



**HAL**  
open science

## Test de type-log rank pour l'évolution de la qualité de vie liée à la santé

Véronique Boisson

► **To cite this version:**

Véronique Boisson. Test de type-log rank pour l'évolution de la qualité de vie liée à la santé. Sciences du Vivant [q-bio]. Université Pierre et Marie Curie - Paris VI, 2008. Français. NNT : . tel-00356750

**HAL Id: tel-00356750**

**<https://theses.hal.science/tel-00356750>**

Submitted on 28 Jan 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

*Spécialité* : Mathématiques  
*Option* : Statistique, Biostatistique

*présentée par*

**Véronique BOISSON**

*pour obtenir le grade de*

DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS VI

Sujet de la thèse

**Test de type log-rank pour l'évolution de la qualité de vie  
liée à la santé**

Soutenue publiquement le 3 décembre 2008 devant le jury composé de

<i>Directeur de thèse</i>	M. Mounir MESBAH	Université Pierre et Marie Curie - Paris VI
<i>Rapporteurs</i>	M. Michel CHAVANCE	INSERM U780 - IFR69, Villejuif
	M. Nozer D. SINGPURWALLA	The Georges Washington University
<i>Examineurs</i>	M. Avner BAR-HEN	Université René Descartes - Paris V
	M. Paul DEHEUVELS	Université Pierre et Marie Curie - Paris VI
	M <sup>me</sup> Catherine HUBER-CAROL	Université René Descartes - Paris V
	M. Giovanni PECCATI	Université de Paris Ouest - Paris X
	M. Rachid SENOUSI	INRA Avignon



« Ô mathématiques sévères, je ne vous ai pas oubliées depuis que vos savantes leçons, plus douces que le miel, filtrèrent dans mon cœur comme une onde rafraîchissante. »

LAUTRÉMONT

« Un nombre n'est pas premièrement un fait. Mais n'est-ce pas plutôt par le nombre que la multitude devient un fait. »

ALAIN

« Il y a évidemment quelque chose qui a été cause de la mort chez le malade qui a succombé, et qui ne s'est pas rencontré chez le malade qui a guéri ; c'est ce quelque chose qu'il faut déterminer, et alors on pourra agir sur ces phénomènes ou les reconnaître et les prévoir exactement ; alors seulement on aura atteint le déterminisme scientifique. »

Claude BERNARD

« Le choix d'un plan thérapeutique, d'une technique de traitement, l'étude de la signification pronostique d'un phénomène, etc. . . , ne peuvent progresser que s'ils sont basés sur des informations dont la signification a été clairement appréciée à partir d'une technique statistique rigoureuse. »

Pierre DENOIX



# Remerciements

*Si la thèse est encore le symbole d'un travail solitaire, les échanges et les collaborations sont indispensables. C'est donc avec plaisir que je remercie ici tous ceux qui m'ont apporté leur concours, leurs conseils et leurs idées.*

*Ma reconnaissance va tout d'abord à mon directeur de thèse le Professeur Mounir Mesbah pour ses précieux conseils, ses encouragements constants et ses observations critiques, dont l'intérêt et les remarques ont été toujours fort stimulants et sans lesquels ce travail n'aurait pu aboutir. Qu'il soit donc très chaleureusement remercié.*

*Je désire adresser une marque toute particulière de ma gratitude au Professeur Paul Deheuvels qui m'a accueillie au sein de son laboratoire et m'a toujours accordé sa confiance et son soutien tout au long de ce travail de thèse.*

*Mes remerciements s'adressent ensuite aux rapporteurs de mon travail de thèse les Professeurs Catherine Huber-Carol et Nozer Singpurwalla ainsi que les Directeurs de recherche Michel Chavance et Rachid Senoussi qui ont bien voulu consacrer du temps à la lecture et à l'examen de cette recherche, me permettant ainsi d'améliorer la qualité et d'envisager de nouvelles voies de développement.*

*J'adresse mes remerciements aux Professeurs Avner Bar-Hen, Paul Deheuvels et Giovanni Peccati pour avoir bien voulu être membre de mon jury de thèse.*

*Je tiens à remercier le Professeur Zhiliang Ying pour m'avoir fait l'honneur de s'intéresser à mon travail et de collaborer en vue de la publication d'un article.*

*Je tiens à remercier l'ensemble de l'Unité INSERM U379 basée à Marseille pour leur aide et conseils concernant la problématique liée au VIH, notamment Patrizia Carriero, Camélia Protopopescu, Bruno Spire, Virginie Villes et tout particulièrement Marie Préau pour son aide inestimable apportée lors de la rédaction de mon projet de bourse.*

*Cette thèse s'inscrit dans un cadre multidisciplinaire et je remercie l'ensemble des membres de la Cohorte ANRS CO8 (voir annexe C) qui sous l'impulsion de Bruno Spire, Geneviève Chêne, Catherine Leport et François Raffi m'a permis d'utiliser leur base de données et ainsi permis le volet applications.*

*J'exprime ma reconnaissance au Professeur Sylvie Chevret ainsi qu'à toute l'équipe du*

*Département de Biostatistique et Informatique Médicale (DBIM) - Hôpital Saint-Louis pour m'avoir initiée à la recherche et guidée lors de mon stage de DEA et plus particulièrement à Raphaël Porcher et Aurélien Latouche pour m'avoir encadrée durant ce stage de DEA et m'avoir confortée dans mon désir de poursuivre en thèse.*

*Je tiens à saluer tous mes nouveaux collègues de l'IUT de Paris - Université René Descartes pour leur accueil sympathique et chaleureux ainsi que pour me permettre de finaliser ma thèse dans des conditions très favorables.*

*Ces années passées au cœur de l'équipe du Laboratoire de Statistique Théorique et Appliquée (LSTA) ont été pour moi particulièrement enrichissantes, c'est la raison pour laquelle je tiens à exprimer ma reconnaissance à tous les Professeurs et Maîtres de conférence du LSTA. Je n'oublierai pas Michel Broniatowski, Giovanni Peccati et Philippe Saint-Pierre. Je garderai un souvenir précieux de l'ensemble des doctorants (Aboubacar, Amadou, Anissa, Bertrand, Boris, Claire, Clara, Glwadys, Hicham, Issam, Jacques, Jean-Baptiste, Julien, Lahcen, Mamadou, Mohamed, Moïse, Mory, Nabil, Nourdine, Olivier, Pierre, Ravan, Salim, Samuella, Ségolène, Vivian, ...) pour leur amitié, soutien, conseils et l'ambiance agréable des moments partagés. J'apporte une mention toute spéciale pour Boutayna, Kaouthar et Bah qui partagent le même directeur de thèse que moi ainsi qu'à François-Xavier pour ses conseils judicieux en SAS<sup>®</sup>.*

*Une thèse ne saurait se faire sans l'aide si précieuse et indispensable du secrétariat Louise Lamart et Anne Durrande ainsi que du professionnalisme du bibliothécaire de l'ISUP Pascal Epron. Je les remercie pour tout du fond du cœur.*

*Je n'oublie pas le plaisir des discussions et échanges fructueux au détour de conférences ou séminaires avec des collègues et amis tels que Véronique Sébille, Sylvia Bacci, Jean-François Dupuy, Mohand Feddag et Jean-Benoît Hardouin.*

*Je voudrais exprimer ma profonde reconnaissance à mes parents, pour m'avoir toujours soutenue et donné les moyens de mener à bien mes études.*

*Enfin, je tiens à remercier l'Agence Nationale de Recherches sur le Sida et les hépatites virales (ANRS) pour le financement qu'elle m'a octroyé et sans lequel cette thèse n'aurait pu être possible.*

# Table des matières

<b>Introduction</b>	<b>5</b>
<b>Notations</b>	<b>9</b>
<b>Abstract</b>	<b>11</b>
<b>I Préliminaires</b>	<b>19</b>
<b>1 Introduction à l'analyse de survie</b>	<b>21</b>
1.1 Fonctions de survie et de risque . . . . .	21
1.2 Données manquantes, censures et troncatures . . . . .	22
1.2.1 Données manquantes . . . . .	22
1.2.2 Schémas de censure . . . . .	24
1.2.3 Troncatures . . . . .	25
1.3 Processus de comptage et constitution des observations . . . . .	25
1.4 L'estimateur de Kaplan-Meier . . . . .	26
1.5 Le modèle semi-paramétrique de Cox . . . . .	26
<b>2 Introduction à l'analyse longitudinale</b>	<b>29</b>
2.1 Modèle linéaire mixte . . . . .	29
2.1.1 Formulation du modèle . . . . .	29
2.1.2 Estimation des paramètres . . . . .	31
2.2 Modèle linéaire mixte pour données longitudinales multivariées . . . . .	31
<b>3 Qualité de vie et analyse de survie</b>	<b>33</b>
3.1 Définition de la qualité de vie . . . . .	33
3.2 Analyse longitudinale de la qualité de vie . . . . .	34
3.3 Modèles de dégradation dans l'industrie . . . . .	35
3.3.1 Présentation et hypothèses . . . . .	35
3.3.2 Processus gaussiens et premier temps de passage . . . . .	36



<b>II</b>	<b>Test non-paramétrique pour l'évolution de la qualité de vie</b>	<b>37</b>
<b>4</b>	<b>Test de type log-rank pour l'évolution de la qualité de vie</b>	<b>39</b>
4.1	Définitions et propriétés de la dégradation de qualité de vie . . . . .	39
4.2	Notations et hypothèses . . . . .	40
4.3	Analyse longitudinale de la dégradation de qualité de vie pour un taux fixé	42
4.3.1	Statistique du score de vraisemblance partielle . . . . .	42
4.3.2	Statistique du test de type log-rank . . . . .	49
4.4	Analyse longitudinale de la dégradation de qualité de vie pour un taux variable	50
4.4.1	Statistique du score de vraisemblance partielle . . . . .	50
4.4.2	Test global de type log-rank . . . . .	52
4.5	Application à une base de données . . . . .	53
<b>5</b>	<b>Simulations</b>	<b>57</b>
<b>III</b>	<b>Application à la Cohorte ANRS CO8</b>	<b>63</b>
<b>6</b>	<b>La Cohorte ANRS CO8</b>	<b>65</b>
6.1	Description de la Cohorte ANRS CO8 . . . . .	65
6.1.1	La Cohorte APROCO . . . . .	65
6.1.2	La Cohorte COPILOTE . . . . .	66
6.2	Rappel des résultats déjà obtenus sur la Cohorte ANRS CO8 . . . . .	66
6.3	Description de la base de données . . . . .	68
6.4	Difficultés rencontrées pour l'étude de la Cohorte ANRS CO8 . . . . .	73
6.4.1	Réponses au questionnaire de qualité de vie à l'inclusion . . . . .	73
6.4.2	Changement de questionnaires et equating . . . . .	73
6.4.3	Problème des dates manquantes . . . . .	79
6.4.4	Problème des données manquantes . . . . .	79
<b>7</b>	<b>Analyse longitudinale de la qualité de vie des patients de la Cohorte ANRS CO8</b>	<b>83</b>
7.1	Résultats obtenus pour le test global de type log-rank . . . . .	83
7.2	Résultats obtenus par le modèle linéaire mixte . . . . .	89
7.3	Discussion . . . . .	89
	<b>Conclusion et perspectives</b>	<b>93</b>
<b>IV</b>	<b>Annexes</b>	<b>95</b>
<b>A</b>	<b>Rappels mathématiques</b>	<b>97</b>
A.1	Processus auto-régressif . . . . .	97
A.2	Mouvement brownien . . . . .	99

A.3	Martingales et processus ponctuels . . . . .	99
A.3.1	Définitions . . . . .	99
A.3.2	Processus de comptage . . . . .	101
A.3.3	Convergence . . . . .	101
A.3.4	Inégalité de Lengart . . . . .	102
<b>B</b>	<b>Programmes informatiques</b>	<b>105</b>
B.1	Test global de type log-rank . . . . .	105
B.1.1	Première dégradation de la qualité de vie . . . . .	105
B.1.2	Dégradation définitive de la qualité de vie . . . . .	115
B.2	Simulations . . . . .	118
<b>C</b>	<b>Groupe d'étude de la Cohorte ANRS CO8</b>	<b>121</b>
	<b>Bibliographie</b>	<b>123</b>



# Introduction

L'avènement des multithérapies antirétrovirales hautement actives qui réduisent et retardent les manifestations pathologiques de l'infection à VIH, a entraîné un allongement de l'espérance de vie, une réduction radicale de la morbidité et un bouleversement dans la prise en charge des patients infectés par le VIH. Cette infection est entrée dans une ère nouvelle qualifiée de « *période de normalisation* » (Setbon, 2000), où les patients sont contraints à un traitement sans interruption sur le long terme. Effectivement, « *La règle générale est qu'il n'y a pas de bénéfice à espérer d'un arrêt du traitement anti-rétroviral chez un patient en succès thérapeutique. En effet, les interruptions de traitement sont suivies d'un rebond de la réplication du VIH et d'une baisse des lymphocytes CD4, d'autant plus que le nadir des lymphocytes CD4 (valeur la plus basse dans l'historique du patient) est plus bas.* » (Yeni, 2006). D'autre part, ces multithérapies montrent d'importantes limites ainsi, le traitement non curatif nécessite une stricte observance indispensable pour maintenir le succès thérapeutique (Paterson *et al.*, 2000). De plus, une toxicité à long terme des antirétroviraux est constatée, se manifestant par l'apparition de nombreux effets indésirables souvent très gênants, persistants au cours du temps et pouvant compromettre la qualité de vie des patients infectés par le VIH. D'autant plus que les traitements les plus puissants peuvent induire davantage d'effets indésirables que des traitements légèrement moins puissants mais mieux supportés. C'est notamment le cas entre la combinaison lopinavir/ritonavir et le nelfinavir (Cvetkovic and Goa, 2003). Ainsi, un problème majeur est de savoir si administrer des traitements plus puissants mais au prix de davantage d'effets indésirables est toujours judicieux ou s'il n'est pas préférable de privilégier la qualité de vie des patients en prescrivant des traitements légèrement moins puissants.

Malgré la prise de conscience récente de la nécessité de s'intéresser non plus seulement à l'efficacité thérapeutique des multithérapies, mais également à la qualité de vie des patients, des questions subsistent. En effet, un nombre important de stratégies thérapeutiques existe et il semble indispensable de pouvoir les comparer en terme de nombre d'effets indésirables ressentis par le patient. Quel est le moment propice pour adapter un traitement, non plus seulement au regard des données biologiques, mais également dans un souci d'améliorer l'acceptation du traitement par le patient ? Dans quelle mesure doit-on apporter un soutien psychologique, en plus du suivi médical ? Ces questions nécessitent que l'on s'intéresse à l'évolution de l'état de santé des patients conjointement à leur qualité de vie. Afin de répondre à ces problématiques un nombre croissant de publications portant

sur l'élaboration de questionnaires de qualité de vie (Leplège *et al.*, 1995, 1998; O'Connell *et al.*, 2003; Shahrier *et al.*, 2003; WHOQOL-HIV, 2004; Wu *et al.*, 2004) et à leurs utilisations en santé publique est observé. Très souvent, l'objectif de ces études porte sur la comparaison de différentes stratégies thérapeutiques en prenant comme critère majeur soit le succès immuno-virologique, soit l'évolution de la qualité de vie (Badia *et al.*, 2004; Cohen *et al.*, 2004; Coplan *et al.*, 2004; Van Leth *et al.*, 2004). Néanmoins dans le cas d'une analyse longitudinale de la qualité de vie des patients infectés par le VIH les outils statistiques existants demeurent insuffisants.

L'objectif de cette thèse consiste en l'élaboration d'un test statistique non-paramétrique permettant d'analyser conjointement les données longitudinales de qualité de vie et de survie par rapport aux variables socio-immuno-virologiques des patients. Cette méthode peut être appliquée aux essais cliniques comme aux études épidémiologiques évaluant différentes stratégies thérapeutiques dans le cadre d'une maladie chronique (infection à VIH, hépatites, diabète, ...). Cet objectif s'intègre à la fois dans les problématiques de l'Agence Nationale de Recherches sur le SIDA et les hépatites virales (ANRS), que dans les grands axes du plan gouvernemental 2007-2011 en santé publique concernant l'amélioration de la qualité de vie des personnes atteintes de maladies chroniques (Bas, 2007).

Généralement lors d'une étude clinique longitudinale, des questionnaires de qualité de vie validés sont donnés à remplir aux patients à des dates préalablement fixées  $t_j$  permettant ainsi de calculer leur score de qualité de vie  $Q_j$ . L'évolution de la qualité de vie peut se traduire par le concept récent de dégradation de qualité de vie, concept non encore défini de manière consensuelle dans la littérature. En effet, la dégradation de qualité de vie à une date  $t_j$ , peut être définie soit par  $D_j = (Q_0 - Q_j)Q_0^{-1}$ , soit par  $D_j = Q_0 - Q_j$ , i.e. soit comme un pourcentage, soit comme une différence mesurant l'écart entre le niveau de qualité de vie lors de l'entrée du patient dans l'étude (baseline) et le niveau de qualité de vie à cette date  $t_j$ . Un taux critique  $x$  de dégradation de qualité de vie à ne pas dépasser peut être fixé. Ainsi, l'événement d'intérêt  $T(x) = \inf\{t_j; D_j \geq x\}$  est la première occurrence où la dégradation de qualité de vie dépasse ce taux critique  $x$  de dégradation de qualité de vie. Pour un taux critique  $x$  de dégradation de qualité de vie préalablement fixé, Awad *et al.* (2002); Mesbah *et al.* (2004) comparent le temps de dégradation de qualité de vie pour deux groupes de patients à l'aide de méthodes classiques d'analyse de survie (Andersen *et al.*, 1993; Fleming and Harrington, 1991; Hill *et al.*, 1996; Hougaard, 2000; Kalbfleisch and Prentice, 2002; Klein and Moeschberger, 2003; Therneau and Grambsch, 2000) telles que l'estimateur de Kaplan-Meier et le test du log-rank.

Le travail de cette thèse, présenté dans le synopsis ci-dessous, s'articule autour de l'élaboration d'un test permettant d'analyser la dégradation de qualité de vie de deux groupes de patients et de son application à une cohorte de patients infectés par le VIH.

Dans le premier chapitre de ce manuscrit, nous rappelons les définitions des fonctions classiquement utilisées en analyse de survie, et les notions de censure et de troncature. Puis nous présentons l'estimateur de Kaplan-Meier et le modèle de Cox.

Dans le deuxième chapitre, nous donnons une description de l'analyse longitudinale et

plus particulièrement des modèles linéaires généralisés.

Dans le troisième chapitre, nous commençons par introduire la notion de qualité de vie liée à la santé en se référant à l'approche globale proposée par l'Organisation Mondiale pour la Santé. Puis nous décrivons les méthodes existantes permettant l'analyse longitudinale de la qualité de vie à savoir la méthode utilisée par Awad *et al.* (2002), celle employée par Heutte and Huber-Carol (2002) ainsi que celle proposée par Jin *et al.* (2006). Enfin, nous abordons brièvement les méthodes d'étude de la dégradation du domaine de l'industrie, car ces techniques présentent un certain nombre d'analogies.

Le quatrième chapitre débute par le concept récent de dégradation de qualité de vie, n'ayant pas encore fait l'objet d'une définition consensuelle. Puis nous étendons la statistique du score de vraisemblance partielle de Cox (1975) afin de prendre en compte un taux  $x$  de dégradation de qualité de vie préalablement fixé. Nous prouvons que lorsque le temps  $t$  est continu, le vecteur du processus de score normalisé  $\{n^{-1/2}\mathbf{U}(\beta_0, t, x); 0 \leq t \leq t_T\}$ , défini à l'instant  $t$  par

$$\frac{1}{\sqrt{n}}\mathbf{U}(\beta_0, t, x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \{\mathbf{Z}_i(u) - \mathbf{E}(\beta_0, u, x)\} N_i(du, x),$$

converge vers un processus gaussien de moyenne nulle et à accroissements indépendants. De plus, nous obtenons les propriétés asymptotiques des estimateurs lorsque le taux  $x$  de dégradation de qualité de vie est préalablement fixé.

Le taux  $x$  de dégradation de qualité de vie est ensuite supposé variable. À l'aide de la théorie des processus empiriques (Pollard, 1984; Shorack and Wellner, 1986; Biliias *et al.*, 1997), nous établissons la convergence en distribution de la statistique du score normalisé  $n^{-1/2}U_n(\cdot)$  vers un processus gaussien de matrice de covariance  $\Gamma$ . Ces travaux nous ont permis de construire, lorsque le taux  $x$  de dégradation de qualité de vie est variable, un test statistique non-paramétrique global de type log-rank permettant de comparer l'évolution longitudinale de la dégradation de qualité de vie pour deux groupes de patients.

Dans le cinquième chapitre, afin de comparer les performances du test introduit au chapitre précédent, nous avons réalisé une étude de simulations. Pour cela, nous avons implémenté à l'aide du logiciel SAS<sup>®</sup> le test non-paramétrique global de type log-rank développé, les codes sont visibles dans la partie annexes.

Dans la partie III, nous illustrons ce test statistique non-paramétrique global de type log-rank par une application sur la Cohorte ANRS CO8 (Protopopescu *et al.*, 2007), Cohorte d'observation française de patients infectés par le VIH-1 ayant pour objectif l'étude des effets sur le long terme des antirétroviraux hautement actifs dans le contexte de la prise en charge de patients mis sous inhibiteur de protéase pour la première fois entre 1997 et 1999. Le questionnaire de qualité de vie remis aux patients a été au début de l'étude l'échelle générique SF-36 remplacée ensuite par l'échelle spécifique WHOQOL-HIV. Afin d'obtenir un unique score de qualité de vie sur l'ensemble des sept dates dont nous disposons nous avons eu recours à des méthodes d'equating. D'autre part, nous avons réalisé une analyse longitudinale classique reposant sur les modèles linéaires mixtes.

Cette thèse a bénéficié d'un financement par l'Agence Nationale de Recherches sur le SIDA et les hépatites virales (ANRS). Une partie des travaux présentés dans ce manuscrit a fait l'objet d'une note parue aux Comptes Rendus de l'Académie des Sciences (Boisson and Mesbah, 2008), et de plusieurs présentations orales lors de conférences (Boisson, 2005, 2008; Boisson *et al.*, 2006; Mesbah *et al.*, 2008; Boisson *et al.*, 2008b,c) dont quatre internationales. Par ailleurs, un article vient d'être soumis (Boisson *et al.*, 2008a).

# Notations

Nous introduisons ci-dessous les conventions d'écriture adoptées dans les différentes parties de ce manuscrit.

## *Abréviations et symboles*

- Les variables aléatoires considérées sont définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ .
- $\mathbb{P}$  la mesure de probabilité attachée à l'espace probabilisable  $(\Omega, \mathcal{A})$ .
- $\mathbb{E}$  l'espérance mathématique associée à  $\mathbb{P}$ .
- $a \wedge b$  le minimum de  $a$  et  $b$ .
- $a \vee b$  le maximum de  $a$  et  $b$ .
- $\mathbf{1}_{\{A\}}$  l'indicatrice de  $A$ .
- $\sigma\{N\}$  la tribu engendrée par  $N$ .
- $\mathcal{F}_t$  filtration à l'instant  $t$ .
- $\Delta F(t) = F(t) - F(t-)$ , si  $F$  est continue à droite.
- $\int \cdots F(dt) = \int \cdots f(t)dt$ , si  $F$  admet une densité.
- $\int \cdots F(dt) = \sum \cdots \Delta F(t)$ , si  $F$  est une fonction de saut.
- Si  $A$  est une matrice de dimension  $n \times p$  alors,
  - $A^t$  désigne la transposée de  $A$ ;
  - si  $n = p$  et si  $A$  est inversible,  $A^{-1}$  désigne l'inverse de  $A$ .
- Si  $\mathbf{a}$  est un vecteur de dimension  $p$ ,  $\mathbf{a}^{\otimes 0} = 1$ ,  $\mathbf{a}^{\otimes 1} = \mathbf{a}$  et  $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^t$ .
- càdlàg, continue à droite et limitée à gauche.

## *Typographie des mathématiques*

- caractère italique régulier, scalaire ;
- caractère italique gras, vecteur et matrice ;
- caractère italique minuscule, variable non aléatoire ;
- caractère italique majuscule, variable aléatoire ;
- caractère grec, paramètre.





# Abstract

## Log-rank-type Test for Evolution of Health Related Quality of Life

The advent of highly active antiretroviral therapies (HAART) delayed the HIV-infection pathological effects and increased life expectancy of HIV-infected patients. Consequently, health policy priorities changed in patients' follow-up. HIV-infection has entered into a new era that we can call "normalisation era" (Setbon, 2000) as now HIV-infected patients have to follow-up a scheduled long-term treatment. Actually, there is no benefit to hope in stopping antiretroviral treatment for a patient in therapeutic success. Furthermore, HAART were shown important limits, so it is widely believed that adherence to an antiretroviral regimen is central to the likelihood that a patient will derive sustained benefit from therapy (Paterson *et al.*, 2000). The variety of treatment options improves the chances of virological success for patients but it also lengthens patients' experience with treatment and their short- and long-term side-effects.

Today, it is well known that, such therapeutic options may provide equivalent levels of sustained virological success, so, the assessment of health related quality of life (HrQoL) has become a priority. Their wellbeing may be influenced not only by their response to treatment but also by treatment-related toxicity. Thereby, very powerful treatments may induce more side-effects than lightly less powerful but better bearable treatments. It is particularly the case between lopinavir/ritonavir and nelfinavir prescription (Cvetkovic and Goa, 2003).

Thus, a major issue is to know if it is always wise to optimize a very powerful treatment to the detriment of HrQoL in HIV-infected patients receiving long-term antiretroviral therapy.

Actually there is a growing number of treatment options that need to be evaluated and also compared using the number of perceived side-effects as outcome variable. Furthermore it may be important to identify when it is worthwhile switching a treatment taking into account both clinical data and patient's acceptance or to bring psychosocial support to improve patients' HrQoL? All answers to these questions ask for jointly studying the clinical status course of HIV-infected patients and their HrQoL status.

Similar situations occur for most of chronic diseases. Despite the recent interest in

both therapeutic efficacy and HrQoL some questions still arise and do need an update answer.

HrQoL has become a major issue for longitudinal clinical or epidemiological studies these last decades. It is particularly the case for chronic diseases such as HIV-infection (Badia *et al.*, 2004; Cohen *et al.*, 2004; Coplan *et al.*, 2004; Van Leth *et al.*, 2004), due to the lack of definitive cure. Long-term treatment of chronic diseases may involve some short- and long-term side-effects which can affect the HrQoL of patients. So, the aim of such studies is an epidemiological surveillance of health, including HrQoL and survival. Such surveillance is principally based on comparison of longitudinal evolution of the HrQoL between different groups of patients. Nevertheless, despite growing importance of HrQoL outcomes and more generally of Patient-Reported Outcomes in medical research (Leplège *et al.*, 1998, 1995; O'Connell *et al.*, 2003; Shahrier *et al.*, 2003; WHOQOL-HIV, 2004; Wu *et al.*, 2004), statistical methods for such longitudinal clinical or epidemiological studies still remain an issue.

The objective of this PhD Thesis is to build a valid non parametric statistical test to model jointly longitudinal data of HrQoL and long-term immunovirological success (by incorporating the dropout process). This method could be extended to both clinical trials for treatment comparisons and epidemiological studies on chronic diseases (for example, HIV-infection, hepatitis, diabetes...). This objective is really in accordance with the French National Agency for Research on AIDS and Viral Hepatitis (ANRS) major issues and fits well into main lines of the French 2007-2011 governmental plan in public health about improvement of HrQoL in patients with chronic diseases (Bas, 2007).

HrQoL is defined by the World Health Organization (1994) as a subjective complex concept reflecting the idea of well-being with respect to the context where the person lives. Usually, in longitudinal clinical studies, patients are asked to fill out questionnaires on predetermined dates  $t_j$ , so it is possible to evaluate their HrQoL  $Q_j$ . The recent concept of degradation of HrQoL can express evolution of HrQoL. This concept is not yet consensually defined, indeed the amount of degradation of HrQoL during time at date  $t_j$  can be defined either as  $D_j = (Q_0 - Q_j)Q_0^{-1}$  or as  $D_j = Q_0 - Q_j$ , i.e. either as a rate, or as a difference reflecting the evolution of HrQoL compared to an initial date (baseline)  $t_0 = 0$ . A threshold  $x$  can be chosen to define the maximum acceptable degradation of HrQoL. Patients are considered as degraded at time  $t_j$  if degradation of HrQoL  $D_j$  is higher than a rate  $x$ . So  $T(x) = \inf\{t : D_j \geq x\}$  is the first time where degradation of HrQoL occurs.

Awad *et al.* (2002); Mesbah *et al.* (2004) presented a method analyzing degradation of HrQoL by defining the time to first occurrence of the event of observing a HrQoL below a fixed rate  $x$ . So, for each value  $x$ , a time to degradation is analyzed by classical methods to compare survival between two groups of patients (Andersen *et al.*, 1993; Fleming and Harrington, 1991; Hill *et al.*, 1996; Hougaard, 2000; Kalbfleisch and Prentice, 2002; Klein and Moeschberger, 2003; Therneau and Grambsch, 2000) like Kaplan Meier estimator and log-rank test. Heutte and Huber-Carol (2002) analyzed complex evolution of HrQoL by a

semi-Markovian multi-states model. Dupuy and Mesbah (2002) proposed a joint distribution of HrQoL and survival-dropout processes in order to model relationship between time of dropout and a longitudinally measured covariate. Jin *et al.* (2006) proposed a semiparametric modeling approach to longitudinal HrQoL data. This approach handles possibly dependent terminal events. When the process of first-hitting-time of degradation of HrQoL is latent and when only death is observed Lee and Whitmore (2006) use a Wiener process. Singpurwalla (2006), using the notion of a hazard potential, rely the competing risk phenomenon (Latouche *et al.*, 2007) and the phenomenon of failures due to ageing or degradation.

Presented in the synopsis below, the PhD thesis aims to elaborate a new non parametric statistical test built on the HrQoL rate of degradation of two groups of patients and is applied on HIV-infected patients cohort.

In the first chapter, we remind some classical definitions of survival analysis and notions about censoring and truncation. Then we present the Kaplan Meier estimate and the Cox model.

In the second chapter, we give a description of longitudinal analysis and more particularly generalized linear models.

In the third chapter, we start by considering the concept of HrQoL according to definition by the World Health Organization (1994). Then we collect available statistical methods used in analyzing longitudinally HrQoL data. So we successively examine methods proposed by Awad *et al.* (2002); Heutte and Huber-Carol (2002); Jin *et al.* (2006) and briefly methods on degradation used in industrial area as well.

In the fourth chapter, we are interested in the relatively recent concept of degradation of HrQoL which is not yet consensually defined.

Let  $Q_t$ , the HrQoL score at time  $t$ , and  $Q_0$ , the baseline HrQoL score. We define, in our context, based on real HrQoL longitudinal study situations, the degradation of quality of life as the first hitting time  $t$  where the longitudinal variable below :

$$D_t = (Q_0 - Q_t)Q_0^{-1}$$

reaches a fixed chosen threshold  $x$ . This is a simple intuitive model based on the consumption of a baseline amount of HrQoL. Some authors in epidemiological or clinical studies choose the same definition without dividing by the baseline score, so  $D_t = Q_0 - Q_t$ .

**In our setting there is no distributional assumption about this degradation process.**

Time  $t \in [0, t_T]$  is supposed continuous. First the critical value  $x$  of degradation of HrQoL is supposed fixed. For patient  $i$ , let  $T_i(x)$  the date of degradation of HrQoL (survival time) and  $C_i$  the right censoring times independent from  $T_i(x)$ . So the  $n$  triplets of observation consist in  $(X_i(x), \delta_i(x), \mathbf{Z}_i)$  where  $X_i(x) = T_i(x) \wedge C_i$ ,  $\delta_i(x) = \mathbf{1}_{(T_i(x) \leq C_i)}$  and  $\mathbf{Z}_i$  the covariate for patient  $i$ .

Information included in the pair  $(X_i(x), \delta_i(x))$  is equivalent to information included in the counting process  $N_i(\cdot, x)$  and the process of censorship  $Y_i(\cdot, x)$  defined respectively by

$$N_i(t, x) = \mathbf{1}_{(X_i(x) \leq t, \delta_i(x)=1)}$$

and

$$Y_i(t, x) = \mathbf{1}_{(X_i(x) \geq t)}.$$

We introduce a generalization of the Cox model (Cox, 1972) taking into account rate  $x$  of degradation of HrQoL. The hazard function, for a patient  $i$ ,  $\lambda_i$  is defined by

$$\lambda_i(t, x/\mathbf{Z}_i(t)) = \lambda_0(t, x) \exp\{\beta' \mathbf{Z}_i(t)\}$$

where  $\beta$  is the unknown vector of regression parameters and  $\lambda_0$  is the unspecified baseline hazard function.

Under an extension of classical notations and assumptions in order to take into account a rate of degradation we obtain,

**Proposition 0.1.** *The normalized vector score process  $\{n^{-1/2} \mathbf{U}(\beta_0, t, x); 0 \leq t \leq t_T\}$  whose value at time  $t$  is*

$$\frac{1}{\sqrt{n}} \mathbf{U}(\beta_0, t, x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \{\mathbf{Z}_i(u) - \mathbf{E}(\beta_0, u, x)\} N_i(du, x),$$

converges weakly in  $(D[0, t_T])^p$  to a mean zero  $p$ -variate Gaussian process so that each component process has independent increment and the covariance function at instant  $t$  for components  $l$  and  $l'$  is

$$\{\Sigma(\beta_0, t, x)\}_{ll'} = \int_0^t \{\mathbf{v}(\beta_0, u, x)\}_{ll'} s^{(0)}(\beta_0, s) \lambda_0(u, x) du.$$

Furthermore, if  $\hat{\beta}$  is a consistent estimator of  $\beta_0$  then,

$$\sup_{0 \leq t \leq t_T} \left\| \frac{1}{n} \int_0^t \sum_{i=1}^n \mathbf{V}(\hat{\beta}, u, x) N_i(du, x) - \Sigma(\beta_0, t, x) \right\|_{\infty} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

We prove the consistency and asymptotic normality of the maximal partial likelihood estimator  $\hat{\beta}$  of  $\beta$  and then we establish the asymptotic distribution of  $\sqrt{n}(\hat{\Lambda}_0 - \Lambda_0)$ , where  $\hat{\Lambda}_0$  defined by

$$\hat{\Lambda}_0(t, x) = \int_0^t \left\{ \sum_{i=1}^n Y_i(u, x) \exp(\hat{\beta}' \mathbf{Z}_i(u)) \right\}^{-1} \left\{ \sum_{i=1}^n N_i(du, x) \right\}$$

is a generalization of the estimator proposed by Breslow (1974).

In the following, we first derive a log-rank-type test to compare two groups of patients when the rate  $x$  of degradation of HrQoL is fixed. Then the rate  $x$  is considered varying.

Let  $A$  and  $B$  be two groups of patients. Covariates are reduced to  $\mathbf{Z}_i = \mathbf{1}_{(i \in B)}$  the indicator variable of group  $B$ . We note, for a patient  $i$ , by  $\tau_i$  the realization of  $T_i(x)$  the

date of degradation of HrQoL. Let  $n_G(t, x) = \sum_{i \in A} \mathbf{1}_{(T_i(x) \geq t)}$  the number of patients at risk in group  $G = \{A, B\}$  at the instant  $t$  for a fixed rate  $x$  of degradation of HrQoL and let  $n(t, x) = n_A(t, x) + n_B(t, x)$  the whole of not degraded patients. The null hypothesis of test is  $H_0 : S_A(t, x) = S_B(t, x)$  the survival functions are identical for the two groups versus  $H_1 : S_A(t, x) \neq S_B(t, x)$ . Define the log-rank-type statistic  $U_n(x)$  and its variance estimator  $\hat{\text{Var}}(U_n(x))$

$$U_n(x) = \sum_{i=1}^n \left\{ \mathbf{Z}_i - \frac{n_B(\tau_i, x)}{n(\tau_i, x)} \right\}$$

and

$$\hat{\text{Var}}(U_n(x)) = \sum_{i=1}^n \frac{n_A(\tau_i, x)n_B(\tau_i, x)}{n^2(\tau_i, x)}.$$

The asymptotic statistic of log-rank-type test obtained for a fixed rate  $x$  of degradation of HrQoL is given as follows,

**Proposition 0.2.** *The normalized log-rank-type statistic  $n^{-1/2}U_n(x)$  converge in distribution to a  $\mathcal{N}(0, \sigma_{U(x)}^2)$ . Furthermore,  $n^{-1}\hat{\text{Var}}(U_n(x))$  converge to  $\sigma_{U(x)}^2$ .*

In a second part of this fourth chapter, we suppose that  $x$ , the rate of degradation of HrQoL, is **unknown and varying but not random**. We obtain,

**Proposition 0.3.** *The normalized log-rank-type statistic  $n^{-1/2}U_n(\cdot)$  converge in distribution to a Gaussian process of covariance function  $\Gamma$ .*

The proof of the Proposition 4.5 is based on the theory of empirical processes (Bilias *et al.*, 1997; Pollard, 1984; Shorack and Wellner, 1986) and more particularly on the general asymptotic theory for the two-parameter Cox score process with staggered entry data established by Bilias *et al.* (1997).

Unfortunately, no explicit estimator  $\hat{\Gamma}$  of  $\Gamma$  is available.

So in order to obtain an approximation of the supremum test statistic  $\sup_{0 < x < 1} |n^{-1/2}U_n(x)|$  we use a graphical perturbation method. Such procedure was previously used to analyze model goodness of fit. Similar graphical perturbation techniques have been developed for Generalized Linear Models (GLM) with independent responses by Su and Wei (1991), for the proportional hazards model with censored survival data by Lin *et al.* (1993), for marginal GLM with dependent response by Lin *et al.* (2002), for generalized linear mixed models by Pan and Lin (2005) and for the Mizon-Richard test procedure by Martinussen *et al.* (2008).

The supremum test statistic  $\sup_{0 < x < 1} |n^{-1/2}U_n(x)|$  is approximated via simulation. Specifically, we consider  $\xi = (\xi_1, \dots, \xi_n)$  a vector of  $n$  independent and identically distributed random variables of  $\mathcal{N}(0, 1)$  and we disrupt the normalized log-rank-type statistic like that,

$$\frac{1}{\sqrt{n}}U_n^*(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \int_0^t \{ \mathbf{Z}_i - \bar{Z}(t, x) \} d\hat{M}_i(t, x).$$

We draw realizations of the disrupted normalized log-rank-type statistic and calculate the statistics on each of many realizations. The  $p$ -value is the proportion of simulated statistics

that exceed the observed test statistic absolute value. Consequently, we obtain the global log-rank-type test.

**Proposition 0.4 (global log-rank-type test).** *Consider  $S \equiv \sup_{0 < x < 1} |n^{-1/2}U_n(x)|$  the supremum normalized log-rank-type statistic and suppose that  $s$  is the observed value of  $S$ . Let  $\hat{S} = \sup_{0 < x < 1} |n^{-1/2}U_n^*(x)|$  the supremum of disrupted normalized log-rank-type statistic. The  $p$ -value of the global log-rank-type test  $\mathbb{P}(S > s)$  can be approximated by  $\mathbb{P}(\hat{S} > s)$ , which can be estimated by a great number of realizations (say 1000 or 10000) from  $\hat{S}$ .*

In the fifth chapter, we realize some simulations to compare performance of the test previously described, i.e. the global log-rank-type test. On purpose, we implement this non parametric test in SAS<sup>®</sup>. Codes of programmes in SAS<sup>®</sup> are shown in the appendix. To that end different data sets which represent longitudinal HrQoL evolution are simulated. We consider that patients' HrQoL evolves linearly. We also consider two scenarios to convey variability of patients' answers compared with the theoretical straight line. In the first one, slope and initial score of HrQoL are disrupted by a white noise consisting of both random effect and fixed effect, whereas in the other one, slope and initial score of HrQoL are disrupted by an autoregressive process (AR(1)) (Doob, 1953; Bosq and Lecoutre, 1987). We take into account missing data and dropout. Furthermore, we consider two cases, in the first one evolutions of the two groups of patients are different whereas in the second we make the null hypothesis  $H_0$ . We examine different situations in varying one of the following parameters : size of sample, number of dates, percentage of each groups, percentage of dropout and missing data and number of simulated curves. We also consider each definition of degradation of HrQoL,  $D_t = (Q_0 - Q_t)Q_0^{-1}$  and  $D_t = Q_0 - Q_t$  and we refer to ANRS C08 data for the value of the different parameters. Simulations show that the test is operational.

In the last part, we apply this proposed log-rank-type test to the French National Agency for Research on AIDS and Viral Hepatitis (ANRS) C08 Cohort (Protopopescu *et al.*, 2007). The French ANRS C08 multicenter Cohort study was established in 1997 to study clinical, immunologic, virologic and socio-behavioral characteristics in HIV-1-infected patients who were beginning a combination antiretroviral therapy (HAART) that included a protease inhibitor (PI) (Carrieri *et al.*, 2006; Protopopescu *et al.*, 2007). Entry in the cohort corresponds to the beginning of combination therapy. Follow-up of longitudinal HrQoL of patients is evaluated by a self-administered questionnaire. During follow-up in the first period the self-administered questionnaire was the generic MOS-SF 36 replaced by the specific WHOQOL-HIV afterwards. For this reason, to obtain an only score of HrQoL at each date we carry out equating methods.

For all variables studying mental HrQoL has increased during the 28 first months and has remained relatively stable afterwards. Variable born in European Union appears like a predictive factor of mental HrQoL. Indeed, patients born in European Union have a better mental HrQoL at inclusion but the inverse phenomenon is recorded from M28.

Variables high-school education and having children seem not to be predictive factors associated with mental HrQoL. Demographic variable gender gives harsh results according to statistical test used, the reason is perhaps the low percentage of women (less than 20%) in this study. Economical variables full or part time employment and stable housing seem not to be predictive factors for mental HrQoL. Nevertheless, variable comfortable housing appears like a predictive factor associated with mental HrQoL. This result is in accordance with Préau *et al.* (2006). The importance of having main partner on evolution of mental HrQoL of HIV-infected patients is shown. Furthermore the support of this main partner is essential, but support of family members and friends seem not to be predictive factors of mental HrQoL. Among viro-immunological and medical variables only the factor antiretroviral naivety is significant on mental HrQoL of HIV-infected patients.

This PhD thesis has been supported by ANRS, a part of this work can be found in the paper Boisson and Mesbah (2008) and also presented during conferences (Boisson, 2005, 2008; Boisson *et al.*, 2006; Mesbah *et al.*, 2008; Boisson *et al.*, 2008b,c). Furthermore, an article has recently been submitted (Boisson *et al.*, 2008a).





**Première partie**

**Préliminaires**



# Chapitre 1

## Introduction à l'analyse de survie

En épidémiologie ou lors d'essais cliniques, l'analyse de survie i.e. la modélisation du temps de survenue d'un événement, apporte un outil principal d'évaluation thérapeutique. L'analyse des données de survie possède deux particularités intrinsèques, d'une part, celle-ci ne concerne que des *variables aléatoires positives* et d'autre part, la présence de *données censurées* est possible. Une conséquence est l'introduction de méthodes spécifiques comme le modèle semi-paramétrique de Cox.

Dans le but de fixer les notations, ce mémoire débute par un rappel des différentes fonctions utilisées en analyse de survie, sans pour autant prétendre à une quelconque exhaustivité. Nous rappelons ensuite les différents schémas de censures ainsi que la constitution des observations ceci nous permettant d'introduire l'estimateur de Kaplan-Meier. Enfin, nous introduisons le modèle semi-paramétrique de Cox.

### 1.1 Fonctions de survie et de risque

Soit  $T$  une variable aléatoire positive définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et représentant le temps écoulé jusqu'à la survenue de l'événement d'intérêt (comme par exemple le décès, l'apparition de symptômes, la guérison, ...). Par la suite, cette variable aléatoire sera appelée **durée de vie**. Sa fonction de répartition  $F(t) = \mathbb{P}(T \leq t)$  représente la probabilité que l'événement d'intérêt se produise durant l'intervalle de temps  $[0, t]$ .

Notons que chaque individu est susceptible de subir une fois et une seule l'événement d'intérêt. De plus, l'observation de la survenue ou non de cet événement constitue la donnée basale pour une modélisation des durées de survie. L'interprétation des durées de survie a conduit à l'introduction de nouvelles fonctions permettant de mieux caractériser sa loi.

La **fonction de survie**  $S(t)$  relative à une variable aléatoire de durée de vie  $T$  est la probabilité que l'événement survienne après l'instant  $t$ . Cette fonction est définie par

$$S(t) = 1 - F(t) = \mathbb{P}(T > t).$$

Par définition,  $S$  est décroissante, continue à droite et limitée à gauche avec  $\lim_{t \rightarrow 0} S(t) = 1$  et  $\lim_{t \rightarrow +\infty} S(t) = 0$ .

Supposons que la loi de  $T$  admette une densité  $f$ , par rapport à la mesure de Lebesgue, nous pouvons définir sa **fonction de risque instantané**,

$$\lambda(t) = \frac{f(t)}{S(t)}.$$

De façon heuristique, la fonction de risque instantané représente la probabilité de subir l'événement à l'instant  $t$  sachant que l'on n'a pas encore subi l'événement à l'instant  $t-$  i.e. juste avant l'instant  $t$ , autrement dit,

$$\forall t \geq 0, \lambda(t) = \lim_{dt \rightarrow 0^+} \frac{\mathbb{P}(t \leq T < t + dt / T \geq t)}{dt}. \quad (1.1)$$

Par conséquent, la fonction de risque instantané traduit l'évolution longitudinale du risque de survenue de l'événement. C'est ainsi que l'allure de cette fonction apporte une information immédiate sur les caractéristiques de l'objet étudié; les parties croissantes de la courbe représentative de  $\lambda$  correspondant à un phénomène de vieillissement, les parties décroissantes à un phénomène de rodage et les parties planes à un phénomène sans vieillissement.

La **fonction de risque cumulé** est donnée, pour tout  $t \in \mathbb{R}^+$ , par

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

La définition de la distribution de probabilité de  $T$  repose de manière équivalente sur sa densité, sa fonction de survie, sa fonction de risque instantané ou sa fonction de risque cumulé (sous réserve de leur existence). En effet, les relations suivantes entre les différentes fonctions définies peuvent être établies. Pour  $t \geq 0$ ,

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{S(t)} = -\frac{d}{dt} \log(S(t)), \\ \Lambda(t) &= \int_0^t \lambda(u) du = -\log(S(t)), \\ S(t) &= \exp(-\Lambda(t)). \end{aligned}$$

Bien que toutes les fonctions caractérisent à elles seules la loi de l'événement d'intérêt, la fonction de risque instantané  $\lambda$  est la plus intéressante. En effet, celle-ci est une description probabiliste du futur immédiat du patient *encore vivant*.

## 1.2 Données manquantes, censures et troncatures

### 1.2.1 Données manquantes

Dans de nombreuses applications statistiques, nous sommes amenés à faire intervenir une variable de durée, c'est-à-dire un délai séparant un instant initial de l'instant où un événement final est observé. C'est notamment le cas si nous souhaitons observer l'effet d'un traitement donné sur une population donnée d'individus. Or, un des problèmes majeurs

dans l'analyse des données longitudinales est causé par les observations incomplètes. De plus, la validité de l'analyse des données en présence d'observations incomplètes dépend du mécanisme associé aux données manquantes. D'après Little and Rubin (1987), différents mécanismes de données manquantes existent, ayant nécessité une classification en trois processus d'abandon.

- **Manquante complètement au hasard (MCAR)**, lorsque le phénomène de réponse est indépendant des données observées et non observées de la variable longitudinale. Par exemple, c'est le cas lorsque l'on fait remplir un auto-questionnaire, à un patient ne remplissant pas un questionnaire mais remplissant les autres questionnaires ;
- **Manquante au hasard (MAR)**, lorsque le phénomène d'abandon est indépendant des données non observées mais dépend de celles observées ;
- **Manquante informative (ou non-ignorable)**, lorsque le phénomène d'abandon dépend des données non observées. Par exemple, lorsque l'on fait remplir un auto-questionnaire, les patients quittant l'étude précocément, peuvent être ceux ayant une moins bonne qualité de vie.

Récemment, un certain nombre de méthodes ont été proposées lorsque les abandons sont non-informatifs pour des données longitudinales. Diggle and Kenward (1994) ont combiné un modèle linéaire permettant de prendre en considération plusieurs covariables pour le processus longitudinal avec un modèle de régression logistique pour l'abandon. Ce modèle logistique tolère la dépendance de l'abandon sur l'observation manquante au temps d'abandon. Molenberghs *et al.* (1997) ont adopté une approche similaire pour les données ordinales longitudinales. Dans d'autres situations, l'abandon dépend du temps plutôt que de la valeur non observée de la covariable. Une relation peut être établie entre le temps d'abandon et les résultats longitudinaux étudiés au travers des effets individuels aléatoires, permettant ainsi de représenter le processus longitudinal. Ceci conduit à des modèles de sélection à coefficients aléatoires (De Gruttola and Tu, 1994; Ribaud *et al.*, 2000; Schuluchter, 1992; Wu and Carroll, 1988). Une catégorie alternative de modèles de distribution conjointe de mesures répétées et de temps d'abandon est appelée modèle-mixte (Hogan and Laird, 1997a; Little, 1995). Cette approche stratifie l'échantillon par temps d'abandon, puis modélise la distribution des mesures répétées en chaque état. Des travaux détaillés de ces diverses approches peuvent être trouvés dans Little and Rubin (1987); Hogan and Laird (1997b); Billot and Mesbah (1999); Verbeke and Molenberghs (2000).

Supposer le processus d'abandon non-informatif n'est généralement pas une hypothèse crédible en recherche clinique. C'est notamment le cas dans l'infection par le VIH où les patients ayant une mauvaise qualité de vie sont susceptibles d'avoir plus de difficultés à compléter les auto-questionnaires administrés. Les patients quittant précocément l'étude sont ceux ayant une moins bonne qualité de vie, le processus d'abandon est informatif. Pour pallier ce problème, Dupuy and Mesbah (2002) ont construit un modèle de Cox pour des données manquantes informatives et dépendant du temps.

Après un résumé de différentes approches proposées pour traiter les abandons, Diggle *et al.* (2007) proposent un modèle linéaire dynamique permettant de compléter la variable

par les incréments escomptés. Les effets aléatoires spécifiques à un patient sont, en l'absence d'abandon, une martingale. De plus, des procédures pour diagnostiquer la crédibilité des hypothèses sont fournies.

Par ailleurs, nous trouvons principalement deux phénomènes dont les effets se traduisent par des données incomplètes ou observées partiellement (Huber-Carol, 2000; Andersen *et al.*, 1993) correspondant effectivement à deux situations distinctes : la censure et la troncature.

### 1.2.2 Schémas de censure

La **variable de censure**, définie par la possibilité de *non-observation* de l'événement, permet de modéliser l'information exacte apportée. Celle-ci se décompose en trois catégories de modèles.

- **Censure à droite**, lorsque nous observons la censure  $C$ , et non la durée de vie  $T$ , et que nous savons que  $T > C$ . Ce modèle est adapté au cas où l'événement considéré est le décès du patient et où la date de fin d'étude est préalablement fixée ;
- **Censure à gauche**, lorsque nous observons la censure  $C$ , et non la durée de vie  $T$ , et que nous savons que  $T < C$ . Un phénomène symétrique au précédent se produit, le patient a déjà subi l'événement lors du bilan. Ce modèle est adapté au cas où l'incorporation d'un patient dans une étude est conditionnée par un *événement initial* ;
- **Censure par intervalles**, lorsque la censure à droite et la censure à gauche sont conjuguées. Dans ce cas l'information apportée par l'expérience se traduit par une appartenance de la durée de vie à un intervalle de temps ( $C_1 < T < C_2$ ). Ce modèle est adapté au cas de suivis périodiques de patients.

Les modèles de censures peuvent être affinés suivant le mécanisme de censure existant. Les différents modes de censure, décrits dans le cadre d'une censure à droite, sont les suivants,

- **La censure non-aléatoire de type I**, étant donné un nombre positif fixé  $c$  et un  $n$ -échantillon  $T_1, \dots, T_n$ , les observations consistent en  $(X_i, \delta_i)$ , où  $X_i = T_i \wedge c$  et  $\delta_i = \mathbf{1}_{\{T_i \leq c\}}$ . Ce modèle souvent utilisé dans les études épidémiologiques, correspond à l'observation de la durée de survie de  $n$  patients au cours d'une expérience de durée prédéterminée  $c$  ;
- **La censure aléatoire de type I**, étant donné un  $n$ -échantillon  $T_1, \dots, T_n$ , il existe une v.a.  $n$ -dimensionnelle  $(C_1, \dots, C_n)$  de  $\mathbb{R}_+^n$  telle que les observations consistent en  $(X_i, \delta_i)$ , où  $X_i = T_i \wedge C_i$  et  $\delta_i = \mathbf{1}_{\{T_i \leq C_i\}}$ . Ce modèle est typiquement utilisé pour les essais thérapeutiques. Nous nous intéressons à une cause de décès ayant lieu au bout d'un temps  $T$  et nous désirons connaître la loi de  $T$  ; cependant, une autre cause aléatoire (décès dû à une autre cause, abandon ...) peut survenir auparavant et par conséquent empêcher l'observation de  $T$  par un mécanisme de censure aléatoire  $C$  ;
- **La censure de type II**, étant donné un nombre positif fixé  $r$  et un  $n$ -échantillon  $T_1, \dots, T_n$ , les observations consistent en  $(X_i, \delta_i)$ , où  $X_i = T_i \wedge T_{(r)}$  et  $\delta_i = \mathbf{1}_{\{T_i \leq T_{(r)}\}}$

avec  $T_{(1)} < T_{(2)} < \dots < T_{(n)}$  sont les statistiques d'ordre. Ce modèle souvent utilisé dans les études de fiabilité, correspond à l'observation de la durée de fonctionnement de  $n$  machines tant que  $r$  d'entre elles ne sont pas tombées en panne.

Désormais, nous nous plaçons dans le cadre d'une **censure à droite, aléatoire de type I et indépendante de la date de survenue de l'événement d'intérêt** (*censure non-informative* ou *ignorable*); ce cadre correspondant à un modèle fréquemment employé dans la pratique.

### 1.2.3 Troncatures

Nous parlons de **troncature** à droite (respectivement à gauche) lorsque la variable d'intérêt d'un patient, n'est pas observable quand elle est supérieure (respectivement inférieure) à un seuil  $c$  fixé.

Dans le cas de la censure la variable  $C$  est observée alors que dans le cas de la troncature l'analyse porte uniquement sur la loi de  $T$  conditionnellement à l'événement  $\{T < c\}$  (respectivement  $\{T > c\}$ ). Pour un exemple de données tronquées nous pouvons consulter (Andersen *et al.*, 1993, p. 152) et pour un exemple de données censurées à gauche nous pouvons consulter Asselineau *et al.* (2007).

## 1.3 Processus de comptage et constitution des observations

Soit un  $n$ -échantillon soumis à une censure à droite. Pour un individu  $i$  nous notons respectivement par  $T_i$  et  $C_i$  les variables aléatoires de durée de vie et de censure. Une observation consiste en  $(X_i, \delta_i)$ , où  $\delta_i = \mathbf{1}_{\{T_i \leq C_i\}}$  est la fonction indicatrice de l'observation réelle de l'événement d'intérêt et  $X_i = T_i \wedge C_i$  est la durée effectivement observée nommée *temps de participation*. L'étude des durées de survie peut être abordée d'une autre façon en représentant l'expérience, pour un individu  $i$  et  $t \geq 0$ , par les processus suivants

$$\begin{cases} N_i(t) = \mathbf{1}_{\{X_i \leq t, \delta_i = 1\}}; \\ Y_i(t) = \mathbf{1}_{\{X_i \geq t\}}. \end{cases}$$

$N_i(t)$  est un **processus de comptage** indiquant si l'événement a été observé avant l'instant  $t$ , tandis que le processus  $Y_i(t)$  caractérise si le sujet  $i$  est *à risque* (i.e. n'a pas subi l'événement) juste avant l'instant  $t$ .

Notons par  $\sigma\{N\}$  la tribu engendrée par  $N$ . D'après (1.1), nous obtenons  $\mathbb{P}(dN_i(t) = 1/\mathcal{F}_{t-}) = \lambda_i(t)Y_i(t)dt$ , où  $\mathcal{F}_t = \sigma\{N_i(u); 0 \leq u \leq t, 1 \leq i \leq n\}$  est la filtration naturelle (i.e. toute l'information disponible à l'instant  $t$ ). Cette relation nous permet d'obtenir la proposition suivante.

**Proposition 1.1.** *Le processus stochastique défini par  $M_i(t) = N_i(t) - \int_0^t \lambda_i(s)Y_i(s)ds$  est une martingale.*

Les processus aléatoires  $\alpha_i(t) = \lambda_i(t)Y_i(t)$  et  $A_i(t) = \int_0^t \lambda_i(s)Y_i(s)ds$  sont désignés respectivement **processus d'intensité** et **processus d'intensité cumulée** de  $N_i$ .  $A_i$  est



également appelé **compensateur** du processus  $N_i$ . Notons que le processus  $N_i$  ne définit pas une martingale.

## 1.4 L'estimateur de Kaplan-Meier

Un estimateur de la fonction de survie, tenant un rôle essentiel dans le cadre des données censurées, est l'estimateur non-paramétrique introduit par Kaplan and Meier (1958). L'**estimateur de Kaplan-Meier** est défini par

$$\hat{S}(t) = \prod_{s \leq t} \left( 1 - \frac{\mathbf{1}_{\{Y(t)\}} \Delta N(s)}{Y(s)} \right),$$

où  $N(t) = \sum_{i=1}^n N_i(t)$ ,  $Y(t) = \sum_{i=1}^n Y_i(t)$  et pour un processus  $X(t)$  càdlàg (continu à droite et limité à gauche)  $\Delta X(t) = X(t) - X(t-)$ .

L'estimateur de Kaplan-Meier est encore appelé *Produit-Limite* car il s'obtient comme la limite d'un produit. En effet, il est fondé sur l'idée intuitive qu'être en vie après l'instant  $t$ , c'est être vivant juste avant l'instant  $t$  et ne pas décéder à l'instant  $t$ . Cette idée se traduit en terme probabiliste par

$$\begin{aligned} S(t) &= \mathbb{P}(T \geq t) \\ &= \mathbb{P}(T \geq t/T \geq t-1) \mathbb{P}(T \geq t-1) \\ &= \dots \\ &= \mathbb{P}(T \geq t/T \geq t-1) \dots \mathbb{P}(T \geq 1/T \geq 0) \mathbb{P}(T \geq 0). \end{aligned}$$

## 1.5 Le modèle semi-paramétrique de Cox

Nous définissons tout d'abord les notations utilisées puis formulons le modèle de Cox. Nous considérons un échantillon de  $n$  individus. Les variables  $T_1, \dots, T_n$  et  $C_1, \dots, C_n$  représentant respectivement les dates de survenue de l'événement et les dates de censure correspondantes sont supposées indépendantes. Nous observons la suite des  $n$  couples de variables  $(X_i, \delta_i)$ , avec  $X_i = T_i \wedge C_i$  et  $\delta_i = \mathbf{1}_{\{T_i \leq C_i\}}$ , ainsi que pour chaque individu  $i$  un ensemble de  $p$  covariables  $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})$ . La durée de survie  $T_i$  dépend des covariables  $\mathbf{Z}_i$ . Soit  $\beta = (\beta_1, \dots, \beta_p)$  le vecteur de dimension  $p$  des paramètres de régression. Le **modèle de Cox (1972)** spécifie que le risque instantané s'écrit

$$\lambda_i(t) = \lambda_0(t) \exp(\beta^t \mathbf{Z}_i),$$

où  $\lambda_0$  est la **fonction de risque de base**.

Le modèle paramétrique de Cox est fréquemment employé lorsque nous cherchons à évaluer l'effet de variables explicatives (*covariables*) sur la durée de survie. En effet, ce modèle exprime, par l'intermédiaire d'une fonction de risque instantané, la relation entre le risque d'occurrence d'un événement (tel qu'un décès, la survenue d'une pathologie, ...) et des covariables. Il s'agit d'un **modèle semi-paramétrique à risques proportionnels**,

- *semi-paramétrique*, du fait de la présence, dans la définition du risque instantané, d'une partie paramétrique (la partie de régression) et d'une partie non-paramétrique (le risque de base) ;
- *à risques proportionnels*, car quels que soient les individus  $i$  et  $j$ , le rapport des risques instantanés ne varie pas au cours du temps. En effet,

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \exp[\beta^t(\mathbf{Z}_i - \mathbf{Z}_j)].$$

Afin d'éliminer le paramètre de nuisance que représente la fonction de risque de base, Cox (1972) considère la **vraisemblance partielle** suivante

$$\mathcal{L}(\beta) = \prod_{i=1}^n \prod_{t \geq 0} \left( \frac{Y_i(t) \exp(\beta^t \mathbf{Z}_i)}{\sum_{j=1}^n Y_j(t) \exp(\beta^t \mathbf{Z}_j)} \right)^{\delta_i}.$$

Nous pouvons montrer que la vraisemblance partielle de Cox a les mêmes propriétés qu'une vraisemblance ordinaire (Andersen and Gill, 1982; Andersen *et al.*, 1993). En temps continu, nous faisons l'hypothèse qu'il n'y a aucun ex-æquo.

Le modèle de Cox a été généralisé à des covariables  $\mathbf{Z}_i$  pouvant dépendre du temps. Dans ce cas, la fonction de risque instantané s'écrit sous la forme suivante,

$$\lambda_i(t) = \lambda_0(t) \exp(\beta^t \mathbf{Z}_i(t)).$$

Le lecteur intéressé par le modèle de Cox avec des covariables pouvant dépendre du temps peut se reporter aux références suivantes Kalbfleisch and Prentice (2002); Klein and Moeschberger (2003); Therneau and Grambsch (2000).



## Chapitre 2

# Introduction à l'analyse longitudinale

L'étude de données longitudinales (répétées) (Diggle *et al.*, 2002; Lindsey, 1993; Verbeke and Molenberghs, 2000) peut être un outil très puissant en permettant d'analyser les modifications inhérentes à un patient au cours du temps. Les mesures répétées possèdent la propriété intrinsèque que les mesures correspondant à un même patient sont corrélées entre elles. Par suite, l'hypothèse de la plupart des techniques statistiques standards, i.e. l'indépendance des observations, est caduque pour les mesures répétées et l'utilisation de ces techniques statistiques pour les mesures répétées conduit à des inférences biaisées (Chavance, 1999). Lorsque nous nous intéressons explicitement à la variabilité inter-individuelle, il est préférable d'utiliser les modèles mixtes plutôt que les modèles marginaux où l'espérance et la variance des observations sont modélisées séparément.

Dans ce chapitre, nous décrivons le modèle linéaire mixte d'abord dans le cadre univarié puis dans le cadre multivarié.

### 2.1 Modèle linéaire mixte

#### 2.1.1 Formulation du modèle

Le **modèle linéaire mixte** introduit par Laird and Ware (1982) est une méthode très répandue pour l'analyse longitudinale de données cliniques et épidémiologiques. Celui-ci consiste à prendre en considération la corrélation entre les mesures répétées des patients pour l'estimation des paramètres de la population, i.e. l'évolution moyenne au cours du temps et de l'association avec les variables explicatives (covariables).

Soient une population de  $N$  patients et  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$  les mesures répétées du patient  $i$ . Le modèle linéaire mixte s'énonce, pour le patient  $i$ , de la manière suivante,

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\gamma_i + \varepsilon_i \text{ avec } \begin{cases} \gamma_i \sim \mathcal{N}(0, \mathbf{G}) \\ \varepsilon_i \sim \mathcal{N}(0, \mathbf{\Sigma}_i) \\ \forall i, \gamma_i \perp \varepsilon_i \end{cases} \quad (2.1)$$

où  $\mathbf{X}_i$  est la matrice des variables explicatives (covariables) de dimension  $n_i \times p$ ,  $\mathbf{Z}_i$  est une sous-matrice de  $\mathbf{X}_i$  de dimension  $n_i \times q$  et  $\varepsilon_i$  est le vecteur d'erreurs du patient  $i$ .  $p$  représente le nombre de variables explicatives et par suite le nombre d'effets fixes  $\beta$ , tandis que  $q$  représente le nombre d'effets aléatoires  $\gamma_i$ . Notons que  $q \leq p$ . De plus, effets aléatoires et erreurs sont indépendants d'un sujet à l'autre et les erreurs sont indépendantes d'une mesure à l'autre chez un même patient. Ceci se formalise,

$$\begin{cases} \forall i, j = 1, \dots, N, \gamma_i \perp \gamma_j \\ \forall i, j = 1, \dots, N, \varepsilon_i \perp \varepsilon_j \\ \forall j, k = 1, \dots, n_i, \varepsilon_{ij} \perp \varepsilon_{ik}. \end{cases}$$

Contrairement aux paramètres fixes  $\beta$ , les effets aléatoires ne sont pas estimables et c'est pourquoi nous supposons que l'ensemble des effets aléatoires  $\gamma_i$  suit une loi normale centrée de matrice de covariance  $\mathbf{G}$ . Par ailleurs, nous supposons généralement que l'erreur de mesure est indépendante pour chaque observation suivant une loi normale centrée. Cela revient à choisir comme matrice de variance de l'erreur de mesure  $\varepsilon_i$  pour l'ensemble des observations d'un patient  $i$ , la matrice diagonale  $\sigma_i = \sigma_\varepsilon^2 I_{n_i}$ .

Contrairement au modèle linéaire classique, qui suppose que la seule source d'aléas provient des erreurs des mesures, le modèle linéaire mixte pose l'hypothèse qu'il existe également une variation individuelle. Ce modèle repose sur le principe de l'existence d'une hétérogénéité naturelle au sein des patients qu'il faut prendre en compte en déterminant des effets aléatoires individuels. Ainsi, le modèle linéaire mixte fait intervenir à la fois des paramètres associés à la population et des paramètres propres à chaque patient ou à un groupe de patients. Par conséquent, le modèle linéaire mixte induit une structure de covariance inter-individuelle.

Nous parlons de modèle linéaire mixte car ce modèle est composé des deux parties :

- les **effets fixes**  $\beta$  identiques pour tous les patients de la population considérée, représentant la tendance moyenne pour la population ;
- les **effets aléatoires**  $\gamma_i$  représentant l'écart de chaque individu  $i$  par rapport à la tendance de la population étudiée.

Lorsque l'évolution des observations des mesures observées chez un patient  $i$  correspond à la réalisation d'un processus stochastique dépendant du temps et propre à chaque individu, nous pouvons concevoir que la seule erreur de mesure ne suffise pas pour modéliser l'ensemble de la variabilité intra-individuelle. Classiquement, cette corrélation diminue au cours du temps. Aussi, pour tenir compte de cette corrélation dite **sérielle** (ou **auto-corrélation**), Diggle (1988) puis Jones and Boadi-Boateng (1991) ont proposé une extension du modèle de Laird and Ware (1982) en incorporant un processus auto-régressif d'ordre 1.

### 2.1.2 Estimation des paramètres

Le modèle linéaire mixte peut se réécrire sous la **forme hiérarchique** suivante

$$\mathbf{Y}_i | \gamma_i \sim \mathcal{N}(\mathbf{X}_i \beta + \mathbf{Z}_i \gamma_i, \boldsymbol{\Sigma}_i) \text{ avec } \gamma_i \sim \mathcal{N}(0, \mathbf{G})$$

mais également sous la **forme marginale** suivante

$$\mathbf{Y}_i \sim \mathcal{N}(\mathbf{X}_i \beta, \mathbf{V}_i = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^t + \boldsymbol{\Sigma}_i)$$

L'écriture marginale est utilisée pour estimer les paramètres du modèle à savoir le vecteur des effets fixes  $\beta$ , les éléments de la matrice  $\mathbf{G}$  des effets aléatoires et les éléments de la matrice  $\boldsymbol{\Sigma}_i$  des erreurs de mesure pour le patient  $i$ . En effet, par la méthode du maximum de vraisemblance nous obtenons la vraisemblance marginale suivante

$$\mathcal{L}(\beta, \mathbf{G}, \boldsymbol{\Sigma}_i) = \prod_{i=1}^N \left( \frac{1}{2\pi} \right)^{n_i/2} |\mathbf{V}_i|^{-1/2} \exp \left( -\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \beta)^t \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta) \right).$$

L'estimateur du maximum de vraisemblance de  $\beta$  est

$$\hat{\beta} = \left( \sum_{i=1}^N \mathbf{X}_i^t \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i^t \mathbf{V}_i^{-1} y_i$$

Les effets aléatoires peuvent être estimés en utilisant la théorie bayésienne (Smith, 1973),

$$\hat{\gamma}_i = \mathbb{E}[\gamma_i | \mathbf{Y}_i = \mathbf{y}_i] = \mathbf{G} \mathbf{Z}_i^t \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta).$$

Le modèle linéaire mixte est devenu une méthode classique pour étudier les variations longitudinales de variables quantitatives. Ceci s'explique par la disponibilité des procédures et au développement de leur utilisation (Littell *et al.*, 1996; Singer, 1998; Thiébaud *et al.*, 2002). Les estimateurs du maximum de vraisemblance des paramètres sont obtenus en maximisant le logarithme de la vraisemblance. Deux algorithmes sont couramment employés : l'algorithme EM (Dempster *et al.*, 1978) ainsi que l'algorithme de Newton-Raphson (Fletcher, 2000). La procédure la plus connue est la procédure MIXED implémentée sous SAS<sup>®</sup>. Cette procédure utilise l'algorithme de Newton-Raphson plus rapide que l'algorithme EM (Lindstrom and Bates, 1988).

## 2.2 Modèle linéaire mixte pour données longitudinales multivariées

Dans un souci de lisibilité, nous développons uniquement le cas bivarié, i.e. avec deux covariables mesurées longitudinalement, car la méthodologie est analogue pour un nombre supérieur de covariables.

Soient  $\mathbf{Y}_i^1 = (Y_i^{11}, \dots, Y_i^{1n_i})$  et  $\mathbf{Y}_i^2 = (Y_i^{21}, \dots, Y_i^{2n_i})$  les vecteurs d'observation pour le patient  $i$ . Le modèle linéaire mixte défini en 2.1 s'écrit pour chaque covariable  $k = 1, 2$

$$\mathbf{Y}_i^k = \mathbf{X}_i^k \beta^k + \mathbf{Z}_i^k \gamma_i^k + \varepsilon_i^k$$

où les matrices  $\mathbf{X}_i^k$  et  $\mathbf{Z}_i^k$  sont des matrices de variables explicatives associées respectivement au vecteur des effets fixes  $\beta^k$  et au vecteur des effets aléatoires  $\gamma_i^k$ ;  $\varepsilon_i^k$  représente le vecteur des erreurs de mesure. En s'intéressant au vecteur de réponses conjoint  $\mathbf{Y}_i = (\mathbf{Y}_i^{1t}, \mathbf{Y}_i^{2t})^t$ , le modèle linéaire mixte bivarié est défini de la façon suivante

$$\mathbf{Y}_i = \begin{pmatrix} \mathbf{Y}_i^1 \\ \mathbf{Y}_i^2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_i^1 & 0 \\ 0 & \mathbf{X}_i^2 \end{pmatrix} \begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_i^1 & 0 \\ 0 & \mathbf{Z}_i^2 \end{pmatrix} \begin{pmatrix} \gamma_i^1 \\ \gamma_i^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \end{pmatrix}$$

avec  $\mathbf{X}_i^k$  une matrice de variables explicatives de dimension  $n_i \times p^k$  ( $k = 1, 2$ ),  $\beta^k$  le vecteur des effets fixes de dimension  $p^k$ ,  $\mathbf{Z}_i^k$  une matrice de dimension  $n_i \times q^k$  des variables explicatives pour les effets aléatoires  $\gamma_i^k$ , de dimension  $q^k$  vérifiant  $q^k \leq p^k$ .

L'écriture de ce modèle linéaire mixte bivarié nous permet de retrouver l'expression du modèle linéaire mixte classique  $\mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{Z}_i \gamma_i + \varepsilon_i$  pour lequel,

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{X}_i^1 & 0 \\ 0 & \mathbf{X}_i^2 \end{pmatrix}, \mathbf{Z}_i = \begin{pmatrix} \mathbf{Z}_i^1 & 0 \\ 0 & \mathbf{Z}_i^2 \end{pmatrix}, \beta = \begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix}, \gamma_i = \begin{pmatrix} \gamma_i^1 \\ \gamma_i^2 \end{pmatrix} \text{ et } \varepsilon = \begin{pmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \end{pmatrix}.$$

La vraisemblance du modèle linéaire mixte bivarié s'écrit de manière similaire à celle d'un modèle linéaire mixte classique et l'estimation des paramètres peut s'obtenir par des procédures identiques à celles utilisées dans le cadre d'un modèle linéaire mixte classique.

Le modèle linéaire mixte est devenu un outil classique en recherche clinique et en épidémiologie car servant à l'étude des changements au cours du temps de covariables. Des procédures d'estimation sont disponibles dans tous les logiciels de statistiques et en particulier pour le logiciel SAS<sup>®</sup> avec la procédure MIXED.

Thiébaud and Jacqmin-Gadda (2004) ont implémenté en SAS un programme permettant la prise en considération des censures à gauche dans le modèle linéaire mixte et présentent une application dans le cadre de l'infection à VIH.

## Chapitre 3

# Qualité de vie et analyse de survie

### 3.1 Définition de la qualité de vie

Les analyses de qualité de vie (QdV) ont commencé à susciter un important intérêt et essor à partir des années 1970 en raison des progrès thérapeutiques ayant contribué à l'augmentation de l'espérance de vie. C'est notamment le cas dans la lutte contre les maladies infectieuses. Parallèlement, certaines maladies sont devenues chroniques, comme les maladies dégénératives et plus tardivement, avec l'arrivée des multithérapies, l'infection à VIH. Les traitements non curatifs de ces maladies chroniques entraînant bien souvent des effets indésirables graves ou invalidants sont jugés sur leur capacité à restaurer ou à préserver la qualité de vie des patients. D'autre part, lorsque l'efficacité clinique des différentes stratégies thérapeutiques concurrentes pour le traitement d'une même pathologie est similaire la qualité de vie des patients est davantage prise en compte par le prescripteur. Ainsi, avec l'introduction des mesures de qualité de vie, le point de vue de la médecine s'est élargi en ne considérant non plus que son objet traditionnel : la santé, mais également la qualité de vie des patients ; c'est particulièrement vrai pour l'infection à VIH.

Dans une perspective de santé publique visant à promouvoir la santé comme « *un état complet de bien-être physique, psychologique et social, et pas seulement comme l'absence de maladie* », l'Organisation Mondiale pour la Santé (World Health Organization, 1994) a défini en 1994 la qualité de vie comme « *La perception qu'a un individu de sa place dans l'existence, dans le contexte de la culture et du système de valeurs dans lesquels il vit, en relation avec ses objectifs, ses attentes, ses normes et ses inquiétudes. Il s'agit d'un large champ conceptuel, englobant de manière complexe la santé physique de la personne, son état psychologique, son niveau d'indépendance, ses relations sociales, ses croyances personnelles et sa relation avec les spécificités de son environnement* ». Malgré cette approche globale proposée la notion de qualité de vie reste difficile à cerner dans sa totalité. Néanmoins, la qualité de vie apparaît comme un concept multidimensionnel (Leplège and Coste, 2001) se structurant généralement autour des quatre dimensions :

- État physique : autonomie, capacités physiques ;



- Sensations somatiques : symptômes, conséquence des traumatismes ou des procédures thérapeutiques, douleurs ;
- État psychologique : émotivité, anxiété, dépression ;
- Domaine social : relations sociales et rapport à l’environnement familial, amical ou professionnel.

Les psychologues, cliniciens et chercheurs en santé publique ont compris la nécessité de prendre en compte les perceptions et les préférences des patients en matière de décisions concernant leur santé. En effet, même si ce sont les médecins qui prescrivent les thérapies, la décision de consulter, de suivre les prescriptions et recommandations ou de rechercher tout autre moyen de répondre à ses besoins revient au patient. Celui-ci établit sa conduite sur son ressenti et non sur la réalité de sa situation clinique. De plus, l’évaluation de la qualité de vie ne peut être réduite à celle de la santé. En effet, lorsqu’un patient est confronté à la perspective d’une longue maladie son instinct de conservation lui fait inévitablement développer des stratégies d’ajustement afin de conserver une certaine satisfaction de vie. De ce fait, certains individus dont le statut fonctionnel et de santé est jugé déplorable ont paradoxalement une bonne qualité de vie, ou inversement (Préau, 2004, p. 29). C’est pourquoi la mesure de la qualité de vie dans le domaine de la santé est un concept subjectif s’intéressant au point de vue des intéressés eux-mêmes. Cependant, la qualité de vie appliquée à la santé ou “*Health Related Quality of Life*” des Anglo-Saxons prend en compte non pas toutes les dimensions de la qualité de vie en général mais celles qui peuvent être modifiées par la maladie ou son traitement.

Par contre, la prise en compte dans les essais thérapeutiques de la qualité de vie est relativement récente. Bien que les méthodes statistiques pour la modélisation de la qualité de vie et de la survie ont eu leur domaine majeur d’application dans les essais sur le cancer, elles ne sont pas encore complètement adaptées à l’étude de l’évolution de la qualité de vie et de la réponse clinique dans le contexte du VIH (Douglas, 1999; Ribaud *et al.*, 2000; Wang *et al.*, 2002).

### 3.2 Analyse longitudinale de la qualité de vie

Awad *et al.* (2002); Mesbah *et al.* (2004) se sont intéressés à l’apparition du premier événement de la dégradation de la qualité de vie pour un certain taux de dégradation de qualité de vie fixé préalablement. Pour chaque taux  $x$  ainsi fixé (par exemple 5%, 10%, 20%, 30%), les constructions de l’estimateur de Kaplan-Meier et d’un test du log-rank ont été réalisées. Les courbes de Kaplan-Meier sont des courbes indiquant aux temps  $t$  la proportion de patients n’ayant pas subi l’événement. Dans le cas de l’infection par le VIH, ces courbes indiquent la proportion de patients ayant une bonne qualité de vie. Le test du log-rank permet la comparaison de courbes de survie.

Pour représenter les évolutions complexes de la qualité de vie, Heutte and Huber-Carol (2002) se sont intéressées à la construction d’un modèle multi-état. Différents états de qualité de vie des patients ont été définis. Un événement est considéré comme une

transition d'un état de qualité de vie vers un autre. Pour prendre en compte la variabilité des temps de séjour dans un état, des processus semi-markoviens ont été utilisés. Ce modèle a été appliqué à la cohorte française SEROCO, cohorte de patients infectés par le VIH, multicentrique, prospective, commencée en 1988. Tous les groupes d'exposition sont représentés dans cette cohorte excepté celui des hémophiles. L'objectif de cette étude est d'étudier l'influence de l'âge sur la progression de l'infection vers le SIDA chez les personnes infectées par le VIH. Les patients âgés de plus de 40 ans à la séroconversion ont progressé plus rapidement vers le SIDA que les patients plus jeunes après la survenue de la première apparition mineure de l'infection. La progression du SIDA vers le décès est la même quel que soit l'âge à la séroconversion (Heutte, 2001).

Jin *et al.* (2006) proposent un modèle de régression semi-paramétrique pour des données longitudinales de qualité de vie, applicable au contexte particulier de maladies chroniques neurodégénératives telle que la maladie d'Alzheimer. Cette approche permet d'appréhender des événements terminaux dépendants. En effet, les patients éprouvent de plus en plus de difficultés, au fil du temps, à réaliser leurs activités quotidiennes. Aussi, les tâches journalières sont mesurées longitudinalement sur des laps de temps au travers d'indicateurs de qualité de vie relative à la santé. Ce modèle offre ainsi la possibilité d'examiner l'effet d'une covariable dépendant ou non du temps (par exemple le sexe, le niveau d'étude) et d'évaluer non-paramétriquement la variation dans les observations due à des facteurs non inclus dans le modèle.

### 3.3 Modèles de dégradation dans l'industrie

#### 3.3.1 Présentation et hypothèses

L'étude de la défaillance (ou survie) de composants et de systèmes est un problème industriel majeur. Le cadre conceptuel classique suppose que la durée de survie  $T$  est définie par le premier temps d'atteinte des limites de performance, i.e.  $T = \sup\{t : \mathcal{D}(t) \in I\}$  où  $\mathcal{D}$  est la fonction de dégradation et  $I$  est l'intervalle des performances garantissant le bon fonctionnement de l'individu ou du système. Dans le domaine industriel, les modèles sont établis sous les hypothèses suivantes :

1. la fonction de dégradation  $\mathcal{D}$  est croissante ;
2. l'intervalle d'observation du processus de dégradation est  $[0, +\infty[$  ;
3. la performance d'un individu est un processus aléatoire à trajectoires presque sûrement croissantes et continues à droite ;
4. la dégradation initiale est  $\mathcal{D}(0) = 0$ , la dégradation des individus est nulle au début de l'étude ;
5. l'intervalle admissible des dégradations est défini par  $I = [0, z_0]$ , signifiant ainsi que la dégradation maximale admissible vaut  $z_0$ . De plus, cette valeur est déterminée par l'utilisateur.

La date de défaillance  $T_0$  est définie comme le premier temps de passage de la dégradation au dessus de la valeur  $z_0$ , i.e.

$$T_0 = \inf\{t : \mathcal{D}(t) \geq z_0\}.$$

Différents modèles existent pour modéliser la loi du processus de dégradation  $\mathcal{D}$ , en fonction de la nature du phénomène concerné ou des données disponibles, ainsi la loi du temps d'atteinte  $T_0$  peut être entre autres modélisée par une loi de Weibull inverse, une loi Log-normale, une loi de Weibull ou une loi gaussienne inverse.

### 3.3.2 Processus gaussiens et premier temps de passage

Les processus gaussiens permettent des calculs explicites de diverses caractéristiques de fiabilité et par suite offrent une alternative efficace pour modéliser les processus de dégradation. Ceux-ci décrivent des trajectoires de dégradation croissantes en moyenne. Notons que les trajectoires du processus gaussien, même avec une tendance linéaire croissante, ne sont pas monotones ; par suite nous observons des périodes durant lesquelles la dégradation est décroissante.

Si la dégradation est régie par un processus de Wiener  $W$  de tendance linéaire  $m$  et de variance  $\sigma^2$  alors, la loi de  $T_0 = \inf\{t : W(t) \geq z_0\}$  est une loi gaussienne inverse  $IG(z_0/m, z_0^2/\sigma^2)$  (Seshadri, 1993) de densité

$$f(t, z_0, \sigma, \mu) = \frac{z_0}{\sqrt{2\pi}\sigma} t^{-3/2} \exp\left(-\frac{(z_0 - \mu t)^2}{2\sigma^2 t}\right).$$

Whitmore (1995) a étendu ce modèle pour le cas de données observées correspondant à des mesures bruitées de la dégradation. Whitmore *et al.* (1998) ont généralisé ce modèle en supposant que  $\{W_t\}$  est un processus stochastique latent mais avec la présence d'un processus partenaire  $\{Y_t\}$  observable. Un état des lieux est disponible dans Lee and Whitmore (2004, 2006).

Singpurwalla (2006) utilisera la notion de potentiel de risque (hazard potentiel) pour relier les modèles de risques compétitifs (Latouche *et al.*, 2007) et les modèles de défaillances dues au vieillissement ou à la dégradation.

## Deuxième partie

# Test non-paramétrique pour l'évolution de la qualité de vie



## Chapitre 4

# Test de type log-rank pour l'évolution de la qualité de vie

### 4.1 Définitions et propriétés de la dégradation de qualité de vie

En général lors d'études cliniques longitudinales les patients remplissent des auto-questionnaires de qualité de vie à des dates préalablement déterminées  $t_j$  ( $j = 0, \dots, T$ ), ainsi il est possible d'évaluer leur score de qualité de vie  $Q_j$ . Les scores de qualité de vie définissent des variables aléatoires mesurées sur des échelles de qualité de vie. Plus le score de qualité de vie est élevé, meilleure est la qualité de vie du patient. L'évolution longitudinale de la qualité de vie par rapport au score initial  $Q_0$  mesuré au temps initial (baseline)  $t_0$  est bien reflétée par la notion de **dégradation de qualité de vie** définie, à l'instant  $t_j$ , soit comme un taux

$$D_j = (Q_0 - Q_j)Q_0^{-1},$$

soit comme une différence

$$D_j = Q_0 - Q_j.$$

Soit  $x$  le taux critique définissant le score maximal acceptable de dégradation de qualité de vie. Un patient est considéré comme dégradé si son score de dégradation de qualité de vie est plus grand que le taux  $x$ . L'**événement d'intérêt**,

$$T(x) = \inf\{t_j : D_j \geq x\},$$

représente la date de survenue de la première fois où le score de dégradation de qualité de vie est supérieur au taux critique  $x$ .

Il est à noter que nous nous plaçons dans un **cadre non-paramétrique** et par conséquent dans un contexte différent de la littérature (rappelé brièvement au chapitre précédent). En effet, nous ne faisons aucune hypothèse aussi bien sur le processus de dégradation de qualité de vie  $D_j$ , que sur le seuil de dégradation de qualité de vie  $x$ .

Pour deux taux différents de dégradation de qualité de vie  $x_1$  et  $x_2$ , les dates de survenue des événements  $T(x_1)$  et  $T(x_2)$  ne sont pas nécessairement les mêmes. Quelques conséquences valables pour les deux définitions de dégradation de qualité de vie peuvent être obtenues.

**Remarque 4.1.**

1.  $D_0 = 0$ , le score de dégradation de qualité de vie est nul à l'origine ;
2. Si  $D_j > 0$  alors le score de qualité de vie à la date  $t_j$  a diminué par rapport au score mesuré au baseline et réciproquement si  $D_j < 0$  alors le score de qualité de vie à la date  $t_j$  est supérieur à celui mesuré au baseline ;
3.  $x_1 > x_2 \Rightarrow \mathbb{P}(D_j > x_1) \leq \mathbb{P}(D_j > x_2)$ , autrement dit plus le taux  $x$  est grand plus l'événement d'intérêt  $T(x)$  se produira tardivement en probabilité.

## 4.2 Notations et hypothèses

Nous considérons une population de  $n$  patients indépendants. Pour des raisons techniques nous supposons que le temps  $t$  est observé de manière continue sur  $[0, t_T]$ . Nous notons par  $D_{it}$  le score de dégradation du patient  $i$  à l'instant  $t$  pour la définition de dégradation de qualité de vie retenue, les résultats mathématiques obtenus sont applicables sur toute définition de dégradation de qualité de vie. Soit  $x$  le taux critique de dégradation de qualité de vie considéré, ainsi la durée de survie correspondante, pour un patient  $i$ , est  $T_i(x) = \inf\{t : D_{it} \geq x\}$ . Notons par  $\tau_i(x)$  la réalisation de  $T_i(x)$ . Soit  $C_i$  la variable de censure à droite indépendante de la durée de survie  $T_i(x)$ . Les observations consistent en les  $n$  triplets  $(X_i(x), \delta_i(x), \mathbf{Z}_i)$  où  $X_i(x) = T_i(x) \wedge C_i$  est l'observation soit du premier temps d'atteinte du seuil  $x$  de dégradation de qualité de vie, soit de la censure,  $\delta_i(x) = \mathbf{1}_{(T_i(x) \leq C_i)}$  est la fonction indicatrice caractérisant si la dégradation de qualité de vie est observée ou non et  $\mathbf{Z}_i$  est le vecteur des covariables. Nous nous plaçons dans une extension, afin de prendre en compte un taux  $x$  de dégradation de qualité de vie, du cadre classique du modèle semi-paramétrique de Cox (1972) i.e., pour un patient  $i$ , la fonction de risque instantané  $\lambda_i$  est définie par

$$\lambda_i(t, x / \mathbf{Z}_i(t)) = \lambda_0(t, x) \exp\{\beta' \mathbf{Z}_i(t)\},$$

où  $\lambda_0$  est la fonction non spécifiée de risque à l'origine et  $\beta$  est le vecteur inconnu de dimension  $p$  des paramètres de régression. Les informations contenues dans la paire  $(X_i(x), \delta_i(x))$  étant équivalentes à la connaissance des informations fournies par le processus de comptage  $N_i(t, x) = \mathbf{1}_{(X_i(x) \leq t, \delta_i(x) = 1)}$  et le processus de censure  $Y_i(t, x) = \mathbf{1}_{(X_i(x) \geq t)}$ , les observations consistent en les  $n$  triplets  $\{N_i(\cdot, x), Y_i(\cdot, x), \mathbf{Z}_i\}$ . Nous sommes en mesure d'étendre le modèle d'intensité multiplicative pour que celui-ci puisse dépendre d'un taux critique  $x$  de dégradation de qualité de vie.

**Définition 4.1.** *Le modèle d'intensité multiplicative dépendant d'un taux  $x$  de dégradation de qualité de vie consiste en les  $n$  triplets d'observations  $\{N_i(\cdot, x), Y_i(\cdot, x), \mathbf{Z}_i\}_{i=1, \dots, n}$  issues de  $n$  patients (ou items) indépendants, d'une filtration continue à droite  $\{\mathcal{F}_t; t \geq 0\}$*

représentant l'information statistique au cours du temps et des  $n$  processus d'intensité  $l_i(t, x) = \lambda_0(t, x)e^{\beta' \mathbf{Z}_i(t)} Y_i(t, x)$ ,  $i = 1, \dots, n$ , respectant les hypothèses suivantes :

1.  $\mathbf{N}(\cdot, x) = (N_1(\cdot, x), \dots, N_n(\cdot, x))'$  est un processus de comptage multivarié, tel que pour tout  $t \geq 0$  et  $i \neq j$ ,

$$\mathbb{P}(\Delta N_i(t, x) = \Delta N_j(t, x) = 1) = 0.$$

2. Pour tout  $i$ ,  $M_i(\cdot, x) = N_i(\cdot, x) - A_i(\cdot, x)$  est une martingale locale pour la filtration  $\{\mathcal{F}_t; t \geq 0\}$ , où  $A_i(\cdot, x)$  est le compensateur défini par

$$A_i(t, x) = \int_0^t \lambda_0(u, x) e^{\beta' \mathbf{Z}_i(u)} Y_i(u, x) du.$$

3. Chaque processus de censure  $Y_i(\cdot, x)$  et chaque processus de covariable  $\mathbf{Z}_i$  sont prévisibles pour la filtration  $\{\mathcal{F}_t; t \geq 0\}$ . De plus, les covariables  $\mathbf{Z}_i$  sont des processus localement bornés.

Nous généralisons également des notations classiques et importantes afin de prendre en considération un taux  $x$  de dégradation de qualité de vie. Pour tout  $k = 0, 1, 2$ ,

$$\mathbf{S}^{(k)}(\beta, t, x) = \frac{1}{n} \sum_{i=1}^n \{\mathbf{Z}_i(t)\}^{\otimes k} Y_i(t, x) \exp\{\beta' \mathbf{Z}_i(t)\}, \quad (4.1)$$

$$\mathbf{E}(\beta, t, x) = \frac{\mathbf{S}^{(1)}(\beta, t, x)}{S^{(0)}(\beta, t, x)} \quad (4.2)$$

et

$$\mathbf{V}(\beta, t, x) = \frac{\mathbf{S}^{(2)}(\beta, t, x)}{S^{(0)}(\beta, t, x)} - \{\mathbf{E}(\beta, t, x)\}^{\otimes 2}, \quad (4.3)$$

avec pour tout vecteur  $\mathbf{X}$ ,  $\mathbf{X}^{\otimes 0} = 1$ ,  $\mathbf{X}^{\otimes 1} = \mathbf{X}$  et  $\mathbf{X}^{\otimes 2} = \mathbf{X}\mathbf{X}'$ . Nous notons également par  $\beta_0$  la vraie valeur de  $\beta$ , par  $D[0, t_T]$  l'espace des fonctions continues à droite et limitées à gauche et par  $\mathbb{P}$  la convergence en probabilité.

Dans la suite, nous supposons que les conditions suivantes sont vérifiées.

(C1) Le temps  $t_T$  est tel que,  $\int_0^{t_T} \lambda_0(u, x) du < \infty$ .

(C2) Pour  $\mathbf{S}^{(j)}$ ,  $j = 0, 1, 2$  définie en (4.1) il existe un voisinage  $\mathcal{B}$  de  $\beta_0$ , un voisinage  $\mathcal{X}$  de  $x$  et respectivement un scalaire, un vecteur et une fonction matricielle  $s^{(0)}$ ,  $\mathbf{s}^{(1)}$  et  $\mathbf{s}^{(2)}$  définis sur  $\mathcal{B} \times [0, t_T] \times \mathcal{X}$  tels que pour  $j = 0, 1, 2$ ,

$$\sup_{u \in [0, t_T], \beta \in \mathcal{B}, x \in \mathcal{X}} \|\mathbf{S}^{(j)}(\beta, u, x) - \mathbf{s}^{(j)}(\beta, u, x)\|_{\infty} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

(C3) Il existe  $\delta > 0$  tel que,

$$n^{-1/2} \sup_{1 \leq i \leq n, 0 \leq u \leq t_T, x \in \mathcal{X}} |\mathbf{Z}_i(u) Y_i(u, x) \mathbf{1}_{\{\beta_0' \mathbf{Z}_i(u) > -\delta |\mathbf{Z}_i(u)|\}}| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$



(C4) Soient  $\mathcal{B}$ ,  $\mathcal{X}$  et  $\mathbf{s}^{(j)}$ ,  $j = 0, 1, 2$ , définie en (C2), posons

$$\mathbf{e} = \frac{\mathbf{s}^{(1)}}{s^{(0)}} \quad \text{et} \quad \mathbf{v} = \frac{\mathbf{s}^{(2)}}{s^{(0)}} - \mathbf{e}^{\otimes 2}$$

alors, pour tout  $\beta \in \mathcal{B}$ ,  $x \in \mathcal{X}$  et  $0 \leq u \leq t_T$ ,

$$\frac{\partial}{\partial \beta} s^{(0)}(\beta, u, x) = \mathbf{s}^{(1)}(\beta, u, x)$$

et

$$\frac{\partial^2}{\partial \beta^2} s^{(0)}(\beta, u, x) = \mathbf{s}^{(2)}(\beta, u, x).$$

(C5) Les fonctions  $\mathbf{s}^{(j)}$ ,  $j = 0, 1, 2$ , sont bornées sur  $\mathcal{B} \times [0, t_T] \times \mathcal{X}$  et de plus,  $s^{(0)}$  ne s'annule pas sur  $\mathcal{B} \times [0, t_T] \times \mathcal{X}$ ; pour  $j = 0, 1, 2$ , la famille de fonctions  $\mathbf{s}^{(j)}(\cdot, u, x)$ ,  $0 \leq u \leq t_T$  et  $x \in \mathcal{X}$  est équicontinue au point  $\beta_0$ .

(C6) La matrice

$$\Sigma(\beta_0, t_T, x) = \int_0^{t_T} \mathbf{v}(\beta_0, u, x) s^{(0)}(\beta_0, u, x) \lambda_0(u, x) du$$

est définie positive.

**Remarque 4.2.** Les conditions introduites sont des généralisations, permettant de prendre en compte un taux  $x$  de dégradation de qualité de vie, de conditions classiques. En effet, les Conditions (C1) et (C3) sont respectivement des extensions de la Condition dite d'intervalle fini et de la Condition de Lindeberg. D'autre part, la Condition (C2) sert à la stabilité asymptotique, tandis que les Conditions (C4)-(C6) servent à assurer la régularité asymptotique.

## 4.3 Analyse longitudinale de la dégradation de qualité de vie pour un taux fixé

### 4.3.1 Statistique du score de vraisemblance partielle pour un taux de dégradation fixé et propriétés asymptotiques

Dans toute cette section, nous supposons que le taux  $x$  de dégradation de qualité de vie est préalablement fixé. Nous établissons la convergence du vecteur du processus de score normalisé vers un processus gaussien, c'est le résultat principal de cette section.

**Théorème 4.1.** (i) Nous supposons que les conditions (C1)-(C6) sont vérifiées alors, le processus du vecteur score normalisé  $\{n^{-1/2} \mathbf{U}(\beta_0, t, x); 0 \leq t \leq t_T\}$ , défini à l'instant  $t$  par

$$\frac{1}{\sqrt{n}} \mathbf{U}(\beta_0, t, x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \{\mathbf{Z}_i(u) - \mathbf{E}(\beta_0, u, x)\} N_i(du, x),$$

converge faiblement dans  $(D[0, t_T])^p$  vers un processus gaussien  $p$ -varié de moyenne nulle tel que, chaque processus est à accroissement indépendant. De plus, la matrice de covariance, à l'instant  $t$  et pour les composantes  $l$  et  $l'$ , vaut

$$\{\Sigma(\beta_0, t, x)\}_{ll'} = \int_0^t \{\mathbf{v}(\beta_0, u, x)\}_{ll'} s^{(0)}(\beta_0, u, x) \lambda_0(u, x) du.$$

(ii) Si de plus,  $\hat{\beta}$  est un estimateur consistant de  $\beta_0$  alors,

$$\sup_{0 \leq t \leq t_T} \left\| \frac{1}{n} \int_0^t \sum_{i=1}^n \mathbf{V}(\hat{\beta}, u, x) N_i(du, x) - \Sigma(\beta_0, t, x) \right\|_{\infty} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

**Preuve du Théorème 4.1** Dans toute la preuve  $t \in [0, t_T]$ .

1. En considérant la martingale  $M_i(t, x) = N_i(t, x) - \int_0^t Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du$  et en posant  $\mathbf{H}_i(\beta_0, u, x) = \mathbf{Z}_i(u) - \mathbf{E}(\beta_0, u, x)$ ,  $\mathbf{U}$  peut s'écrire,

$$\mathbf{U}(\beta_0, t, x) = \sum_{i=1}^n \int_0^t \mathbf{H}_i(\beta_0, u, x) M_i(du, x).$$

De même, la  $l$ -ième composante,  $U_l(\beta_0, t, x)$  du processus de score peut être réécrite par

$$U_l(\beta_0, t, x) = \sum_{i=1}^n \int_0^t H_{il}(\beta_0, u, x) M_i(du, x),$$

avec  $H_{il}(\beta_0, u, x) = Z_{il}(u) - E_l(\beta_0, u, x)$  et  $E_l(\beta_0, u, x) = \frac{\sum_{j=1}^n Y_j(u, x) Z_{jl}(u) \exp\{\beta_0' \mathbf{Z}_j(u, x)\}}{\sum_{j=1}^n Y_j(u, x) \exp\{\beta_0' \mathbf{Z}_j(u, x)\}}$ .

Posons  $\mathbf{U}^{(n)} = n^{-1/2} \mathbf{U}$ . En remarquant que deux processus de comptage  $N_i(\cdot, x)$  n'ont pas leurs sauts à la même date nous obtenons

$$\begin{aligned} \langle U_l^{(n)}(\beta_0, \cdot, x), U_l^{(n)}(\beta_0, \cdot, x) \rangle(t) &:= \frac{1}{n} \sum_{i=1}^n \int_0^t H_{il}^2(\beta_0, u, x) Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \\ &= \int_0^t \frac{1}{n} \sum_{i=1}^n \{Z_{il}(u) - E_l(\beta_0, u, x)\}^2 Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \\ &= \int_0^t \{\mathbf{V}(\beta_0, u, x)\}_l S^{(0)}(\beta_0, u, x) \lambda_0(u, x) du, \end{aligned}$$

où  $\{\mathbf{V}(\beta_0, u, x)\}_l$  est le  $l$ -ième élément diagonal de la matrice  $\mathbf{V}(\beta_0, u, x)$  définie en (4.3). Or, d'après les conditions (C1), (C2) et (C5),

$$\langle U_l^{(n)}(\beta_0, \cdot, x), U_l^{(n)}(\beta_0, \cdot, x) \rangle(t) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \int_0^t \{\mathbf{v}(\beta_0, u, x)\}_l s^{(0)}(\beta_0, u, x) \lambda_0(u, x) du.$$

Nous obtenons de même,

$$\langle U_l^{(n)}(\beta_0, \cdot, x), U_l^{(n)}(\beta_0, \cdot, x) \rangle(t) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \int_0^t \{\mathbf{v}(\beta_0, u, x)\}_{ll'} s^{(0)}(\beta_0, u, x) \lambda_0(u, x) du.$$

Il reste seulement à vérifier la généralisation, afin de prendre en compte un taux  $x$  de dégradation de qualité de vie, de la Condition de Lindeberg. Posons  $\mathbf{H}_i^{(n)} = n^{-1/2} \mathbf{H}_i$  alors,

$$U_l^{(n)}(\beta_0, t, x) = \sum_{i=1}^n \int_0^t H_{il}^{(n)}(\beta_0, u, x) M_i(du, x)$$

et

$$U_{l,\varepsilon}^{(n)}(\beta_0, t, x) = \sum_{i=1}^n \int_0^t H_{il}^{(n)}(\beta_0, u, x) \mathbf{1}_{\{|H_{il}^{(n)}(\beta_0, u, x)| > \varepsilon\}} M_i(du, x)$$

par conséquent,

$$\begin{aligned}
\langle U_{l,\varepsilon}^{(n)}, U_{l',\varepsilon}^{(n)} \rangle(t) &:= \sum_{i=1}^n \int_0^t \{H_{il}^{(n)}(\beta_0, u, x)\}^2 \mathbf{1}_{\{|H_{il}^{(n)}(\beta_0, u, x)| > \varepsilon\}} Y_i(u, x) e^{\beta'_0 \mathbf{Z}_i(u)} \lambda_0(u, x) du \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^t \{H_{il}(\beta_0, u, x)\}^2 \mathbf{1}_{\{\frac{1}{\sqrt{n}} |H_{il}(\beta_0, u, x)| > \varepsilon\}} Y_i(u, x) e^{\beta'_0 \mathbf{Z}_i(u)} \lambda_0(u, x) du \\
&\leq \frac{4}{n} \sum_{i=1}^n \int_0^{t_T} |Z_{il}(u)|^2 \mathbf{1}_{\{\frac{1}{\sqrt{n}} |Z_{il}(u)| > \frac{\varepsilon}{2}\}} Y_i(u, x) e^{\beta'_0 \mathbf{Z}_i(u)} \lambda_0(u, x) du \\
&\quad + 4 \int_0^{t_T} |E_l(\beta_0, u, x)|^2 \mathbf{1}_{\{\frac{1}{\sqrt{n}} |E_l(\beta_0, u, x)| > \frac{\varepsilon}{2}\}} S^{(0)}(\beta_0, u, x) \lambda_0(u, x) du, \quad (4.4)
\end{aligned}$$

car  $H_{il}(\beta_0, u, x) = Z_{il}(u) - E_l(\beta_0, u, x)$  et pour tout  $(a, b) \in \mathbb{R}^2$ ,  $|a - b|^2 \mathbf{1}_{\{|a-b| > \varepsilon\}} \leq 4|a|^2 \mathbf{1}_{\{|a| > \frac{\varepsilon}{2}\}} + 4|b|^2 \mathbf{1}_{\{|b| > \frac{\varepsilon}{2}\}}$ .

Soit  $\varepsilon' > 0$  fixé, les Conditions (C2) et (C5) impliquent que lorsque  $n$  est suffisamment grand il existe un ensemble  $A_1$  tel que  $P(A_1) \geq 1 - \varepsilon'$  et vérifiant pour tout  $0 \leq u \leq t_T$ ,  $\mathbf{1}_{\{\frac{1}{\sqrt{n}} |E_l(\beta_0, u, x)| > \frac{\varepsilon}{2}\}} = 0$  aussi, le second terme de (4.4) converge en probabilité vers 0.

Soient les deux événements disjoints  $B_{1i}^l(u, x) = \left\{ \frac{1}{\sqrt{n}} |Z_{il}(u)| > \frac{\varepsilon}{2}, \beta'_0 \mathbf{Z}_i(u) > -\delta |\mathbf{Z}_i(u)| \right\}$  et  $B_{2i}^l(u, x) = \left\{ \frac{1}{\sqrt{n}} |Z_{il}(u)| > \frac{\varepsilon}{2}, \beta'_0 \mathbf{Z}_i(u) \leq -\delta |\mathbf{Z}_i(u)| \right\}$  nous obtenons  $\left\{ \frac{1}{\sqrt{n}} |Z_{il}(u)| > \frac{\varepsilon}{2} \right\} = B_{1i}^l(u, x) \cup B_{2i}^l(u, x)$ . La Condition (C3) implique l'existence d'au moins un  $\delta > 0$  tel que pour  $\varepsilon' > 0$  fixé et pour  $n$  suffisamment grand, il existe un ensemble  $A_2$  tel que  $P(A_2) \geq 1 - \varepsilon'$  et pour  $0 \leq u \leq t_T, 1 \leq i \leq n$  et  $x \in \mathcal{X}$ ,  $\mathbf{1}_{\{B_{1i}^l(u, x)\}} Y_i(u, x) = 0$  par conséquent,  $\int_0^{t_T} \frac{1}{n} \sum_{i=1}^n |Z_{il}(u)|^2 \mathbf{1}_{\{B_{1i}^l(u, x)\}} Y_i(u, x) e^{\beta'_0 \mathbf{Z}_i(u)} \lambda_0(u, x) du$  converge vers 0 en probabilité.

Pour montrer que  $\int_0^{t_T} \frac{1}{n} \sum_{i=1}^n |Z_{il}(u)|^2 \mathbf{1}_{\{B_{2i}^l(u, x)\}} Y_i(u, x) e^{\beta'_0 \mathbf{Z}_i(u)} \lambda_0(u, x) du$  converge vers 0 en probabilité, nous remarquons que ce terme est borné par

$$\int_0^{t_T} \frac{1}{n} \sum_{i=1}^n |Z_{il}(u)|^2 e^{-\delta |\mathbf{Z}_i(u)|} \mathbf{1}_{\{|Z_{il}(u)| > \sqrt{n} \frac{\varepsilon}{2}\}} \lambda_0(u, x) du. \quad (4.5)$$

Puisque  $\lim_{u \rightarrow \infty} u^2 e^{-\delta u} = 0$  quand  $\delta > 0$ , pour tout  $x \in \mathcal{X}$ , tout  $\eta > 0$  fixé et pour  $n$  suffisamment grand quel que soit  $i$ ,  $\mathbf{1}_{\{|Z_{il}(u)| > \sqrt{n} \frac{\varepsilon}{2}\}} |Z_{il}(u)|^2 e^{-\delta |\mathbf{Z}_i(u)|} < \eta$ . Ainsi (4.5) est borné par  $\eta \int_0^{t_T} \lambda_0(u, x) du$  et donc (4.5) converge en probabilité vers 0 quand  $n$  tend vers l'infini.

Nous venons de montrer que pour tout  $\varepsilon > 0$ ,  $\langle U_{l,\varepsilon}^{(n)}, U_{l',\varepsilon}^{(n)} \rangle(t)$  converge en probabilité vers 0 et pour tout  $(l, l') \in \{1, \dots, r\}^2$  et tout  $t \geq 0$ ,  $\langle U_l^{(n)}, U_{l'}^{(n)} \rangle(t)$  converge en probabilité vers  $\int_0^t \{\mathbf{v}(\beta_0, u, x)\} W S^{(0)}(\beta_0, u, x) \lambda_0(u, x) du$ , il s'en suit la partie (i) du théorème.

**2.** Le processus de comptage  $\bar{N}(\cdot, x) = \sum_i N_i(\cdot, x)$  admettant pour compensateur  $\int_0^{t_T} \sum_i Y_i(u, x) e^{\beta'_0 \mathbf{Z}_i(u)} \lambda_0(u, x) du$ , l'Inégalité de Lengart peut s'écrire,

$$\begin{aligned}
\forall c, \delta > 0, \forall t \geq 0, P \left\{ \frac{1}{n} \bar{N}(t_T, x) \geq c \right\} &\leq \frac{\delta}{c} + P \left\{ \int_0^{t_T} \frac{1}{n} \sum_{i=1}^n Y_i(u, x) e^{\beta'_0 \mathbf{Z}_i(u)} \lambda_0(u, x) du > \delta \right\} \\
&= \frac{\delta}{c} + P \left\{ \int_0^{t_T} S^{(0)}(\beta_0, u, x) \lambda_0(u, x) du > \delta \right\}.
\end{aligned}$$

D'après le Théorème des grands nombres,  $n^{-1}\bar{N}(t_T, x)$  converge presque sûrement, il s'en suit que la  $\lim_{n \rightarrow \infty} P\{n^{-1}\bar{N}(t_T, x) > c\}$  existe pour tout  $c$  fini. D'après la Condition (C2), le terme de droite de l'inégalité converge vers  $\frac{\delta}{c} + P\left\{\int_0^{t_T} s^{(0)}(\beta_0, u, x)\lambda_0(u, x)du > \delta\right\}$ . Puisque  $\delta$  est choisi arbitrairement, prendre  $\delta > \int_0^{t_T} s^{(0)}(\beta_0, u, x)\lambda_0(u, x)du$  entraîne

$$\lim_{c \uparrow \infty} \lim_{n \rightarrow \infty} P\{n^{-1}\bar{N}(t_T, x) > c\} = 0. \quad (4.6)$$

Soit  $\mathcal{I}$  la matrice d'information de Fisher définie par  $\mathcal{I}(\beta, t, x) := -\frac{\partial}{\partial \beta} \mathbf{U}(\beta, t, x)$ . Comme  $\mathcal{I}(\beta, t, x) = \int_0^t \mathbf{V}(\beta, t, x) \bar{N}_i(du, x)$ , nous en déduisons que pour tout estimateur  $\hat{\beta}$ ,

$$\begin{aligned} & \left\| \frac{1}{n} \mathcal{I}(\hat{\beta}, t_T, x) - \Sigma(\beta_0, t_T, x) \right\|_{\infty} \\ & \leq \left\| \int_0^{t_T} \{\mathbf{V}(\hat{\beta}, u, x) - \mathbf{v}(\hat{\beta}, u, x)\} \frac{1}{n} \bar{N}(du, x) \right\|_{\infty} \\ & \quad + \left\| \int_0^{t_T} \{\mathbf{v}(\hat{\beta}, u, x) - \mathbf{v}(\beta_0, u, x)\} \frac{1}{n} \bar{N}(du, x) \right\|_{\infty} \\ & \quad + \left\| \int_0^{t_T} \{\mathbf{v}(\beta_0, u, x)\} \frac{1}{n} \left\{ \bar{N}(du, x) - \sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \right\} \right\|_{\infty} \\ & \quad + \left\| \int_0^{t_T} \mathbf{v}(\beta_0, u, x) \{S^{(0)}(\beta_0, u, x) - s^{(0)}(\beta_0, u, x)\} \lambda_0(u, x) du \right\|_{\infty}. \end{aligned} \quad (4.7)$$

La Condition (C2) et la majoration de la Condition (C5) impliquent que pour un estimateur consistant  $\hat{\beta}$ ,  $\sup_{0 \leq u \leq t_T} \left\| \mathbf{V}(\hat{\beta}, u, x) - \mathbf{v}(\hat{\beta}, u, x) \right\|_{\infty}$  converge vers 0 en probabilité lorsque  $n$  tend vers l'infini. Cette limite couplée à l'Équation (4.6) établit que le premier terme de droite de l'Inégalité (4.7) converge vers 0 en probabilité quand  $n$  tend vers l'infini. L'Équation (4.6) et l'équicontinuité des familles  $\mathbf{s}^{(j)}$  impliquent que le second terme de droite de (4.7) converge également vers 0. Par une application directe des Conditions (C1), (C2) et (C5) le quatrième terme est négligeable asymptotiquement.

La partie (2) de l'Inégalité de Lengart entraîne que pour chaque paire  $j, k$ ,

$$\begin{aligned} & P \left\{ \left| \int_0^{t_T} \{\mathbf{v}(\beta_0, u, x)\}_{jk} \frac{1}{n} M(du, x) \right| > \rho \right\} \\ & \leq \frac{\delta}{\rho^2} + P \left\{ \frac{1}{n} \int_0^{t_T} [\{\mathbf{v}(\beta_0, u, x)\}_{jk}]^2 S^{(0)}(\beta_0, u, x) \lambda_0(u, x) du > \delta \right\}. \end{aligned}$$

Les Conditions (C1), (C2) et (C5) impliquent aussi que le terme de gauche de l'inégalité converge vers 0 quand  $n$  tend vers l'infini, il s'en suit la partie (ii) du théorème. ■

**Remarque 4.3.** Comme  $\mathcal{I}(\beta, t, x) = \int_0^t \sum_{i=1}^n \mathbf{V}(\beta, u, x) N_i(du, x)$ , où  $\mathcal{I}$  est la matrice d'information de Fisher, la partie (ii) du Théorème 4.1 montre que  $n^{-1}\mathcal{I}(\beta_0, t, x)$  et  $n^{-1}\mathcal{I}(\hat{\beta}, t, x)$  sont deux estimateurs uniformément consistants sur  $[0, t_T]$  et de variance  $\Sigma(\beta_0, t_T, x)$  pour le processus de score en  $\beta = \beta_0$ .

Nous obtenons la consistance et la normalité asymptotique de  $\hat{\beta}$  l'estimateur de  $\beta$ .

**Théorème 4.2 (Consistance).** Soient  $\hat{\beta}$  l'estimateur du maximum de vraisemblance partielle de  $\beta$  et  $\beta_0$  la vraie valeur du paramètre  $\beta$  pour le modèle d'intensité multiplicative alors,  $\hat{\beta}$  converge en probabilité lorsque  $n$  tend vers l'infini vers  $\beta_0$ .

**Preuve du Théorème 4.2** Le théorème repose sur la convergence du log de la vraisemblance partielle pour une fonction maximisée en  $\beta = \beta_0$  et sur le lemme

**Lemme 4.1 (Andersen and Gill (1982)).** Considérons  $E$  un sous-ensemble ouvert convexe de  $\mathbb{R}^p$ ,  $F_1, F_2, \dots$ , une suite de fonctions aléatoires concaves de  $E$  et  $f$  une fonction de  $E$  telle que pour tout  $x \in E$ ,  $\lim_{n \rightarrow \infty} F_n(x) = f(x)$  en probabilité alors,

1. la fonction  $f$  est concave,
2. pour tout sous-ensemble compact  $A$  de  $E$ ,  $\sup_{x \in A} |F_n(x) - f(x)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$ ,
3. si  $F_n$  admet un unique maximum en  $X_n$  et si  $f$  est unique en  $x$  alors,  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} x$ .

Nous commençons par montrer la convergence du log de la vraisemblance partielle en exprimant une limite comme une martingale. Soit  $nX_n(\beta, \cdot, x)$  le processus représentant la différence du log des vraisemblances partielles sur  $[0, t]$ , évalué en  $\beta$  arbitraire et pour la vraie valeur  $\beta_0$  nous avons,

$$\begin{aligned} X_n(\beta, t, x) &= \frac{1}{n} [\mathcal{L}(\beta, t, x) - \mathcal{L}(\beta_0, t, x)] \\ &= \frac{1}{n} \left[ \sum_{i=1}^n \int_0^t (\beta - \beta_0)' \mathbf{Z}_i(u) dN_i(\cdot, x)(u) \right. \\ &\quad \left. - \int_0^t \log \left( \frac{\sum_{i=1}^n Y_i(u, x) e^{\beta' \mathbf{Z}_i(u)}}{\sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)}} \right) \bar{N}(du, x) \right]. \end{aligned}$$

Considérons

$$\begin{aligned} A_n(\beta, t, x) &= \frac{1}{n} \left[ \sum_{i=1}^n \int_0^t (\beta - \beta_0)' \mathbf{Z}_i(u) Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \right. \\ &\quad \left. - \int_0^t \sum_{i=1}^n \log \left( \frac{S^{(0)}(\beta, u, x)}{S^{(0)}(\beta_0, u, x)} \right) Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \right], \end{aligned}$$

alors,  $X_n(\beta, t, x) - A_n(\beta, t, x) = \frac{1}{n} \sum_{i=1}^n \int_0^t \left\{ (\beta - \beta_0)' \mathbf{Z}_i(u) - \log \left( \frac{S^{(0)}(\beta, u, x)}{S^{(0)}(\beta_0, u, x)} \right) \right\} M_i(du, x)$ .

Le processus  $X_n(\beta, \cdot, x) - A_n(\beta, \cdot, x)$  est une martingale locale de carré intégrable avec comme processus prévisible de variation en  $t$ ,

$$\begin{aligned} &\langle X_n(\beta, \cdot, x) - A_n(\beta, \cdot, x), X_n(\beta, \cdot, x) - A_n(\beta, \cdot, x) \rangle(t) \\ &= \frac{1}{n^2} \sum_{i=1}^n \int_0^t \left\{ (\beta - \beta_0)' \mathbf{Z}_i(u) - \log \left( \frac{S^{(0)}(\beta, u, x)}{S^{(0)}(\beta_0, u, x)} \right) \right\}^2 Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \\ &= \frac{1}{n} \int_0^t \left\{ (\beta - \beta_0)' \mathbf{S}^{(2)}(\beta_0, u, x) (\beta - \beta_0) - 2(\beta - \beta_0)' \mathbf{S}^{(1)}(\beta_0, u, x) \log \left( \frac{S^{(0)}(\beta, u, x)}{S^{(0)}(\beta_0, u, x)} \right) \right. \\ &\quad \left. + \left[ \log \left( \frac{S^{(0)}(\beta, u, x)}{S^{(0)}(\beta_0, u, x)} \right) \right]^2 S^{(0)}(\beta_0, u, x) \right\} \lambda_0(u, x) du. \end{aligned}$$

Les Conditions (C1), (C2) et (C5) entraînent que  $n\langle X_n(\beta, \cdot, x) - A_n(\beta, \cdot, x), X_n(\beta, \cdot, x) - A_n(\beta, \cdot, x) \rangle(t_T)$  converge vers une limite finie, de plus l'Inégalité de Lengart implique que  $\lim_{n \rightarrow \infty} X_n(\beta, t_T, x) - A_n(\beta, t_T, x) = 0$  en probabilité. Puisque  $A_n(\beta, t_T, x)$  converge vers

$$A(\beta, t_T, x) = \int_0^{t_T} \left\{ (\beta - \beta_0)' \mathbf{s}^{(1)}(\beta_0, u, x) - \log \left( \frac{s^{(0)}(\beta, u, x)}{s^{(0)}(\beta_0, u, x)} \right) s^{(0)}(\beta, u, x) \right\} \lambda_0(u, x) du,$$

$X_n(\beta, t_T, x)$  converge en probabilité vers la même limite tant que  $\beta \in \mathcal{B}$ .  $X_n(\beta, t_T, x)$  est une fonction concave de  $\beta$  et admet un unique maximum. En utilisant les conditions (C1), (C2) et (C5) on peut montrer que  $A(\beta, t_T, x)$  admet un unique maximum en  $\beta = \beta_0$ . Le théorème s'en suit par le Lemme A.2 point (3). ■

**Théorème 4.3 (Normalité asymptotique).** *Soit  $\Sigma(\beta_0, t, x)$  définie comme dans la Condition (C6) alors,  $\sqrt{n}(\hat{\beta} - \beta_0)$  converge en distribution lorsque  $n$  tend vers l'infini, vers une variable gaussienne  $p$ -variée centrée de matrice de covariance  $\{\Sigma(\beta_0, t_T, x)\}^{-1}$ .*

**Preuve du Théorème 4.3** Le score pour une  $\mathbf{U}$  statistique est  $\mathbf{U}(\hat{\beta}, t_T, x) = \mathbf{U}(\beta, t_T, x) - \mathcal{I}(\beta^*, t_T, x)(\hat{\beta} - \beta_0)$ , avec  $\beta^*$  situé sur le segment défini par  $\hat{\beta}$  et  $\beta_0$ . Puisque  $\mathbf{U}(\hat{\beta}, t_T, x) = 0$  nous obtenons  $n^{-1}\mathcal{I}(\beta^*, t_T, x)n^{1/2}(\hat{\beta} - \beta_0) = n^{-1/2}\mathbf{U}(\beta, t_T, x)$ . La partie (1) du Théorème 4.1 établit que  $n^{-1/2}\mathbf{U}(\beta, t_T, x)$  est asymptotiquement normal avec comme matrice de covariance  $\Sigma(\beta_0, t_T, x)$ . La consistance de  $\hat{\beta}$  entraîne que  $\beta^*$  converge vers  $\beta$  en probabilité lorsque  $n$  tend vers l'infini. La preuve de la partie (2) du Théorème 4.1 montre que  $\mathcal{I}(\beta^*, t_T)$  converge en probabilité vers une matrice non singulière  $\Sigma(\beta_0, t_T, x)$ . D'après le Théorème de Slutsky le résultat est immédiat. ■

Afin, de compléter les propriétés asymptotiques des estimateurs, nous introduisons une généralisation, prenant en compte un taux  $x$  de dégradation de qualité de vie, de l'estimateur  $\hat{\Lambda}_0$  de la fonction de risque cumulée  $\Lambda_0$  donné par Breslow (1974). Cette généralisation de l'estimateur de Breslow (1974) est définie par

$$\hat{\Lambda}_0(t, x) = \int_0^t \left\{ \sum_{i=1}^n Y_i(u, x) \exp(\hat{\beta}' \mathbf{Z}_i(u)) \right\}^{-1} \left\{ \sum_{i=1}^n N_i(du, x) \right\}.$$

Notons que l'estimateur de Breslow (1974) est l'estimateur de la fonction de risque cumulée le plus utilisé. Nous établissons la convergence faible de  $\sqrt{n}(\hat{\Lambda}_0 - \Lambda_0)$ .

**Théorème 4.4 (Convergence faible).** *Nous supposons que les conditions sont vérifiées alors,  $\sqrt{n}(\hat{\Lambda}_0 - \Lambda_0)$  converge faiblement sur  $D[0, t_T]$  vers un processus gaussien de moyenne nulle, à accroissement indépendant et de matrice de covariance,*

$$\int_0^t \{s_0(\beta_0, u, x)\}^{-1} \lambda_0(u, x) du + \mathbf{Q}'(\beta_0, t, x) \Sigma^{-1}(\beta_0, t_T, x) \mathbf{Q}(\beta_0, t, x),$$

où  $\mathbf{Q}(\beta_0, t, x) = \int_0^t \mathbf{e}(\beta_0, u, x) \lambda_0(u, x) du$ .

**Preuve du Théorème 4.4** Pour démontrer le théorème on utilise la décomposition suivante,

$$\begin{aligned} \sqrt{n}\{\hat{\Lambda}_0(t) - \Lambda_0(t)\} &= \sqrt{n} \int_0^t \left\{ \frac{1}{\sum_{i=1}^n Y_i(u, x) e^{\hat{\beta}' \mathbf{Z}_i(u)}} - \frac{1}{\sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)}} \right\} \bar{N}(du, x) \\ &+ \sqrt{n} \left\{ \int_0^t \frac{1}{\sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)}} \bar{N}(du, x) - \Lambda_0^*(t) \right\} \\ &+ \sqrt{n} \{ \Lambda_0^*(t) - \Lambda_0(t) \}, \end{aligned} \quad (4.8)$$

avec  $\Lambda_0^*(t) = \int_0^t \mathbf{1}_{(\bar{Y}(u, x) > 0)} \lambda_0(u, x) du$ ,  $\bar{N}(u, x) = \sum_{i=1}^n N_i(u, x)$  et  $\bar{Y}(u, x) = \sum_{i=1}^n Y_i(u, x)$ .

L'idée principale de la preuve est de montrer la convergence en distribution des deux premiers termes vers deux processus gaussiens asymptotiquement indépendants, et de montrer que le troisième terme est asymptotiquement négligeable.

Comme  $\sqrt{n} \{ \Lambda_0^*(t) - \Lambda_0(t) \} = \sqrt{n} \int_0^t \mathbf{1}_{(\bar{Y}(u, x) = 0)} d\Lambda_0(u)$  nous en déduisons que le troisième terme est asymptotiquement négligeable.

Le deuxième terme, la martingale locale  $\int_0^t \left\{ \sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \right\}^{-1} M(du, x)$ , converge vers un processus gaussien de variance  $\int_0^t \{s^{(0)}(\beta, u, x)\}^{-1} \lambda_0(u, x) du$ .

Un développement en série de Taylor à l'ordre deux de  $\left\{ \sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \right\}^{-1}$  au point  $\beta = \beta_0$  permet de réécrire le premier terme  $\mathbf{H}'(\beta^*, t, x) \sqrt{n}(\hat{\beta} - \beta_0)$ , où  $\beta^*$  est situé sur le segment compris entre  $\hat{\beta}$  et  $\beta_0$  et où  $\mathbf{H}$  est le vecteur colonne défini par  $\mathbf{H}(\beta, t, x) = - \int_0^t n^{-1} \mathbf{S}^{(1)}(\beta, u, x) \{S^{(0)}(\beta, u, x)\}^{-2} \bar{N}(du, x)$ . Nous allons établir la convergence du vecteur  $\mathbf{H}(\beta^*, \cdot, x)$  vers  $\int \mathbf{s}^{(1)}(\beta, u, x) (s^{(0)}(\beta, u, x))^{-1} \lambda_0(u, x) du$ . Puisque  $\beta^*$  est situé sur le segment compris entre  $\hat{\beta}$  et  $\beta_0$  et que  $\hat{\beta}$  est consistant il est raisonnable d'étudier la limite de  $\mathbf{H}(\beta_0, \cdot, x)$ . Nous avons,

$$\begin{aligned} \mathbf{H}(\beta_0, t, x) &= - \int_0^t \frac{1}{n} \frac{\mathbf{S}^{(1)}(\beta_0, u, x)}{(S^{(0)}(\beta_0, u, x))^2} \bar{N}(du, x) \\ &= - \int_0^t \frac{1}{n} \frac{\mathbf{S}^{(1)}(\beta_0, u, x)}{(S^{(0)}(\beta_0, u, x))^2} \bar{M}(du, x) \\ &\quad - \int_0^t \frac{1}{n} \frac{\mathbf{S}^{(1)}(\beta_0, u, x)}{(S^{(0)}(\beta_0, u, x))^2} \sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du. \end{aligned}$$

Le premier terme de l'expression de  $\mathbf{H}(\beta_0, t, x)$  est une martingale locale de carré intégrable et de processus prévisible

$$\begin{aligned} &\int_0^t \frac{1}{n^2} \left[ \frac{\mathbf{S}^{(1)}(\beta_0, u, x)}{(S^{(0)}(\beta_0, u, x))^2} \right] \sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \\ &= \int_0^t \frac{1}{n} \left[ \frac{\mathbf{S}^{(1)}(\beta_0, u, x)}{(S^{(0)}(\beta_0, u, x))^2} \right] S^{(0)}(\beta_0, u, x) \lambda_0(u, x) du, \end{aligned}$$

qui converge en probabilité vers 0 pour tout  $t$ . De plus, le premier terme de l'expression de  $\mathbf{H}(\beta_0, t, x)$  est un processus de moyenne nulle convergeant vers un processus gaussien et par suite celui-ci converge en probabilité vers 0. Le second terme de l'expression de  $\mathbf{H}(\beta_0, t, x)$

converge, sous les conditions de régularité (C1)-(C6), vers  $-\int_0^t \frac{s^{(1)}(\beta_0, u, x)}{s^{(0)}(\beta_0, u)} \lambda_0(u) du$ . Nous en déduisons que  $\mathbf{H}$  converge vers  $-\int_0^t \frac{s^{(1)}(\beta_0, u, x)}{s^{(0)}(\beta_0, u, x)} \lambda_0(u, x) du$ , comme de plus,  $\sqrt{n}(\hat{\beta} - \beta_0)$  est asymptotiquement normal de matrice de covariance  $\Sigma^{-1}(\beta_0, t_T, x)$ , la variance asymptotique du premier terme de (4.8) est

$$\mathbf{Q}'(\beta_0, t, x) \Sigma^{-1}(\beta_0, t_T, x) \mathbf{Q}(\beta_0, t, x).$$

Comme  $\hat{\beta}$  est une fonction du score  $\mathbf{U}(\beta, t_T, x)$  au temps  $t_T$ , pour établir l'indépendance asymptotique des deux premiers termes de (4.8), il suffit de montrer que le second terme est indépendant du score. Le processus de score  $\mathbf{U}(\beta_0, \cdot, x)$  est une martingale locale dont la  $l$ -ième composante admet la représentation intégrale,  $U_l(\beta_0, \cdot, x) = \sum_{i=1}^n \int_0^\cdot \{Z_{il}(u) - E_l(\beta_0, u, x)\} M_i(du, x)$ , et par conséquent,

$$\begin{aligned} & \left\langle U_l(\beta_0, \cdot, x), \int_0^\cdot \left\{ \sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \right\}^{-1} M(du, x) \right\rangle (t) \\ &= \sum_{i=1}^n \int_0^\cdot \left\{ \frac{Z_{il}(u) - E_l(\beta_0, u, x)}{\sum_{i=1}^n Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)}} \right\} Y_i(u, x) e^{\beta_0' \mathbf{Z}_i(u)} \lambda_0(u, x) du \\ &= \int_0^t \left\{ \frac{S_l^{(1)}(\beta_0, u, x)}{S^{(0)}(\beta_0, u, x)} - E_l(\beta_0, u, x) \right\} \lambda_0(u, x) du \\ &= 0. \end{aligned}$$

■

### 4.3.2 Statistique du test de type log-rank pour un taux de dégradation de qualité de vie fixé

Considérons  $A$  et  $B$  deux groupes de patients (par exemple, mis sous des traitements différents). Nous supposons que les covariables sont réduites à  $\mathbf{Z}_i = \mathbf{1}_{(i \in B)}$ , la fonction indicatrice d'appartenance au groupe  $B$  qui ne dépend ni du temps, ni du taux  $x$  de dégradation de qualité de vie. Notons  $n_G(t, x) = \sum_{i \in G} \mathbf{1}_{(T_i(x) \geq t)}$ , ( $G = A, B$ ) le nombre de patients du groupe  $G$  à risque, i.e. n'ayant pas subi de dégradation, à l'instant  $t$  et  $n(t, x) = n_A(t, x) + n_B(t, x)$  le nombre total de patients non dégradés à l'instant  $t$ . Pour un patient  $i$  nous notons  $\tau_i$  la réalisation de l'événement  $T_i(x)$ . Dans cette section, pour un taux de dégradation  $x$  préalablement fixé, nous cherchons à tester l'égalité des fonctions de survie de deux groupes  $A$  et  $B$  de patients. Par conséquent, l'hypothèse nulle à tester est  $H_0 : S_A(t, x) = S_B(t, x)$  contre l'hypothèse alternative  $H_1 : S_A(t, x) \neq S_B(t, x)$ . Soient  $U_n(x)$  la statistique de type log-rank et  $\hat{\text{Var}}(U_n(x))$  l'estimateur de sa variance définis respectivement par

$$U_n(x) = \sum_{i=1}^n \left\{ \mathbf{Z}_i - \frac{n_B(\tau_i, x)}{n(\tau_i, x)} \right\}$$

et

$$\hat{\text{Var}}(U_n(x)) = \sum_{i=1}^n \frac{n_A(\tau_i, x) n_B(\tau_i, x)}{n^2(\tau_i, x)}.$$



Nous obtenons la convergence asymptotique de la statistique de type log-rank.

**Théorème 4.5.**  $n^{-1/2}U_n(x)$  converge en distribution vers  $\mathcal{N}(0, \sigma_{U(x)}^2)$ . De plus,  $n^{-1}\hat{\text{Var}}(U_n(x))$  converge vers  $\sigma_{U(x)}^2$ .

**Preuve du Théorème 4.5** C'est une conséquence du Théorème 4.1. ■

## 4.4 Analyse longitudinale de la dégradation de qualité de vie pour un taux variable

### 4.4.1 Statistique du score de vraisemblance partielle

Nous supposons maintenant que le taux  $x$  de dégradation de qualité de vie est variable. Sous les notations et hypothèses précédemment introduites nous obtenons la proposition suivante.

**Théorème 4.6.**  $n^{-1/2}U_n(\cdot)$  converge en distribution vers un processus gaussien de fonction de covariance  $\Gamma$ .

La preuve du Théorème 4.6 est une application indirecte du théorème central fonctionnel suivant,

**Théorème 4.7 (Bilias *et al.* (1997)).** Soient  $x$  le temps calendaire courant et  $t$  le temps de survie. Nous supposons que  $x \geq t$  et nous définissons le processus de scores à deux indices de temps  $t$  et  $x$ ,

$$U(\beta; t, x) = \sum_{i=1}^n \int_0^t [\mathbf{Z}_i(u) - \mathbf{E}(\beta; u, x)] N_i(du, x). \quad (4.9)$$

Pour des raisons de stabilité sur  $\mathbf{E}$ , nous restreignons  $x$  à l'intervalle  $[0, x^*]$  où  $x^*$  satisfait

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i(x^*, x^*)] > 0.$$

Soient  $b$  un vecteur fixé de  $\mathbb{R}^p$  et  $\beta_0$  la vraie valeur du vecteur des paramètres de régression. Nous supposons que le vrai paramètre est  $\beta + b/\sqrt{n}$ . Sous les conditions de régularité suivantes

1. Il existe une constante  $B$  telle que  $\|\mathbf{Z}_i(0)\|_{L_1} + \int_0^{x^*} \|d\mathbf{Z}_i(t)\|_{L_1} \leq B$  ;
2. Pour tout  $k = 0, 1, 2$ , il existe  $\Gamma_k(t, x)$  tel que pour tout  $(t, x) \in \{(t, x) : 0 \leq t \leq x \leq x^*\}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \mathbf{Z}_i^{\otimes k}(t) Y_i(t, x) \exp\{\beta^x \mathbf{Z}_i(t)\} \right] = \Gamma_k(t, x),$$

le processus  $n^{-1/2}U(\beta; \cdot, \cdot)$  converge en distribution vers  $\xi + \mu b$ , où  $\xi$  est un processus gaussien de moyenne nulle et de fonction de covariance

$$\mathbb{E}[\xi(t_1, x_1)\xi(t_2, x_2)] = \int_0^{t_1 \wedge t_2} w(u, x_1)w(u, x_2) \left[ \Gamma_2(u, x_1 \wedge x_2) - \frac{\Gamma_1^{\otimes 2}(u, x_1 \wedge x_2)}{\Gamma_0(u, x_1 \wedge x_2)} \right] \lambda_0(u) du,$$

avec  $w_n$  une suite de fonctions convergeant uniformément sur  $\{(t, x) : 0 \leq t \leq x \leq x^*\}$  vers 0 et où

$$\mu(t, x) = \int_0^t \left[ \Gamma_2(u, x) - \frac{\Gamma_1^{\otimes 2}(u, x)}{\Gamma_0(u, x)} \right] \lambda_0(u) du.$$

**Éléments de preuve du Théorème 4.7** Nous pouvons réécrire

$$n^{-1/2}U(\beta; t, x) = n^{-1/2} \sum_{i=1}^n \int_0^t [\mathbf{Z}_i(u) - \mathbf{E}(\beta_0; u, x)] M_i(\beta_n; du, x) \quad (4.10)$$

$$+ n^{-1/2} \sum_{i=1}^n \int_0^t [\mathbf{Z}_i(u) - \mathbf{E}(\beta_0; u, x)] \times [\exp(\beta'_n \mathbf{Z}_i(u)) - \exp(\beta'_0 \mathbf{Z}_i(u))] Y_i(u, x) \lambda_0(u) du. \quad (4.11)$$

Le premier terme de la partie droite de l'Égalité 4.10 converge vers  $\xi$  tandis que le second terme converge uniformément vers  $b \int_0^t \left[ \Gamma_2(u, x) - \frac{\Gamma_1^{\otimes 2}(u, x)}{\Gamma_0(u, x)} \right] \lambda_0(u) du$ . ■

Le Théorème 4.7 montre que le processus de score à deux indices de temps,  $t$  et  $x$  mesurés sur un même calendrier,  $n^{-1/2}U(\beta; \cdot, \cdot)$  converge en distribution vers un processus gaussien. Nous allons appliquer d'une manière indirecte ce théorème pour montrer que la statistique normalisée de type log-rank  $n^{-1/2}U(\cdot)$  converge vers un processus gaussien. En effet, nous ne pouvons appliquer directement le Théorème 4.7 car la statistique normalisée de type log-rank dépend d'un taux  $x$  de dégradation de qualité de vie et du temps  $t$  variant chacun dans des domaines séparés.

**Preuve du Théorème 4.6** Les patients  $i$  étant indépendants, les variables aléatoires  $\mathbf{Z}_i = \mathbf{1}_{(i \in B)}$  sont indépendantes. À l'instant  $T_i(x)$ ,  $U_n(x)$  peut être réécrite

$$\begin{aligned} U_n(x) &= \sum_{i=1}^n \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j \mathbf{1}_{(T_j(x) \geq T_i(x))}}{\sum_{j=1}^n \mathbf{1}_{(T_j(x) \geq T_i(x))}} \right\} \\ &= \sum_{i=1}^n \int \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j \mathbf{1}_{(T_j(x) \geq T_i(x))}}{\sum_{j=1}^n \mathbf{1}_{(T_j(x) \geq T_i(x))}} \right\} \delta \mathbf{1}_{(T_i(x)=t)} \\ &= \sum_{i=1}^n \int \{ \mathbf{Z}_i - \bar{Z}(t, x) \} \{ d\mathbf{1}_{(T_i(x) \leq t)} - \mathbf{1}_{(T_i(x) \geq t)} d\Lambda_x(t) \} \end{aligned}$$

avec la notation  $\bar{Z}(t, x) = \frac{\sum_{j=1}^n \mathbf{Z}_j \mathbf{1}_{(T_j(x) \geq T_i(x))}}{\sum_{j=1}^n \mathbf{1}_{(T_j(x) \geq T_i(x))}}$  et où  $\Lambda_x(t)$  est la fonction de risque cumulé mesurée à l'instant  $T_i(x)$ . Il est facile de voir que  $M_i(t, x) = \mathbf{1}_{(T_i(x) \leq t)} - \int_0^t \mathbf{1}_{(T_i(x) \geq s)} d\Lambda_x(s)$  est une martingale de compensateur  $\mathbf{1}_{(T_i(x) \leq t)}$ . Par conséquent,

$$\frac{1}{\sqrt{n}} U_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int \{ \mathbf{Z}_i - \bar{Z}(t, x) \} M_i(dt, x).$$

Pour tout taux  $x$  de dégradation de qualité de vie fixé,  $n^{-1/2}U_n(x)$  est une variable aléatoire gaussienne; nous allons montrer que lorsque  $x$  varie,  $n^{-1/2}U_n(\cdot)$  est un processus gaussien.

Sous l'hypothèse  $H_0$  les variables aléatoires  $T_i(x)$  sont indépendantes et identiquement distribuées. Posons  $g_Z(t, x) = \lim_{n \rightarrow \infty} \bar{Z}(t, x)$ , nous obtenons

$$\frac{1}{\sqrt{n}} U_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int \{\mathbf{Z}_i - g_Z(t, x)\} M_i(dt, x) + op(1).$$

Nous obtenons une somme de  $n$  variables aléatoires indépendantes et de moyenne nulle. De plus,  $T_i(\cdot)$  est une fonction monotone (en effet, si  $x_1$  est plus grand que  $x_2$  alors,  $T_i(x_1)$  est plus grand que  $T_i(x_2)$ ). Puisque  $\mathbf{Z}_i = \mathbf{1}_{(i \in B)}$ , qu'il existe une constante  $C$  vérifiant  $\|\mathbf{Z}_i\|_1 + \int_0^{t^*} \|d\mathbf{Z}_i\|_1 \leq C$  et que pour tout  $k = 0, 1, 2$ , et pour tout  $t, x$ , la limite quand  $n$  tend vers l'infini de  $\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \mathbf{Z}_i^{\otimes k} Y_i(t, x) \exp(\beta'_0 \mathbf{Z}_i) \right]$  existe, par application indirecte du Théorème 4.7  $n^{-1/2} U_n(\cdot)$  converge en distribution vers un processus gaussien de fonction de covariance  $\Gamma$  définie par,

$$\begin{aligned} \Gamma(x, x') &= \lim_{n \rightarrow \infty} \mathbb{E} \left( \frac{1}{n} U_n(x) U_n(x') \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \int \{(\mathbf{Z}_i - \bar{Z}(t, x)) M_i(dt, x)\} \int \{(\mathbf{Z}_i - \bar{Z}(t, x')) M_i(dt, x')\} \right]. \end{aligned}$$

■

Soit  $\hat{\Gamma}$  l'estimateur de  $\Gamma$  défini comme suit,

$$\hat{\Gamma}(x, x') = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \int \{(\mathbf{Z}_i - \bar{Z}(t, x)) \hat{M}_i(dt, x)\} \int \{\mathbf{Z}_i - \bar{Z}(t, x) \hat{M}_i(dt, x')\} \right].$$

**Proposition 4.1.**  $\lim_{n \rightarrow \infty} \hat{\Gamma}(x, x') = \Gamma(x, x')$ .

**Preuve de la Proposition 4.1** La limite est une conséquence de la consistance des estimateurs. ■

#### 4.4.2 Test global de type log-rank

Malheureusement aucun estimateur  $\hat{\Gamma}$  de  $\Gamma$  n'est connu. Aussi, afin d'obtenir une approximation du supremum de la statistique du test de type log-rank  $\sup_{0 < x < 1} |n^{-1/2} U_n(x)|$  nous utilisons la méthode graphique. En effet, cette procédure indique les imprécisions du modèle.

Soit  $(\xi_1, \dots, \xi_n)$  un vecteur de  $n$  variables aléatoires indépendantes et identiquement distribuées de loi  $\mathcal{N}(0, 1)$ . De plus, nous supposons que les  $\xi_j$  sont indépendants des individus  $i$  et par suite des données observées  $(X_i(x), \delta_i(x), \mathbf{Z}_i)$ . Considérons le score perturbé de type log-rank suivant,

$$\frac{1}{\sqrt{n}} U_n^*(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \int \{\mathbf{Z}_i - \bar{Z}(t, x)\} \hat{M}_i(dt, x).$$

Puisque la matrice de covariance du score perturbé de type log-rank

$$\begin{aligned} & \text{Cov} \left( \frac{1}{\sqrt{n}} U_n^*(x), \frac{1}{\sqrt{n}} U_n^*(x') \right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \int \{\mathbf{Z}_i - \bar{Z}(t, x)\} \hat{M}_i(dt, x) \int \{\mathbf{Z}_i - \bar{Z}(t, x')\} \hat{M}_i(dt, x') \right] \\ &= \hat{\Gamma}(x, x') \xrightarrow[n \rightarrow \infty]{} \Gamma(x, x'). \end{aligned}$$

Il s'en suit que  $n^{-1/2} U_n^*(x)$  a asymptotiquement la même covariance et par conséquent la même distribution asymptotique que  $n^{-1/2} U_n(x)$ . Aussi nous pouvons obtenir une approximation de  $n^{-1/2} U_n(x)$  en simulant un grand nombre de réalisations (i.e. 1000, 10000) à partir de  $n^{-1/2} U_n^*(x)$  en générant le vecteur des variables aléatoires  $(\xi_1, \dots, \xi_n)$  tandis que les données observées  $(X_i(x), \delta_i(x), \mathbf{Z}_i)$  sont fixes. Ceci nous donne un test statistique non-paramétrique permettant d'analyser l'effet d'un traitement par rapport à l'évolution longitudinale jointe de la qualité de vie et de la survie.

**Théorème 4.8 (Test global de type log-rank).** *Soient  $S \equiv \sup_{0 < x < 1} |n^{-1/2} U_n(x)|$  le supremum de la statistique de test de type log-rank et  $s$  l'observation de  $S$ . Considérons  $\hat{S} = \sup_{0 < x < 1} |n^{-1/2} U_n^*(x)|$  le supremum du score perturbé de type log-rank. La  $p$ -value du test global de type log-rank  $\mathbb{P}(S > s)$  peut être approximée par  $\mathbb{P}(\hat{S} > s)$  à partir d'un grand nombre de réalisations (i.e. 1000, 10000) de  $\hat{S}$ .*

**Remarque 4.4.** *Les simulations sont réalisées sous l'hypothèse nulle à partir de la courbe observée. Sous  $H_0$ , le processus  $U_n(x)$  est gaussien de moyenne nulle et par conséquent nous obtenons bien un test et son degré de signification.*

**Remarque 4.5.** *La technique graphique a déjà été utilisée pour le modèle linéaire généralisé (GLM) avec réponses indépendantes par Su and Wei (1991), pour le modèle de risque proportionné avec données de survie censurées par Lin et al. (1993), pour le modèle linéaire généralisé marginal avec réponses dépendantes par Lin et al. (2002), pour le modèle linéaire mixte généralisé par Pan and Lin (2005) et pour la procédure de test de Mizon-Richard par Martinussen et al. (2008). Sánchez et al. (2008) proposent et évaluent une classe de procédures graphiques ainsi que les tests associés afin d'examiner la distribution et la linéarité des hypothèses. La classe de procédures graphiques introduite est le modèle linéaire pour équations complexes.*

## 4.5 Application à une base de données

Dans cette section, nous illustrons le test statistique proposé sur une base de données issue d'un essai clinique sur le traitement du cancer colorectal métastatique. (Awad *et al.*, 2002; Mesbah *et al.*, 2004). Dans cet essai clinique, la qualité de vie est mesurée à l'aide de l'instrument QLQ-C30 (Aaronson *et al.*, 1993). C'est un auto-questionnaire comportant trente items. Par transformation linéaire nous obtenons quinze échelles réparties de la

façon suivante : cinq échelles fonctionnelles, neuf échelles de symptômes et une échelle de statut global de la santé / qualité de vie notée QL. Nous analysons uniquement l'échelle QL et nous caractérisons la dégradation de qualité de vie comme un état absorbant défini à l'instant  $t$  par  $D_t = (Q_0 - Q_t)Q_0^{-1}$ .

Dans cette illustration, la  $\hat{p}$ -value du test global de type log-rank est obtenue par 1000 réalisations de  $\hat{S}$ . Pour chaque illustration graphique le processus observé est donné par une courbe en trait plein tandis que les trente courbes simulées sont représentées en pointillé.

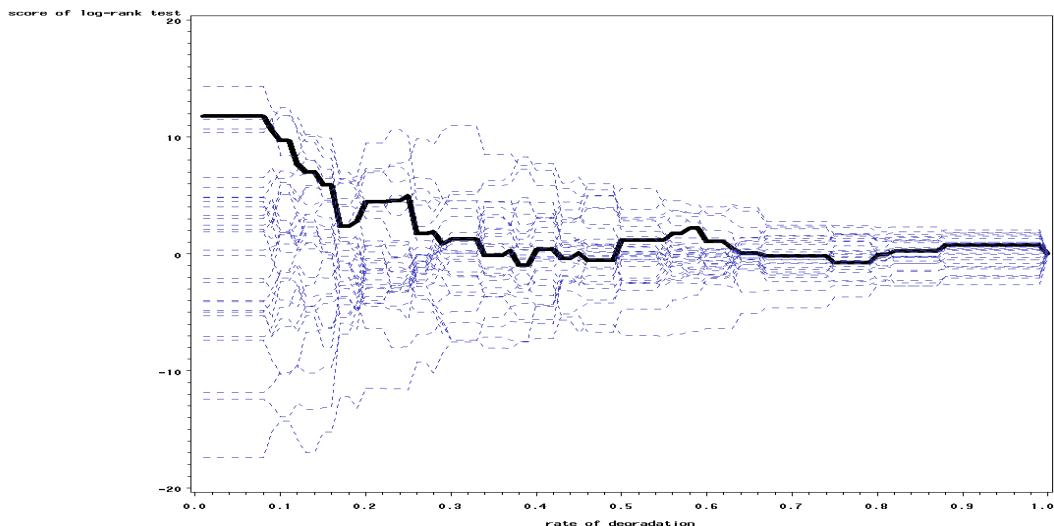


FIG. 4.1 – Score du processus  $n^{-1/2}U_n$  en fonction du taux de dégradation de QdV. Le temps de dégradation de QdV est exprimé en jours.

Cent trente deux patients sont suivis pendant un an dans le groupe A et cent trente et un dans le groupe B. La figure 4.1 représente le score du processus  $n^{-1/2}U_n$  en fonction du taux de dégradation de qualité de vie lorsque le temps de dégradation de qualité de vie est exprimé en jours. La  $\hat{p}$ -value du test global de type log-rank est égale à 0,094.

En raison de la durée des cycles du protocole et également d'une fréquence d'évolution de la qualité de vie différente entre les deux groupes, l'analyse de la dégradation de qualité de vie est effectuée par rapport à des intervalles correspondant à des périodes de traitement chimiothérapique de sept semaines. En effet, ceci permet une meilleure synchronisation du schéma des questionnaires entre les deux groupes de traitements. La figure 4.2 représente le score du processus  $n^{-1/2}U_n$  en fonction du taux de dégradation de qualité de vie lorsque le temps de dégradation de qualité de vie est exprimé en cycles. La  $\hat{p}$ -value du test global de type log-rank est égale à 0,032 aussi le test donne une différence plus significative entre les deux groupes. Ceci est cohérent avec les résultats obtenus par Awad *et al.* (2002) où aucun test global n'a été utilisé.

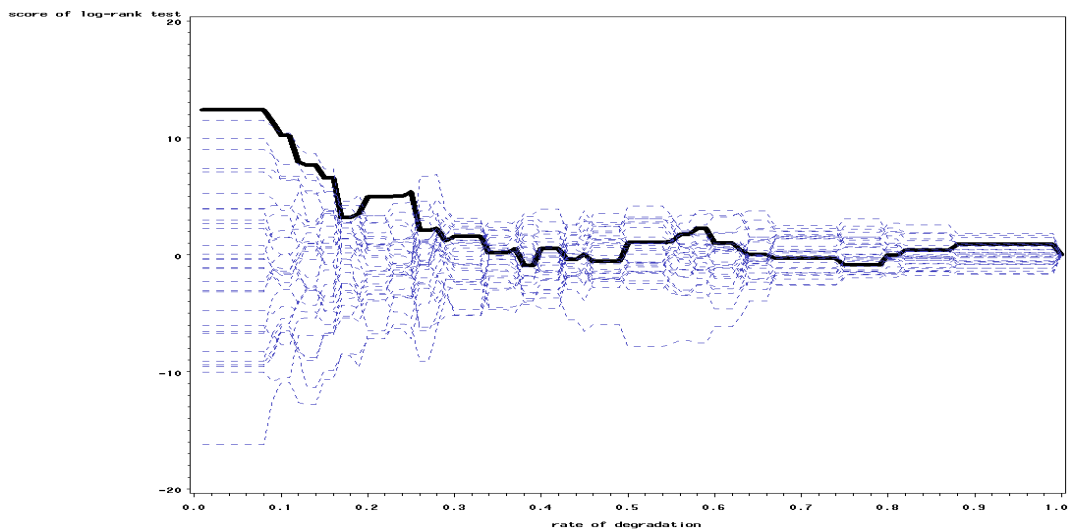


FIG. 4.2 – Score du processus  $n^{-1/2}U_n$  en fonction du taux de dégradation de QdV. Le temps de dégradation de QdV est exprimé en cycles de 7 semaines.



## Chapitre 5

# Simulations

Des simulations sont réalisées, à l'aide du logiciel SAS<sup>®</sup>, afin de comparer les performances du test global de type log-rank décrit dans le chapitre précédent. Ainsi, des jeux de données représentant l'évolution longitudinale de la qualité de vie de patients répartis entre deux groupes notés respectivement 0 et 1, ont été simulés. Pour cela, nous avons considéré que la qualité de vie des patients évolue de façon linéaire, i.e à l'instant  $t$ ,  $Q_t = a_0 + b_0t$  pour le groupe 0 et  $Q_t = a_1 + b_1t$  pour le groupe 1. Afin de représenter la variabilité des réponses des patients par rapport à la droite théorique, deux scénarii ont été envisagés. Pour le premier, la pente et le score de qualité de vie initial sont perturbés par un bruit blanc comportant un effet aléatoire ainsi qu'un effet fixe, tandis que pour le deuxième la pente et le score de qualité de vie initial sont perturbés par un processus auto-régressif d'ordre 1 (AR(1)) (Doob, 1953; Bosq and Lecoutre, 1987), dont la définition est rappelée en annexes. D'autre part, pour les deux scénarii étudiés les phénomènes de données manquantes et d'abandons sont pris en considération et simulés par des lois de Bernoulli de paramètre respectif 0,90 et 0,95. Les différents programmes des simulations figurent dans le chapitre programmes informatiques des annexes.

De plus, nous avons considéré deux grands cas de figure, l'un s'intéresse à des évolutions dissemblables entre les deux groupes de patients tandis que l'autre est associé à l'hypothèse nulle  $H_0$ . Ainsi,

1. dans le premier cas de figure nous avons posé les valeurs suivantes pour les paramètres  $a_0 = 28$ ,  $a_1 = 40$ ,  $b_0 = 1$  et  $b_1 = 15$  ;
2. dans le second cas de figure les paramètres deviennent  $a_0 = a_1 = 28$  et  $b_0 = b_1 = 1$ .

Enfin, nous avons étudié chaque cas de figure en jouant à chaque fois sur l'un des paramètres suivants : la taille de l'échantillon, le nombre de dates de visite, le pourcentage de patients inclus par groupe, le pourcentage d'abandons et de données manquantes et pour finir le nombre de courbes simulées. L'étude incorpore en parallèle les deux définitions de dégradation de qualité de vie  $D_t = (Q_0 - Q_t)Q_0^{-1}$  et  $D_t = Q_0 - Q_t$ . De plus, les valeurs de référence des différents paramètres sur lesquels nous jouons sont calquées sur les données issues de la Cohorte ANRS CO8 présentée dans le prochain chapitre. Les résultats obtenus sont répertoriés dans les tableaux qui suivent.



TAB. 5.1 – Résultats des simulations pour le premier cas de figure ( $a_0 = 28$ ,  $a_1 = 40$ ,  $b_0 = 1$  et  $b_1 = 15$ ) lorsque le nombre de patients varie. Constantes : 7 dates de visite, 100 courbes perturbées, présence de données manquantes et d'abandons, même taille pour chaque groupe de patients

Nombre de patients	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
30	0,05	0,02	0,02	0,02
300	0,07	0,01	0,03	< 0,01
500	0,06	0,01	0,01	0,03
1000	0,04	0,02	0,03	0,03

TAB. 5.2 – Résultats des simulations pour le second cas de figure ( $a_0 = a_1 = 28$  et  $b_0 = b_1 = 1$ ) lorsque le nombre de patients varie. Constantes : 7 dates de visite, 100 courbes perturbées, présence de données manquantes et d'abandons, même taille pour chaque groupe de patients

Nombre de patients	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
30	0,83	0,95	0,41	0,56
300	0,72	0,74	0,48	0,49
500	0,31	0,47	0,47	0,46
1000	0,23	0,16	0,55	0,43

TAB. 5.3 – Résultats des simulations pour le premier cas de figure ( $a_0 = 28$ ,  $a_1 = 40$ ,  $b_0 = 1$  et  $b_1 = 15$ ) lorsque le nombre de dates de visite varie. Constantes : 300 patients, 100 courbes perturbées, présence de données manquantes et d'abandons, même taille pour chaque groupe de patients

Nombre de dates	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
5	0,08	0,02	0,06	0,02
7	0,07	0,01	0,03	< 0,01
10	0,02	0,01	< 0,01	0,01

TAB. 5.4 – Résultats des simulations pour le second cas de figure ( $a_0 = a_1 = 28$  et  $b_0 = b_1 = 1$ ) lorsque le nombre de dates de visite varie. Constantes : 300 patients, 100 courbes perturbées, présence de données manquantes et d'abandons, même taille pour chaque groupe de patients

Nombre de dates	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
5	0,71	0,60	0,78	0,70
7	0,72	0,74	0,48	0,49
10	0,61	0,43	0,26	0,29

TAB. 5.5 – Résultats des simulations pour le premier cas de figure ( $a_0 = 28$ ,  $a_1 = 40$ ,  $b_0 = 1$  et  $b_1 = 15$ ) lorsque le pourcentage de patients inclus par groupe varie. Constantes : 7 dates de visite, 300 patients, 100 courbes perturbées, présence de données manquantes et d'abandons

Pourcentage	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
20%	0,06	< 0,01	0,01	0,10
33%	0,02	0,02	0,03	0,03
50%	0,07	0,01	0,03	< 0,01

TAB. 5.6 – Résultats des simulations pour le second cas de figure ( $a_0 = a_1 = 28$  et  $b_0 = b_1 = 1$ ) lorsque le pourcentage de patients inclus par groupe varie. Constantes : 7 dates de visite, 300 patients, 100 courbes perturbées, présence de données manquantes et d'abandons

Pourcentage	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
20%	0,77	0,47	0,07	0,04
33%	0,34	0,32	0,40	0,47
50%	0,72	0,74	0,48	0,49

TAB. 5.7 – Résultats des simulations pour le premier cas de figure ( $a_0 = 28$ ,  $a_1 = 40$ ,  $b_0 = 1$  et  $b_1 = 15$ ) suivant la présence ou non de données manquantes et d'abandons. Constantes : 7 dates de visite, 300 patients, 100 courbes perturbées, même taille pour chaque groupe de patients

Abandons	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
Sans abandon	0,05	0,07	0,04	0,06
Avec abandons	0,07	0,01	0,03	< 0,01

TAB. 5.8 – Résultats des simulations pour le second cas de figure ( $a_0 = a_1 = 28$  et  $b_0 = b_1 = 1$ ) suivant la présence ou non de données manquantes et d'abandons. Constantes : 7 dates de visite, 300 patients, 100 courbes perturbées, même taille pour chaque groupe de patients

Abandons	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
Sans abandon	0,48	0,32	0,32	0,24
Avec abandons	0,72	0,74	0,48	0,49

TAB. 5.9 – Résultats des simulations pour le premier cas de figure ( $a_0 = 28$ ,  $a_1 = 40$ ,  $b_0 = 1$  et  $b_1 = 15$ ) lorsque le nombre de courbes perturbées varie. Constantes : 7 dates de visite, 300 patients, présence de données manquantes et d'abandons, même taille pour chaque groupe de patients

Courbes perturbées	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
100	0,07	0,01	0,03	< 0,01
1000	0,05	0,05	0,03	0,05

TAB. 5.10 – Résultats des simulations pour le second cas de figure ( $a_0 = a_1 = 28$  et  $b_0 = b_1 = 1$ ) lorsque le nombre de courbes perturbées varie. Constantes : 7 dates de visite, 300 patients, présence de données manquantes et d'abandons, même taille pour chaque groupe de patients

$a_0 = a_1 = 28$ et $b_0 = b_1 = 1$				
Courbes perturbées	$\hat{p}$ -value du cas bruit blanc		$\hat{p}$ -value du cas AR(1)	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$	$(Q_0 - Q_t)Q_0^{-1}$	$Q_0 - Q_t$
100	0,72	0,74	0,48	0,49
1000	0,21	0,34	0,45	0,50

Nous pouvons constater que le nombre de dates de visite influe sur la puissance du test global de type log-rank. Ainsi, plus le nombre de dates de visite est important, plus le test global de type log-rank est puissant (Tables 5.3 et 5.4). La taille de l'échantillon ne semble modifier la puissance du test que dans le cas où la pente et le score de qualité de vie initial sont perturbés par un bruit blanc (Tables 5.1 et 5.2). Une grande disparité entre la taille respective de chaque groupe de patients rend difficilement interprétable le test global de type log-rank (Tables 5.5 et 5.6). Le présence ou non de données manquantes et d'abandon ne semble pas jouer sur la puissance du test global de type log-rank sauf dans le cas où la pente et le score de qualité de vie initial sont perturbés par un bruit blanc (Tables 5.7 et 5.8). Notons cependant que les abandons et les données manquantes sont distribués de manière complètement aléatoire, ce qui ne reflète pas bien la réalité de ce phénomène. Dès la réalisation d'une centaine de courbes perturbées le test global de type log-rank est tout à fait interprétable (Tables 5.9 et 5.10).

Les résultats obtenus en utilisant l'une ou l'autre des définitions de dégradation de qualité de vie  $D_t = (Q_0 - Q_t)Q_0^{-1}$  ou  $D_t = Q_0 - Q_t$  sont en adéquation. Ceci n'est pas très surprenant car nous avons simulé la qualité de vie à la date  $t$  à l'aide d'un modèle linéaire, i.e.  $Q_t = a_G + b_G t$  ( $G = 0$  ou  $1$ ). Ainsi, les deux définitions de la dégradation de qualité de vie peuvent se réécrire respectivement par  $D_t = (Q_0 - Q_t)Q_0^{-1} = -(b_G/a_G)t$  et par  $D_t = Q_0 - Q_t = -b_G t$ . Par conséquent, quelle que soit la définition de dégradation de qualité de vie retenue, celle-ci évoluera de façon parallèle.

Les différentes tables montrent que le test global de type log-rank donnent des résultats

en accord avec ceux escomptés dans le sens où sous l'hypothèse nulle  $H_0$  les résultats ne sont pas significatifs tandis que lorsque l'évolution de la qualité de vie est différente entre les deux groupes de patients les valeurs des  $\hat{p}$ -values du test global de type log-rank donnent des résultats significatifs pour un risque de 10%. Le test global de type log-rank n'est pas très puissant.



## Troisième partie

# Application à la Cohorte ANRS CO8 (APROCO - COPILOTE) de patients infectés par le VIH



## Chapitre 6

# La Cohorte ANRS CO8

Les progrès dans le domaine des multithérapies antirétrovirales hautement actives ont transformé l'infection à VIH en maladie chronique contraignant les patients à un traitement au long cours. Or, ces traitements présentent de graves limites comme la survenue de nombreux et importants effets indésirables pouvant altérer la qualité de vie des patients. Par suite, l'intérêt du corps médical ne porte plus seulement sur l'action thérapeutique et l'état de santé des patients mais aussi sur le ressenti de la qualité de vie par les patients eux-mêmes. Afin d'étudier la qualité de vie, se trouvant à l'interface de la médecine et de la psychologie, des bases de données sont constituées. C'est notamment le cas pour la Cohorte ANRS CO8 de patients infectés par le VIH.

Après une description de la Cohorte ANRS CO8, nous rappelons les principaux résultats et avancées obtenus sur cette Cohorte. Puis nous exposons les caractéristiques sociocomportementales et immunovirologiques des patients inclus dans cette Cohorte ainsi que les difficultés sous-jacentes à l'analyse de la dégradation de qualité de vie des patients constituant la Cohorte ANRS CO8.

### 6.1 Description de la Cohorte ANRS CO8

La Cohorte ANRS CO8 a connu deux grandes périodes dans son histoire. En effet, celle-ci a été initialisée en 1997 lors des premières prescriptions des multithérapies antirétrovirales hautement actives sous le nom de Cohorte APROCO. En 2002, devant l'intérêt suscité par la connaissance à très long terme des effets des antirétroviraux la Cohorte APROCO a été reconduite sous le nom de Cohorte COPILOTE. À l'heure actuelle dans un souci d'harmonisation, de la part de l'Agence Nationale de Recherche sur le Sida et les hépatites virales, la Cohorte APROCO-COPILOTE a été rebaptisée Cohorte ANRS CO8.

#### 6.1.1 La Cohorte APROCO

La Cohorte APROCO (1997) « Anti-PROtéase COhorte » est une cohorte d'observation nationale, prospective, multicentrique et multidisciplinaire constituée de 1281 pa-



tients adultes volontaires infectés par le VIH-1 ayant initié un traitement par inhibiteur de protéase entre 1997 et 1999. L'objectif principal de cette Cohorte est d'étudier l'évolution clinique et biologique des patients infectés par le VIH-1 ayant débuté un traitement avec antiprotéase dans le contexte de la pratique des prescriptions antirétrovirales en 1997. Une question spécifique concerne l'étude des obstacles, en particulier liés à l'observance et à la tolérance à long terme des antirétroviraux hautement actifs.

L'équipe médicale assure les suivis médicaux ainsi que le recueil de données aux dates définies par le protocole d'observation mis en place. Par ailleurs, un questionnaire a été proposé aux patients inclus dans l'étude à M0 (date de l'initialisation du traitement), ainsi qu'à M1, M4, puis tous les 8 mois. Le questionnaire administré, que les patients remplissent par eux-mêmes sans aide médicale, contient une évaluation de la qualité de vie, l'échelle générique SF-36 (Leplège *et al.*, 1995, 1998). Cette échelle permet de calculer deux scores, un score agrégé de qualité de vie physique et un score agrégé de qualité de vie mentale.

### 6.1.2 La Cohorte COPILOTE

La Cohorte COPILOTE (2002) « COHorte de Patients mis sous Inhibiteur de protéase, suivi LOng TERme » a inclus 716 patients en 2003 et est la continuité de la Cohorte APROCO. L'objectif principal de la Cohorte COPILOTE est d'étudier l'évolution clinique et biologique, ainsi que l'observance et la tolérance aux antirétroviraux hautement actifs de patients infectés par le VIH-1 et mis sous inhibiteur de protéase pour la première fois entre 1997 et 1999.

Le suivi dans la Cohorte COPILOTE s'effectue à la suite de la Cohorte APROCO de façon à ce que le premier bilan de la Cohorte COPILOTE coïncide avec un bilan annuel, i.e. M48, M60 ou M72, du calendrier de la Cohorte APROCO. De plus, pour la Cohorte COPILOTE l'échelle de qualité de vie proposée, tous les 12 mois, aux patients est l'échelle spécifique WHOQOL-HIV (Carrieri *et al.*, 2003; WHOQOL-HIV, 2004). En effet, l'échelle WHOQOL-HIV est spécifique pour l'étude des patients infectés par le VIH traités par multithérapie contrairement à l'échelle générique SF-36.

## 6.2 Rappel des résultats déjà obtenus sur la Cohorte ANRS CO8

La Cohorte APROCO ayant permis de montrer l'efficacité des antirétroviraux (Lewden *et al.*, 2002) a suscité un grand intérêt pour l'identification des facteurs pronostiques de la réponse immunovirologique et de la survie. L'observance est le déterminant majeur de la réponse virologique précoce tandis que la charge virale, le taux de CD4 et l'âge sont des facteurs prédictifs pour la réponse immunologique précoce (Le Moing *et al.*, 2001, 2005). Une observance initiale à 100% est cruciale pour obtenir sur le long terme une réponse virologique mais également un succès immunologique (Carrieri *et al.*, 2003). L'approche longitudinale de la Cohorte APROCO a mis en évidence le phénomène dynamique de l'ob-

servance thérapeutique. Si certains patients initialement observants deviennent à d'autres moments moins bien observants, le phénomène inverse est également constaté pour d'autres patients (Carrieri *et al.*, 2001). L'observance est expliquée par le vécu thérapeutique des patients lors des quatre premiers mois de traitement par inhibiteur de protéase et ceci indépendamment du niveau de qualité de vie pré-thérapeutique, du succès immunovirologique et de l'ancienneté de l'infection par le VIH (Spire *et al.*, 2002). Le risque de devenir moins bien observant augmentant de 6% pour chaque nouvel effet indésirable ressenti par le patient, le nombre d'effets indésirables perçus joue un rôle important dans le maintien d'une bonne observance. Le syndrome lipodystrophique provoquant une modification de l'image corporelle est un facteur de risque de rupture d'observance (Carrieri *et al.*, 2001). Une mauvaise observance à plus ou moins longue échéance peut être due à un état dépressif mais également à une mauvaise qualité de vie (Préau, 2004). Aussi si l'on souhaite maintenir une observance maximale afin de garantir l'efficacité des stratégies thérapeutiques et limiter les phénomènes de résistances aux traitements acquises graduellement par le VIH, l'étude de la qualité de vie est indispensable. La qualité de vie des patients augmente durant la première année qui suit la mise sous inhibiteur de protéase et se stabilise ensuite (Protopopescu *et al.*, 2007). Les facteurs influençant la qualité de vie ont été recherchés. Les patients contaminés par usage de drogue présentent une moins bonne qualité de vie que les autres patients infectés par le VIH. Les patients en stade SIDA lors de l'initialisation du traitement antirétroviral hautement actif (HAART) paraissent avoir une meilleure qualité de vie mentale mais cet effet ne perdure pas (Préau, 2004). Une mauvaise relation médecin-patient, un soutien social défectueux ainsi que l'apparition de lipodystrophies chez les femmes, de lipoatrophies pour les hommes affectent la qualité de vie des patients (Préau *et al.*, 2004). Les prédicteurs d'une mauvaise qualité de vie sont la co-infection VIH-VHC, un faible taux de CD4, le nombre d'effets indésirables, lipodystrophies exceptées (Protopopescu *et al.*, 2007). Un changement de traitement a un effet délétère sur la qualité de vie mentale des patients, confirmant l'importance de l'étude tant sur le plan médical que psychosocial lors de la comparaison des différentes stratégies thérapeutiques. Les facteurs associés aux données manquantes ont été recherchés et il apparaît qu'une charge virale indétectable, un niveau d'étude inférieur au baccalauréat ou un nombre important d'effets indésirables dus au traitement sont des facteurs favorisant le non-rendu de questionnaires (Valerio, 2004). Les données manquantes portant sur la non-observance, après cinq ans de suivi, principalement dues aux omissions des items relatifs à l'observance, mais aussi au non-respect des visites médicales prévues, induisent un biais non négligeable. Le motif des facteurs liés à l'observance durant la phase de maintenance après ajustement du biais diffère du motif obtenu sans correction de celui-ci. Il apparaît ainsi que les patients nés hors de l'Union Européenne sont plus observants, et que le schéma de prise de médicaments en deux fois par jour est le mieux respecté (Carrieri *et al.*, 2006). Par ailleurs, sous l'hypothèse que le modèle est paramétrique, les abandons n'ont pas d'effet significatif sur les estimateurs longitudinaux de la qualité de vie (Protopopescu *et al.*, 2007). Un problème crucial est de savoir s'il est préférable de prescrire des traitements plus puissants mais

engendrant davantage d'effets indésirables ou s'il n'est pas bénéfique d'administrer des combinaisons légèrement moins puissantes afin de privilégier la qualité de vie des patients. En effet, par exemple, la combinaison lopinavir/ritonavir (Kaletra) en induisant peu de résistances apparaît plus efficace que le nelfinavir. Par contre, cette combinaison provoque davantage de difficultés pour le patient à suivre le traitement, pouvant entraîner la survenue d'effets indésirables comme des troubles gastro-intestinaux avec cependant peu d'effets indésirables graves imputables au lopinavir et pouvant amener des arrêts de traitement (Brunet-François *et al.*, 2007). Par ailleurs, le syndrome lipodystrophique apparaît avec une occurrence identique en utilisant la combinaison lopinavir/ritonavir ou le nelfinavir (Cvetkovic and Goa, 2003).

### 6.3 Description de la base de données

Les caractéristiques socio-démographiques et cliniques à l'inclusion des 343 patients restant de la Cohorte ANRS CO8 sont présentées dans la Table 6.1. Les patients sont principalement des hommes, nés en Union Européenne et vivant dans des logements confortables. Les patients sont en grande majorité âgés entre 30 et 50 ans avec une moyenne d'âge de 40 ans. Deux tiers des patients ont un partenaire stable mais seulement 40% des patients ont au moins un enfant. 40% des patients ont un niveau d'étude supérieur ou égal au BAC et 60% ont un emploi. Lors de leur inclusion dans la Cohorte ANRS CO8 tous les patients débutent une combinaison antirétrovirale hautement active incluant un inhibiteur de protéase (IP). 42% des patients n'avaient jamais reçu de traitement antirétroviral auparavant. À l'inclusion, la durée moyenne (et l'intervalle de confiance) depuis le diagnostic de l'infection à VIH est de 54 (4-105) mois. Les patients inclus dans la Cohorte ANRS CO8 ont été principalement infectés par contact homosexuel ou hétérosexuel et 23% des patients sont classés en stade SIDA avéré. À l'inclusion, un peu moins de 7% des patients ont une charge virale indétectable pour un taux moyen (et un intervalle de confiance) de CD4+ de 316 (92-540) copies par millimètre cube.

La relation médecin-patient, dont l'évolution est décrite dans la Table 6.2, est un facteur psychosocial fondamental de la qualité de vie des personnes vivant avec le VIH (Préau, 2006). À l'entrée dans l'étude 95% des patients font état de relations avec leur médecin plutôt bonnes voire même pour 60% très bonnes. Ce rapport médecin-patient de bonne qualité se confirme au cours du temps. L'évolution du soutien matériel et psychologique apporté aux patients est reportée dans la Table 6.3. En effet, il est important de souligner le rôle du soutien social qu'il s'agisse d'un soutien de la part du partenaire principal, de la famille, des amis car celui-ci est fortement lié à une meilleure qualité de vie (Préau *et al.*, 2004). Psychologiquement, la moitié des patients se sentent beaucoup soutenus par leur partenaire principal et plus d'un tiers se déclarent beaucoup soutenus par les membres de leur famille ou par leurs ami(e)s et proches. Par contre, d'un point de vue matériel si un peu moins de 40% des patients se considèrent soutenus par leur partenaire principal, entre 40 et 50% n'ont pas du tout de soutien matériel de la part des membres de leur famille,

TAB. 6.1 – **Caractéristiques socio-démographiques et cliniques des patients de la Cohorte ANRS CO8 à l'inclusion (M0)**

Caractéristiques des patients	pourcentage de réponses positives	taille de la population
Sexe masculin	80,47	343
Déclare être né(e) en UE	82,80	343
A au moins un enfant	39,50	319
Niveau supérieur au BAC	39,94	313
Possède un emploi	59,57	323
Est propriétaire ou locataire	87,65	324
Logement		
très confortable	45,14	319
assez confortable	44,51	319
peu confortable	7,52	319
pas du tout confortable	2,82	319
Partenaire principal	68,52	324
Âge - <i>ans</i> ( $\mu$ ( $\sigma$ ))	39,81 (9,92)	343
moins de 30 ans	10,50	343
entre 30 et 35 ans	22,45	343
entre 35 et 40 ans	24,20	343
entre 40 et 50 ans	28,57	343
plus de 50 ans	14,29	343
Stade clinique		
A	44,14	324
B	32,72	324
C	23,15	324
Groupe de transmission		
toxicomane	9,88	324
homosexuel	45,37	324
hétérosexuel	32,74	324
autres et données manquantes	12,04	324
Naïf de tout traitement	42,28	324
Durée depuis le diagnostic de la séropositivité - <i>mois</i> ( $\mu$ ( $\sigma$ ))	54,17 (50,47)	322

$\mu$ =moyenne ;  $\sigma$ =écart-type

TAB. 6.2 – Évolution des caractéristiques liées à l'alcool, médicales et de la dépression des patients de la Cohorte ANRS C08 (en pourcentage)

	M0	M72	M84	M96	M108
<b>CONSOMMATION D'ALCOOL</b>					
Alcool 4 dernières semaines	n=321	n=339	n=339	n=267	n=212
jamais	12,15	16,41	17,99	18,73	16,51
moins d'un fois par mois	16,20	16,81	15,04	16,85	18,87
1 à 4 fois par mois	28,97	34,22	33,63	31,84	31,13
2 à 6 fois par semaine	21,81	18,58	17,11	17,60	17,45
tous les jours	20,87	13,57	16,22	14,98	16,04
Nombre de verres d'alcool	n=260	n=280	n=272	n=218	n=176
1 ou 2 verres	66,15	73,57	70,96	79,82	71,02
3 ou 4 verres	26,64	19,64	22,06	15,60	23,86
5 ou 6 verres	5,38	4,64	4,04	3,67	3,41
7 verres ou plus	1,92	2,14	2,94	0,92	1,70
<b>DONNÉES MÉDICALES</b>					
État des relations avec le médecin		n=338	n=334	n=268	n=211
très mauvaises		0,00	0,00	0,37	0,00
plutôt mauvaises		0,30	0,30	0,00	0,00
neutres		4,44	5,39	3,73	4,27
plutôt bonnes		35,21	34,43	31,34	29,38
très bonnes		60,06	59,88	64,55	66,35
Charge virale	n=323	n=342	n=341	n=229	n=97
indélectable	6,81	68,71	72,43	75,55	72,16
Sous IP d'après auto-questionnaire		n=343	n=343	n=269	n=215
oui		39,07	40,23	43,12	44,19
Nombre de prises		n=305	n=315	n=250	n=188
1 prise		14,43	21,90	30,00	31,38
2 prises		77,05	73,02	66,40	66,49
3 prises ou plus		8,52	5,08	3,60	2,13
<b>DÉPRESSION</b>					
Dépression dichotomique	n=303	n=334	n=337	n=264	n=208
dépression	38,61	30,24	32,05	31,44	26,44

n=taille de la population ; IP=inhibiteur de protéase

TAB. 6.3 – Évolution du soutien matériel et psychologique des 343 patients de la Cohorte ANRS CO8 (en pourcentage)

	M0	M72		M84		M96		M108	
		psycho	mat	psycho	mat	psycho	mat	psycho	mat
Par le partenaire principal									
n	316	300	291	288	278	235	277	184	177
pas du tout	3,16	9,67	18,21	7,19	15,83	11,06	21,59	7,61	16,95
un peu	3,80	6,67	10,31	4,17	12,59	6,81	5,73	4,35	5,65
moyennement	8,54	9,00	7,22	12,50	10,79	11,91	12,33	9,78	12,43
beaucoup	51,58	53,67	39,18	58,68	39,93	48,09	33,48	57,07	39,55
pas concerné(e)	32,91	21,00	25,09	17,36	20,86	22,13	26,87	21,20	25,42
Par les membres de la famille									
n	306	319	313	325	317	255	248	202	195
pas du tout	32,68	26,02	40,58	21,23	44,16	20,39	41,94	20,30	40,00
un peu	10,13	10,97	12,78	12,92	9,46	10,98	14,11	11,39	12,31
moyennement	11,44	11,60	11,50	16,92	11,67	14,90	8,06	12,87	9,23
beaucoup	43,14	33,54	16,61	33,85	14,51	34,12	13,31	41,58	19,49
pas concerné(e)	2,61	17,87	18,53	15,08	20,19	19,61	22,58	13,86	18,97
Par les amis et les proches									
n	292	326	319	327	320	255	248	205	196
pas du tout	19,18	17,48	48,59	17,43	48,75	21,18	51,21	20,49	47,45
un peu	10,62	17,18	8,78	11,93	9,69	10,59	6,85	9,27	6,63
moyennement	20,21	15,95	7,21	18,04	9,69	15,29	7,66	15,12	10,20
beaucoup	46,58	35,89	14,42	35,17	7,81	34,90	9,27	37,07	9,69
pas concerné(e)	3,42	13,50	21,00	17,43	24,06	18,04	25,00	18,05	26,02

n=taille de la population ; mat=matériel ; psycho=psychologique

de leurs ami(e)s et des proches. Par ailleurs, entre M0 et M108 un tiers des patients ont présenté un état dépressif au moins une fois.

À M72, 39% des patients sont toujours sous inhibiteur de protéase. Il est important de noter que si à l'inclusion seulement 7% des patients présentent une charge virale indétectable, à M72 ce pourcentage est dix fois plus élevé passant ainsi à 69%. Le taux de patients présentant une charge virale indétectable reste ensuite toujours au dessus de 72%. Parallèlement, le taux moyen de CD4+ de 316 copies par millimètre cube à l'inclusion passe et se maintient à environ 600 copies par millimètre cube dès M72, confirmant l'efficacité thérapeutique des puissantes combinaisons antirétrovirales. Par ailleurs, le nombre de symptômes ressentis par les patients augmente progressivement entre M72 et M108 passant de 8 symptômes dont 3 signalés gênants à 9 symptômes dont 4 perçus comme gênants.

La Table 6.4 expose l'évolution des différents scores de qualité de vie des patients.

TAB. 6.4 – Évolution du nombre de symptômes, du statut immuno-virologique et des scores de dépression et de qualité de vie des 343 patients de la Cohorte ANRS CO8

	M0	M28	M44	M72	M84	M96	M108
	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )
<b>ÉVOLUTION DES SYMPTÔMES</b>							
nombre de TM	2,67 (2,70)	2,66 (2,74)	2,96 (2,87)	2,90 (2,84)			
nombre de symptômes hors TM	5,55 (4,10)	6,15 (4,23)	6,11 (4,22)	6,15 (4,21)			
nombre de tous symptômes	8,22 (5,76)	8,81 (6,07)	9,07 (6,10)	9,05 (6,23)			
nombre de TM gênants	1,08 (2,02)	1,21 (2,17)	1,59 (2,32)	1,35 (2,25)			
nombre de symptômes hors TM gênants	2,02 (2,96)	2,04 (2,97)	2,31 (3,11)	2,45 (3,20)			
nombre de tous symptômes gênants	3,10 (4,28)	3,25 (4,69)	3,90 (4,62)	3,80 (4,70)			
<b>STATUT IMMUNO-VIROLOGIQUE</b>							
CD4 bouché	316,36 (224,63)		582,42 (307,14)	594,92 (321,48)	607,83 (301,62)	605,46 (262,00)	
<b>DÉPRESSION</b>							
score de dépression	17,17 (11,10)		14,71 (10,93)	15,24 (11,99)	14,50 (11,03)	13,61 (10,76)	
<b>SCORES DE QUALITÉ DE VIE</b>							
<b>SF-12</b>							
PCS-12	49,07 (9,15)	51,09 (7,17)	50,93 (7,96)	49,30 (8,14)			
MCS-12	42,88 (11,32)	46,65 (10,88)	45,98 (11,04)	46,11 (11,76)			
<b>WHOQOL-HIV</b>							
domaine physique	14,96 (3,15)	15,02 (3,08)	15,03 (3,06)	15,02 (3,31)			
domaine psychologique	14,77 (3,02)	14,86 (2,99)	14,78 (2,88)	14,99 (2,78)			
niveau d'indépendance	15,58 (3,48)	15,66 (3,45)	15,80 (3,27)	15,88 (3,38)			
relation sociale	14,84 (3,31)	15,08 (3,16)	14,91 (3,07)	15,27 (3,12)			
environnement	15,75 (2,53)	15,66 (2,62)	15,54 (2,64)	15,80 (2,64)			
spiritualité/religion/croyance personnelle	14,28 (3,50)	14,45 (3,46)	14,40 (3,31)	14,45 (3,17)			

TM=trouble métabolique ; PCS-12=score agrégé physique SF-12 ; MCS-12=score agrégé mental SF-12  
 $\mu$ =moyenne ;  $\sigma$ =écart-type

Entre M0 et M72 le questionnaire utilisé est l'échelle générique SF-12 incluse sans perte d'information dans l'échelle générique SF-36 et permettant de calculer deux scores agrégés de qualité de vie l'une physique (PCS-12) et l'autre mentale (MCS-12). Le score agrégé de qualité de vie physique est légèrement supérieur au score agrégé de qualité de vie mentale. De plus, les deux scores agrégés de qualité de vie augmentent entre les dates M0 et M28 et restent relativement stables ensuite. Entre M72 et M108 le questionnaire administré a été remplacé par le questionnaire WHOQOL-HIV spécifique à l'infection à VIH. Cet outil permet de mesurer six domaines de qualité de vie. Les scores de qualité de vie obtenus pour chacun des six domaines sont relativement similaires et constants au cours du temps.

## **6.4 Difficultés rencontrées pour l'étude de la Cohorte ANRS CO8**

### **6.4.1 Réponses au questionnaire de qualité de vie à l'inclusion**

Les 343 patients restant dans l'étude n'ont pas tous répondu au questionnaire de qualité de vie lors de leur inclusion dans la Cohorte ANRS CO8, ce qui ne permet pas de calculer pour ces patients les scores ultérieurs de dégradation de qualité de vie. Par conséquent, l'étude inclura seulement les 279 patients ayant répondu lors de leur entrée dans la Cohorte ANRS CO8 au questionnaire de qualité de vie.

Les caractéristiques socio-démographiques et cliniques des patients ayant répondu au questionnaire de qualité de vie lors de leur inclusion, présentées dans la Table 6.5, ne montrent pas de différences notables avec l'ensemble des patients de la Cohorte ANRS CO8. Il en va de même quant à l'évolution des caractéristiques liées à l'alcool, médicales et de la dépression indiquées à la Table 6.6, ainsi que pour l'évolution du soutien matériel et psychologique reportée en Table 6.7, et pour l'évolution du nombre de symptômes, du statut immuno-virologique et des scores de dépression et de qualité de vie mentionnée en Table 6.8. En conséquence, la restriction de l'analyse aux seuls 279 patients ayant complété le questionnaire de qualité de vie lors de leur incorporation dans la Cohorte ANRS CO8 ne devrait pas influencer sur la pertinence des résultats de l'étude de la dégradation de qualité de vie des patients inclus dans la Cohorte ANRS CO8.

### **6.4.2 Changement de questionnaires et equating**

Contrairement à un grand nombre d'études sur l'observance et la qualité de vie qui sont faites de façon transversale, i.e. à un moment donné du temps, l'étude de la Cohorte ANRS CO8 est prospective et permet donc le suivi des patients dans le temps. Durant ce suivi des questionnaires de qualité de vie sont donnés à remplir aux patients. Les auto-questionnaires ont changé en cours d'étude en effet, durant les cinq premières années le questionnaire de qualité de vie est le SF-12 tandis que durant les quatre dernières années le questionnaire de qualité de vie est le WHOQOL-HIV. Ce remplacement de questionnaire s'explique par le fait que le SF-12, destiné à l'évaluation des services, des besoins et des



TAB. 6.5 – Caractéristiques socio-démographiques et cliniques des patients de la Cohorte ANRS CO8 ayant répondu au questionnaire de qualité de vie à l'inclusion (M0)

Caractéristiques des patients	pourcentage de réponses positives	taille de la population
Sexe masculin	79,57	279
Déclare être né(e) en UE	85,30	279
A au moins un enfant	37,82	275
Niveau supérieur au BAC	43,33	270
Possède un emploi	62,37	279
Est propriétaire ou locataire	88,17	279
Logement		
très confortable	46,74	276
assez confortable	43,48	276
peu confortable	7,25	276
pas du tout confortable	2,54	276
Partenaire principal	68,46	279
Âge - ans ( $\mu$ ( $\sigma$ ))	39,51 (9,72)	279
moins de 30 ans	11,47	279
entre 30 et 35 ans	23,30	279
entre 35 et 40 ans	22,94	279
entre 40 et 50 ans	28,32	279
plus de 50 ans	13,98	279
Stade clinique		
A	44,80	279
B	32,62	279
C	22,58	279
Groupe de transmission		
toxicomane	10,04	279
homosexuel	47,67	279
hétérosexuel	30,47	279
autres et données manquantes	11,83	279
Naïf de tout traitement	40,86	279
Durée depuis le diagnostic de la séropositivité - mois ( $\mu$ ( $\sigma$ ))	56,40 (50,44)	277

$\mu$ =moyenne ;  $\sigma$ =écart-type

TAB. 6.6 – **Évolution des caractéristiques liées à l'alcool, médicales et de la dépression des patients de la Cohorte ANRS CO8 ayant répondu au questionnaire de qualité de vie lors de leur inclusion (en pourcentage)**

	M0	M72	M84	M96	M108
<b>CONSOMMATION D'ALCOOL</b>					
Alcool 4 dernières semaines	n=278	n=277	n=276	n=216	n=182
jamais	12,59	15,52	17,03	18,06	14,84
moins d'une fois par mois	16,19	18,05	17,03	18,52	20,88
1 à 4 fois par mois	29,86	33,94	33,70	31,48	30,22
2 à 6 fois par semaine	21,94	18,77	16,30	17,13	18,13
tous les jours	19,42	13,72	15,94	14,81	15,93
Nombre de verres d'alcool	n=224	n=230	n=223	n=178	n=155
1 ou 2 verres	68,30	74,35	72,20	82,58	72,90
3 ou 4 verres	25,00	19,57	21,08	12,92	21,29
5 ou 6 verres	4,46	4,35	4,48	3,37	3,87
7 verres ou plus	2,23	1,74	2,24	1,12	1,94
<b>DONNÉES MÉDICALES</b>					
État des relations avec le médecin		n=277	n=274	n=216	n=182
très mauvaises		0,00	0,00	0,46	0,00
plutôt mauvaises		0,36	0,36	0,00	0,00
neutres		3,97	5,47	4,17	3,85
plutôt bonnes		36,82	36,86	34,26	30,77
très bonnes		58,84	57,30	61,11	65,38
Charge virale	n=279	n=278	n=277	n=190	n=84
indétectable	7,89	67,99	71,48	75,26	71,43
Sous IP d'après auto-questionnaire		n=279	n=279	n=216	n=184
oui		39,43	40,50	43,98	44,02
Nombre de prises		n=249	n=256	n=203	n=164
1 prise		14,06	21,48	31,53	32,93
2 prises		77,91	74,22	66,01	65,85
3 prises ou plus		8,03	4,30	2,46	1,22
<b>DÉPRESSION</b>					
Dépression dichotomique	n=272	n=274	n=276	n=214	n=180
dépression	37,13	29,56	31,16	29,44	25,56

n=taille de la population ; IP=inhibiteur de protéase

TAB. 6.7 – Évolution du soutien matériel et psychologique des 279 patients de la Cohorte ANRS CO8 ayant répondu au questionnaire de qualité de vie à l'entrée dans la Cohorte (en pourcentage)

	M0	M72		M84		M96		M108	
		psycho	mat	psycho	mat	psycho	mat	psycho	mat
Par le partenaire principal									
n	274	247	243	233	226	189	184	156	153
pas du tout	2,55	7,69	16,87	7,30	15,49	10,05	18,48	4,49	15,03
un peu	3,28	6,48	9,88	3,00	12,39	7,41	5,43	3,21	6,54
moyennement	8,76	8,50	7,41	12,02	10,62	10,58	13,59	8,97	9,80
beaucoup	53,28	53,44	37,86	57,94	38,94	47,62	33,15	60,26	40,52
pas concerné(e)	32,12	23,89	27,98	19,74	22,57	24,34	29,35	23,08	28,10
Par les membres de la famille									
n	265	262	259	267	263	210	204	174	171
pas du tout	32,83	24,43	41,31	20,60	46,01	19,52	41,67	20,11	40,35
un peu	9,06	11,45	12,74	14,23	9,89	11,43	15,20	11,49	14,04
moyennement	12,08	12,98	11,58	17,60	10,65	16,19	8,82	13,22	7,02
beaucoup	43,77	32,44	14,29	31,84	12,55	32,86	10,29	40,23	18,13
pas concerné(e)	2,26	18,70	20,08	15,73	20,91	20,00	24,02	14,94	20,47
Par les amis et les proches									
n	252	269	266	268	264	209	205	177	174
pas du tout	17,06	14,87	48,87	17,54	50,00	21,05	51,22	18,64	46,55
un peu	10,71	17,10	9,02	11,19	9,09	10,05	7,32	9,04	6,90
moyennement	23,41	17,10	6,77	18,28	9,85	15,79	7,80	15,82	10,34
beaucoup	45,24	36,06	13,16	35,07	6,06	33,97	6,83	37,29	8,62
pas concerné(e)	3,57	14,87	22,18	17,91	25,00	19,14	26,83	19,21	27,59

n=taille de la population ; mat=matériel ; psycho=psychologique

TAB. 6.8 – Évolution du nombre de symptômes, du statut immuno-virologique et des scores de dépression et de qualité de vie des 279 patients de la Cohorte ANRS CO8 ayant répondu au questionnaire de qualité de vie lors de leur inclusion

	M0	M28	M44	M72	M84	M96	M108
	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )
<b>ÉVOLUTION DES SYMPTÔMES</b>							
nombre de TM				2,61 (2,69)	2,60 (2,67)	2,93 (2,83)	3,02 (2,84)
nombre de symptômes hors TM				5,63 (4,09)	6,13 (3,94)	6,09 (4,07)	6,41 (4,28)
nombre de tous symptômes				8,23 (5,73)	8,73 (5,67)	9,02 (5,89)	9,42 (6,31)
nombre de TM gênants				1,11 (2,03)	1,22 (2,18)	1,55 (2,22)	1,41 (2,25)
nombre de symptômes hors TM gênants				2,09 (2,91)	2,04 (2,79)	2,24 (3,01)	2,60 (3,35)
nombre de tous symptômes gênants				3,19 (4,30)	3,27 (4,38)	3,79 (4,41)	4,02 (4,84)
<b>STATUT IMMUNO-VIROLOGIQUE</b>							
CD4 bouché	325,59 (225,25)			583,81 (300,89)	590,53 (307,32)	607,86 (295,56)	610,42 (267,45)
<b>DÉPRESSION</b>							
score de dépression	16,77 (10,81)			14,55 (10,94)	14,54 (11,24)	14,02 (10,69)	13,34 (10,27)
<b>SCORES DE QUALITÉ DE VIE</b>							
<b>SF-12</b>							
PCS-12	49,07 (9,15)	51,36 (6,88)	51,24 (7,90)	49,36 (8,19)			
MCS-12	42,88 (11,32)	46,46 (10,79)	46,33 (10,77)	46,19 (11,50)			
<b>WHOQOL-HIV</b>							
domaine physique				15,01 (3,09)	15,07 (3,05)	15,12 (3,06)	15,07 (3,26)
domaine psychologique				14,82 (2,88)	14,90 (2,91)	14,92 (2,80)	14,96 (2,67)
niveau d'indépendance				15,68 (3,38)	15,58 (3,50)	15,91 (3,20)	15,87 (3,44)
relation sociale				14,94 (3,16)	15,10 (3,14)	15,05 (2,90)	15,24 (3,07)
environnement				15,76 (2,53)	15,74 (2,65)	15,61 (2,64)	15,79 (2,54)
spiritualité/religion/croyance personnelle				14,25 (3,41)	14,40 (3,42)	14,38 (3,31)	14,30 (3,19)

TM=trouble métabolique; PCS-12=score agrégé physique SF-12; MCS-12=score agrégé mental SF-12

$\mu$ =moyenne;  $\sigma$ =écart-type

traitements, adopte le point de vue associé des experts et des patients, il s'agit donc plus de mesure de statut de santé ou de qualité de vie liée à la santé que de qualité de vie perçue par les patients eux-mêmes (Lepège and Coste, 2001). Ainsi, les échelles de qualité de vie génériques : SF-36 et SF-12 validées et largement utilisées chez des patients infectés par le VIH ont fini par montrer leurs limites, elles ne permettent pas de différencier précisément les différents états de qualité de vie physique et ne prennent pas en compte le rapport à la sexualité, par exemple. L'échelle WHOQOL-HIV est issue d'une échelle générique de qualité de vie à laquelle il a été rajouté cinq questions spécifiques des problématiques de l'infection par le VIH. Ainsi, le WHOQOL-HIV est un outil spécifique de mesure de la qualité de vie des patients infectés par le VIH (WHOQOL-HIV, 2004).

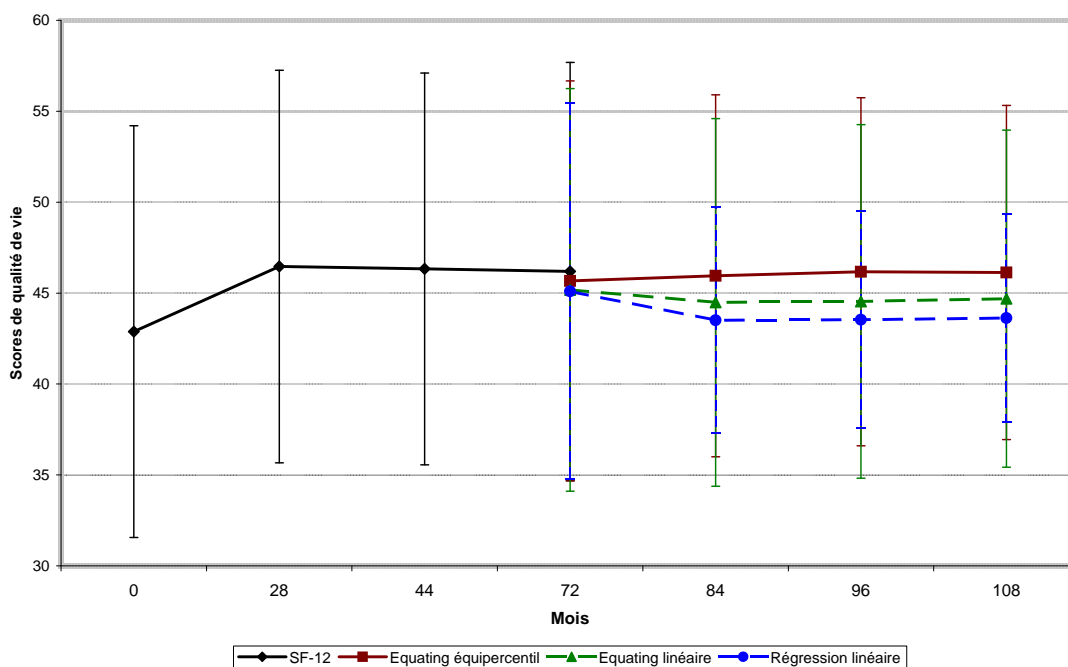


FIG. 6.1 – Évolution des différents scores mentaux de qualité de vie obtenus par equating des patients de la Cohorte ANRS CO8 ayant répondu lors de leur inclusion au questionnaire de qualité de vie

Afin d'obtenir un unique questionnaire de qualité de vie durant le suivi des patients le score agrégé mental du SF-12 a été estimé par equating à partir du domaine psychologique du WHOQOL-HIV. Trois méthodes d'equating ont été utilisées : l'equating équipercentil, l'equating linéaire et la régression linéaire. Les résultats de ces différentes méthodes sont montrés en rouge dans la Table 6.9 pour l'ensemble des patients de la Cohorte et dans

la Table 6.10 pour les patients ayant répondu au questionnaire de qualité de vie initial. L'évolution longitudinale des différents scores mentaux de qualité de vie est tracée dans la Figure 6.1. Le score agrégé mental du SF-12 observé entre les dates M0 et M72 est dessiné en noir tandis que les trois scores obtenus, entre les dates M72 et M108, par equating équipercatil, equating linéaire et régression linéaire sont tracés respectivement en rouge, en vert et en bleu. Nous pouvons voir que le score mental de qualité de vie augmente durant les 28 premiers mois et devient relativement stable ensuite. D'autre part, les différentes méthodes d'equating utilisées donnent des résultats proches. Toutefois à M72, date où les patients ont rempli les deux questionnaires de qualité de vie, les scores obtenus par equating sont légèrement inférieurs au score agrégé mental du SF-12, l'equating équipercatil fournit la valeur la plus proche de celle observée. Le score final analysé présenté en bleu dans la Table 6.10 est défini par le score agrégé mental du SF-12 lorsque celui-ci est observé et par son estimation par equating équipercatil autrement ; ce score étant le plus proche de la valeur observée en M72.

### 6.4.3 Problème des dates manquantes

Le Groupe d'étude de la Cohorte APROCO (APROCO, 1997) a fixé le calendrier du suivi de la qualité de vie des patients par auto-questionnaire à compléter lors de l'inclusion (M0), à M4 puis tous les huit mois jusqu'à M60. Pour la poursuite de l'étude dans la Cohorte COPILOTE (COPILOTE, 2002) le Groupe d'étude a programmé la délivrance des auto-questionnaires tous les douze mois avec le premier questionnaire remis correspondant à un suivi annuel (M48, M60 ou M72) du protocole antérieur. Ce protocole a connu des réajustements au cours de l'étude, par exemple à M48 un auto-questionnaire de qualité de vie a concerné le suivi des patients présentant un ou plusieurs troubles métaboliques. D'autre part, la date M60 correspond au questionnaire d'entrée dans la Cohorte COPILOTE où le SF-12 et le WHOQOL-HIV ont été posés selon les inclusions des patients à M48, M60 ou M72.

Le Comité Scientifique de la Cohorte ANRS CO8 a communiqué pour cette étude les données immunovirologiques et sociocomportementales collectées aux dates M0, M28, M44, M72, M84, M96 et M108, attachant une attention toute particulière à l'analyse des données les plus récentes issues de la Cohorte COPILOTE.

### 6.4.4 Problème des données manquantes

La Cohorte ANRS CO8 n'est pas exempte des difficultés inhérentes à tout suivi sur le long terme de patients que sont les données manquantes pouvant être dues soit aux données manquantes intermittentes, soit aux abandons et perdus de vue. Ainsi, sur les 1281 patients inclus à l'initiation de la Cohorte ANRS CO8, seuls 343 patients sont toujours présents dont seulement 279 patients peuvent être retenus pour une analyse longitudinale de la dégradation de qualité de vie.

Cette déperdition a fait l'objet d'une étude réalisée par Protopopescu *et al.* (2007) dont

TAB. 6.9 – Évolution des scores de qualité de vie obtenus par equating des 343 patients de la Cohorte ANRS CO8

	M0	M28	M44	M72	M84	M96	M108
	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )
<b>SCORES DE QUALITÉ DE VIE OBSERVÉS</b>							
Score agrégé mental SF-12	42,88 (11,32)	46,65 (10,88)	45,98 (11,04)	46,11 (11,76)			
Domaine psychologique du WHOQOL-HIV				14,77 (3,02)	14,86 (2,99)	14,78 (2,88)	14,99 (2,78)
<b>SCORES DE QUALITÉ DE VIE OBTENUS PAR EQUATING</b>							
MCS-12 equating équipercentil				45,65 (10,34)	45,85 (10,20)	45,70 (9,89)	46,25 (9,60)
MCS-12 equating linéaire				44,04 (10,48)	44,35 (10,40)	44,06 (10,02)	44,80 (9,68)
MCS-12 régression linéaire				43,24 (6,45)	43,42 (6,39)	43,25 (6,16)	43,70 (5,95)
<b>SCORE DE QUALITÉ DE VIE ANALYSÉ</b>							
MCS-12 equating équipercentil	42,88 (11,32)	46,65 (10,88)	45,98 (11,04)	46,11 (11,76)	45,85 (10,20)	45,70 (9,89)	46,25 (9,60)

$\mu$ =moyenne ;  $\sigma$ =écart-type

TAB. 6.10 – Évolution des scores de qualité de vie obtenus par equating des 279 patients de la Cohorte ANRS CO8 ayant rempli lors de l'inclusion le questionnaire de qualité de vie

	M0	M28	M44	M72	M84	M96	M108
	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )	$\mu$ ( $\sigma$ )
<b>SCORES DE QUALITÉ DE VIE OBSERVÉS</b>							
Score agrégé mental SF-12	42,88 (11,32)	46,46 (10,79)	46,33 (10,77)	46,19 (11,50)	14,90 (2,91)	14,92 (2,80)	14,96 (2,67)
Domaine psychologique du WHOQOL-HIV				14,82 (2,88)			
<b>SCORES DE QUALITÉ DE VIE OBTENUS PAR EQUATING</b>							
MCS-12 equating équipercantil				45,67 (11,00)	45,95 (9,95)	46,17 (9,57)	46,13 (9,19)
MCS-12 equating linéaire				45,18 (11,07)	44,49 (10,11)	44,54 (9,72)	44,69 (9,27)
MCS-12 régression linéaire				45,10 (10,35)	43,51 (6,22)	43,54 (5,98)	43,63 (5,71)
<b>SCORE DE QUALITÉ DE VIE ANALYSÉ</b>							
MCS-12 equating équipercantil	42,88 (11,32)	46,46 (10,79)	46,33 (10,77)	46,19 (11,50)	45,95 (9,95)	46,17 (9,57)	46,13 (9,19)

$\mu$ =moyenne ;  $\sigma$ =écart-type



il ressort que, sous l'hypothèse d'un modèle paramétrique, les abandons n'influencent pas de manière significative les estimateurs longitudinaux de la qualité de vie des patients infectés par le VIH.

## Chapitre 7

# Analyse longitudinale de la qualité de vie des patients de la Cohorte ANRS C08

Dans ce chapitre, nous étudions l'évolution sur une période de neuf ans de la dégradation de qualité de vie des patients de la Cohorte ANRS C08 infectés par le VIH et mis entre 1997 et 1999 sous traitement antirétroviral hautement actif comportant un inhibiteur de protéase. Pour cela, le score de qualité de vie considéré est le score agrégé mental du SF-12 equating équipercantil décrit dans le chapitre précédent. L'évolution longitudinale du score agrégé mental du SF-12 equating équipercantil par rapport aux différentes caractéristiques mesurées à l'inclusion pour les 279 patients ayant répondu au questionnaire de qualité de vie initial est montré dans la Table 7.1. La figure 7.1 trace l'évolution longitudinale de la qualité de vie mentale des patients de la Cohorte ANRS C08 ayant complété le questionnaire de qualité de vie au moment de leur inclusion dans l'étude. Pour toutes les covariables le score mental de qualité de vie est croissant durant les 28 premiers mois et est relativement stable après, alors qu'aucune évolution particulière du score de qualité de vie mentale ne transparaît.

Nous commençons par présenter les résultats obtenus lorsque nous appliquons le test global de type log-rank afin d'étudier la dégradation de qualité de vie des patients puis nous effectuons une analyse longitudinale classique sur les patients inclus dans la Cohorte ANRS C08. Enfin, nous procédons à une comparaison et discussion sur les résultats.

### 7.1 Résultats obtenus pour le test global de type log-rank

Les Figures 7.2 et 7.3 montrent, pour les différentes covariables mesurées lors de l'inclusion des patients de la Cohorte ANRS C08, l'évolution du score du processus du test global de type log-rank en fonction du taux de dégradation de qualité de vie. La définition de la dégradation de qualité de vie, à l'instant  $t$ , est donnée respectivement par  $D_t = (Q_0 - Q_t)Q_0^{-1}$  et  $D_t = Q_0 - Q_t$ . De plus, les représentations graphiques des Figures

TAB. 7.1 – Évolution longitudinale du score mental de qualité de vie obtenue par equating équipercentil par rapport aux différentes covariables des patients infectés par le VIH inclus dans la Cohorte ANRS CO8 et ayant complété à l'inclusion le questionnaire de qualité de vie

	M0 n=279 $\mu$ ( $\sigma$ )	M28 n=222 $\mu$ ( $\sigma$ )	M44 n=223 $\mu$ ( $\sigma$ )	M72 n=230 $\mu$ ( $\sigma$ )	M84 n=224 $\mu$ ( $\sigma$ )	M96 n=256 $\mu$ ( $\sigma$ )	M108 n=199 $\mu$ ( $\sigma$ )
Sexe							
homme	43,84 (10,91)	47,55 (10,16)	47,50 (10,04)	46,83 (9,61)	46,83 (9,61)	47,17 (9,28)	46,84 (8,63)
femme	39,13 (12,18)	41,87 (12,22)	41,58 (12,37)	41,72 (12,60)	42,43 (10,58)	42,58 (9,83)	43,44 (10,79)
Déclare être né(e) en UE							
oui	45,98 (9,39)	46,36 (11,09)	46,00 (10,84)	45,21 (11,03)	45,75 (9,84)	46,14 (9,52)	45,98 (9,39)
non	41,56 (12,59)	47,10 (8,73)	48,62 (10,17)	48,45 (10,52)	47,11 (10,62)	46,35 (9,95)	46,97 (8,09)
Enfant							
oui	42,62 (11,74)	46,86 (11,03)	45,24 (12,03)	45,14 (11,41)	45,01 (10,34)	44,99 (10,62)	44,88 (9,54)
non	42,97 (11,15)	46,37 (10,62)	46,74 (10,12)	46,10 (10,73)	46,55 (9,63)	46,83 (8,97)	47,12 (8,84)
Niveau supérieur au BAC							
oui	43,21 (11,47)	47,61 (10,50)	46,93 (10,82)	47,32 (10,62)	47,28 (9,55)	47,57 (9,81)	46,88 (8,72)
non	42,58 (11,40)	44,92 (11,11)	45,98 (10,71)	44,40 (11,31)	44,97 (10,14)	45,03 (9,38)	45,32 (9,58)
Emploi							
oui	43,74 (10,70)	46,66 (10,69)	47,12 (10,60)	47,05 (10,67)	46,74 (10,73)	46,89 (8,99)	47,07 (8,74)
non	41,46 (12,19)	46,05 (11,03)	44,95 (11,01)	43,85 (11,24)	44,33 (11,27)	45,29 (10,22)	44,49 (9,79)
Propriétaire ou locataire							
oui	43,23 (11,22)	46,40 (10,59)	46,39 (10,47)	45,62 (10,94)	45,83 (9,85)	45,96 (9,43)	46,03 (9,18)
non	40,27 (11,89)	47,96 (12,86)	47,75 (13,52)	46,05 (11,58)	46,84 (10,82)	48,13 (10,74)	47,10 (9,52)
Partenaire principal							
oui	43,97 (10,85)	46,65 (11,04)	46,56 (10,51)	45,86 (10,03)	46,01 (9,77)	46,26 (9,68)	46,09 (9,57)
non	40,51 (12,00)	46,00 (10,27)	45,83 (12,38)	45,25 (10,97)	45,81 (10,39)	45,81 (10,39)	46,22 (8,49)
Naïf de tout traitement à l'inclusion							
oui	39,53 (11,39)	45,86 (11,48)	44,85 (10,99)	45,42 (10,39)	45,80 (9,48)	46,59 (9,11)	45,72 (7,90)
non	45,19 (10,70)	46,83 (10,36)	47,34 (10,54)	45,84 (11,43)	45,04 (10,29)	45,90 (9,88)	46,30 (9,69)
CV indétectable							
oui	44,73 (12,47)	46,09 (13,31)	47,03 (14,79)	44,91 (12,16)	46,25 (11,41)	46,73 (9,52)	47,47 (9,41)
non	42,72 (11,22)	46,48 (10,67)	46,26 (10,35)	45,73 (10,92)	45,92 (9,84)	46,13 (9,60)	45,99 (9,19)

n=nombre de patients ;  $\mu$ =moyenne ;  $\sigma$ =écart-type

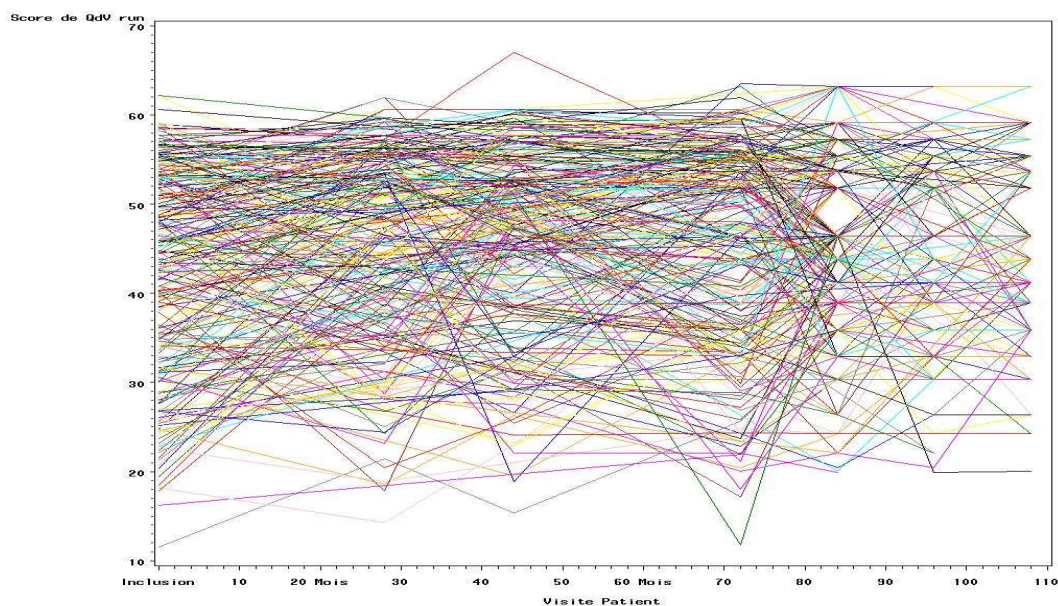


FIG. 7.1 – Évolution pour chaque patient du score mental de qualité de vie obtenue par equating équipercenil des patients de la Cohorte ANRS CO8 ayant répondu au questionnaire de qualité de vie lors de leur entrée dans la Cohorte

7.2 et 7.3 sont réalisées en simulant 30 processus. La courbe observée est indiquée en trait plein alors que les courbes perturbées sont montrées en pointillé bleu.

La Table 7.2 donne pour chaque covariable mesurée lors de l'inclusion dans la Cohorte ANRS CO8 des patients ayant rendu le questionnaire de qualité de vie initial, la valeur des  $\hat{p}$ -values du test global de type log-rank. Ces  $\hat{p}$ -values sont calculées par perturbations de 1000 courbes. Afin d'étudier l'influence des patients n'ayant pas complété le questionnaire de qualité de vie initial, la valeur des  $\hat{p}$ -values du test global de type log-rank des 343 patients inclus dans la Cohorte ANRS CO8 sont montrées dans la Table 7.3 lorsque la dégradation de qualité de vie est définie, à l'instant  $t$ , comme le taux  $D_t = (Q_0 - Q_t)Q_0^{-1}$ . Il faut néanmoins préciser que dans ce cas le questionnaire de qualité de vie nommé  $q_0$  correspond au premier questionnaire de qualité de vie rendu par le patient et n'est donc plus nécessairement le questionnaire de qualité de vie donné à remplir lors de l'inclusion.

Il apparaît que les résultats obtenus, dans la Table 7.2, en considérant l'une ou l'autre des définitions de dégradation de qualité de vie donnent des résultats concordants. Ainsi, les variables déclare être né(e) en Union Européenne, niveau supérieur au BAC, partenaire principal et naïf de tout traitement à l'inclusion semblent être des facteurs prédictifs de la qualité de vie mentale des patients infectés par le VIH. L'étude concernant les variables enfant et propriétaire ou locataire demande à être approfondie. Tandis que les variables sexe, emploi et charge virale indétectable à l'inclusion semblent être des variables non significatives pour la qualité de vie mentale des patients infectés par le VIH.

Les conclusions dégagées par l'analyse des 343 patients incorporés dans la Cohorte

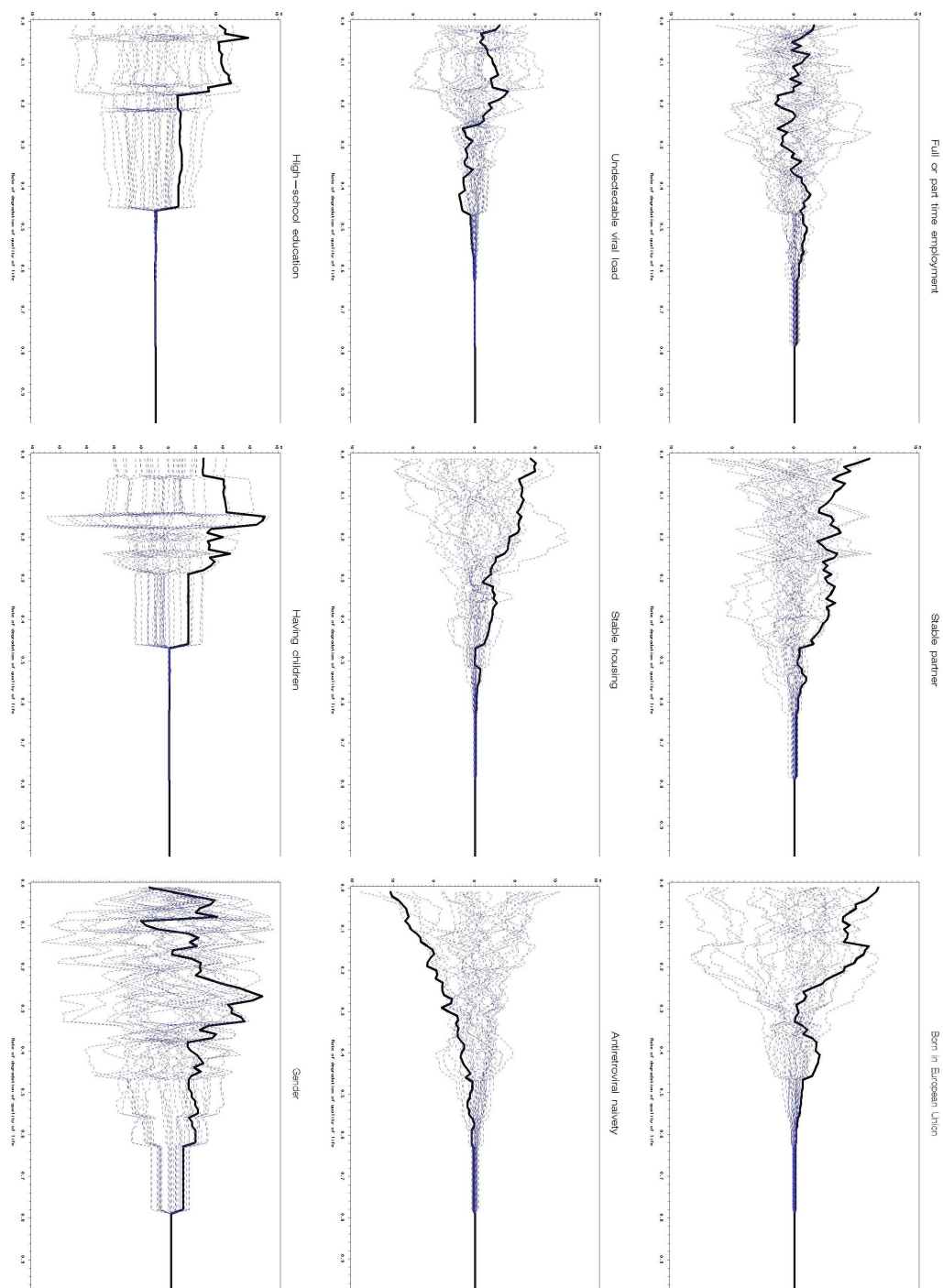


FIG. 7.2 – Score du processus du test global de type log-rank en fonction du taux de dégradation de qualité de vie lorsque  $D_t = (Q_0 - Q_t)Q_0^{-1}$

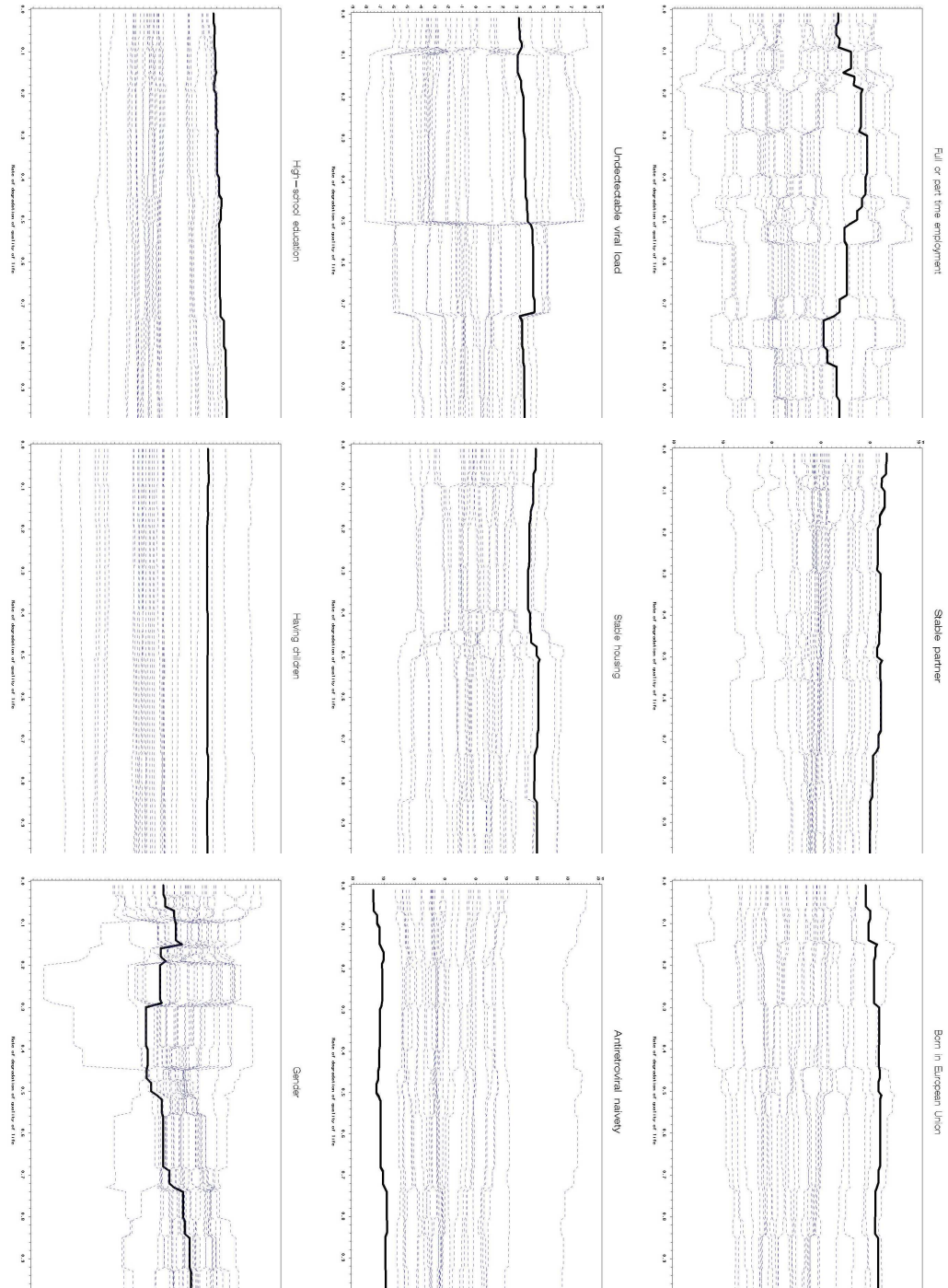


FIG. 7.3 – Score du processus du test global de type log-rank en fonction du taux de dégradation de qualité de vie lorsque  $D_t = Q_0 - Q_t$

TAB. 7.2 –  $\hat{p}$ -value du test global de type log-rank pour le score mental de qualité de vie obtenue par equating équipercenil des 279 patients de la Cohorte ANRS CO8 ayant complété le questionnaire initial de qualité de vie

	$\hat{p}$ -value	
	$D_t = (Q_0 - Q_t)Q_0^{-1}$	$D_t = Q_0 - Q_t$
Sexe	0,33	0,67
Déclare être né(e) en UE	0,08	0,06
Enfant	0,13	0,18
Niveau supérieur au BAC	0,09	0,10
Emploi	0,93	0,45
Propriétaire ou locataire	0,18	0,11
Partenaire principal	0,06	0,07
Naïf de tout traitement à l'inclusion	0,06	0,07
CV indétectable à l'inclusion	0,53	0,46

TAB. 7.3 –  $\hat{p}$ -value du test global de type log-rank pour le score mental de qualité de vie obtenue par equating équipercenil des 343 patients de la Cohorte ANRS CO8 lorsque  $D_t = (Q_0 - Q_t)Q_0^{-1}$

	$\hat{p}$ -value
Sexe	0,05
Déclare être né(e) en UE	0,09
Enfant	0,16
Niveau supérieur au BAC	0,06
Emploi	0,03
Propriétaire ou locataire	0,09
Partenaire principal	0,05
Naïf de tout traitement à l'inclusion	0,08
CV indétectable à l'inclusion	0,40

ANRS CO8 (Table 7.3) sont similaires exceptées pour les variables sexe et emploi aux conclusions obtenues lors de l'analyse des patients ayant complété le questionnaire de qualité de vie initial (Table 7.2). En effet, le test global de type log-rank, lors de l'étude des patients ayant retourné à l'inclusion le questionnaire de qualité de vie donne les variables sexe et emploi non significatives tandis que les résultats obtenus sur l'ensemble de la Cohorte pour ces variables apparaissent significatives. En conséquence, les variables sexe et emploi semblent avoir influées sur le fait d'avoir répondu ou non au questionnaire de qualité de vie initial.

## 7.2 Résultats obtenus par le modèle linéaire mixte

Les covariables apparaissant significatives lors des analyses univariées ont été retenues pour une analyse longitudinale classique reposant sur les modèles linéaires mixtes de l'évolution de la qualité de vie. Cette analyse a été effectuée à l'aide de la procédure proc mixed du logiciel SAS<sup>®</sup>, sans donner de contraintes quant à la forme de la matrice de covariance. Les  $p$ -values des effets aléatoires sont fournies par le test de Student tandis que les  $p$ -values des effets fixes sont obtenues par le test de Fischer. Les résultats des effets fixes et aléatoires sont visibles dans la Table 7.4.

Il ressort de cette analyse longitudinale classique reposant sur les modèles linéaires mixtes que les variables dates de visite, sexe, confort du logement, naïf de tout traitement à l'inclusion et le soutien par le partenaire principal sont des variables prédictives indépendantes de l'évolution du score de qualité de vie mentale des patients infectés par le VIH. Les variables enfant, emploi, propriétaire ou locataire, charge virale indétectable à l'inclusion, stade clinique à l'inclusion, soutien par les membres de la famille et soutien par les ami(e)s et les proches apparaissent comme non significatives. Enfin, nous ne pouvons pas conclure en ce qui concerne la significativité des variables déclare être né(e) en Union Européenne, partenaire principal et niveau supérieur au BAC.

## 7.3 Discussion

Pour toutes les covariables le score mental de qualité de vie connaît une évolution croissante significative durant les 28 premiers mois et reste relativement stable après. Ce résultat est en accord avec Protopopescu *et al.* (2007) qui "*showed a two-phase evolution in health related quality of life scores with a change in the slope one year after initiation therapy*". Si nous n'obtenons pas la même date de changement de pente, la raison en est due à l'absence de ces données à la date M12.

Commençons par signaler que si les réponses apportées d'une part par l'analyse longitudinale classique reposant sur les modèles linéaires mixtes et d'autre part par le test global de type log-rank concernent toutes deux la qualité de vie, le processus sous-jacent analysé diffère. En effet, le test global de type log-rank permet d'apprécier si la dégradation de qualité de vie évolue de façon similaire entre deux groupes de patients tandis que l'ana-



TAB. 7.4 – Facteurs pronostiques de l'évolution de la qualité de vie mentale des 279 patients infectés par le VIH et ayant rempli le questionnaire initial de qualité de vie

	Estimation ( $\sigma$ )	$Pr >  t ^a$	$Pr > F^b$
Intercept	37,72 (4,63)	<0,001	
Dates de visite	0,02 (0,01)	<0,001	<0,001
Sexe	-4,92 (1,24)	<0,001	<0,001
Déclare être né(e) en UE	-2,69 (1,38)	0,05	0,05
Enfant	-0,72 (1,09)	0,51	0,51
Niveau supérieur au BAC	1,62 (1,01)	0,11	0,11
Emploi	1,10 (1,05)	0,30	0,30
Confort du logement	2,46 (0,68)	<0,001	<0,001
Propriétaire ou locataire	0,21 (1,53)	0,89	0,89
Naïf de tout traitement à l'inclusion	-2,86 (1,03)	<0,001	<0,001
CV indétectable à l'inclusion	-1,31 (1,76)	0,46	0,46
Stade clinique à l'inclusion	-0,06 (0,63)	0,92	0,92
Partenaire principal	2,97 (1,50)	0,05	0,05
Soutien par le partenaire principal	2,03 (0,79)	0,01	0,01
Soutien par les membres de la famille	-0,31 (0,40)	0,44	0,44
Soutien par les ami(e)s et les proches	0,33 (0,47)	0,48	0,48

$\sigma$ =écart-type ; <sup>a</sup>test de Student ; <sup>b</sup>test de Fischer

lyse longitudinale classique reposant sur les modèles linéaires mixtes s'intéresse à comparer l'évolution de la qualité de vie entre deux groupes de patients. Ainsi, la notion de dégradation n'apparaît pas dans l'analyse longitudinale classique reposant sur les modèles linéaires mixtes.

La variable déclare être né(e) en Union Européenne est un facteur prédictif de l'évolution de la qualité de vie mentale des patients infectés par le VIH. Ainsi, les patients déclarant être nés hors de l'Union Européenne ont une moins bonne qualité de vie mentale initiale que les autres patients mais dès la date M28 le phénomène inverse est constaté.

Le rôle primordial sur l'évolution de la qualité de vie du fait d'avoir un partenaire principal est mis en évidence. De plus, le soutien apporté par ce partenaire principal est également un facteur significatif de la qualité de vie des patients infectés par le VIH. Par contre, le rôle du soutien par les membres de la famille, des ami(e)s et des proches n'apparaît pas comme significatif.

Parmi les variables viro-immunologiques et médicales considérées dans l'étude, seul le fait d'être naïf de tout traitement à l'inclusion influence l'évolution de la qualité de vie mentale des patients infectés par le VIH.

Le fait d'avoir au moins un enfant ne semble pas avoir d'influence sur la qualité de vie mentale des patients infectés par le VIH.

Si le fait d'être propriétaire ou locataire ne semble pas avoir d'impact sur l'évolution de la qualité de vie mentale, le confort du logement apparaît comme un facteur significatif de la qualité de vie mentale des patients infectés par le VIH. Ceci est en accord avec les résultats décrits par Préau *et al.* (2006).

Les variables sexe et niveau d'étude supérieur au BAC donnent des résultats discordants selon le test utilisé. De plus pour la variable démographique sexe, le panel de sujets sélectionnés, i.e. ont répondu ou non au questionnaire de qualité de vie initial, influe sur les résultats relatifs à cette variable. Ceci est peut-être dû au fait que les femmes sont nettement minoritaires (moins de 20%) dans l'étude.

La variable socio-économique emploi n'apparaît pas comme un facteur prédictif de l'évolution de la qualité de vie des patients infectés par le VIH. Néanmoins, lorsque nous considérons l'ensemble des patients, i.e. qu'ils aient ou non répondu au questionnaire de qualité de vie initial, le test global de type log-rank donne cette variable comme étant un facteur prédictif de l'évolution de la qualité de vie des patients infectés par le VIH. Le travail en France des personnes infectées par le VIH a été étudié dans son aspect socio-économique par Dray-Spira *et al.* (2007), mais sans le versant qualité de vie.



# Conclusions et perspectives

Notre projet de thèse est issu d'une problématique concrète, la nécessité de s'intéresser non plus seulement à l'efficacité et à la tolérance sur le long terme des traitements antirétroviraux hautement actifs mais également à l'évolution de la qualité de vie des patients infectés par le VIH.

Aussi pour répondre à cette problématique, dans ce travail, avons nous construit un test non-paramétrique de type log-rank permettant l'analyse longitudinale jointe de la dégradation de qualité de vie et de la survie pour deux groupes de patients par rapport à une covariable. Ce test ne suppose aucune hypothèse aussi bien pour le processus de dégradation de qualité de vie  $D_t$  que pour le seuil critique de dégradation de qualité de vie  $x$ , contrairement à la littérature sur la dégradation. En effet, nous utilisons seulement le fait que la statistique du log-rank  $U_n(x)$  converge vers un processus gaussien. Pour montrer cette propriété lorsque  $x$  varie nous appliquons de façon indirecte les travaux de Biliias *et al.* (1997). Le test non-paramétrique de type log-rank obtenu repose sur une application des méthodes graphiques. En raison de la non-connaissance de la forme explicite de la matrice de variance, celui-ci permet seulement d'obtenir une valeur approchée de la  $p$ -value, dont l'intervalle de confiance reste à préciser. Pour l'améliorer, il serait intéressant de calculer par bootstrap la matrice de variance empirique afin d'obtenir une  $p$ -value pour le test global de type log-rank.

Une autre piste d'étude serait d'appliquer les travaux de O'Quigley (2003, 2008) afin d'obtenir un test de type Khmaladze pour l'analyse de la dégradation de qualité de vie. Le test de type Khmaladze est obtenu par transformation du processus de score en un pont brownien et utilise les propriétés du supremum d'un pont brownien. Ce test de type Khmaladze permettrait d'obtenir directement la valeur de la  $p$ -value.

Nous avons implémenté en SAS<sup>®</sup> le test global de type log-rank élaboré, néanmoins il serait intéressant de concrétiser le travail par la conception de nouveaux macro-programmes SAS<sup>®</sup> utilisables directement en pratique. Notons que ce test statistique programmé en SAS<sup>®</sup> peut être utilisé sur toute base de données longitudinales de qualité de vie dans le domaine des pathologies chroniques. De plus, une étude par simulations du test non-paramétrique de type log-rank montre que celui-ci fonctionne.

D'autre part, il serait intéressant de généraliser ce test non-paramétrique de type log-rank afin de pouvoir prendre en compte :

1. plusieurs groupes de patients ;
2. plusieurs covariables.

Une application, qu'il reste à peaufiner, de ce test global de type log-rank a été réalisée sur la Cohorte ANRS C08 de patients infectés par le VIH mis sous multithérapies et suivis depuis 1997-1999.

Quatrième partie

**Annexes**



# Annexe A

## Rappels mathématiques

### A.1 Processus auto-régressif

L'état d'un système dépendant du temps et du hasard se représente mathématiquement sous la forme d'un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et d'une fonction

$$(t, \omega) \rightarrow X(t, \omega)$$

de  $\mathbb{R}^+ \times \Omega$  à valeurs dans  $E$  l'état du système. Si  $t$  est fixé, l'état du système est une variable aléatoire  $X(t, \omega)$ . En revanche, si  $\omega$  est fixé nous obtenons une évolution particulière du système, dont les états successifs sont caractérisés par la fonction

$$t \rightarrow X(t, \Omega)$$

nommée *trajectoire*. Afin de pouvoir étudier une trajectoire particulière nous sommes amenés à introduire la notion de processus stochastique.

**Définition A.1.** Un **processus stochastique**  $X$ , défini sur un ensemble d'indices  $T$  à valeurs dans l'espace mesurable  $(E, \mathcal{E})$ , est constitué par

$$X = (\Omega, \mathcal{A}, \mathbb{P}, (X_t)_{t \in T})$$

où  $(\Omega, \mathcal{A}, \mathbb{P})$  est un espace probabilisé appelé espace de base et où  $X_t$  est une famille de variables aléatoires définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $(E, \mathcal{E})$ .

**Définition A.2.** Nous appellerons **processus à accroissements indépendants stationnaires** un processus stochastique  $(X_t)_{t \in T}$  tel que

1.  $\forall s < t, X_t - X_s$  est indépendant de la tribu  $\mathcal{F}_s = \sigma\{X_u, u \leq s\}$ ;
2. la loi de  $X_t - X_s$  ne dépend que de  $t - s$ ;
3.  $(X_t)_{t \in T}$  est à trajectoire càdlàg (continues à droite et pourvues de limites à gauche).

Dans la suite de cette section nous nous restreindrons à la classe particulière des processus stochastiques du second ordre, cadre mathématique classique de l'étude des signaux aléatoires.



**Définition A.3.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé, un processus  $(X_t)_{t \in \mathbb{R}}$  est un **processus du second ordre** si pour tout  $t$ ,  $X_t \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ . Autrement dit si,

$$\forall t, \quad \mathbb{E}(|X_t|^2) < +\infty.$$

Le processus est dit **centré** si pour tout  $t$ ,  $\mathbb{E}[X_t] = 0$ , sa covariance est définie par la fonction

$$K(s, t) = \mathbb{E}[X_s \overline{X_t}].$$

**Définition A.4.** Un processus centré du second ordre  $(X_t)_{t \in \mathbb{R}}$  est nommé **processus stationnaire** si sa covariance  $K$  est telle que

$$K(t + s, s) = \mathbb{E}[X_{t+s} \overline{X_s}]$$

ne dépend que de  $t$ .

Cet emploi très particulier du terme stationnaire ne signifie nullement que les lois marginales, i.e. les lois des vecteurs  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ , sont invariantes par translation du temps, mais seulement que leurs moments d'ordre 2 le sont. Si les lois marginales sont invariantes par translation du temps, nous parlons alors de processus strictement stationnaires.

Avant d'introduire la notion de processus auto-régressif d'ordre  $p$ , nous rappelons le concept de bruit blanc.

**Définition A.5.** Un **bruit blanc**  $W_t$  est un processus gaussien réel centré de covariance  $K_W(s, t) = \delta_0(t - s)$  où  $\delta_0$  est la fonction de Dirac à l'origine.

Remarquons que  $K_W$  ne dépend que de  $t - s$  aussi,  $W_t$  est un processus stationnaire tel que pour  $s \neq t$ ,  $W_t$  et  $W_s$  sont orthogonaux. Nous sommes maintenant en mesure d'introduire la notion de processus auto-régressif d'ordre  $p$ .

**Définition A.6.** Un processus stationnaire  $X_t$  est appelé **processus auto-régressif d'ordre  $p$**  s'il existe un bruit blanc  $W_t$  et  $p + 1$  nombres  $a_j$  tels que

$$W_t = \sum_{j=0}^p a_j x_{t-j}, \quad a_0 \neq 0.$$

Dans le cas particulier d'un processus auto-régressif d'ordre 1  $X_t$ , la définition peut se réécrire

$$X_t = \alpha X_{t-1} + \varepsilon_t$$

où  $\alpha$  est un nombre et où  $\varepsilon_t$  est un bruit blanc.

## A.2 Mouvement brownien

Le mouvement brownien est le plus célèbre et le plus important des processus stochastiques et ceci pour plusieurs raisons.

Pour des raisons historiques d'abord, la découverte qui s'est faite durant la période 1900-1930 fut le premier cas où le calcul des probabilités s'appliquait à la description d'un phénomène physique indépendamment de tout hasard, à savoir la description du mouvement d'une petite particule dans un fluide soumise à des chocs moléculaires dus à l'agitation thermique

Pour des raisons mathématiques ensuite, ce processus est en effet un objet mathématique remarquable sur lequel s'accumule un nombre considérable de propriétés.

**Définition A.7.** *Un processus stochastique  $(X_t)_{t \in \mathbb{R}}$  est appelé **mouvement brownien (standard)** si les trois conditions suivantes sont satisfaites,*

1.  $X(0) = 0$  ;
2.  $(X_t)_{t \in \mathbb{R}}$  est à accroissements indépendants et stationnaires ;
3. pour tout  $t > 0$ , la variable aléatoire  $X(t)$  suit la loi normale  $\mathcal{N}(0, \sqrt{t})$ .

Ce processus qui doit son nom au botaniste Robert Brown est également désigné **processus de Wiener**.

**Proposition A.1.** *Avec une probabilité égale à un,  $X(t)$  est une fonction continue de  $t$  ; autrement dit, presque toute trajectoire du mouvement brownien est continue. De plus, la fonction aléatoire  $X(t)$  est localement continue en probabilité et n'est dérivable en aucun point  $t \in \mathbb{R}$ .*

**Définition A.8.** *Soit  $(X_t)_{t \in \mathbb{R}}$  un mouvement brownien standard. Le processus  $(Y_t)_{0 \leq t \leq 1}$  défini par  $Y(t) := X(t) - tX(1)$  est appelé **pont brownien**.*

**Proposition A.2.** *Les fonctions d'espérance et de variance du pont brownien sont données respectivement par,*

1.  $\forall 0 \leq t \leq 1, \mathbb{E}[Y(t)] = 0$  ;
2.  $\forall 0 \leq s \leq t \leq 1, \text{Cov}(Y(s), Y(t)) = s(1 - t)$ .

## A.3 Martingales et processus ponctuels

### A.3.1 Définitions

**Définition A.9.** *Soit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ ,  $t \in \mathbb{N}$  ou  $\mathbb{R}^+$ . Une **filtration** est une famille  $\mathcal{F}_t$  de tribus,  $t \in \mathbb{N}$  ou  $\mathbb{R}^+$ , telle que*

$$\forall s \leq t, \quad \mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{A}.$$

Un choix naturel est de considérer l'information apportée au cours du temps comme une filtration, c'est-à-dire l'histoire de l'expérience depuis le temps 0 jusqu'à l'instant  $t$  inclus. Remarquons cependant, que ce n'est pas le temps chronologique qui est utilisé. En effet, chaque patient a un temps 0 qui correspond à sa date d'entrée dans l'étude.

**Définition A.10.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé muni d'une filtration  $\mathcal{F} = \{\mathcal{F}_t; t \geq 0\}$ . On définit la tribu sur  $\mathbb{R}^+ \times \Omega$  engendrée par tous les ensembles de la forme

- $[0] \times A; A \in \mathcal{F}_0,$
- et  $]a, b] \times A, 0 \leq a < b \leq \infty, A \in \mathcal{F}_a,$

comme étant la tribu prévisible engendrée par la filtration  $\{\mathcal{F}_t; t \geq 0\}$ .

$X$  est un **processus prévisible** par rapport à une filtration  $\mathcal{F}$ , si en tant qu'application de  $\mathbb{R}^+ \times \Omega$  dans  $\mathbb{R}$ , il est mesurable par rapport à la tribu prévisible engendrée par  $\mathcal{F}$ .

**Définition A.11.** Soient un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $\mathcal{F}_t$  une filtration.  $X = X(t, \omega)$  est un **processus  $\mathcal{F}_t$ -adapté** si pour tout  $t$ ,  $X_t$  est  $\mathcal{F}_t$ -mesurable.

**Définition A.12.** Soient  $X = \{X_t; t \geq 0\}$ , un processus stochastique continu à droite et limité à gauche (càdlàg) et  $F = \{\mathcal{F}_t; t \geq 0\}$  une filtration, tous deux définis sur le même espace probabilisé.  $X$  est une **martingale** par rapport à la filtration  $\mathcal{F}$  si l'on a

1.  $X$  est  $\mathcal{F}$ -adaptée,
2.  $\mathbb{E}(|X_t|)$  est finie, pour  $t \geq 0$ ,
3.  $\mathbb{E}(X_{t+s}/\mathcal{F}_t) = X_t$ , presque sûrement, pour tout  $t$  et  $s$  positifs.

**Définition A.13.** Soit  $x$ , un processus vérifiant les propriétés 1 et 2 de la définition précédente et tel que  $\mathbb{E}(X_{t+s}/\mathcal{F}_t) \geq X_t$  (resp.  $\mathbb{E}(X_{t+s}/\mathcal{F}_t) \leq X_t$ ), presque sûrement, pour  $t$  et  $s$  positifs. On dit alors que  $X$  est une **sous-martingale** (resp. **sur-martingale**).

**Exemple A.1.** Soit  $\mathcal{F}$  la filtration engendrée par  $\mathbb{N}$ , i.e.  $\mathcal{F}_t = \sigma(\{X \leq s\}, 0 \leq s \leq t)$ . L'intégrale d'un processus prévisible  $B$  sur  $(\Omega, \mathcal{A}, P, F)$  par rapport à l'intensité  $I$  est

$$\int_0^t B(s)dI(s) = \int_0^t B(s)Y(s)\lambda(s)ds.$$

L'intégrale de  $B$  par rapport à la martingale  $M = N - I$  est une variable aléatoire. Si  $B$  est un processus prévisible alors  $\int_0^t B(s)dM(s)$  est une martingale par rapport à la filtration  $\mathcal{F}$ . Comme  $N(0) = I(0) = 0$ , la propriété 3 des martingales implique  $\forall t > 0, \forall B$  processus prévisible,

$$\begin{aligned} \mathbb{E}[M(t)] &= \mathbb{E}[\mathbb{E}(M(t)/\mathcal{F}_0)] = \mathbb{E}[M(0)] = 0, \\ \mathbb{E} \left[ \int_0^t B(s)dM(s) \right] &= \mathbb{E} \left[ \mathbb{E} \left\{ \int_0^t B(s)dM(s)/\mathcal{F}_0 \right\} \right] = 0. \end{aligned}$$

**Exemple A.2 (temps d'arrêt).** Une variable aléatoire réelle  $T$  sur un espace  $(\Omega, \mathcal{A}, P, \mathbb{F})$  est un **temps d'arrêt** sur  $\mathbb{R}^+$  si l'ensemble  $\{T \leq t\}$  est dans  $\mathcal{F}_t$  pour  $t \geq 0$ ;  $T$  est un temps d'arrêt prévisible s'il existe une suite croissante  $(T_n)_{n \in \mathbb{N}}$  de temps d'arrêt qui annonce  $T : T_n < T$  pour tout  $n$  et  $T_n \rightarrow T$ .

### A.3.2 Processus de comptage

**Définition A.14.** Un **processus de comptage**  $N$  est un processus càdlàg, adapté, nul en zéro, croissant et ayant des sauts d'amplitude 1.

**Proposition A.3.** Soit  $N(t)$  un processus de comptage. Il existe un processus  $\Lambda(t)$  prévisible, croissant, continu à droite et nul en zéro tel que

$$M(t) = N(t) - \Lambda(t)$$

soit une martingale.

**Définition A.15.**  $\Lambda(t)$  s'appelle le **compensateur** de  $N(t)$ , ou encore son processus d'intensité cumulé.

**Définition A.16.** Soit  $M_p$  l'espace des mesures ponctuelles, c'est-à-dire qui sont sommes finies ou dénombrables sans point d'accumulation de masse de Dirac, sur  $(E, \mathcal{E})$ . On munit  $M_p$  de la tribu  $\mathcal{M}_p$  engendrée par les applications  $m \rightarrow m(A)$  pour  $A \in \mathcal{E}$ . Alors, un **processus ponctuel**  $N$  est une variable aléatoire à valeurs dans  $(M_p, \mathcal{M}_p)$ ,

$$(\Omega, \mathcal{A}, \mathbb{P}) \xrightarrow{N} (M_p, \mathcal{M}_p).$$

**Proposition A.4.** Soient  $N$  un processus ponctuel de dimension 1 et  $\Lambda$  son compensateur. Si  $N$  est absolument continu, alors  $N$  possède une intensité, i.e. il existe un processus prévisible  $\lambda$  tel que

$$\forall t, \Lambda(t) = \int_0^t \lambda(s) ds.$$

### A.3.3 Convergence

**Théorème A.1.** Soient  $r$  mouvements browniens dépendants du temps  $W_1^*, \dots, W_r^*$ . Plus spécifiquement  $(W_1^*, \dots, W_r^*)$  est un processus gaussien  $r$ -varié à incréments indépendants,  $W_l^*(0) = 0$  p.s. et pour tout  $0 \leq s \leq t$ ,  $\mathbb{E}W_l^*(t) = 0$  et  $\mathbb{E}W_l^*(t)W_{l'}^*(t) = C_{ll'}(s)$ , où  $C_{ll'}(s)$  est une fonction continue pour tout  $(l, l') \in \{1, \dots, r\}^2$ .

Pour tout  $n$ ,  $\{N_{i,l}^{(n)}; i = 1, \dots, n\}$  est un processus de comptage multivarié pour  $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, P)$ . Le compensateur  $A_i^{(n)}$  de  $N_i^{(n)}$  est continu. Nous supposons de plus que  $\{H_{i,l}^{(n)}; i = 1, \dots, n, l = 1, \dots, r\}$  vérifie pour tout  $i, l$ ,  $H_{i,l}^{(n)}$  est un processus  $\mathcal{F}_t$ -prévisible localement borné. Considérons le vecteur de martingales locales de carrés intégrables  $(U_1^{(n)}, \dots, U_r^{(n)})$  où pour  $l = 1, \dots, r$  et pour  $t \geq 0$ ,

$$U_l^{(n)}(t) = \sum_{i=1}^n \int_0^t H_{i,l}^{(n)}(u) d\{N_i^{(n)}(u) - A_i^{(n)}(u)\}.$$

Supposons pour tout  $(l, l') \in \{1, \dots, r\}^2$  et pour tout  $t > 0$ ,

$$\langle U_l^{(n)}, U_{l'}^{(n)} \rangle(t) = \sum_{i=1}^n \int_0^t H_{i,l}^{(n)}(u) H_{i,l'}^{(n)}(u) dA_i^{(n)}(u) \xrightarrow[n \rightarrow \infty]{P} C_{ll'}(t), \quad (\text{A.1})$$

et

$$\forall \varepsilon > 0, \langle U_{l,\varepsilon}^{(n)}, U_{l,\varepsilon}^{(n)} \rangle(t) = \sum_{i=1}^n \int_0^t \{H_{i,l}^{(n)}(u)\}^2 \mathbf{1}_{\{|H_{i,l}^{(n)}(u)| > \varepsilon\}} dA_i^{(n)}(u) \xrightarrow[n \rightarrow \infty]{P} 0. \quad (\text{A.2})$$

Alors,

$$(U_1^{(n)}, \dots, U_r^{(n)}) \Rightarrow (W_1^*, \dots, W_r^*) \text{ dans } (D[0, \tau])^r \text{ lorsque } n \rightarrow \infty.$$

### A.3.4 Inégalité de Lenglar

Nous donnons l'énoncé de l'Inégalité de Lenglar (1977).

**Théorème A.2 (Inégalité de Lenglar (1977)).** *Soit  $X$  un processus adapté continu à droite et  $Y$  un processus prévisible non décroissant tel que  $Y(0) = 0$ . Supposons pour tout temps d'arrêt borné  $T$ ,*

$$\mathbb{E}\{|X(T)|\} \leq \mathbb{E}\{Y(T)\}.$$

Alors, pour tout temps d'arrêt  $T$  et tout  $\varepsilon, \eta > 0$ ,

$$P \left\{ \sup_{t \leq T} |X(t)| \geq \varepsilon \right\} \leq \frac{\eta}{\varepsilon} + P\{Y(T) \geq \eta\}.$$

**Corollaire A.1.** *Soient  $N$  un processus de comptage et  $M = N - A$  la martingale locale de carré intégrable correspondante. Supposons que  $H$  est un processus adapté continu à gauche et admettant une limite à droite ou de façon plus générale un processus prévisible localement borné. Alors pour tout temps d'arrêt  $T$  tel que  $P\{T < \infty\} = 1$ , et pour tout  $\varepsilon, \eta > 0$ ,*

$$P \left\{ \sup_{t \leq T} \left( \int_0^t H(u) dM(u) \right)^2 \geq \varepsilon \right\} \leq \frac{\eta}{\varepsilon} + P \left\{ \int_0^T H^2(u) d\langle M, M \rangle(u) \geq \eta \right\}.$$

**Lemme A.1.** *Soit  $N$  un processus de comptage univarié admettant un compensateur continu  $A$ , soit  $M = N - A$  et soit  $H$  un processus prévisible localement borné. Alors, pour tout  $\delta, \rho > 0$  et tout  $t \geq 0$ ,*

1.

$$P\{N(t) \geq \rho\} \leq \frac{\delta}{\rho} + P\{A(t) \geq \delta\},$$

2.

$$P \left\{ \sup_{0 \leq y \leq t} \left| \int_0^y H(u) dM(u) \right| \geq \rho \right\} \leq \frac{\delta}{\rho^2} + P \left\{ \int_0^t H^2(u) dA(u) \geq \delta \right\}.$$

**Remarque** Le Lemme A.1 est un cas spécial de l'Inégalité de Lenglar.

**Lemme A.2.** *Soit  $E$  un sous-ensemble ouvert convexe de  $\mathbb{R}^p$  et soient  $F_1, F_2, \dots$ , une suite de fonctions aléatoires concaves de  $E$  et  $f$  une fonction de  $E$  telle que pour tout  $x \in E$ ,*

$$\lim_{n \rightarrow \infty} F_n(x) = f(x)$$

en probabilité. Alors,

1. la fonction  $f$  est concave,
2. pour tout sous-ensemble compact  $A$  de  $E$ ,

$$\sup_{x \in A} |F_n(x) - f(x)| \xrightarrow{n \rightarrow \infty} 0$$

en probabilité,

3. si  $F_n$  admet un unique maximum en  $X_n$  et si  $f$  est unique en  $x$  alors,  $X_n \xrightarrow{n \rightarrow \infty} x$  en probabilité.



## Annexe B

# Programmes informatiques

### B.1 Test global de type log-rank

#### B.1.1 Première dégradation de la qualité de vie

La macro `tauxdegradation` présentée dans les pages ci-après

1. trace l'ensemble des graphiques perturbés en pointillé ainsi que la courbe réelle en trait plein ;
2. fournit la valeur de la  $\hat{p}$ -value du test global de type log-rank au niveau de la table baptisée `test`.

La syntaxe de cette macro est la suivante :

```
%tauxdegradation(degrascor,nombre_taux,nombre_courbes_perturbées);
```

où

1. `nombre_taux` est la précision voulue concernant le nombre de valeurs équiréparties dans l'intervalle  $[0, 1]$  de taux de dégradation ;
2. `nombre_courbes_perturbées` est le nombre de graphiques perturbés.

Cette macro fait appel à la table `degrascor` calculant pour chaque patient son score de dégradation de qualité de vie, mais pour cela il faut préciser le nom de la base de données à étudier, ici nommée `base`. Cette base de données doit comporter :

1. le nom de la variable de qualité de vie, nommée `q` ;
2. le nom de la variable temps, nommée `t` ;
3. le nom de la variable codant les patients, nommée `n` ;
4. le nom de la variable groupe, codée en 0 et 1, nommée `gp`.

Remarques

- a. Les variables doivent être classées par individu ;
- b. et pour chaque patient les dates ordonnées par ordre croissant.



```

/*****
/*      CALCUL DU SCORE DE DEGRADATION DE QUALITE DE VIE      */
*****/

data Inclusion; /* Pour chaque patient nous créons une variable q0 égale au
score de qualité de vie à l'inclusion */
    set Base;
    if t=0; q0=q;
    drop q;
run;

data Scorline;
    merge Inclusion Base;
    by n;
run;

data Degrascor; /* Calcul du score de dégradation d pour chaque visite et pour
chaque patient */
    label d='Score de dégradation de qualité de vie';
    set Scorline;
    d=(q0-q)/q0;
run;

/*****
/*      CALCUL DE LA DATE DE DEGRADATION POUR UN TAUX X      */
*****/

/* Dans cette partie nous calculons la date de dégradation. L'événement
considéré est la première fois où le score de dégradation (pour un patient)
dépasse le taux de dégradation x considéré. Si un patient n'a pas subi
d'événement alors nous prenons la date de sa dernière visite, nous considérons
également ce patient comme censuré. */

%MACRO Score(tabin,tabout,x,y);
%put &x;

/* Le but de la MACRO SCORE est de calculer pour un taux de dégradation x les
différents poids de la statistique de test aux différentes dates d'événement.
La statistique de test est celle se trouvant dans l'article.
    En entrée : TABIN c'est la table DEGRASCOR appelée au travers de la
macro TAUXDEGRADATION.
    En sortie : TABOUT table donnant pour le taux de dégradation x les
différents poids de la statistique de test aux différentes dates d'événement.
    Paramètres : X est le taux de dégradation critique considéré.
                Y sert à donner un nom différent pour chacune des colonnes
indiquant les valeurs des différents poids de la statistique de test. en
effet, chaque colonne étant calculée pour un taux de dégradation x différent.
*/

data Degradscore; /* Création d'une variable indicatrice indiquant si un
patient est dégradé ou non à la date donnée */
    set &tabin;
    d0=(d>&x); /* 0 si le patient est à risque pour le taux de dégradation x
et 1 sinon */
    drop q q0;
run;

```

```

proc sql;
  create table Date1 as select n, gp, d, t, d0
  from degradscore
  group by n
  having d0=max(d0);
run;

data Copie;
  set Date1;
  t1=t;
run;

proc sql;
  create table Date2 as select n, d0, gp, t1
  from Copie
  group by n
  having t1=max(t1);
run;

proc sql;
  create table Date3 as select n, d0, gp, t
  from Copie
  group by n
  having t=min(t);
run;

data Degradate;
  merge Date2 Date3; by n;
  label d_event='Date de la survenue de l'événement du patient';
  if d0=0 then d_event=t1; else d_event=t; /* d_event est la date
d'événement ; si un patient n'a pas subi de dégradation c'est la date de sa
dernière visite sinon c'est la première fois où son score de dégradation
dépasse le taux x fixé.
  Remarque d0 sert de variable de censure valant 0 si le patient est à
risque et 1 sinon. */
  keep n gp d0 d_event;
run;

/*****
/*          CALCUL DU NOMBRE DE PATIENTS A RISQUE ET DEGRADES          */
/*          CHAQUE DATE POUR UN TAUX X          */
*****/

/* Dans cette partie nous calculons à chaque date le nombre de patients à
risque ainsi que le nombre de sujets dégradés dans chaque groupe. Nous
considérons toutes les dates où un questionnaire de qualité de vie est donné à
remplir aux patients. */

data _nullpat_;
  set degradate;
  call symput('NbPat',_n_);
run;

%put &NbPat; /* NbPat représente le nombre total de patients */

data patA;
  set degradate;

```

```

    if gp=0 then nA=n; else nA=0;
run;

proc sort data=pata out=c_patA nodupkey;
    by nA; /* Le nodupkey sert à ne pas avoir de répétition pour la valeur de
la variable (ici nA) */
run;

data _nullpatA;
    set c_patA;
    call symput('NbPata',_n_);
run;

%put &NbPata; /* NbPata représente le nombre de patients plus un du groupe 0
*/

data patB;
    set degradate;
    if gp=1 then nB=n; else nB=0;
run;

proc sort data=patb out=c_patB nodupkey;
    by nB;
run;

data _nullpatB;
    set c_patB;
    call symput('NbPatb',_n_);
run;

%put &NbPatb; /* NbPatb représente le nombre de patients plus un du groupe 1
*/

/* CALCUL A CHAQUE DATE DU NOMBRE DE PATIENTS A RISQUE ET DEGRADES DANS LE
GROUPE 0 */

data pat_degrad_A;
    set degradate;
    if d0*(1-gp)=1 then degrada=d_event; else degrada='.'; /* Création de la
variable degrada qui garde uniquement les dates d'événements correspondant à
des censures du groupe 0 */
    n_risque_A=0;
run;

proc sql; /* Création de la variable n_degrad_A qui compte le nombre de sujets
dégradés à chaque date où a eu lieu une dégradation dans le groupe 0.
(Remarque pour l'instant tous les sujets censurés ou appartenant au groupe 1
sont comptabilisés dans une même date) */
    create table vpatA as select n, gp, d0, d_event, degrada, n_risque_A,
count(*) as n_degrad_A
    from pat_degrad_A
    group by degrada;
run;

proc sort data=vpata out=pata;
    by d_event;
run;

```

```

proc iml; /* Calcule à chaque date le nombre de sujets à risque du groupe 0 */
  use pata;
  read all var {n gp d0 d_event degrada n_degrad_A n_risque_A} into pata;
  pata[1,7]=&NbPata-1; /* On initialise le nombre de sujets à risque du
groupe 0 au nombre de sujets du groupe 0 */
  do i=1 to &NbPat-1; /* Si le patient appartient au groupe 1 alors
n_risque_A=n_risque_A et si le patient appartient au groupe 0 alors
n_risque_A=n_risque_A-1 */
    if pata[i,2]=1 then pata[i+1,7]=pata[i,7];
    else pata[i+1,7]=pata[i,7]-1;
  end;
  do i=1 to &NbPat; /* Si le patient appartient au groupe 1 ou si le
patient est censuré alors n_degrad_A=0 */
    if (1-pata[i,2])*pata[i,3]=0 then pata[i,6]=0;
  end;
  create n_degrad_A from pata

[colname={'n','gp','d0','d_event','degrada','n_degrad_A','n_risque_A'}];
  append from pata;
run;

/* Pour chaque date d'événement nous ne conservons que la valeur maximale du
nombre de patients dégradés, puis du nombre de patients à risque */
proc sql;
  create table n_degrad_A2 as select n, d0, d_event, degrada, n_degrad_A,
n_risque_A
  from n_degrad_A
  group by d_event
  having n_degrad_A=max(n_degrad_A);
run;

proc sql;
  create table nb_degrad_A as select n, d0, d_event, degrada, n_degrad_A,
n_risque_A
  from n_degrad_A2
  group by d_event
  having n_risque_A=max(n_risque_A);
run;

proc sql;
  create table degrad_A as select n, d0, d_event, degrada, n_degrad_A,
n_risque_A
  from nb_degrad_A
  group by d_event
  having n=max(n);
run;

/* CALCUL A CHAQUE DATE DU NOMBRE DE PATIENTS A RISQUE ET DEGRADEES DANS LE
GROUPE 1 */
/* Voir les remarques et commentaires du groupe 0 */

data pat_degrad_B;
  set degradate;
  if d0*gp=1 then degradb=d_event; else degradb='.';
  n_risque_B=0;
run;

```

```

proc sql;
    create table vpatB as select n, gp, d0, d_event, degradb, n_risque_B,
        count(*) as n_degrad_B
    from pat_degrad_B
    group by degradb;
run;

proc sort data=vpatab out=patb;
    by d_event;
run;

proc iml;
    use patb;
    read all var {n gp d0 d_event degradb n_degrad_B n_risque_B} into patb;
    patb[1,7]=&NbPat-1;
    do i=1 to &NbPat-1;
        if patb[i,2]=0 then patb[i+1,7]=patb[i,7];
        else patb[i+1,7]=patb[i,7]-1;
    end;
    do i=1 to &NbPat;
        if patb[i,2]*patb[i,3]=0 then patb[i,6]=0;
    end;
    create n_degrad_B from patb

[colname={'n', 'gp', 'd0', 'd_event', 'degradb', 'n_degrad_B', 'n_risque_B'}];
    append from patb;
run;

proc sql;
    create table n_degrad_B2 as select n, d0, d_event, degradb, n_degrad_B,
n_risque_B
    from n_degrad_B
    group by d_event
    having n_degrad_B=max(n_degrad_B);
run;

proc sql;
    create table nb_degrad_B as select n, d0, d_event, degradb, n_degrad_B,
n_risque_B
    from n_degrad_B2
    group by d_event
    having n_risque_B=max(n_risque_B);
run;

proc sql;
    create table degrad_B as select n, d0, d_event, degradb, n_degrad_B,
n_risque_B
    from nb_degrad_B
    group by d_event
    having n=max(n);
run;

/* CREATION DE LA TABLE RECAPITULATIVE */

```

```

data risque; /* Nous créons une table où pour chaque date d'événement nous
avons le nombre de sujets à risque ainsi que le nombre de patients dégradés
pour chacun des deux groupes */
    merge degrad_A degrad_B;
    by d_event;
    keep d_event n_risque_A n_risque_B n_degrad_A n_degrad_B;
run;

/*****
/*      CALCUL DU SCORE DU TEST DU LOG-RANK      */
*****/

/* Dans cette partie nous calculons pour chaque taux de dégradation considéré
la valeur du score du test. Nous n'effectuons pas de normalisation. */

/* CREATION DE TOUTES LES TABLES POUR LES DIFFERENTS TAUX DE DEGRADATION x */

data &TABOUT;
    set risque;
    n_risque=n_risque_A+n_risque_B; /* Nombre total de patients à risque */
    n_degrad=n_degrad_A+n_degrad_B; /* Nombre total de patients dégradés */
    &y=n_degrad_B-n_degrad*n_risque_B/n_risque; /* Fournit les différents
poids obtenus aux différentes dates d'événements pour un taux y donné */
    keep d_event &y;
run;

%MEND score;

%MACRO tauxdegradation(tabin,nbseuil,nbgraph);

/* La MACRO TAUXDEGRADATION
   En entrée : TABIN c'est la table DEGRASCOR.
   Paramètres : NBSEUIL le nombre de valeurs de taux de dégradation considéré
pour le graphique.
   Les différentes valeurs sont réparties uniformément.
   NBGRAPH le nombre de graphiques simulés.
   En sortie nous obtenons
       1) les différents graphiques, réel (en rouge) et simulés (en bleu)
       2) la répartition des scores de qualité de vie
       3) le test pour savoir si la différence entre les deux groupes est
significative ou non. */

%do i=1 %to %eval(&nbseuil-1);
    %score(&tabin,pat&i,&i/%eval(&nbseuil),pds&i);
%end;

%score(&tabin,pat&nbseuil,1,pds&nbseuil);

/* REUNION DE TOUTES LES TABLES DES POIDS PRECEDEMMENT CALCULES */

data pdsreunion; /* Réunion des différents poids calculés */
    merge
    %do j=1 %to &nbseuil;
        pat&j
    %end;;
    by d_event;
    drop d_event;

```

```

run;

proc transpose data=pdsreunion out=Tpdsreunion;
run;

data sc_logrank;
  set Tpdsreunion;
  s=sum(of col:); /* Calcul du score sans normalisation */
  tx=0; graph=0;
  keep s tx graph; /* s est la valeur du score */
run;

/*****
/*      SIMULATIONS DU TEST DU LOG-RANK      */
*****/

/* Dans cette section nous perturbons la courbe observée */

%do i=1 %to &nbgraph; /* On perturbe à l'aide d'une N(0,1) les différents
poids obtenus */
data simul&i;
  set pdsreunion;
  nor=rand('Normal'); /* Simulation d'une loi normale centrée réduite */
  %do j=1 %to &nbseuil;
    pds&j=pds&j*nor;
  %end;
  drop y;
run;

proc transpose data=simul&i out=Tsimul&i;
run;

data sc_logrank_simul&i; /* Calcul des scores perturbés */
  set Tsimul&i;
  s=sum(of col:);
  tx=1/&nbseuil; graph=&i;
  keep s tx graph;
run;

proc iml; /* On attribue pour chaque valeur du score perturbé la valeur du
taux de dégradation correspondant */
  use sc_logrank_simul&i;
  read all var {s tx graph} into sc_logrank_simul&i;
  do j=1 to &nbseuil;
    sc_logrank_simul&i[j,2]=j/&nbseuil;
  end;
  create tx_degrad&i from sc_logrank_simul&i
    [colname={'s','tx','graph'}];
  append from sc_logrank_simul&i;
run;

%end;

proc iml; /* On attribue pour chaque valeur du score la valeur du taux de
dégradation correspondant */
  use sc_logrank;
  read all var {s tx graph} into sc_logrank;

```

```

do j=1 to &nbseuil;
  sc_logrank[j,2]=j/&nbseuil;
end;
create tx_degrad0 from sc_logrank
  [colname={'s','tx','graph'}];
append from sc_logrank;
run;

/*****
/*      GRAPHISME SCORE DU LOG-RANK EN FONCTION DU TAUX DE DEGRADATION      */
*****/

/* Dans cette partie nous effectuons la représentation graphique de la valeur
du score en fonction du taux de dégradation ainsi que l'histogramme de la
valeur absolue du maximum des différentes valeurs du score obtenues par
perturbation par une normale centrée réduite. */

data courbes;
label s='score';
label tx='rate of degradation of quality of life';
  %do i=0 %to &nbgraph;
    set tx_degrad&i; output;
  %end;
run;

goptions reset=all; /* Supprime les différentes options pour les graphiques
qui auraient pu être utilisés précédemment */
proc gplot data=courbes; /* Représentation du score du log-rank en fonction du
taux de dégradation */
  symbol1 interpol=join color=black line=1 width=6 repeat=1;
  symbol2 interpol=join color=moderateblue width=1 line=20 repeat=&nbgraph;
  plot s*tx=graph / nolegend;
/* Le nolegend supprime les commentaires de représentations graphiques pour
chacun des graphiques */
run;

/*****
/*      TESTS ET P-VALUE      */
*****/

/* calcul du pourcentage de courbes perturbées dont le supremum de la valeur
du score du log-rank est supérieur à la valeur du supremum du score du log-
rank pour la vraie courbe. */

data absolue;
  set courbes;
  s=abs(s); /* Calcul de la valeur absolue */
  drop tx;
run;

proc sql;
  create table sc_max as select s, graph
  from absolue
  group by graph
  having s=max(s); /* Calcul du maximum de la valeur absolue pour les
différents graphiques */
run;

```



```

data int;
  set sc_max;
  ind=0;
run;

proc iml;
  use int;
  read all var {s graph ind} into int;
  do i=1 to &nbgraph+1;
    int[i,3]=(int[i,1]>int[1,1]); /* Création de la variable indicatrice
valant 0 si la i-ème valeur du score est inférieure à celle observée, i.e.
celle de la première ligne. */
  end;
  create compare from int
    [colname={'s','grah','ind'}];
  append from int;
run;

data indicatrice;
  set compare;
  keep ind;
run;

proc transpose data=indicatrice
  out=Tindicatrice;
run;

data test;
  set Tindicatrice;
  t=sum(of coll-col&nbgraph)/&nbgraph; /* On obtient la valeur du test; i.e.
le pourcentage de courbes perturbées dont le supremum de la valeur du score du
log-rank est supérieur à la valeur du supremum du score du log-rank pour la
vraie courbe.*/
  keep t;
run;

%MEND tauxdegradation;

%tauxdegradation(degrascor,100,1000);

```

### B.1.2 Dégradation définitive de la qualité de vie

Jusqu'à présent nous nous sommes intéressés à la première occurrence de la dégradation de qualité de vie, i.e. l'événement d'intérêt est défini par  $T(x) = \inf\{t : d \geq x\}$ . Mais cette dégradation de qualité de vie peut être passagère et induire un biais dans l'interprétation des résultats. Par conséquent nous pouvons également nous préoccuper de la dégradation définitive de qualité de vie des patients, à savoir la date où pour la première fois la dégradation de qualité de vie dépasse un seuil  $x$  mais se maintient tout le temps supérieure à ce taux  $x$ . Un problème majeur se pose, l'événement d'intérêt dégradation définitive de qualité de vie ne définit pas un temps d'arrêt aussi ne pouvons nous pas obtenir de résultats mathématiques relatifs à cette définition. Par contre il est possible d'ajuster la macro tauxdegradation à la notion de dégradation définitive de qualité de vie.

Si nous nous intéressons à l'événement de la dégradation définitive de qualité de vie il faut remplacer la section « calcul de la date de dégradation pour un taux  $x$  » par le code programme ci-après

```

%MACRO Score(tabin,tabout,x,y);
%put &x;

data degradscore;
  set degrascor;
  d0=(d>x); /* 0 si le patient est à risque */
  drop s s0;
run;

data copie;
  set degradscore;
  t1=t;
run;

proc sql;
  create table date1 as select n, gt, t1, d0
  from copie
  group by n
  having t1=max(t1); /* Si patient censuré i.e. d0=0 alors, d_event=t1
*/
run;

proc sort data=degradscore out=inverse;
  by descending t;
run;

proc sort data=inverse out=count;
  by n;
run;

data _nulle_;
  set count;
  call symput('NbLigne',_n_);
run;

%put &NbLigne; /* Nombre de lignes dans la table degradscore */

data copie2;
  set count;
  s=d0; gp=(gt='B');
  drop gt;
run;

proc iml;
  use copie2;
  read all var {d n t d0 s gp} into copie2;
  do i=1 to &NbLigne-1;
    if copie2[i,2]=copie2[i+1,2] then if copie2[i,5]=0 then
copie2[i+1,5]=0;
  end;
  create somme from copie2
  [colname={'d','n','t','d0','s','gp'}];
  append from copie2;
run;

proc sql;
  create table mins as select n, gp, s, t
  from somme
  group by n
  having s=max(s);
run;

```

```
proc sql;
  create table date2 as select n, gp, t
  from mins
  group by n
  having t=min(t); /* Si patient dégradé définitivement i.e. d0=1 alors,
d_event=t*/
run;

data degradate;
  set date2;
  set datel;
  label d_event='date de la survenue de l événement du patient';
  if d0=0 then d_event=t1; else d_event=t;
  if d_event<0 then d_event='.'; else d_event=d_event;
  keep n gp d0 d_event;
run;
```

## **B.2 Simulations**

Dans cette section, nous présentons les programmes informatiques nous ayant permis de simuler une base de données de patients répartis en deux groupes. L'évolution du score de qualité de vie est modélisée par une droite dont la pente et le score initial varient d'un groupe à l'autre. De plus, les phénomènes de données manquantes et d'abandon sont également modélisés.

La partie du programme qui diffère lorsque l'évolution de la qualité de vie suit un AR(1) est placée à la fin.

```

data Patients; /* Création d'une table contenant les patients, les dates de
visite */
label n='Patients'
      t='Dates of visits'
      gp='Group';
n=0; gp=0;
do i=0 to 278;
      n=n+1; output;
end;
run;

data Normales; /* Création des bruits blancs pour chaque patient */
set Patients;
      subject=rannor(2347);
      L0=subject; eps0=rannor(756789); NS0=L0+eps0;
      L1=subject; eps1=rannor(563789); NS1=L1+eps1;
      L2=subject; eps2=rannor(623789); NS2=L2+eps2;
      L3=subject; eps3=rannor(156789); NS3=L3+eps3;
      L4=subject; eps4=rannor(5789); NS4=L4+eps4;
      L5=subject; eps5=rannor(56789); NS5=L5+eps5;
      L6=subject; eps6=rannor(256789); NS6=L6+eps6;
keep n Ns0 Ns1 Ns2 Ns3 Ns4 Ns5 Ns6;
run;

proc transpose data=Normales out=Normales_t; by n; run;

data Rename; /* Recodage de la variable temps et création des groupes
d'appartenance de chaque patient */
set Normales_t;
      if _name_="NS0" then t=0;
      if _name_="NS1" then t=1;
      if _name_="NS2" then t=3;
      if _name_="NS3" then t=4;
      if _name_="NS4" then t=5;
      if _name_="NS5" then t=6;
      if _name_="NS6" then t=7;
      if n>150 then gp=1; else gp=0;
run;

data qdv; /* Création des droites d'évolution de la qualité de vie pour les
deux groupes */
set Rename;
      label q='Score of quality of life';
      label t='Dates of visits';
      a0=44; a1=39; b0=1; b1=1.1;
      if gp=0 then q=a0+coll+(b0+coll)*t; else q=a1+coll+(b1+coll)*t;
      keep n q t gp;
run;

data Abandon1; /* Simulation des dates d'abandon ainsi que des données
manquantes */
set qdv;
      Ber_abandon=rand('Bernoulli',0.95); /* Simulation des dates d'abandon à
l'aide d'une Bernoulli */
      Ber_manquante=rand('Bernoulli',0.9); /* Simulation des données
manquantes à l'aide d'une Bernoulli */

```

```

        if t<=2 then Ber_abandon=1; /* Pour ne pas avoir d'abandon durant les
trois premières dates */
        if t=0 then Ber_manquante=1; /* Pour ne pas commencer par une donnée
manquante */
run;

proc iml; /* Création des abandons à partir des dates d'abandon simulées */
    use Abandon1;
    read all var {n t gp q Ber_abandon Ber_manquante} into abandon1;
    do i=2 to 1953;
        if Abandon1[i,1]=Abandon1[i-1,1] then
Abandon1[i,5]=Abandon1[i,5]*Abandon1[i-1,5];
    end;
    create Abandon2 from Abandon1
        [colname={'n','t','gp','q','Ber_abandon','Ber_manquante'}];
    append from Abandon1;
run;
quit;

data Bruitblanc; /* Création de la table finale */
    label q='Score of quality of life';
    label t='Dates of visits';
    set Abandon2;
    q=q*Ber_abandon*Ber_manquante;
    if q=0 then q='.';
run;

goptions reset=all;
proc gplot data=Bruitblanc; /* Représentation des courbes de qualité de vie de
chaque patient */
    symbol interpol=join repeat=279;
    plot q*t=n / nolegend;
run;
quit;

/*****

proc iml; /* Création de données de qualité de vie des patients suivant un
AR(1) */
    use Rename;
    read all var {n coll t gp} into Rename;
    do i=1 to 1953;
        if Rename[i,3]=0 then Rename[i,2]=Rename[i,5];
        else Rename[i,2]=Rename[i,4]*Rename[i-1,2]+Rename[i,2];
    end;
    create qdv from Rename [colname={'n','q','t','gp'}];
    append from Rename;
run;
quit;

```

## Annexe C

# Groupe d'étude de la Cohorte ANRS CO8

### **Comité Scientifique :**

**Comité directeur :** *Principaux investigateurs :* C. Leport, F. Raffi, *Methodologie :* G. Chêne, R. Salamon, *Science sociale :* J-P. Moatti, J. Pierret, B. Spire, *Virologie :* F. Brun-Vézinet, H. Fleury, B. Masquelier, *Pharmacologie :* G. Peytavin, R. Garraffo; *Autres membres :* D. Costagliola, P. Dellamonica, C. Katlama, L. Meyer, D. Salmon, A. Sobel;

**Comité de validation des événements :** L. Cuzin, M. Dupon, X. Duval, V. Le Moing, B. Marchou, T. May, P. Morlat, C. Rabaud, A. Waldner-Combernoux;

**Coordination du projet :** F. Collin-Filleul;

**Représentants de l'ANRS :** N. Job-Spira, M. Trumeau;

**Observateurs :** C. Perronne;

**Groupe de recherche clinique :** V. Le Moing, C. Lewden.

### **Monitoring des données et analyses statistiques :**

J. Biemar, S. Boucherit, AD. Bouhnik, C. Brunet-François, M.P. Carrieri, F. Couturier, J.L. Ecobichon, V. Guiyedi, P. Kurkdji, S. Martiren, M. Préau, C. Protopescu, C. Roy, J. Surzyn, A. Taieb, V. Villes, C. Wallet.

### **Promoteur :**

Agence Nationale de Recherches sur le Sida et les hépatites virales (ANRS, Action Coordonnée n°7).

### **Autres subventions :**

Collège des Universitaires de Maladies Infectieuses et Tropicales (CMIT ex APPIT), Sidaction Ensemble contre le Sida, et laboratoires : Abbott, Boehringer-Ingelheim, Bristol-Myers Squibb, Gilead, Glaxo- SmithKline, Roche.

### **Centres cliniques (Coordinateurs) :**

Amiens (Pr JL. Schmit), Angers (Dr JM. Chennebault), Belfort (Dr JP. Faller),



Besançon (Pr JL. Dupond, Dr JM. Estavoyer, Dr Drobachef), Bobigny (Pr O. Bouchaud), Bordeaux (Pr M. Dupon, Pr Longy-Boursier, Pr P. Morlat, Pr JM. Ragnaud), Bourg-en-Bresse (Dr P. Granier), Brest (Pr M. Garré), Caen (Pr R. Verdon), Compiègne (Dr D. Merrien), Corbeil Essonnes (Dr A. Devidas), Créteil (Pr A. Sobel), Dijon (Pr H. Portier), Garches (Pr C. Perronne), Lagny (Dr P. Lagarde), Libourne (Dr J. Ceccaldi), Lyon (Pr D. Peyramond), Meaux (Dr C. Allard), Montpellier (Pr J. Reynes), Nancy (Pr T. May), Nantes (Pr F. Raffi), Nice (Pr JG Fuzibet, Pr P. Dellamonica), Orléans (Dr P. Arsac), Paris (Pr E. Bouvet, Pr F. Bricaire, Pr P. Bergmann, Pr J. Cabane, Dr J. Monsonogo, Pr P.M. Girard, Pr L. Guillevin, Pr S. Herson, Pr C. Leport, Pr MC. Meyohas, Pr J.M. Molina, Pr G. Pialoux, Pr D. Salmon), Poitiers (Pr B. Becq-Giraudon), Reims (Pr R. Jaussaud), Rennes (Pr C. Michelet), Saint-Etienne (Pr F. Lucht), Saint-Mandé (Pr T. Debord), Strasbourg (Pr JM. Lang), Toulon (Dr JP. De Jaureguiberry), Toulouse (Pr B. Marchou), Tours (Pr JM. Besnier).

# Bibliographie

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., De Hues, J. C. J. M., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M., and Takeda, F. (1993). For the European Organization for Research and Treatment of Cancer Study Group on Quality of Life, “The European Organization for Research and Treatment of Cancer QLQ-C30 : A Quality-of-Life Instrument for use in International Clinical Trials in Oncology”. *J. Nat. Cancer Instit.*, **85**, 365–376.
- Andersen, P. K. and Gill, R. D. (1982). Cox’s Regression Model for Counting Processes : A Large Sample Study. *Ann. Stat.*, **10**, 1100–1120.
- Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag, New York.
- APROCO (1997). *Protocole APROCO “Anti PROtéase COhorte”, Cohorte de patients infectés par le VIH ayant débuté un traitement avec Inhibiteurs de Protéase*. Agence Nationale de Recherche sur le Sida ; Association des Professeurs de Pathologie Infectieuse et Tropicale ; Institut National de la Santé et de la Recherche Médicale.
- Asselineau, J., Thiébaud, R., Perez, P., Pinganaud, G., and Chêne, G. (2007). Analysis of Left-Censored Quantitative Outcome : Example of Procalcitonin Level. *Rev Epidemiol. Sante Publique*, **55**, 213–220.
- Awad, L., Zuber, E., and Mesbah, M. (2002). Applying Survival Data Methodology to Analyse Longitudinal Quality of Life. In M. Mesbah, B. Cole, and M. Lee, editors, *Statistical Methods for Quality of Life Studies : Design, Measurements and Analysis*, pages 231–243. Kluwer Academic, Boston.
- Badia, X., Podzamczar, D., Morel, I., Roset, M., Armaiz, J. A., Lonca, M., Casiro, A., Roson, B., Gatell, J. M., and Best Qol Study Group (2004). Health-Related Quality of Life in HIV Patients Switching to Twice-Daily Indinavir/Ribonavir Regimen or Continuing with Three-Times-Daily Indinavir Based Therapy. *Antivir. Ther.*, **9(6)**, 979–985.

- Bas, P. (2007). *Plan 2007-2011 pour l'amélioration de la qualité de vie des personnes atteintes de maladies chroniques*. Ministère de la Santé, de la Jeunesse, des Sports et de la Vie Associative.
- Bilias, Y., Gu, M., and Ying, Z. (1997). Towards a General Asymptotic Theory for Cox Model with Staggered Entry. *Ann. Statist.*, **25(2)**, 662–682.
- Billot, L. and Mesbah, M. (1999). Analyse d'un essai de Qualité de Vie longitudinal en cancérologie avec prise en compte des abandons. In *Actes des Journées Vannes-Bordeaux*.
- Boisson, V. (2005). Analyse longitudinale de la qualité de vie de patients infectés par le VIH, dans un but comparatif. In *Working group on Biostatistics and Latent Variables. Measure and Statistical Analysis of Quality and Duration of Life*. Paris.
- Boisson, V. (2008). Log-rank Test for Longitudinal Analysis of Latent Quality of Life Data. In *International Workshop on Statistical Latent Variable Models in Health Sciences*. Perugia (Italie).
- Boisson, V. and Mesbah, M. (2008). Test du log-rank et statistique du score de vraisemblance partielle pour une durée de survie dépendant d'un taux de dégradation. *C. R. Acad. Sci. Paris, Ser. I*, **346(9-10)**, 567–570.
- Boisson, V., Mesbah, M., and Peccati, G. (2006). Log-rank Test for Longitudinal Degradation of Quality of Life. In *International Conference on Degradation, Damage, Fatigue and Accelerated Life Models in Reliability Testing*. ISTIA. Angers.
- Boisson, V., Mesbah, M., and Ying, Z. (2008a). Log-rank-type Test for Evolution of Health Related Quality of Life. *Submit*.
- Boisson, V., Mesbah, M., Spire, B., Chêne, G., Raffi, F., and Leport, C. (2008b). A Nonparametric Statistical Test for Joint Longitudinal Evolution of Health Related Quality of Life and Survival of HIV-Infected Patients. In *European Seminar - Mathematical Methods for Survival Analysis, Reliability and Quality of Life*. Bordeaux.
- Boisson, V., Mesbah, M., Spire, B., Chêne, G., Raffi, F., and Leport, C. (2008c). A Nonparametric Statistical Test for Treatment Effect on Joint Longitudinal Health Related Quality of Life and Survival. In *Conference on Epidemiology and Biometry 2008 - Recent Statistical Method on Epidemiology*. Paris.
- Bosq, D. and Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle*. Économie et Statistiques Avancées. Economica, Paris.
- Breslow, N. E. (1974). Covariance Analysis of Censored Survival Data. *Biometrics*, **30**, 89–99.

- Brunet-François, C., Taieb, A., Masquelier, B., Le Moing, V., Lewden, C., Dellamonica, P., Cuzin, L., Allavena, C., Spire, B., Chêne, G., Raffi, F., and Groupe d'étude APROCO-COPILOTE (ANRS CO8) (2007). Lopinavir/ritonavir chez le patient infecté par le VIH en situation d'échec virologique prolongé : évolution immunovirologique et tolérance chez 121 patients de la cohorte ANRS CO8 Aproco-Copilote. *Médecine et maladies infectieuses*, page [Epub ahead of print].
- Carrieri, M. P., Cailleton, V., Le Moing, V., Spire, B., Dellamonica, P., Bouvet, E., Raffi, F., Jounot, V., Moatti, J. P., and APROCO Study Group (2001). The Dynamic of Adherence to Highly Active Antiretroviral Therapy : Results From the French National APROCO Cohort. *JAIDS*, **28**, 232–239.
- Carrieri, M. P., Raffi, F., Lewden, C., Sobel, A., Michelet, C., Cailleton, V., Chêne, G., Leport, C. Moatti, J. P., Spire, B., and APROCO Study Group (2003). Impact of Early Versus Late Adherence to Highly Active Antiretroviral Therapy on Immuno-Virological Response : a 3-Year Follow-up Study. *Antivir. Ther.*, **8(6)**, 585–594.
- Carrieri, M. P., Leport, C., Protopopescu, C., Cassuto, J. P., Bouvet, E., Peyramond, D., Raffi, F., Moati, J. P., Chêne, G., and Spire, B. (2006). Factors Associated with Nonadherence to Highly Active Antiretroviral Therapy. a 5-Year Follow-up Analysis with Correction for the Bias Induced by Missing Data in the Treatment Maintenance Phase. *JAIDS*, **41(4)**, 477–485.
- Chavance, M. (1999). Modeling Correlated Data in Epidemiology : Mixed or Marginal Model? *Rev. Epidemiol. Sante Publique*, **47**, 535–544.
- Cohen, C. J., Clumeck, N., Molina, J. M., Thompson, M., Patel, K., Wintfeld, N., and Green, J. (2004). Health-Related Quality of Life with Enfuvirtide (ENF ; T-20) in Combination with an Optimised Background Regimen. *JAIDS*, **37(1)**, 1140–1146.
- COPILOTE (2002). *Protocole COPILOTE "COhorte de Patients mis sous Inhibiteurs de protéase, suivi LOng TERme"*. Agence Nationale de Recherche sur le Sida ; Association des Professeurs de Pathologie Infectieuse et Tropicale ; Institut National de la Santé et de la Recherche Médicale.
- Coplan, P. M., Cook, J. R., Carides, G. W., Heyse, J. F., Wu, A. W., Hammer, S. M., Nguyen, B. Y., Meibohm, A. R., Di Nubile, M. J., and AIDS Clinical Trials Group 320 Study Team (2004). Impact of Indinavir on the Quality of Life in Patients Advanced HIV Infection Treated with Zidovudine and Lamivudine. *Clin Infect Dis.*, **39(3)**, 426–433.

- Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). *J. Roy. Statist. Soc. Ser. B*, **34**, 187–220.
- Cox, D. R. (1975). Partial Likelihood. *Biometrika*, **62(2)**, 269–276.
- Cvetkovic, R. S. and Goa, K. L. (2003). Lopinavir/Ritonavir : A Review of its use in the Management of HIV Infection. *Drugs*, **63(8)**, 769–802.
- De Gruttola, V. and Tu, X. M. (1994). Modelling Progression of CD4-Lymphocyte Count and its Relationship to Survival Time. *Biometrics*, **50**, 1003–1014.
- Dempster, A., Laird, N., and Rubin, D. (1978). Maximum Likelihood From Incomplete Data via EM Algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B*, **39**, 1–38.
- Diggle, P. J. (1988). An Approach to the Analysis of Repeated Measurements. *Biometrics*, **44**, 959–971.
- Diggle, P. J. and Kenward, M. G. (1994). Informative Dropout in Longitudinal Data Analysis (with discussion). *Appl. Statist.*, **43**, 49–93.
- Diggle, P. J., Heagerty, P. J., Liang, K.-Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, volume 25 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, second edition.
- Diggle, P. J., Farewell, D., and Henderson, R. (2007). Analysis of Longitudinal Data Analysis with Drop-out : Objectives, Assumptions and a Proposal (with discussion). *Appl. Statist.*, **56**, 499–550.
- Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.
- Douglas, J. A. (1999). Item Response Models for Longitudinal Quality of Life Data in Clinical Trials. *Stat. Med.*, **18**, 2917–2931.
- Dray-Spira, R., Lert, F., and the VESPA Study Group (2007). Living and Working with HIV in France in 2003 : Results from the ANRS-EN12-VESPA Study. *AIDS*, **21(Suppl 1)**, S29–S36.
- Dupuy, J. F. and Mesbah, M. (2002). Joint Modeling of Event Time and non-Ignorable Missing Longitudinal Data. *Lifetime Data Anal.*, **8**, 99–115.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York.
- Fletcher, R. (2000). *Practical Methods of Optimization*. John Wiley & Sons Inc., New York, second edition.

- Heutte, N. (2001). *Modèles semi-markoviens et données de survie*. Thèse de doctorat, Université René Descartes - Paris V, France.
- Heutte, N. and Huber-Carol, C. (2002). Semi-Markov Models for Quality of Life Data with Censoring. In M. Mesbah, B. Cole, and M. Lee, editors, *Statistical Methods for Quality of Life Studies : Design, Measurements and Analysis*. Kluwer Academic, Boston.
- Hill, C., Com-Nougué, C., Kramar, A., Moreau, T., O'Quigley, J., Senoussi, R., and Chastang, C. (1996). *Analyse statistique des données de survie*. Médecine-Science Flammarion, Paris, second edition.
- Hogan, J. W. and Laird, N. M. (1997a). Mixture Models for the Joint Distribution of Repeated Measurements and Event Times. *Stat. Med.*, **16**, 239–257.
- Hogan, J. W. and Laird, N. M. (1997b). Model-Based Approaches to Analysing Incomplete Longitudinal and Failure Time Data. *Stat. Med.*, **16**, 259–272.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.
- Huber-Carol, C. (2000). Recent Advances in Reliability Theory. In N. Limnios and M. Nikulin, editors, *Methodologie, practice and inference*, chapter 19. Birkhäuser.
- Jin, Z., Liu, M., Albert, S., and Ying, Z. (2006). Analysis of Longitudinal Health-Related Quality of Life Data with Terminal Events. *Lifetime Data Anal.*, **12**, 169–190.
- Jones, R. H. and Boadi-Boateng, F. (1991). Unequally Spaced Longitudinal Data with AR(1) Serial Correlation. *Biometrics*, **47**, 161–175.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Probability and Statistics. John Wiley & Sons Inc., New York, second edition.
- Kaplan, E. and Meier, P. (1958). Nonparametric Estimation From Incomplete Observations. *J. Am. Stat. Assoc.*, **53**, 457–481.
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis : Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, New York, second edition.
- Laird, N. and Ware, J. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, **38**, 963–974.
- Latouche, A., Boisson, V., Chevret, S., and Porcher, R. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Stat. Med.*, **26**(5), 965–974.

- Le Moing, V., Chêne, G., and Carrieri, M. P. (2001). Clinical, Biologic, and Behavioural of Early Immunologic and Virologic Response in HIV-Infected Patients Initiating Protease Inhibitors. *JAIDS*, **27**(4), 372–376.
- Le Moing, V., Chêne, G., Spire, B., Raffi, F., Leport, C., and groupe d'étude APROCO/COPILOTE (2005). Outcome of HIV-Infected Patients After 5 Years of Antiretroviral Therapy Including a Protease Inhibitor : the APROCO/COPILOTE Cohort. *Presse Med.*, **Jun 4 ;34(10 Suppl) :1S**, 31–37.
- Lee, M.-L. T. and Whitmore, G. A. (2004). First Hitting Time Models for Lifetime Data. *Handbook of Statistics*, **23**, 537–543.
- Lee, M.-L. T. and Whitmore, G. A. (2006). Threshold Regression for Survival Analysis : Modeling Event Times by a Stochastic Process Reaching a Boundary. *Statistical Science*, **21**(4), 501–513.
- Lenglart, E. (1977). Relation de domination entre deux processus. *Annal de l'Institut Henri Poincaré*, **13**, 171–179.
- Leplège, A. and Coste, J. (2001). *Mesure de la Santé Perceptuelle et de la Qualité de Vie : méthodes et applications*. Éditions Estem, Paris.
- Leplège, A., Mesbah, M., and Marquis, P. (1995). Analyse préliminaire des propriétés psychométriques de la version française d'un questionnaire international de mesure de qualité de vie : le MOS SF-36. *Rev. Epidemiol. Sante Publique*, **43**, 371–379.
- Leplège, A., Ecosse, E., Verdier, A., and Pergener, T. V. (1998). The French SF-36 Health Survey : Translation, Cultural Adaptation and Preliminary Psychometric Evaluation. *Journal of Clinical Epidemiology*, **51**(11), 1013–1023.
- Lewden, C., Raffi, F., Cuzin, L., Cailleton, V., Vilde, J. L., Chêne, G., Allavena, C., Salomon, R., Leport, C., and APROCO Study Group (2002). Factors Associated with Mortality in Human Immunodeficiency Virus Type 1-Infected Adults Initiating Protease Inhibitor-Containing Therapy : Role of Education Level and Early Transaminase Level Elevation (aproco-anrs ep11 study). *The Journal of Infectious Diseases*, **186**(5), 710–714.
- Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika*, **80**(3), 557–572.
- Lin, D. Y., Wei, L. J., and Ying, Z. (2002). Model-Checking Techniques Based on Cumulative Residuals. *Biometrics*, **58**, 1–12.
- Lindsey, J. K. (1993). *Models for Repeated Measurements*, volume 10 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York. Oxford Science Publications.

- Lindstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *JASA*, **83**, 1014–1022.
- Littell, R. C., Milliken, G., Stroup, W. W., and Wolfinger, R. D. (1996). SAS System for Mixed Models. In *SAS Institute*, North Carolina.
- Little, R. J. A. (1995). Modelling the Dropout Mechanism in Repeated-Measures Studies. *J. Am. Stat. Assoc.*, **90**, 1112–1121.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley, New York.
- Martinussen, T., Aalen, O. O., and Scheike, T. (2008). The Mizon-Richard Encompassing Test for the Cox and Aalen Additive Hazards Models. *Biometrics*, **64**, 164–171.
- Mesbah, M., Dupuy, J.-F., Heutte, N., and Awad, L. (2004). Joint Analysis of Longitudinal Quality of Life and Survival Processes. In *Advances in survival analysis*, volume 23 of *Handbook of Statist.*, pages 689–728. Elsevier, Amsterdam.
- Mesbah, M., Boisson, V., and Spire, B. (2008). Longitudinal Analysis of Partially Equated Health Related Quality of Life Scales. In *The First International Symposium on Biopharmaceutical Statistics*. Shanghai (Chine).
- Molenberghs, G., Kenward, M. G., and Lesaffre, E. (1997). The Analysis of Longitudinal Ordinal Data with Nonrandom Dropout. *Biometrika*, **84**, 33–44.
- O’Connell, K., Skevington, S., and Saxena, S. (2003). Preliminary Development of the World Health Organisation’s Quality of Life HIV Instrument (WHOQOL-HIV) : Analysis of the Pilot Version. *Social Science and Medicine*, **57(7)**, 1259–1275.
- O’Quigley, J. (2003). Khmaladze-Type Graphical Evaluation of the Proportional Hazards Assumption. *Biometrika*, **90**, 577–584.
- O’Quigley, J. (2008). *Proportional Hazards Regression*. Springer, New York.
- Pan, Z. and Lin, D. Y. (2005). Goodness-of-Fit Methods for Generalized Linear Mixed Models. *Biometrics*, **61**, 1000–1009.
- Paterson, D. L., Swindells, S., Mohr, J., Brester, M., Vergis, E. N., Squier, C., Wagener, M. M., and Singh, N. (2000). Adherence to Protease Inhibitor Therapy and Outcomes in Patients with HIV Infection. *Ann. Intern. Med.*, **133**, 21–30.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.



- Préau, M. (2004). *Évaluation psychosociale de la qualité de vie des patients infectés par le VIH*. Thèse de doctorat, Université Aix-Marseille I, France.
- Préau, M. (2006). Qualité de vie et syndrome lipodystrophique chez les patients infectés par le VIH. *L'encéphale*, **32(1)**, 713–719.
- Préau, M., Leport, C., Salmon-Ceron, D., Carrieri, M. P., Portier, H., Chêne, G., Spire, B., Choutet, P. Raffi, M., Morin, M., and the APROCO Study Group (2004). Health-Related Quality of Life and Patient-Provider Relationships in HIV-Infected Patients During the First Three Years After Starting PI-Containing Antiretroviral Treatment. *AIDS Care*, **16(5)**, 651–663.
- Préau, M., Protopopescu, C., Spire, B., Dellamonica, P., Poizot-Martin, I., Villes, V., Carrieri, M. P., and le groupe d'étude MANIF-2000 (2006). Qualité de vie des patients co-infectés par le VIH et VHC. *Rev Epidemiol Sante Publique*, **54**, 1S33–1S43.
- Protopopescu, C., Marcellin, F., Spire, B. Préau, M., Verdon, R., Peyramond, D., Raffi, F., Chêne, G., Leport, C., and Carrieri, M. P. (2007). Health-Related Quality of Life in HIV-1-Infected Patients on HAART : A Five-Years Longitudinal Analysis Accounting for Dropout in the APROCO-COPILOTE Cohort ANRS CO-8. *Quality of Life Research*, **16**, 557–591.
- Ribaudo, H. J., Thompson, S. G., and Allen-Mersh, T. G. (2000). A Joint Analysis of Quality of Life and Survival Using a Random Effect Selection Model. *Stat. Med.*, **19**, 3237–3250.
- Sánchez, B. N., Houseman, E. A., and Ryan, L. M. (2008). Residual-Based Diagnostics for Structural Equation Models. *Biometrics*, (DOI : [10.1111/j.1541-0420.2008.01022.x](https://doi.org/10.1111/j.1541-0420.2008.01022.x)).
- Schuluchter, M. D. (1992). Methods for the Analysis of Informatively Censored Longitudinal Data. *Stat. Med.*, **11**, 1861–1870.
- Seshadri, V. (1993). *The Inverse Gaussian Distribution : A Case Study in Exponential Families*. Oxford, New York.
- Setbon, M. (2000). La normalisation paradoxale du SIDA. *Revue Française de Sociologie*, **41(1)**, 61–78.
- Shahrier, J., Delate, T., Hays, R. D., and Coons, S. J. (2003). Commentary on Using the SF-36 or MOS-HIV in Studies of Persons with HIV Disease. *Health Quality of Life Outcomes*, **July**, 1–25.
- Shorack, G. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley, New York.

- Singer, J. D. (1998). Using SAS Proc Mixed to Fit Multilevel Models, Hierarchical Models, Individual Growth Models. *Journal and Educational and Behavioral Statistics*, **24**, 323–355.
- Singpurwalla, N. D. (2006). On Competing Risk and Degradation Processes. *IMS Lecture Notes - Monograph Series - 2nd Lehmann Symposium - Optimality*, **49**, 229–240.
- Smith, A. F. M. (1973). A General Bayesian Linear Model. *J. Roy. Statist. Soc. Ser. B*, **35**, 67–75.
- Spire, B., Duran, S., Souville, M., Leport, C., Raffi, F., and Moatti, J. P. (2002). Adherence to Highly Active Antiretroviral Therapies (HAART) in HIV-Infected Patients : From a Predictive to a Dynamic Approach. *Social Science and Medicine*, **54**, 1481–1496.
- Su, J. Q. and Wei, L. J. (1991). A Lack-of-Fit Test for the Mean Function in A Generalized Linear Model. *JASA*, **86(414)**, 420–426.
- Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data*. Statistics for Biology and Health. Springer, New York.
- Thiébaud, R. and Jacqmin-Gadda, H. (2004). Mixed Models for Longitudinal Left-Censored Repeated Measures. *Computer Methods and Programs in Biomedicine*, **74**, 255–260.
- Thiébaud, R., Jacqmin-Gadda, H., Chêne, G., Leport, C., and Commenges, D. (2002). Bivariate Linear Mixed Models Using SAS Proc Mixed. *Computer Methods and Programs in Biomedicine*, **69**, 249–256.
- Valerio, L. (2004). *Modélisation des observations incomplètes dans une cohorte de patients infectés par le VIH*. Mémoire de D.E.A. de santé publique option biostatistique, Université Paris-Sud, René Descartes et Panthéon-Sorbonne, France.
- Van Leth, F., Conway, B., Laplume, H., Martin, D., Fisher, M., Jelaska, A., Wit, F. W., and Lange, J. M. (2004). Quality of Life in Patients Treated with First-Time Antiretroviral Therapy Containing Nevirapine and/or Efavirenz. *Antivir Ther.*, **9(5)**, 721–728.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. Springer-Verlag, New York.
- Wang, C., Douglas, J., and Anderson, S. (2002). Item Response Models for Joint Analysis of Quality of Life and Survival. *Stat. Med.*, **21**, 129–142.

- Whitmore, G. A. (1995). Estimating Degradation by a Wiener Diffusion Process Subject to Measurement Error. *Lifetime Data Anal.*, **1(3)**, 307–319.
- Whitmore, G. A., Crowder, M. J., and Lawless, J. F. (1998). Failure Inference from a Marker Process Based on a Bivariate Wiener Model. *Lifetime Data Anal.*, **4**, 229–251.
- WHOQOL-HIV (2004). WHOQOL-HIV for Quality of Life Assessment Among People Living with HIV and AIDS : Results From the Field Test. *AIDS Care*, **16(7)**, 882–889.
- World Health Organization (1994). *Report of the Working Group on Concepts and Principles of Health Promotion*, Copenhagen : WHO edition.
- Wu, A. W., Dave, N. B., Diener-West, M., Sorensen, S., Huang, I. C., and Revicki, D. A. (2004). Measuring Validity of Self-Reported Symptoms Among People with HIV. *AIDS Care*, **16(7)**, 876–881.
- Wu, M. C. and Carroll, R. J. (1988). Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modelling the Censoring Process. *Biometrics*, **44**, 175–188.
- Yeni, P. (2006). *Prise en charge médicale des personnes infectées par le VIH - Rapport 2006 des recommandations du Groupe d'experts sous la direction du Professeur Yeni*. Médecine-Sciences Flammarion, Paris.