient. Cette dépendance non mesurable est souvent causée par des facteurs environnementaux ou génétiques. La fragilité explique la dépendance dans le sens où si elle est connue, alors les évènements sont indépendants. En d'autres termes, les données de survie d'un même groupe sont indépendantes conditionnellement à la fragilité de ce groupe.

Par ailleurs, dans la plupart des cohortes, l'information concernant la cause de décès des patients est mal connue. Soit cette information n'est pas connue du tout, ceci car le certificat de décès est incomplet ou l'autopsie n'a pas été faite, soit cette information est mal renseignée. Dans le dernier cas, ceci provient d'un problème souvent délicat qui est celui de la connaissance de la véritable cause de décès du patient. En effet, prenons l'exemple d'une personne décédant dans un accident de la route. Cette personne peut soit être décédée à cause de l'accident lui-même soit être décédée, par exemple, à cause d'un accident vasculaire cérébral au moment où elle conduisait. Il est donc difficile, dans certains cas, de classer les décès. Ainsi, un des problèmes majeurs en analyse de survie est de pouvoir mesurer la part des décès imputables à la pathologie étudiée. Dans le cadre de la transplantation rénale par exemple, les décès liés à un cancer peuvent être d'origine indépendante de la transplantation ou liés aux traitements immunosuppresseurs. Ces deux causes peuvent même être imbriquées. Ainsi, une solution pour remédier à ce problème de classification des décès est de recourir aux modèles de survie relative. La notion de survie relative a été introduite au début des années 1940 par Berkson (8). La survie relative se définit comme le ratio entre la survie observée d'une population étudiée et la survie attendue d'une population de référence. La survie attendue d'une population de référence peut être, par exemple, la survie de la population nationale française. Chaque individu de la population étudiée peut être comparé à un individu ayant les mêmes caractéristiques que lui (sexe, âge, etc...) dans la population de référence. Ainsi, l'intérêt de cette méthode est qu'elle donne une mesure de survie nette, c'est-à-dire l'excès de mortalité due à la maladie étudiée. Pour pouvoir estimer la survie nette, la cause spécifique du décès étudié est supposée indépendante des autres causes de décès, mais elle n'est pas connue. Classiquement, lorsque les causes de décès sont connues, il est possible d'étudier la mortalité spécifique en considérant soit un modèle

de survie où les décès indépendants de la maladie sont censurés (20), soit un modèle à risques compétitifs où les décès peuvent être distingués selon leur cause (30). Le principal avantage de la survie relative est qu'elle permet d'estimer la survie spécifique à la maladie étudiée sans nécessité de connaître l'information sur la cause de décès.

De plus, la plupart des analyses de survie font généralement l'hypothèse d'indépendance entre le délai de survenue de l'évènement et le délai de survenue de la censure. Cette hypothèse d'indépendance entre la censure et l'évènement est une hypothèse forte qui induit un biais dans l'estimation de la survie si elle n'est pas respectée. En analyse de survie, plusieurs cas de censure sont possibles. Tout d'abord, les patients perdus de vue sont ceux dont l'état n'est pas connu à la date de point. Ils représentent alors une perte d'information et sont source de biais. Dans le cas des patients VIH, par exemple, ceux qui abandonnent le suivi sont généralement les cas les plus graves (50). En effet, le traitement est souvent lourd avec des effets secondaires multiples et provoque ainsi des arrêts de suivi. De plus, il a été observé que ce sont les patients les plus malades qui arrêtent leurs traitements soit parce qu'ils ne supportent plus la thérapie soit parce qu'ils n'ont plus d'espoir. Dans le cas de l'asthme, au contraire, il semble que les patients perdus de vue soient des patients bien contrôlés qui ne ressentent pas le besoin de venir consulter (16). Dans ces deux exemples, le processus de censure est clairement lié à l'évènement étudié puisqu'il apporte une information sur cet évènement. D'autre part, l'autre cas de censure, la censure administrative, est seulement engendrée par la fin de l'étude. Cette censure est indépendante de l'évènement et est alors considérée comme non-informative.

L'objectif de cette thèse est de prendre en compte les trois problématiques rencontrées précédemment dans les modèles de survie. Ceci s'illustrera au travers d'une base de données concernant des transplantés rénaux. La thèse s'organisera de la façon suivante.

Le premier chapitre présente la base de données : il s'agit de la base de données DIVAT (Données Informatisées et VALidées en Transplantation). Tout d'abord, comme nous nous intéressons aux transplantations rénales, le contexte clinique de ce domaine est donné. Ensuite, la base est présentée. Pour avoir un premier aperçu des données, nous procédons à des statistiques élémentaires sur cette base. Après analyse de ces statistiques et au travers des questions posées par les médecins, plusieurs problématiques sont révélées propre à cette pathologie. Elles seront ainsi détaillées clairement à la fin de ce premier chapitre. Par la suite, nous devons donc les prendre en compte dans les modèles proposés.

Le deuxième chapitre est un rappel des principales notions de survie, des principaux modèles utilisés en analyse de survie. Ce chapitre rappelle les cinq fonctions de survie, le concept de la censure, les méthodes d'estimations des modèles non-paramétrique, semi-paramétrique et paramétrique. Dans la base DIVAT, nous nous intéressons principalement au décès des patients transplantés rénaux ayant leur greffon fonctionnel. Le temps d'intérêt est donc le délai entre la greffe et la survenue du décès. Ainsi, des premières analyses de survie nous permettront d'appréhender l'évène-

ment d'intérêt.

Par ailleurs, dans la base DIVAT, nous possédons les renseignements concernant les patients greffés dans quatre centres différents. Ensuite, la question se pose de savoir s'il y a un effet centre de transplantation, c'est-à-dire si la survie des patients est différente selon les centres dans lesquels ils sont transplantés. Nous estimerons donc un modèle de fragilité dans lequel la fragilité représente la dépendance des patients greffés dans un même centre. De plus, durant la période de suivi de ces patients transplantés (de 1990 à 2006), les stratégies thérapeutiques ont été modifiées. Néanmoins, il est difficile de quantifier ces changements. Nous voulons donc savoir si ces changements thérapeutiques ont eu un effet sur la survie des transplantés rénaux, c'est-à-dire si la survie d'un patient dépend de son année de transplantation. Nous optons donc, ici aussi, pour un modèle de fragilité dans lequel la fragilité modélise la dépendance des patients greffés la même année. Le chapitre 3 sera donc consacré aux modèles de fragilité.

Le chapitre 4 présentera en détails la notion de survie relative ainsi que les modèles les plus cités. De plus, nous proposons un modèle multiplicatif de survie relative paramétrique. Comme les renseignements concernant les causes de décès des patients rénaux ne sont pas disponibles, nous appliquons ces différents modèles de survie relative à la base DIVAT.

Ensuite, il était pertinent de mélanger les deux problématiques précédentes : l'hétérogénéité des individus et le manque d'information concernant les causes des décès des patients. Nous proposons donc un modèle multiplicatif de survie relative avec fragilité dans le chapitre 5.

Enfin, le chapitre 6 aborde la problématique de la censure. Dans la base DIVAT, les deux types censures rencontrées sont la censure administrative engendrée par le gel de la base et la censure provoquée par le rejet de la greffe observé chez les patients. En effet, les patients greffés qui rejettent leur greffon retournent en dialyse. A ce moment-là, ces patients changent de centre de soins, ils ne sont donc plus suivis dans le centre de transplantation d'origine. Les informations concernant les patients rejetant leur greffe ne sont alors plus disponibles. Il y a donc une censure de ces patients. De plus, il semblerait que les patients qui rejettent leur greffon présentent un risque de décéder plus important (60). Nous ne pouvons donc supposer que cette censure soit indépendante de l'évènement décès, ainsi elle est considérée comme informative. Dans ce chapitre, nous proposons un modèle multi-états permettant de prendre en compte cette censure informative.

Pour clore ce mémoire, le dernier chapitre concerne les discussions, conclusions et perspectives du travail élaboré au cours de cette thèse.

DONNÉES ET PROBLÉMATIQUES CLINIQUES



SOMMAIRE

1.1	CONTEXTE CLINIQUE	7
1.2	DONNÉES	8
1.3	STATISTIQUES DESCRIPTIVES ET ÉLÉMENTAIRES	11
1.3.1	Statistiques générales	12
1.3.2	Statistiques par centre de transplantation	13
1.3.3	Statistiques par année de transplantation	15
1.3.4	Corrélation entre les covariables	15
1.4	PROBLÉMATIQUES CLINIQUES	16
	DISCUSSION-CONCLUSION	18

Ce chapitre présente et explore succinctement la base de données qui sert d'appui dans toute cette thèse. Tout d'abord, le contexte clinique concernant le domaine de la transplantation rénale est donné. Ensuite, une présentation générale des données est effectuée. Il s'agit de la base de données nommée DIVAT (Données Informatisées et VALidées en Transplantation). Cette base est nationale et les informations relatives aux transplantés rénaux sont incommensurables. Nous choisissons donc d'explicitier seulement les principales variables, celles utiles pour les diverses études auxquelles nous allons procéder. Ensuite, des statistiques élémentaires sur cette base sont établies afin d'avoir un premier aperçu des données. De cette partie résulte, par exemple, la proportion des décès, celle des rejets de greffe, le nombre d'hommes transplantés, leur âge, leur critère immunologique, etc ... Après analyse de ces statistiques et afin de répondre aux questions des médecins, nous posons clairement les différentes problématiques auxquelles nous allons répondre dans ce mémoire. Un plan explicité de cette thèse est établi à travers les problématiques et la conclusion de ce chapitre.

1.1 CONTEXTE CLINIQUE

Les reins sont aussi importants pour la santé que le coeur, le foie ou les poumons. Ils débarrassent le sang de ses déchets par le biais de l'urine. Ceux-ci régularisent les quantités d'eau et de divers minéraux dont l'organisme a besoin pour se maintenir en santé. Ils produisent aussi des hormones qui contrôlent d'autres fonctions de l'organisme. De nombreux autres organes dépendent des reins pour leur bon fonctionnement.

Il y a insuffisance rénale aiguë (IRA) lorsque les reins cessent soudainement de fonctionner. Dans ce cas, la fonction rénale revient habituellement à la normale, mais il peut être nécessaire d'avoir recours à des traitements de dialyse jusqu'à ce que les reins recommencent à faire leur travail. Il existe plusieurs causes de l'insuffisance rénale aiguë. On parle d'insuffisance rénale chronique (IRC) quand il y a une détérioration lente et progressive de la fonction rénale. Cette détérioration est en général irréversible. L'IRC survient lorsque les néphrons, de minuscules filtres qui débarrassent les reins de divers déchets, cessent de fonctionner. Les néphrons peuvent être endommagés par des maladies comme le diabète et l'hypertension. Si ces maladies sont bien contrôlées, il est possible de retarder et même de prévenir l'insuffisance rénale terminale.

L'insuffisance rénale chronique ne se guérit pas. Dans la phase initiale, il peut suffire, pour ralentir la détérioration des reins, de faire attention à son alimentation, de prendre des médicaments et de bien contrôler sa tension artérielle. Toutefois, quand les reins ne fonctionnent qu'à moins de 10 à 20% de leur capacité normale, il devient nécessaire d'avoir recours à la dialyse ou à une transplantation rénale pour rester en vie. C'est le stade d'insuffisance rénale terminale (IRT).

Comparée aux différentes techniques de dialyse, la transplantation rénale procure non seulement une qualité de vie supérieure mais aussi une espérance de vie plus longue : la survie des patients diabétiques (tant de type 1 que de type 2) greffés d'un rein est en effet supérieure à celle de diabétiques demeurés en dialyse, appariés pour le sexe, l'âge, la durée du diabète et la durée du traitement par hémodialyse jusqu'à la greffe.

Même rare, le décès d'un patient peut survenir à différents moments : lors de l'attente d'une greffe, lors de l'opération chirurgicale, ou après la transplantation avec ou sans rein fonctionnel. Les causes de décès des transplantés avec rein fonctionnel sont : cardiovasculaires, infectieuses, inconnues, autres que liées au problème rénal, tumorales, dues à une attaque cérébrale, à une anomalie digestive, accidentelles ou suicidaires. Un des objectifs de ce mémoire est de tester plusieurs facteurs de risque sur l'évènement "décès des patients greffés rénaux avec leur rein fonctionnel". Ces facteurs, qu'ils soient propres aux donneurs ou aux receveurs, sont nombreux. Les principaux, détaillés par la suite, sont : l'âge et le sexe du receveur et du donneur, l'ischémie froide, le nombre d'incompatibilités HLA, le taux d'immunisation anti HLA, le délai de reprise de fonction du greffon, l'année de la greffe et la maladie initiale. La plupart de ces

facteurs sont plutôt connus pour être explicatifs des rejets de greffon, cependant, au vu des causes de décès avec un rein fonctionnel, ils peuvent aussi être soupçonnés liés au décès.

1.2 DONNÉES

Les données sont extraites de la cohorte DIVAT (Données Informatisées et VALidées en Transplantation), coordonnée par l'Institut de Transplantation et de Recherche en Transplantation (ITERT) du CHU de Nantes. Cet institut est le premier centre français et européen de greffe de reins (171 greffes en 2004). Le recueil des données est effectué en temps réel, regroupant les paramètres du suivi biologique et médical des patients greffés d'un rein. Les CHU de Paris Saint-Louis, Nancy, Montpellier et Toulouse se sont associés à Nantes pour mettre en commun cet outil de recueil et créer une base de données unique fonctionnant en réseau, en temps réel via Internet et selon la même méthodologie de saisie des données. La banque est validée par un audit annuel. Actuellement, la base en réseau DIVAT porte sur plus de 7000 patients.

Afin de mieux percevoir les futurs résultats de ce mémoire, nous proposons une description des principales covariables disponibles dans DIVAT (celles intégrées dans les analyses).

- ★ **Le retard au démarrage des greffons (DGF).** C'est le nombre de jours nécessaires au greffon pour obtenir une clairance à la créatinine calculée selon la formule de Cockcroft plus ou moins égale à 1 mL par minute. En résumé, c'est le nombre de jours nécessaires au greffon pour reprendre ses fonctions.
- ★ **L'incompatibilité HLA.** Elle désigne la compatibilité entre le donneur et le receveur sur les antigènes d'histocompatibilité HLA A, B et DR. C'est un critère immunologique d'appariement considéré comme base des systèmes de répartition par les transplantateurs depuis plus de 30 ans. Malgré une large remise en cause de leur influence à la lumière de l'amélioration générale de l'efficacité et de la tolérance des traitements immunosuppresseurs, ces antigènes demeurent importants. La plupart des organisations de régulation accordent une "priorité universelle" aux malades parfaitement compatibles (zéro incompatibilités HLA A, B, DR). Elles utilisent, à des degrés divers, la compatibilité HLA dans leurs systèmes de répartition.
- ★ **L'immunisation anti HLA.** C'est une information concernant le risque immunologique des patients, elle est renseignée en terme de taux. Plus ce taux est élevé chez un patient plus celui-ci est considéré comme un patient à risque immunologique élevé.
- ★ **La maladie initiale.** C'est le type de maladie ayant conduit à l'insuffisance rénale terminale. Les différentes maladies possibles sont

classées comme récidives ou non.

- ★ **L'ischémie froide.** Elle correspond au temps de conservation du greffon entre le clampage et le déclampage. Elle est évaluée en minutes. La réduction préférable de cette ischémie froide peut se faire en organisant différemment les prélèvements, la répartition des organes et les greffes.
- ★ **La néoplasie.** Elle renseigne sur les antécédents de cancer avant la greffe. Le bilan cardiovasculaire est fondamental car ces antécédents sont une source de mortalité après greffe.
- ★ **Le Traitement d'induction.** C'est le traitement immunosuppresseif d'induction au moment de la transplantation. Ils sont au nombre de quatre dans notre base de données. Ces traitements ont pour but de diminuer le risque de rejet de greffe. Cependant il a été observé que chacune de ces drogues immunosuppresseives contribue à une réduction de la résistance aux infections ainsi qu'au risque de développer un cancer.
- ★ **La relation donneur / receveur.** Elle définit la relation entre le donneur et le receveur et le type de greffe. Le greffon peut provenir d'un donneur vivant ou d'un donneur décédé. Dans le premier cas de transplantation, le rein provient d'une personne en bonne santé dont le groupe sanguin et le groupe tissulaire sont compatibles avec ceux du receveur. Le donneur vivant peut être un membre de la famille (donneur vivant apparenté) ou il peut aussi s'agir d'une personne qui n'est pas génétiquement apparentée (donneur vivant non apparenté).
- ★ **L'âge du donneur et du receveur.** Le nombre de transplantations rénales réalisées chez le sujet âgé est en constante augmentation depuis ces dix dernières années. On observe d'une part une augmentation de la prévalence de l'insuffisance rénale chronique chez les sujets âgés, à l'origine d'une augmentation de la demande. D'autre part, les résultats de la transplantation sont supérieurs à ceux de la dialyse en termes de survie et de qualité de vie. Un des inconvénients de cette pratique, en ces temps de pénurie d'organes, est de pénaliser les sujets plus jeunes en les privant de greffons ou en prolongeant leur temps d'attente. Afin de réduire ce risque, le pool des donneurs a été élargi à des sujets plus âgés et des programmes visant à réserver ces reins âgés pour des receveurs âgés ont vu le jour. Les premiers résultats à court et moyen termes de ce type de transplantation sont encourageants avec une excellente survie des patients et des greffons (86 et 80% à un an respectivement) et une bonne fonction rénale (52).
La mortalité après la greffe chez les receveurs d'une transplantation rénale âgés est due principalement à la maladie cardiovasculaire, à un taux accru d'infections et au cancer (56). La greffe rénale, même si elle est effectuée avec un donneur limite, c'est-à-dire ayant au

moins un des critères suivants : donneur de plus de 60 ans, antécédents d'hypertension, antécédents de diabète, décès par accident vasculaire cérébral, fait plus que doubler la survie des patients de 60 ans et plus par rapport à ceux qui restent en dialyse.

Par ailleurs, les infections bactériennes restent la première cause de mortalité en transplantation rénale chez l'enfant. L'amélioration de la survie globale passe probablement par une plus grande vigilance vis-à-vis des infections nosocomiales et par une limitation de l'immunosuppression initiale (34).

- ★ **Le sexe du donneur et du receveur.** Le sexe du donneur n'entre pas en ligne de compte dans les critères de sélection pour une greffe. Cependant, les premières études qui rapportent l'effet du sexe sur la survie des greffons remontent aux années 1990 et montrent une moins bonne survie des greffons à court et à long terme quand un rein de donneur femme est transplanté chez un receveur homme (45). Ces observations ont été confirmées ultérieurement sur de plus larges séries qui confirment que la survie des premières transplantations est significativement moins bonne quand le rein provient d'un donneur féminin quel que soit le sexe du receveur (63). L'explication la plus répandue de cet effet du sexe du donneur sur la survie des greffons concerne l'inadéquation entre la "dose de néphrons" et la masse du receveur (42). En effet, les études anatomiques sur les reins de femmes montrent qu'ils sont moins gros et ont moins de néphrons que les reins d'hommes mais cette observation n'est plus valable lorsque la taille du rein est corrigée par la surface corporelle (47). Cependant, l'inadéquation entre la taille du rein et la masse du receveur, ne semble pas être le seul facteur explicatif de l'effet du sexe du donneur sur la survie du greffon. En effet, quelques études montrent que l'incidence du rejet aigu est plus élevée chez les receveurs masculins qui ont reçu un rein de femme. Une hypothèse, issue de modèles animaux, suggère que les reins de femelles expriment plus d'antigènes HLA et seraient ainsi plus antigéniques que les reins de mâles. L'effet du sexe est aussi retrouvé chez le receveur indépendamment du donneur (46), (58). Cet effet a été récemment confirmé sur une population de 512 premières greffes de reins dans un seul centre où il est montré que la perte des greffons, due à une néphropathie chronique du transplant, est corrélée à la créatinémie, à la présence d'anticorps anti HLA de classe II, au rejet aigu, et au sexe masculin du receveur (15). Une hypothèse hormonale pourrait expliquer, au moins en partie, cette observation.
- ★ **L'année de greffe.** Les stratégies thérapeutiques ont été considérablement modifiées aux alentours des années 1996-1998. De plus, durant la même période, l'amélioration de la logistique d'organisation des greffes a permis de réduire très significativement la durée de l'ischémie froide, principal facteur de risque de survenue d'un retard au démarrage des greffons. Enfin, les techniques de dépistage des anticorps anti HLA pré-transplantation et le typage des motifs antigéniques HLA des donneurs et des receveurs avant greffe uti-

lisent depuis ces dix dernières années des techniques hautement performantes et résolutive permettant d'éviter les rejets humoraux hyper aigus, très délétères pour les greffons, et d'améliorer la compatibilité tissulaire entre les donneurs et les receveurs. Cependant durant la même période, l'âge des donneurs et celui des receveurs se sont considérablement accrus, n'allant cette fois-ci pas dans le sens d'une amélioration des résultats des greffes dans ces tranches d'âge élevé.

Toutes les covariables décrites ci-dessus sont résumées et leurs codages sont donnés dans le tableau (1.2).

Nom	Code
Age du receveur	1 si ≥ 55 ans, 0 sinon
Sexe du receveur	1 si femme, 0 sinon
Date de transplantation	1 si ≥ 1996 , 0 sinon
DGF	1 si ≥ 6 jours, 0 sinon
Incompatibilité HLA	1 si ≥ 4 , 0 sinon
Taux d'immunisation Anti HLA	1 si $\geq 0\%$, 0 sinon
Traitement d'induction	1,2,3 ou 4
Maladie initiale	1 si maladie à récurrence, 0 sinon
Antécédents cardiologiques du receveur	1 si oui, 0 sinon
Relation donneur / receveur	1 si cadavre, 0 sinon
Ischémie	1 si ≥ 24 heures, 0 sinon
Néoplasie	1 si oui, 0 sinon
Age du donneur	1 si ≥ 55 ans, 0 sinon
Sexe du donneur	1 si femme, 0 sinon
Centre de transplantation	1, 2, 3 ou 4

TAB. 1.1 – *Caractéristiques des patients transplantés et leur codage.*

Pour des raisons d'éthique, les noms des centres ne sont pas mentionnés. Nous les appellerons donc par des numéros.

1.3 STATISTIQUES DESCRIPTIVES ET ÉLÉMENTAIRES

Les différents graphiques et tableaux donnés dans ce paragraphe nous permettent d'avoir un aperçu général des données.

Le gel de la base a été effectué en octobre 2006. A cette date, la base contenait 4659 patients. Cependant, nous avons fait le choix de supprimer les patients ayant des valeurs manquantes dans leurs caractéristiques. Le tableau (1.2) récapitule le nombre de valeurs manquantes et la proportion que celles-ci représentent par covariable. Nous remarquons que les covariables "anti-HLA" et "néoplasie" sont celles qui contiennent le plus de valeurs manquantes. Nous pouvons nous permettre de supprimer les patients ayant des valeurs manquantes car le nombre de patients dans la base reste assez élevé (2864) et que les valeurs manquantes peuvent être considérées complètement aléatoires des valeurs qui auraient été obser-

vées. Ceci se vérifie puisque les répartitions et les proportions de chaque information (décès, rejet, âge du receveur, ...) entre la base extraite et la base "allégée" restent les mêmes (voir les tableaux (1.3) et (1.4)). Nous allons donc, à partir de maintenant, explorer la base de données sans les valeurs manquantes.

Cette base "allégée" comporte 2864 patients qui ont subi une première greffe de rein. Dès que ceux-ci rejettent leur première greffe, ils repartent en dialyse et nous n'avons aucune information ultérieure c'est-à-dire qu'un éventuel décès ou une éventuelle re-transplantation n'est pas renseigné.

Covariables	NA	
	Effectifs	%
Sexe Receveur	0	0,00%
Age Receveur	0	0,00%
DGF	732	15,71%
Année de la greffe	0	0,00%
Sexe Donneur	594	12,75%
Age Donneur	542	11,63%
Incompatibilité D/R	564	12,11%
Anti HLA	1380	29,62%
Traitement	0	0,00%
Maladie initiale	0	0,00%
Antécédent Cardio	719	15,43%
Relation	501	10,75%
Ischémie	585	12,56%
Néoplasie	1235	26,51%
Centre	0	0,00%

TAB. 1.2 – Détails des effectifs des valeurs manquantes des covariables.

1.3.1 Statistiques générales

Le tableau (1.3) montre que 7,54% des patients décèdent après une transplantation de rein et que 14,53% retournent en dialyse. Le temps minimum entre le moment de la transplantation et un décès ou un retour en dialyse est de 0,0055 ans soit environ 2 jours.

Base	Evènement	Effectifs	%	Minimum	Maximum	Moyenne	Médiane
Extraite (4659)	Rejet	629	13,50%	0,0027	16,2600	4,3620	3,5840
	Décès	333	7,15%	0,0027	16,1900	4,2670	3,2930
	Censure	3697	79,35%	0,0082	16,7600	6,7620	5,8710
"Allégée" (2864)	Rejet	416	14,53%	0,0055	16,2600	5,0360	4,4220
	Décès	216	7,54%	0,0055	16,1900	4,7530	3,8410
	Censure	2232	77,93%	0,0521	16,7600	6,8470	6,0070

TAB. 1.3 – Détails des effectifs et répartition des transitions.

Le tableau (1.4) donne différentes informations concernant les patients transplantés rénaux : environ 62% des patients sont des hommes et 66% de l'ensemble des transplantés sont âgés de moins de 55 ans. Par ailleurs, plus de 50% des patients ont été transplantés dans le centre 2 et la plupart des transplantations (environ 70%) ont eu lieu après 1996. Les donneurs

de rein sont en général des hommes (environ 64%) et sont âgés de moins de 55 ans (75,45%).

Covariables		Base Initiale		Base "Allégée"	
		Total 4659	100,00%	Total 2864	100,00%
Sexe Receveur	Homme	2882	61,86%	1782	62,22%
	Femme	1777	38,14%	1082	37,78%
Age Receveur	< 55 ans	3167	67,98%	1890	65,99%
	≥ 55 ans	1492	32,02%	974	34,01%
DGF	< 6 jours	2639	67,20%	1872	65,36%
	≥ 6 jours	1288	32,80%	992	34,64%
Année de la greffe	< 1996	1397	29,98%	867	30,27%
	≥ 1996	3262	70,02%	1997	69,73%
Sexe Donneur	Homme	2558	62,93%	1840	64,25%
	Femme	1507	37,07%	1024	35,75%
Age Donneur	< 55 ans	3033	73,67%	2161	75,45%
	≥ 55 ans	1084	26,33%	703	24,55%
Incompatibilité D/R	< 4	2518	61,49%	1694	59,15%
	≥ 4	1577	38,10%	1170	40,85%
Anti HLA	< 0	2707	82,56%	2351	82,09%
	≥ 0	572	17,44%	513	17,91%
Traitement	0	798	17,13%	24	0,84%
	1	2070	44,43%	1572	54,89%
	2	653	14,02%	566	19,76%
	3	1138	24,43%	702	24,51%
Maladie initiale	0	3169	68,02%	1846	64,64%
	1	1490	31,98%	1018	35,54%
Antécédent Cardio	0	652	16,55%	541	18,89%
	1	3288	83,45%	2323	81,11%
Relation	0	316	7,60%	175	6,11%
	1	3842	92,40%	2689	93,89%
Ischémie	< 24	2281	55,98%	1523	53,18%
	≥ 24	1793	44,02%	1341	46,82%
Néoplasie	0	3243	94,71%	2724	95,11%
	1	181	5,29%	140	4,89%
Centre	1	1132	24,30%	921	32,16%
	2	1543	33,12%	1463	51,08%
	3	962	20,65%	339	11,84%
	4	1022	21,94%	141	4,92%

TAB. 1.4 – Détails des effectifs des modalités des covariables.

1.3.2 Statistiques par centre de transplantation

Les tables suivantes donnent différentes informations concernant les greffes de rein par centre hospitalier.

La table (1.5) montre qu'il y a plus de décès dans le centre 2; par ailleurs, il y a aussi une plus grande proportion de retour en dialyse.

La table (1.6) permet de mieux percevoir les caractéristiques des patients transplantés par centre hospitalier. Le centre 2, dans lequel les pro-

Covariables		Centre 1		Centre 2		Centre 3		Centre 4	
		921	100,00%	1463	100,00%	339	100,00%	141	100,00%
Décès	0	856	92,94%	1326	90,64%	328	96,76%	138	97,87%
	1	65	7,06%	137	9,36%	11	3,24%	3	2,13%
Rejet	0	792	85,99%	1199	81,95%	321	94,69%	136	96,45%
	1	129	14,01%	264	18,05%	18	5,31%	5	3,55%

TAB. 1.5 – Détails des effectifs de patients décédés et retournés en dialyse par centre de transplantation.

portions de décès et de retour en dialyse sont les plus fortes, est le centre dont le nombre de transplantations est le plus important. De plus, ce centre greffe :

- * plus de personnes âgées,
- * plus de patients ayant plus de 4 incompatibilités avec leur donneur,
- * plus de patients ayant un taux d'immunisation positif ou nul,
- * plus de patients ayant des antécédents cardiologiques,
- * plus de patients ayant un donneur homme,

que les autres centres. Par ailleurs, pour le centre 2, le nombre de greffes dont l'ischémie est supérieure à 24 heures est plus important et il y a aussi une plus grande proportion de greffons qui ont mis plus de 6 jours à reprendre fonction.

Covariables		Centre 1		Centre 2		Centre 3		Centre 4	
		921	100,00%	1463	100,00%	339	100,00%	141	100,00%
Sexe Receveur	Homme	571	62,00%	921	62,95%	200	59,00%	90	63,83%
	Femme	350	38,00%	542	37,05%	139	41,00%	51	36,17%
Age Receveur	< 55 ans	645	70,03%	905	61,86%	241	71,09%	99	70,21%
	≥ 55 ans	276	29,97%	558	38,14%	98	28,91%	42	29,79%
DGF	< 6 jours	679	73,72%	858	58,65%	221	65,19%	114	80,85%
	≥ 6 jours	242	26,28%	605	41,35%	118	34,81%	27	19,15%
Année de la greffe	< 1996	292	31,70%	535	36,57%	20	5,90%	20	14,18%
	≥ 1996	629	68,30%	928	63,43%	319	94,10%	121	85,82%
Sexe Donneur	Homme	575	62,43%	976	66,71%	203	59,88%	86	60,99%
	Femme	346	37,57%	487	33,29%	136	40,12%	55	39,01%
Age Donneur	< 55 ans	747	81,11%	1056	72,18%	243	71,68%	115	81,56%
	≥ 55 ans	174	18,89%	407	27,82%	96	28,32%	26	18,44%
Incompatibilité D/R	< 4	587	63,74%	726	49,62%	286	84,37%	95	67,38%
	≥ 4	334	36,26%	737	50,38%	53	15,63%	46	32,62%
Anti HLA	< 0	831	90,23%	1124	76,83%	268	79,06%	128	90,78%
	≥ 0	90	9,77%	339	23,17%	71	20,94%	13	9,22%
Traitement	0	4	0,43%	5	0,34%	10	2,95%	5	3,55%
	1	652	70,79%	708	48,39%	147	43,36%	65	46,10%
	2	164	17,81%	304	20,78%	73	21,53%	25	17,73%
	3	101	10,97%	446	30,49%	109	32,15%	46	32,62%
Maladie initiale	0	638	69,27%	910	62,20%	215	63,42%	83	58,87%
	1	283	30,73%	553	37,80%	124	36,58%	58	41,13%
Antécédent Cardio	0	222	24,10%	216	14,76%	72	21,24%	31	21,99%
	1	699	75,90%	1247	85,24%	267	78,76%	110	78,01%
Relation	0	46	4,99%	95	6,49%	21	6,19%	13	9,22%
	1	875	95,01%	1368	93,51%	318	93,81%	128	90,78%
Ischémie	< 24	573	62,21%	624	42,65%	231	68,15%	95	67,38%
	≥ 24	348	37,79%	839	57,35%	108	31,86%	46	32,62%
Néoplasie	0	878	95,33%	1386	94,75%	329	97,05%	131	92,91%
	1	43	4,67%	77	5,26%	10	2,95%	10	7,09%

TAB. 1.6 – Détails des effectifs des différentes modalités des covariables par centre de transplantation.

1.3.3 Statistiques par année de transplantation

Les greffes ont lieu entre janvier 1990 et octobre 2006. La table (1.7) et la table (A.1) en Annexe page 130 donnent un aperçu des transplantations effectuées par année. Nous remarquons qu'après 1996 la proportion des décès et des retours en dialyse s'amointrit. Ceci est dû, entre autre, aux changements des stratégies et moyens thérapeutiques ainsi qu'au faible recul pour les dernières années. De plus, d'un point de vue général, deux observations sont assez remarquables. Premièrement, les traitements administrés aux patients diffèrent selon les années. Ensuite, les différents centres de transplantations n'enregistrent pas le même nombre de greffes suivant les années.

Années	Décès		Rejet		
	0	1	0	1	
1990	172 100%	152 88,37%	20 11,63%	107 62,21%	65 37,79%
1991	154 100%	128 83,22%	26 16,88%	104 67,53%	50 32,47%
1992	153 100%	130 84,97%	23 15,03%	103 67,32%	50 32,68%
1993	141 100%	107 75,89%	34 24,11%	116 82,27%	25 17,73%
1994	111 100%	96 86,49%	15 13,51%	82 73,87%	29 26,13%
1995	136 100%	121 88,97%	15 11,03%	96 70,59%	40 29,41%
1996	147 100%	135 91,84%	12 8,16%	117 79,59%	30 20,41%
1997	155 100%	145 93,55%	10 6,45%	131 84,52%	24 15,48%
1998	192 100%	182 94,79%	10 5,21%	167 86,98%	25 13,02%
1999	151 100%	144 95,47%	7 4,63%	136 90,07%	15 9,93%
2000	201 100%	193 96,02%	8 3,98%	184 91,54%	17 8,46%
2001	184 100%	174 94,67%	10 5,43%	165 89,67%	19 10,33%
2002	258 100%	244 94,67%	14 5,43%	246 95,35%	12 4,65%
2003	212 100%	206 97,17%	6 2,83%	210 99,06%	2 0,94%
2004	210 100%	208 99,05%	2 0,95%	199 94,76%	11 5,24%
2005	194 100%	190 97,94%	4 2,06%	192 98,97%	2 1,03%
2006	93 100%	93 100,00%	0 0,00%	93 100,00%	0 0,00%

TAB. 1.7 – Effectifs de patients décédés et retournés en dialyse par année de transplantation.

1.3.4 Corrélation entre les covariables

Le tableau (1.8) montre, grâce au test du Chi-Deux d'indépendance, que beaucoup de covariables sont corrélées entre elles (en gras). Il va donc

falloir être vigilant quant aux choix des covariables significatives et à l'interprétation des prochains résultats.

Covariables	Sexe Recepteur	Age Recepteur	DGF	Année de la greffe	Sexe Donneur	Age Donneur	Incompatibilité D/R	Anti HLA	Traitement	Maladie initiale	Antécédent Cardio	Relation	Ischémie	Néoplasie	Centre
Sexe Recepteur
Age Recepteur	0,998
DGF	0,003	0,000
Année de la greffe	0,293	0,061	0,092
Sexe Donneur	0,370	0,821	0,015	0,000
Age Donneur	0,899	0,000	0,000	0,000	0,003
Incompatibilité D/R	0,012	0,000	0,000	0,095	0,832	0,001
Anti HLA	0,000	0,282	0,000	0,000	0,392	0,091	0,066
Traitement	0,446	0,000	0,248	0,000	0,052	0,000	0,028	0,000
Maladie i	0,000	0,000	0,431	0,000	0,073	0,000	0,641	0,710	0,022
Antécédent Cardio	0,000	0,000	0,346	0,000	0,955	0,000	0,005	0,167	0,000	0,354
Relation	0,642	0,000	0,000	0,000	0,000	0,011	0,000	0,000	0,000	0,000	0,345	0,833	.	.	.
Ischémie	0,915	0,479	0,000	0,000	0,004	0,592	0,003	0,000	0,000	0,465	0,872	0,000	.	.	.
Néoplasie	0,361	0,000	0,523	0,409	0,282	0,000	0,084	0,639	0,039	0,113	0,099	0,213	0,335	.	.
Centre	0,571	0,000	0,000	0,000	0,034	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,191	0,000	0,191

TAB. 1.8 – Test du Chi-deux d'indépendance entre les covariables.

1.4 PROBLÉMATIQUES CLINIQUES

Dans le domaine de la transplantation rénale, le principal objectif est la survie du greffon. Cependant pour que ce greffon vive, il faut aussi que le transplanté vive. Un des intérêts des médecins porte donc sur les décès des patients ayant subi une transplantation de rein. Nous allons donc chercher à connaître les chances de survie des patients après leur transplantation ainsi que les caractéristiques des patients les plus à risques de décéder après une greffe de rein. Par contre, nous nous intéressons principalement à la survie des transplantés ayant leur greffon fonctionnel. De plus, les patients étudiés sont des patients ayant subi une seule transplantation de rein.

La base DIVAT possède de nombreux renseignements concernant les patients transplantés rénaux. Cependant, l'information concernant les causes des décès des patients n'est pas connue. Lorsqu'un patient décède, il est difficile de classer son décès. En effet, affirmer ou infirmer qu'un décès est imputable au fait que le patient ait été transplanté est assez délicat. Les causes de décès souvent rencontrées après une transplantation rénale sont cardiovasculaires ou infectieuses mais ces causes sont-elles vraiment dues à la greffe? Il existe une méthode qui permet d'estimer la survie nette d'une pathologie c'est-à-dire la survie d'une cause donnée lorsque toutes les autres causes de décès ont été éliminées. Cette méthode est celle de la survie relative. Un des avantages de cette méthode, que nous verrons dans le chapitre 4, est qu'elle ne requiert pas la connaissance des

causes de décès spécifiques à la pathologie étudiée.

Par ailleurs, la base DIVAT est une base multi-centrique. Ici, nous possédons les renseignements concernant les patients greffés dans quatre centres différents. Dans ce chapitre, nous avons remarqué que la proportion de décès n'est pas la même dans chaque centre. La question posée est de savoir si cet effet centre est un biais généré par la sélection des patients ou s'il est dû à une autre cause. Le chapitre 3 présente une approche, celle de la fragilité, qui permet de prendre en compte la dépendance entre plusieurs individus d'un même groupe. Cette dépendance n'étant pas mesurable, elle est modélisée par une variable aléatoire que l'on nomme donc la fragilité. C'est le cas dans la base DIVAT, en effet, les patients transplantés dans un même centre peuvent avoir un lien qui n'est pas forcément évaluable.

De plus, durant la période de suivi de ces patients transplantés (de 1990 à 2006), les stratégies thérapeutiques ont été modifiées. Dans ce chapitre, nous avons remarqué qu'après 1996 la proportion des décès s'amointrit. Cela est naturellement dû au fait que, pour les premières années, il y a plus de recul donc plus d'évènements sont enregistrés. Cependant, en plus de prendre en compte l'effet de l'année de transplantation comme covariable, il nous a paru intéressant de modéliser cet effet au moyen de la fragilité. Effectivement, nous savons que les moyens thérapeutiques ont changé au cours des années 90, néanmoins il est difficile de quantifier ces changements.

Ensuite, il nous a paru intéressant de pouvoir combiner les deux problématiques précédentes : information indisponible des causes de décès des greffés rénaux et dépendance des patients issus de même groupe. En effet, nous voulons prendre en compte la dépendance entre les patients greffés dans un même centre (respectivement la même année) mais nous voulons, en même temps, estimer la survie nette des patients greffés. Nous verrons donc, au chapitre 5, un moyen d'associer l'approche de la survie relative et celle de la fragilité.

Enfin, nous savons que certaines covariables sont connues pour être à la fois liées à la perte du greffon et au décès. De plus, dans la base DIVAT, les patients rejetant leur greffon retournent en dialyse, changent de centre de soins et donc l'information sur leur suivi est perdue. Dans ce cas, cette perte de suivi ne peut être considérée comme une censure classique. Effectivement, les censures souvent rencontrées en analyse de survie sont les censures administratives ou planifiées, elles sont alors considérées comme indépendantes de l'évènement étudié. Cependant, dans notre cas, il ne serait pas adéquat de considérer la censure liée au retour en dialyse comme une censure non informative. Dans le chapitre 6, nous verrons donc un moyen de prendre en compte ce problème de la censure informative.

DISCUSSION-CONCLUSION

Ce chapitre donne un descriptif des données que nous allons traiter dans ce mémoire. Il va aussi nous servir de référence pour suggérer certains modèles plutôt que d'autres ainsi que pour appuyer certains résultats et conclusions.

Nous allons, tout au long de ce mémoire, nous intéresser au décès des patients ayant subi une transplantation de rein qui ont leur greffon fonctionnel ainsi qu'à leurs caractéristiques. Les questions posées seront celles que l'on rencontre classiquement en analyse de survie. Le chapitre 2 est un rappel des principales notions de survie. Il possède une partie application qui va permettre de répondre aux questions habituelles en survie.

Après les statistiques élémentaires de la base DIVAT, plusieurs conclusions sont relevées. Tout d'abord, la survie des patients greffés rénaux est soupçonnée différente selon leur centre de transplantation. Ainsi, le chapitre 3 décrira comment prendre en compte cette particularité grâce aux modèles de fragilité. Par ailleurs, la survie des patients est constatée hétérogène selon les années de transplantation, nous aurons donc aussi recours à la fragilité pour modéliser cette spécificité.

Ensuite, comme souvent en analyse de survie, les causes de décès concernant les patients ne sont pas disponibles. Ceci parce que les registres ne sont pas assez complets ou parce que les autopsies n'ont pu être faites. Il est assez donc difficile de définir si le décès d'un patient greffé est imputable à la transplantation rénale. Ainsi, l'estimation de la survie n'est pas souvent exacte dans le sens où elle prend en compte tous les décès même ceux qui ne sont pas dus à la pathologie étudiée. Le chapitre 4 nous permettra d'apprécier comment estimer la survie nette des patients transplantés rénaux au travers des modèles de survie relative.

Par la suite, les problématiques résolues aux chapitres 3 et 4 vont être combinées dans le chapitre 5. En effet, l'estimation de la survie nette des patients transplantés en considérant que ces patients sont dépendants selon leur centre ou leur année de transplantation mérite un intérêt.

Pour finir, les cliniciens ont remarqué que certaines covariables sont à la fois liées à la perte du greffon et au décès du patient. Par ailleurs, l'information des patients après leur rejet de greffe n'est plus disponible. Ceci est donc un cas de censure informative. Le chapitre 6 donnera donc un moyen de considérer la perte du greffon dans la survie des patients transplantés rénaux.

SOMMAIRE

2.1	PREMIÈRES DÉFINITIONS	21
2.1.1	Les fonctions de survie	21
2.1.2	Les mécanismes de censure	22
2.2	LES MODÈLES CLASSIQUES DE SURVIE	22
2.2.1	Modélisation non-paramétrique	22
2.2.2	Modélisation paramétrique	23
2.2.3	Modélisation semi-paramétrique	24
2.3	MODÈLE STRATIFIÉ	25
2.4	STRATÉGIE DE SÉLECTION DES COVARIABLES	25
2.5	APPLICATION AUX TRANSPLANTÉS RÉNAUX	26
2.5.1	Survie non paramétrique	26
2.5.2	Modèle paramétrique	27
2.5.3	Modèles de Cox	31
	DISCUSSION-CONCLUSION	34

DANS de nombreuses études médicales, la survie est le critère principal d'évaluation thérapeutique. Analyser les données de survie c'est s'intéresser à l'apparition d'un évènement au cours du temps. Comme dans toute analyse statistique, nous voulons décrire les observations, donc fournir les estimations de paramètres pertinents, comparer la survie de plusieurs groupes de sujets, expliquer et prédire la durée de survie en fonction de certains facteurs. Le terme de durée de survie est employé pour désigner le temps qui s'écoule jusqu'à la survenue d'un évènement particulier. Cet évènement n'est pas forcément le décès, il peut s'agir, par exemple, d'une rechute et la durée de survie est, dans ce cas, le délai de rémission, ou il peut s'agir de la guérison, et la durée de survie représente alors le délai qui sépare le diagnostic de la guérison.

Ce chapitre est un recueil de différentes notions d'analyse de survie nécessaires à la compréhension de ce mémoire. Dans une première partie, les définitions et fonctions propres à la survie sont données. De plus, nous énonçons brièvement les notions de censure. Ensuite, nous présentons les principales méthodes pour estimer la survie : la modélisation

non-paramétrique, la paramétrique et la semi-paramétrique. Pour finir, ces modèles classiques de l'analyse de survie sont appliqués à la base de données DIVAT. En effet, nous nous intéressons à la survie des patients transplantés ayant leur greffon fonctionnel.

2.1 PREMIÈRES DÉFINITIONS

2.1.1 Les fonctions de survie

Les modèles de survie permettent d'étudier le délai jusqu'à la survenue d'un évènement particulier : l'âge de déclaration d'une maladie, le délai de rémission ou la mort par exemple. L'analyse porte donc sur une variable aléatoire non négative. De plus, nous considérons ici que le cas où T est continue, c'est-à-dire que la probabilité de l'évènement à chaque temps t est infiniment petite.

Cinq fonctions équivalentes définissent la loi du délai de survenue d'un évènement. Soit T , la variable aléatoire continue positive caractérisant ce délai.

- ★ $S(t)$, la **fonction de survie** de T . C'est la probabilité que l'évènement ne se produise pas avant un temps t :

$$S(t) = P(T \geq t). \quad (2.1)$$

$S(t)$ est une fonction monotone décroissante et continue, telle que :

$$S(0) = 1 \text{ et } S(\infty) = 0.$$

- ★ $F(t)$, la **fonction de répartition** de T . C'est la probabilité que l'évènement se produise avant un temps t :

$$F(t) = P(T < t) = 1 - S(t). \quad (2.2)$$

$F(t)$ est une fonction monotone croissante et continue, telle que :

$$F(0) = 0 \text{ et } F(\infty) = 1.$$

- ★ $f(t)$, la **densité de probabilité** de T . C'est la probabilité que l'évènement se produise dans un petit intervalle de temps après t :

$$\begin{aligned} f(t) &= \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt)}{dt} \\ &= -\frac{\partial S(t)}{\partial t} \\ &= \frac{\partial F(t)}{\partial t}. \end{aligned} \quad (2.3)$$

Cette fonction n'est pas forcément bornée à 1 comme le serait une probabilité.

- ★ $\lambda(t)$, la **fonction de risque instantané** de l'évènement, ou plus rigoureusement le taux instantané de l'évènement :

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T > t)}{dt}. \quad (2.4)$$

C'est la probabilité que l'évènement se produise dans un petit intervalle de temps après l'instant t , sachant qu'aucun évènement ne se produit jusqu'à l'instant t . La fonction $\lambda(t)$ vérifie les relations suivantes :

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{\partial}{\partial t} \ln[S(t)].$$

★ $H(t)$, la **fonction de risque cumulée** :

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln(S(t)). \quad (2.5)$$

2.1.2 Les mécanismes de censure

La particularité des données de survie réside dans le fait que ces données sont souvent incomplètes à cause du phénomène de censure. Les méthodes statistiques pour ce genre de données doivent alors tenir compte des données censurées. Différents types de censure sont rencontrés.

- ★ La censure à gauche. Elle se produit lorsqu'on commence à observer un échantillon à un point dans le temps où certains sujets pourraient déjà avoir vécu l'évènement.
- ★ La censure par intervalle. Elle se rencontre, par exemple, quand on constate l'apparition d'un symptôme lors d'une visite médicale : il est apparu depuis la visite précédente, mais on ne sait pas exactement quand.
- ★ La censure à droite qui est la censure la plus courante dans les études de survie. Il existe différentes causes de censure à droite : un individu est considéré comme censuré exclu vivant (si l'évènement se produit après la date de point), décédé pour cause de risques compétitifs (décès non imputable à la maladie tel qu'un accident de la route par exemple) ou perdu de vue (le patient peut décider d'aller se faire soigner ailleurs par exemple et donc on ne le revoit plus).

2.2 LES MODÈLES CLASSIQUES DE SURVIE

Dans cette section, les différents estimateurs de la fonction de survie $S(t)$ sont brièvement présentés : l'estimateur non-paramétrique de Kaplan-Meier, les estimateurs associés à des modèles paramétriques (Exponentiel et Weibull en particulier) et l'estimateur semi-paramétrique (modèle de Cox).

2.2.1 Modélisation non-paramétrique

L'estimateur non-paramétrique le plus utilisé de la fonction de survie est l'estimateur de Kaplan-Meier (38). Il repose sur l'idée intuitive suivante : "être encore en vie après l'instant t , c'est être en vie juste avant t et ne pas mourir à l'instant t ". Cet estimateur consiste donc, à chaque instant t , au rapport du nombre de sujets à risque moins le nombre de décédés sur

le nombre de sujets à risque. L'estimation de la fonction de survie, $\hat{S}(t)$, s'écrit donc

$$\hat{S}(t) = \sum_{t_i \leq t} \frac{n_i - m_i}{n_i}$$

où m_i est le nombre de décès au temps t_i et n_i le nombre de sujets à risque au temps t_i^- .

Cet estimateur est aujourd'hui utilisé en routine et sert souvent de première analyse. Cependant, celui-ci ne permet de prendre en compte l'effet des covariables que par stratification, *i.e.* on estime une fonction de survie par modalité de covariable. L'inconvénient est que ceci entraîne une réduction de la puissance, il est donc difficile de prendre en compte les covariables dans ce type de modèle. De plus, avec la modélisation non-paramétrique, l'information concernant les patients censurés est perdue.

2.2.2 Modélisation paramétrique

Les modèles paramétriques permettent d'écrire la distribution des durées de survie comme une fonction paramétrique du temps. Celle-ci peut s'exprimer en fonction d'un ou plusieurs facteurs de risque susceptibles d'influencer la survie. L'approche paramétrique permet d'évaluer de façon simple l'importance propre de chaque facteur de la survie; elle permet aussi d'estimer de façon explicite les différentes fonctions de distribution des durées de survie (densité de probabilité, fonction de risque instantané, fonction de survie).

Les modèles paramétriques les plus souvent utilisés sont les modèles Exponentiel, Weibull et Weibull Généralisé. En effet, ces modèles sont emboîtés, ils sont donc comparables par des tests statistiques classiques. De plus, la distribution de Weibull Généralisée généralise les deux autres et possède de bonnes propriétés. Selon la valeur des paramètres, la fonction de risque peut-être constante, monotone (croissante ou décroissante), ou non-monotone (\cup ou \cap).

- ★ La loi **Exponentielle**, $\mathcal{E}(\sigma)$: le risque est constant au cours du temps :

$$\lambda_{\mathcal{E}}(t) = \frac{1}{\sigma} \quad \forall \sigma \in \mathbb{R}_*^+. \quad (2.6)$$

- ★ La loi de **Weibull**, $\mathcal{W}(\sigma, \nu)$: le risque est monotone au cours du temps :

$$\lambda_{\mathcal{W}}(t) = \nu \left(\frac{1}{\sigma} \right)^\nu t^{\nu-1} \quad \forall \nu, \sigma \in \mathbb{R}_*^+. \quad (2.7)$$

- ★ La loi de **Weibull Généralisée**, $\mathcal{WG}(\sigma, \nu, \theta)$: le risque est non-monotone au cours du temps :

$$\lambda_{\mathcal{WG}}(t) = \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma} \right)^\nu \right)^{\frac{1}{\theta}-1} \frac{\nu}{\sigma} \left(\frac{t}{\sigma} \right)^{\nu-1} \quad \forall \nu, \sigma, \theta \in \mathbb{R}_*^+. \quad (2.8)$$

Les paramètres de tels modèles sont obtenus en maximisant la vraisemblance.

La vraisemblance se définit comme le produit des contributions de chaque observation : décès ou censure. Son écriture générale est la suivante :

$$V = \prod_{j=1}^n f(t_j)^{\delta_j} S(t_j)^{1-\delta_j} \quad (2.9)$$

avec $j = 1, \dots, n$, l'indice des individus et δ_j , l'indicatrice de l'évènement (i.e., $\delta_j = 1$ si l'évènement pour l'individu j est observé, 0 sinon). Si $\delta_j = 1$ alors t_j représente le temps de survenue de l'évènement pour le sujet j sinon son temps de participation. Cette écriture de vraisemblance suppose l'indépendance entre tous les individus ainsi que l'indépendance entre la censure et l'évènement. Dans le cas de la base DIVAT, la pertinence de ces deux hypothèses sera examinée.

2.2.3 Modélisation semi-paramétrique

Cette approche est un compromis entre les approches non-paramétrique et paramétrique. Le modèle de Cox (20) permet la prise en compte simultanée de plusieurs variables pour expliquer la survie sans donner aux fonctions de survie des formes paramétriques précises. Ce modèle est largement utilisé dans les analyses de données longitudinales mais aussi dans les enquêtes de cohorte et essais thérapeutiques. La particularité de cette approche consiste à ne pas faire d'hypothèse sur la fonction de risque de base. De plus, ce modèle fait l'hypothèse de proportionnalité des risques. Il s'écrit ainsi :

$$\lambda(t, Z) = \lambda_0(t) \exp(\beta'Z) \quad (2.10)$$

où $\lambda_0(t)$ est le risque de base, Z le vecteur des covariables observées et β le vecteur de paramètres de régression associé à ces covariables. λ_0 n'est pas estimée. L'effet des covariables est multiplicatif de la fonction de risque et indépendant du temps. L'utilisation de la fonction Exponentielle permet de conserver la contrainte de positivité de la fonction de risque. Par ailleurs, cette méthode a l'avantage d'interpréter l'exponentiel des paramètres de régression en terme de risques relatifs.

Remarque 2.1 *Il est possible de définir une distribution paramétrique pour la fonction de risque de base $\lambda_0(\cdot)$. Dans ce cas, il s'agit d'un modèle de régression paramétrique.*

En reprenant les mêmes notations que dans le paragraphe précédent, la vraisemblance partielle de Cox s'écrit :

$$VP = \prod_{j=1}^n \left\{ \frac{\exp(\beta'z_j)}{\sum_{k:t_k \geq t_j} \exp(\beta'z_k)} \right\}^{\delta_j} \quad (2.11)$$

où k représente les individus encore à risque au temps t_j .

Remarque 2.2 *Pour vérifier l'hypothèse de proportionnalité des risques d'une covariable, une méthode possible est d'estimer les survies de chaque modalité de cette covariable puis de tracer les courbes $\log(-\log(\hat{S}(t)))$ de ces différentes modalités. La proportionnalité des risques est vérifiée si les courbes ne se croisent pas et présentent un écart constant au cours du temps (40).*

2.3 MODÈLE STRATIFIÉ

Lorsque l'hypothèse de proportionnalité des risques n'est pas vérifiée pour une variable explicative et que l'on souhaite tout de même prendre en compte cette covariable, une solution est de stratifier le modèle sur cette covariable. La stratification divise l'échantillon en plusieurs groupes distincts, autant que de modalité de la covariable. Chacun de ces groupes est supposé posséder une fonction de risque différente mais les coefficients de régression pour les autres variables explicatives étudiées sont communs.

Pour un individu j appartenant à la strate s , la fonction de risque s'écrit :

$$\lambda_s(t_{js}, z_{js}) = \lambda_{0s}(t_{js}) \exp(\beta z_{js}) \quad (2.12)$$

avec $\lambda_{0s}(\cdot)$ la fonction de risque de base commune à tous les individus de la strate s . z_{js} est le vecteur des covariables ne contenant pas la variable sur laquelle le modèle est stratifié et β est le vecteur des coefficients de régression commun à tous les sujets. La vraisemblance de tout l'échantillon est le produit des vraisemblances de chaque strate.

2.4 STRATÉGIE DE SÉLECTION DES COVARIABLES

L'objectif d'une modélisation est d'expliquer au mieux le phénomène d'intérêt à l'aide d'un modèle le plus parcimonieux possible. Le choix d'un tel modèle nécessite de tester l'apport éventuel de paramètres supplémentaires. Pour ce faire, nous utilisons deux tests : le test de Wald (pour la sélection des covariables en univarié) et le test du rapport de vraisemblance (pour la sélection des covariables en multivarié).

- ★ **Stratégie 1 : Analyse univariée puis multivariée.** Tout d'abord, nous estimons un modèle univarié pour chaque covariable à tester. Toutes les covariables ayant une p -value inférieure à 0,20 dans leur modèle univarié seront insérées dans le modèle multivarié. A partir de là, nous optons pour une stratégie descendante qui consiste à éliminer une à une les covariables les moins significatives (p -value $> 0,05$). La procédure s'arrête lorsque tous les facteurs du modèle ont un risque de première espèce inférieur au seuil. A chaque étape de cette stratégie descendante, nous vérifions si les estimations des autres coefficients ne varient pas trop fortement ($\simeq 30\%$). Ceci pour mesurer la stabilité des estimations et identifier de possibles facteurs de confusion.

- ★ **Stratégie 2 : Analyse multivariée.** Cette procédure est une étape de vérification et de confirmation de la parcimonie du modèle choisi par l'approche précédente (Stratégie 1). Elle consiste à incorporer toutes les covariables dans le modèle d'intérêt et, comme dans l'analyse précédente, à l'élimination une à une de celle-ci.

Remarque 2.3 Pour la sélection des covariables, un modèle univarié est un modèle dans lequel une seule covariable est testée et un modèle multivarié est un modèle dans lequel plusieurs covariables sont testées.

Toutes les analyses ainsi que tous les graphiques présentés dans ce mémoire ont été réalisés à l'aide du logiciel R. La fonction `optim()` est utilisée pour maximiser la logvraisemblance et estimer la valeur des paramètres ainsi que leur matrice hessienne. Cette fonction utilise l'algorithme de Quasi-Newton (14). L'initialisation des paramètres propres aux distributions est établie grâce à la fonction `gnls()`, qui utilise la méthode des moindres carrés généralisés non linéaires, à partir des survies estimées par la méthode non paramétrique de Kaplan-Meier (38).

2.5 APPLICATION AUX TRANSPLANTÉS RÉNAUX

Grâce aux premières notions mentionnées ci-dessus, ce paragraphe engage les analyses initiales de la survie des patients transplantés rénaux.

2.5.1 Survie non paramétrique

Nous commençons par estimer la survie des patients au moyen de la méthode non paramétrique. La figure (2.1) montre que, selon le modèle de Kaplan-Meier, la probabilité de survivre 17 ans après une greffe de rein est de l'ordre de 81% environ. Les greffes étant observées de 1990 à 2006, 17 ans correspond au recul de la base.

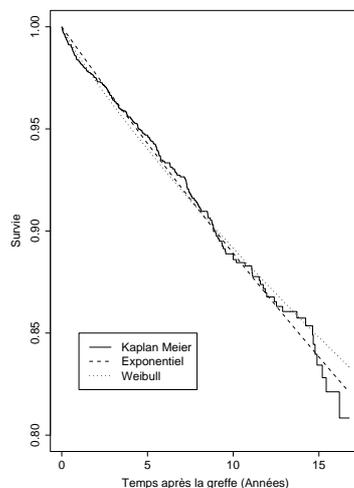


FIG. 2.1 – Survie estimée par différents modèles.

D'après les statistiques élémentaires de la section (1.3) page 11, les survies des patients par centre de transplantation semblent être différentes. Le graphique (2.2) permet d'avoir un aperçu des survies selon les quatre centres. Visuellement, il confirme effectivement une disparité entre ces différentes survies.

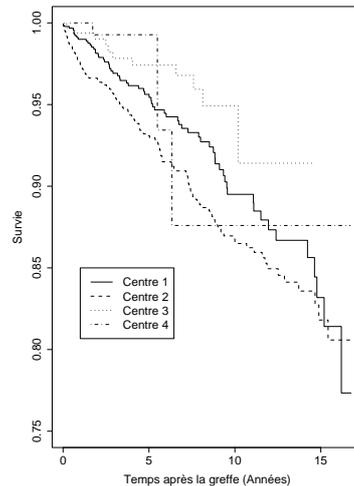


FIG. 2.2 – Survies par centre estimées par la méthode de Kaplan Meier.

Le graphique concernant les différentes survies des patients transplantés par année de greffe ne sera pas exposé ici du fait de sa non clarté (17 années donc trop de courbes sur le graphe).

2.5.2 Modèle paramétrique

Sans covariable

Tout d'abord, la survie des patients transplantés est modélisée par une loi de Weibull (équation (2.7)). Ensuite, nous effectuons un test du rapport de vraisemblance pour savoir si la distribution de survie est Exponentielle, *i.e.* nous testons $\{H_0 : \nu = 1\}$. La statistique de ce test vaut 3,712 correspondant à une p -value supérieure à 0,05, ce résultat nous amène donc à ne pas rejeter l'hypothèse selon laquelle cette survie suit une loi Exponentielle. Le graphique (2.1) permet de visualiser ces deux modélisations. Nous remarquons que ces courbes sont très proches et donc que les deux modélisations sont quasiment similaires.

Le graphique (2.3) est la représentation des différentes estimations de survie par centre de transplantation. Il permet de distinguer, une fois encore, que les patients transplantés dans le centre 2 ont une survie plus faible par rapport aux autres centres. Cependant, nous avons vu dans le chapitre exploratoire de la base de données que le profil des patients greffés dans le centre 2 est différent des autres avec des patients plus à risque (plus âgés, plus d'antécédents cardiologiques, ...).

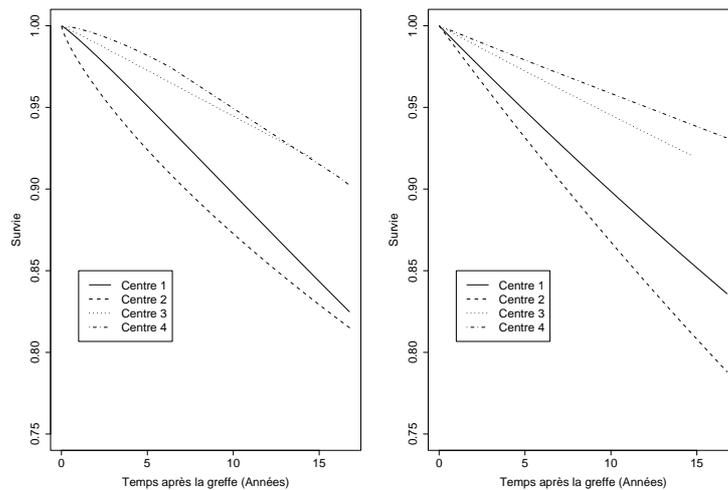


FIG. 2.3 – Survies par centre estimées par les modèles paramétriques : Weibull (à gauche) et Exponentiel (à droite).

Avec covariables

Pour la modélisation avec les covariables, nous prenons une distribution de Weibull pour la fonction de risque. Ensuite, nous effectuons un test du rapport de vraisemblance entre le modèle paramétrique de Weibull et le modèle paramétrique Exponentiel pour savoir si la distribution de survie est Exponentielle, *i.e.* nous testons $\{H_0 : \nu = 1\}$. La statistique de ce test vaut 7,836 correspondant à une p -value inférieure à 0,01. Nous ne validons pas l'hypothèse selon laquelle la fonction de risque est Exponentielle.

Le tableau (2.1) donne les différentes covariables retenues selon les stratégies de sélection. Les quatre covariables significatives en analyse multivariée sont le sexe et l'âge du receveur, le dgf et l'année de la greffe.

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur	×		
Type de maladie initiale	×		
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie	×		
Néoplasie	×		
Age du donneur	×		
Sexe du donneur			
Année de la greffe	×	×	×
Centre	×		

TAB. 2.1 – Modèle paramétrique de Weibull : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).

Les figures (2.4) et (2.5) permettent de vérifier visuellement l'hypothèse de proportionnalité des risques pour les modèles univariés de chaque co-variable. Ces modèles univariés sont bien à risques proportionnels.

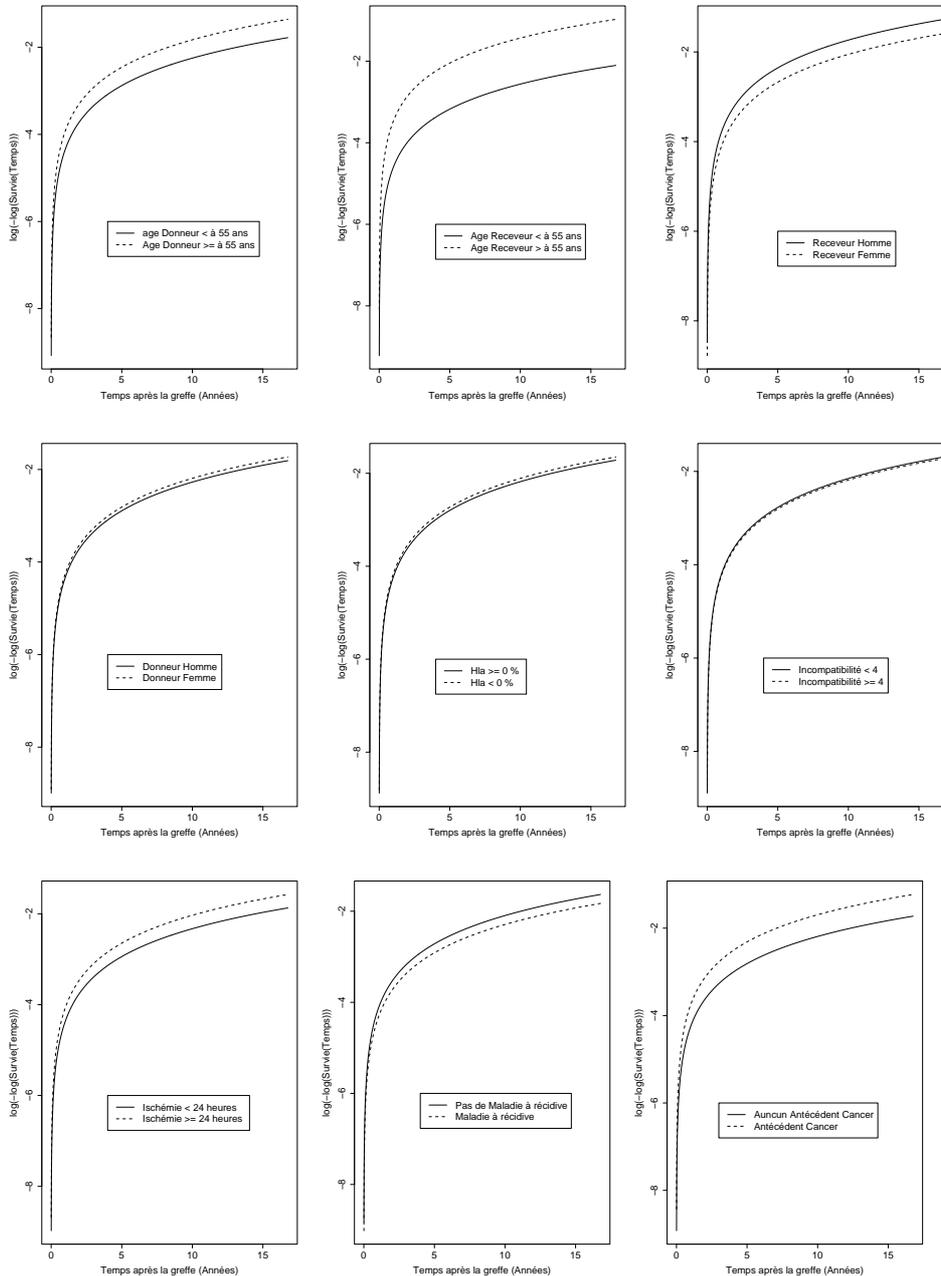


FIG. 2.4 – Test graphique de l'hypothèse de proportionnalité des risques avec le modèle de Weibull.

Le tableau (2.2) montre que les hommes ont environ 1,4 fois plus de chances de décéder après une transplantation de rein que les femmes. Les greffés de plus de 55 ans ont environ 3,1 fois plus de risque de décéder que les moins de 55 ans. Les patients dont le greffon met plus de 6 jours à reprendre fonction ont environ 1,5 fois plus de chance de décéder que ceux dont le greffon met moins de 6 jours. De plus, les patients trans-

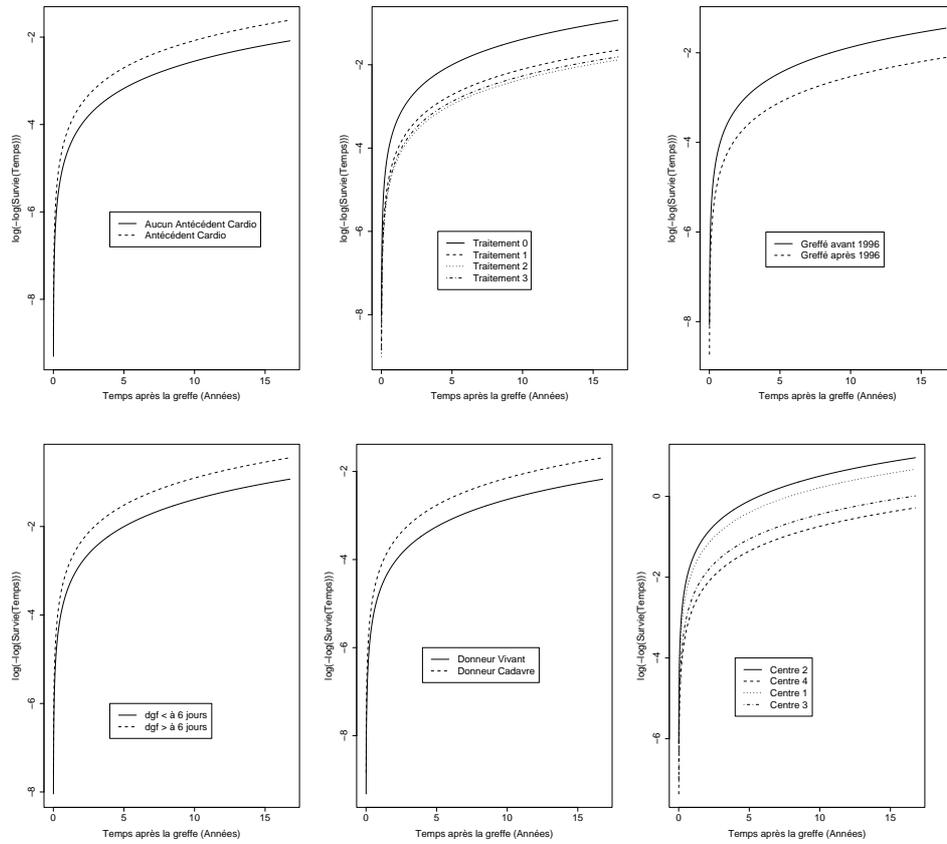


FIG. 2.5 – Test graphique de l'hypothèse de proportionnalité des risques avec le modèle de Weibull.

Paramètres	coef	exp(coef)	se(coef)	p-value
σ	111,691	.	34,650	.
ν	0,848	.	0,053	< 0,05*
Sexe du receveur	-0,333	0,717	0,148	0,024
Age du receveur	1,123	3,074	0,138	0,000
DGF	0,372	1,450	0,137	0,007
Année greffe	-0,680	0,507	0,145	0,000

* Test du rapport de vraisemblance : $\{H_0 : \nu = 1\}$.

TAB. 2.2 – *Modèle paramétrique de Weibull : estimations des paramètres (LogVraisemblance = -1122,80).*

plantés avant 1996 ont environ 2 fois plus de chance de décéder que ceux transplantés après 1996.

La figure (2.6) permet de visualiser les fonctions de risque de décès au cours du temps selon les modalités des covariables. Les différents graphiques confirment les résultats du tableau (2.2). Par ailleurs, ces fonctions de risques n'excèdent pas 0,05 sauf celle pour l'âge du receveur (0,10).

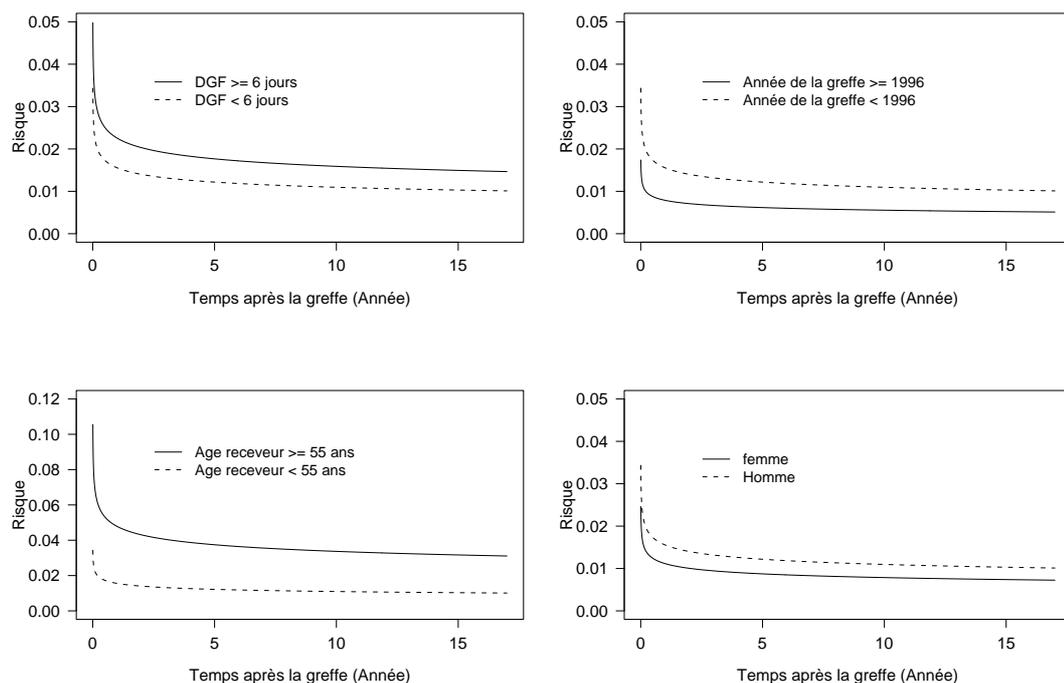


FIG. 2.6 – *Fonction de risque selon les modalités des covariables avec le modèle de Weibull.*

2.5.3 Modèles de Cox

Dans cette section, pour des raisons de futures comparaisons, plusieurs modèles de Cox sont proposés :

- * **un modèle de Cox classique** (tableaux (2.3) et (2.4)),
- * **un modèle de Cox stratifié sur les centres de transplantation** (tableaux (2.5) et (2.6)),

- ★ **un modèle de Cox stratifié sur les années de greffe** (tableaux (2.7) et (2.8)).

Les 3 modèles donnent quasiment les mêmes conclusions (mêmes covariables significatives et même risque relatif pour chaque covariable). De plus, ce sont les mêmes résultats que la modélisation du risque de base par une loi de Weibull. Ainsi, pour alléger le mémoire, nous ne donnons que les tableaux concernant les différents modèles de Cox (et pas les graphiques comme fait pour le modèle de Weibull).

D'après les résultats des estimations des trois modèles de Cox :

- ★ les transplantés hommes ont environ 1,4 fois plus de risque de décéder que les transplantés femmes,
- ★ les patients greffés après 55 ans ont environ 3,1 fois plus de risque de décéder que ceux greffés avant 55 ans,
- ★ les patients dont le greffon met plus de six jours à reprendre fonction ont environ 1,4 fois plus de risque de décéder que ceux dont le greffon met moins de six jours.

Par ailleurs, pour le modèle de Cox classique et le modèle de Cox stratifié sur les centres de transplantation, les patients greffés avant 1996 ont environ 1,7 fois plus de risque de décéder que ceux greffés après 1996.

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur			
Type de maladie initiale			
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie	×		
Néoplasie			
Age du donneur	×		
Sexe du donneur			
Année de la greffe	×	×	×
Centre	×		

TAB. 2.3 – *Modèle de Cox classique : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).*

Covariables	coef	exp(coef)	se(coef)	p-value
Sexe du receveur	-0,338	0,713	0,148	0,022
Age du receveur	1,127	3,087	0,138	0,000
DGF	0,375	1,454	0,137	0,006
Année greffe	-0,640	0,527	0,151	0,000

TAB. 2.4 – *Modèle de Cox classique : estimations des paramètres (LogVraisemblance = -1525,71).*

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur	×		
Type de maladie initiale			
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie			
Néoplasie			
Age du donneur	×		
Sexe du donneur			
Année de la greffe	×	×	×

TAB. 2.5 – Modèle de Cox stratifié sur les centres de transplantations : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).

Covariables	coef	exp(coef)	se(coef)	p-value
Sexe Receveur	-0,346	0,708	0,148	0,019
Age Receveur	1,115	3,049	0,139	0,000
DGF	0,350	1,418	0,140	0,013
Année Greffe	-0,524	0,592	0,154	0,001

TAB. 2.6 – Modèle de Cox stratifié sur les centres de transplantations : estimations des paramètres (LogVraisemblance = -1334,44).

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur			
Type de maladie initiale			
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie			
Néoplasie			
Age du donneur	×		
Sexe du donneur			
Centre			

TAB. 2.7 – Modèle de Cox stratifié sur les années de greffe : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).

Covariable	coef	exp(coef)	se(coef)	p-value
Sexe Receveur	-0,333	0,716	0,148	0,024
Age Receveur	1,114	3,046	0,139	0,000
DGF	0,361	1,434	0,138	0,009

TAB. 2.8 – Modèle de Cox stratifié sur les années de greffe : estimations des paramètres (LogVraisemblance = -1009,301).

DISCUSSION-CONCLUSION

Ce chapitre est une amorce aux analyses de survie des patients transplantés rénaux. Les différents résultats montrent que les caractéristiques des patients les plus à risque de décéder après une transplantation de rein avec un rein fonctionnel sont : les hommes, les plus de 55 ans, les patients dont le greffon met plus de 6 jours à reprendre fonction et les transplantés avant 1996.

Etant donnés les graphiques ainsi que les résultats concernant la survie des patients, *i.e.* les survies sont différentes selon les centres et la covariable "année de transplantation" apparaît toujours significative, nous nous sommes questionnés sur un moyen de modéliser ces deux facteurs autrement qu'en les prenant comme simples covariables dans les modèles ou qu'en stratifiant les modèles sur ceux-ci. En effet, tous les modèles vus dans ce chapitre supposent l'indépendance entre les individus observés. Seulement, on pourrait supposer que, suivant le centre de transplantation (respectivement l'année), ces patients sont dépendants. Cependant, cette dépendance n'est pas mesurable. Pour ce faire, nous nous sommes intéressés aux modèles de fragilité qui permettent de résoudre ce genre de problématique. Les modèles de fragilité seront donc l'objet du prochain chapitre.

Par ailleurs, une autre hypothèse des analyses de survie traditionnelles peut être remise en cause : l'indépendance entre la censure et le processus étudié, ici le décès. Dans ce cas, la censure est dite non informative, ceci car elle se produit indépendamment de l'évènement étudié. Cette hypothèse d'indépendance entre la censure et l'évènement est une hypothèse forte qui induit un biais dans l'estimation de la survie si elle n'est pas respectée. De plus, cette hypothèse n'est pas souvent vérifiée dans la réalité. Nous verrons donc, dans un prochain chapitre (chapitre 6), comment et pourquoi prendre en compte la censure informative dans le cas de la base DIVAT.

SOMMAIRE

3.1	MODÈLES DE FRAGILITÉ "UNIVARIÉS"	39
3.2	MODÈLES DE FRAGILITÉ "MULTIVARIÉS"	41
3.2.1	Le modèle de fragilité partagée	42
3.2.2	Le modèle de fragilité corrélée	44
3.3	LE MODÈLE DE FRAGILITÉ PROPOSÉ	45
3.4	APPLICATION AUX TRANSPLANTÉS RÉNAUX	48
3.4.1	Modèle de Cox avec fragilité	48
3.4.2	Modèle de régression de Weibull avec fragilité	50
	DISCUSSION-CONCLUSION	55

LA notion de fragilité offre un moyen pratique d'introduire des effets aléatoires, de dépendance et d'hétérogénéité non observée, dans les modèles de données de survie. Dans sa forme la plus simple, une fragilité est un facteur de proportionnalité aléatoire non observé qui modifie la fonction de risque d'un individu ou d'un groupe de personnes liées. Fondamentalement, le concept de fragilité remonte aux travaux de Greenwood et Yule (31) sur la "prédisposition à l'accident". Ce n'est qu'en 1979 que le terme de fragilité lui-même a été présenté par Vaupel *et al.* (57) dans les modèles de survie. Cependant, en 1978, sans utiliser le terme de fragilité, Clayton (19) avait déjà largement promu son concept par son application sur l'incidence des maladies chroniques dans les familles.

Les modèles de fragilité sont des extensions du modèle à risques proportionnels de Cox (20). En général, dans la plupart des applications cliniques, l'analyse de survie suppose implicitement que la population étudiée est homogène. Ceci signifie que toutes les personnes échantillonnées dans ce genre d'étude sont soumises, en principe, aux mêmes risques (par exemple, risque de décès, risque de récurrence). Dans de nombreuses applications, la population étudiée ne peut pas être supposée homogène, mais doit être considérée comme un échantillon hétérogène, c'est-à-dire un mélange de personnes avec différents risques. Par exemple, dans de nombreux cas, il est impossible de mesurer toutes les covariables pertinentes liées à la maladie d'intérêt. Les raisons sont souvent économiques

ou liées à la méconnaissance de l'importance de certaines covariables sur la pathologie étudiée. Plus fondamentalement, il y aura toujours une part d'explication inconnue liée au hasard pur. L'approche de la fragilité est un concept de modélisation statistique qui vise à représenter l'hétérogénéité, provoquée par les covariables non mesurées. En termes statistiques, un modèle de fragilité est un modèle à effet aléatoire pour données de survie, dans lequel l'effet aléatoire (la fragilité) a un effet multiplicatif sur la fonction de risque de base. On peut distinguer deux grandes catégories de modèles de fragilité : les modèles que nous appelons "univariés" (un seul évènement par individu) et les modèles que nous appelons "multivariés" (par exemple, la récurrence d'un même évènement chez un individu, l'occurrence d'une maladie dans les familles, l'observation de plusieurs évènements chez un même individu, ...). Dans le premier cas, un modèle de survie avec des temps de survie univariés est utilisé pour décrire l'influence des covariables non observées dans un modèle à risques proportionnels (hétérogénéité). La variabilité des données de survie est divisée en deux parties : une qui dépend des facteurs de risque et est donc théoriquement prévisible, et une autre partie qui est initialement imprévisible, même si toute l'information significative est connue. L'avantage de la séparation de ces deux sources de variabilité est que l'hétérogénéité peut expliquer certains résultats inattendus ou donner une autre interprétation de certains résultats. Plus intéressant, cependant, c'est le second cas, lorsque les modèles multivariés peuvent être considérés. Là, on cherche à tenir compte de la dépendance des temps d'évènements groupés, par exemple pour les durées de vie des patients dans un essai clinique multi-centre, causée par des conditions spécifique du centre, Andersen *et al.* (7). Une façon naturelle de modéliser la dépendance des temps d'évènements groupés est l'introduction d'un effet aléatoire spécifique-groupe, la fragilité. Cet effet aléatoire explique la dépendance dans le sens où si la fragilité était connue, alors les évènements seraient indépendants. En d'autres termes, les durées de vie sont indépendantes conditionnellement à la fragilité. Cette approche peut être utilisée pour les temps de survie d'unités liées, comme les membres d'une famille ou les observations récurrentes sur une même personne. Les modèles de fragilité univariés peuvent être étendus à des modèles multivariés et ceux-ci sont donc examinés dans ce chapitre.

Ainsi, le présent chapitre est un *tour d'horizon* des modèles de fragilité. Les deux premières sections de ce chapitre concernent les différentes notions de fragilité : univariée et multivariée. Les principales formules de survie sont données. Ensuite, nous appliquons ce concept de fragilité à la base de données DIVAT. Comme la survie des patients greffés est suspectée différente selon leur centre de transplantation et que le lien entre les patients d'un même centre n'est pas observable, nous posons un modèle avec fragilité dans lequel cette fragilité définit la dépendance des patients issus d'un même centre. Il en est de même pour les patients transplantés la même année. Un modèle dont la variable de fragilité caractérise l'année de la transplantation est construit. La section "application" comporte deux parties. La première concerne le terme de fragilité inclus dans le modèle de Cox et la deuxième, la fragilité incluse dans un modèle de régression

paramétrique de Weibull. Nous incluons la fragilité dans ces deux types de modèles car nous voulons comparer les résultats concernant cette fragilité dans les deux modèles. Nous cherchons à savoir si elle est significative ou pas, c'est-à-dire s'il y a un effet du centre de transplantation (respectivement année de transplantation) quel que soit le modèle utilisé. Dans chaque partie, deux modèles de fragilité sont proposés : l'un sur les centres de transplantation et l'autre sur les années de transplantation.

3.1 MODÈLES DE FRAGILITÉ “UNIVARIÉS”

Généralement, les modèles de survie supposent que les données de survie des patients sont de formes indépendantes et que chaque distribution des temps de survie individuels des patients est la même (temps d'évènements indépendants et identiquement distribués). Cette hypothèse de base implique une population homogène. Toutefois, dans de nombreuses situations pratiques, les patients observés diffèrent notablement. L'effet d'une drogue, d'un traitement ou l'influence de diverses variables explicatives peut être très différent entre des sous-groupes de patients. Pour rendre compte d'une telle hétérogénéité non observée dans la population étudiée, Vaupel *et al.* (57) ont présenté des modèles de fragilité univariés pour l'analyse de survie. L'idée maîtresse est que les individus possèdent des fragilités différentes, et que les patients les plus fragiles vont mourir plus tôt que les autres. En conséquence, il y a une sélection systématique des individus robustes, *i.e.* les patients avec une faible fragilité. Lorsque les taux de mortalité sont estimés, on peut s'intéresser à la façon dont ces taux changent au cours du temps ou de l'âge. Très souvent, lors d'une maladie, par exemple, nécessitant une intervention chirurgicale, on observe que la fonction de risque augmente au début, atteint un maximum, puis diminue (intensité unimodale) et reprend les valeurs “normales” de risque de décès liées à l'âge (illustration figure 3.1). Plus le patient vit après la manifestation de la maladie, plus ses chances de revenir à un risque “naturel” augmentent. Il est probable que les intensités unimodales, qui retrouvent le risque “naturel”, soient souvent le résultat d'un processus de sélection agissant dans une population hétérogène et ne reflètent pas la mortalité individuelle. Le taux de mortalité de la population hétérogène, sans en tenir compte, peut décliner tout simplement parce que les personnes à risque élevé (grande fragilité) sont déjà mortes. La force de mortalité d'un individu donné pourrait être sur-évaluée à cause de l'hétérogénéité de l'échantillon étudié quand cette hétérogénéité n'est pas prise en compte.

Si les facteurs de protection ou les facteurs de risque sont connus, ceux-ci peuvent être inclus dans l'analyse en utilisant le modèle à risques proportionnels, qui est de la forme de l'équation (2.10). Le “confort mathématique” de ce modèle réside dans la séparation des effets du vieillissement dans la fonction de risque de base des effets des covariables dans la partie paramétrique.

Il existe deux principales raisons pour lesquelles il est souvent impossible d'inclure tous les facteurs importants sur le plan individuel dans l'analyse. Parfois, il y a trop de variables explicatives à prendre en compte dans le modèle, et dans d'autres cas, les chercheurs ne connaissent pas ou ne sont pas capables de mesurer toutes les covariables. Dans les deux cas, il existe deux sources de variation dans les données de survie : la *variabilité* représentée par les facteurs de risques mesurables, qui est donc théoriquement “prévisible” ; et l'*hétérogénéité* causée par les covariables inconnues, qui est donc théoriquement imprévisible, à moins de connaître toutes les informations pertinentes (ce qui, comme on peut le montrer,

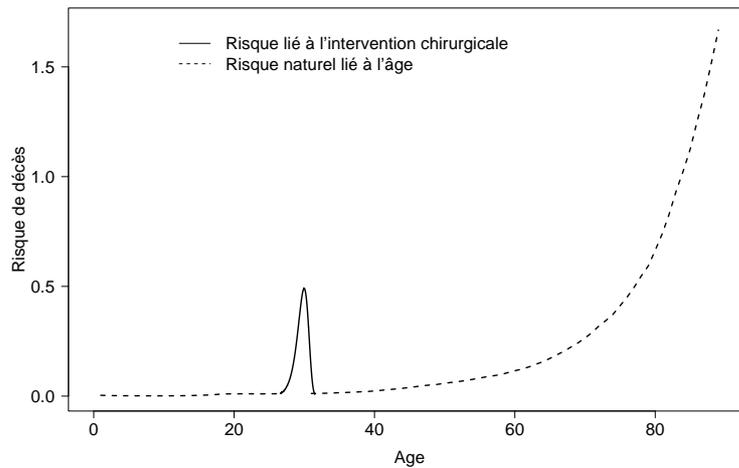


FIG. 3.1 – Exemple de risque de décès pour un patient subissant une intervention chirurgicale.

ne sera jamais le cas). Il y a des avantages à séparer ces deux sources de variation car l'hétérogénéité contrairement à la variabilité peut expliquer certains résultats "inattendus" ou peut fournir une autre explication de certains résultats ; comme par exemple dans le cas où les risques sont non proportionnels.

Dans un modèle à risques proportionnels, la négligence d'un sous-ensemble de covariables importantes conduit à des estimations biaisées des coefficients de régression et du taux de risque. La raison d'un tel biais réside dans le fait que le taux de risque temps-dépendant se traduit par des changements différents dans la composition de la population étudiée dans le temps selon les variables explicatives.

Supposons un échantillon étudié dans lequel certaines personnes "éprouvent" un plus grand risque d'évènement que d'autres, intuitivement ces personnes expérimenteront l'évènement plus tôt. Ainsi, le reste des personnes à risque a tendance à former un groupe plus ou moins sélectionné avec un risque d'évènement plus faible. Une estimation du taux de risque individuel, sans tenir compte de la fragilité non observée, devrait donc sous-estimer la véritable fonction de risque et l'ampleur de la sous-estimation devrait augmenter à mesure que le temps avance.

Le modèle de fragilité univarié est une extension du modèle de Cox dans lequel le risque d'un individu dépend en plus d'une variable aléatoire non observable Y , qui agit multiplicativement sur la fonction de risque de base λ :

$$\begin{aligned}\lambda(t|Y, Z) &= Y\lambda(t|Z) \\ &= Y\lambda_0(t) \exp(\beta Z).\end{aligned}\tag{3.1}$$

La fragilité Y est une variable aléatoire variant avec la population et qui diminue ($Y < 1$) ou augmente ($Y > 1$) le risque individuel. Le point le

plus important ici est que la fragilité est non observable. La fonction de survie S respective est donnée par

$$\begin{aligned} S(t|Y, Z) &= \exp(-\Lambda(t|Y, Z)), \\ &= \exp\left(-\int_0^t \lambda(u|Y, Z) du\right), \\ &= \exp\left(-Y \exp(\beta Z) \int_0^t \lambda_0(u) du\right). \end{aligned} \quad (3.2)$$

$S(t|Y, Z)$ peut être interprétée comme la fraction d’individus survivant au temps t après le début du suivi sachant le vecteur de covariables observables Z et la fragilité Y . Notons, que les relations (3.1) et (3.2) décrivent le même modèle en utilisant différentes notions. Jusqu’à présent, le modèle a été décrit au niveau des individus. La fonction de survie de la population totale est la moyenne des fonctions de survie sur la fonction de fragilité (3.2). Elle peut être considérée comme la fonction de survie d’un individu tiré aléatoirement, et correspond à ce qui est réellement observé. Ce qui peut être observé dans la population est le résultat net d’un certain nombre de personnes avec différents valeurs de Y .

Plusieurs choix pour la distribution de la fragilité sont possibles. Les distributions de fragilité les plus souvent utilisées sont les suivantes :

- ★ La distribution **Gamma** (Clayton, (19); Vaupel *et al.* (57)).
Cette famille de distributions inclut entre autres les lois exponentielles, les lois de sommes de variables aléatoires indépendantes suivant une même loi exponentielle, ainsi que la loi du χ^2 .
- ★ La distribution **Positive Stable** (Hougaard (36)).
Une condition des distributions stables est que la somme de n copies indépendantes de la distribution doit avoir la même distribution que la distribution multipliée par un facteur d’échelle. La particularité de la distribution positive stable est que le paramètre d’échelle est égal au paramètre de forme. Ainsi, il s’agit d’une distribution à un seul paramètre.
- ★ La distribution **PVF (Power Variance Function)** (Hougaard, (35)).
Cette famille de distributions est une distribution à trois paramètres regroupant, en outre, les distributions Gamma et Stable Positive.
- ★ La distribution de **Poisson Composée** (Aalen (1), (2)).
Cette distribution résulte d’une somme d’un nombre aléatoire de variables Gamma indépendantes. Le nombre de termes dans la somme est distribuée selon une loi de Poisson.
- ★ La distribution **Log-Normale** (McGilchrist et Aisbett, (43)).
Une variable aléatoire Y est dite distribuée de façon log-normale lorsque $X = \log Y$ suit une distribution normale.

Comme certaines de ces distributions possèdent plusieurs écritures selon les valeurs de leurs paramètres, elles ne sont pas détaillées.

3.2 MODÈLES DE FRAGILITÉ “MULTIVARIÉS”

Une deuxième application importante des modèles de fragilité concerne le domaine des données de survie multivariées. Ce type de données est

obtenu, par exemple, si les durées de vie (ou de temps d'apparition d'une maladie) de familles (jumeaux, parent-enfant), ou des évènements tels que les infections récurrentes pour un même individu sont prises en compte. Dans de tels cas, l'indépendance entre les temps de survie à l'intérieur d'un groupe ne peut pas être présumée. Les modèles multivariés sont capables de prendre en compte la présence de dépendance entre ces évènements. Une approche très couramment utilisée et générale est de préciser l'indépendance entre certaines données observées conditionnellement à un ensemble de variables latentes ou non observées, (Hougaard (37)). La dépendance dans la structure du modèle multivarié résulte de la présence d'une variable latente dans les modèles conditionnels pour des temps de survie observés plusieurs fois. Par exemple, prenons $S(t_1|Y, Z_1)$ et $S(t_2|Y, Z_2)$ les fonctions de survie conditionnelles de deux personnes avec différents vecteurs de covariables observées Z_1 et Z_2 , respectivement, (voir l'équation (3.2)). L'utilisation d'une hypothèse de distribution pour les variables latentes (par exemple, en utilisant une distribution Gamma, Log-Normale ou Stable) induit alors un modèle pour les données observées multivariées. Dans le cas d'observations appariées, la fonction de survie à deux dimensions est de la forme

$$S(t_1, t_2|Z_1, Z_2) = \int_0^\infty S(t_1|Y = y, Z_1)S(t_2|Y = y, Z_2)g(y)dy, \quad (3.3)$$

où g représente la densité de la fragilité Y . En effet, comme les temps de survie des individus 1 et 2 d'un même groupe sont supposées indépendantes sachant la fragilité de ce groupe :

$$\begin{aligned} S(t_1, t_2|Y, Z_1, Z_2) &= P(T_1 > t_1, T_2 > t_2|Y, Z_1, Z_2) \\ &= P(T_1 > t_1|Y, Z_1)P(T_2 > t_2|Y, Z_2) \\ &= S(t_1|Y, Z_1)S(t_2|Y, Z_2). \end{aligned}$$

Ainsi, la fonction de survie à deux dimensions s'écrit :

$$\begin{aligned} S(t_1, t_2|Z_1, Z_2) &= \int_0^\infty S(t_1, t_2|Y = y, Z_1, Z_2)g(y)dy \\ &= \int_0^\infty S(t_1|Y = y, Z_1)S(t_2|Y = y, Z_2)g(y)dy. \end{aligned}$$

Dans le cas de jumeaux, $S(t_1, t_2)$ désigne la proportion de paires de jumeaux avec le jumeau 1 survivant à t_1 et le jumeau 2 survivant à t_2 . Les modèles de fragilité pour les données de survie multivariées sont obtenus sous l'hypothèse d'indépendance conditionnelle en précisant les variables latentes qui agissent multiplicativement sur les risques de base.

3.2.1 Le modèle de fragilité partagée

Le modèle de fragilité partagée est pertinent pour des temps d'évènements d'individus corrélés, les organes doubles et les mesures répétées. Les individus dans un groupe sont censés partager la même fragilité Y , et c'est pourquoi ce modèle est appelé modèle de fragilité partagée. Il a

été présenté par Clayton (19) et étudié de manière approfondie par Hougaard (37). Les temps de survie sont supposés être indépendants conditionnellement à la fragilité partagée (commune). Pour faciliter la présentation, nous considérons le cas de groupes avec des paires d'individus (par exemple les temps événements de jumeaux ou de parent-enfant). Les extensions aux données multivariées sont naturelles. Conditionnellement à la fragilité Y , la fonction de risque d'un individu dans une paire est de la forme $Y \lambda_0(t) \exp(\beta Z)$, où la valeur de Y est commune aux deux individus de la paire, et est donc la cause de la dépendance entre le temps de survie au sein des couples. L'indépendance des temps de survie dans une paire correspond à une distribution de fragilité dégénérée ($Y = 1$, $\text{Var}(Y) = \sigma^2 = 0$). Dans tous les autres cas avec $\sigma^2 > 0$, la dépendance est positive par construction du modèle. Conditionnellement à Y , la fonction de survie bivariée est donnée par

$$S(t_1, t_2|Y) = S(t_1)^Y S(t_2)^Y. \quad (3.4)$$

En effet,

$$\begin{aligned} S(t_1, t_2|Y) &= S(t_1|Y)S(t_2|Y) \\ &= \exp\left(-\int_0^{t_1} \lambda(u|Y)du\right) \exp\left(-\int_0^{t_2} \lambda(u|Y)du\right) \\ &= \exp\left(-Y \int_0^{t_1} \lambda(u)du\right) \exp\left(-Y \int_0^{t_2} \lambda(u)du\right) \\ &= \exp\left(-\int_0^{t_1} \lambda(u)du\right)^Y \exp\left(-\int_0^{t_2} \lambda(u)du\right)^Y \\ &= S(t_1)^Y S(t_2)^Y. \end{aligned}$$

Dans la plupart des applications, il est supposé que la distribution de fragilité (c'est-à-dire la variable aléatoire Y) est une distribution Gamma de moyenne 1 et de variance σ^2 . La moyenne de la fonction de survie conditionnelle, sous cette hypothèse, conduit à des fonctions de survie de la forme

$$S(t_1, t_2) = (S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1)^{-1/\sigma^2}. \quad (3.5)$$

La démonstration étant un peu longue, elle se situe donc en Annexe page 128. Notons que toutes les formules énoncées pour le cas bivarié s'étendent au cas multivarié.

La fragilité partagée explique la corrélation entre les sujets dans les groupes. Cependant, elle comporte certaines limites. Tout d'abord, elle suppose que les facteurs non observés sont les mêmes dans le groupe, ce qui ne reflète pas toujours la réalité. Par exemple, à certains moments, il peut être inapproprié de supposer que tous les partenaires dans un groupe partagent tous leurs facteurs de risque non observés. Deuxièmement, la dépendance entre les temps de survie au sein du groupe est basée sur les distributions marginales des temps de survie. Toutefois, lorsque les covariables sont présentes dans un modèle à risques proportionnels avec fragilité distribuée selon une loi Gamma, le paramètre de dépendance et l'hétérogénéité de la population sont confondus (Clayton et Cuzick, (18)).

Ceci implique que la distribution jointe ne peut être différenciée des distributions marginales (Hougaard, (35), (36)). Troisièmement, dans la plupart des cas, une fragilité à une dimension ne peut que provoquer une association positive au sein du groupe. Toutefois, il y a certaines situations où les temps de survie des sujets, dans le même groupe sont associés négativement. Par exemple, dans l'Étude de Transplantation de Coeur de Stanford, plus une personne doit attendre pour un coeur disponible, moins il ou elle est susceptible de survivre après la transplantation. Par conséquent, les temps d'attente et les temps de survie peuvent être négativement associés. Pour éviter les limitations mentionnées ci-dessus les modèles de fragilité corrélée ont été développés. Nous évoquons ceux-ci brièvement car il en existe un grand nombre.

3.2.2 Le modèle de fragilité corrélée

A l'origine, les modèles de fragilité corrélée ont été développés pour l'analyse des données de temps d'évènements bivariés, dans laquelle deux variables aléatoires associées sont utilisées pour caractériser l'effet de la fragilité pour chaque paire. Par exemple, une variable aléatoire est attribuée pour le partenaire 1 et une pour le partenaire 2 afin qu'ils ne soient plus contraints d'avoir une fragilité commune. Ces deux variables sont associées et ont une distribution conjointe. La connaissance de l'une d'entre elles n'implique pas nécessairement la connaissance de l'autre. Il n'y a plus de restriction sur le type de corrélation. Ces deux variables peuvent également être négativement associées, ce qui induit une association négative entre les temps de survie.

Nous présentons brièvement un exemple de modèle de fragilité corrélée, il s'agit de l'étude de la mortalité de jumeaux danois de Yashin et Iachine (61). Dans leur article, ils posent T_i et Y_i ($i = 1, 2$) les temps de survie et les fragilités de deux jumeaux, respectivement. De plus, les risques individuels propres aux deux jumeaux s'écrivent $\lambda(Y_i, t) = Y_i \lambda_0(t)$, $i = 1, 2$. L'hypothèse fondamentale de ce modèle est que les temps de survie T_1 et T_2 sont conditionnellement indépendants sachant Y_1 et Y_2 avec $Y_i = \sum_{j=1}^n Y_{ij}$ ($i = 1, 2$), où Y_{ij} et Y_{im} sont des variables aléatoires indépendantes, non-négatives et Gamma-distribuées ($\forall j, m, j \neq m$). Ils supposent $Y_{1j} = W_{0j} + W_{1j}$ et $Y_{2j} = W_{0j} + W_{2j}$, où W_{ij} , $i = 0, 1, 2$ et $j = 1, 2, \dots, n$, sont des variables aléatoires indépendantes, non-négatives et Gamma-distribuées : $W_{0j} \sim G(k_{0j}, \lambda)$ et $W_{1j}, W_{2j} \sim G(k_{1j}, \lambda)$. Ainsi, les deux fragilités Y_{1j} et Y_{2j} sont corrélées. Nous ne mentionnons pas toute la théorie découlant de ces hypothèses, seulement, il résulte de ce modèle la distribution de survie bivariée de la forme

$$S(t_1, t_2) = \frac{S(t_1)^{1-\rho} S(t_2)^{1-\rho}}{(S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1)^{\rho/\sigma^2}} \quad (3.6)$$

où ρ représente la corrélation entre le jumeau 1 et le jumeau 2.

D'autres modèles de fragilité Gamma corrélée ont été appliqués au cas des jumeaux (Zahl (62), Wienke *et al.* (59), ...).

Les exemples de l'utilisation des modèles de fragilité multivariés sont divers et soulignent l'importance de cette famille de modèles statistiques

pour la survie des données. De plus, pour chaque situation, un modèle différent peut être créé.

3.3 LE MODÈLE DE FRAGILITÉ PROPOSÉ

Dans cette partie, nous présentons le modèle de fragilité développé pour notre problématique clinique. Nous choisissons une structure de modèle de fragilité partagée. En effet, nous sommes en présence de patients peuvent être groupés selon des variables. Nous avons suspecté, dans le chapitre 2, que les patients issus du même centre de transplantation sont dépendants. Il en est de même dans le cas des patients transplantés la même année. La fragilité est donc nécessaire pour représenter la dépendance des patients issus d'un même groupe (centre ou année de transplantation).

Nous considérons un échantillon constitué de n groupes indicés par i ($i = 1, \dots, n$). Chacun de ces groupes contient n_i individus repérés par l'indice j ($j = 1, \dots, n_i$). Les individus issus d'un même groupe sont considérés dépendants entre eux. Cette dépendance est contenue dans les termes de fragilité propres à chaque groupe i . Nous les notons y_i .

Pour l'individu j du groupe i , nous posons T_{ij} le temps de survie égal au minimum entre le temps de censure et d'apparition de l'évènement. Nous supposons la censure non informative. De même, nous posons $Z_{ij} = (Z_{ij1}, \dots, Z_{ijp})$ le vecteur de p covariables considérées comme indépendantes du temps.

Conditionnellement au terme de fragilité y_i et aux covariables observées Z_{ij} , la fonction de risque de l'individu j du groupe i est de la forme (Aalen (1) et Klein (39)) :

$$\lambda(t_{ij}|y_i) = y_i \lambda_0(t_{ij}) \exp(\beta z_{ij}) \quad (3.7)$$

où $\lambda_0(t_{ij})$ est la fonction de risque de base et $\beta = (\beta_1, \dots, \beta_p)'$ le vecteur des coefficients de régression associés à z_{ij} .

A partir de (3.7), la survie au temps t_{ij} sachant la fragilité du groupe i est

$$S(t_{ij}|y_i) = \exp[-y_i \Lambda(t_{ij})] \quad (3.8)$$

avec $\Lambda(t_{ij})$ la fonction de risque cumulée pour l'individu j du groupe i .

La survie marginale au temps t_{ij} peut s'écrire sous la forme (Hougaard (37)) :

$$\begin{aligned} S(t_{ij}) &= E[S(t_{ij}|y_i)] \\ &= E[\exp(-y_i \Lambda(t_{ij}))] \\ &= L[\Lambda(t_{ij})] \end{aligned} \quad (3.9)$$

avec L la transformée de Laplace d'une variable U :

$$\begin{aligned} L(a) &= \int_0^{\infty} \exp(-au)g(u)du \\ &= E[\exp(-aU)]. \end{aligned}$$

De même, la survie multivariée pour le groupe i s'écrit :

$$S(t_{i1}, \dots, t_{in_i}) = L \left[\sum_{j=1}^{n_i} L^{-1}(S(t_{ij})) \right]. \quad (3.10)$$

En effet, d'après les équations (3.4) et (3.8)

$$\begin{aligned} S(t_{i1}, \dots, t_{in_i} | y_i) &= S(t_{i1} | y_i) \dots S(t_{in_i} | y_i) \\ &= \exp[-y_i(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i}))]. \end{aligned} \quad (3.11)$$

Ainsi,

$$S(t_{i1}, \dots, t_{in_i}) = L[\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i})]. \quad (3.12)$$

Grâce à l'équation (3.9), le risque cumulé pour l'individu j du groupe i peut aussi s'écrire de la manière suivante :

$$\Lambda(t_{ij}) = L^{-1}[S(t_{ij})]. \quad (3.13)$$

En remplaçant l'équation (3.13) dans l'expression (3.12), on retrouve bien la formule (3.10) de la survie multivariée pour le groupe i . Une démonstration plus détaillée se trouve en Annexe page 125.

A partir de (3.10), nous déduisons la fonction de densité multivariée :

$$\begin{aligned} f(t_{i1}, \dots, t_{in_i}) &= (-1)^{n_i} \frac{\partial^{n_i}}{\partial t_{i1} \dots \partial t_{in_i}} S(t_{i1}, \dots, t_{in_i}) \\ &= (-1)^{n_i} \left(\prod_{j=1}^{n_i} \lambda(t_{ij}) \right) L^{(n_i)}(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i})). \end{aligned} \quad (3.14)$$

Pour faciliter la lecture, la démonstration de ce résultat est donnée en Annexe page 126.

Soient δ_{ij} , l'indicatrice de décès pour l'individu j du groupe i et $\delta_i = \sum_{j=1}^{n_i} \delta_{ij}$ le nombre total de décès dans le groupe i . Nous supposons la censure indépendante de la fragilité, *i.e.* la censure des patients ne dépend pas du groupe auquel ils appartiennent. Ainsi, la version censurée de la fonction de densité (3.14) est :

$$(-1)^{\delta_i} \left(\prod_{j=1}^{n_i} \lambda(t_{ij})^{(\delta_{ij})} \right) L^{(\delta_i)}(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i})). \quad (3.15)$$

La preuve est fournie en Annexe page 127.

La vraisemblance s'écrit alors :

$$V = \prod_{i=1}^n \left\{ \left(\prod_{j=1}^{n_i} (\lambda_0(t_{ij}) \exp(\beta z_{ij}))^{\delta_{ij}} \right) (-1)^{\delta_i} L^{\delta_i} \left(\sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta z_{ij}) \right) \right\} \quad (3.16)$$

où $\Lambda_0(t_{ij}) = \int_0^{t_{ij}} \lambda_0(x) dx$ est le risque de base cumulé.

La logvraisemblance de l'équation (3.16) découle naturellement :

$$\begin{aligned} \log(V) = & \sum_{i=1}^n \left\{ \sum_{j=1}^{n_i} \delta_{ij} [\log(\lambda_0(t_{ij})) + \beta z_{ij}] \right. \\ & \left. + \log \left((-1)^{\delta_i} L^{\delta_i} \left(\sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta z_{ij}) \right) \right) \right\}. \quad (3.17) \end{aligned}$$

Pour obtenir la fonction de logvraisemblance finale, il suffit alors de définir la fonction de risque de base $\Lambda_0(\cdot)$ et la distribution des termes de fragilité.

Nous supposons que l'unité statistique est le groupe et qu'elle suit une loi Gamma, $G(\gamma, \gamma)$, dont la densité est donnée par :

$$g(y) = \frac{\gamma^\gamma y^{\gamma-1} \exp(-\gamma y)}{\Gamma(\gamma)}, \quad \forall \gamma \geq 0 \quad (3.18)$$

où $\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} \exp(-t) dt$ est la fonction Gamma. Nous choisissons une loi Gamma de paramètre d'échelle et de forme γ pour la variable de fragilité pour éviter le problème de la surparamétrisation et obtenir une espérance égale à 1 (Aalen (3)). La variance de la variable de fragilité est donc égale à $1/\gamma$. De plus, la distribution Gamma est l'une des plus utilisées car elle possède de bonnes propriétés mathématiques et, dans notre cas, la transformée de Laplace pour une telle loi est égale à $L(a) = \gamma^\gamma (\gamma + a)^{-\gamma}$.

Concernant le risque, le choix de la fonction de risque de base est important : soit il est inconnu, il s'agit de la logique de Cox, soit il s'agit d'un risque dans un modèle multiplicatif général et alors il peut être paramétré. Le dernier exemple est celui de la distribution de Weibull :

$$\lambda_0(t) = \nu \left(\frac{1}{\sigma} \right)^\nu t^{(\nu-1)}, \quad t \geq 0, \quad \sigma \text{ et } \nu > 0. \quad (3.19)$$

En introduisant (3.18) et (3.19) dans l'équation de la logvraisemblance (3.17), nous obtenons la logvraisemblance suivante :

$$\begin{aligned} \log(V) = & \sum_{i=1}^n \left\{ \sum_{j=1}^{n_i} \delta_{ij} (\log(\nu) - \nu \log(\sigma) + (\nu - 1) \log(t_{ij}) + \beta z_{ij}) \right. \\ & + \gamma \log(\gamma) + \log \Gamma(\gamma + \delta_i) - \log \Gamma(\gamma) \\ & \left. - (\gamma + \delta_i) \log \left(\gamma + \sum_{j=1}^{n_i} \left[\left(\frac{t_{ij}}{\sigma} \right)^\nu + \exp(\beta z_{ij}) \right] \right) \right\}. \quad (3.20) \end{aligned}$$

Nous avons donc trois paramètres à estimer sans compter ceux relatifs aux covariables.

Le modèle que nous venons de présenter est un modèle de fragilité Gamma partagée dont la fonction de risque de base suit une distribution de Weibull. Nous allons maintenant l'appliquer à la base de données DI-VAT.

3.4 APPLICATION AUX TRANSPLANTÉS RÉNAUX

Il semble que la survie des transplantés rénaux soit différente selon leur centre de transplantation ainsi que selon les années de transplantation. Nous testons donc deux modèles : un avec une variable de fragilité portant sur les centres de transplantation et un portant sur les années de transplantation. Ceci permet de modéliser la corrélation des individus transplantés dans les mêmes centres et les mêmes années, respectivement. Dans un premier temps, nous utilisons un modèle semi-paramétrique de Cox pour tester ces modèles. Ensuite, nous utilisons le modèle proposé, c'est-à-dire le modèle paramétrique de Weibull avec fragilité Gamma partagée. Les modèles seront comparés et interprétés. Dans les deux types de modèle, l'effet de la fragilité est testé au moyen d'un rapport de vraisemblance entre le modèle en question et le même modèle sans fragilité.

3.4.1 Modèle de Cox avec fragilité

Pour estimer les paramètres de ce modèle, nous utilisons la fonction *coxph* disponible sous le logiciel R. L'option *frailty* de cette fonction permet d'estimer un modèle de Cox avec fragilité Gamma partagée. Le modèle s'écrit ainsi :

$$\lambda(t|Y, Z) = Y\lambda_0(t) \exp(\beta'Z).$$

où Y correspond à la variable de fragilité.

Dans cette partie, nous ne spécifions pas le terme concernant la fragilité dans les tableaux de résultats. Notre intérêt porte seulement sur son effet.

Fragilité sur les centres de transplantation

Ici, la variable de fragilité modélise la dépendance des individus appartenant au même centre de transplantation. Le centre de transplantation n'est donc pas testé comme covariable dans ce modèle mais au travers de l'effet aléatoire. Les tableaux (3.1) et (3.2) résument les résultats de cette modélisation. Le tableau (3.1) permet de visualiser les différentes covariables significatives après les diverses méthodes de sélection et le tableau (3.2) récapitule les estimations des paramètres des covariables significatives dans le modèle de Cox multivarié.

Les tableaux (3.1) et (3.2) montrent que les résultats sont quasiment identiques à ceux des modèles de Cox obtenus dans le chapitre 2 (le modèle de Cox classique et les modèles de Cox stratifiés). De plus, nous

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité			
Anti HLA			
Traitement immunosuppresseur	×		
Type de maladie initiale	×		
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie			
Néoplasie	×		
Age du donneur	×		
Sexe du donneur			
Année de la greffe	×	×	×

TAB. 3.1 – *Modèle de Cox avec variable de fragilité sur les centres de transplantation : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).*

Paramètres	coef	exp(coef)	se	p-value
Sexe du receveur	-0,339	0,713	0,148	0,022
Age du receveur	1,126	3,084	0,138	0,000
DGF	0,371	1,450	0,137	0,007
Année de la greffe	-0,636	0,530	0,151	0,000

TAB. 3.2 – *Modèle de Cox avec Fragilité sur les centres de transplantation ($\log(V) = -1525,397$).*

remarquons que ce sont toujours les mêmes covariables qui sont significatives par rapport à ces modèles. Il s'agit des covariables âge et sexe du receveur, année de la greffe et DGF. L'estimation des effets de ces covariables sur la survie des patients transplantés diffère très légèrement (de l'ordre de 0,002 environ) par rapport à leur estimation dans le modèle de Cox classique. Par ailleurs, la statistique du test de rapport de vraisemblance entre le modèle de Cox avec fragilité et celui sans fragilité vaut 0,616, correspondant à une p -valeur supérieur à 0,50. Ceci montre que l'ajout d'un terme de fragilité dans le modèle de Cox ne semble pas nécessaire. Ainsi, il ne semble pas y avoir d'effet centre de transplantation. De plus, la variance de la fragilité vaut 0,002 et sa p -value vaut 0,32. Par contre, en prenant en compte la dépendance entre les individus d'un même centre de transplantation, être un homme, être âgé de plus de 55 ans, avoir un greffon qui met plus de six jours à reprendre fonction et avoir été greffé avant 1996 sont des facteurs de risque de décès après une transplantation de rein. Cette conclusion concernant les facteurs de risque est la même que celle des modèles de Cox sans fragilité obtenus au chapitre 2.

Fragilité sur les années de transplantation

Dans ce paragraphe, la variable de fragilité modélise la corrélation des patients greffés la même année. Ainsi, l'année de transplantation n'est pas testée comme facteur dans ce modèle. Les résultats sont visibles dans les tableaux (3.3) et (3.4).

Les covariables significatives dans cette analyse sont l'âge et le sexe

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur	×		
Type de maladie initiale	×		
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie	×		
Néoplasie	×		
Age du donneur	×		
Sexe du donneur			
Centre			

TAB. 3.3 – Modèle de Cox avec variable de fragilité sur les années de greffe : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).

Covariables	coef	exp(coef)	se	p-value
Sexe du receveur	-0,330	0,719	0,148	0,026
Age du receveur	1,111	3,038	0,139	0,000
DGF	0,366	1,442	0,138	0,008

TAB. 3.4 – Modèle de Cox avec Fragilité sur les années de transplantation ($\log(V) = -1522,27$).

du receveur ainsi que le DGF. Ici encore, les résultats des estimations des paramètres sont très proches de ceux d'un modèle de Cox classique. De plus, la statistique du test de rapport de vraisemblance entre le modèle de Cox avec fragilité et celui sans fragilité vaut 25,122, correspondant à une p -value inférieure à 0,001. Ceci montre que l'ajout d'un terme de fragilité dans le modèle de Cox semble significatif. Il semble donc y avoir un effet année de transplantation. Ainsi, la dépendance entre les patients greffés la même année de transplantation se confirme. Par ailleurs, la variance de la variable de fragilité vaut 0,082 et sa p -value vaut 0,02. Sachant cette dépendance des individus transplantés la même année, être un homme, avoir plus de 55 ans et avoir un greffon qui met plus de six jours à reprendre fonction sont des facteurs de risque de décès après une greffe de rein.

3.4.2 Modèle de régression de Weibull avec fragilité

Dans cette partie, nous introduisons la notion de fragilité dans un modèle paramétrique de Weibull. Il s'agit du modèle que nous avons proposé dans la partie théorique de ce chapitre. Comme précédemment, nous proposons deux modèles, un dans lequel la fragilité représente la liaison entre les patients greffés dans les mêmes centres de transplantation et un autre dans lequel elle représente la liaison entre les patients greffés les mêmes années.

Fragilité sur les centres de transplantation

Nous testons ici le modèle avec la fragilité sur les centres de transplantation. Le tableau (3.5) montre que l'âge, le sexe du receveur, le DGF et l'année de transplantation sont des facteurs ayant une influence significative sur le risque de décéder après une greffe de rein sachant que la dépendance des patients greffés dans un même centre est prise en considération.

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur			
Type de maladie initiale			
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie			
Néoplasie			
Age du donneur	×		
Sexe du donneur			
Année de la greffe	×	×	×

TAB. 3.5 – Modèle de Weibull avec variable de fragilité sur les centres de transplantation : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).

Paramètres	coef	exp(coef)	se	p-value
σ	145,760	.	84,139	.
ν	0,850	.	0,053	< 0,05*
γ	1,556	.	0,294	> 0,05*
Sexe du receveur	-0,338	0,713	0,148	0,022
Age du receveur	1,108	3,028	0,138	0,000
DGF	0,345	1,413	0,140	0,014
Année de la greffe	-0,596	0,551	0,149	0,000

* Test du rapport de vraisemblance : $\{H_0 : \nu = 1\}$.

• Test du rapport de vraisemblance entre le modèle avec fragilité et le même modèle sans fragilité.

TAB. 3.6 – Modèle de Weibull avec fragilité sur les centres de transplantation ($\log(V) = -1123,90$).

Le tableau (3.6) confirme les résultats des modèles de Cox avec fragilité. Il ne semble pas y avoir d'effet centre de transplantation d'après les tests de rapport de vraisemblance. En effet, la statistique du test de rapport de vraisemblance entre le modèle de Weibull avec fragilité et celui sans fragilité (estimé au chapitre 2) vaut $-2,192$, ce qui correspond à une p -valeur supérieure à $0,05$. Nous testons par ailleurs $\{H_0 : \nu = 1\}$, *i.e.* si la fonction de risque de décès suit une distribution Exponentielle. Un rapport de vraisemblance entre un modèle de Weibull avec fragilité et un modèle Exponentiel avec fragilité est établi. La statistique de ce test vaut $4,55$ correspondant à une p -valeur inférieure à $0,05$. Ce résultat nous conduit à rejeter l'hypothèse selon laquelle la fonction de risque de décès

est Exponentielle. De plus, sachant que la liaison entre les patients transplantés dans un même centre est prise en considération :

- * les hommes ont environ 1,402 fois plus de risque de décéder que les femmes,
- * les greffés de plus de 55 ans ont environ 1,108 fois plus de risque de décéder que les moins de 55 ans,
- * les patients dont le greffon met plus de six jours à reprendre fonction ont environ 1,413 fois plus de risque de décéder que ceux dont le greffon met moins de six jours,
- * les patients greffés avant 1996 ont approximativement 1,815 fois plus de risque de décéder que ceux greffés après 1996.

Ces risques diffèrent de l'ordre de 0,001 de ceux trouvés avec un modèle de Weibull sans fragilité (chapitre 2).

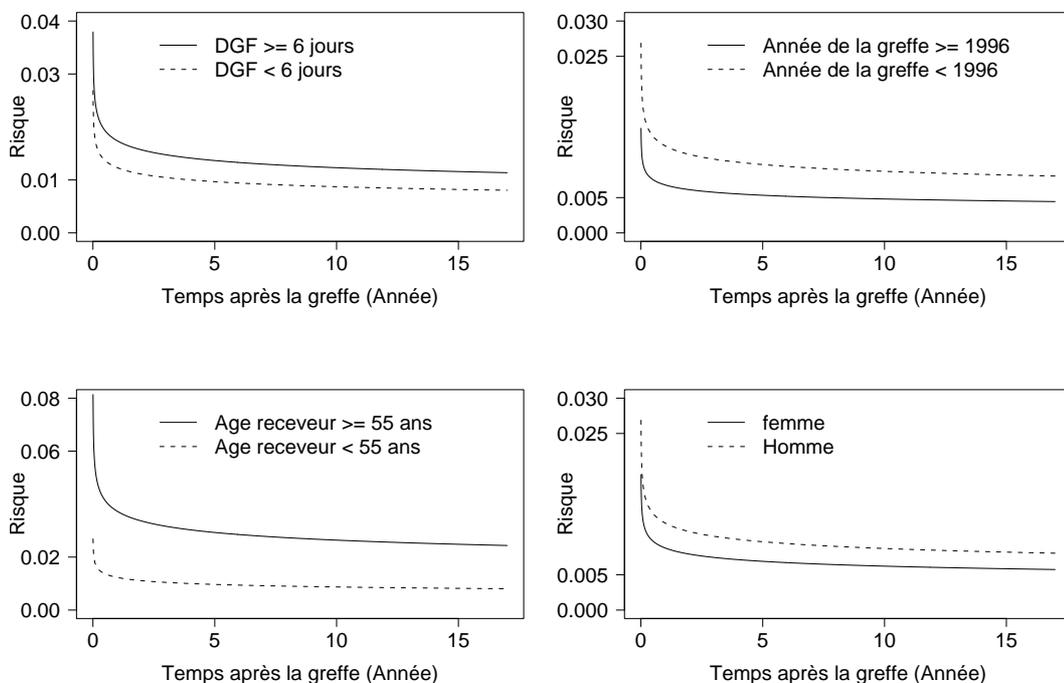


FIG. 3.2 – Fonction de risque selon les modalités des covariables sous le modèle de Weibull avec fragilité sur les centres de transplantation.

Le graphique (3.2) montre les fonctions de risque au cours du temps selon les modalités des différentes covariables significatives (âge, sexe du receveur, DGF et année de transplantation). Le risque de décès d'un homme est d'environ 0,03 juste après la greffe et se stabilise à 0,008 cinq ans après la greffe. Pour une femme, ce risque varie d'environ 0,02 jusqu'à 0,006. Par ailleurs, un patient greffé avant 55 ans a un risque de décès de 0,03 juste après la greffe jusqu'à 0,008 au delà de cinq ans alors que ce risque est de 0,08 jusqu'à 0,02 aux mêmes périodes pour un patient greffé après 55 ans. Un patient dont le greffon met moins de 6 jours à reprendre fonction a environ 0,03 risque de décéder juste après la transplantation jusqu'à environ 0,008 cinq ans après la transplantation. Par contre, celui dont le greffon met plus de 6 jours à reprendre fonction varie d'environ

0,04 juste après la greffe et se stabilise à 0,01 cinq ans après la greffe. Enfin, le risque de décès d'un patient greffé avant 1996 est d'environ 0,03 après la transplantation jusqu'à une stabilisation à 0,008 cinq ans après la transplantation. Pour un transplanté après 1996, le risque de décès varie de 0,015 juste après la greffe jusqu'à environ 0,004. En conclusion, être un homme, être âgé de plus de 55 ans, être greffé avant 1996 et avoir un greffon qui met plus de six jours à reprendre fonction sont des facteurs de risque de décès dès les premiers instants après la transplantation et ce jusqu'à environ 17 ans après. De plus, de manière générale, ces risques ont tendance à se stabiliser environ cinq ans après la greffe de rein. Nous remarquons aussi, que ces risques de décès sont quasiment du même ordre que ceux estimés par le modèle de Weibull sans fragilité du chapitre 2.

Fragilité sur les années de transplantation

Dans ce paragraphe, la fragilité porte sur les années de transplantation. Le tableau (3.7) montre que, sachant la prise en compte de la dépendance par année de transplantation des patients, les facteurs significatifs d'un décès après une greffe de rein sont l'âge, le sexe du receveur et le DGF.

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur			
Type de maladie initiale			
Antécédents cardiologiques	×		
Relation donneur / receveur			
Ischémie			
Néoplasie			
Age du donneur	×		
Sexe du donneur			
Centre			

TAB. 3.7 – Modèle de Weibull avec variable de fragilité sur les années de transplantation : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).

Paramètres	coef	exp(coef)	se	p-value
σ	172,491	.	70,912	.
ν	0,855	.	0,054	< 0,05*
γ	1,477	.	0,062	< 0,05•
Sexe du receveur	-0,322	0,725	0,148	0,029
Age du receveur	1,112	3,041	0,139	0,000
DGF	0,358	1,431	0,138	0,009

* Test du rapport de vraisemblance : $\{H_0 : \nu = 1\}$.

• Test du rapport de vraisemblance entre le modèle avec fragilité et le même modèle sans fragilité.

TAB. 3.8 – Modèle de Weibull avec Fragilité sur les années de transplantation ($\log(V) = -1102,92$).

Les résultats des estimations de ce modèle se trouvent dans la table (3.8). Cette table confirme les résultats des modèles de Cox avec fragilité

étant donné les résultats des tests de rapport de vraisemblance, il semblerait y avoir un effet année de transplantation. On peut donc conclure à une corrélation entre les patients greffés la même année. En effet, la statistique du test de rapport de vraisemblance entre le modèle de Weibull avec fragilité et celui sans fragilité (estimé au chapitre 2) vaut 62,23, correspondant à une p -value inférieure à 0,001. Par ailleurs, la statistique correspondant au test $\{H_0 : \nu = 1\}$, *i.e* la fonction de risque de décès suit une distribution Exponentielle, vaut 6,5 correspondant à une p -valeur inférieure à 0,02. Nous concluons donc que la fonction de risque de décès suit bien une distribution de Weibull.

De plus, sachant que les patients greffés la même année sont corrélés :

- ★ les hommes ont environ 1,380 fois plus de risque de décéder que les femmes,
- ★ les greffés de plus de 55 ans ont environ 1,112 fois plus de risque de décéder que les moins de 55 ans,
- ★ les patients dont le greffon met plus de six jours à reprendre fonction ont environ 1,431 fois plus de risque de décéder que ceux dont le greffon met moins de six jours.

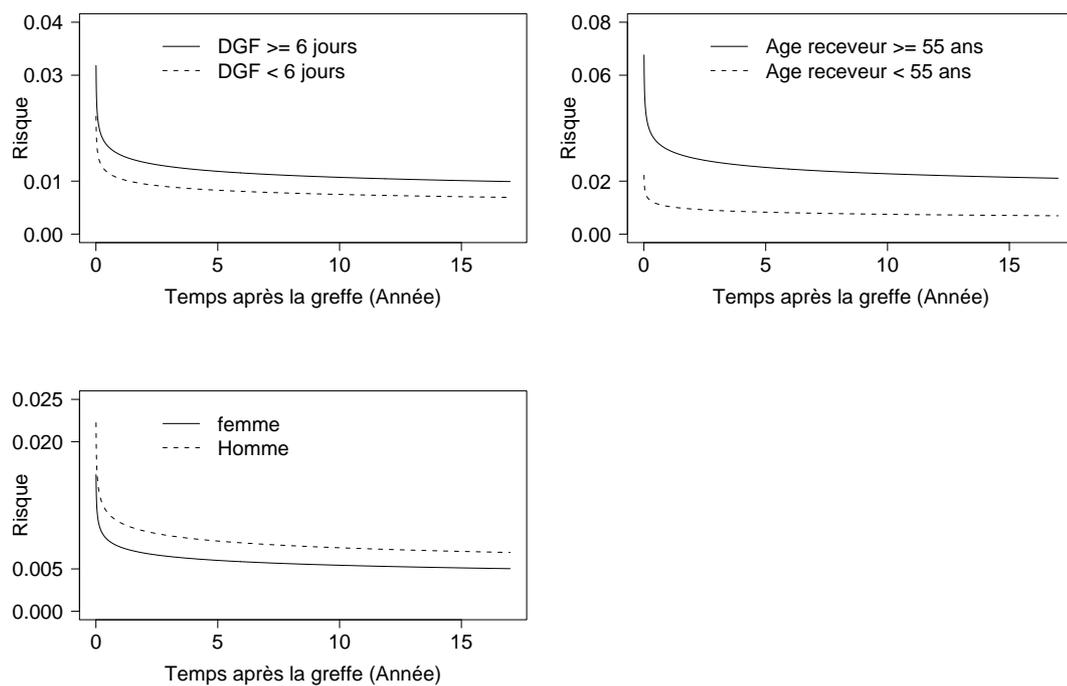


FIG. 3.3 – Fonction de risque selon les modalités des covariables sous le modèle de Weibull avec fragilité sur les années de transplantation.

Le graphique (3.3) montre les fonctions de risque au cours du temps selon les modalités des différentes covariables significatives (âge, sexe du receveur et DGF). Ces risques sont quasiment les mêmes que ceux observés pour le modèle dont la fragilité porte sur les centres de transplantation. Ces résultats confirment que les patients les plus à risque de décéder après une greffe de rein sont les hommes, les plus de 55 ans, ceux dont le greffon met plus de 6 jours à reprendre fonction. Ceci dès les

premiers instants après la transplantation et jusqu'à plus de 16 ans après. De même dans ce cas, les risques ont tendance à se stabiliser cinq ans après la greffe de rein. Nous remarquons, encore une fois, que ces risques de décès sont à peu près identiques à ceux du modèle de Weibull sans fragilité estimé au chapitre 2.

DISCUSSION-CONCLUSION

Ce chapitre introduit la notion de fragilité ainsi que ses différentes situations d'application. La fragilité est une variable aléatoire qui permet, entre autre, de prendre en compte l'hétérogénéité des patients issus d'un même groupe par rapport à d'éventuels sous groupes homogènes. Cette hétérogénéité est considérée comme non observable et elle reflète souvent des facteurs environnementaux ou des facteurs génétiques. Les groupes d'individus peuvent être des familles, des jumeaux ou par exemple un évènement observé avec répétition pour une même personne.

Dans l'exploration de la base DIVAT effectuée au chapitre 2, l'hypothèse classique des modèles de survie selon laquelle les patients transplantés sont indépendants est remise en cause. En effet, la survie des patients semble différente selon les centres et les années de transplantation. Cependant, il n'est pas possible de mesurer concrètement ces deux facteurs. Par exemple, nous savons que les stratégies de transplantation sont différentes selon les centres et que, de plus, elles ont évolué au cours du temps, mais nous ne pouvons pas vraiment mesurer ce facteur. C'est en quelque sorte un facteur caché dans les facteurs année et centre de transplantation. Ainsi, la fragilité permet de remédier à ce problème. Nous avons donc utilisé cette fragilité pour modéliser d'une part la dépendance des greffés issus d'un même centre de transplantation et d'autre part la dépendance des greffés la même année. Ces deux types de fragilité sont introduites dans deux modèles distincts : un modèle de Cox et un modèle de Weibull. Le modèle de Weibull est celui que nous proposons. Nous avons donc procédé à la construction de quatre modèles. Nous avons choisi de tester un modèle de Cox avec fragilité pour analyser si les résultats concernant la fragilité, à savoir si elle est significative, sont les mêmes que ceux du modèle proposé. Les résultats concernant les significativités des covariables et de la fragilité sont effectivement les mêmes entre les modèles de Weibull et de Cox. Retenons pour la suite de ce mémoire, qu'en utilisant des modèles de Cox et de Weibull comprenant la fragilité, l'hypothèse de la dépendance des patients greffés dans un même centre de transplantation n'est pas confirmée. Par contre, celle concernant la corrélation entre les patients greffés la même année est vérifiée. Il semble donc que la survie des patients soit différente selon les années de transplantation, par contre il ne semble pas qu'elle soit différente selon les centres dans lesquels sont transplantés les patients.

Dans chacun de ces modèles, la fragilité est modélisée par une distribution Gamma. En effet, cette distribution possède de bonnes propriétés. Un de ces avantages réside dans ses facilités calculatoires, ceci au travers des transformées de Laplace. Cependant, d'autres distributions peuvent être

utilisées pour modéliser cette fragilité : Positive Stable, Log-normale, ... Le choix de la distribution pour l'effet aléatoire est un problème majeur dans les modèles de fragilité. Même si chaque distribution possède ses atouts, dans la plupart des analyses effectuées dans la littérature, il ressort que la distribution Gamma est plutôt satisfaisante (en particulier Hougaard (37)). C'est pourquoi nous n'utilisons que celle-ci pour modéliser la fragilité dans les divers modèles de ce mémoire.

LA SURVIE RELATIVE

4

SOMMAIRE

4.1	DÉFINITION DES DIFFÉRENTES NOTIONS DE SURVIE	59
4.2	MÉTHODE D'ESTÈVE <i>et al.</i>	60
4.3	MODÈLE D'HAKULINEN ET TENKANEN	62
4.4	MODÈLE D'ANDERSEN <i>et al.</i>	63
4.5	TABLE DE MORTALITÉ	64
4.5.1	Définition et présentation	64
4.5.2	Analyse des tables vitales	65
4.6	CHOIX D'UN MODÈLE DE SURVIE RELATIVE	66
4.7	APPLICATION AUX TRANSPLANTÉS RÉNAUX	68
4.7.1	Résultats avec le modèle d'Andersen	68
4.7.2	Résultats avec le modèle paramétrique	71
	DISCUSSION-CONCLUSION	74

La notion de survie relative est apparue au début des années 40 et fut introduite par Berkson (8). Brièvement, elle se définit comme le ratio entre la survie observée et la survie attendue . En 1950, le concept fut défini de façon plus précise par Berkson et Gage (9). Les résultats de leur travaux ont permis d'obtenir une mesure objective de la proportion de patients qui décèdent des conséquences directes ou indirectes d'une maladie dans une population donnée. Les premières méthodes de calcul univariées furent introduites par Ederer et Heise (23), Ederer *et al.* (22) et Hakulinen (32). Elles sont plus connues sous les noms de Ederer II, Ederer I et Hakulinen, respectivement. Ces méthodes reposent essentiellement sur le calcul de la survie attendue et sur des méthodes d'estimation non paramétriques. Ensuite, des approches fondées sur la modélisation de la fonction de risque firent leur apparition. Breslow (11) a été le premier à proposer une modélisation multiplicative. Par la suite, d'autres modèles multiplicatifs ont été développés, notamment par Andersen et Vaeth (5). Par contre, plusieurs auteurs (Andersen et Vaeth (5), Buckley (13), Hakulinen et Tenkanen (33)) ont jugé plus naturel d'employer une modélisation additive. En effet, de tels modèles sont généralement plus plausibles d'un point de vue biologique et donnent un meilleur ajustement des données, surtout pour les données de survie de type populationnel. Ensuite, la modélisation multivariée fit son apparition. Elle a permis de combler

les besoins d'estimation de la survie, en fonction de plusieurs facteurs pronostiques. Breslow *et al.* (12), Andersen *et al.* (6) sont, entre autre, les auteurs de modèles multiplicatifs. Pour les modèles additifs, les plus connus sont ceux de Pocock *et al.* (48), Hakulinen et Tenkanen (33) et Estève *et al.* (25). Cependant, les modèles additifs restent les modèles les plus utilisés à cause de l'interprétation plus aisée des résultats.

Pour commencer ce chapitre, nous donnons les différentes notions de survie ainsi que la définition de la survie relative. Les différentes théories concernant les principaux modèles de survie relative, Estève *et al.*(25), Hakulinen et Tenkanen (33) et Andersen *et al.* (6), sont présentées. Ensuite, un modèle multiplicatif de survie relative paramétrique est exposé. En effet, c'est celui-ci qui nous sera nécessaire dans la suite de ce mémoire. De plus, nous présentons les informations que peuvent, entre autre, fournir les tables vitales produites par l'INSEE. Par ailleurs, le choix concernant la structure (additive ou multiplicative) d'un modèle de survie relative est discuté. Enfin, nous appliquons les deux types de modèles multiplicatifs à la base de données DIVAT. En effet, l'information sur la cause de décès des patients transplantés rénaux n'est pas fournie. Les différents résultats de l'application sont alors analysés en détails afin de mieux comprendre et d'apprécier les différents types d'interprétation obtenu pour chacun des modèles étudiés.

4.1 DÉFINITION DES DIFFÉRENTES NOTIONS DE SURVIE

Lorsque l'on aborde le sujet de l'analyse de survie d'un groupe, il faut au préalable préciser le type de survie dont il s'agit. En effet, on peut distinguer deux notions : la survie brute ou la survie nette (Estève *et al.* (26)). La survie brute (ou survie globale) représente la survie prenant en compte toutes les causes possibles de décès. Par exemple, une personne présentant un cancer du sein est soumise à une certaine force de mortalité liée à son cancer mais elle est également soumise à d'autres forces de mortalité, spécifiques ou non à sa personne. C'est aussi le cas pour l'ensemble de la population générale. En effet, une personne décédant dans un accident de la route peut soit être décédée par l'accident lui-même soit être décédée, par exemple, par un accident vasculaire cérébral au moment où elle conduisait. Le concept de la survie nette diffère dans la mesure où la pathologie faisant l'objet de l'étude de survie n'est pas l'unique source de mortalité des individus du groupe étudié. La survie nette représente donc la survie d'une cause donnée lorsque toutes les autres causes de décès ont été éliminées. Pour pouvoir l'estimer, on considère que la cause spécifique de décès que l'on étudie est indépendante des autres causes de décès. La survie nette peut être estimée par deux méthodes :

- * La méthode de la survie spécifique nécessite de savoir si la cause du décès est directement imputable à la pathologie étudiée, les décès pour d'autres causes sont alors traités comme des données censurées. Cependant, cette méthode est rarement utilisable car elle nécessite de connaître la cause exacte du décès. Cela est souvent rendu difficile à cause de l'imprécision des certificats de décès et de la faible fréquence des autopsies (et de l'indépendance, en conséquence, des autres causes de décès).
- * La méthode de la survie relative, ou survie corrigée, ne nécessite pas de savoir si la cause du décès est directement imputable à la pathologie étudiée puisque la survie observée du collectif étudié est « corrigée » par sa survie attendue issue des tables de mortalité de la population générale. En outre, l'information sur la cause réelle de décès n'est pas nécessaire. Seulement la date du décès est requise. Par conséquent, l'emploi de la survie relative permet d'éviter les problèmes d'imprécision ou de non-disponibilité des certificats de décès ainsi que de l'incertitude quant à la cause du décès. Le calcul de la survie relative demande seulement de savoir que le sujet a eu un diagnostic de maladie et qu'il était vivant ou mort à un moment précis.

La survie relative se définit comme le ratio entre la survie observée d'une population étudiée (survie brute définie précédemment) et la survie attendue d'une population de référence. Chaque individu de la population étudiée peut être comparé à un individu ayant les mêmes caractéristiques que lui (sexe, âge, etc...) dans la population de référence. La survie relative, $S_{\text{Rel}}(t)$, vise donc à « corriger » la survie observée du groupe étudié, $S_{\text{Obs}}(t)$, par la survie attendue dans une population générale en l'absence de la maladie étudiée, $S_{\text{Att}}(t)$. Pour l'instant, pour alléger les premières définitions, les différentes notions ne sont notées qu'en fonction du temps

t (temps depuis le diagnostic de la pathologie étudiée). La survie relative au temps t vaut :

$$S_{\text{Rel}}(t) = \frac{S_{\text{Obs}}(t)}{S_{\text{Att}}(t)}. \quad (4.1)$$

Ainsi, au temps t , la survie relative peut-être interprétée comme la proportion de sujets malades qui auraient survécu au moins au temps t et ce, dans la situation hypothétique où la maladie étudiée serait la seule cause possible de décès.

En notant respectivement $\lambda_{\text{Obs}}(t)$, $\lambda_{\text{Att}}(t)$ et $\lambda_{\text{Rel}}(t)$ le risque instantané observé du collectif suivi, attendu, et relatif à la pathologie étudiée, on déduit de la relation (4.1) :

$$\begin{aligned} \log S_{\text{Rel}}(t) &= \log S_{\text{Obs}}(t) - \log S_{\text{Att}}(t) \\ - \int_0^t \lambda_{\text{Rel}}(u) du &= - \int_0^t \lambda_{\text{Obs}}(u) du + \int_0^t \lambda_{\text{Att}}(u) du \\ \int_0^t \lambda_{\text{Obs}}(u) du &= \int_0^t \lambda_{\text{Att}}(u) du + \int_0^t \lambda_{\text{Rel}}(u) du \\ \int_0^t \lambda_{\text{Obs}}(u) du &= \int_0^t (\lambda_{\text{Att}}(u) + \lambda_{\text{Rel}}(u)) du \\ \lambda_{\text{Obs}}(t) &= \lambda_{\text{Att}}(t) + \lambda_{\text{Rel}}(t). \end{aligned} \quad (4.2)$$

Pour le calcul des $S_{\text{Att}}(t)$ ou $\lambda_{\text{Att}}(t)$, plusieurs possibilités existent. Une éventualité logique et évidente est de les considérer comme des quantités connues obtenues à partir des tables de mortalité de la population générale pour des sous-populations présentant des caractéristiques identiques à celles du groupe suivi, exceptée la maladie étudiée, telles que le sexe et l'âge (Ederer *et al.* (22), Hakulinen (32)). Les fonctions sont alors traitées comme des constantes, et non des estimations.

L'autre alternative est de les considérer comme des quantités inconnues, *i.e.* on modélise la survie attendue (ou le risque) par une distribution paramétrique. Dans les problématiques démographiques, la distribution de Gompertz-Makeham est souvent utilisée (distribution à trois paramètres). Sinon, bien que non présentée ici, la distribution de Weibull est adaptée à la modélisation des tables de mortalité de la population générale. Comme les tables de mortalité sont facilement disponibles (grâce à l'INSEE) et qu'une vraie valeur est toujours préférée à une estimation, nous choisissons de considérer les $S_{\text{Att}}(t)$ ou $\lambda_{\text{Att}}(t)$ à partir des tables de mortalité.

Pour exprimer la survie relative en fonction de facteurs pronostiques, les deux principales méthodes d'estimation reposant sur la modélisation de l'équation (4.2) sont celles de Estève *et al.* (25) et de Hakulinen et Tenkanen (33). Ces méthodes sont basées sur une approche de maximisation de la vraisemblance pour l'estimation des paramètres du modèle.

4.2 MÉTHODE D'ESTÈVE *et al.*

Une approche bien connue pour la modélisation de la survie relative, appelée approche par maximisation de la vraisemblance, a été proposée

par Estève *et al.* (25). Ce travail venait en fait parfaire les idées de Pocock *et al.* (48), Buckley (13) et Andersen *et al.* (6). Par cette approche, c'était la première fois que des formules explicites étaient diffusées pour l'implantation de la méthode. Comme son nom l'indique, cette méthode emploie la maximisation de la vraisemblance afin d'estimer les paramètres du modèle de survie relative et ce, à l'aide de données individuelles.

Selon Estève *et al.*, la fonction de risque relative s'écrit :

$$\lambda_{\text{Rel}}(t, z) = \exp(\beta z) \sum_{k=1}^m \tau_k I_k(t), \quad (4.3)$$

où t est le temps depuis le diagnostic de la pathologie étudié, $I_k(t)$ est une fonction indicatrice pour l'intervalle k et τ_k représente le risque instantané relatif à la pathologie étudiée. τ_k est supposé constant sur l'intervalle k . m est le nombre total d'années de suivi. En référence à l'équation (4.2), la fonction de risque observé s'écrit ainsi :

$$\lambda_{\text{Obs}}(t; a, z) = \lambda_{\text{Att}}(a + t, z_s) + \exp(\beta z) \sum_{k=1}^m \tau_k I_k(t), \quad (4.4)$$

avec a l'âge au diagnostic de la pathologie étudiée et z_s un sous-vecteur de z , un vecteur de covariables. En général, z_s correspond à l'âge et au sexe. Le problème statistique consiste à estimer β et τ en utilisant la méthode du maximum de vraisemblance.

Soit $(t_i, a_i, \delta_i, z_i)$ un échantillon observé où

- * $i = 1, \dots, n$ est l'indice des individus,
- * t_i est le temps de survie depuis le diagnostic pour le sujet i ,
- * a_i est l'âge au diagnostic du sujet i ,
- * δ_i est l'indicatrice de décès, *i.e.* $\delta_i = 1$ si t_i est le moment du décès et $\delta_i = 0$ si le temps de survie est censuré à t_i ,
- * et z_i est le vecteur de covariables pour le sujet i .

La fonction de vraisemblance qui est associée au modèle de l'équation (4.2) est donnée par

$$V = \sum_{i=1}^n \exp\left(\int_0^{t_i} \lambda_{\text{Obs}}(s; a_i, z_i) ds\right) \{\lambda_{\text{Obs}}(t_i; a_i, z_i)\}^{\delta_i}, \quad (4.5)$$

avec $\lambda_{\text{Obs}}(t_i; a_i, z_i)$ la fonction de risque de décès pour l'individu i au temps t_i diagnostiqué à a_i années et possédant les caractéristiques z_i . Pour estimer β et τ , les dérivées premières et secondes de la logvraisemblance correspondante sont calculées naturellement.

Il existe de multiples extensions et cas particuliers du modèle d'Estève *et al.*. Un des cas particuliers se place dans la situation des données agrégées. Il considère que l'âge au diagnostic ne fait pas partie du vecteur de covariables définissant le sujet i . En d'autres mots, z_i ne dépend pas de a_i . Ce cas particulier suppose que z_i comporte K catégories distinctes, ce qui implique que, pour chacune des strates, une table de mortalité doit être

disponible. Une des extensions du modèle d'Estève *et al.* consiste à estimer le modèle de survie relative à partir de données divisées en J intervalles de suivi qui peuvent être des années, des mois ou des jours (Dickman *et al.* (21)). Plus explicitement, le temps d'un individu est réparti sur les j intervalles de suivi définis au début de l'étude. Par ailleurs, l'hypothèse initiale de ce modèle est que le nombre de décès observé dans chaque intervalle suit un processus de Poisson. Une autre extension du modèle d'Estève *et al.* est l'approche d'Hakulinen et Tenkanen (33) que nous allons donc voir dans la prochaine partie.

4.3 MODÈLE D'HAKULINEN ET TENKANEN

La méthode d'estimation de la survie relative présentée précédemment repose sur la modélisation de la fonction de risque. Le modèle proposé par Hakulinen et Tenkanen (33) repose sur la modélisation de la survie relative. Cette méthode permet donc une modélisation plus directe de la survie.

Supposons que le nombre de décès δ_{jk} pour l'intervalle j de la strate k est une variable aléatoire de loi binomiale :

$$\delta_{jk} \sim \text{Bin}(n_{jk}, 1 - p_{\text{Obs},jk}),$$

où $p_{\text{Obs},jk}$ est la proportion de survie observée d'un intervalle spécifique, pour la strate k et l'intervalle j . Cette formulation montre que le modèle de survie relative peut s'écrire comme un modèle linéaire généralisé, caractérisé, pour une strate et un intervalle de temps spécifique, par une erreur de structure binomiale et une fonction de lien log-logarithmique non standard $\ln \left\{ -\ln \left(\frac{p_{\text{Obs},jk}}{p_{\text{Att},jk}} \right) \right\}$. Ainsi,

$$\ln \left\{ -\ln \left(\frac{p_{\text{Obs},jk}}{p_{\text{Att},jk}} \right) \right\} = \beta'z \quad (4.6)$$

où $p_{\text{Att},jk}$ est la proportion de survie attendue d'un intervalle spécifique, pour la strate k et l'intervalle j . Cette probabilité se définit par :

$$p_{\text{Att},jk} = \frac{1}{n_{jk}} \sum_{i=1}^{n_{jk}} p_{\text{Att},jk}(i) \quad (4.7)$$

où n_{jk} est le nombre d'individus en vie au début de l'intervalle j (*i.e.* à t_{j-1}) de la strate k et

$$p_{\text{Att},jk}(i) = \begin{cases} [1 - \lambda_{\text{Att},k}(a_{ji})]^{(t_j - t_{j-1})} & \text{si } y_{jik} > 0, \\ 0 & \text{sinon} \end{cases}$$

est la probabilité de survie attendue pour l'individu i de la strate k pour l'intervalle de suivi j avec :

- * $\lambda_{\text{Att},k}(a_{ji})$ le taux de mortalité annuel attendu pour l'individu i de la strate k à l'âge a_{jik} et
- * y_{jik} est le temps à risque total pour tous les individus à risque à l'âge j de la strate k .

Cette expression rejoint la définition du calcul de survie attendue de la méthode Ederder II (23).

La méthode d'Hakulinen et Tenkanen présentée ici implique que le taux de survie par strate et intervalle de temps est la moyenne des probabilités de survie attendue pour les individus vivants au début de l'intervalle. Ce mode de calcul s'avère être un handicap lorsque les intervalles de temps sont longs. Aussi, l'utilisation des strates entraînerait, selon Giorgi (28), un manque de puissance et des estimations moins précises en cas de petits effectifs ou en cas de survie à très long terme.

Maintenant que nous avons présenté les modèles de survie additifs les plus fréquemment utilisés, nous allons présenter les modèles de survie relative multiplicatifs. Nous avons choisi de présenter celui qui est à la base des différents travaux de ce mémoire.

4.4 MODÈLE D'ANDERSEN *et al.*

Andersen *et al.* (6) ont choisi une structure multiplicative de modèle.

La définition d'un modèle de survie relative multiplicatif est la suivante :

$$\lambda_{\text{Obs}}(t) = \lambda_{\text{Att}}(t) \cdot \lambda_{\text{Rel}}(t). \quad (4.8)$$

Un tel modèle ne suppose pas que le risque observé est toujours plus grand que le risque de la population cependant il a une interprétation moins évidente que le modèle additif.

Le modèle d'Andersen suppose $\lambda_{\text{Rel}}(t) = \lambda_{\text{Rel},0}(t) \exp(\beta Z)$ (modèle de Cox). Le risque observé devient alors :

$$\lambda_{\text{Obs}}(t) = \lambda_{\text{Att}}(t) \lambda_{\text{Rel},0}(t) \exp(\beta Z). \quad (4.9)$$

Ce modèle peut être réécrit dans la forme suivante :

$$\lambda_{\text{Obs}}(t) = \lambda_{\text{Rel},0}(t) \exp(\beta Z + \ln \lambda_{\text{Att}}(t)). \quad (4.10)$$

A partir de cette écriture, on met en évidence la structure d'un modèle de Cox avec une variable additionnelle qui dépend du temps. Sachant que les tables de mortalité de la population générale sont souvent structurées en intervalles annuels et afin d'ajuster le modèle en utilisant les procédures du modèle de Cox, les données doivent être divisées en intervalles annuels, avec les $\lambda_{\text{Att}}(t)$ mis à jour sur chacun des intervalles.

On s'est inspiré de ce modèle pour la suite de ce mémoire. La différence réside dans le choix d'une fonction paramétrique pour modéliser la fonction de risque de base relative à la maladie étudiée, *i.e.* $\lambda_{\text{Rel},0}(t)$. Le risque de décès observé pour un individu $j \in \{1, \dots, n\}$ (n est l'effectif total des individus étudiés), est :

$$\lambda_{\text{Obs}}(t_j) = \lambda_{\text{Att}}(t_j + a_{j0}) \lambda_{\text{Rel}}(t_j) \quad (4.11)$$

avec t_j le minimum entre le temps d'entrée dans l'étude et le décès ou la censure pour l'individu j et a_{j0} son âge à l'entrée dans l'étude. Pour estimer les paramètres de ce modèle, nous utilisons l'estimateur du maximum de vraisemblance. La vraisemblance de ce modèle s'écrit :

$$V = \prod_{j=1}^n \lambda_{\text{Obs}}(t_j)^{\delta_j} S_{\text{Obs}}(t_j) \quad (4.12)$$

avec δ_j l'indicatrice de décès. Pour obtenir la fonction de logvraisemblance finale, il suffit alors de définir les différentes fonctions qui la composent. Ainsi, pour la fonction de risque $\lambda_{\text{Rel},0}$, nous choisissons une distribution de Weibull de paramètres ν et σ (formule (2.7)). Nous considérons les $(\lambda_{\text{Att}}(t_j + a_{j0}))$ comme des quantités connues, *i.e.* provenant des tables vitales. La survie observée étant égale à :

$$S_{\text{Obs}}(t_j) = \exp(-\Lambda_{\text{Obs}}(t_j)), \quad (4.13)$$

le risque observé cumulé de décès s'écrit donc ainsi :

$$\begin{aligned} \Lambda_{\text{Obs}}(t_j) &= \int_0^{t_j} \lambda_{\text{Obs}}(u) du \\ &= \int_0^{t_j} \lambda_{\text{Att}}(u + a_{j0}) \lambda_{\text{Rel}}(u) du \\ &= \sum_{k=0}^{na_j} \lambda_{\text{Att}}(t_{j,k} + a_{j0}) \left(\int_{t_{j,k}}^{t_{j,(k+1)}} \lambda_{\text{Rel},0}(u) du \right) \exp(\beta Z_j) \\ &= \sum_{k=0}^{na_j} \lambda_{\text{Att}}(t_{j,k} + a_{j0}) \left(\frac{1}{\sigma^\nu} (t_{j,(k+1)}^\nu - t_{j,k}^\nu) \right) \exp(\beta Z_j). \end{aligned} \quad (4.14)$$

Ce risque s'écrit comme une somme d'aires de la fonction de risque car les taux de mortalité des tables vitales sont constants suivant les tranches d'âge, le sexe et les années. na_j représente ainsi le nombre d'années de survie jusqu'à t_j .

Nous maximisons donc la logvraisemblance suivante :

$$\log(V) = \sum_{j=1}^n \{ \delta_j \log(\lambda_{\text{Obs}}(t_j)) - \Lambda_{\text{Obs}}(t_j) \} \quad (4.15)$$

avec Λ_{Obs} calculé grâce à la formule (4.14). Ce modèle possède juste deux paramètres de plus à estimer par rapport au modèle d'Andersen. Nous verrons dans la partie appliquée que, le fait de poser une fonction paramétrique pour $\lambda_{\text{Rel},0}$, ne change pas les estimations des paramètres des autres variables.

4.5 TABLE DE MORTALITÉ

4.5.1 Définition et présentation

Une table de mortalité annuelle suit le cheminement d'une génération fictive de 100 000 nouveau-nés à qui l'on fait subir aux divers âges les conditions de mortalité observées sur les diverses générations réelles,

durant l'année étudiée. Pour éviter les aléas des tables annuelles et pour disposer d'une table détaillée par âge aussi précise que possible, on calcule également une table de mortalité couvrant une période de trois années.

La table (4.1) est un extrait de la table de mortalité pour les années de 2002 à 2004.

Age x	Homme			Femme		
	$S(x)$	$q(x)$	$E(x)$	$S(x)$	$q(x)$	$E(x)$
0	100000	471	75,70	100000	374	82,95
1	99529	37	75,06	99626	32	82,26
2	99492	27	74,09	99594	21	81,29
3	99465	22	73,11	99574	15	80,31
⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	14702	18522	3,83	32817	13994	4,53
91	11979	20362	3,59	28225	15804	4,19
92	9540	22240	3,38	23764	17790	3,88
⋮	⋮	⋮	⋮	⋮	⋮	⋮
104	138	46030	2,57	464	45465	1,69

TAB. 4.1 – Table de Mortalité (2002-2004).

$S(x)$ représente les survivants à l'âge x . $q(x)$ est le quotient de mortalité pour 100000 survivants à l'âge x . Ce quotient à un âge donné mesure "la probabilité", pour les personnes survivantes à cet âge, de décéder avant l'âge suivant ($x + 1$). Notons que, pour nos futures analyses, pour obtenir le taux de mortalité de la population générale, nous diviserons $q(x)$ par 100000 afin d'obtenir une "réelle" probabilité (*i.e.* un nombre compris entre 0 et 1). Enfin, $E(x)$ représente le nombre d'années restant à vivre au delà de l'âge x .

Les tables de mortalité sont fournies par l'INSEE (Institut National de la Statistique et des Études Économiques).

4.5.2 Analyse des tables vitales

La période de suivi observée pour la base DIVAT est de 1990 à 2006. Nous allons donc, pour l'explication des futures analyses, exploiter les tables vitales de cette période de suivi.

Nous disposons de tables vitales annuelles de 1990 à 1997 et de tables vitales calculées sur trois ans : 1998-2000, 2000-2002 et 2002-2004 que nous notons 1998, 2000 et 2002, respectivement. Bien que l'étude de suivi de la base DIVAT soit jusqu'en 2006, nous avons dû nous limiter à des tables vitales jusqu'en 2004 car l'INSEE n'a pas encore publié de tables au-delà de 2004. Ceci n'est pas problématique car, comme nous allons le voir, depuis 1990, les statistiques de mortalité générale de la population française n'évoluent pas beaucoup.

Afin de mieux percevoir les statistiques de mortalité des français, nous avons tracé la probabilité de survivre à chaque âge, par sexe et par année. Ceci correspond à la figure (4.1) page 67 qui comporte donc 11 graphiques et deux courbes sur chaque graphique. Nous pouvons donc constater que la probabilité de survivre d'un homme âgé de 80 ans en 1992 était de l'ordre de 0,4. Il lui restait, en moyenne, 7,13 années de survie. De plus, le graphique (4.1) montre que, sur les différentes années, les survies de la population nationale française sont quasiment identiques. La survie des femmes étant visiblement toujours supérieure à celle des hommes de l'âge d'environ 20 ans et ce jusqu'à 100 ans.

4.6 CHOIX D'UN MODÈLE DE SURVIE RELATIVE

Il est toujours délicat et récurrent de lancer le débat concernant le choix d'un modèle à structure multiplicative ou additive. En général, il s'établit selon les objectifs de l'étude et les données dont nous disposons.

Quand les facteurs de risque d'un évènement sont recherchés, ce qui est le cas dans les études étiologiques (cohorte, cas-témoin), les risques sont naturellement multiplicatifs. Dans ce genre d'étude, on cherche à évaluer le risque de survenue d'un évènement entre deux groupes. Ainsi, pour estimer ce ratio de risque, il apparaît plus facile d'utiliser un modèle à structure multiplicative. Par contre, en santé publique, où c'est plutôt la prévalence (nombre de nouveaux cas) qui est étudiée, les risques ont tendance à être additifs.

Un des principaux désavantages des modèles multiplicatifs de survie relative est qu'ils ne se traduisent pas explicitement en terme de survie comme les modèles additifs (*i.e.* pour les additifs : on passe facilement des formules de risque aux formules de survie et vice-versa).

Cependant, dans les modèles additifs classiques, Estève *et al.* (25) et Hakulinen et Tenkanen (33), du fait des paramètres τ_k à considérer, l'effectif total des paramètres à estimer peut augmenter rapidement si le nombre d'années de suivi est important et/ou si de nombreuses co-variables entrent en jeu. Une autre limitation de ces modèles additifs classiques est que la survie observée (brute) doit toujours être plus petite que la survie de la population générale (pour que la survie relative reste inférieure ou égale à 1). Ceci peut être excessif dans le cas d'une pathologie avec peu de décès. Ainsi, d'autres approches de modèles de survie relative à structure additive, utilisant des fonctions flexibles, ont été proposées : Bolard *et al.* (10), Giorgi *et al.* (29), Lambert *et al.* (41), (44), Remontet *et al.* (51). Ces différentes approches permettent de s'affranchir des contraintes liées au choix du nombre et de l'amplitude des intervalles ainsi que des problèmes engendrés dans les situations d'évènements rares. Ces auteurs utilisent notamment les fonctions splines ou dérivées (B-splines, splines cubiques restreintes) pour modéliser, entre autres, la fonction de risque relatif de base. Néanmoins, même si ces fonctions offrent une grande flexibilité pour ajuster les données sans nécessité de

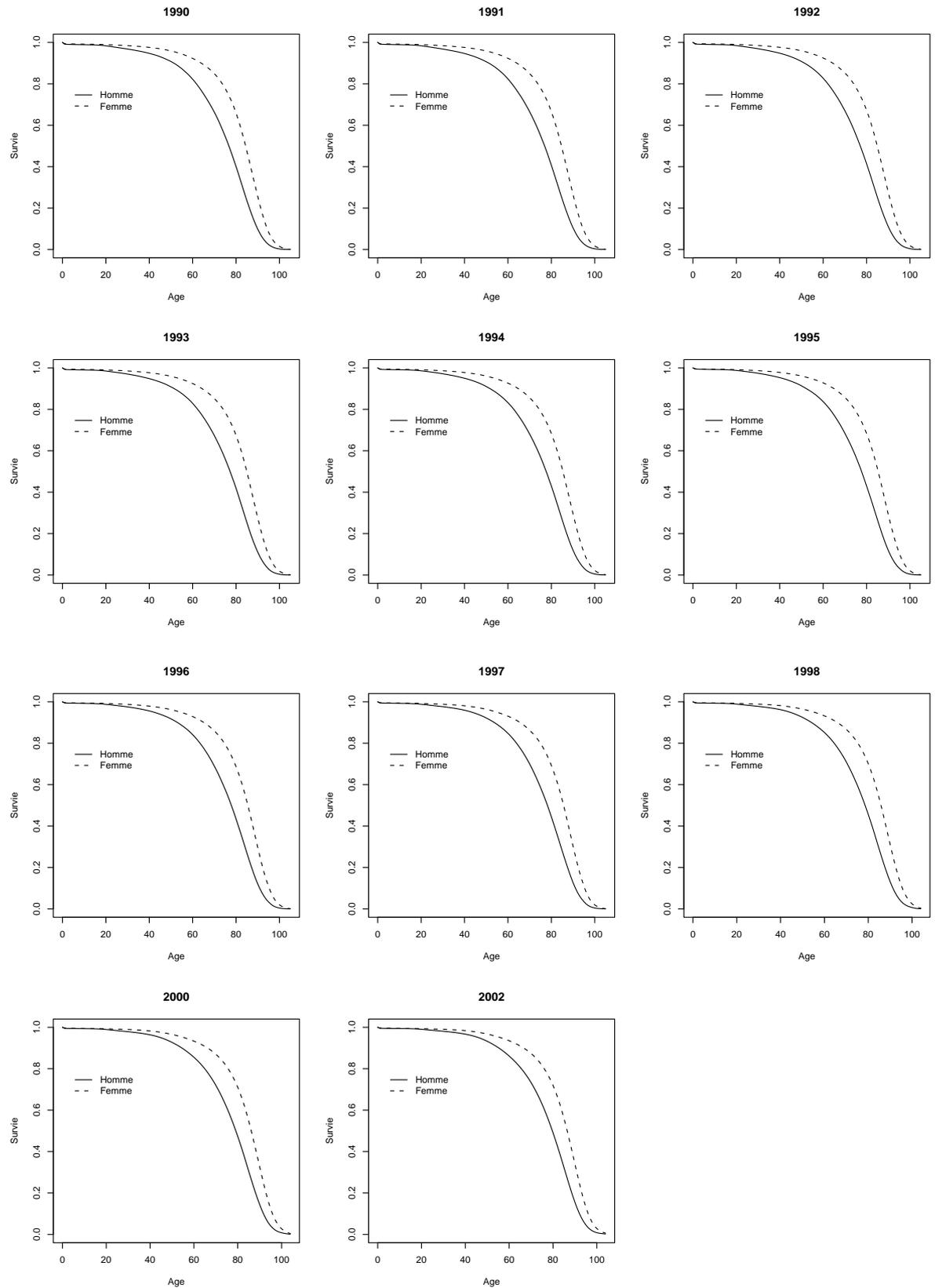


FIG. 4.1 – Graphiques de la survie attendue par âge, sexe et année.

faire des hypothèses sur la forme de l'évolution du ratio des risques au cours du temps, elles nécessitent de faire des choix sur les noeuds, *i.e.* leur nombre, leur localisation ainsi que leur degré. En général, ces critères sont fixés *a priori*. Dans ce cas, ces choix sont argumentés cliniquement ou par le fait que le nombre de noeuds choisi est suffisant pour couvrir une large collection de courbes. Cependant, compte tenu de l'arbitraire du choix du nombre de noeuds et de leur localisation, ces deux paramètres peuvent être considérés comme des variables à estimer. Cela est possible mais augmente rapidement le nombre de paramètres à estimer. En règle générale, tout cela montre la difficulté qu'il y a à définir des seuils et quelle est la signification exacte d'un seuil. Par exemple, à un noeud, y-a-t-il statistiquement un changement de risque ? (ceci dépendant fortement du modèle choisi et du degré de la spline).

Dans la cohorte DIVAT, les décès après une transplantation de rein (*i.e.* évènement à étudier) sont très peu nombreux (surtout après l'année 1998). Nous ne pouvons donc utiliser les modèles de survie additifs classiques. En effet, la survie observée des patients, estimée par un modèle de Weibull par exemple, est inférieure à la survie attendue issues des tables vitales. Restreindre l'étude de la cohorte aux premières années de transplantation serait aberrant compte tenu du nombre d'années de suivi que la base DIVAT possède et du fait que les stratégies de transplantation ont changé à partir de 1996 – 1998. Dans la suite de ce mémoire, comme une large littérature a d'ores et déjà exploré les modèles additifs et leur contraintes, nous nous focalisons donc sur les modèles de survie relative à structure multiplicative.

4.7 APPLICATION AUX TRANSPLANTÉS RÉNAUX

Dans ce paragraphe, nous appliquons les deux types de modèles multiplicatifs de survie relative à la base de données DIVAT. Le premier modèle est celui d'Andersen et le second est celui dont la fonction de risque relative de base $\lambda_{\text{Rel},0}$ est modélisée par une distribution paramétrique. L'application de ces deux modèles à la base DIVAT a des fins de comparaison et de prise de décision concernant un des deux modèles pour la suite du mémoire.

4.7.1 Résultats avec le modèle d'Andersen

La table (4.2) récapitule les estimations des paramètres du modèle d'Andersen. Ce modèle est implémenté grâce au package *RelSurv* (créé par Pohar *et al.* (49)).

Sachant la prise en compte du risque de mortalité toutes maladies confondues, le sexe du receveur, l'âge du receveur ainsi que l'année de la greffe sont des facteurs ayant un effet sur la survie des patients transplantés rénaux. Cependant, les paramètres des covariables sexe et de l'âge du receveur ont des signes opposés par rapport à ceux du modèle de Cox estimés au chapitre 2 (voir table (2.4)). Or, l'interprétation des effets de ces deux covariables dans ce modèle de survie relative n'est pas la

même que pour un modèle Cox. Ce n'est donc pas à interpréter comme une inversion des effets de ces covariables. En effet, l'âge et le sexe du receveur et l'année de la greffe interviennent aussi dans le risque attendu issu des tables de mortalité de la population générale. Ces covariables sont donc prises en compte deux fois dans le modèle : une fois dans le risque relatif au groupe de patients greffés et une fois dans le risque attendu du groupe de référence. Nous nous sommes posés la question de savoir s'il était nécessaire de prendre en compte l'âge, le sexe du receveur et l'année de la greffe dans le risque relatif. Cependant, même si elles interviennent déjà dans le risque attendu, il est tout à fait plausible qu'elle intervienne dans le risque relatif à la transplantation rénale. Ces trois covariables peuvent être des covariables d'ajustement. Nous ne les analyserons donc pas comme des covariables classiques.

Par ailleurs, la table (4.2) montre que la covariable DGF est significative. Cette covariable a la même signification que ce soit pour le risque relatif à la transplantation que pour le risque observé des transplantés. En effet, pour le risque relatif à la transplantation, le ratio des risques entre les deux groupes est :

$$\begin{aligned} \frac{\lambda_{\text{Rel}}(t, Z_{\text{DGF}} = 1)}{\lambda_{\text{Rel}}(t, Z_{\text{DGF}} = 0)} &= \frac{\lambda_{\text{Rel},0}(t) \exp(\beta_{\text{DGF}})}{\lambda_{\text{Rel},0}(t)} \\ &= \exp(\beta_{\text{DGF}}). \end{aligned}$$

C'est l'écriture classique d'un ratio de risque dans le cas d'un modèle de Cox. De plus, le ratio des risques concernant le risque observé des transplantés est :

$$\begin{aligned} \frac{\lambda_{\text{Obs}}(t, Z_{\text{DGF}} = 1)}{\lambda_{\text{Obs}}(t, Z_{\text{DGF}} = 0)} &= \frac{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \lambda_{\text{Rel}}(t, Z_{\text{DGF}} = 1)}{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \lambda_{\text{Rel}}(t, Z_{\text{DGF}} = 0)} \\ &= \frac{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \lambda_{\text{Rel},0}(t) \exp(\beta_{\text{DGF}})}{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \lambda_{\text{Rel},0}(t)} \\ &= \frac{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \exp(\beta_{\text{DGF}})}{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année de la greffe} = y)} \\ &= \exp(\beta_{\text{DGF}}). \end{aligned}$$

Ces rapports de risques montrent que l'interprétation de la covariable DGF est la même dans le risque observé des transplantés et dans le risque relatif aux transplantés. Ceci signifie que les patients dont le greffon met plus de 6 jours à reprendre fonction ont plus de risque de décéder que les patients dont le greffon met moins de 6 jours. Cependant, nous remarquons que ce risque est d'environ 1,387 alors que pour le modèle de Cox, vu au chapitre 2, il est de 1,454. Cette remarque implique qu'après correction sur les facteurs âge, sexe du receveur et année de la greffe, le risque des patients dont le greffon met plus de six jours à reprendre fonction par rapport aux patients dont le greffon met moins de six jours est diminué par rapport au risque estimé dans un modèle de Cox classique.

Covariables	coef	exp(coef)	se	p-value
Sexe du receveur	0,529	1,697	0,148	3,5e - 04
Age du receveur	-0,402	0,669	0,138	3,6e - 03
DGF	0,327	1,387	0,137	1,7e - 02
Année de la greffe	-0,621	0,538	0,151	3,8e - 05

TAB. 4.2 – Estimation des paramètres du modèle d'Andersen ($\log(\mathcal{L}) = -1497,027$).

Comme l'interprétation des covariables âge, sexe du receveur et année de la greffe est équivoque en ce qui concerne le risque relatif aux transplantés, nous nous intéressons maintenant à leurs interprétations dans l'analyse du risque observé. Le risque observé est donc le risque "non corrigé" par les facteurs âge, sexe et année intervenant dans le risque attendu, *i.e.* risque de toutes les pathologies confondues dans la population générale.

Cependant, comme les patients sont âgés de 18 à 80 ans et qu'ils sont greffés entre les années 1990 et 2006, l'interprétation des résultats est riche. Ainsi, nous n'interpréterons pas le ratio des risques observés pour les covariables âge du receveur et année de la greffe. Par contre, pour se rendre compte des interprétations possibles du risque observé, le ratio des risques observés pour la covariable sexe du receveur sera analysé.

Le ratio du risque observé des patients de plus de 55 ans sur le risque observé des patients de moins de 55 ans (même sexe et même année de greffe) s'écrit :

$$\begin{aligned}
 \frac{\lambda_{\text{Obs}}(t, Z_{\text{Age}} = 1)}{\lambda_{\text{Obs}}(t, Z_{\text{Age}} = 0)} &= \frac{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} \geq 55, \text{Année de la greffe} = y) \lambda_{\text{Rel}}(t, Z_{\text{Age}} = 1)}{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} < 55, \text{Année de la greffe} = y) \lambda_{\text{Rel}}(t, Z_{\text{Age}} = 0)} \\
 &= \frac{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} \geq 55, \text{Année de la greffe} = y) \lambda_{\text{Rel},0}(t) \exp(\beta_{\text{Age}})}{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} < 55, \text{Année de la greffe} = y) \lambda_{\text{Rel},0}(t)} \\
 &= \frac{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} \geq 55, \text{Année de la greffe} = y) \exp(\beta_{\text{Age}})}{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} < 55, \text{Année de la greffe} = y)}.
 \end{aligned}$$

Le ratio du risque observé des patients greffés après 1996 sur le risque observé de ceux greffés avant 1996 ayant les mêmes caractéristiques que les précédents (âge et sexe) est :

$$\frac{\lambda_{\text{Obs}}(t, Z_{\text{Année greffe}} = 1)}{\lambda_{\text{Obs}}(t, Z_{\text{Année greffe}} = 0)} = \frac{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année greffe} \geq 1996) \exp(\beta_{\text{Année greffe}})}{\lambda_{\text{Att}}(\text{Sexe} = s, \text{Age à la greffe} = a, \text{Année greffe} < 1996)}.$$

Enfin, le ratio du risque observé des greffés femme par rapport au risque observé des greffés homme aux mêmes âges et années de transplantation s'écrit :

$$\frac{\lambda_{\text{Obs}}(t, Z_{\text{Sexe}} = 1)}{\lambda_{\text{Obs}}(t, Z_{\text{Sexe}} = 0)} = \frac{\lambda_{\text{Att}}(\text{Sexe} = 1, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \exp(\beta_{\text{Sexe}})}{\lambda_{\text{Att}}(\text{Sexe} = 0, \text{Age à la greffe} = a, \text{Année de la greffe} = y)}.$$

La figure (4.2) permet de visualiser ce ratio par année et par âge. Nous remarquons que ce ratio varie d'environ 0,4 à 1,4 sur l'ensemble des années de greffe et à différents âges. D'après ces graphiques, les hommes ont un risque observé plus important de décéder après une greffe de rein que les femmes, ceci quelle que soit l'année de transplantation.

Ainsi, cette partie application explore les interprétations concernant le risque relatif aux transplantés rénaux mais aussi celles concernant le risque observé chez ces transplantés. La différence entre ces deux risques réside dans le fait que l'un est spécifique à la transplantation rénale alors que l'autre prend en compte toutes les pathologies auxquelles les greffés peuvent aussi être exposés.

Pour le risque relatif à la transplantation, seule la covariable DGF est interprétable (les autres covariables étant considérées comme des covariables d'ajustement). Par contre, pour le risque observé des transplantés rénaux, seule la covariable sexe du receveur est interprétée. En effet, les interprétations des résultats peuvent s'avérer riche et le choix d'interpréter certains résultats concernant des âges ou des années spécifiques est laissé aux médecins.

Nous allons maintenant analyser le modèle de survie relative multiplicatif avec la fonction de risque de base relatif à la greffe paramétrique. Les résultats de ce modèle ne seront pas aussi détaillés que pour le modèle d'Andersen. En effet, nous ne présenterons que les résultats spécifiques au risque relatif à la transplantation.

4.7.2 Résultats avec le modèle paramétrique

Dans cette partie, nous appliquons le modèle dont la fonction de risque de base relatif à la greffe, *i.e.* $\lambda_{\text{Rel},0}$, est modélisée par une distribution de Weibull $\mathcal{W}(\sigma, \nu)$. Ici, nous nous sommes chargés nous même de l'implémentation informatique de ce modèle. La table (4.3) donne les résultats des estimations des paramètres de ce modèle.

Tout d'abord, avant de tester les covariables, nous faisons le test $\{H_0 : \nu = 1\}$ pour savoir si une distribution Exponentielle $\mathcal{E}(\sigma)$ (formule (2.6)) n'est pas suffisante pour $\lambda_{\text{Rel},0}$. Grâce au test du rapport de vraisemblance qui vaut 19,87, correspondant à une p -valeur inférieure à 0,0001, nous n'acceptons pas l'hypothèse que le risque de base relative suive une loi Exponentielle. Nous choisissons donc le modèle avec la fonction de risque de base relative modélisée par une distribution de Weibull.

Ensuite, nous testons les covariables et remarquons que les covariables significatives sont l'âge et le sexe du receveur, le DGF et l'année de la greffe. Les estimations des différents paramètres sont assez proches du modèle d'Andersen. Concernant la covariable DGF, les personnes dont le greffon met plus de six jours à reprendre fonction ont 1,372 fois plus de risque de décéder que les personnes dont le greffon met moins de six jours. Ce risque est plus faible que celui trouvé avec le modèle d'Andersen.

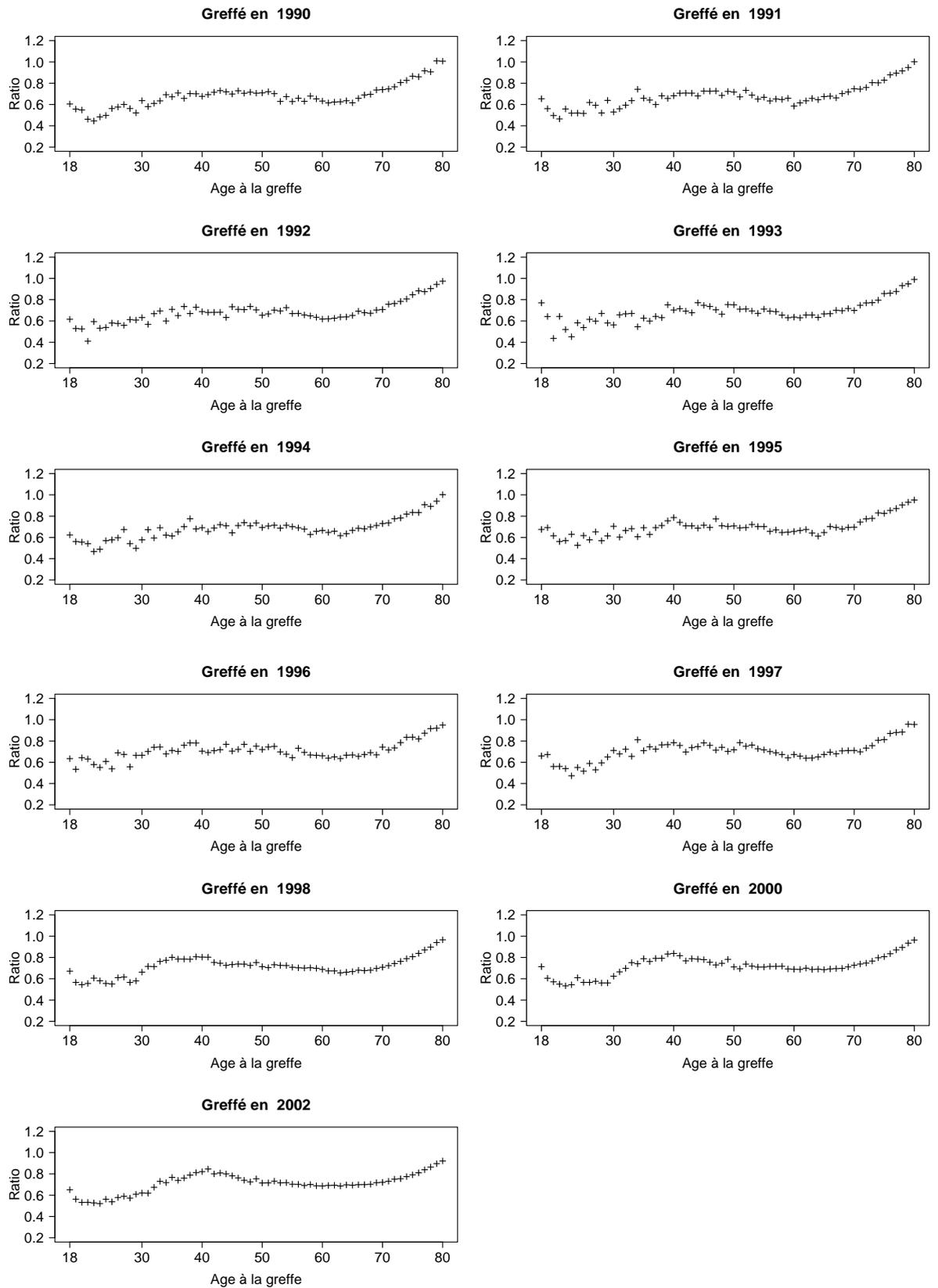


FIG. 4.2 – Modèle d'Andersen : Graphique du ratio du risque observé des femmes sur le risque observé des hommes ayant les mêmes caractéristiques (âge, année de greffe).

sen. Cependant, il est aussi adéquat de comparer ce modèle de survie relative paramétrique de Weibull avec le modèle paramétrique de Weibull sans survie relative du chapitre 2. Ceci pour montrer l'apport de la survie relative dans un modèle de régression multiplicatif. Les covariables apparaissant significatives sont les mêmes pour les deux modèles. De plus, le risque relatif aux transplantés rénaux modélisé par le modèle avec survie relative est plus petit que le risque relatif estimé avec le modèle sans survie relative pour lequel il est de 1,45 (table (2.2)).

Paramètres	coef	exp(coef)	se	p-value
σ	0,265	.	0,113	.
ν	0,699	.	0,046	< 0,05*
Sexe du receveur	0,496	1,642	0,148	0,001
Age du receveur	-0,368	0,692	0,138	0,008
DGF	0,316	1,372	0,137	0,021
Année de la greffe	-0,636	0,529	0,146	0,000

* Test du rapport de vraisemblance : $\{H_0 : \nu = 1\}$.

TAB. 4.3 – Estimation des paramètres du modèle multiplicatif de survie relative avec distribution de Weibull pour la fonction de risque de base relative ($\log(\mathcal{L}) = -1106,217$).

Le graphique (4.3) montre les risques relatifs à la greffe selon la covariable DGF en fonction du temps après la greffe. Pour les patients dont le greffon met plus de six jours à reprendre fonction, le risque de décès, environ deux jours après la greffe (0,0055 années), atteint presque 12 alors que pour les patients dont le greffon met moins de six jours à reprendre fonction ce risque est de l'ordre de 8. Par ailleurs, le modèle de régression paramétrique de Weibull montre que les risques de décès, environ deux jours après la greffe, sont de l'ordre de 0,05 et 0,03 pour les patients dont le greffon met plus de six jours à reprendre fonction et ceux dont le greffon met moins de six jours, respectivement. Nous remarquons donc, que le modèle de Weibull avec survie relative diminue le ratio de risque pour la covariable DGF par rapport au modèle paramétrique de Weibull sans survie relative. Cependant, ce modèle montre que le risque de décès relatif à la transplantation est plus élevé les premiers instants après la greffe que le risque de décès non corrigé par les facteurs âge, sexe du receveur et année de la greffe. En bref, le modèle avec survie relative diminue le ratio des risques pour la covariable DGF mais il augmente le risque de décès relatif à la maladie lors des premiers instants après la transplantation. Cependant, pour le modèle de survie relative, nous remarquons que, quel que soit la modalité de la covariable DGF, les risques de décès relatif à la greffe se stabilisent autour de la valeur 2 environ cinq ans après la greffe. Par contre, pour le modèle paramétrique sans survie relative, ces mêmes risques se stabilisent aussi cinq ans après la greffe mais autour de la valeur 0,02. Même au delà de cinq ans après la transplantation, le risque de décès relatif à la greffe est supérieur au risque de décès observé avec un modèle paramétrique classique.

Au vu des résultats quasiment identiques entre les deux modèles multiplicatifs, celui d'Andersen et celui paramétrique, nous choisissons pour la suite de ce mémoire de considérer le modèle multiplicatif dont le risque

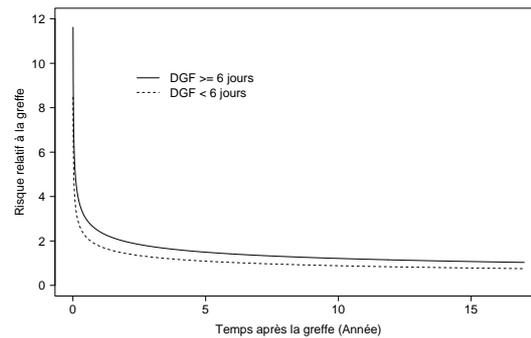


FIG. 4.3 – Graphique de la fonction de risque relatif selon la covariable DGF.

relatif à la pathologie est considéré paramétrique. Ce choix est actuellement possible car les paramètres sont devenus estimables grâce à diverses méthodes statistiques.

DISCUSSION-CONCLUSION

Dans ce chapitre, nous avons défini les différents modèles de survie relative existants. Nous avons succinctement exploré les modèles de Estève *et al.* (25) et de Hakulinen et Tenkanen (33) pour les modèles additifs et le modèle de Andersen *et al.* (6) pour les modèles multiplicatifs. Ce sont les modèles les plus connus en survie relative. Nous avons pu remarquer les différences entre les modèles additifs et les multiplicatifs. D'une part, la formulation de la survie relative dans le cas des modèles additifs est plus naturelle que dans le cas des modèles multiplicatifs. C'est une raison pour laquelle ces modèles sont souvent préférés aux modèles multiplicatifs. D'autre part, l'interprétation d'un modèle additif est plus directe. Cependant, ces modèles de survie relative additifs classiques ont quelques limitations. Vu qu'une fonction de risque relatif de base est modélisée pour chaque intervalle-temps, le nombre de paramètres peut rapidement augmenter quand le nombre d'années de suivi d'une base est importante et/ou que l'on souhaite tester plusieurs covariables. Par ailleurs, une contrainte est que la survie observée du groupe de patients étudiés soit toujours inférieure à la survie attendue du groupe de référence. C'est une des limitations des modèles additifs classiques qui nous a amenés à ne pas les utiliser pour notre application. Néanmoins, pour résoudre ces problématiques, d'autres modèles de survie relative additifs ont été proposés dans la littérature. Ils sont énumérés dans la partie 4.6 de ce chapitre. Ces différents modèles ne sont pas détaillés du fait que nous avons choisi de nous concentrer sur les modèles de survie relative multiplicatifs. Et ceci, dans tout le mémoire. De plus, nous avons proposé d'utiliser un modèle multiplicatif de survie relative dont la fonction de risque de base relatif à la pathologie étudiée est paramétrique. Ce qui n'est pas le cas dans le modèle d'Andersen. Cependant, en utilisant ces deux modèles de survie relative multiplicatifs pour la base de données DIVAT, nous nous sommes rendus compte que les résultats sont quasiment similaires. Ainsi, en vue d'alléger les calculs théoriques de l'inférence statistique et

de faciliter l'implémentation informatique, nous choisissons de prendre le modèle multiplicatif de survie relative paramétrique en dépit du modèle d'Andersen.

Dans la suite de ce mémoire, nous introduisons le concept de survie relative à tous les modèles proposés. Nous verrons que le fait d'introduire la survie relative n'engendre pas de complications particulières sauf peut-être le fait d'alourdir certains calculs théoriques.

SURVIE RELATIVE MULTIPLICATIVE AVEC FRAGILITÉ

SOMMAIRE

5.1	LE MODÈLE	79
5.2	APPLICATION AUX TRANSPLANTÉS RÉNAUX	80
5.2.1	Fragilité sur les centres de transplantation	81
5.2.2	Fragilité sur les années de transplantation	83
	DISCUSSION-CONCLUSION	86

La plupart des modèles de survie relative supposent que la population étudiée est homogène (c'est-à-dire que l'unité statistique indépendante est l'individu). Or, dans certaines situations, cette hypothèse n'est pas réaliste. En effet, l'échantillon étudié peut être composé de sous-groupes d'individus ayant des caractéristiques communes qui ne sont pas forcément mesurables (génétiques, environnementales, etc). Ces groupes peuvent par exemple être des familles, des jumeaux, ou dans notre cas des individus transplantés dans les mêmes centres hospitaliers. Ainsi, dans ces situations, comme déjà expliqué au chapitre 3, la fragilité intervient.

Dans ce chapitre, nous proposons d'introduire dans un modèle de survie relative de type multiplicatif un terme de fragilité permettant de prendre en compte ce type d'hétérogénéité non observée. Cette fragilité est un effet aléatoire qui décrit la dépendance entre les individus d'un même groupe et permet donc de prendre en compte les facteurs de risque non observés communs à plusieurs individus. Le modèle de survie relative avec fragilité est approprié aux problématiques rencontrées dans cette base. Tout d'abord, comme vu dans le chapitre 2, les patients issus d'un même centre de transplantation sont suspectés dépendants. Il en est de même pour ceux greffés la même année. Par ailleurs, dans cette base de données, les informations concernant les causes des décès des patients transplantés avec leur rein fonctionnel ne sont pas renseignées. Ainsi, nous recourons à la fragilité et à la survie relative pour résoudre ces deux problématiques.

Song et al. (55) ont déjà proposé un modèle relatif multiplicatif pour données de survie corrélées. Ils considèrent un modèle de fragilité partagée, identique à celui que nous avons choisi au chapitre 3. Dans leur modèle, ils introduisent les notions de survie relative. Cependant, le vecteur des covariables est appelé le risque relatif à la pathologie et le risque de base est le risque connu de la population de référence. Notre modèle diffère du précédent car le risque relatif à la pathologie est le produit du risque relatif de base et du vecteur des covariables. De plus, dans notre modèle, le risque de la population de référence est le risque connu des tables vitales fournies par l'INSEE. Ainsi, notre modèle a l'avantage de bien distinguer le risque relatif à la pathologie et le risque de la population de référence. En effet, le vecteur des covariables inclut les caractéristiques propres aux transplantés rénaux et le risque de la population de référence provient de données réelles de la population nationale française.

Dans une première partie de ce chapitre nous présentons le modèle de survie relative avec fragilité. Comme ce modèle mélange deux notions, celle de la survie relative vue au chapitre 4 et celle de la fragilité vue au chapitre 3, ce chapitre n'introduit pas vraiment de nouvelles formules. Par contre, il fait souvent référence aux chapitres 3 et 4. Ensuite, nous appliquons le modèle proposé aux données de la base DIVAT. Comme expliqué précédemment, ce modèle est bien approprié aux problématiques de cette base. Par ailleurs, nous faisons souvent référence au modèle de fragilité ainsi qu'au modèle de survie relative. Ceci pour apprécier les apports du modèle proposé, *i.e.*, si la fragilité dans un modèle de survie relative apporte de nouveaux résultats par rapport à un modèle de survie relative sans fragilité, et, si la survie relative dans un modèle de fragilité apporte de nouveaux résultats par rapport à un modèle de fragilité sans survie relative.

5.1 LE MODÈLE

Soit un échantillon constitué de n groupes indicés par i ($i = 1, \dots, n$). Chacun de ces groupes contient n_i individus repérés par l'indice j ($j = 1, \dots, n_i$). Nous introduisons une variable de fragilité y_i permettant de modéliser la dépendance entre les individus d'un même groupe.

Par analogie avec l'équation (3.7), soit

$$\lambda_{\text{Obs}}(t_{ij}|z_{ij}, y_i) = y_i \lambda_{\text{Obs}}(t_{ij}|z_{ij}) \quad (5.1)$$

la fonction de risque observée à l'instant t_{ij} pour l'individu j du groupe i . Le vecteur des covariables de cet individu est $z_{ij} = (z_{ij1}, \dots, z_{ijp})'$ et la fragilité propre au groupe i est notée y_i . t_{ij} représente le délai minimum entre l'inclusion dans l'étude et le décès ou la censure.

Comme nous introduisons la notion fragilité dans les modèles de survie relative, nous retrouvons la théorie du chapitre 3 concernant la fragilité avec les notions propres à la survie relative du chapitre 4.

Sous l'hypothèse d'indépendance des individus, nous choisissons le modèle multiplicatif de survie relative de la même forme que l'équation (4.11). Ici, le risque observé de l'individu j appartenant au groupe i s'écrit :

$$\lambda_{\text{Obs}}(t_{ij}|z_{ij}) = \lambda_{\text{Att}}(t_{ij} + a_{ij0}) \underbrace{\lambda_{\text{Rel},0}(t_{ij}) \exp(\beta z_{ij})}_{\lambda_{\text{Rel}}(t_{ij}|z_{ij})} \quad (5.2)$$

où a_{ij0} représente l'âge à l'inclusion dans l'étude de l'individu j du groupe i . $\lambda_{\text{Att}}(t_{ij} + a_{ij0})$ est la fonction de risque de la population générale de même sexe, même âge $t_{ij} + a_{ij0}$ et né la même année que l'individu j du groupe i . Ici encore, nous considérons les $\lambda_{\text{Att}}(t_{ij} + a_{ij0})$ comme des quantités connues. $\beta = (\beta_1, \dots, \beta_p)'$ est un vecteur de coefficients de régression inconnus associés à z_{ij} . $\lambda_{\text{Rel},0}(t_{ij})$ est la fonction de risque relatif de base spécifique à la pathologie étudiée.

Dans le cas présent, pour définir un modèle multiplicatif de survie relative avec fragilité, l'équation (5.2) devient :

$$\lambda_{\text{Obs}}(t_{ij}|z_{ij}, y_j) = y_i \lambda_{\text{Att}}(t_{ij} + a_{ij0}) \lambda_{\text{Rel},0}(t_{ij}) \exp(\beta z_{ij}) \quad (5.3)$$

avec la contrainte que l'espérance de la variable aléatoire (y_i) soit égale à 1. Notons que le terme aléatoire est uniquement lié à la fonction de risque, permettant ainsi de prendre en compte la fragilité de certains groupes. L'effet des covariables (effet fixe) est considéré homogène quels que soient ces groupes. Autrement dit, l'appartenance aux différents groupes est supposée indépendante de l'effet des covariables.

Nous ne redémontrons pas les formules des différentes fonctions de survie vues au chapitre 3. Elles sont identiques aux notations près, *i.e.* $S(\cdot)$ devient $S_{\text{Obs}}(\cdot)$. Ainsi, tous les calculs relatifs à la survie restent les mêmes.

Comme vu au chapitre 3 concernant la fragilité, la logvraisemblance s'écrit :

$$\log(\mathcal{L}) = \sum_{i=1}^n \left\{ \sum_{j=1}^{n_i} D_{ij} \log(\lambda_{\text{Obs}}(t_{ij})) + \log \left((-1)^{D_i} L^{(D_i)} \left(\sum_{j=1}^{n_i} \Lambda_{\text{Obs}}(t_{ij}) \right) \right) \right\}. \quad (5.4)$$

Pour obtenir la fonction de logvraisemblance finale, nous reprenons les mêmes distributions que dans la section 4.4 du chapitre 4 concernant la survie relative. Pour la fonction de risque relative de base, il s'agit d'une distribution de Weibull :

$$\lambda_{\text{Rel},0}(t) = \nu \left(\frac{1}{\sigma} \right)^\nu t^{(\nu-1)}, \quad t \geq 0, \quad \sigma \text{ et } \nu > 0.$$

Nous obtenons donc le risque cumulé observé suivant :

$$\Lambda_{\text{Obs}}(t_{ij}) = \sum_{k=0}^{na_{ij}} \lambda_{\text{Att}}(t_{ijk} + a_{ij0}) \left(\frac{1}{\sigma^\nu} (t_{ij(k+1)}^\nu - t_{ijk}^\nu) \right) \exp(\beta z_{ij})$$

avec na_{ij} le nombre d'années de survie jusqu'à t_{ij} .

Nous supposons les y_i indépendants et identiquement distribués selon une loi de Gamma, $G(\gamma, \gamma)$, la densité est donnée par l'équation (3.18) page 47.

Ainsi la logvraisemblance s'écrit :

$$\log(\mathcal{L}) = \sum_{i=1}^n \left\{ \sum_{j=1}^{n_i} D_{ij} \log(\lambda_{\text{Obs}}(t_{ij})) + \gamma \log(\gamma) + \log \Gamma(\gamma + D_i) - \log \Gamma(\gamma) - (\gamma + D_i) \log \left(\gamma + \sum_{j=1}^{n_i} \Lambda_{\text{Obs}}(t_{ij}) \right) \right\}. \quad (5.5)$$

Ici, nous avons trois paramètres à estimer plus ceux relatifs aux covariables. Il y en a le même nombre que pour le modèle de fragilité (chapitre 3) puisque la fonction de risque attendu de la population de référence est considérée comme connue grâce aux tables vitales fournies par l'INSEE.

Nous allons maintenant appliquer ce modèle à la base de données DIVAT. Dans un premier temps, la fragilité modélise la dépendance des individus greffés dans les mêmes centres de transplantation. Ensuite, elle modélisera la dépendance des patients transplantés les mêmes années.

5.2 APPLICATION AUX TRANSPLANTÉS RÉNAUX

Dans cette partie, nous appliquons le modèle proposé précédemment aux données DIVAT. Comme nous suspectons deux sources de dépendance parmi les patients transplantés, le centre et l'année de transplantation, nous effectuons deux modèles : un dont la fragilité modélise la corrélation intra-centre et un autre dont la fragilité modélise la dépendance des patients transplantés les mêmes années.

5.2.1 Fragilité sur les centres de transplantation

Comme dans cette partie le modèle est celui dont la fragilité porte sur les centres de transplantation, nous ne testons pas les centres comme facteur potentiellement explicatif du décès. La table (5.1) montre les covariables ressortant significatives suivant les différentes stratégies de sélection. Sachant le lien entre les patients greffés dans le même centre et sachant le risque de décès de toutes les pathologies confondues causé par l'âge, le sexe et l'année, la covariable spécifique du décès après une transplantation rénale est le DGF.

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA	×		
Traitement immunosuppresseur	×		
Type de maladie initiale			
Antécédents cardiologiques			
Relation donneur / receveur			
Ischémie			
Néoplasie			
Age du donneur			
Sexe du donneur			
Année de la greffe	×	×	×

TAB. 5.1 – *Modèle Multiplicatif de Survie Relative avec Fragilité (MSRF) sur les centres de transplantation : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).*

Paramètres	coef	exp(coef)	se	p-value
σ	0,349	.	0,251	.
ν	0,700	.	0,046	< 0,05*
γ	1,562	.	0,294	> 0,05•
Sexe du receveur	0,491	1,634	0,148	0,001
Age du receveur	-0,390	0,677	0,138	0,005
DGF	0,295	1,343	0,140	0,035
Année de la greffe	-0,571	0,565	0,150	0,000

* Test du rapport de vraisemblance : $\{H_0 : \nu = 1\}$.

• Test du rapport de vraisemblance entre le modèle avec fragilité et le même modèle sans fragilité.

TAB. 5.2 – *Analyse du modèle Multiplicatif de Survie Relative avec Fragilité (MSRF) sur les centres de transplantation ($\log(V) = -1166,305$).*

Le tableau (5.2) montre que les résultats concernant les effets fixes du modèle sont à peu près identiques à ceux du modèle multiplicatif de survie relative sans fragilité (de l'ordre de 0,001). Par contre, les résultats de ce modèle diffèrent du modèle de fragilité sans survie relative. Les estimations des effets des covariables âge et sexe du receveur sont de signes opposés. Comme nous l'avons mentionné au chapitre 4, l'apport de la notion de survie relative corrige l'estimation de la survie observée du groupe de patients étudié. Ce retournement de situation, *i.e.* inversion des signes des paramètres entre les deux modèles, n'implique pas un

retournement des effets des covariables. Mais, les variables âge, sexe et année calendaire sont plutôt considérées comme des variables d'ajustement et non plus comme des variables explicatives du décès au sens classique de la survie. Nos résultats du modèle étudié dans le chapitre prédisent donc que : sachant les effets de l'âge, du sexe et de l'année de la greffe sur le décès toutes causes confondues et sachant la dépendance des patients greffés dans le même centre, avoir un greffon qui met moins de 6 jours à reprendre fonction est un facteur protecteur du décès après une transplantation rénale.

Nous testons la significativité de la fragilité par un test du rapport de vraisemblance entre le modèle multiplicatif de survie relative avec fragilité et celui sans fragilité. La statistique de ce test vaut $-120,175$ correspondant à une p -valeur supérieur à $0,05$. Ce test indique donc que la fragilité n'est pas significative. Il n'y a donc pas de lien entre les patients issus de même centre de transplantation. Nous avons déjà obtenu ce résultat avec le modèle de fragilité de Cox et celui de Weibull présentés au chapitre 3.

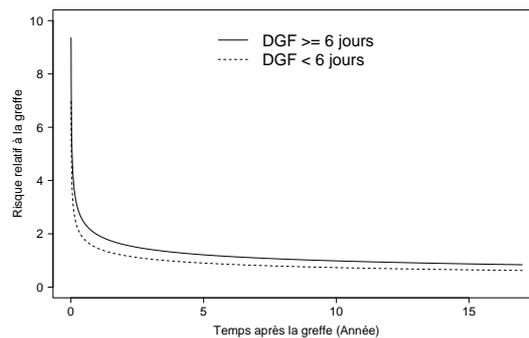


FIG. 5.1 – Fonction de risque selon les modalités des covariables sous le modèle multiplicatif de survie relative de Weibull avec fragilité sur les centres de transplantation.

La figure (5.1) représente le risque de décès relatif aux greffés en fonction du temps et selon les modalités de la covariable DGF. Ce risque est de l'ordre de 9,5 juste après la transplantation pour les patients ayant leur greffon qui met plus de six jours à reprendre fonction alors que pour ceux dont le greffon met moins de six jours, ce risque vaut environ 7. Pour les deux modalités du DGF, ce risque tend à se stabiliser au-delà de cinq ans après la transplantation. Pour les greffés ayant un DGF inférieur à six jours, le risque de décès se stabilise à environ 0,6 alors qu'il se stabilise à environ 0,8 pour les patients dont le DGF est supérieur à six jours.

Comme vu au chapitre 4, l'interprétation des covariables âge, sexe du receveur et année de la greffe est équivoque en ce qui concerne le risque relatif aux transplantés, nous nous intéressons donc seulement à leurs interprétations dans l'analyse du risque observé. De plus, comme l'interprétation des résultats de chaque covariable peut s'avérer dense, nous n'interprétons que les résultats concernant la covariable sexe du receveur.

Le risque observé est donc le risque “non corrigé” par les facteurs âge, sexe et année intervenant dans le risque attendu, *i.e.* risque de toutes les pathologies confondues dans la population générale.

Pour cette covariable, nous choisissons d’interpréter les résultats concernant certaines années de transplantation, ceci dans l’objectif d’alléger les interprétations. De plus, le choix des années pour lesquelles nous faisons les interprétations est complètement arbitraire mais il peut être fait selon l’intérêt des médecins.

Le ratio correspondant à la covariable sexe du receveur est le suivant :

$$\frac{\lambda_{\text{Obs}}(t|Z_{\text{Sexe}} = 1)}{\lambda_{\text{Obs}}(t|Z_{\text{Sexe}} = 0)} = \frac{\lambda_{\text{Att}}(\text{Sexe} = 1, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \exp(\beta_{\text{Sexe}})}{\lambda_{\text{Att}}(\text{Sexe} = 0, \text{Age à la greffe} = a, \text{Année de la greffe} = y)}$$

La figure (5.2) illustre ce ratio pour $y = (1994, 2002)$. Elle contient donc deux graphiques (un pour chaque année). Ces deux graphiques sont quasiment similaires et montrent que le ratio est inférieur à 1. Ainsi, entre 18 et 80 ans, être une femme constitue un facteur protecteur pour le décès compte tenu de tous les autres risques de décès possibles, *i.e.* dus à d’autres pathologies ou simplement décès accidentels ou naturels.

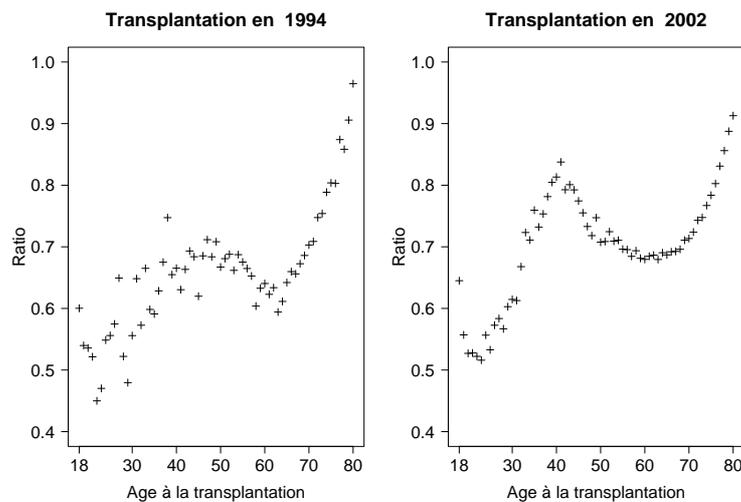


FIG. 5.2 – Modèle de survie relative avec fragilité sur les centres : Ratio pour la covariable sexe du receveur.

Nous allons maintenant appliquer le modèle fragilisé de survie relative avec la fragilité modélisant la dépendance des patients transplantés la même année.

5.2.2 Fragilité sur les années de transplantation

Comme la fragilité porte ici sur les années de transplantation, nous ne testons pas les années comme facteur potentiellement explicatif du décès. La table (5.3) donne les covariables significatives suivant les différentes stratégies de sélection. Sachant les effets de l’âge, du sexe et de l’année calendaire sur la mortalité générale et sachant la dépendance entre les

patients greffés la même année, la covariable spécifique du décès est le DGF.

Covariables	Stratégie 1		Stratégie 2
	Univarié	Multivarié	Multivarié
Sexe du receveur	×	×	×
Age du receveur	×	×	×
DGF	×	×	×
Incompatibilité Anti HLA			
Traitement immunosuppresseur			
Type de maladie initiale			
Antécédents cardiologiques			
Relation donneur / receveur			
Ischémie			
Néoplasie			
Age du donneur			
Sexe du donneur			
Centre de greffe			

TAB. 5.3 – *Modèle Multiplicatif de Survie Relative avec Fragilité (MSRF) sur les années de transplantation : covariables retenues (×) après les différentes stratégies de sélection (explications page 25).*

Paramètres	coef	exp(coef)	se	<i>p</i> -value
σ	2,152	.	1,032	.
ν	0,701	.	0,047	< 0,05*
γ	1,476	.	0,062	> 0,05•
Sexe du receveur	0,514	1,671	0,148	0,001
Age du receveur	-0,382	0,682	0,139	0,006
DGF	0,299	1,348	0,138	0,030

* Test du rapport de vraisemblance : $\{H_0 : \nu = 1\}$.

• Test du rapport de vraisemblance entre le modèle avec fragilité et le même modèle sans fragilité.

TAB. 5.4 – *Résultats du modèle Multiplicatif de Survie Relative avec Fragilité (MSRF) sur les années de transplantation ($\log(\mathcal{L}) = -1143,358$).*

Le tableau (5.4) montre que les résultats concernant les effets fixes du modèle sont à peu près identiques à ceux du modèle multiplicatif de survie relative sans fragilité (de l'ordre de 0,001). Par contre, comme pour le modèle avec la fragilité modélisant la dépendance intra-centre, les résultats de ce modèle diffèrent du modèle de fragilité sans survie relative, *i.e.* inversion des signes des paramètres des covariables âge et sexe entre les deux modèles. Les résultats de ce modèle de survie relative fragilisé sur les années de transplantation concluent donc que : sachant les effets de l'âge, du sexe et de l'année de la greffe sur le décès toutes causes confondues et sachant la dépendance des patients greffés les mêmes années, avoir un greffon qui met plus de 6 jours à reprendre fonction est un facteur de risque du décès après une transplantation rénale.

Nous testons la significativité de la fragilité par un test du rapport de vraisemblance entre le modèle multiplicatif de survie relative avec fragilité et celui sans fragilité. La statistique de ce test vaut $-74,281$ correspondant à une *p*-valeur supérieure à 0,05. Ce test indique donc

que la fragilité n'est pas significative. Il n'y a donc pas de lien entre les patients greffés la même année. Ce résultat diffère des précédents à savoir le modèle de fragilité de Cox et celui de Weibull présentés au chapitre 3. En effet, ces modèles prédisent un effet année de transplantation. Ici, en prenant en compte les effets de l'âge, du sexe et de l'année calendaire sur la mortalité générale, *i.e.* toutes pathologies confondues, il ne semble pas y avoir d'effet année de transplantation. Ceci peut s'expliquer par le fait que l'effet année est déjà pris en considération au travers du risque attendu de décès issu de la population générale.

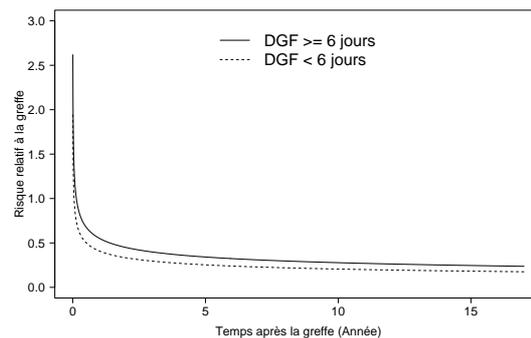


FIG. 5.3 – Fonction de risque selon les modalités des covariables sous le modèle multiplicatif de survie relative de Weibull avec fragilité sur les années de transplantation.

La figure (5.3) représente le risque de décès relatif aux greffés en fonction du temps et selon les modalités de la covariable DGF. Ce risque est de l'ordre de 2,5 juste après la transplantation pour les patients ayant leur greffon qui met plus de six jours à reprendre fonction alors que pour ceux dont le greffon met moins de six jours, ce risque vaut environ 1,7. Pour les deux modalités du DGF, ce risque tend à se stabiliser environ cinq ans après la transplantation. Pour les greffés ayant un DGF inférieur à six jours, le risque de décès se stabilise à environ 0,15 alors qu'il se stabilise à environ 0,23 pour les patients dont leur DGF est supérieur à six jours.

Pour le risque observé, comme précédemment, nous choisissons d'interpréter seulement les résultats concernant la covariable sexe du receveur.

Le ratio correspondant à cette covariable s'écrit de la même façon que précédemment :

$$\frac{\lambda_{\text{Obs}}(t|Z_{\text{Sexe}} = 1)}{\lambda_{\text{Obs}}(t|Z_{\text{Sexe}} = 0)} = \frac{\lambda_{\text{Att}}(\text{Sexe} = 1, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \exp(\beta_{\text{Sexe}})}{\lambda_{\text{Att}}(\text{Sexe} = 0, \text{Age à la greffe} = a, \text{Année de la greffe} = y)}$$

La figure (5.4) illustre ce ratio pour $y = (1994, 2002)$. Ces deux graphiques montrent que le ratio est inférieur à 1. Ainsi, entre 18 et 80 ans, être une femme constitue un facteur protecteur pour le décès compte tenu de tous les autres risques de décès possibles.

Encore une fois, nous remarquons que les différentes analyses que nous pouvons effectuer avec la survie relative sont riches : analyse du

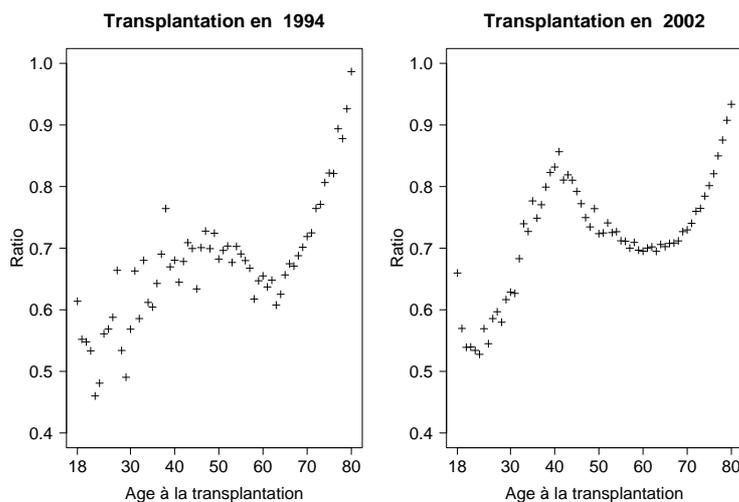


FIG. 5.4 – Modèle de survie relative avec fragilité sur les années : Ratio pour la covariable sexe du receveur.

risque relatif à la pathologie étudiée ou analyse du risque observé du groupe étudié.

DISCUSSION-CONCLUSION

Dans ce chapitre, nous avons considéré dans un même modèle la notion de survie relative vue au chapitre 4 et la notion de fragilité explicitée au chapitre 3. La survie relative permet d'estimer la survie propre aux transplantés rénaux en éliminant toutes les autres causes de décès. Quant à la fragilité, elle permet de prendre en compte l'hétérogénéité du groupe de patients étudié, cette hétérogénéité est considérée comme non observable. Il peut s'agir de facteurs génétiques ou environnementaux. La fragilité estime donc un effet de groupe qui n'est pas observable. Ce groupe peut représenter des individus appartenant à une même famille ou des observations répétées pour un même individu. Dans le cas de la base DIVAT, comme un effet intra-centre (respectivement un effet année de transplantation) est suspecté dès le chapitre 1, la fragilité est utilisée pour modéliser la dépendance des patients greffés dans un même centre (respectivement la dépendance des patients transplantés la même année).

Le modèle présenté dans ce chapitre apporte de nouveaux résultats par rapport aux précédents. En effet, il ne valide pas l'effet année de transplantation suspecté au chapitre 1. Cet effet est pourtant confirmé par le modèle de fragilité de Cox et celui de Weibull. Ici, en prenant en compte les effets dus à la mortalité générale (âge, sexe, année), la dépendance des patients greffés la même année n'est pas validée. Ceci peut s'expliquer par le fait que l'effet année est pris en considération dans le risque de la population générale. Cependant, le modèle de survie relative avec fragilité sur les centres de transplantation confirme, tout comme les modèles de fragilité de Cox et de Weibull, qu'il n'y a pas d'effet centre de

transplantation. Ainsi, les individus ne sont pas corrélés selon le centre dans lequel ils sont transplantés.

Après de tels résultats, la question est de savoir quel modèle choisir. En effet, deux choix s'offrent à nous.

Le premier est de considérer le modèle de survie relative. Ce modèle a l'avantage d'estimer la survie spécifique à la maladie. De plus, il prend en compte des valeurs réelles de mortalité à travers les tables vitales fournies par l'INSEE. L'année de transplantation est donc implicitement prise en compte dans ce modèle. Par ailleurs, l'estimation de ce modèle ne nécessite pas l'ajout de paramètres supplémentaires par rapport à un modèle de survie classique. Le seul problème peut être l'estimation du risque cumulé relatif de base. En effet, cette estimation peut manquer de précision dans le sens où ce risque cumulé est calculé grâce à une somme d'aires de la fonction de risque relatif de base (équation (4.14) du chapitre 4). Ceci est une estimation annuelle et peut entraîner une perte de précision. En effet, il y a une différence, par exemple, entre 2,1 et 2,8 années de survie.

Le deuxième choix possible est de considérer le modèle de survie avec la fragilité modélisant la dépendance des patients greffés les mêmes années. Cependant, ce modèle nécessite l'estimation d'un paramètre supplémentaire, celui de la fragilité, par rapport à un modèle classique de survie. La fragilité est utilisée pour modéliser l'effet de l'année de transplantation qui n'est pas une variable vraiment explicite. En effet, le facteur année de transplantation englobe d'autres facteurs pas forcément mesurables tels que les stratégies de transplantation par exemple.

Nous remarquons que ces deux modèles prennent en compte l'année de transplantation mais que le modèle de fragilité nécessite l'estimation d'un paramètre supplémentaire. De plus, l'AIC (Akaike Information Criterion) du modèle de survie relative vaut 2224,434 et celui du modèle de fragilité portant sur les années de transplantation vaut 2217,84. Ce critère nous amène donc à choisir le modèle de fragilité car il minimise l'AIC.

Cependant, dans la suite de ce mémoire, notre choix s'oriente plutôt sur le modèle de survie relative. En effet, ce modèle est préféré car il estime une survie nette, spécifique aux transplantés rénaux tout en considérant les effets de l'âge, du sexe et de l'année *via* les tables vitales de l'INSEE. Le risque d'erreur d'estimation avec ce modèle est moindre que le modèle avec fragilité. En effet, dans le modèle de fragilité, il faut estimer le paramètre propre à la fragilité alors que dans le modèle de survie relative il n'y a pas d'autres paramètres à estimer. De plus, le recul pour les dernières années observées n'étant pas assez grand (effet cohorte limité à 2006), ceci peut fausser l'estimation de la fragilité, il est donc préférable de prendre l'effet de l'année en utilisant la survie relative.

PRISE EN COMPTE DE LA CENSURE INFORMATIVE

SOMMAIRE

6.1	QU'EST-CE QU'UN MODÈLE SEMI-MARKOVIENT ?	93
6.2	LE MODÈLE PROPOSÉ	94
6.2.1	Notations et hypothèses	94
6.2.2	Introduction des covariables	96
6.2.3	Ecriture de la vraisemblance	96
6.3	TEST DE L'HYPOTHÈSE D'HOMOGENÉITÉ	98
6.3.1	La statistique de test	98
6.4	APPLICATION AUX TRANSPLANTÉS RÉNAUX	100
6.4.1	Stratégie de sélection des covariables	101
6.4.2	Estimation des paramètres du modèle	101
6.4.3	Test de l'hypothèse d'homogénéité	107
	DISCUSSION-CONCLUSION	113

LA plupart des analyses de survie font généralement l'hypothèse d'indépendance entre le délai de survenue de l'évènement et le délai de survenue de la censure. Dans certaines situations telle que les censures non planifiées ou non administratives, cette hypothèse n'est pas valable. Dans ce chapitre, nous proposons une méthode pour considérer la censure comme informative dans un modèle de survie relative multiplicatif.

Cette méthode s'inspire des modèles multi-états. Ce type de modèles recourt aux notions d'état et de processus pour décrire un phénomène. La notion de processus permet de définir un ensemble d'états successivement occupés à chaque temps d'observation par un patient et ainsi représente l'évolution clinique de celui-ci à travers les différents stades d'une pathologie. Grâce aux modèles multi-états, de nombreuses dynamiques complexes peuvent être étudiées.

Il existe plusieurs types de modèles multi-états. Les plus simples, cas des modèles progressifs qui sont des modèles de succession d'états sans retour possible, supposent que l'information sur l'état présent renseigne

sur les états précédents. Néanmoins, lorsque la dynamique du processus est plus complexe et qu'elle comprend des états réversibles, des hypothèses sur la trajectoire des patients doivent être établies. Pour cela, il existe les modèles markoviens qui supposent que l'information sur les états précédents est résumée par l'état présent.

Plusieurs formes de modèles markoviens sont possibles selon le choix de l'échelle du temps. En effet, en notant $\alpha(t, d)$ les intensités de transition où t représente la durée de suivi et d la durée passée dans l'état (temps depuis la dernière transition), plusieurs cas se présentent :

- * si les intensités de transition ne dépendent ni du temps chronologique ni de la durée dans l'état, i.e. $\alpha(t, d) = \alpha$, le modèle markovien est dit homogène par rapport au temps,
- * si $\alpha(t, d) = \alpha(t)$ i.e. les intensités de transition ne dépendent que du temps chronologique, c'est un modèle markovien non-homogène,
- * si $\alpha(t, d) = \alpha(d)$, le modèle est semi-markovien homogène par rapport au temps t ,
- * si le modèle dépend de t et de d alors il est appelé semi-markovien non homogène.

Le dernier cas, c'est-à-dire la gestion des deux échelles de temps, demeure un problème assez complexe.

Lorsque l'échelle de temps la plus adéquate est choisie, la nature des intensités de transitions doit être définie. Différentes approches (paramétrique, semi-paramétrique et non paramétrique) sont possibles pour les modéliser.

Comme, en analyse de survie, plusieurs cas de censure sont possibles, les modèles multi-états peuvent intervenir en considérant, par exemple, une censure comme un état. Tout d'abord, les patients perdus de vue sont ceux dont l'état n'est pas connu à la date de point. Ils représentent alors une perte d'information et sont source de biais : dans le cas des patients VIH, par exemple, ceux qui abandonnent le suivi sont généralement les cas les plus graves (50). En effet, le traitement est souvent lourd avec des effets secondaires multiples et provoque ainsi des arrêts de suivi. De plus, il a été observé que ce sont les patients les plus malades qui arrêtent leurs traitements soit parce qu'ils ne supportent plus la thérapie soit parce qu'ils n'ont plus d'espoir. Dans le cas de l'asthme, au contraire, il semble que les patients perdus de vue soient des patients bien contrôlés qui ne ressentent pas le besoin de venir consulter (16). Dans ces deux exemples, le processus de censure est clairement lié à l'évènement étudié puisqu'il apporte une information sur cet évènement. D'autre part, l'autre cas de censure, la censure administrative, est seulement engendrée par la fin de l'étude. Cette censure est indépendante de l'évènement et est alors considérée comme non-informative.

Dans la base DIVAT des transplantés rénaux, le temps étudié est le délai entre la greffe et le décès du patient. Cependant, ce temps peut être censuré pour deux raisons : soit la censure est non-informative puisqu'elle est due au choix de la date de point, soit elle peut être informative si elle est due à un retour en dialyse. En effet, lors d'un retour en dialyse, le suivi du patient n'est plus disponible mais l'étude de Wolfe *et al.* (60)

montre un risque de décès plus important après un retour en dialyse. Par ailleurs, comme le temps observé pour chaque patient est le délai entre la greffe et le décès ou le retour en dialyse, il peut être considéré comme le temps passé dans l'état "bonne santé après la greffe" avant de transiter à une certaine date dans l'état décès ou dans l'état retour en dialyse.

Le chapitre est organisé de la façon suivante. Un rappel de la théorie des processus semi-markoviens est énoncé. A partir de cette théorie, le modèle est appliqué aux transplantés rénaux. En effet, comme le temps observé pour chaque patient peut être considéré comme la durée passée dans un état appelé "bonne santé après la greffe", nous choisissons un processus semi-markovien pour modéliser la dynamique des patients subissant une transplantation rénale. De plus, comme nous l'avons précisé dans les chapitres précédents, les différentes causes de décès ne sont pas disponibles dans la base DIVAT. Ainsi, la notion de survie relative est incluse dans le modèle multi-états. Par ailleurs, inspiré par le travail de thèse de Foucher Y. (27), nous proposons une alternative à la statistique de test de Pearson proposé par Aguirre-Hernandez et Farewell (4). Ceci permet de tester l'hypothèse d'homogénéité du modèle proposé.

6.1 QU'EST-CE QU'UN MODÈLE SEMI-MARKOVIAN ?

Soit E un espace d'états discret. Pour chaque sujet, $X = \{X_0, X_1, \dots, X_n\}$ sont les n états successifs observés où X_0 est l'état initial et X_n est le dernier état après n transitions. On suppose que le nombre total d'états possible est fini et que $X \in E$. On note X_r l'état occupé par le processus après la transition r et on dira que la séquence $X = \{X_r, r \geq 0\}$ forme une **chaîne de Markov**.

Remarque 6.1 *Une chaîne de Markov est un processus stochastique possédant la propriété markovienne. Dans un tel processus, la prédiction du futur à partir du présent ne nécessite pas la connaissance du passé, i.e. :
Si la distribution de probabilité conditionnelle de X_{r+1} sur les états passés est une fonction de X_r seul, alors :*

$$P(X_{r+1} = x | X_0, X_1, X_2, \dots, X_r) = P(X_{r+1} = x | X_r)$$

où x est un état quelconque du processus. L'identité ci-dessus identifie la probabilité markovienne.

Les probabilités de transition de l'état e vers l'état f associées à la chaîne de Markov X sont notées P_{ef} . Elles sont définies par :

$$P_{ef} = P(X_{r+1} = f | X_r = e). \quad (6.1)$$

La probabilité de transition du processus markovien (6.1) correspond à la probabilité que le processus soit dans l'état f à la transition $(r + 1)$ sachant qu'il était dans l'état e à la transition r . Le processus est supposé **homogène** puisque P_{ef} ne dépend pas du temps.

L'état e est dit **absorbant** si :

$$\begin{cases} P_{ef} = 0 \text{ si } e \neq f \\ P_{ef} = 1 \text{ si } e = f. \end{cases}$$

Au contraire, si e n'est pas absorbant :

$$\begin{cases} P_{ef} > 0 \text{ si } e \neq f \\ P_{ef} = 0 \text{ si } e = f. \end{cases}$$

Un état absorbant est un état duquel on ne peut jamais sortir (ie $T_{r+1} - T_r = \infty$). Notons que si le dernier état de la chaîne, X_n , n'est pas absorbant, le temps T_{n+1} est censuré à droite et l'état suivant X_{n+1} n'est pas connu.

De façon analogue à l'analyse de survie et aux équations (2.1) ... (2.5), les différentes fonctions de survie peuvent être écrites (ne sont exposées que celles utilisées dans la suite du chapitre).

Les temps d'entrée successifs dans les états X sont notés $T = \{T_0, T_1, \dots, T_n\}$ où T_r est le temps d'entrée dans l'état r . Pour chaque individu, le couple $(T, X) = \{(T_r, X_r) : r \geq 0\}$ est observé. $T_{r+1} - T_r$ est la durée passée dans l'état X_r avant de transiter en X_{r+1} .

$F_{ef}(t)$ est la fonction de répartition des temps de séjour dans l'état e sachant l'état suivant f . Elle est définie telle que :

$$F_{ef}(t) = P(T_{r+1} - T_r \leq t | X_{r+1} = f, X_r = e). \quad (6.2)$$

Le couple (T, X) est un **processus semi-markovien** si la distribution des temps de séjour $(T_{r+1} - T_r)$ satisfait la condition suivante :

$$\begin{aligned} P(T_{r+1} - T_r \leq t, X_{r+1} = f | X_0, T_0, X_1, T_1, \dots, X_r = e, T_r) \\ = P(T_{r+1} - T_r \leq t, X_{r+1} = f | X_r = e) \end{aligned} \quad (6.3)$$

i.e., tout le passé est résumé par l'état présent.

D'après le théorème de Bayes et les définitions (6.1) et (6.2) :

$$\begin{aligned} P(T_{r+1} - T_r \leq t, X_{r+1} = f | X_0, T_0, X_1, \dots, X_r = e, T_r) \\ = P(T_{r+1} - T_r \leq t | X_{r+1} = f, X_r = e) P(X_{r+1} = f | X_r = e) \\ = F_{ef}(t) P_{ef}. \end{aligned} \quad (6.4)$$

La fonction de survie, *i.e.* la probabilité d'être dans l'état e jusqu'à l'instant t avant de transiter dans l'état f , s'écrit ainsi :

$$S_{ef}(t) = P(T_{r+1} - T_r \geq t | X_{r+1} = f, X_r = e). \quad (6.5)$$

De même, la fonction de densité qui représente la probabilité de transiter de l'état e vers l'état f dans un petit intervalle de temps après t est notée :

$$f_{ef}(t) = \lim_{h \rightarrow 0^+} \frac{P(t < T_{r+1} - T_r < t + h | X_{r+1} = f, X_r = e)}{h}. \quad (6.6)$$

Dans la partie suivante, nous proposons de présenter la dynamique du processus semi-markovien choisi ainsi que la théorie qui s'ensuit.

6.2 LE MODÈLE PROPOSÉ

Dans un premier temps, nous exposons les différentes notations et hypothèses du modèle semi-markovien proposé. Ensuite, nous montrons comment nous avons choisi d'introduire les covariables dans ce modèle. Enfin, l'écriture de la vraisemblance correspondant au modèle est donnée.

6.2.1 Notations et hypothèses

Soit n individus observés et indexés par j ($j = 1, \dots, n$). Soit E , l'ensemble des états pouvant être observés pour chaque individu, $E = \{0, I, D\}$.

Définissons ces 3 états.

- ★ L'état initial, *i.e.*, l'état dans lequel les individus entrent dans l'étude. Tous les individus entrent dans ce même état, noté "Etat 0". Il peut correspondre, par exemple, au diagnostic d'une maladie. De plus, un individu peut rester dans cet état jusqu'à la date de point.

- ★ Le décès, représentant l'évènement d'intérêt dans les modèles de survie relative. Cet état est noté "Etat D".
- ★ Un état compétitif, noté "Etat I". Lorsqu'un individu passe dans cet état, il n'est plus suivi ensuite, *i.e.*, nous ne savons pas ce qu'il devient. L'information sur le décès possible de cet individu n'est pas connue. Cependant, il n'est pas possible de considérer cet état comme une simple censure à droite car nous supposons que cet état modifie le risque de décès (60).

Ces états sont résumés dans la figure 6.1.

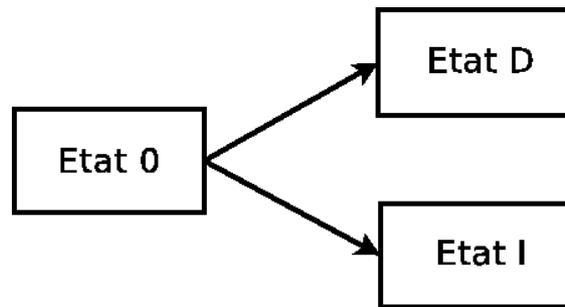


FIG. 6.1 – Modèle à trois états (1 transitoire et 2 absorbants).

Pour chaque individu, on observe au plus une transition. Quand un individu entre dans l'étude, son état initial est noté X_0 . Pour tous les individus $X_0 = 0$. Chaque individu peut soit passer de l'état 0 à l'état D ou I , soit rester dans l'état 0 jusqu'à la fin du suivi. Cette censure est supposée être indépendante des évènements D et I . De plus, pour tous les individus, la transition de l'état I à l'état D ne sera jamais observée puisque le suivi des patients n'est pas continué après un retour en dialyse. Cette transition entre les deux états n'est donc pas étudiée en considérant D et I comme deux échecs compétitifs. Soit X_1 l'état suivant l'état initial.

Pour chaque individu, nous observons un temps T et un état e à la fin du suivi avec $e \in E$. De plus, P^e la probabilité d'être dans l'état e après l'état 0 : $P^e \in [0;1], \forall e = D, I$. Par conséquent, $P^I + P^D = 1$. Soulignons que les probabilités de transition ne dépendent pas du temps. Ainsi, le modèle proposé est un modèle semi-markovien homogène.

Notons la fonction de densité, qui représente la probabilité de transiter vers l'état final en t sachant que l'état final est l'état e , comme :

$$f^e(t) = \lim_{dt \rightarrow 0} \frac{P(t < T < t + dt | X_1 = e)}{dt}, \quad e = D, I. \quad (6.7)$$

De cette fonction de densité, nous pouvons en déduire les autres fonctions de survie habituelles. L'écriture de la fonction de risque est donc la suivante :

$$\lambda^e(t) = \lim_{dt \rightarrow 0} \frac{P(t < T < t + dt | T \geq t, X_1 = e)}{dt}, \quad e = D, I. \quad (6.8)$$

Ce risque représente la probabilité de transiter en t , sachant que l'état e suit et que le processus était dans l'état 0 jusqu'en t .

6.2.2 Introduction des covariables

Pour prendre en compte d'éventuelles covariables dans le modèle, nous reprenons le principe de proportionnalité des risques (20). De plus, dans la fonction de risque de décès, nous considérons le risque de mortalité de la population générale, toutes maladies confondues. Les deux fonctions de risque (décès et censure informative) pour un individu j , $j = 1, \dots, n$, s'écrivent donc sous la forme de modèle de régression :

$$\lambda_{\text{Obs}}^I(t_j|Z_j^I) = \lambda_{\text{Obs},0}^I(t_j) \exp(\beta^I Z_j^I) \quad (6.9)$$

$$\lambda_{\text{Obs}}^D(t_j|Z_j^D) = \lambda_{\text{Att}}^D(t_j + a_{j0}) \underbrace{\lambda_{\text{Rel},0}^D(t_j) \exp(\beta^D Z_j^D)}_{\lambda_{\text{Rel}}^D(t_j|Z_j^D)} \quad (6.10)$$

avec a_{j0} qui représente l'âge de l'individu j à l'inclusion. Rappelons aussi que :

- ★ $\lambda_{\text{Att}}^D(t_j + a_{j0})$ est la fonction de risque de la population générale de même sexe, même âge $t_j + a_{j0}$ et même année de naissance que l'individu j observé.
- ★ $\lambda_{\text{Rel},0}^D$ est la fonction de risque relatif de base spécifique à la maladie étudiée.
- ★ $\lambda_{\text{Obs},0}^I$ est la fonction de risque de base observé à la censure informative étudiée.
- ★ Z_j^I et Z_j^D sont les vecteurs de covariables associés à la censure informative et au décès, respectivement.
- ★ β^I et β^D sont les vecteurs de coefficients de régression inconnus associés à Z_j^I et Z_j^D , respectivement.

A partir de ces deux fonctions de risque, les fonctions de survie s'écrivent :

$$S_{\text{Obs}}^I(t_j|Z_j^I) = \exp\left(-\Lambda_{\text{Obs}}^I(t_j|Z_j^I)\right) \quad (6.11)$$

$$S_{\text{Obs}}^D(t_j|Z_j^D) = \exp\left(-\Lambda_{\text{Obs}}^D(t_j|Z_j^D)\right) \quad (6.12)$$

avec Λ^I et Λ^D les fonctions de risques cumulées de censure informative et de décès. Λ^I dépendra de la fonction de risque de base choisie. Par contre, Λ^D , en plus de dépendre de la fonction de risque de base, s'écrit comme déjà vu au chapitre 4 :

$$\begin{aligned} \Lambda_{\text{Obs}}^D(t_j|Z_j^D) &= \int_0^{t_j} \lambda_{\text{Obs}}^D(u|Z_j^D) du \\ &= \sum_{k=0}^{na_j} \lambda_{\text{Att}}^D(t_{j,k} + a_{j0}) \left(\int_{t_{j,k}}^{t_{j,(k+1)}} \lambda_{\text{Rel},0}^D(u) du \right) \exp(\beta^D Z_j^D). \end{aligned}$$

6.2.3 Ecriture de la vraisemblance

Soient δ_j^I et δ_j^D les indicatrices de censure informative et de décès pour l'individu j avec $j = 1, \dots, n$.

Dans notre contexte, trois types d'observations peuvent être distingués :

★ l'individu j est censuré informativement au temps t_j ($\delta_j^I = 1$) :

$$P^I f_{\text{Obs}}^I(t_j), \quad (6.13)$$

★ l'individu j décède au temps t_j ($\delta_j^D = 1$) :

$$P^D f_{\text{Obs}}^D(t_j), \quad (6.14)$$

★ l'individu j est censuré au temps t_j (avant de passer dans les deux autres états, $1 - \delta_j^I - \delta_j^D = 1$) :

$$\begin{aligned} P(T > t_j) &= P(T > t_j, X_1 = I) + P(T > t_j, X_1 = D) \\ &= P(T > t_j | X_1 = I)P(X_1 = I) + P(T > t_j | X_1 = D)P(X_1 = D) \\ &= S_{\text{Obs}}^I(t_j)P^I + S_{\text{Obs}}^D(t_j)P^D. \end{aligned} \quad (6.15)$$

En faisant le produit de toutes ces contributions, on obtient la vraisemblance suivante :

$$V = \prod_{j=1}^n \left(P^I f_{\text{Obs}}^I(t_j) \right)^{\delta_j^I} \left(P^D f_{\text{Obs}}^D(t_j) \right)^{\delta_j^D} \left(P^I S_{\text{Obs}}^I(t_j) + P^D S_{\text{Obs}}^D(t_j) \right)^{(1 - \delta_j^I - \delta_j^D)}.$$

Ainsi, la logvraisemblance s'écrit :

$$\begin{aligned} \log(V) &= \sum_{j=1}^n \delta_j^I \left(\log(P^I) + \log(\lambda_{\text{Obs}}^I(t_j)) + \log(S_{\text{Obs}}^I(t_j)) \right) \\ &\quad + \delta_j^D \left(\log(P^D) + \log(\lambda_{\text{Obs}}^D(t_j)) + \log(S_{\text{Obs}}^D(t_j)) \right) \\ &\quad + (1 - \delta_j^I - \delta_j^D) \left(P^I S_{\text{Obs}}^I(t_j) + P^D S_{\text{Obs}}^D(t_j) \right). \end{aligned} \quad (6.16)$$

Pour obtenir la fonction de logvraisemblance finale, il suffit alors de définir les différentes fonctions qui la composent.

Pour les fonctions de risque, $\lambda_{\text{Obs},0}^I$ et $\lambda_{\text{Obs},0}^D$, nous choisissons de les modéliser par une distribution de Weibull (voir équation (2.7))

Ici encore, nous choisissons des quantités connues provenant des tables vitales pour les taux de mortalité des individus de la population générale. Le risque cumulé de décès s'écrit ainsi :

$$\Lambda_{\text{Obs}}^D(t_j) = \sum_{k=0}^{na_j} \frac{1}{\sigma_D^{v_D}} (t_{j,(k+1)}^{v_D} - t_{j,k}^{v_D}) \lambda_{\text{Att}}^D(t_{j,k} + a_{j0}) \exp(\beta_D Z_j^D).$$

Pour respecter les contraintes des probabilités de transitions de la chaîne de Markov ($P^e \in [0; 1]$ et $P^I = 1 - P^D$), P^I et P^D sont estimées à partir du coefficient α et de la fonction logit :

$$\ln \left(\frac{P^I}{1 - P^I} \right) = \alpha \text{ d'où } P^I = \left(\frac{\exp(\alpha)}{1 + \exp(\alpha)} \right) \text{ et } P^D = \left(\frac{1}{1 + \exp(\alpha)} \right).$$

Ainsi, voici une version plus décomposée de la logvraisemblance :

$$\begin{aligned}
& \log(V) \\
&= \sum_{j=1}^n \left\{ \delta_j^I \left[\alpha - \log(1 + \exp(\alpha)) + \log(v_I) - v_I \log(\sigma_I) + (v_I - 1) \log(t_j) - \left(\frac{t_j}{\sigma_I} \right)^{v_I} \exp(\beta^I Z_j^I) \right] \right. \\
&+ \delta_j^D \left[-\log(1 + \exp(\alpha)) + \log(\lambda_{\text{Att}}^D(t_j + a_{i0})) + \log(v_D) - v_D \log(\sigma_D) + (v_D - 1) \log(t_j) \right. \\
&\quad \left. \left. - \sum_{k=0}^{na_j} \lambda_{\text{Att}}^D(t_{j,k} + a_{i0}) \frac{1}{\sigma^v} (t_{j,(k+1)}^v - t_{j,k}^v) \exp(\beta^D Z_j^D) \right] \right. \\
&+ (1 - \delta_j^I - \delta_j^D) \left[\frac{\exp(\alpha)}{1 + \exp(\alpha)} \exp\left(-\left(\frac{t_j}{\sigma_I}\right)^{v_I} \exp(\beta^I Z_j^I)\right) \right. \\
&\quad \left. \left. + \frac{1}{1 + \exp(\alpha)} \exp\left(-\sum_{k=0}^{na_j} \lambda_{\text{Att}}^D(t_{j,k} + a_{i0}) \frac{1}{\sigma^v} (t_{j,(k+1)}^v - t_{j,k}^v) \exp(\beta^D Z_j^D)\right) \right] \right\}.
\end{aligned}$$

Comme dans les chapitres précédents, la procédure pour estimer les différents paramètres du modèle est celle du maximum de vraisemblance.

Nous allons maintenant présenter un test pour l'hypothèse d'homogénéité du modèle proposé.

6.3 TEST DE L'HYPOTHÈSE D'HOMOGENÉITÉ

Aguirre-Hernandez et Farewell (4) ont défini une statistique de type Pearson pour examiner la qualité d'ajustement d'un modèle markovien stationnaire à temps continu d'ordre 1. Cette statistique est construite pour les modèles dont les forces de transition dépendent de covariables. Elle est aussi adaptée aux données dont l'espacement entre les mesures et le nombre total d'observations peut varier d'un individu à l'autre. De plus, les auteurs ne supposent pas de distribution asymptotique et proposent une méthode de bootstrap pour calculer la distribution de la statistique de test.

Notre problématique est équivalente puisque le processus d'observation d'un individu est complètement aléatoire, chaque individu possédant des temps de transition différents. De plus, le modèle dépend de covariables. Cependant quelques différences persistent : le modèle est semi-markovien et les temps d'apparition des événements absorbants sont exactement renseignés. En nous inspirant des travaux de thèse de Foucher (27), nous proposons une variante à la statistique de test de type Pearson proposée par Aguirre-Hernandez et Farewell (4). L'hypothèse d'une distribution asymptotique de cette statistique étant trop forte, nous définissons une méthode de bootstrap semi-paramétrique. Rappelons, tout d'abord, le principe du fonctionnement de cette statistique.

6.3.1 La statistique de test

Dans le modèle multi-états défini précédemment, les taux de transitions dépendent des profils des patients (covariables), des temps de transitions

et des séquences d'états observés. Tous ces facteurs doivent être considérés dans la statistique de test.

- ★ **Regroupement en fonction des covariables.** Comme toutes les covariables utilisées sont qualitatives ou discrètes, ce classement est assez direct. Les catégories ainsi définies sont dénotées par l'indice c ($c = 1, \dots, C$).
- ★ **Regroupement en fonction des événements étudiés.** Si le nombre d'états possibles est K , alors en considérant que dans le cas semi-markovien le passage d'un état vers lui-même est impossible, il existe un maximum de $K^2 - K$ transitions observables. En fonction de la structure multi-états étudiée, cet effectif peut être restreint. C'est par exemple le cas si le modèle est uni-directionnel ou si certains états sont absorbants. Les événements seront notés par l'indice k ($k = 1, \dots, K$).
- ★ **Regroupement en fonction des temps d'incidence des événements.** Pour obtenir des effectifs équilibrés, les quantiles des temps d'apparition des événements sont utilisés pour définir les intervalles de classement. On définit L le nombre d'intervalles avec comme bornes $(t_0, t_1, \dots, t_l, \dots, t_L)$. Ce découpage en fonction du temps chronologique permet de tester la stationnarité du processus.

Soit $e_{l,k,c}$ le nombre de transitions attendues dans la cellule relative aux catégories (l, k, c) . De la même manière, posons $n_{l,k,c}$ le nombre de transitions observées pour cette cellule. La statistique d'adéquation de type Pearson, permettant de tester l'adéquation du modèle multi-états est égale à :

$$G = \sum_{l=1}^L \sum_{k=1}^K \sum_{c=1}^C \frac{(n_{l,k,c} - e_{l,k,c})^2}{e_{l,k,c}}. \quad (6.17)$$

La distribution de la statistique G (6.17) dépend à la fois de la taille de l'échantillon n , du nombre de cellules indépendantes du tableau de contingence R et du nombre de paramètres estimés dans le modèle semi-markovien (ι). Si R est petit ($R > \iota$) et n grand, la statistique est supposée suivre approximativement une loi du Chi-deux à $(R - \iota)$ degrés de liberté. Cependant, ces hypothèses ne sont pas valides dans notre cas où R et ι deviennent vite assez grands par rapport à la taille de l'échantillon. L'approximation du Chi-deux n'est donc pas adéquate *a priori*.

La définition de la distribution exacte est d'autant plus difficile que le nombre total d'observations n'est pas fixe. En effet, l'entrée dans un état absorbant au temps t_j termine l'observation d'un sujet pour le reste des intervalles $[t_l; t_{l+1}] \forall t_j < t_l$. L'approximation du Chi-deux serait adéquate en considérant le nombre total d'observations fixe, ainsi la probabilité $P(n_{1,1,1} = N_{1,1,1}, \dots, n_{L,K,C} = N_{L,K,C})$ serait la somme des distributions multinomiales indépendantes et non-identiques.

Des simulations peuvent être utilisées pour fournir des intervalles d'estimation ou des p -values de certains tests lorsque la distribution asymptotique n'est pas disponible. Une méthode d'estimation de la distribution de (6.17) est la génération de B échantillons indépendants de bootstrap, à partir du modèle à tester sous l'hypothèse nulle, selon laquelle le modèle semi-markovien est homogène. On calcule alors une statistique pour chacun de ces échantillons. Quand B tend vers l'infini, la distribution de bootstrap de (6.17) approchera la vraie distribution de la statistique de test sous l'hypothèse nulle (24). A partir de la statistique de test calculée sur l'échantillon initial, la p -value peut alors être calculée à partir de la distribution de la statistique.

6.4 APPLICATION AUX TRANSPLANTÉS RÉNAUX

Dans le cas de la base DIVAT, les deux types de censure sont : la censure administrative qui correspond à la date de point et la censure liée au retour en dialyse, les individus n'étant plus suivis après cette date. La censure administrative est considérée comme non-informative, contrairement à la censure à la date de retour en dialyse. De nombreuses études ont pu montrer que les facteurs explicatifs peuvent être à la fois liés au retour en dialyse et au décès. De plus, les travaux de Wolfe *et al.* (60) ont montré que la mortalité était plus importante après le retour en dialyse.

La figure (6.2) permet de schématiser les différents états que peuvent subir chacun des transplantés rénaux. L'état initial pour chaque patient est "transplanté en bonne santé" puis soit il retourne en dialyse soit il décède soit il est toujours en bonne santé à la date de point.

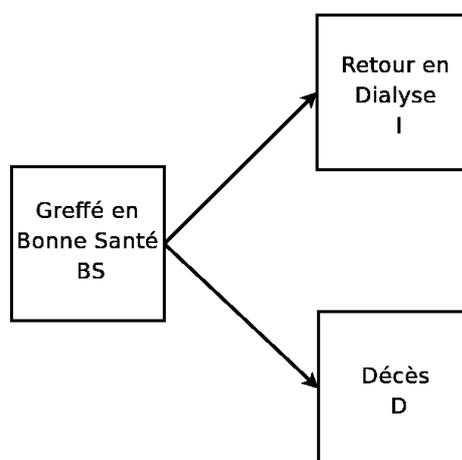


FIG. 6.2 – Schéma des trois états possibles pour chaque patient.

Comme nous l'avons déjà fait remarqué précédemment, les causes de décès des transplantés rénaux de la base DIVAT ne sont pas renseignées, c'est pourquoi nous proposons dans ce modèle (ainsi que dans tous ceux de ce mémoire) d'inclure la notion de survie relative. De plus, afin de percevoir l'apport de cette survie relative, nous choisissons d'appliquer à

la base de données un modèle semi-markovien comme proposé ci-dessus mais sans prendre en compte la survie relative, *i.e.* sans tenir compte des taux de mortalité de la population générale dans la fonction de risque de décès. Avant de présenter les résultats des deux modèles, nous expliquons la procédure pour sélectionner les covariables. En effet, celle-ci diffère légèrement de celle présentée dans le chapitre 2. Cette stratégie s'appliquera aux deux modèles semi-markoviens (avec et sans survie relative).

6.4.1 Stratégie de sélection des covariables

Pour obtenir le modèle le plus parcimonieux possible, les variables explicatives sont choisies selon une méthode de sélection spécifique. La première étape consiste à estimer un modèle de semi-Markov pour chaque covariable à l'aide du maximum de vraisemblance. Notons que sept paramètres sont estimés pour chaque modèle : un paramètre pour les probabilités de transition (α), deux paramètres pour chacune des deux fonctions de Weibull (σ_e, ν_e) et un paramètre pour l'effet de la covariable, ceci pour chaque transition e (le décès et le retour en dialyse). Dans la cohorte, quinze covariables ont été sélectionnées comme potentiellement importantes. Par conséquent, quinze modèles univariés sont estimés avec la même variable explicative pour chaque transition (le décès et le retour en dialyse). À la fin de cette étape, toutes les covariables ayant leur paramètre estimé avec une p -valeur du test de Wald inférieure à 20% sont retenues et incluses dans un modèle multivarié. Le modèle multivarié constitue la deuxième étape de la stratégie de sélection. Les covariables sont choisies par la sélection descendante grâce au test de Wald. La procédure de sélection s'arrête quand toutes les covariables ont leur p -valeur du test de Wald inférieure à 5%. Une fois les deux étapes accomplies, les paramètres des fonctions de risque ν_e sont testés égaux à 1 grâce au test de rapport de vraisemblance. Si l'hypothèse nulle est satisfaite, la distribution est considérée comme Exponentielle, un cas particulier de la distribution de Weibull. Nous ne présentons pas de test pour les paramètres α et σ_e car σ_e est contraint d'être positif et α n'a pas de contrainte.

6.4.2 Estimation des paramètres du modèle

Dans les analyses univariées, 18 covariables (décès et retour en dialyse confondus) sont sélectionnées pour les deux types de modèles semi-markoviens (avec et sans survie relative). Ce ne sont pas les mêmes pour les deux transitions. Après la procédure descendante, seulement neuf covariables sont finalement retenues pour chaque modèle (quatre pour le décès et cinq pour le retour en dialyse). La table (6.1) résume ces différents résultats.

Le tableau (6.2) récapitule les estimations des différents paramètres des deux modèles, avec et sans survie relative. Seules les covariables significatives sont données dans cette table.

Nous analysons les différents résultats des estimations des deux modèles : les probabilités de transition dans les états, les choix des

Covariables	Sans Survie Relative				Avec Survie Relative			
	Univarié		Multivarié		Univarié		Multivarié	
	R	D	R	D	R	D	R	D
Sexe du receveur		×		×		×		×
Age du receveur		×		×		×		×
DGF	×	×		×	×	×		×
Incompatibilité	×		×		×		×	
Anti HLA	×		×		×		×	
Traitement immunosuppresseur								
Type de maladie initiale								
Antécédents cardiologiques		×				×		
Relation donneur / receveur	×				×			
Ischémie	×	×			×	×		
Néoplasie		×				×		
Age du donneur	×	×	×		×	×	×	
Sexe du donneur	×		×		×		×	
Année de la greffe	×	×	×	×	×	×	×	×
Centre	×	×			×	×		

TAB. 6.1 – Modèles semi-markoviens : covariables retenues après les différentes stratégies de sélection (R = Retour en dialyse et D = Décès).

	Sans Survie Relative ($\log(\mathcal{L}) = -3081,64$)				Avec Survie Relative ($\log(\mathcal{L}) = -3067,708$)			
	coef	exp(coef)	se(coef)	p-value	coef	exp(coef)	se(coef)	p-value
α	0,259	—	0,275	—	0,073	—	0,222	—
Décès :								
σ_D	47,170	—	18,426	—	0,105	—	0,052	—
ν_D	0,845	—	0,056	< 0,05*	0,694	—	0,048	< 0,05*
Age du receveur	1,138	3,121	0,152	0,000	-0,354	0,702	0,150	0,019
Sexe du receveur	-0,336	0,715	0,161	0,005	0,502	1,652	0,160	0,002
DGF	0,432	1,540	0,153	0,000	0,352	1,422	0,150	0,019
Année de la greffe	-0,611	0,543	0,158	0,000	-0,572	0,565	0,158	0,000
Retour en dialyse :								
σ_I	38,818	—	10,296	—	0,031	—	0,007	—
ν_I	0,926	—	0,046	< 0,05*	—	—	—	> 0,05*
Age du donneur	0,469	1,598	0,135	0,001	0,490	1,632	0,139	0,000
Sexe du donneur	0,273	1,313	0,116	0,019	0,279	1,322	0,119	0,019
Incompatibilité	0,390	1,477	0,115	0,001	0,415	1,515	0,118	0,000
Anti HLA	0,306	1,358	0,127	0,016	0,306	1,358	0,132	0,020
Année de la greffe	-0,743	4,7598	0,122	0,000	-0,728	0,483	0,123	0,000

* Test du rapport de vraisemblance : $\{H_0 : \nu_e = 1\}$.

TAB. 6.2 – Estimation des paramètres des deux modèles semi-markoviens.

distributions des fonctions de risque et les coefficients des covariables significatives.

Modèle semi-markovien sans survie relative

Pour le modèle semi-markovien sans survie relative, $\alpha = 0,259$ donc $P^I = (\exp(0,259)/(1 + \exp(0,259))) = 0,436$. Ce nombre correspond à probabilité de transiter dans l'état de retour en dialyse. La probabilité de transiter dans l'état décès est de 0,564.

Pour les transitions dans les états décès et retour en dialyse, nous avons supposé des fonctions de risque de Weibull. Mais, nous allons tester si ces fonctions de risque ne suivent pas plutôt des distributions Exponentielles. Dans un premier temps, pour la transition dans l'état décès, nous faisons un test de rapport de vraisemblance entre un modèle avec des fonctions de risque de Weibull pour les deux transitions et un modèle avec une fonction de risque Exponentielle pour le décès et une fonction de risque de Weibull pour le retour en dialyse, *i.e.* $\{H_0 : \nu_D = 1\}$. La statistique de ce test vaut 5,123 ce qui correspond à une p -valeur inférieure à 5%. Nous rejetons donc l'hypothèse nulle suivant laquelle la fonction de risque pour la transition dans l'état décès suit une distribution Exponentielle. Ensuite, nous faisons le même genre de test pour la fonction de risque pour la transition dans l'état retour en dialyse, *i.e.* nous faisons un test de rapport de vraisemblance entre un modèle avec des fonctions de risque de Weibull pour les deux transitions et un modèle avec une fonction de risque de Weibull pour le décès et une fonction de risque Exponentielle pour le retour en dialyse, *i.e.* $\{H_0 : \nu_I = 1\}$. La statistique de ce test vaut 4,408 ce qui correspond à une p -valeur inférieure à 5%. Nous rejetons donc l'hypothèse nulle suivant laquelle la fonction de risque pour la transition dans l'état retour en dialyse suit une distribution Exponentielle. En conclusion, les fonctions de risque pour le retour en dialyse et pour le décès suivent des distributions de Weibull.

Nous nous intéressons maintenant aux covariables significatives pour les deux types de transition. Pour le retour en dialyse, ce sont les covariables âge et sexe du donneur, incompatibilité, anti HLA et année de la greffe qui ressortent significatives. Les facteurs de risque de retourner en dialyse après une greffe de rein sont :

- * avoir un donneur de plus de 55 ans,
- * avoir un donneur femme,
- * avoir plus de 4 incompatibilités avec le donneur,
- * avoir un taux d'immunisation supérieur à 0% et
- * avoir été greffé avant 1996.

Les covariables ayant un effet sur le décès sont l'âge et le sexe du receveur, l'année de transplantation et le DGF. Les interprétations des résultats restent semblables à celles d'un modèle de type Cox (voir le chapitre 2). D'après le modèle semi-markovien sans survie relative,

- * les hommes ont environ 1,4 fois plus de risque de décéder que les femmes après une greffe,

- * les transplantés à plus de 55 ans ont environ 3,1 fois plus de risque de décéder que ceux transplantés à moins de 55 ans,
- * les transplantés dont le greffon reprend fonction plus de 6 jours après la greffe ont environ 1,5 fois plus de risque de décéder que les transplantés dont le greffon reprend fonction avant 6 jours,
- * enfin, les patients transplantés avant 1996 ont environ 1,8 fois plus de chance de décéder que ceux transplantés après 1996.

La figure (6.3) montre les fonctions de risque de décès au cours du temps selon les modalités des covariables. Les différents graphiques confirment les résultats du tableau (6.2). Par ailleurs, ces fonctions de risques n'excèdent pas 0,12 sauf celle pour l'âge du receveur (0,25). De plus, ces risques de décès sont élevés dès les premiers instants après la greffe et ils ont tendance à se stabiliser environ cinq ans après la greffe.

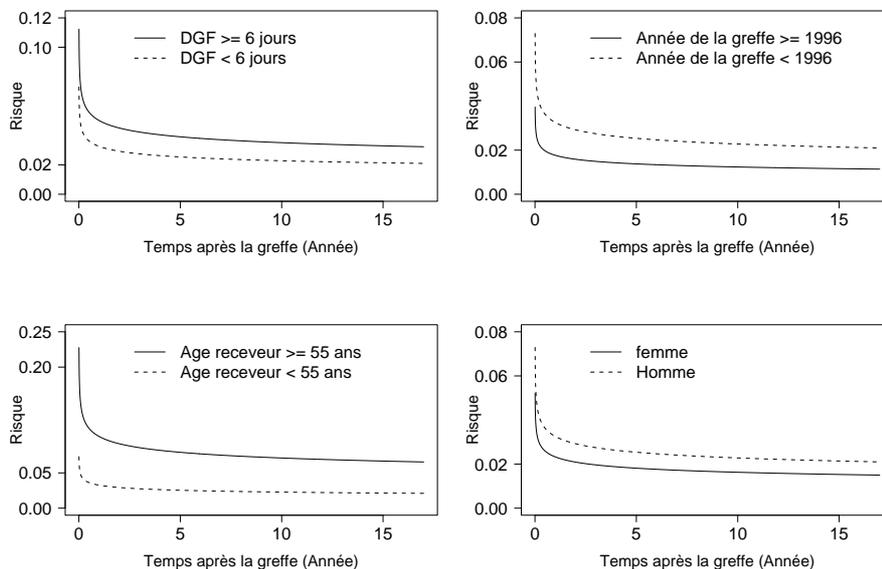


FIG. 6.3 – Modèle semi-markovien : Graphique de la fonction de risque relatif à la greffe selon les modalités des covariables.

Par ailleurs, nous remarquons que, pour les passages dans les deux états (décès et retour en dialyse), la covariable année de la greffe apparaît significative. Il semblerait donc que le décès et le retour en dialyse soient dépendants au travers de cette covariable.

Le prochain paragraphe concerne l'interprétation des résultats du modèle semi-markovien avec survie relative.

Modèle semi-markovien avec survie relative

Avec le modèle semi-markovien avec survie relative, la probabilité de transiter dans l'état retour en dialyse est de 0,482 ($\alpha = 0,073$). Ainsi, la probabilité de transiter dans l'état décès est de 0,518. Bien que ces résultats soient proches avec ceux du modèle semi-markovien sans survie relative, la probabilité de transiter dans l'état décès est plus faible quand

on considère le modèle semi-markovien avec survie relative.

Pour les transitions dans les états décès et retour en dialyse, nous avons supposé des fonctions de risque de Weibull. Mais, comme fait pour le modèle semi-markovien sans survie relative, nous testons si ces fonctions de risque ne suivent pas plutôt des distributions Exponentielle. Tout d'abord, pour la fonction de la transition dans l'état décès, nous faisons un test de rapport de vraisemblance entre un modèle avec des fonctions de risque de Weibull pour les deux transitions et un modèle avec une fonction de risque Exponentielle pour le décès et une fonction de risque de Weibull pour le retour en dialyse. La statistique de ce test vaut 6.862 ce qui correspond à une p -valeur inférieure à 5%. Nous rejetons donc l'hypothèse nulle suivant laquelle la fonction de risque pour la transition dans l'état décès suit une distribution Exponentielle. Ensuite, nous faisons un test de rapport de vraisemblance entre un modèle avec des fonctions de risque de Weibull pour les deux transitions et un modèle avec une fonction de risque de Weibull pour le décès et une fonction de risque Exponentielle pour le retour en dialyse. La statistique de ce test vaut 2,474 ce qui correspond à une p -valeur supérieure à 5%. Ainsi, nous ne rejetons pas l'hypothèse nulle suivant laquelle la fonction de risque pour la transition dans l'état retour en dialyse suit une distribution Exponentielle. Nous concluons donc que la fonction de risque pour le retour en dialyse suit une distribution Exponentielle et celle pour le décès suit une distribution de Weibull.

Nous analysons maintenant les coefficients des différentes covariables significatives. Pour le retour en dialyse, ce sont les covariables âge du donneur, sexe du donneur, incompatibilité, anti HLA et année de la greffe qui ressortent significatives. Nous retrouvons les mêmes facteurs de risque de retour en dialyse que pour le modèle semi-markovien sans survie relative à savoir :

- * avoir un donneur âgé de plus de 55 ans,
- * avoir un donneur femme,
- * avoir plus de 4 incompatibilités avec le donneur,
- * avoir un taux d'immunisation supérieur à 0% et
- * avoir été greffé avant 1996.

Nous nous intéressons maintenant aux covariables pour le décès. Les covariables ressortant significatives sont l'âge et le sexe du receveur, le DGF et l'année de la greffe. Cependant, comme vu dans le chapitre 4, l'analyse des covariables âge et sexe du receveur et année de la greffe n'est pas la même que dans un modèle de survie classique (*i.e.* sans survie relative). L'âge et le sexe du receveur ainsi que l'année de la greffe ne peuvent pas être considérées comme des covariables classiques. Ceci car elles interviennent déjà dans le risque de mortalité de la population générale. Elles ont plutôt un rôle d'ajustement. Lorsqu'un modèle avec survie relative est considéré, nous pouvons interpréter deux risques de décès différents : le risque de décès du groupe observé et le risque de décès relatif à la pathologie. Ainsi, nous analysons les deux types de risques. Cependant, pour la covariable DGF, l'analyse reste habituelle (voir chapitre 4). De plus, l'effet de cette covariable est la même que l'on se place du point de vue du

risque observé du groupe des transplantés ou du point du risque relatif à la transplantation rénale. Ainsi, seule la covariable DGF est analysée pour le risque relatif à la greffe. Pour le risque observé des transplantés, nous choisissons, comme au chapitre 4, d'interpréter seulement les résultats concernant la covariable sexe du receveur .

Ce modèle confirme qu'avoir un greffon mettant plus de 6 jours à reprendre fonction est un facteur de risque pour le décès. Le ratio des risques relatifs est de 1,422. Pour le modèle semi-markovien sans survie relatif, ce ratio est de 1,54. La figure (6.4) montre le risque de décès relatif aux transplantés rénaux en fonction du temps après la greffe et selon la covariable DGF. Nous remarquons que le risque relatif pour les transplantés dont le greffon met plus de six jours à reprendre fonction est d'environ 23 dès les premiers instants après la greffe alors que ce même risque pour les transplantés dont le greffon met moins de six jours à reprendre fonction est de l'ordre de 16. Cinq ans après la greffe, ces risques se stabilisent à 2 pour les greffés dont le DGF est supérieur à six jours et à 1,4 pour ceux dont le DGF est inférieur à six jours. Par ailleurs, comparé au modèle semi-markovien sans survie relative, ces risques relatifs aux transplantés sont de l'ordre de 0,1 et 0,07, respectivement. Ici, quel que soit le temps que met le greffon à reprendre fonction, l'ajout de la survie relative dans le modèle semi-markovien montre que le risque de décès spécifique à la greffe de rein est nettement supérieur au risque de décès du groupe observé qui ne distingue pas le risque propre aux greffés rénaux. De plus, pour le modèle qui ne prend pas en compte la censure informative, *i.e.* le modèle multiplicatif de survie relative proposé au chapitre 4, les risques relatifs aux transplantés sont de l'ordre de 12 pour les patients dont le greffon met plus de six jours à reprendre fonction et 8 pour ceux dont le greffon met moins de six jours. Ces risques se stabilisent cinq ans après la greffe à environ 1 et 0,75, respectivement. Nous remarquons donc que la prise en compte de la censure informative dans le modèle de survie relative augmenterait le risque de décès relatif aux transplantés.

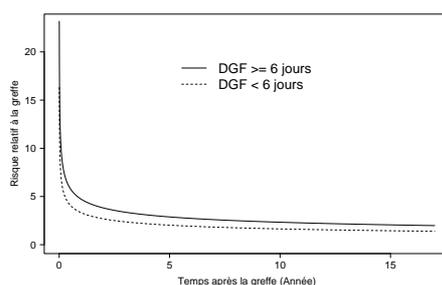


FIG. 6.4 – Modèle semi-markovien avec survie relative : Graphique de la fonction de risque relatif à la greffe selon les modalités de la covariable DGF.

Comme vu au chapitre 4, l'interprétation des covariables âge, sexe du receveur et année de la greffe est équivoque en ce qui concerne le risque relatif aux transplantés, nous nous intéressons donc, uniquement, à leurs interprétations dans l'analyse du risque observé. De plus, comme dit précédemment, nous n'interprétons que les résultats concernant la covariable sexe du receveur. Le risque observé est donc le risque "non corrigé" par

les facteurs âge, sexe et année intervenant dans le risque attendu, *i.e.* risque de toutes les pathologies confondues dans la population générale.

Pour cette covariable, nous choisissons d'interpréter les résultats de certaines années de transplantation, ceci dans l'objectif d'alléger les interprétations. En effet, les patients sont greffés entre 1990 et 2006 et ils ont entre 18 et 80 ans. De plus, le choix des années pour lesquelles nous faisons les interprétations est complètement arbitraire mais il peut être fait selon l'intérêt de l'étude.

Le ratio correspondant à la covariable sexe du receveur est le suivant :

$$\frac{\lambda_{\text{Obs}}^D(t|Z_{\text{Sexe}} = 1)}{\lambda_{\text{Obs}}^D(t|Z_{\text{Sexe}} = 0)} = \frac{\lambda_{\text{Att}}^D(\text{Sexe} = 1, \text{Age à la greffe} = a, \text{Année de la greffe} = y) \exp(\beta_{\text{Sexe}}^D)}{\lambda_{\text{Att}}^D(\text{Sexe} = 0, \text{Age à la greffe} = a, \text{Année de la greffe} = y)}$$

La figure (6.5) illustre ce ratio pour $y = (1994, 2002)$. Elle contient donc deux graphiques (un pour chaque année). Ces deux graphiques sont quasiment similaires et montrent que le ratio est inférieur à 1. Ainsi, entre 18 et 80 ans, être une femme constitue un facteur protecteur pour le décès. Ceci, en prenant en compte tous les autres risques de décès possibles, *i.e.* dus à d'autres pathologies ou simplement décès accidentels ou naturels.

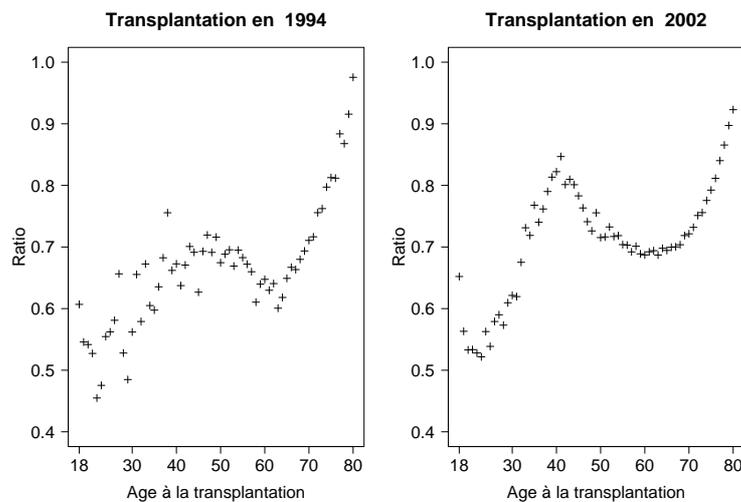


FIG. 6.5 – Modèle semi-markovien avec survie relative : Ratio pour la covariable sexe du receveur.

Nous remarquons donc, encore une fois, les différentes analyses que nous pouvons effectuer avec la survie relative : analyse du risque relatif à la pathologie étudiée ou analyse du risque observé du groupe étudié. Comme déjà vu, la dernière analyse tient compte de tous les autres facteurs de décès possibles en plus de celui de la pathologie étudiée.

6.4.3 Test de l'hypothèse d'homogénéité

Nous procédons maintenant au fondement du test de l'hypothèse d'homogénéité du modèle semi-markovien dans le cas des transplantés rénaux. Ce

test sera appliqué pour les deux modèles semi-markoviens (avec et sans survie relative).

Définition du tableau de contingence

Concernant le regroupement des événements, les temps d'apparition des deux états absorbants de notre modèle sont exactement connus. Ceci permet de facilement les classer selon le temps depuis la greffe. Nous considérons ainsi deux types d'observations ($K = 2$) : passage de l'état 0 vers l'état I et le passage de l'état 0 vers l'état D . K représente donc le nombre d'états absorbants et terminaux.

Nous avons choisi quatre intervalles de temps ($L = 4$) pour obtenir des effectifs assez nombreux dans chaque cellule. Les bornes sont définies de telle sorte que les effectifs d'événements absorbants soient répartis de façon égale : ($l = 1, \dots, 4$), $t_l = \{0, 0055; 0, 1, 4435; 4, 1315; 7, 3720; 16, 2650\}$.

La table (6.3) montre les différents effectifs selon les covariables et les transitions subies par les patients. Aucun regroupement selon les modalités des covariables n'est cliniquement intuitif. Par ailleurs, la stratégie consistant à grouper les individus selon certains facteurs de risque aboutirait à des cellules de faibles effectifs. En effet, plus le groupement des patients a de croisements de caractéristiques (*i.e.* âge, sexe, DGF, *etc*) plus le nombre de patients par groupement est petit. De plus, l'effet des covariables est testé par la statistique du rapport de vraisemblance. Les observations seront ainsi étudiées quel que soit le profil des patients ($C = 1$).

Le principal objectif de la statistique de test (6.17) est donc d'examiner l'hypothèse d'homogénéité du modèle semi-markovien proposé.

Calcul des effectifs

En reprenant les expressions (6.13) et (6.14), la probabilité qu'un individu h transite vers l'état k ($k = I, D$) au temps t depuis la greffe est égale à :

$$P_k f_k(t).$$

Ainsi, par intégration, la probabilité que cet événement k se produise entre t_{l-1} et t_l ($l = 1, \dots, L$) est égale à :

$$P_k \int_{t_{l-1}}^{t_l} f_k(t) dt. \quad (6.18)$$

En prenant en compte la censure à droite, l'effectif attendu de transitions dans la cellule (l, k) s'écrit :

$$e_{l,k} = \sum_{R(t_{l-1})} P_k \int_{t_{l-1}}^{\min(t_h, t_l)} f_k(t) dt. \quad (6.19)$$

où $\min(t_h, t_l)$ est le minimum entre l'éventuel temps de censure à droite pour le sujet h et la borne de l'intervalle (t_l). $R(t_{l-1})$ dénote la somme sur

Covariables	Transitions						
	0 → I		0 → D		0 → Censure		
	416	100%	216	100%	2232	100%	
Sexe Receveur	Homme	247	59,38%	149	68,98%	1386	62,10%
	Femme	169	40,63%	67	31,02%	846	37,90%
Age Receveur	< 55 ans	289	69,47%	93	43,06%	1508	67,56%
	≥ 55 ans	127	30,53%	123	56,94%	724	32,44%
DGF	< 6 jours	253	60,82%	118	54,63%	1501	67,25%
	≥ 6 jours	163	39,18%	98	45,37%	731	32,75%
Année de la greffe	< 1996	259	62,26%	133	61,57%	475	21,28%
	≥ 1996	157	37,74%	83	38,43%	1757	78,72%
Sexe Donneur	Homme	264	63,46%	141	65,28%	1435	64,29%
	Femme	152	36,54%	75	34,72%	797	35,71%
Age Donneur	< 55 ans	321	77,16%	163	75,46%	1677	75,13%
	≥ 55 ans	95	22,84%	53	24,54%	555	24,87%
Incompatibilité D/R	< 4	224	53,85%	135	62,50%	1335	59,81%
	≥ 4	192	46,16%	81	37,50%	897	40,19%
Anti HLA	< 0	302	72,60%	169	78,24%	1880	84,23%
	≥ 0	114	27,40%	47	21,76%	352	15,77%
Traitement	0	2	0,48%	3	1,39%	19	0,85%
	1	290	69,71%	149	68,98%	1133	50,76%
	2	87	20,91%	35	16,20%	444	19,89%
	3	37	8,89%	29	13,43%	636	28,49%
Maladie i	0	250	60,10%	145	67,13%	1451	65,01%
	1	166	39,90%	71	32,87%	781	34,99%
Antécédent Cardio	0	89	21,39%	32	14,81%	420	18,82%
	1	327	78,61%	184	85,19%	1812	81,18%
Relation	0	13	3,13%	6	2,78%	156	6,99%
	1	403	96,87%	210	97,22%	2076	93,01%
Ischémie	< 24	182	43,75%	90	41,67%	1251	56,05%
	≥ 24	234	56,25%	126	58,33%	981	43,95%
Néoplasie	0	399	95,91%	202	93,52%	2123	95,12%
	1	17	4,09%	14	6,48%	109	4,88%
Centre	1	129	31,01%	65	30,09%	727	32,57%
	2	264	63,46%	137	63,43%	1062	47,58%
	3	18	4,33%	11	5,09%	310	13,89%
	4	5	1,20%	3	1,39%	133	5,96%

TAB. 6.3 – Détails des effectifs par covariable des différentes transitions.

tous les individus qui n'ont pas été censurés avant le temps t_{l-1} .

L'effectif de transitions observées est simplement égal au nombre d'événements k se produisant dans l'intervalle :

$$n_{l,k} = \sum_{R(t_{l-1})} \mathbb{1}_{\{t_h \leq t_l \text{ et } X_h = k\}} \quad (6.20)$$

où $\mathbb{1}_a = 1$ si la condition a est vraie et 0 sinon.

Bootstrap

Comme nous l'avons vu plus tôt, cette méthode permet d'estimer la distribution de la statistique (6.17) sans formuler d'hypothèses asymptotiques sur celle-ci. Ce calcul se déroule en plusieurs étapes.

- ★ **Génération de B échantillons de bootstrap** chacun de taille n . Chaque individu h^* ($h^* = 1, \dots, n$) n'est pas simulé, mais il est tiré aléatoirement (avec remise) dans l'échantillon initial.
- ★ **Simulation de la trajectoire de chaque individu h^*** sous l'hypothèse nulle selon laquelle son évolution obéit au modèle multi-états estimé initialement. Tous les individus entrent dans l'état 0. A partir du paramètre estimé dans le modèle, α , le second état (étant aussi le dernier) est simulé à l'aide d'une loi binomiale de paramètres $\{P_k, k = I, D\}$.
- ★ **Simulation des temps de passage** dans les états sous l'hypothèse nulle. Comme vu dans la thèse de Foucher Y. (27), ceci est possible à partir des fonctions de répartitions inverses des lois de passages dans les différents états et d'un générateur aléatoire fournissant des réalisations d'une variable aléatoire uniforme.

Pour le modèle semi-markovien avec survie relative, la fonction de répartition inverse pour la transition vers l'état retour en dialyse est tout simplement la fonction de répartition inverse de la loi Exponentielle *i.e.* :

$$(F_{\text{Obs}}^I(u_{h^*} | Z_{h^*}^I))^{-1} = -\frac{\log(1 - u_{h^*})}{\sigma_I \exp(\hat{\beta}^I Z_{h^*}^I)}. \quad (6.21)$$

Cependant, pour le modèle sans survie relative, il s'agit de la fonction de répartition de la loi de Weibull, *i.e.* :

$$F_{\text{Obs}}^I(u_{h^*} | Z_{h^*}^I)^{-1} = \sigma_I \exp\left(\frac{1}{\nu_I} \left(\log(-\log(1 - u_{h^*})) - \hat{\beta}^I Z_{h^*}^I\right)\right). \quad (6.22)$$

Par contre, pour le modèle avec survie relative et le passage dans l'état décès, nous ne pouvons obtenir de formulation explicite pour la fonction de répartition inverse correspondante. En effet, les différentes étapes de calculs pour trouver la fonction de répartition in-

verse sont les suivantes :

$$\begin{aligned}
u_{h^*} &= F_{\text{Obs}}^D(t_{h^*} | Z_{h^*}^D) \\
u_{h^*} &= 1 - S_{\text{Obs}}^D(t_{h^*} | Z_{h^*}^D) \\
u_{h^*} &= 1 - \exp\left(-\Lambda_{\text{Obs}}^D(t_{h^*} | Z_{h^*}^D)\right) \\
u_{h^*} &= 1 - \exp\left(-\sum_{k=0}^{na_{h^*}} \frac{1}{\sigma_D^{vD}} (t_{h^*,(k+1)}^{vD} - t_{h^*,k}^{vD}) \right. \\
&\quad \left. \times \lambda_{\text{Att}}^D(t_{h^*,k} + a_{h^*0}) \exp(\beta_D Z_{h^*}^D)\right)
\end{aligned}$$

$$\log(1 - u_{h^*}) = -\frac{1}{\sigma_D^{vD}} \sum_{k=0}^{na_{h^*}} (t_{h^*,(k+1)}^{vD} - t_{h^*,k}^{vD}) \lambda_{\text{Att}}^D(t_{h^*,k} + a_{h^*0}) \exp(\beta_D Z_{h^*}^D).$$

Ainsi, pour résoudre ce problème, à partir des n variables aléatoires uniformes sur $[0, 1]$, u_{h^*} , on cherche na_{h^*} tel que :

$$-\frac{\sigma_D^{vD} \log(1 - u_{h^*})}{\exp(\hat{\beta}^D Z_{h^*}^D)} = \sum_{k=0}^{na_{h^*}} \lambda_{\text{Att}}^D(t_{h^*,k} + a_{h^*0}) (t_{h^*,(k+1)}^{vD} - t_{h^*,k}^{vD}). \quad (6.23)$$

Pour le modèle semi-markovien avec survie relative, les temps simulés pour les transitions vers l'état décès sont donc des nombres discrets. Ceci entraîne donc une perte de précision.

- ★ **Estimation du modèle** à partir de l'échantillon de bootstrap. Pour chaque individu h^* de l'échantillon de bootstrap, on dispose, grâce aux simulations précédentes, de son état final ainsi que de son temps de passage dans cet état. Si, par exemple, le temps simulé pour un individu est supérieur au nombre d'années de suivi observé maximal, alors cet individu sera censuré dans l'état 0. De plus, le temps d'observation de ce même individu correspondra alors à la différence entre la date du gel de la base et sa date de transplantation.

A partir des contributions individuelles et de la fonction de vraisemblance (6.16), le modèle correspondant à l'échantillon de bootstrap peut alors être estimé.

- ★ **Calcul de la p-value du test.** Pour chaque couple composé de l'échantillon de bootstrap et du modèle semi-markovien estimé à partir de ce dernier, la statistique de test G_b^* ($b = 1, \dots, B$) est calculée à partir des expressions (6.17) et (6.19). La p -value, c'est-à-dire la probabilité de rejeter l'hypothèse nulle alors que cette dernière est vraie, est alors facilement calculable :

$$p = B^{-1} \sum_{b=1}^B \mathbb{1}_{\{G_b^* \geq G\}} \quad (6.24)$$

où $\mathbb{1}_{\{G_b^* \geq G\}}$ vaut 1 si la statistique G_b^* est supérieure ou égale à G la statistique calculée à partir de l'échantillon initial, et 0 sinon.

Résultats

Pour chaque intervalle de temps choisi ainsi que pour chaque transition, le tableau (6.4) présente les effectifs observés et théoriques associés aux deux modèles estimés dans ce chapitre.

Temps (en années)	Transitions	Observés	Attendus	
			Sans SR	Avec SR
[0,0055; 1,4435]	$0 \rightarrow I$	99	91,268	91,235
	$0 \rightarrow D$	57	52,314	23,686
[1,4435; 4,1315]	$0 \rightarrow I$	94	110,839	110,876
	$0 \rightarrow D$	55	55,313	46,526
[4,1315; 7,3720]	$0 \rightarrow I$	116	76,417	76,355
	$0 \rightarrow D$	50	36,171	27,697
[7,3720; 16,2650]	$0 \rightarrow I$	107	62,793	62,546
	$0 \rightarrow D$	54	30,080	31,460

TAB. 6.4 – Tableau de contingence des transitions observées et attendues vers les états terminaux pour les deux modèles de semi-Markov.

Pour le modèle sans survie relative, la statistique du test vaut 79,57. Le nombre d'itérations bootstrap est 10 000. 14 statistiques de bootstrap sont supérieures ou égales à G , correspondant à une p -value de 0,0014. La table (6.5) permet d'avoir un aperçu de la distribution des statistiques de bootstrap. En considérant un risque de première espèce de 5%, le modèle semi-markovien sans survie relative ne semble pas homogène sur le temps. De plus, nous remarquons que les décès contribuent à 31,08% de la valeur de la statistique de ce test et le retour en dialyse à 68,92%. Le modèle semble donc plus adapté à la prédiction du décès.

Minimum	1er Quartile	Mediane	Moyenne	3ème Quartile	Maximum
20,65	41,10	46,65	47,55	53,09	92,18

TAB. 6.5 – Quartiles de la distribution de bootstrap de la statistique de test pour le modèle semi-markovien sans survie relative.

Pour le modèle semi-markovien avec survie relative, la statistique du test vaut 100.56. 7411 itérations bootstrap sont réalisées et la p -value du test est de 0,7411. La table (6.6) donne les quartiles de la distribution de ces statistiques. Le modèle semi-markovien avec survie relative semble donc homogène sur le temps. La prise en compte de la survie relative dans le modèle semi-markovien semble donc homogénéiser sur le temps. Par ailleurs, les décès contribuent à 82,49% de la valeur de la statistique de ce test et le retour en dialyse à 17,51%. Le modèle semble donc plus adapté à la prédiction du retour en dialyse.

Par ailleurs, nous remarquons grâce au tableau (6.4) que la principale différence entre les deux modèles semi-markovien (avec et sans survie relative) provient des effectifs attendus pour la transition vers l'état décès dans l'intervalle de temps [0,0055; 1,4435]. Ceci vient du fait que les temps simulés pour la transition vers l'état décès par le modèle semi-markovien

Minimum	1er Quartile	Mediane	Moyenne	3ème Quartile	Maximum
64,28	100,10	112,20	114,10	126,00	197,00

TAB. 6.6 – Quartiles de la distribution de bootstrap de la statistique de test pour le modèle semi-markovien avec survie relative.

avec survie relative ne sont pas aussi précis que ceux simulés par celui sans survie relative. En effet, celui sans survie relative permet de simuler des temps en jours alors que le modèle avec survie relative ne le permet pas, il simule des temps en années. De plus, cet intervalle de temps ne comprend qu'une seule année entière (*i.e.* un an entre le début et la fin de l'intervalle) contrairement aux autres intervalles.

En conclusion, le modèle semi-markovien avec survie relative semble homogène sur le temps alors que le modèle sans survie relative ne l'est pas. Dans le cas des transplantés rénaux, l'insertion de la notion de survie relative dans le modèle semi-markovien semble l'homogénéiser sur le temps.

DISCUSSION-CONCLUSION

La problématique de la prise en compte du retour en dialyse nous a conduit au choix du modèle semi-markovien.

Ce modèle a l'avantage de pouvoir modéliser toute la dynamique clinique d'une pathologie quelle qu'elle soit. Il suffit juste que les données suivent une structure identique à celle proposée (même nombre d'états et mêmes possibilités de transitions dans les différents états). Ici, nous avons choisi de modéliser le retour en dialyse comme un état à part entière et de le considérer comme absorbant, ceci car nous n'avons aucune information sur le devenir des patients qui rejettent leur greffe. Cependant, dans d'autres situations (VIH, asthme, ...), l'arrêt des consultations, le changement ou l'interruption du traitement au bout d'une certaine date peuvent aussi être considérés comme des états de censure informative. Nous proposons donc un modèle à trois états dont deux absorbants, le décès et le retour en dialyse. Comme nous analysons les greffés rénaux, l'état initial de chaque patient est considéré comme greffé en bonne santé. Le fait et le moyen d'estimer les probabilités de transition dans les états décès et retour en dialyse montrent que ces états dépendent bien l'un de l'autre. En effet, la probabilité de transiter dans un des deux états dépend de la probabilité de transiter dans l'autre. Cela montre que les risques de décès ou de retour en dialyse sont bien pris en compte dans un même modèle. Ceci est différent du fait de modéliser les deux risques par des modèles de Cox et de conclure ensuite à une dépendance entre ces risques si les mêmes covariables significatives sont obtenues dans chacun des deux modèles. Dans ce cas, la dépendance est trouvée par l'intermédiaire des covariables mais elle n'est pas prise en compte.

Par ailleurs, le modèle semi-markovien a l'avantage de modéliser la durée passée dans les états par des fonctions explicites. Dans l'application, on s'intéresse à la durée passée dans l'état "greffé en bonne santé" avant

de transiter dans l'état décès ou dans l'état retour en dialyse. Les deux fonctions pour les deux transitions possibles sont d'ailleurs différentes.

De plus, comme nous avons choisi une modélisation paramétrique, les facteurs protecteurs et de risque de décès ont pu être déterminés. Les résultats montrent que ce sont les mêmes qu'avec les autres modèles ne prenant pas en compte la censure informative. Cependant, avec ce modèle multi-états, nous ne faisons pas l'hypothèse d'indépendance entre le décès et la censure et ainsi nos estimations ne sont pas biaisées. De plus, avec ce modèle, nous obtenons aussi les facteurs explicatifs du retour en dialyse. Ce modèle apporte donc des informations complémentaires par rapport aux modèles que nous avons fait précédemment.

Par ailleurs, l'addition de la survie relative dans le modèle semi-markovien, qui est une nouveauté par rapport à ce genre de modèle réputé, marque encore une fois ses avantages. Les facteurs explicatifs du décès peuvent être analysés plus précisément que dans un modèle sans survie relative, *i.e* ils peuvent être interprétés pour le risque de décès observé (toutes pathologies confondues) et pour le risque de décès spécifique à la pathologie étudiée. En effet, nous pouvons apprécier les différents effets des covariables soit au travers du risque de décès relatif à la pathologie étudiée soit au travers du risque observé du groupe de patients étudiés. Le premier met en valeur le risque de décès propre à la pathologie étudiée tandis que le second estime le risque de décès des patients subissant la pathologie étudiée mais il tient compte des autres causes de décès possibles (accidents de la route, décès naturels, ...). Naturellement, l'attention des médecins porte plutôt sur la connaissance du risque spécifique à la pathologie étudiée. Dans ce cas, l'interprétation du risque de décès relatif à la maladie est préféré. Dans l'application du modèle proposé à la base DIVAT, nous présentons les résultats concernant les deux types de risque : relatif et observé. Pour l'analyse du risque observé, nous avons choisi de présenter les résultats pour certains âges et certaines années. Cependant, selon les besoins de l'étude, d'autres résultats sur différents âges et années peuvent être fournis.

D'autre part, dans notre cas, il semblerait que l'adjonction de la survie relative homogénéise le modèle semi-markovien sur le temps. Ceci peut s'expliquer par le fait que les tables vitales sont réalisées par année calendaire, sexe, âge d'où la prise en compte du temps chronologique.

Cependant, un des résultats de notre analyse reste quelque peu curieux, il s'agit de la distribution concernant la fonction de risque pour la transition dans l'état retour en dialyse. En effet, dans le modèle sans survie relative, il s'agirait d'une distribution de Weibull alors que dans le modèle avec survie relative d'une distribution Exponentielle. Nous avons procédé à des tests de rapport de vraisemblances pour choisir entre les distributions de Weibull et Exponentielle. La question se pose de savoir si cela est dû au choix du risque de première espèce (5%) ou si c'est un effet dû à l'ajout de la survie relative.

En outre, nous supposons qu'il existe un lien entre le risque de retour

en dialyse et le risque de décès, ceci par le biais de covariables. Dans notre base de données, il semble que les patients perdus de vue apportent de l'information sur le décès par l'intermédiaire de la covariable année de la transplantation. L'inconvénient est que nous ne pouvons pas quantifier ce lien. En effet, on se doute que si tous les facteurs sont identiques pour la censure informative et le décès alors le lien est fort et dans le cas inverse il y a peu de chance d'avoir un lien. Il serait donc très intéressant de pouvoir définir ce lien mais cela nous semble possible seulement dans le cas où les informations concernant les patients qui retournent en dialyse sont connues.

DISCUSSION-CONCLUSION- PERSPECTIVES

Cette thèse a abordé trois grandes problématiques souvent rencontrées en analyse de survie. Les modèles proposés pour résoudre ces problématiques sont appliqués aux transplantés rénaux de la base DIVAT.

La première problématique concerne le cas de la dépendance non mesurable entre certains individus. Ces individus partagent, par exemple, des facteurs environnementaux ou génétiques communs qui ne sont pas observables. La solution pour résoudre un tel problème est d'inclure une variable aléatoire, appelée la fragilité, dans les modèles de survie. Le chapitre 3 introduit les modèles de fragilité. Il existe essentiellement deux types de modèles de fragilité : les univariés qui permettent de modéliser l'effet de covariables non observées et les multivariés qui permettent de modéliser la dépendance non mesurable entre les individus issus de même groupe. Il existe une multitude de modèles de fragilité s'adaptant à toutes sortes de problématique de dépendance. La théorie propre à ces modèles est présentée dans ce chapitre. Ensuite, un modèle est proposé et appliqué à la base de données DIVAT. En effet, dans cette base, les statistiques descriptives et les premières analyses de survie nous ont conduit à penser que les survies des patients transplantés sont différentes selon les centres et les années de transplantation. Le modèle proposé est un modèle de fragilité partagée et il est appliqué deux fois à la base de données : la première permet de prendre en compte la dépendance des patients issus du même centre et la seconde la dépendance des patients greffés la même année. Le modèle suppose donc, que, sachant la fragilité, c'est-à-dire la connaissance du centre de transplantation (respectivement la connaissance de l'année de transplantation), les temps de survie des patients greffés dans le même centre (respectivement la même année) sont indépendants. Les résultats de ces modélisations confirment l'effet année de transplantation mais pas celui concernant le centre de transplantation. Ainsi, les patients greffés la même année semblent liés par un facteur non mesurable provoqué par l'année de transplantation. Il peut s'agir, par exemple, des stratégies de transplantation. En effet, nous savons qu'elles ont évolué au cours du temps mais nous ne pouvons réellement quantifier ces changements.

Dans ce modèle, nous avons choisi une distribution Gamma pour la variable de fragilité. Ceci car cette distribution possède de bonnes propriétés mathématiques et que les calculs théoriques engendrés par la fragilité dans les modèles de survie sont facilités par les transformées de Laplace de cette loi. De plus, comme souvent effectué pour éviter le problème de surparamétrisation, nous prenons les paramètres d'échelle et de forme de

cette loi identiques. Cependant, bien que la distribution Gamma soit réputée dans la littérature, il serait intéressant de tenter d'autres distributions dans notre cas et d'établir des comparaisons. Hougaard (37) a effectué une étude détaillée des propriétés des modèles de chaque famille de distributions (Gamma, Positive Stable, PVF, ...). Il a établi essentiellement trois sortes de comparaisons entre ces modèles : comparaison théorique, comparaison de l'ajustement et comparaison des différentes mesures de dépendance. De manière brève, il ressort de ces analyses, que le choix d'une distribution pour la fragilité est assez délicat et dépend du problème étudié. Même si chaque distribution a ses avantages, le choix d'une distribution Gamma se révèle correct et c'est pourquoi cette distribution est largement utilisée dans les études.

Par ailleurs, le modèle proposé suppose une fragilité à une dimension. Néanmoins, cette fragilité peut seulement induire une dépendance positive à l'intérieur des groupes. Dans certaines situations, les temps de survie pour les individus appartenant à un même groupe sont négativement associés à une covariable (ils ne partagent pas la même fragilité). Dans le cas des transplantés rénaux, plus un patient doit attendre un rein disponible pour la greffe, plus sa probabilité de rejeter son greffon est importante. Ceci implique, d'après Wolfe *et al.*, un temps de survie après la transplantation plus faible. Dans ce cas, les temps d'attente et les temps de survie après la transplantation peuvent être négativement associés. Il serait donc pertinent d'estimer un modèle de fragilité à plusieurs dimensions otant la restriction du type de corrélation entre les temps de survie des différents patients.

De plus, notre modèle suppose une fragilité constante au cours du temps, cela signifie que la dépendance entre les individus issus d'un même groupe est la même au cours du temps. Cependant, il se pourrait que la dépendance diminue au cours du temps. En effet, plus les patients survivent après la greffe plus la dépendance à laquelle ils étaient soumis lors de leur transplantation s'amointrit. Si, par exemple, les transplantés rénaux sont dépendants à cause de facteurs environnementaux ou stratégiques liés à l'année, alors, plus ces patients survivent après la greffe, plus ils ont fait face à leurs facteurs de dépendance. Ainsi, leur dépendance due à ces facteurs a tendance à diminuer avec le temps. Un modèle de fragilité avec fragilité dépendant du temps peut donc s'avérer pertinent (dans ce cas, la fragilité devient un processus).

La deuxième problématique souvent rencontrée en analyse de survie concerne la méconnaissance des causes de décès des patients. Ceci soit parce que les certificats de décès ne sont pas complets soit parce que les autopsies ne sont pas toujours effectuées. Cependant, même lorsque la cause de décès d'un patient est connue, il est difficile de classer son décès. Ceci est le cas dans la base des transplantés rénaux. Les causes de décès ne sont pas renseignées, mais, même si elles l'étaient, il serait difficile d'attribuer, ou non, un décès à la transplantation. En effet, les décès des transplantés dus à un cancer peuvent être indépendants de la transplantation, mais, ils peuvent aussi être dus aux traitements immunosuppresseurs qui leur sont administrés. Il est donc difficile de classer les décès dans ces cas-là. Pour résoudre ce problème, nous avons eu re-

cours aux modèles de survie relative. Ce genre de modèle a l'avantage de pouvoir estimer une survie propre à la pathologie étudiée sans nécessité de connaître l'information concernant les causes de décès des patients. Il existe essentiellement deux types de modèles de survie relative : les additifs et les multiplicatifs. Cependant, nous nous sommes plus particulièrement intéressés aux modèles de type multiplicatif. En effet, une limite des modèles additifs est que la survie observée du groupe de patients étudiés doit être inférieure à celle observée dans la population générale c'est-à-dire toutes les pathologies confondues. Dans la base DIVAT, ceci est vérifié jusqu'à cinq années après une greffe, mais au-delà, la survie observée des transplantés rénaux est supérieure à celle de la population nationale française. A partir de ces remarques, deux choix se sont présentés dans notre cas : le premier consiste à utiliser les modèles additifs sur les cinq premières années de suivi (sachant que le recul de la base est d'environ 17 ans) et le deuxième est l'utilisation des modèles multiplicatifs. Dans le but d'utiliser toute l'information disponible, l'analyse de la survie sur un recul de 17 ans nous a semblé préférable à l'analyse de la survie sur les cinq premières années, nous avons donc opté pour les modèles multiplicatifs. Par ailleurs, les conclusions qu'apportent le modèle multiplicatif de survie relative sont assez denses et sont à analyser avec précaution. Tout d'abord, les covariables liées à la survie de la population de référence, c'est-à-dire l'âge et le sexe des patients ainsi que l'année, et qui sont aussi des covariables du risque relatif à la pathologie étudiée, ne sont plus considérées comme des covariables classiques de l'analyse de survie. Comme ces covariables interviennent de deux manières dans le modèle, explicitement dans le risque relatif à la pathologie et implicitement dans le risque attendu de la population de référence, elles sont plutôt considérées comme des covariables d'ajustement. De plus, les modèles de survie relative permettent d'interpréter deux types de risque : le risque observé de la population étudiée (risque brut) et le risque relatif à la pathologie de cette population (risque net). Le premier comprend tous les autres risques de décès auxquels la population étudiée est exposée alors que le second ne reflète que le risque de décès lié à la pathologie en question. Le choix de l'interprétation plus approfondie de l'un ou de l'autre dépend de la personne concernée par les résultats. Nous supposons que s'il s'agit d'un médecin, il sera plus intéressé par le risque relatif à la pathologie étudiée. L'application du modèle multiplicatif de survie relative aux transplantés rénaux montre que le délai de reprise de fonction du greffon (DGF) est un facteur spécifique du décès après une transplantation rénale. Les patients dont le greffon met plus de six jours à reprendre fonction sont plus à risque de décéder que ceux dont le greffon met moins de six jours. Ce facteur ressort aussi significatif dans les analyses de survie classique mais, grâce au modèle de survie relative, nous savons qu'il est réellement spécifique du décès après une transplantation rénale.

Dans le modèle proposé, une modélisation paramétrique pour le risque de base spécifique à la pathologie est choisie. Nous optons alors pour calculer le risque cumulé observé de l'échantillon étudié grâce à la somme des aires sous la courbe de la fonction de risque relatif de base. Cependant, cette estimation du risque cumulé relatif de base peut manquer de

précision dans le sens où il s'agit d'une estimation annuelle. En effet, il y a une différence entre, par exemple, prédire 2,1 et 2,8 années de survie. Il serait donc intéressant d'évaluer cette possible perte de précision en estimant d'une autre manière le risque cumulé observé de l'échantillon étudié.

Par ailleurs, même si les modèles de survie relative datent de 1940 environ, ils n'ont pas été exploités dans toutes les problématiques qu'engendrent l'analyse de survie. Étant donnée la multitude de modèles en analyse de survie, il serait intéressant d'introduire cette notion de survie relative dans d'autres modèles de survie. Cette thèse propose donc de résoudre certaines problématiques connues de l'analyse de survie en les combinant avec la survie relative.

Ainsi, nous avons choisi de mélanger les deux problématiques précédentes dans un même modèle, c'est-à-dire prendre en considération le fait que nous ne connaissons pas les causes des décès des patients transplantés rénaux et que, de plus, ces patients peuvent être liés entre eux par des facteurs non observables. Le chapitre 5 concerne donc le modèle multiplicatif de survie relative avec fragilité. Nous reprenons donc la théorie du modèle multiplicatif de survie relative proposé au chapitre 4 et nous ajoutons un terme aléatoire, la fragilité, comme facteur multiplicatif de la fonction de risque observé de la population étudiée. Le terme de fragilité, comme au chapitre 3, permet, dans un premier modèle, de prendre en considération la dépendance des patients issus du même centre transplantation et, dans un second modèle, la dépendance des patients greffés la même année. Les résultats de ces deux modélisations appliquées aux cas des transplantés rénaux apportent une nouvelle conclusion par rapport aux modèles de fragilité sans survie relative. Tout d'abord, comme au chapitre 3, l'effet centre de transplantation n'est pas confirmé. Ensuite, la deuxième modélisation ne valide pas l'effet année de transplantation. Cette conclusion diffère de celle du modèle estimé au chapitre 3. Ainsi, les patients transplantés rénaux ne semblent pas liés par un facteur caché de l'année de transplantation lorsque les risques attendus de la population nationale française sont considérés dans le modèle, et donc, implicitement l'année de la greffe. À la fin de ce chapitre, nous nous sommes donc posé la question du choix de modèle à adapter : le modèle avec fragilité portant sur l'année de la greffe ou le modèle multiplicatif de survie relative. Ce choix n'est pas évident car les deux modèles prennent en considération l'effet année de transplantation. Le modèle de survie relative peut être remis en cause par la façon dont il estime le risque relatif cumulé observé mais le modèle de fragilité nécessite l'estimation d'un paramètre supplémentaire. Seul le fait que l'insertion d'un paramètre supplémentaire dans un modèle peut engendrer des erreurs d'estimation, nous fait opter pour le modèle de survie relative.

La troisième problématique étudiée dans cette thèse est celle concernant la censure. Il existe essentiellement deux sortes de censure. La première, la censure administrative, est engendrée par le gel de la base. Elle n'est pas considérée comme informative, c'est-à-dire qu'elle n'apporte pas d'information sur l'évènement étudié. Par contre, la deuxième

censure rencontrée dans la base DIVAT, ne peut être considérée comme non informative. En effet, elle est engendrée par la perte du greffon des patients. Quand un patient est transplanté, il se peut qu'il rejette son greffon et, dans ce cas, il retourne en dialyse et change de centre de soins. Il n'est donc plus suivi dans le même centre que celui dans lequel il a été transplanté. De ce fait, l'information concernant ce patient est perdue de la base de données. Nous sommes donc dans un cas de censure. Cependant, cette censure est considérée comme informative car elle ne peut être supposée indépendante de l'évènement étudié. Ceci est appuyé par l'étude de Wolfe *et al.* (60) qui montre un risque de décès plus important chez les patients transplantés rénaux qui retournent en dialyse. Dans cette thèse, pour résoudre ce problème de censure informative, nous avons opté pour un modèle multi-états. Tout d'abord parce que ce type de modèle permet de prendre en considération toute la dynamique clinique d'une pathologie, ensuite, parce que les fonctions de risque des différents évènements peuvent être modélisées par des distributions différentes. Par ailleurs, nous avons choisi un modèle à trois états. Le premier correspond à l'état bonne santé après la greffe. Les deux autres états sont le décès et le retour en dialyse, ils sont considérés comme des états absorbants (une fois entré dans ces états, un individu ne peut en sortir). Comme la transition de l'état retour en dialyse vers l'état décès n'est jamais observé dans cette base, il n'y a donc que deux probabilités de transition et deux fonctions de risques à estimer (de l'état "greffé en bonne santé" vers l'état "retour en dialyse" ou de l'état "greffé en bonne santé" vers l'état "décès"). A des fins de comparaison, nous avons estimés deux types de modèles multi-états : un avec survie relative et un sans survie relative. Les résultats de ces deux modèles donnent les facteurs explicatifs d'un décès après une transplantation mais également ceux d'un retour en dialyse. Le lien entre le décès et le retour en dialyse se déduit donc des covariables. De plus, les résultats des deux modèles multi-états (avec et sans survie relative) confirment qu'avoir un greffon qui met plus de six jours à reprendre fonction est un facteur de risque de décès. Par ailleurs, nous avons voulu tester l'hypothèse d'homogénéité de ces modèles sur le temps par une adaptation du test d'adéquation de Aguirre-Hernandez et Farewell (4). Au vu des résultats des tests (un test pour chaque type de modèle multi-états), il semble que le modèle multi-états avec survie relative soit homogène sur le temps alors que celui sans survie relative ne le soit pas. L'adjonction de la survie relative semble homogénéiser le modèle multi-états sur le temps. Ceci peut s'expliquer par le fait que les tables vitales sont réalisées par année calendaire, sexe, âge d'où la prise en compte du temps chronologique.

Une limite de notre modèle est qu'il ne peut pas réellement quantifier le lien entre la censure et le décès. Ce lien est seulement déduit par l'intermédiaire des covariables. En effet, nous supposons que si toutes les covariables explicatives d'un décès sont aussi les covariables explicatives de la censure (dans notre cas, le retour en dialyse) alors le lien entre la censure et le décès est fort. Par contre, si aucune covariable significative n'est identique pour les deux évènements, alors nous concluons que la censure n'apporte pas d'information sur le décès. Dans l'application aux transplantés rénaux, il est assez difficile de conclure. En effet, il n'y a

qu'une seule covariable significative pour le décès et pour le retour en dialyse, il s'agit de l'année de la greffe (ceci pour les deux types de modèles multi-états). Nous sommes donc limités à conclure à un lien entre le retour en dialyse et l'évènement décès mais nous ne pouvons donner une mesure de ce lien.

Par ailleurs, pour résoudre ce problème de censure informative, d'autres méthodes existent. Une, en particulier, peut être appliquée à nos données. Il s'agit de la méthode IPCW (Inverse Probability Censoring Weight) introduite par Robins (53), puis utilisée notamment par Robins *et al.* (54) ainsi que Castelli *et al.* (17). Le principe de cette méthode est d'étudier le risque de censure en fonction des covariables afin de déduire des poids spécifiques à chaque individu. Les poids sont ensuite intégrés dans les estimateurs classiques (Kaplan-Meier) pour modifier le nombre de personnes à risque afin de prendre en compte les personnes censurées. Ainsi, le phénomène de censure informative est pris en compte.

D'un point de vue clinique, nous remarquons que, dans tous les modèles proposés, le temps de retard au démarrage des greffons (DGF) est un facteur explicatif de décès après une transplantation de rein. Dans ce mémoire, nous l'avons étudié de façon catégorielle, c'est-à-dire qu'il vaut 1 s'il est supérieur à 6 jours et 0 sinon. Cependant, il serait sans doute intéressant d'exploiter l'effet de ce facteur de manière continue, c'est-à-dire en jours, dans les modèles proposés.

ANNEXES

A

SOMMAIRE

A.1 DÉMONSTRATIONS AVEC LA FRAGILITÉ	125
A.2 TABLEAU D'EFFECTIFS SELON LES ANNÉES DE GREFFE	130

A.1 DÉMONSTRATIONS AVEC LA FRAGILITÉ

$$\star S(t_{ij}) = E [S(t_{ij}|\mathbf{y}_i)] = L[\Lambda(t_{ij})].$$

La survie individuelle peut s'écrire :

$$\begin{aligned} S(t_{ij}) &= \int_0^\infty S(t_{ij}|Y = u)g(u)du \quad (g \text{ est la densité de } Y) \\ &= E [S(t_{ij}|u)] \quad (\text{l'espérance porte sur } y) \\ &= E [\exp(-Y(\Lambda(t_{ij})))]. \end{aligned}$$

La transformée de Laplace d'une variable U est

$$L(a) = \int_0^\infty \exp(-au)g(u)du.$$

D'où :

$$S(t_{ij}) = L[\Lambda(t_{ij})]. \quad \blacksquare$$

$$\star S(t_{i1}, \dots, t_{in_i}) = L \left[\sum_{j=1}^{n_i} L^{-1}(S(t_{ij})) \right] ..$$

Commençons par démontrer ceci :

$$\begin{aligned} S(t_{i1}, \dots, t_{in_i}|Y) &= \exp[-Y(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i}))] \\ \text{avec } \Lambda(t_{ij}) &= \int_0^{t_{ij}} \lambda(u)du, \quad j = 1, \dots, n_i. \end{aligned}$$

Conditionnellement à la variable de fragilité, la survie multivariée vaut :

$$\begin{aligned} S(t_{i1}, \dots, t_{in_i}|Y) &= P(T_{i1} > t_{i1}, \dots, T_{in_i} > t_{in_i}|Y) \\ &= P(T_{i1} > t_{i1}|Y) \dots P(T_{in_i} > t_{in_i}|Y) \\ &= S(t_{i1}|Y) \dots S(t_{in_i}|Y) \\ &= \exp \left(- \int_0^{t_{i1}} \lambda(u|Y)du - \dots - \int_0^{t_{in_i}} \lambda(u|Y)du \right) \\ &= \exp \left(- \int_0^{t_{i1}} Y(\omega)\lambda(u)du - \dots - \int_0^{t_{in_i}} Y(\omega)\lambda(u)du \right) \\ &= \exp[-Y(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i}))]. \end{aligned}$$

Ainsi :

$$\begin{aligned} S(t_{i1}, \dots, t_{in_i}) &= E [S(t_{i1}, \dots, t_{in_i}|Y)] \\ &= E [\exp[-Y(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i}))]] \\ &= L [\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i})]. \end{aligned}$$

On a :

$$S(t_{ij}) = L[\Lambda(t_{ij})].$$

Ainsi :

$$\Lambda(t_{ij}) = L^{-1}[S(t_{ij})].$$

Donc :

$$S(t_{i1}, \dots, t_{in_i}) = L \left[\sum_{j=1}^{n_i} L^{-1}(S(t_{ij})) \right]. \quad \blacksquare$$

$$\star f(t_{i1}, \dots, t_{in_i}) = (-1)^{n_i} \left(\prod_{j=1}^{n_i} \lambda(t_{ij}) \right) L^{(n_i)}(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i})).$$

La densité multivariée pour le groupe i vaut :

$$\begin{aligned} f(t_{i1}, \dots, t_{in_i}) &= (-1)^{n_i} \frac{\partial^{n_i}}{\partial t_{i1} \dots \partial t_{in_i}} S(t_{i1}, \dots, t_{in_i}) \\ &= (-1)^{n_i} \frac{\partial^{n_i}}{\partial t_{i1} \dots \partial t_{in_i}} \int_0^\infty S(t_{i1}, \dots, t_{in_i} | Y = u) g(u) du \\ &= (-1)^{n_i} \frac{\partial^{n_i}}{\partial t_{i1} \dots \partial t_{in_i}} \int_0^\infty S(t_{i1} | Y = u) \dots S(t_{in_i} | Y = u) g(u) du \\ &= (-1)^{n_i} \frac{\partial^{n_i}}{\partial t_{i1} \dots \partial t_{in_i}} \int_0^\infty \left(- \int_{t_{i1}}^\infty f(t_{i1} = v_{i1} | Y = u) dv_{i1} \right) \\ &\quad \dots \left(- \int_{t_{in_i}}^\infty f(t_{in_i} = v_{in_i} | Y = u) dv_{in_i} \right) g(u) du \\ &= (-1)^{n_i} \frac{\partial^{n_i}}{\partial t_{i1} \dots \partial t_{in_i}} (-1)^{n_i} \int_{t_{i1}}^\infty \dots \int_{t_{in_i}}^\infty \int_0^\infty f(t_{i1} = v_{i1} | Y = u) \\ &\quad \dots f(t_{in_i} = v_{in_i} | Y = u) g(u) dv_{i1} \dots dv_{in_i} du \\ &= (-1)^{n_i} \int_0^\infty (-1)^{n_i} f(t_{i1} = v_{i1} | Y = u) \dots f(t_{in_i} = v_{in_i} | Y = u) g(u) du \\ &= (-1)^{n_i} \int_0^\infty (-1)^{n_i} \lambda(t_{i1} = v_{i1} | Y = u) S(t_{i1} = v_{i1} | Y = u) \\ &\quad \dots \lambda(t_{in_i} = v_{in_i} | Y = u) S(t_{in_i} = v_{in_i} | Y = u) g(u) du \\ &= (-1)^{n_i} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) \right\} \int_0^\infty (-1)^{n_i} Y^{n_i} S(t_{i1}, \dots, t_{in_i} | Y = u) g(u) du \\ &= (-1)^{n_i} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) \right\} \int_0^\infty (-1)^{n_i} Y^{n_i} \exp(-Y \Lambda(t_{i1}, \dots, t_{in_i})) g(u) du \\ &= (-1)^{n_i} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) \right\} E[(-1)^{n_i} Y^{n_i} \exp(-Y \Lambda(t_{i1}, \dots, t_{in_i}))] \\ &= (-1)^{n_i} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) \right\} L^{(n_i)}(\Lambda(t_{i1}) + \dots + \Lambda(t_{in_i})). \quad \blacksquare \end{aligned}$$

$$\star (-1)^{\delta_i} \left(\prod_{j=1}^{n_i} \lambda(t_{ij})^{(\delta_{ij})} \right) L^{(\delta_i)} (\Lambda(t_{i1}) + \cdots + \Lambda(t_{in_i})).$$

Pour la preuve, nous nous plaçons dans le cas bivarié :

Soit $U_j = \inf(T_j, C_j)$, $j = 1, 2$. U est le temps observé, T le temps de l'évènement et C le temps de censure. Il y a donc quatre possibilités :

Cas 1 : **Si** $U_1 = C_1$ et $U_2 = C_2$, ie nous observons deux censures et $T_1 > C_1$ et $T_2 > C_2$. **Alors** la contribution à la vraisemblance est $P(T_1 > t_1, T_2 > t_2) = S(t_1, t_2)$.

Cas 2 : **Si** $U_1 = C_1$ et $U_2 = T_2$, ie nous observons une censure pour l'individu 1 et un évènement pour l'individu 2 et que $T_1 > C_1$ et $T_2 < C_2$. **Alors** :

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{P(T_1 > t_1, T_2 > t_2) - P(T_1 > t_1, T_2 > t_2 + h)}{h} \\ &= \lim_{h \rightarrow 0} \frac{P(T_1 > t_1, t_2 < T_2 < t_2 + h)}{h} \\ &= \lim_{h \rightarrow 0} \frac{F_{\text{Obs } t_1}(t_2 + h) - F_{t_1}(t_2)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \int_{t_1}^{\infty} \left(\int_{t_2}^{t_2+h} f(u, v) dv \right) du \\ &= \lim_{h \rightarrow 0} \int_{t_1}^{\infty} \frac{1}{h} \left(\int_{t_2}^{t_2+h} f(u, v) dv \right) du \\ &= \int_{t_1}^{\infty} f(u, t_2) du \\ &= - \frac{\partial S(t_1, t_2)}{\partial t_2} \end{aligned}$$

Dans ce cas, la contribution à la vraisemblance est $-\frac{\partial S(t_1, t_2)}{\partial t_2}$. (En fait, nous dérivons la survie bivariée par rapport au temps où il y a un évènement).

Cas 3 : **Si** $U_1 = T_1$ et $U_2 = C_2$, ie nous observons un évènement pour l'individu 1 et une censure pour l'individu 2 et $T_1 < C_1$ et $T_2 > C_2$. **Alors**, de façon analogue au cas précédent, la contribution à la vraisemblance est $-\frac{\partial S(t_1, t_2)}{\partial t_1}$.

Cas 4 : **Si** $U_1 = T_1$ et $U_2 = T_2$, ie nous observons deux évènements et $T_1 < C_1$ et $T_2 < C_2$. **Alors** la contribution à la vraisemblance est $(-1)^2 \frac{\partial^2 S(t_1, t_2)}{\partial t_2 \partial t_1}$. ■

$$\star S(t_1, t_2 | Y) = S_1(t_1)^Y S_2(t_2)^Y.$$

Les survies des individus 1 et 2 d'un même groupe sont indé-

pendantes sachant la fragilité de ce groupe.

$$\begin{aligned}
 S(t_1, t_2|Y) &= P(T_1 > t_1, T_2 > t_2|Y) \\
 &= P(T_1 > t_1|Y)P(T_2 > t_2|Y) \\
 &= S(t_1|Y)S(t_2|Y) \\
 &= \exp\left(-\int_0^{t_1} \lambda(u|Y)du\right) \exp\left(-\int_0^{t_2} \lambda(u|Y)du\right) \\
 &= \exp\left(-Y \int_0^{t_1} \lambda(u)du\right) \exp\left(-Y \int_0^{t_2} \lambda(u)du\right) \\
 &= \exp\left(-\int_0^{t_1} \lambda(u)du\right)^Y \exp\left(-\int_0^{t_2} \lambda(u)du\right)^Y \\
 &= S_1(t_1)^Y S_2(t_2)^Y. \quad \blacksquare
 \end{aligned}$$

$$\star S(t_1, t_2) = (S_1(t_1)^{-\sigma^2} + S_2(t_2)^{-\sigma^2} - 1)^{-1/\sigma^2}.$$

Y suit une distribution Gamma de moyenne 1 et de variance σ^2 , nous choisissons, par commodité, de poser $\sigma^2 = 1/\gamma$, dans ce cas, la fragilité suit une Gamma(γ, γ). La survie bvariée d'une telle distribution vaut :

$$S(t_1, t_2) = \frac{\gamma^\gamma}{(\gamma + M_1(t_1) + M_2(t_2))^\gamma}.$$

En effet : $S(t_1, t_2) = L(M_1(t_1) + M_2(t_2))$ et la p -ème transformée de Laplace d'une distribution Gamma(γ, γ) est :

$$L^{(p)}(s) = (-1)^p \gamma^\gamma (\gamma + s)^{-(\gamma+p)} \frac{\Gamma(\gamma + p)}{\Gamma(\gamma)}.$$

Ici $p = 0$ donc $L(s) = \gamma^\gamma (\gamma + s)^{-\gamma}$.

Par ailleurs, la survie marginale pour l'individu 1 s'écrit :

$$\begin{aligned}
 S_1(t_1) &= S(t_1, 0) \\
 &= \lim_{t_2 \rightarrow 0} \gamma^\gamma (\gamma + M_1(t_1) + M_2(t_2))^{-\gamma} \\
 &= \gamma^\gamma (\gamma + M_1(t_1))^{-\gamma}.
 \end{aligned}$$

En effet : $M_2(0) = 0$.

Et donc,

$$\begin{aligned}
 S_1(t_1)^{-\frac{1}{\gamma}} &= \gamma^{-1} (\gamma + M_1(t_1)) \\
 M_1(t_1) &= \left(\frac{1}{\gamma}\right)^{-1} S_1(t_1)^{-\frac{1}{\gamma}} - \gamma \\
 &= \gamma \left\{ S_1(t_1)^{-\frac{1}{\gamma}} - 1 \right\}.
 \end{aligned}$$

La fonction de risque marginale cumulée pour l'individu 2 se calcule de la même façon. Une expression alternative de la survie bvariée

basée sur les lois marginales :

$$\begin{aligned}
S(t_1, t_2) &= L(M_1(t_1) + M_2(t_2)) \\
&= L\left(L^{-1}(S_1(t_1)) + L^{-1}(S_2(t_2))\right) \\
&= L\left(\gamma \left\{S_1(t_1)^{-\frac{1}{\gamma}} - 1\right\} + \gamma \left\{S_2(t_2)^{-\frac{1}{\gamma}} - 1\right\}\right) \\
&= \gamma^\gamma \left[\gamma + \gamma \left\{S_1(t_1)^{-\frac{1}{\gamma}} - 1\right\} + \gamma \left\{S_2(t_2)^{-\frac{1}{\gamma}} - 1\right\}\right]^{-\gamma} \\
&= \gamma^\gamma \left[\gamma S_1(t_1)^{-\frac{1}{\gamma}} + \gamma S_2(t_2)^{-\frac{1}{\gamma}} - \gamma\right]^{-\gamma} \\
&= \gamma^{\gamma-\gamma} \left[S_1(t_1)^{-\frac{1}{\gamma}} + S_2(t_2)^{-\frac{1}{\gamma}} - 1\right]^{-\gamma} . \\
&= \left[S_1(t_1)^{-\frac{1}{\gamma}} + S_2(t_2)^{-\frac{1}{\gamma}} - 1\right]^{-\gamma} .
\end{aligned}$$

D'où, la survie bivariée

$$S(t_1, t_2) = (S_1(t_1)^{-\sigma^2} + S_2(t_2)^{-\sigma^2} - 1)^{-1/\sigma^2}. \quad \blacksquare$$

A.2 TABLEAU D'EFFECTIFS SELON LES ANNÉES DE GREFFE

Traitement				Anti HLA		Incomp		Age Donneur		Sexe Donneur		DGF		Age Receveur		Sexe Receveur		Cov	Année
3	2	1	0	≥ 0	< 0	≥ 4	< 4	≥ 55	< 55	F	H	≥ 6	< 6	≥ 55	< 55	F	H		
0	0	172	0	57	115	54	118	21	151	34	138	49	123	45	127	66	106	172	1990
0.00	0.00	100.00	0.00	33.14	66.86	31.40	68.61	12.21	87.79	19.77	80.23	28.49	71.51	26.16	73.84	38.37	61.63	%	
4	5	145	0	37	117	53	101	13	141	49	105	59	95	47	107	59	95	154	1991
2.60	3.25	94.16	0.00	24.03	75.97	34.42	65.58	08.44	91.56	31.82	68.18	38.31	61.69	30.52	69.48	38.31	61.69	%	
0	5	148	0	32	121	56	97	18	135	44	109	52	101	48	105	53	100	153	1992
0.00	3.27	96.73	0.00	20.92	79.08	36.60	63.40	11.77	88.24	28.76	71.24	33.99	66.01	31.37	68.63	34.64	65.36	%	
0	19	120	2	34	107	55	86	15	126	47	94	60	81	52	89	44	97	141	1993
0.00	13.48	85.11	1.42	24.11	75.89	39.01	60.99	10.64	89.36	33.33	66.67	42.55	57.45	36.88	63.12	31.21	68.79	%	
6	22	82	1	29	82	65	46	15	96	36	75	46	65	35	76	39	72	111	1994
5.41	19.82	73.87	0.90	26.13	73.87	58.56	41.44	13.51	86.49	32.43	67.57	41.44	58.56	31.53	68.47	35.14	64.86	%	
26	33	75	2	44	92	51	85	23	113	48	88	54	82	46	90	54	82	136	1995
19.12	24.26	55.15	1.47	32.35	67.65	37.50	62.50	16.91	83.09	35.29	64.71	39.71	60.29	33.82	66.18	39.71	60.29	%	
10	69	68	0	28	119	63	84	25	122	52	95	60	87	46	101	56	91	147	1996
6.80	46.94	46.26	0.00	19.05	80.95	42.86	57.14	17.01	82.99	35.37	64.63	40.82	59.18	31.29	68.71	38.10	61.90	%	
0	69	86	0	27	128	63	92	27	128	44	111	46	109	52	103	58	97	155	1997
0.00	44.52	55.48	0.00	17.42	82.58	40.65	59.35	17.42	82.58	28.39	71.61	29.68	70.32	33.55	66.45	37.42	62.58	%	
7	59	117	9	35	157	65	127	35	157	74	118	60	132	44	148	72	120	192	1998
3.65	30.73	60.94	4.69	18.23	81.77	33.85	66.15	18.23	81.77	38.54	61.46	31.25	68.75	22.92	77.08	37.50	62.5	%	
23	55	73	0	20	131	61	90	37	114	54	97	60	91	40	111	66	85	151	1999
15.23	36.42	48.34	0.00	13.25	86.75	40.40	59.60	24.50	75.50	35.76	64.24	39.74	60.27	26.49	73.51	43.71	56.29	%	
66	45	87	3	40	161	79	122	49	152	74	127	67	134	61	140	80	121	201	2000
32.84	22.89	43.28	1.49	19.90	80.10	39.30	60.70	24.38	75.62	36.82	63.18	33.33	66.67	30.35	69.65	39.80	60.20	%	
95	37	52	0	25	159	61	123	55	129	85	99	67	117	63	121	77	107	184	2001
51.63	20.11	28.26	0.00	13.59	86.41	33.15	66.85	29.89	70.11	46.20	53.80	36.41	63.59	34.24	65.76	41.85	58.15	%	
109	38	110	1	33	225	96	162	68	190	86	172	84	174	83	175	94	164	258	2002
42.25	14.73	42.64	0.39	12.79	87.21	37.21	62.79	26.36	73.64	33.33	66.67	32.56	67.44	32.17	67.83	36.43	63.57	%	
87	41	81	3	15	197	106	106	70	142	99	113	68	144	91	121	72	140	212	2003
41.04	19.34	38.21	1.42	7.08	92.92	50.00	50.00	33.02	66.98	46.70	53.30	32.08	67.93	42.92	57.08	33.96	66.04	%	
92	27	89	2	22	188	115	95	101	109	80	130	68	142	94	116	73	137	210	2004
43.81	12.86	42.38	0.95	10.48	89.52	54.76	45.24	48.10	51.91	38.10	61.91	32.38	67.62	44.76	55.24	34.76	65.24	%	
103	36	54	1	24	170	87	107	92	102	76	118	58	136	94	100	90	104	194	2005
53.09	18.56	27.84	0.52	12.37	87.63	44.85	55.15	47.42	52.58	39.18	60.82	29.90	70.10	48.45	51.55	46.39	53.61	%	
74	6	13	0	11	82	40	53	39	54	42	51	34	59	33	60	29	64	93	2006
79.57	6.45	13.98	0.00	11.83	88.17	43.01	56.99	41.94	58.06	45.16	54.84	36.56	63.44	35.48	64.52	31.18	68.82	%	

Continu page suivante ...

... Suite de la page précédente															
Centre				Néoplasie		Ischémie		Relation		Cardio		Maladie i		Cov	
4	3	2	1	1	0	≥ 24	< 24	1	0	1	0	1	0	1	0
20	1	101	50	5	167	120	52	171	1	131	41	67	105	172	1990
11.63	0.58	58.72	29.07	2.91	97.09	69.77	30.23	99.42	0.58	76.16	23.84	38.95	61.05	%	
0	1	89	64	7	147	103	51	152	2	121	33	58	96	154	1991
0.00	0.65	57.79	41.56	4.55	95.45	66.88	33.12	98.70	1.30	78.57	21.43	37.66	62.34	%	
0	4	97	52	4	149	97	56	149	4	120	33	79	74	153	1992
0.00	2.61	63.40	33.99	2.61	97.39	63.40	36.60	97.39	2.61	78.43	21.57	51.63	48.37	%	
0	9	90	42	10	131	77	64	137	4	108	33	47	94	141	1993
0.00	6.38	63.83	29.79	7.09	92.91	54.61	45.39	97.16	2.84	76.60	23.40	33.33	66.67	%	
0	2	72	37	4	107	59	52	110	1	92	19	48	63	111	1994
0.00	1.80	64.86	33.33	3.60	96.40	53.15	46.85	99.10	0.90	82.88	17.12	43.24	56.76	%	
0	3	86	47	8	128	85	51	129	7	97	39	56	80	136	1995
0.00	2.21	63.24	34.56	5.88	94.12	62.50	37.50	94.85	5.15	71.32	28.68	41.18	58.82	%	
0	19	87	41	1	146	83	64	141	6	111	36	59	88	147	1996
0.00	12.93	59.18	27.89	0.68	99.32	56.46	43.54	95.92	4.08	75.51	24.49	40.14	59.86	%	
0	22	85	48	5	150	54	101	150	5	117	38	57	98	155	1997
0.00	14.19	54.84	30.97	3.23	96.77	34.84	65.16	96.77	3.23	75.48	24.52	36.77	63.23	%	
0	58	71	63	11	181	78	114	184	8	130	62	69	123	192	1998
0.00	30.21	36.98	32.81	5.73	94.27	40.63	59.38	95.83	4.17	67.71	32.29	35.94	64.06	%	
0	35	63	53	10	141	62	89	143	8	120	31	49	102	151	1999
0.00	23.18	41.72	35.10	6.62	93.38	41.06	58.94	94.70	5.30	79.47	20.53	32.45	67.55	%	
0	54	91	56	5	196	97	104	193	8	171	30	67	134	201	2000
0.00	26.87	45.27	27.86	2.49	97.51	48.26	51.74	96.02	3.98	85.07	14.93	33.33	66.67	%	
1	23	97	63	9	175	69	115	171	13	160	24	62	122	184	2001
0.54	12.50	52.72	34.24	4.89	95.11	37.50	62.50	92.93	7.07	86.96	13.04	33.70	66.30	%	
56	20	100	82	11	247	102	156	235	23	212	46	87	171	258	2002
21.71	7.75	38.76	31.78	4.26	95.74	39.53	60.47	91.09	8.91	82.17	17.83	33.72	66.28	%	
59	11	79	63	13	199	86	126	190	22	188	24	71	141	212	2003
27.83	5.19	37.26	29.72	6.13	93.87	40.57	59.43	89.62	10.38	88.68	11.32	33.49	66.51	%	
4	16	104	86	23	187	70	140	187	23	187	23	68	142	210	2004
1.90	7.62	49.52	40.95	10.95	89.05	33.33	66.67	89.05	10.95	89.05	10.95	32.38	67.62	%	
1	31	101	61	10	184	67	127	170	24	174	20	45	149	194	2005
0.52	15.98	52.06	31.44	5.15	94.85	34.54	65.46	87.63	12.37	89.69	10.31	23.20	76.80	%	
0	30	50	13	4	89	32	61	77	16	84	9	29	64	93	2006
0.00	32.26	53.76	13.98	4.30	95.70	34.41	65.59	82.80	17.20	90.32	9.68	31.18	68.82	%	

TAB. A.1 – Détails des effectifs des modalités des différentes covariables par année de greffe.

BIBLIOGRAPHIE

- [1] OO Aalen. Heterogeneity in survival analysis. *Statistics in Medicine*, 7(11) :1121–1137, 1988. (Cité pages 41 et 45.)
- [2] OO Aalen. Modelling heterogeneity in survival analysis by the Compound Poisson distribution . *Annals of Applied Probability*, 4 (2) :951–972, 1992. (Cité page 41.)
- [3] OO Aalen. Effects of frailty in survival analysis. *Statistical Methods for Medical Research*, 3 :227–243, 1994. (Cité page 47.)
- [4] R Aguirre-Hernandez and VT Farewell. A Pearson-type goodness-of-fit test for stationnary and time-continuous Markov regression models. *Statistics in Medicine*, 21 :1899–1911, 2002. (Cité pages 91, 98 et 121.)
- [5] P Andersen and M Vaeth. Simple parametric and nonparametric models for excess and relative mortality. *Biometrics*, 45(2) :523–535, 1989. (Cité page 57.)
- [6] PK Andersen, K Borch-Johnsen, T Deckert, A Green, P Hougaard, N Keiding, and S Kreiner. A Cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics*, 41 :921–932, 1985. (Cité pages 58, 61, 63 et 74.)
- [7] PK Andersen, JP Klein, and MJ Zhang. Testing for centre effects in multi-centre survival studies : a Monte Carlo comparison of fixed and random effects tests. *Statistics in Medicine*, 18 :1489–1500, 1999. (Cité page 36.)
- [8] J Berkson. The calculation of survival rates. In *W Walters et H.K. Gray et P.J. T. éditeurs : Carcinoma and other malignant lesions of the stomach*, pages 467–484, 1942. (Cité pages 2 et 57.)
- [9] J Berkson and R Gage. *Calculation of survival rates for cancer*, 25 :270–286. (Cité page 57.)
- [10] P Bolard, C Quantin, J Esteve, J Faivre, and M Abrahamowicz. Modelling time-dependant hazard ratios in relative survival : Application to colon cancer. *Journal of Clinical Epidemiology*, 54 :986–996, 2001. (Cité page 66.)
- [11] N Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review*, 43 :45–58, 1975. (Cité page 57.)

- [12] NE Breslow, JH Lubin, P Marek, and B Langholz. Multiplicative models and cohort analysis. *Journal of the American Statistical Association*, 78 :1–12, 1983. (Cité page 58.)
- [13] JD Buckley. Additive and multiplicative models for relative survival rates. *Biometrics*, 40 :51–62, 1984. (Cité pages 57 et 61.)
- [14] RH Byrd, P Lu, J Nocedal, and C Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, 16 :1190–1208, 1995. (Cité page 26.)
- [15] EF Campos, H Tedesco-Silva, PG Machado, M Franco, JO Medina-Pestana, and M Gerbase-DeLima. Post-transplant anti-HLA class II antibodies as risk factor for late kidney allograft failure. *American Journal of Transplantation*, 10 :2316–2320, 2006. (Cité page 10.)
- [16] C Castelli. *Modèles semi-markoviens et méthodes de régression dans les analyses de coût-efficacité : application au cancer colo-rectal*. Thèse de doctorat en Biostatistique, 2007. (Cité pages 3 et 90.)
- [17] C Castelli, P Saint-Pierre, and JP Daures. Informative censoring in survival analysis and application to asthma. *Far East Journal of Theoretical Statistics*, 19(2) :203–217, 2006. (Cité page 122.)
- [18] D Clayton and J Cuzick. *The semi-parametric Pareto model for regression analysis of survival times*. Proceedings of the Centenary Session of the International Statistical Institute, Amsterdam, 1985. (Cité page 43.)
- [19] DG Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65 :141–151, 1978. (Cité pages 2, 35, 41 et 43.)
- [20] DR Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society : Series B*, 34 :187–220, 1972. (Cité pages 1, 3, 24, 35 et 96.)
- [21] PW Dickman, A Sloggett, M Hills, and T Hakulinen. Regression models for relative survival. *Statistics in medicine*, 23 :51–64, 2004. (Cité page 62.)
- [22] F Ederer, LM Axtell, and SJ Cutler. The relative survival rate : A statistical methodology. *National Cancer Institute*, 1961. (Cité pages 57 et 60.)
- [23] F Ederer and H Heise. The effect of eliminating deaths from cancer in general population survival rates. *Rapport technique, National Cancer Institute, methodological note 11, End result Evaluation Section.*, 1959. (Cité pages 57 et 63.)
- [24] B Efron and RJ Tibshirani. *An introduction to the Bootstrap*. Chapman and Hall, New-York, 1993. (Cité page 100.)

- [25] J Estève, E Benhamou, M Croasdale, and L Raymond. Relative Survival And The Estimation Of Net Survival : Elements For Further Discussion. *Statistics in Medecine*, 9 :529–538, 1990. (Cité pages 58, 60, 61, 66 et 74.)
- [26] J Estève, E Benhamou, and L Raymond. Statistical methods in descriptive epidemiology. *IARC Scientific Publication n128, International Agency for Research on Cancer, Lyon, 1994.* (Cité page 59.)
- [27] Y Foucher. *Modèles semi-markoviens : Application à l'analyse de l'évolution de pathologies chroniques.* Thèse de doctorat en Biostatistique, 2007. (Cité pages 91, 98 et 110.)
- [28] R Giorgi. *Analyses comparatives des méthodes de survie et extensions d'un modèle régressif de survie relative : prise en compte de la non proportionnalité des risques par des fonctions B-splines et développement d'une méthode d'analyse bayésienne.* Thèse de doctorat en Biostatistique, 2002. (Cité page 63.)
- [29] R Giorgi, M Abrahamowicz, C Quantin, P Bolard, J Estève, J Gouvernet, and J Faivre. A relative regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine*, 22 :2767–2784, 2003. (Cité page 66.)
- [30] DW Gjertson, DM Dabrowska, X Cui, and JM Cecka. Four Causes of Cadaveric Kidney Transplant Failure : A Competing Risk Analysis. *American Journal of Transplantation*, 2 :84–93, 2002. (Cité page 3.)
- [31] M Greenwood and GU Yule. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 83 :255–279, 1920. (Cité pages 1 et 35.)
- [32] T Hakulinen. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics*, 38 :933–942, 1982. (Cité pages 57 et 60.)
- [33] T Hakulinen and L Tenkanen. Regression analysis of relative survival rates. *Journal of the Royal Statistical Society Series C*, 36 :309–317, 1987. (Cité pages 57, 58, 60, 62, 66 et 74.)
- [34] K Harzallah, D Floret, X Martin, and P Cochat. Mortalité en transplantation rénale pédiatrique : 15 ans d'expériences = Mortality in pediatric renal transplant : experience of 15 years. *Archives de pédiatrie*, 11(8) :916–920, 2004. (Cité page 10.)
- [35] P Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73 :671–678, 1986a. (Cité pages 41 et 44.)
- [36] P Hougaard. A class of multivariate failure time distributions. *Biometrika*, 73 :671–678, 1986b. (Cité pages 41 et 44.)
- [37] P Hougaard. *Analysis of multivariate survival data.* Springer, New-York, 2000. (Cité pages 42, 43, 45, 56 et 118.)

- [38] EL Kaplan and P Meier. Non parametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53 :457–481, 1958. (Cité pages 22 et 26.)
- [39] JP Klein. Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm. *Biometrics*, 48 :795–806, 1992. (Cité page 45.)
- [40] DG Kleinbaum and M Klein. *Survival Analysis - A Self-Learning Text (Second Edition)*. Springer, London, Statistics for Biology and Health, 2005. (Cité page 25.)
- [41] PC Lambert, LK Smith, DR Jones, and JL Botha. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine*, 24 :3871–3885, 2005. (Cité page 66.)
- [42] HS Mackenzie, H Azuma, HG Rennke, NL Tilney, and BM Brenner. Renal mass as a determinant of late allograft outcome : insights from experimental studies in rats. *Kidney International Supplement*, 52 :S38–42, 1995. (Cité page 10.)
- [43] CA McGilchrist and CW Aisbett. Regression with frailty in survival analysis. *Biometrics*, 47 :461–466, 1991. (Cité page 41.)
- [44] CP Nelson, PC Lambert, IB Squire, and DR Jones. Flexible parametrics models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26 :5486–5498, 2007. (Cité page 66.)
- [45] J Neugarten, T Srinivas, V Tellis, S Silbiger, and Greenstein S. The effect of donor gender on renal allograft survival. *Journal of the American Society of Nephrology*, 7 :318–324, 1996. (Cité page 10.)
- [46] JR Nyengaard and TF Bendtsen. Glomerular number and size in relation to age, kidney weight, and body surface in normal man. *Anatomical Record*, 232 :194–201, 1992. (Cité page 10.)
- [47] O Oudar, M Elger, L Bankir, D Ganten, U Ganten, and W Kriz. Differences in rat kidney morphology between males, females and testosterone-treated females. (Cité page 10.)
- [48] SJ Pocock, SM Gore, and GR Kerr. Long-term survival analysis : The curability of breast cancer. *Statistics in Medicine*, 1 :93–104, 1982. (Cité pages 58 et 61.)
- [49] M Pohar and J Stare. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine*, 81 :272–278, 2006. (Cité page 68.)
- [50] JM Raboud, JSG Montaner, A Thorne, J Singer, and MT Schechter. Impact of missing data due to dropouts on estimates of the treatment effect in a randomized trial of antiretroviral therapy for hiv-infected individuals. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 12(1) :46–55, 1996. (Cité pages 3 et 90.)

- [51] L Remontet, N Bossard, A Belot, J Estève, and the French network of cancer registries FRANCIM. An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. *Statistics in medicine*, 26 :2214–2228, 2007. (Cité page 66.)
- [52] JP Rerolle, JC Szlag, M Diaconita, Saada, JC Aldigier, and Y Le Meur. Renal transplantation in the elderly : a survey. *Néphrologie et Thérapeutique*, 2(1) :8–14, 2006. (Cité page 9.)
- [53] JM Robins. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, pages 24–33, 1993. (Cité page 122.)
- [54] JM Robins and DM Finkelstein. Correcting for noncompliance and dependent censoring in a AIDS clinical trial with Inverse Probability of Censoring Weighted (IPCW) log-rank tests. *Biometrics*, 56 :779–788, 2000. (Cité page 122.)
- [55] JX Song, JT Wassel, and A Kapadia. Relative mortality for correlated lifetime data. *Computational Statistics and Data Analysis*, 45 :849–864, 2004. (Cité page 78.)
- [56] CE Stigant and JS Zaltzamn. La greffe rénale chez les personnes âgées. *Néphrologie, Conférences scientifiques*, 1(6), 2000. (Cité page 9.)
- [57] JW Vaupel, KG Manton, and E Stallard. The Impact of the Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16 :439–454, 1979. (Cité pages 2, 35, 39 et 41.)
- [58] P Vereerstraeten, M Wissing, L De Pauw, D Abramowicz, and P Kinnaert. Male recipients of kidneys from female donors are at increased risk of graft loss from both rejection and technical failure. *Clinical Transplantation*, 13 :181–186, 1999. (Cité page 10.)
- [59] A Wienke, N Holm, A Skytthe, and A.I Yashin. The heritability of mortality due to heart diseases : a correlated frailty model applied to Danish twins. *Twin Research*, 4 :266–274, 2001. (Cité page 44.)
- [60] RA Wolfe, VB Ashby, EL Milford, AO Ojo, RE Ettenger, LYC Agodoa, PJ Held, and FK Port. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *The New England Journal of Medicine*, 341(23) :1725–30, 1999. (Cité pages 4, 90, 95, 100 et 121.)
- [61] AI Yashin and IA Iachine. Genetic Analysis of Durations : Correlated Frailty Model Applied to Survival of Danish Twins. *Genetic Epidemiology*, 12 :529–538, 1995. (Cité page 44.)
- [62] PH Zahl. Frailty Modelling For The Excess Hazard. *Statistics in Medicine*, 16 :1573–1585, 1997. (Cité page 44.)

- [63] M Zeier, B Dohler, G Opelz, , and E Ritz. The effect of donor gender on graft survival. *Journal of the American Society of Nephrology*, 13 :2570–2576, 2002. (Cité page 10.)