Université Victor Segalen Bordeaux 2

Année 2009 Thèse n° 1658

THESE

pour le

DOCTORAT DE L'UNIVERSITE BORDEAUX 2

Mention : Sciences, Technologies, Santé

Option : Epidémiologie et Santé Publique

Présentée et soutenue publiquement

Le 8 décembre 2009

Par

Etienne DANTAN

Né le 11 novembre 1982 à Niort

Modèles conjoints pour données longitudinales et données de survie incomplètes appliqués à l'étude du vieillissement cognitif

Membres du jury

Monsieur Daniel Commenges, Directeur de Recherche, Bordeaux	Président
Monsieur Emmanuel Lesaffre, Professeur, Louvain	Rapporteur
Monsieur Christian Lavergne, Professeur, Montpellier	Rapporteur
Madame Véronique Sébille, Maître de Conférences, Nantes	Examinatrice
Monsieur Jean-François Dartigues, Professeur, Bordeaux	Examinateur
Madame Hélène Jacqmin-Gadda, Directrice de Recherche, Bordeaux	Directrice de thèse

A Chloé

Remerciements

A Monsieur Daniel Commenges

Vous me faites un très grand honneur en présidant ce jury. Je souhaite ici vous exprimer tout le plaisir que j'ai eu à travailler au sein de l'équipe de recherche que vous dirigez. J'y ai trouvé une ambiance de travail extrêmement stimulante avec un esprit d'émulation, d'entraide et d'intérêt mutuel, ceci dans une atmosphère relationnelle vraiment agréable.

A Monsieur Emmanuel Lesaffre

You honor me to judge this work. I have much admiration for the quality of your work on longitudinal data analysis and missing data analysis. I am very grateful for your presence among my judges.

A Monsieur Christian Lavergne

Vous me faîtes un grand honneur d'être rapporteur de mon travail de recherche. Votre grande connaissance des modèles de survie à structure cachée et votre regard sur les modèles de mélange m'apporteront beaucoup. Veuillez trouver ici l'expression de ma reconnaissance.

A Madame Véronique Sébille

Je vous suis très reconnaissant d'avoir bien voulu participer à mon jury de thèse. Votre présence est un honneur, soyez en remerciée.

A Monsieur Jean-François Dartigues

Au cours de ces trois années de recherche, j'ai pu mesurer à quel point vos travaux et votre expérience dans la recherche sur le vieillissement cognitif sont une référence. Soyez vivement remercié de l'honneur que vous me faîtes en participant à l'évaluation de ce travail de recherche. Je vous en suis profondément reconnaissant.

A Madame Hélène Jacqmin-Gadda

Je suis vraiment très heureux d'avoir eu cette grande chance de travailler avec toi. Tes compétences et ton expérience dans les domaines de la statistique et de l'épidémiologie m'ont énormément appris. Au cours de ces trois années, tu as toujours su me guider avec tes conseils autant théoriques que techniques et m'encourager avec beaucoup de patience, ce travail n'aurait jamais pu voir le jour sans ton investissement permanent. Ta gentillesse, ta disponibilité et ton enthousiasme m'ont infiniment aidé. Je ne pourrai jamais assez te remercier pour tout ce que tu m'as apporté.

A Cécile. Tes travaux sont à l'origine d'une grande partie de ce travail. Merci d'avoir partagé avec moi ton expérience et tes connaissances.

A Pierre, toujours disponible et accueillant pour des échanges qui m'ont beaucoup apporté.

A l'ensemble des personnes de l'équipe biostat. Merci pour votre accueil et votre sympathie tout au long de ces années. Merci à Guillaume toujours là pour nous dépatouiller les PC.

A l'équipe Paquid. Je remercie chaleureusement Luc pour sa disponibilité et ses réflexions sur mes questions concernant le vieillissement cognitif.

A l'ensemble des personnes du bureau 45 : Linda, Cécile, Julia, Andrey, tous dans le même bateau.

A tous ceux que j'ai rencontré pendant ces années à l'ISPED.

A mes amis qui m'ont toujours soutenu (Nico, Thib, Céline, Marie, Brice, Laure, Sam, Vinc',...) et qui ont toujours été présents alors même que j'étais à bloc dans mon travail. Spéciale dédicace à Virginie : à toi de jouer...

A tous ceux que je n'ai pas mis dans la parenthèse précédente et qui devraient s'y trouver.

A mes deux frangins et à toute ma famille.

Enfin MERCI du fond du coeur à mes parents que j'ai impliqué directement dans cette thèse (qu'est-ce qu'il y avait comme photes d'aurtografes). Merci pour tout le temps que vous avez consacré à m'aider et même à essayer de comprendre le pourquoi du comment alors que c'était pas gagné, merci pour votre soutien inconditionnel tout au long de ces années d'étude. Je vous dédie cette thèse.

Le plus fort de mes remerciements est pour Chloé qui a tout supporté. Ta présence est indispensable. Merci d'être là tous les jours.

Résumé : Dans l'étude du vieillissement cérébral, le suivi des personnes âgées est soumis à une forte sélection avec un risque de décès associé à de faibles performances cognitives. La modélisation de l'histoire naturelle du vieillissement cognitif est complexe du fait de données longitudinales et données de survie incomplètes. Par ailleurs, un déclin accru des performances cognitives est souvent observé avant le diagnostic de démence sénile, mais le début de cette accélération n'est pas facile à identifier. Les profils d'évolution peuvent être variés et associés à des risques différents de survenue d'un événement ; cette hétérogénéité des déclins cognitifs de la population des personnes âgées doit être prise en compte. Ce travail a pour objectif d'étudier des modèles conjoints pour données longitudinales et données de survie incomplètes afin de décrire l'évolution cognitive chez les personnes âgées. L'utilisation d'approches à variables latentes a permis de tenir compte de ces phénomènes sous-jacents au vieillissement cognitif que sont l'hétérogénéité et l'accélération du déclin. Au cours d'un premier travail, nous comparons deux approches pour tenir compte des données manquantes dans l'étude d'un processus longitudinal. Dans un second travail, nous proposons un modèle conjoint à état latent pour modéliser simultanément l'évolution cognitive et son accélération pré-démentielle, le risque de démence et le risque de décès.

Mots clés : modèles mixtes, données manquantes, modèles conjoints, modèle multi-états, état latent, vieillissement cognitif, démence, décès.

Abstract : In cognitive ageing study, older people are highly selected by a risk of death associated with poor cognitive performances. Modeling the natural history of cognitive decline is difficult in presence of incomplete longitudinal and survival data. Moreover, the non observed cognitive decline acceleration begining before the dementia diagnosis is difficult to evaluate. Cognitive decline is higly heterogeneous, e.g. there are various patterns associated with different risks of survival event. The objective is to study joint models for incomplete longitudinal and survival data to describe the cognitive evolution in older people. Latent variable approaches were used to take into account the non-observed mecanismes, e.g. heterogeneity and decline acceleration. First, we compared two approaches to consider missing data in longitudinal data analysis. Second, we propose a joint model with a latent state to model cognitive evolution and its pre-dementia acceleration, dementia risk and death risk.

Key words : mixed model, missing data, joint model, multi-state model, latent state, cognitive ageing, dementia, death.

Laboratoire :

Centre de Recherche Epidémiologie et Biostatistique, Inserm U897

Université Victor Segalen Bordeaux 2

146, rue Léo Saignat 33076 BORDEAUX

Table des matières

Chapitre 1

Introduction

Avec l'allongement de l'espérance de vie et les progrès scientifiques qui ont permis de transformer les maladies autrefois mortelles en maladies de longue durée, l'étude des maladies chroniques est devenue un véritable enjeu de santé publique de toutes les sociétés modernes. Ce sont des maladies d'évolution lente, au long cours, souvent associées à une invalidité et à la menace de complications graves. De l'asthme infantile au diabète du jeune adulte, une maladie chronique, quelle qu'elle soit, détériore la qualité de vie. Elle peut entraîner des handicaps lourds dont la gestion (traitements, soins complémentaires, régimes, etc.) a un retentissement considérable sur la vie quotidienne du patient et de son entourage. L'éventail des maladies chroniques recouvre de nombreuses pathologies :

- les maladies neurodégénératives comme la maladie d'Alzheimer ou la maladie de Parkinson,
- les troubles mentaux de longue durée tels que la schizophrénie ou la psychose maniaco-dépressive,
- les maladies transmissibles persistantes comme l'infection par le Virus de l'Immunodéficience Humaine (VIH) ou l'hépatite C,
- les maladies rares comme la mucoviscidose, la drépanocytose et les myopathies,
- ainsi que le cancer, le diabète, les maladies cardio-vasculaires, l'asthme, les bronchites chroniques, l'insuffisance rénale et bien d'autres.

Du fait du vieillissement de la population et de l'augmentation de la probabilité d'exposition aux facteurs de risques de maladies chroniques, nous assistons à une augmentation très importante de l'incidence (nombre de nouveaux cas d'une pathologie survenus dans

un intervalle de temps, le plus souvent exprimé par an) et de la prévalence (nombre de cas existants à un moment donné) des maladies chroniques.

Le travail de recherche présenté dans ce manuscrit a pour objet de répondre à certains des problèmes statistiques posés par l'étude du développement de maladies chroniques. Ce travail a été réalisé dans le contexte du vieillissement cognitif normal et pathologique des personnes âgées. Il est centré sur l'étude de la démence sénile et prolonge le développement de méthodes statistiques adaptées pour l'analyse de données longitudinales et de données de survie.

1.1 Epidémiologie du vieillissement cognitif

Le nombre de personnes âgées ne cessant de croître, les sociétés modernes doivent répondre au problème de la perte d'autonomie physique, psychique et sociale, associée au déclin cognitif de ces personnes. Le vieillissement cognitif des personnes âgées peut être normal avec des troubles restant légers, ou bien pathologique avec des troubles de plus en plus marqués au cours du temps. Parmi les pathologies liées au vieillissement cérébral, les maladies neurodégénératives constituent un groupe hétérogène et complexe de pathologies chroniques d'évolution progressive. Le processus en cause consiste généralement en une détérioration du fonctionnement des cellules nerveuses aboutissant à leur mort cellulaire ; ce processus est appelé perte neuronale. Les symptômes cliniques sont variés allant d'une atteinte prédominante des fonctions psychiques et intellectuelles, aboutissant à la démence comme dans la maladie d'Alzheimer, à des anomalies motrices prédominantes comme dans la sclérose latérale amyotrophique ou la maladie de Parkinson, ou encore à l'association des deux comme dans la chorée de Huntington ou la maladie de Creutzfeldt-Jacob.

1.1.1 Histoire naturelle du vieillissement cognitif

Dans le cas du vieillissement normal, l'origine des troubles cognitifs légers n'est pas clairement établie. Il n'est pas évident de distinguer si ce déclin cognitif léger est irrémédiable ou s'il correspond à un manque de stimulation cognitive lié au contexte psycho-social de la vieillesse. Dans le cas du déclin cognitif pathologique, il s'agit d'un processus d'évolution au long cours dont le début se situe bien souvent longtemps avant le diagnostic de

la maladie. Les troubles cognitifs liés à l'âge se traduisent par une perte progressive de la mémoire, du langage, de l'attention, de la capacité d'abstraction ou de l'orientation dans le temps et l'espace. Les sujets passent successivement par différents stades de la maladie pouvant aller de l'état normal à l'état le plus grave (le décès), en passant éventuellement par l'état démentiel diagnostiqué. La démence s'inscrit dans un processus continu de détérioration cognitive. L'évolution de la plupart des démences est progressive avec une aggravation des troubles cognitifs, une apparition de nouveaux troubles notamment au niveau du comportement et de la personnalité, une aggravation de la dépendance avec au stade sévère une altération des activités essentielles de la vie quotidienne (toilette, habillage, alimentation, locomotion).

Les différents types de démence

Le terme de démence désigne les maladies neurobiologiques caractérisées par une altération des capacités cognitives. Selon l'Organisation Mondiale de la Santé (OMS), un syndrome démentiel est défini par une altération progressive de la mémoire et de l'idéation, suffisamment marquée pour handicaper les activités de la vie de tous les jours, apparue depuis au moins 6 mois et associée à un trouble d'au moins une des fonctions suivantes : le langage, l'attention, les fonctions visuo-spatiales, les fonctions exécutives (c'est-à-dire les fonctions d'anticipation, d'initiation et de planification des tâches), la conscience de soi et de son environnement, les praxies (capacités gestuelles) et les gnosies (capacités à reconnaître les êtres vivants et les objets). Plusieurs types de démence sénile sont définis en fonction du processus causal. On distingue les démences non-dégénératives (les démences vasculaires par exemple) et les démences dégénératives. Parmi celles-ci, on retrouve la démence à corps de Lewy, la démence associée à la maladie de Parkinson, la démence fronto-temporale et la plus fréquente d'entre elles la maladie d'Alzheimer qui représente plus de deux tiers des cas de démence. Le processus neurodégénératif responsable de la maladie d'Alzheimer correspond à la formation, entre les neurones, de plaques amyloïdes, et à l'intérieur des neurones d'agrégats de protéines *tau* formant les dégénérescences neurofibrillaires. L'atrophie corticale résultante touche d'abord le lobe temporal interne (et notamment l'hippocampe) puis les cortex associatifs frontaux et temporo-pariétaux à un stade plus avancé. Enfin, l'association entre une maladie d'Alzheimer et un accident vas-

culaire définit les démences dites mixtes. Le plus souvent, les diagnostics de démence de personnes âgées sont établis à l'aide du manuel diagnostique et statistique des troubles mentaux (DSM IV-R) qui reprend le concept proposé par l'OMS.

Les fonctions cognitives atteintes

Il est maintenant bien admis que le déficit cognitif est le symptôme prédictif le plus sensible d'une démence sénile. Concernant la nature des processus cognitifs qui se détériorent, la mémoire épisodique est l'une des dimensions de la cognition qui décline en premier (Small et al., 2000; Bäckman et al., 2001). Certaines études ont montré qu'un déficit cognitif pouvait également être détecté pour d'autres aspects de la cognition bien avant que les critères diagnostiques de démence ne soient réunis (Dartigues et al., 1997; Schmand et al., 2000; Tierney et al., 2005; Storandt et al., 2006) : par exemple, les performances de mémoire visuelle (Howieson et al., 1997), celles de mémoire sémantique et lexicale (Linn et al., 1995) ou encore la fluence verbale (Masur et al., 1994). D'autres travaux ont également montré que le déficit cognitif pré-clinique pouvait être en partie lié à une réduction des activités de la vie quotidienne (Barberger-Gateau et al., 1999). Cependant la plupart de ces études ne considèrent qu'une seule mesure avant le diagnostic de démence. Bien qu'elles aient largement contribué à l'amélioration des connaissances sur l'évolution cognitive précédant le diagnostic de démence, il n'est pas possible d'établir si les faibles performances d'un futur dément sont le reflet d'un état pré-existant de faibles performances cognitives ou bien s'il s'agit d'un processus dynamique et évolutif associé au déclin cognitif. Des études cas-témoins avec peu de mesures répétées confirment la présence d'un déficit cognitif sans pour autant valider la notion de déclin au cours du temps (Bäckman et al., 2001; Chen et al., 2001). Ces travaux ne permettent pas non plus de déterminer le début du déclin. Avec un suivi suffisamment long de personnes âgées, des études longitudinales ont permis de montrer l'apparition successive de différents déficits cognitifs et de symptômes dépressifs au cours d'une phase prodromale de la démence particulièrement longue et progressive (Amieva et al., 2008). Plusieurs années avant le diagnostic clinique de démence, les sujets déments souffrent déjà d'un déclin pathologique de leurs performances cognitives (Joseph et al., 1999; Hall et al., 2000). Une étude a montré que ce déclin était relativement lent dans une première phase et s'accélèrait entre 5 et

8 ans avant le diagnostic de démence. L'âge à l'accélération du déclin cognitif varie selon les études et le type de fonction cognitive étudiée (Hall et al., 2000, 2003).

La phase d'accélération du déclin cognitif et le syndrome MCI

Plusieurs études ont donc montré un déficit cognitif accru chez les sujets déments pouvant être observé bien avant le diagnostic clinique de démence. Ce déficit provient d'un déclin cognitif pré-diagnostique. Plusieurs chercheurs ont tenté d'établir des critères définissant une entité clinique précédant le diagnostic de démence afin de pouvoir mettre en place des stratégies de prévention adaptées. Le syndrome appelé "Mild Cognitive Impairment" (MCI) a été défini dans ce sens (Flicker et al., 1991; Petersen et al., 1999). Dans le contexte du vieillissement cognitif, le MCI est formellement défini comme une altération légère de la mémoire sans diagnostic positif de démence avec des activités du quotidien préservées. Le MCI est un syndrome défini à l'origine pour qualifier les sujets vus en consultation mémoire qui présentaient un état intermédiaire entre le vieillissement normal et la maladie d'Alzheimer. Il a été défini dans l'objectif d'identifier les sujets en phase pré-démentielle avant leur diagnostic positif de démence. Cet état correspond pour certains auteurs à un état transitionnel sans réversibilité entre le vieillissement normal et la démence chez des patients présentant un léger déficit cognitif sans diagnostic positif de démence. Cependant, d'autres considèrent le MCI comme un état latent réversible qui correspondrait à un facteur de risque de la démence (Ritchie et al., 2001; Larrieu et al., 2002). Par exemple, Visser et al. (2006) ont étudié le risque de démence en fonction de l'âge au diagnostic de MCI ; ils montrent notamment que la majorité des sujets MCI avant 85 ans n'évolue pas vers une démence sénile. Ce résultat est appuyé par d'autres études basées sur un suivi long de personnes âgées (Grober et al., 2000; Ganguli et al., 2004) et montrant qu'une proportion importante de sujets diagnostiqués MCI a une évolution cognitive stable au cours du temps et que les critères définissant un MCI peuvent ne plus être remplis par un sujet au cours du temps, ce sujet revenant vers un statut cognitif normal. La notion de MCI est largement discutée dans la littérature (Morris et al., 2001; Jicha et al., 2006) car elle ne permet pas d'établir explicitement un état de pré-démence durant lequel il serait possible d'intervenir en amont de la démence. Le diagnostic de MCI ne semble pas spécifique pour déterminer une phase prodromale de la démence. Dubois et

al. (2007) proposent donc d'autres critères d'évaluation de stades précoces de la démence basés sur un examen clinique des déficits de mémoire épisodique afin d'établir une entité clinique plus opérationnelle pour l'intervention en santé publique.

La distinction entre déclin cognitif normal et déclin cognitif pathologique est primordiale pour la compréhension du processus de dégradation conduisant à la démence et pour la recherche d'outils permettant de détecter plus précocément une démence. Jusqu'à présent, les études menées sur le vieillissement cognitif ont abouti à des résultats assez variables. L'identification d'un déclin des fonctions cognitives dans le vieillissement normal varie suivant la mesure de cognition étudiée (Ganguli et al., 1996; Jacqmin-Gadda et al., 1997a) et notamment s'il s'agit de tests psychométriques dont la réalisation est limitée en temps (Salthouse, 1996; Jacqmin-Gadda et al., 1997b), c'est-à-dire incluant une composante de vitesse.

1.1.2 Facteurs de risque de la démence sénile

L'identification des facteurs de risque de la démence est déterminante dans la mise en place d'une politique de prévention adaptée. Cela permet de cibler des populations à risque différentes et d'adapter la prise en charge en fonction de la population cible. Les principales études des facteurs de risque biologiques, génétiques, environnementaux ou socio-démographiques ont été réalisées à partir de données de cohortes prospectives. La maladie d'Alzheimer et les maladies apparentées du vieillissement cérébral sont des maladies multi-factorielles. La mise en évidence de facteurs de risque du déclin cognitif est difficile tant le déclin cognitif est un processus complexe. Certains facteurs de risque peuvent influer sur le risque de survenue d'une démence sénile et/ou sur l'évolution cognitive.

Le premier facteur déterminant de la survenue d'une démence sénile est l'âge (Letenneur et al., 1994). Il a également été montré que les femmes ont un risque de survenue d'une démence plus élevé que les hommes au-delà de 75 ans (Letenneur et al., 1999; Fratiglioni et al., 2000; Joly et al., 2009). Toutefois, il n'est pas évident de déterminer si ces différences sont intrinsèquement liées à des raisons biologiques et/ou hormonales ou bien d'ordre socio-culturel.

Concernant les facteurs génétiques, il a été montré que l'allèle $\epsilon 4$ du gène codant pour l'apolipoprotéine E (apoE) est un facteur de risque de la survenue d'une démence (Farrer et al., 1997). L'apoE est une protéine transporteuse de lipides codée par un gène situé sur le chromosome 19. Plusieurs auteurs ont montré que le fait de porter l'allèle $\epsilon 4$ de l'apoE était associé à un risque plus élevé de développer une maladie d'Alzheimer. Cependant, le variant $\epsilon 4$ de l'apoE n'est ni nécessaire, ni suffisant pour développer une démence d'Alzheimer ; c'est simplement un facteur de risque. Une forme particulière de la maladie d'Alzheimer par transmission génétique est à considérer à part : il s'agit d'une démence dégénérative de transmission autosomique dominante avec des mutations au niveau des chromosomes 21, 14, et 1, responsables de maladies d'Alzheimer extrêmement précoces (avant 60 ans). Cette forme particulière représente moins de 1% des cas correspondant à des formes monogéniques.

La plupart des études réalisées à partir de données d'incidence montre une association entre un niveau d'éducation bas et un risque accru de démence (Katzman, 1993; Letenneur et al., 1999; Karp, 2004; Jacqmin-Gadda et al., 2006). Le niveau d'éducation a également une influence sur la forme de l'évolution. Des études récentes suggèrent que les sujets de haut niveau d'études ont une accélération plus brutale de leur déclin cognitif dans les dernières années précédant le diagnostic de démence par rapport aux sujets de faible niveau d'études (Amieva et al., 2005; Jacqmin-Gadda et al., 2006; Scarmeas et al., 2006). Stern et al. (1994) suggèrent que cette différence entre les sujets de haut niveau d'études et ceux de faible niveau d'études provient d'une plus grande capacité de réserve cognitive des sujets de haut niveau d'études. Cependant toutes les études épidémiologiques ne retrouvent pas ces résultats concernant l'effet du niveau d'éducation sur le vieillissement cognitif. La signification du niveau d'éducation comme facteur de risque de la survenue d'une démence reste très discutée et mérite d'être précisée.

Bien d'autres facteurs de risque de survenue de démence ont été suspectés pour leur effet sur la survenue ou sur le processus de déclin : certains facteurs sociaux avec, par exemple, la fréquence et le type des activités de loisirs ou sociales (Fabrigoule et al., 1995), des facteurs nutritionnels tels que la consommation de poisson (Larrieu et al., 2004), le stress oxydant (Helmer et al., 2003; Luchsinger et Mayeux, 2004a) et des facteurs de risques vasculaires tels que l'hypertension artérielle pour la démence non-dégénérative,

l'hypercholestérolémie, le diabète ou encore le statut tabagique (De la Torre, 2004; Luchsinger et Mayeux, 2004b; Cowppli-Bony et al., 2006). La solitude et la dépression chronique pourraient augmenter le risque de développer une démence d'Alzheimer (Ownby et al., 2006). Des travaux ont également porté sur l'impact de l'aluminium sur la survenue d'une démence (Rondeau et al., 2000; Flaten, 2001) et suggèrent un risque plus élevé de démence pour des sujets exposés à de l'aluminium présent dans l'eau de boisson.

La recherche de facteurs de risque s'intéresse aujourd'hui aux facteurs de risque modifiables. Il s'agit d'identifier des facteurs de risque ou des facteurs protecteurs pour agir préventivement. Cela représente un réel enjeu de santé publique et une part très importante des travaux épidémiologiques portant sur le vieillissement cognitif des personnes âgées. Cependant, la démence étant un processus de dégradation continu et progressif, l'étude de l'association entre ces facteurs et le diagnostic de démence ne permet pas de comprendre entièrement le mécanisme d'implication de ces facteurs dans le développement d'une démence. Il est primordial d'étudier le déclin des fonctions cognitives au cours du temps en plus du risque de survenue d'une démence sénile afin de mieux comprendre l'histoire naturelle de la maladie et d'évaluer les mécanismes faisant intervenir les facteurs de risque.

1.2 Problèmes méthodologiques dans l'étude du vieillissement cérébral

1.2.1 Outils de mesure de la cognition

Le niveau cognitif d'un sujet peut être évalué par différents tests psychométriques. L'évolution dans le temps des résultats à ces tests reflète le maintien ou la détérioration du niveau cognitif d'un sujet. Le test du "Mini-Mental Score Examination" (MMSE) introduit par Folstein et al. (1975) est l'un des plus populaires auprès des cliniciens en raison de sa rapidité de passation. Il est constitué de 30 items et permet d'évaluer principalement l'orientation dans le temps et l'espace, le langage, les capacités visuo-constructives, la mémoire immédiate et la mémoire différée. Le test d'Isaacs (Isaacs Set Test ou IST)

(Isaacs et Kennie, 1973) évalue la fluence verbale, les fonctions exécutives et la mémoire sémantique avec un score allant de 0 à 40. Le test de Benton (Benton Visual Retention Test ou BVRT) (Benton, 1965) évalue la mémoire visuelle avec un score allant de 0 à 15. Le "Digit Symbol Subsitution Test" de Wechsler (DSSTW) (Wechsler, 1981) caractérise les performances cognitives d'un sujet en terme d'attention et de raisonnement logique avec un score allant de 0 à 90. Pour chacun de ces 4 tests psychométriques, un score bas caractérise de faibles performances cognitives. D'autres tests psychométriques existent comme le test des 5 mots, celui de Grober & Buschke ou encore le test de l'horloge. Chacun de ces tests évalue des dimensions différentes de la cognition telles que la mémoire (épisodique, de travail, sémantique, à court terme, à long terme, visuelle, etc.), le langage, l'attention, les fonctions visuo-spatiales, les fonctions exécutives et l'abstraction. L'échelle des IADL (Instrumental Activities of Daily Living) évalue l'autonomie des personnes âgées dans la vie courante.

Les tests psychométriques sont donc des marqueurs de la démence. Ils doivent être sensibles, c'est-à-dire permettre de distinguer les sujets les uns des autres, et surtout de mettre en évidence l'évolution cognitive d'un sujet âgé. Pour l'étude du vieillissement cognitif, comme pour beaucoup de maladies chroniques, les données dont on dispose sont des mesures répétées au cours du temps d'un marqueur représentant une mesure avec erreur de la quantité étudiée. De plus, les conditions de passation des tests psychométriques font intervenir des dimensions humaines, relationnelles et émotionnelles conduisant à des mesures imparfaites. Les tests psychométriques peuvent être considérés comme des mesures bruitées d'une ou plusieurs quantités latentes : la cognition. Dans certains cas, le choix d'un test psychométrique s'impose car il évalue une fonction cognitive particulière mais la plupart sont des tests évaluant plusieurs dimensions de la cognition. Le choix d'un seul test psychométrique est alors arbitraire si on s'intéresse à l'évolution du niveau cognitif global et peut faire varier les résultats trouvés (Jacqmin-Gadda et al., 1997a; Morris et al., 2001; Amieva et al., 2005). Des méthodes d'analyses multivariées doivent être envisagées pour considérer la corrélation des différents tests psychométriques et avoir une évaluation plus fiable de la cognition des sujets âgés. Par ailleurs, ces marqueurs psychométriques sont souvent non gaussiens et par conséquent difficiles à prendre en compte.

1.2.2 Etude du déclin cognitif pré-démentiel

Le processus de vieillissement cognitif peut être perturbé par la survenue de la démence sénile. Il est bien admis qu'il existe une relation étroite entre la survenue d'une démence et l'altération des fonctions cognitives et que certains facteurs de risque ont une influence sur la survenue d'une démence et/ou sur les scores psychométriques.

Pour distinguer l'évolution cognitive normale et l'évolution pathologique menant à la démence sénile, le déclin cognitif des sujets normaux pourrait être comparé avec celui des futurs déments. Deux groupes de sujets peuvent être définis en fonction de leur diagnostic de démence effectué après une durée de suivi pré-determinée, ce qui suppose que les sujets soient vus à cette date. Les évolutions cognitives de ces 2 groupes de sujets peuvent ensuite être comparées. Ce type d'étude présente plusieurs limites liées aux données incomplètes. D'une part, les sujets en phase pré-diagnostique de démence souffrent d'un déclin de leurs performances cognitives plusieurs années avant le diagnostic clinique de démence (Hall et al., 2003; Jacqmin-Gadda et al., 2006). Il peut donc y avoir une part de sujets en phase de déclin cognitif pré-démentiel parmi les sujets non-déments (censuré à droite de la démence). Cela peut induire une sur-estimation du déclin cognitif dans le groupe de sujets "normaux" et une augmentation de la variance des estimations (Sliwinski et al., 1996, 2003a). D'autre part, pour être inclus dans l'analyse, les sujets doivent être vus à la visite choisie comme date de point pour le diagnostic de démence. Il peut exister un biais de sélection car les sujets présentant un déclin cognitif marqué semblent avoir un risque plus élevé de sortie d'étude (Jacqmin-Gadda et al., 1997a) (cf. section 1.2.5). L'analyse séparée des deux groupes de sujets ne permet donc pas d'évaluer correctement ce déclin pré-démentiel et rend difficile l'évaluation de l'accélération du déclin cognitif.

L'étude séparée de l'évolution cognitive normale et de l'évolution pré-diagnostique pouvant induire des biais dans les analyses de la détérioration cognitive, il est nécessaire d'étudier la population dans son ensemble. De nouveaux développements méthodologiques sont nécessaires pour tenir compte de l'évolution pré-démentielle sans négliger les phénomènes de sélection et ce afin de mieux comprendre le processus de vieillissement cognitif menant à une démence. Le lien unissant la démence et le déclin cognitif suggère l'intérêt d'une approche capable de modéliser ces deux dimensions ensemble. Une modélisation conjointe de la survenue d'une démence et de l'évolution des tests psychométriques au

cours du temps vise à mieux analyser ce lien. Un de nos objectifs est de proposer un modèle conjoint permettant d'étudier l'influence des facteurs de risque simultanément sur les deux composantes et d'en mesurer les effets respectifs.

1.2.3 Non-linéarité du processus d'évolution cognitive

Dans le cas de la détérioration cognitive, le processus d'évolution considéré est souvent non linéaire (Amieva et al., 2005). L'évolution peut être considérée en deux phases : une première durant laquelle le déclin cognitif est lent, une seconde durant laquelle le déclin s'accélère jusqu'à devenir pathologique. De plus, l'évolution des maladies chroniques est un processus soumis à une forte variabilité individuelle. Il est donc nécessaire de considérer des outils permettant à la fois de bien calibrer la forme générale de l'évolution et de distinguer l'évolution normale de l'évolution pathologique, ainsi que de mesurer les variations inter-individuelles.

L'âge au moment de l'accélération du déclin cognitif diffère selon certains facteurs de risque mais aussi selon le type de fonction cognitive étudiée (Hall et al., 2000, 2003). Il est facile d'imaginer que des dimensions de la cognition puissent avoir des évolutions fondamentalement différentes. Certains travaux épidémiologiques présentent la mémoire épisodique comme étant une des dimensions qui décline en premier, alors que l'apraxie et l'agnosie seraient généralement plus tardives. Ceci rejoint l'idée qu'il puisse exister plusieurs processus latents correspondant à différentes dimensions cognitives et évoluant de manière différente. Pour mieux comprendre le processus d'évolution conduisant à la démence, il est donc important de tenir compte de la variabilité autour de l'âge à l'accélération du déclin par des outils de modélisation souples et dynamiques. Dans ce sens, les modèles à état latent constituent une approche intéressante.

1.2.4 Hétérogénéité inter-individuelle des processus d'évolution

Parmi les différents profils d'évolution cognitive, nous distinguons l'évolution normale et l'évolution pathologique. Cependant, certains sujets ont des performances cognitives anormales sans pour autant développer une démence sénile. Le déclin cognitif des personnes âgées est fortement hétérogène notamment parce qu'il existe plusieurs types de démence avec des étiologies différentes (Bird et al., 1989). Certains travaux épidémiolo-

giques suggèrent que la démence ne serait pas la seule source d'hétérogénéité du vieillissement cérébral. Même au sein d'une population de sujets en phase pré-diagnostique de démence, il existerait une hétérogénité des déclins cognitifs (Valdois et al., 1990; Brooks et Yesavage, 1995; Hall et al., 2000; Palmer et al., 2002). Il est également possible qu'il existe différents profils de vieillissement cognitif non pathologique.

Cette hétérogénéité inter-individuelle peut refléter des différences biologiques spécifiques au sujet comme une prédisposition génétique ou bien venir d'une fragilité acquise au cours du temps. Par exemple, on observe une accélération du déclin cognitif plus tardive mais plus accentuée chez les sujets de haut niveau d'études que chez les sujets de faible niveau d'études (Amieva et al., 2005; Jacqmin-Gadda et al., 2006) qui peut s'expliquer par le concept de capacité de réserve décrit par Stern et al. (1994). Cette hétérogénéité inter-individuelle induit une différence pour les risques de développer une démence ou de décéder ainsi que sur le risque d'accélération du vieillissement cognitif. L'étude du processus d'évolution cognitive requiert donc un ajustement sur ces facteurs de risque afin d'évaluer au mieux cette hétérogénéité inter-individuelle. Cependant, les variables observées ne sont souvent qu'une partie des facteurs de risque pertinents. De plus, il est possible que certains facteurs ne soient pas inclus dans les analyses s'ils ne sont pas suspectés d'être influents, ce qui génère d'autres sources d'hétérogénéité. Il est possible de parler d'hétérogénéité sous-jacente pour laquelle un ajustement classique n'est pas envisageable. Cette hétérogénéité non observée doit être prise en compte par des méthodes d'analyse plus complexes comme par exemple les modèles à classes latentes (cf. section 2.3.3) permettant de considérer des groupes d'évolution cognitive différents.

1.2.5 Données incomplètes

Dans l'étude des maladies chroniques à l'aide de données longitudinales, un suivi en continu et durant la vie entière est en pratique impossible dans la mesure où le recueil de données est soumis à des contraintes d'organisation. Les données dont on dispose sont alors incomplètes. La plupart du temps, le mécanisme conduisant à des données incomplètes est complexe. L'ignorabilité de ce processus, concept introduit par Rubin (1976), doit être étudiée. Plusieurs mécanismes peuvent être distingués concernant soit les données de survie, soit les mesures répétées au cours du temps. Seul le problème des données

manquantes de la variable réponse est considéré dans ce travail ; celui des variables explicatives manquantes n'est pas traité.

Censure et troncature en analyse de survie

En ce qui concerne le recueil des données de survie, les sujets peuvent ne pas avoir connu l'événement d'intérêt (la maladie) à la fin de la durée du suivi. Une durée de suivi est dite *censurée à droite* si le sujet n'a pas été diagnostiqué malade à sa dernière observation. La censure à droite est le cas le plus fréquent de données incomplètes en analyse de survie (Kaplan-Meier, 1958; Cox, 1972). Sur le même principe, certains sujets auront déjà connu l'événement d'intérêt au début du suivi, il s'agit de sujets malades à l'inclusion. Ce sont des cas prévalents. La date de survenue de la maladie n'est donc pas connue. Ils sont dits *censurés à gauche*. La plupart du temps, les sujets ne sont pas suivis en temps continu ; le recueil des données est effectué au cours de visites successives durant lesquelles un diagnostic de maladie peut être établi. L'information disponible ne correspond pas à la date exacte de survenue de la maladie. Au mieux, nous savons que la maladie a débuté entre la visite de diagnostic et la visite précédente. On parle alors de *censure par intervalle*. Enfin, une donnée de survie est dite tronquée si elle est conditionnelle à un événement. On parle de *troncature à gauche* si l'événement d'intérêt d'un individu n'est observable que si son temps de survenue est supérieur à une certaine valeur. De manière analogue, on parle de *troncature à droite* si le temps d'événement d'un individu n'est observable que si son temps de survenue est inférieur à une certaine valeur. On parle de *troncature par intervalle* quand les deux mécanismes sont combinés.

Beaucoup d'études sur le vieillissement cognitif des personnes âgées sont réalisées à l'aide de cohortes prospectives dites incidentes. Bien souvent, les sujets prévalents à l'inclusion sont exclus de l'étude. Il y a donc troncature à gauche car l'étude de la survenue d'une démence est effectuée à partir d'une population dont le temps de survenue est supérieur à la date de visite initiale. Dans ces cohortes incidentes de personnes âgées, le recueil des temps d'événement est fortement censuré à droite. Plusieurs causes de censure à droite peuvent être identifiées : le protocole de l'étude suppose une durée totale du suivi qui est une censure administrative, le sujet peut décéder avant de connaître l'événement d'intérêt, le sujet peut refuser de poursuivre l'étude jusqu'à son terme ou être perdu de

vue par les investigateurs. De plus, les diagnostics de démence réalisés au cours de visites génèrent un phénomène de censure par intervalle.

En analyse de survie, la présence de censure et de troncature constitue des limites aux analyses effectuées et doit être intégrée dans l'écriture de la vraisemblance du modèle. De plus, le phénomène de censure à droite nécessite parfois des développements spécifiques si le processus de censure est associé au processus d'intérêt. On parle alors de *censure informative* pouvant induire des biais dans les analyses. La fin de l'étude constitue une censure à droite administrative qui est un processus indépendant du processus d'évolution longitudinale et donc n'influe pas sur l'estimation des paramètres du modèle. Ce type de censure peut être pris en compte dans l'écriture du modèle de manière simple. En revanche, les sorties d'étude, ainsi que les décès (cf. section 1.2.6), associés au déclin cognitif génèrent des mécanismes de censure informative.

Données manquantes au cours du suivi

Dans les études longitudinales, les données manquantes peuvent être monotones ou bien intermittentes. Les données manquantes sont monotones lorsqu'une donnée manquante au temps t implique que toutes les observations suivantes du sujet sont manquantes jusqu'à la fin de l'étude. Les données manquantes monotones sont également appelées sorties d'étude. Les sorties d'étude impactent à la fois la qualité du suivi longitudinal et le recueil des temps d'événement. Les sorties d'étude sont à l'origine de mécanismes de censure à droite. Lorsque les données manquantes peuvent être suivies par une observation, elles sont dites intermittentes.

En ce qui concerne le recueil des données longitudinales, l'impact des données manquantes au cours du suivi ne doit pas être négligé. Une classification des données manquantes, développée en section 2.2.5, a été définie par Little et Rubin (2002) :

- les données sont manquantes complètement aléatoirement (MCAR pour Missing Completely At Random) si la probabilité qu'une donnée soit manquante est indépendante du processus d'intérêt qu'il soit observé ou non.
- les données sont manquantes aléatoirement (MAR pour Missing At Random) si la probabilité qu'une donnée soit manquante est indépendante des valeurs non observées du processus d'intérêt (le score cognitif au temps présent), conditionnellement

aux valeurs observées (les scores cognitifs passés)

– les données sont manquantes non aléatoirement ou informatives (MNAR pour Missing Not At Random) si la probabilité qu'une donnée soit manquante est dépendante des réponses non observées.

Dans l'analyse du suivi d'un sujet, la présence de données manquantes est une limite aux analyses réalisées. Les données incomplètes pouvant être associées au processus d'évolution, certains développements doivent être envisagés pour en tenir compte. On parle alors de *données manquantes informatives* pouvant induire des biais dans les analyses. L'association entre évolution et sortie d'étude a été discutée dans la littérature (Little, 1995; Little et al., 2000). Dans le contexte du vieillissement cérébral, les sorties d'étude au sein des cohortes de personnes âgées ne sont pas ignorables car elles semblent avoir un lien étroit avec le niveau cognitif des sujets (Jacqmin-Gadda et al., 1997a; Rabbit et al., 2004). Rabbit et al. (2005) soulignent également le fait que le déclin cognitif à des âges avancés peut être sous-estimé si le processus de sortie d'étude est négligé. De plus, les phénomènes de sortie d'étude sont plus fréquents chez les personnes âgées que dans la population adulte : les maladies, les entrées en institution et les décès étant plus fréquents. La modélisation de l'évolution cognitive et de la survenue d'un événement doit donc tenir compte de ces mécanismes de censure informative en lien avec l'altération des performances cognitives.

1.2.6 Surmortalité

Le phénomène de sortie d'étude est d'autant plus important dans le contexte du vieillissement cognitif qu'une part non-négligeable de sujets décède au cours du suivi. De plus, des travaux ont montré que le décès est souvent précédé d'un déclin cognitif marqué : on parle de déclin cognitif terminal. Bosworth et al. (1999) ont montré l'existence d'un lien entre performances cognitives et risque de décès mais leur relation varie selon la dimension cognitive considérée. Small et Bäckman (1997) ont étudié les dimensions cognitives prédictives de la survenue d'un décès et ont montré que la mémoire épisodique et la fluence verbale étaient particulièrement sensibles. Wilson et al. (2003) confirment l'accélération du déclin cognitif dans les années précédant le décès et met en relief que cette accélération concerne notamment les dimensions de mémoire épisodique, mémoire sémantique,

mémoire de travail et capacité visuo-spatiale. Small et al. (2003) montrent que les sujets évoluant vers le décès ont de plus faibles performances cognitives que les survivants. Rabbitt et al. (2008) montrent des résultats similaires et évoque le fait que l'analyse du risque de décès puisse être biaisée si les autres causes de sortie d'étude ne sont pas considérées.

Une fois la démence déclarée, la médiane de survie est estimée globalement à 4,5 ans, les femmes décédant plus tardivement que les hommes (Helmer et al., 2001). La survenue d'une démence majore le risque de décès des personnes âgées, multipliant ce risque par 1,8, avec notamment une augmentation du risque de décès par maladie cérébrovasculaire et pathologie respiratoire.

Le phénomène de surmortalité liée à la détérioration cognitive a largement été présenté dans la littérature. Le décès constitue donc une véritable source de censure informative à droite. Des outils d'analyse de l'évolution cognitive et/ou du risque de survenue d'une démence doivent être développés pour tenir compte de cette censure informative induite par le décès. Ce problème de censure à droite informative est d'autant plus important qu'il est couplé à un problème de censure par intervalle, les diagnostics de démence étant établis au cours de visites successives en temps discret. En effet, si les diagnostics de démence étaient connus à leur âge exact, l'étude du déclin cognitif pré-démentiel pourrait se faire de manière simple. Mais les sujets en phase pré-démentielle semble avoir un surrisque de décès. Un sujet peut donc décéder entre une visite avec diagnostic de démence négatif et la visite suivante au cours de laquelle il aurait pu être diagnostiqué dément. Le décès induit donc une censure à droite informative pour la démence, dont il est difficile de tenir compte. La seule information disponible concernant un sujet de ce type est qu'il est décédé après une visite à diagnostic de démence négatif. Il apparait donc nécessaire d'utiliser des outils de modélisation pouvant tenir compte de l'imbrication de ces différents phénomènes, à savoir, d'une part, l'évolution cognitive, la survenue de la démence et les problèmes de censure par intervalle ; et d'autre part, la survenue du décès et les problèmes de censure à droite informative.

1.3 Le projet PAQUID

Durant les 25 dernières années, plusieurs grandes cohortes prospectives (Framingham Study, Seattle Longitudinal Study, Rotterdam Study, Twin Study of Aging, ...) ont été mises en place dans le but d'étudier le vieillissement cérébral des personnes âgées. Les éléments de connaissance sur le déclin cognitif, présentés dans les sections précédentes, ont en partie pu être obtenus à partir de ces études neurologiques et psychologiques.

Initiée en 1988, la cohorte française PAQUID (pour Personnes Agées Quid ?) a été l'une des premières cohortes européennes portant sur l'étude du vieillissement cognitif des personnes âgées. L'étude PAQUID est une cohorte prospective constituée de 3777 personnes âgées de plus de 65 ans, réparties sur 75 communes des départements de la Dordogne et de la Gironde. La représentativité de la population générale a été assurée par un tirage au sort stratifié selon l'âge, le sexe ainsi que la taille des unités urbaines de résidence. Les sujets étaient inclus dans l'étude s'ils étaient âgés d'au moins 65 ans au 31 décembre 1987, s'ils ne se trouvaient pas en institution au moment du recueil des données initiales et s'ils étaient inscrits sur les listes électorales. La méthodologie de l'étude PAQUID a été décrite par Dartigues et al. (1992).

Objectifs

La cohorte PAQUID est une étude épidémiologique dont l'objectif général est de s'intéresser au vieillissement cérébral et fonctionnel après 65 ans. Plusieurs objectifs spécifiques ont été définis :

- décrire l'état fonctionnel des personnes âgées en terme de déficience et ainsi distinguer les modalités d'évolution normale et pathologique,
- étudier l'impact de facteurs de risque sur le déclin cognitif,
- identifier, au moyen des tests neuropsychologiques, des individus susceptibles de développer une démence et à haut risque de détérioration physique ou intellectuelle chez lesquels une action préventive ou thérapeutique serait possible,
- estimer l'incidence et la prévalence de la démence, et plus particulièrement celle de la maladie d'Alzheimer,
- étudier l'évolution des démences incidentes en terme de dépendance, d'entrée en institution et de mortalité,

- étudier la symptomatologie dépressive du sujet âgé, et sa liaison avec la survenue d'une démence,
- étudier la perte d'autonomie du sujet âgé et déterminer des facteurs propres au sujet ou à son environnement qui entraînent un risque élevé d'entrée en institution à court-terme, chez des sujets vivant auparavant à leur domicile.

Dans cette étude, en plus de la visite initiale, les sujets ont été rencontrés à leur domicile 3 ans, 5 ans, 8 ans, 10 ans, 13 ans, 15 ans et 17 ans après le début du suivi. Les sujets girondins ont effectué une visite supplémentaire à 1 an. Des informations concernant les aspects socio-démographiques, l'état de santé et l'environnement social ont été relevés. Les performances cognitives des sujets ont été mesurées à l'aide de tests psychométriques dont le MMSE, le BVRT, l'IST (cf. section 1.2.1). Le diagnostic de démence a été porté suivant les critères du manuel diagnostique et statistique des troubles mentaux dans sa troisième version (DSM-III R) et des critères développés par la "National Institute of Neurological and Communicative Disorders and Stroke - Alzheimer Disease and Related Disorders Association" (NINCDS-ADRDA) permettant de définir une démence d'Alzheimer. A chaque visite, un dépistage de la démence est effectué par l'enquêteur. Si ce diagnostic est positif, un neurologue réalise un examen clinique pour confirmer ou infirmer le diagnostic de démence et préciser son étiologie (maladie d'Alzheimer, démence vasculaire, etc.).

Des études de test-retest sur des courtes périodes de temps ont mis en évidence une augmentation du score MMSE indiquant la possibilité d'un effet d'apprentissage (Tombaugh et McIntyre, 1992). Jacqmin-Gadda et al. (1997a) retrouvent cette augmentation lorsque le test est effectué à la visite à 1 an. L'intervalle de temps de 1 an réduit l'hypothèse d'un effet d'apprentissage. Il peut s'agir d'un effet de stress engendré par la visite initiale. Quelle que soit son origine, cet effet est réel et a été appelé "effet de première passation".

Le projet PAQUID a été conçu davantage pour estimer l'incidence que la prévalence de la démence. En effet, le tirage au sort des sujets a exclu les personnes âgées en institution où se trouve une proportion supérieure de cas de démences par rapport à la population à domicile. Les cas prévalents (déments à la visite initiale) dans le projet PAQUID ne sont donc pas représentatifs. De plus, la probabilité de refus de participation est probablement

supérieure parmi les déments. Il existe un biais de sélection. Il est donc judicieux d'exclure les déments prévalents : le problème de censure à gauche se transforme en un problème de troncature à gauche dont il faut tenir compte dans la modélisation. Les sujets étant vus à des visites pré-établies, les diagnostics de démence ne peuvent se faire qu'en temps discret, ce qui induit un phénomène de censure par intervalle. L'information sur l'âge de décès d'un sujet est recueillie en temps continu par un appel téléphonique au médecin ou à un proche, même si les sujets ont quitté l'étude. En pratique, le phénomène de censure pour le décès est une censure administrative correspondant à la durée de l'étude, c'est-à-dire 15 ans sur les données utilisées dans ce travail puisque l'informatisation des observations au suivi à 17 ans n'était pas terminée. Les âges de censure pour la démence et le décès peuvent donc être différents.

La prévalence des démences et de la maladie d'Alzheimer a été estimée à partir des données de suivi à 10 ans (Ramaroson et al., 2003). Les participants étant maintenant âgés de 75 ans et plus, ces estimations ne concernent que les personnes de cette tranche d'âge. Ces estimations sont présentées suivant l'âge et le sexe dans la table 1.1.

Tab. 1.1 : Prévalence de la démence et de la maladie d'Alzheimer en %

	Homme		Femme		Total	
	Demence	MA	Démence	MA	Démence	MA
$75-79$ ans	7.7	4.6	5.7	3.7	6.5	4.1
$80-84$ ans	12.5	9.6	16.6	15.3	15.1	13.2
$85-89$ ans	22.9	15.2	29.9	23.8	27.6	21.0
90 ans et plus	27.0	21.6	52.8	46.5	47.0	40.9
Total	13.2	9.1	20.5	17.1	17.8	14.2

Avec le suivi à 10 ans de la cohorte PAQUID, l'incidence de la démence a été estimée à 22 pour 1000 personnes-années et l'incidence de la maladie d'Alzheimer est de 15 pour 1000 personnes-années pour les sujets âgés de plus de 75 ans. L'analyse de l'incidence de la démence en fonction de l'âge et du sexe avait montré des résultats intéressants (Commenges et al., 1998; Letenneur et al., 1999) qui ont été confirmés plus récemment

à l'aide du suivi à 13 ans (Joly et al., 2009). Avant 75 ans, l'incidence est plus élevée chez les hommes que chez les femmes. A l'aide de ces estimations de l'incidence et des données de recensement de la population française, on estime à plus de 855 000 le nombre de personnes âgées de plus de 65 ans atteintes d'une maladie d'Alzheimer ou d'un autre syndrome démentiel et le nombre de nouveaux cas est évalué à 225 000 chaque année (OPEPS, 2006).

Les données de la cohorte PAQUID sont celles qui sont utilisées dans l'ensemble des applications présentées dans la suite de ce travail. L'étude de la cohorte PAQUID comporte un certain nombre des problèmes méthodologiques évoqués précédemment tels que l'hétérogénéité de la population des personnes âgées, la non-linéarité de l'évolution, les données incomplètes englobant les problèmes de censure et de troncature ainsi que les problèmes de sorties d'étude informatives et de surmortalité.

1.4 Réflexions générales

Bien que les études épidémiologiques existantes sur le vieillissement cérébral aient permis l'amélioration des connaissances sur le déclin cognitif, son histoire naturelle reste mal connue. Le déclin cognitif survenant en phase pré-diagnostique de démence a été décrit dans la littérature (Amieva et al., 2005; Hall et al., 2000, 2001) mais sa forme et la manière dont il se distingue du déclin cognitif normal ne sont pas encore bien établies. Les mécanismes d'implication des facteurs de risque d'un déclin, d'une démence ou d'un décès ne sont pas non plus clairement identifiés.

Actuellement, la recherche médicale porte sur le développement de nouveaux traitements permettant d'améliorer les performances cognitives des patients pré-déments et déments (Ringman et Cummings, 2003). Les efforts de recherche sur le diagnostic précoce de la démence sénile sont indispensables afin d'initier les traitements le plus tôt possible pour agir sur le processus de dégradation des fonctions cognitives. Un travail de recherche épidémiologique est également nécessaire concernant les facteurs de risque modifiables afin de mettre en place une politique de prévention et d'améliorer la prise en charge des sujets déments. L'établissement d'un diagnostic précoce est donc une nécessité car le volet curatif du vieillissement pathologique est vecteur d'espoir.

La contribution des statisticiens à la recherche globale sur le processus de vieillissement cérébral, ses causes, ses conséquences et son traitement doit porter sur le développement d'outils permettant de modéliser le déclin cognitif afin d'améliorer les connaissances de la maladie, de préciser le rôle des différents facteurs de risque et de permettre un dépistage précoce de la démence. Les méthodes statistiques doivent tenir compte de problèmes méthodologiques liés aux caractéristiques des données, de la population et des phénomènes étudiés. Les outils d'analyse doivent notamment être adaptés pour considérer des processus d'évolution non-linéaires et hétérogènes, des données incomplètes et des phénomènes de sélection liés à la surmortalité des sujets présentant un déclin cognitif ou une démence. Un grand nombre des problèmes méthodologiques évoqués sont communs à l'étude de beaucoup d'autres maladies chroniques comme l'étude de l'infection par le VIH, l'évolution des pathologies cancéreuses.

1.5 Objectif et plan du mémoire

L'objectif de ce travail est d'étudier et d'étendre des méthodes statistiques pour l'analyse d'un processus longitudinal. Ce travail, spécifiquement développé dans le contexte du vieillissement cognitif des personnes âgées, a pour but de répondre à un certain nombre de problèmes méthodologiques, à savoir :

- l'évolution non-linéaire,
- la modélisation de l'hétérogénéité de la population,
- la distinction entre évolution normale et évolution pathologique et notamment l'identification de l'accélération du déclin cognitif,
- l'association de l'évolution cognitive avec la survenue de la démence et du décès,
- la prise en compte des données incomplètes associées à la censure, la troncature, les sorties d'étude ou la surmortalité.

En ce qui concerne l'application de ce travail au contexte du vieillissement cérébral, l'objectif est d'analyser l'histoire naturelle de l'évolution cognitive des personnes âgées ainsi que d'étudier l'impact de facteurs de risque sur le déclin cognitif et la survenue de la démence en tenant compte des problèmes de sélection liés à la mortalité. Le travail de recherche effectué porte sur le développement de modèles conjoints permettant l'étude

simultanée de plusieurs processus d'intérêt.

Dans le chapitre 2, nous proposons une revue de la littérature des méthodes statistiques utilisées et développées dans ce travail. La première partie expose les notions indispensables à la compréhension de l'analyse de la survenue d'événement. Nous développons ensuite la méthodologie concernant l'analyse de données longitudinales. Puis, nous présentons les modèles conjoints pour étudier simultanément l'évolution d'un marqueur quantitatif et la survenue d'un événement. Enfin, nous terminons cette section par la présentation de méthodes d'estimation de ces modèles.

Dans le chapitre 3, nous avons réalisé une analyse de sensibilité d'un modèle non-linéaire pour données longitudinales multivariées à processus latent en comparant deux approches pour traiter des sorties d'étude informatives : celle des Pattern Mixture Model et celle des modèles à classes latentes.

Dans le chapitre 4, nous développons un modèle conjoint à état latent pour données longitudinales et données de survie. Ce modèle combine un modèle bi-phasique pour décrire l'évolution cognitive et un modèle multi-états à état latent pour décrire les risques de survenue de la démence et du décès.

Le chapitre 5 est une synthèse des approches statistiques proposées mettant en relief leurs points de convergence. Nous discutons des intérêts et des limites sur un plan général de ce type de modélisation.

Chapitre 2

Etat des connaissances

2.1 Analyse de données de survie

L'analyse de données de survie est un domaine des statistiques qui trouve sa place dans tous les champs d'application où l'on étudie la survenue d'un événement. L'objectif de cette analyse réside dans l'analyse du délai de survenue d'un événement dans un ou plusieurs groupes d'individus. Dans le domaine biomédical, plusieurs événements sont intéressants à étudier : le développement d'une maladie, la réponse à un traitement donné, la rechute d'une maladie ou le décès.

2.1.1 Modèles de survie

En analyse de survie, nous nous attachons à décrire la distribution des temps de survie, à comparer la survie de plusieurs groupes de sujets et à étudier l'impact de facteurs de risque sur le délai de survenue de l'événement d'intérêt. Cet événement peut être associé à un changement d'état au cours de la progression de la maladie. Dans le cas le plus simple, l'analyse de survie peut se voir comme l'étude du passage irréversible entre deux états fixés (Figure 2.1).

Fig. 2.1 : Structure du modèle de survie simple

Définitions et notations

La variable étudiée T^* est la durée de survie, définie comme le délai écoulé entre l'état 0 et l'état 1. Pour définir ce délai, il est nécessaire de définir une date d'origine qui est la date de début du phénomène étudié. Par exemple, dans l'étude de l'évolution d'une maladie, la date d'origine T_{orig} est la date de début de la maladie et si on s'intéresse à l'âge du sujet à la survenue de l'événement, la date d'origine sera la date de naissance du sujet ($T_{\text{orig}} = 0$). Chaque individu peut avoir une date d'origine différente. Il est également nécessaire de définir une échelle de temps. La durée de survie T^* est une variable aléatoire positive et continue. Plusieurs fonctions caractérisent la distribution du temps de survenue de l'événement T^* :

- la fonction de densité

  On note $f(t)$ la fonction de densité de T^* à valeurs dans $\mathbb{R}^+$ définie par

  $$f(t) = \lim_{dt \to 0} \frac{P(t \leq T^* < t + dt)}{dt}$$

  $f(t)dt$ est la probabilité de décéder entre t et $t + dt$ pour un sujet.

- la fonction de répartition

  La fonction de répartition $F(t)$ mesure la probabilité de connaître l'événement au plus tard en t :

  $$F(t) = P(T^* < t) = \int_{T_{\text{orig}}}^{t} f(v)dv$$

  La fonction de répartition est croissante avec $F(T_{\text{orig}}) = 0$ et $\lim_{t \to +\infty} F(t) = 1$

- la fonction de survie

  La fonction de survie $S(t)$ mesure la probabilité de ne pas connaître l'événement avant t :

  $$S(t) = P(T^* \geq t) = 1 - F(t) = \int_{t}^{+\infty} f(v)dv$$

  La fonction de survie est décroissante avec $S(T_{\text{orig}}) = 1$ et $\lim_{t \to +\infty} S(t) = 0$

- la fonction de risque

  La fonction de risque $\alpha(t)$ est définie par

  $$\alpha(t) = \lim_{dt \to 0} \frac{P(t \leq T^* < t + dt | T \geq t)}{dt}$$

  $\alpha(t)dt$ est la probabilité de décéder entre t et $t+dt$ pour un sujet, conditionnellement au fait que ce sujet n'ait pas connu l'événement à l'instant t.

La fonction $\alpha(t)$ est également appelée intensité de transition et traduit le risque de passer de l'état 0 à l'état 1. Elle peut également s'écrire de la manière suivante

$$\alpha(t) = \frac{f(t)}{S(t)}$$

Il s'en suit

$$S(t) = \exp\left(-\int_{T_{\text{orig}}}^{t} \alpha(v)dv\right)$$

– la fonction de risque cumulé

La fonction de risque cumulé $\Lambda(t)$ est définie par

$$\Lambda(t) = \int_{T_{\text{orig}}}^{t} \alpha(v)dv$$

Toutes ces fonctions (f, S, F, α, Λ) permettent de décrire la distribution de la durée de survie T^*. La fonction de risque instantané $\alpha(t)$ est une des fonctions les plus intéressantes car elle donne une vision probabiliste du futur d'un sujet n'ayant pas encore connu l'événement d'intérêt. La fonction de risque instantané a également l'avantage de refléter des différences entre les modèles souvent moins lisibles sur les fonctions de répartition ou les fonctions de survie. De plus, si l'âge est choisi comme temps de base, la fonction de risque peut être assimilée à l'incidence d'une maladie en fonction de l'âge. La distribution du temps de survie que l'on cherche à modéliser est généralement inconnue. Deux approches sont possibles pour estimer cette distribution : l'inférence paramétrique et l'inférence non-paramétrique.

Censure et troncature

En analyse de survie, les données recueillies sont la plupart du temps incomplètes : la censure et la troncature sont des mécanismes inhérents au recueil des données de survie. Il est nécessaire d'en tenir compte dans l'écriture de la vraisemblance du modèle. Dans la section 2.1, la variable durée de survie T^* est définie comme le délai écoulé entre la date d'origine T_{orig} et la date de survenue de l'événement. Dans le cas de données de survie censurées, la durée de survie T^* n'est pas observée pour tous les sujets et on définit alors l'événement par le couple (T, δ), δ étant l'indicatrice d'observation de l'événement et T le

temps correspondant à l'information connue (le temps d'événement ou le temps de censure).

– Censure à droite

Une durée de survie est dite censurée à droite si l'individu n'a pas connu l'événement d'intérêt à sa dernière visite. La censure à droite est l'exemple le plus fréquent d'observation incomplète en analyse de survie, et a largement été décrite dans la littérature (Andersen et al., 1993). Les conditions sous lesquelles la vraisemblance classique pour données de survie censurées à droite est valide ont également été étudiées (Lagakos, 1979; Kalbfleisch et Prentice, 1980). Formellement, la durée de survie d'un événement pour un individu est définie par le couple (T, δ) où

$$T = \min(T^*, C)$$

et

$$\delta = \begin{cases} 1 \text{ si } T^* \leq C \\ 0 \text{ si } T^* > C \end{cases}$$

avec la durée de vie T^* et la date de censure C supposées indépendantes, le processus de censure est alors non-informatif (Lawless, 1982). Si $\delta = 1$, le sujet subit l'événement et est observé. Si $\delta = 0$, le sujet est dit censuré à droite : au lieu d'observer T^*, on observe une valeur C avec pour seule information le fait que T^* soit supérieure à C.

– Censure à gauche

Une durée de survie est dite censurée à gauche si l'individu a déjà connu l'événement d'intérêt avant l'entrée dans l'étude. Formellement, la durée de survie pour un individu est définie par le couple (T, δ) où

$$T = \max(T^*, C)$$

et

$$\delta = \begin{cases} 0 \text{ si } T^* \leq C \\ 1 \text{ si } T^* > C \end{cases}$$

avec la durée de vie T^* et la date de censure C supposées indépendantes. Si $\delta = 1$, le sujet subit l'événement et est observé. Si $\delta = 0$, le sujet est dit censuré à gauche : au lieu d'observer T^*, on observe une valeur C avec pour seule information le fait que T^* soit inférieure à C. Pour traiter la censure à gauche, Andersen et al. (1993) ont proposé d'inverser l'échelle de temps pour se rapporter alors à un problème de censure à droite. Ceci n'est possible que dans le cas de données exclusivement censurées à gauche.

– Censure par intervalle

Souvent, le recueil des données se fait au cours de visites successives. Cela induit un phénomène de censure par intervalle si, au lieu d'observer la variable T^*, on observe deux valeurs C_1 et C_2 ($C_1 < C_2$) et si la seule information disponible sur T^* est que $C_1 < T^* < C_2$. La distribution de T^* est supposée indépendante des valeurs C_1 et C_2. Ce type de censure a déjà été étudié en analyse de survie (Odell et al., 1992; Sun et al., 2001; Chi et Tseng, 2002) ainsi que dans le contexte des modèles multi-états (Satten et Sternberg, 1999; Commenges, 2002; Joly et Commenges, 1999; Joly et al., 2002).

– Troncature

Une donnée de survie est dite tronquée si elle est conditionnelle à un événement. Il s'agit de troncature à gauche si un temps de survie T n'est observable qu'à la condition $T > U$, U étant une variable supposée indépendante de T. Par symétrie, la troncature à droite est définie si un temps de survie T n'est observable qu'à la condition $T < U$. Il est également possible d'avoir la combinaison des deux mécanismes, on parle alors de troncature par intervalle. De nombreux travaux ont été effectués sur l'analyse de données tronquées (Lagakos et al., 1988; Kalbfleisch et Lawless, 1989; Klein et Moeschberger, 1997).

Estimation

En analyse de survie, l'approche paramétrique consiste à considérer que la distribution de la durée de survie étudiée suit une distribution théorique connue. Un modèle de survie paramétrique est donc un modèle dans lequel la fonction $\alpha(t;\theta)$ est une fonction mathématique qui dépend du vecteur de paramètres θ. Kalbfleisch et Prentice (1980) ou encore Lawless (1982) décrivent un grand nombre de modèles paramétriques pouvant être utilisés en analyse de survie. Cox et Oakes (1984) discutent des critères permettant de choisir entre plusieurs modèles paramétriques. Les lois paramétriques considérées doivent être des lois adaptées pour des variables aléatoires positives ; c'est notamment pour cette raison que la loi normale ne convient pas dans ce contexte. Les modèles paramétriques les plus utilisés sont les modèles de Weibull, exponentiel, log-normal. L'estimation des paramètres θ du modèle peut se faire par la méthode du maximum de vraisemblance. L'étude du temps de survie T^* est réalisée à partir d'un échantillon de N sujets défini comme une suite de variables aléatoires $(T_1, T_2, ..., T_N)$ indépendantes, de même loi de probabilité. La vraisemblance est la probabilité d'observer un échantillon particulier $(t_1, t_2, ..., t_n)$. La vraisemblance est une fonction des paramètres du modèle et s'écrit :

$$L(\theta) = \prod_{i=1}^{N} l_i(\theta)$$

où $l_i(\theta)$ est la contribution de la i-ème observation à la vraisemblance.

L'expression de $l_i(\theta)$ dépend de la nature de l'observation. Dans le cas où l'observation est exacte, la contribution d'un sujet à la vraisemblance s'écrit $l_i(\theta) = f(t_i;\theta)$. Dans le cas où la censure est non informative, c'est-à-dire dans le cas où la durée de vie T^* est indépendante de la censure C, les contributions individuelles à la vraisemblance s'écrivent simplement :

– si l'observation est censurée à droite par c_i, alors

$$l_i(\theta) = S(c_i;\theta)$$

– si l'observation est censurée à gauche par c_i, alors

$$l_i(\theta) = 1 - S(c_i;\theta) = F(c_i;\theta)$$

– si l'observation est censurée par intervalle par $[c_{1i}, c_{2i}]$, alors

$$l_i(\theta) = S(c_{1i},\theta) - S(c_{2i},\theta)$$

En présence de phénomènes de troncature, les contributions individuelles à la vraisemblance peuvent également s'écrire simplement en les conditionnant par la probabilité de l'événement définissant la troncature. Cela revient à diviser la contribution individuelle à la vraisemblance par cette probabilité :

– si l'observation est tronquée à gauche par u_i, alors

$$l_i(\theta) = \frac{f(t_i;\theta)}{S(u_i;\theta)}$$

– si l'observation est tronquée à droite par u_i, alors

$$l_i(\theta) = \frac{f(t_i;\theta)}{1 - S(u_i;\theta)} = \frac{f(t_i;\theta)}{F(u_i;\theta)}$$

– si l'observation est tronquée par intervalle $[u_{1i},u_{2i}]$, alors

$$l_i(\theta) = \frac{f(t_i;\theta)}{S(u_{1i};\theta) - S(u_{2i};\theta)}$$

Par extension de ces différents cas, il est possible de définir la contribution à la vraisemblance dans des cas correspondant à des combinaisons de censure et troncature, comme le cas suivant fréquemment rencontré :

– si l'observation est censurée à droite par c_i et tronquée à gauche par u_i, alors

$$l_i(\theta) = \frac{S(c_i;\theta)}{S(u_i;\theta)}$$

L'estimateur du modèle est obtenu en calculant la valeur de θ qui rend maximale la fonction de vraisemblance $L(\theta)$. Le calcul de la log-vraisemblance du modèle est souvent plus pratique :

$$l(\theta) = \log\big(L(\theta)\big) = \sum_{i=1}^{N} \log\big(l_i(\theta)\big)$$

L'approche paramétrique induit des hypothèses sur la distribution de données, difficiles à vérifier avec des données incomplètes. Une alternative possible est l'utilisation de méthodes non-paramétriques. L'estimateur non-paramétrique le plus simple de la fonction de distribution est la distribution empirique. Il correspond à l'estimateur non-paramétrique du maximum de vraisemblance pour des observations complètes. De nombreux estimateurs ont été développés afin de considérer les mécanismes de censure et troncature. Les

plus connus sont l'estimateur de la fonction de survie de (Kaplan-Meier, 1958) et celui de la fonction de risque cumulé de Nelson-Aalen (Nelson, 1972; Aalen, 1975) pour traiter des données censurées à droite. L'estimateur de Kaplan-Meier peut être adapté pour des données à la fois tronquées à gauche et censurées à droite ou des données censurées à gauche. D'autres développements ont été proposés pour considérer des données tronquées à gauche, tronquées à droite ou encore censurées par intervalle (Peto, 1973; Turnbull, 1976; Woodroofe, 1985; Lagakos et al., 1988; Andersen et al., 1993). Un inconvénient majeur de ces estimateurs non-paramétriques du maximum de vraisemblance est de fournir une estimation en temps discret. Il est difficile d'en déduire une estimation de la fonction de risque ; le lissage de la fonction de risque estimée *a posteriori* est une solution envisageable.

Une autre approche non-paramétrique a été développée pour obtenir un estimateur non-paramétrique de la fonction de risque sans faire d'hypothèse forte sur la forme de la distribution des temps de survie : l'approche par vraisemblance pénalisée (Silverman, 1985). La log-vraisemblance est pénalisée par un terme qui est d'autant plus grand que la fonction de risque est peu lisse.

Modèle de régression

En analyse de survie, l'un des objectifs est d'étudier la survie en fonction des groupes à risque et d'évaluer l'impact de facteurs de risque sur le délai de survenue de l'événement d'intérêt. Pour prendre en compte les différences de survie entre les groupes à risque, les modèles incluent des variables explicatives représentant les facteurs de risque. Le modèle le plus utilisé en analyse de survie pour étudier l'impact de variables explicatives sur le risque de survenue d'un événement est le modèle à risques proportionnels, également appelé modèle de Cox (1972). Ce modèle est dit semi-paramétrique car il permet d'établir une relation paramétrique entre les facteurs de risque de l'événement et la distribution des durées de survie sans imposer à celle-ci une forme paramétrique. Il existe d'autres approches semi-paramétriques comme les modèles de survie accélérée (Kalbfleisch et Prentice, 1980; Cox et Oakes, 1984). Le modèle des risques proportionnels exprime une relation entre la fonction de risque instantané α et un vecteur de variables explicatives $x = (x_1, ..., x_p)$:

$$\alpha(t, x) = \alpha^0(t) r(\beta, x)$$

où $\beta' = (\beta_1, ..., \beta_p)'$ est le vecteur des coefficients de régression et $\alpha^0(t)$ est la fonction de

risque instantané de base. Plus précisément, $\alpha^0(t)$ est le risque instantané de connaître l'événement des sujets pour lequel toutes les covariables x_i sont égales à 0. La fonction $r(\beta, x)$ dépend des caractéristiques x du sujet, et cette dépendance est mesurée par les coefficients β. Cox (1972) a proposé la formulation suivante $r(\beta, x) = \exp(\beta' x)$ de manière à obtenir une fonction de risque instantané positive et sans contrainte sur les coefficients β. Le modèle s'écrit alors :

$$\alpha(t, x) = \alpha^0(t)\exp(\beta' x)$$

Les modèles de régression peuvent être définis de manière paramétrique : le risque instantané de base $\alpha^0(t)$ est défini par une distribution paramétrique. L'approche par maximisation de la vraisemblance peut donc être utilisée pour estimer les paramètres de régression β. Mais les modèles de régression peuvent également être non-paramétriques avec un risque instantané de base $\alpha^0(t)$ non-paramétrique. Dans cette situation, Cox (1972) a proposé d'estimer les paramètres de régression β en utilisant la vraisemblance dite partielle du modèle et ne nécessitant pas l'estimation de la fonction de risque de base. Une approche alternative est la vraisemblance pénalisée permettant à la fois l'estimation du risque de base $\alpha^0(t)$ et des paramètres de régression β.

Modèle pour données corrélées

Les modèles de survie ont été étendus dans le contexte de données multivariées de survie afin de tenir compte de la corrélation entre des temps de survie. Lin et Wei (1984) et Wei et al. (1989) se sont intéressés à une formulation marginale de la survie pour des événements répétés alors que Lee et al. (1992) ont développé cette approche pour des données groupées. Cette approche, issue des modèles marginaux de type GEE (Generalized Estimating Equation) développée à l'origine dans le contexte de données longitudinales (Liang et Zeger, 1986), consiste à spécifier la fonction de risque marginale des temps de survie corrélés sans modéliser de façon explicite la structure de dépendance entre les temps de survie. D'autres développements en analyse de survie ont permis de prendre en compte des données corrélées (groupées, répétées ou récurrentes) : les modèles à fragilité. Le plus souvent, les modèles à fragilité ont été définis par une extension du modèle des risques proportionnels avec

$$\alpha(t, x|z_i) = z_i\alpha^0(t)\exp(\beta' x)$$

où z_i est un effet aléatoire spécifique à chaque groupe (ou à chaque sujet), appelé variable de fragilité, qui agit multiplicativement sur le risque, de sorte qu'une valeur élevée de cette variable augmente le risque. La littérature concernant ces modèles est abondante (Clayton, 1978; Clayton et Cuzyck, 1985; Nielsen, 1992; Yashin et al., 1995; Korsgaard et Andersen, 1998). La variable de fragilité représente l'ensemble des facteurs de risque non observés et communs à un même groupe (ou à un même sujet) qui vont fragiliser les individus d'un même groupe (ou les mesures répétées d'un même sujet) et être responsables de la dépendance des sujets dans le groupe (des mesures d'un sujet). Ce type de modèle de survie à effets aléatoires permet de quantifier la variabilité entre les groupes et la dépendance intra-groupe ou encore la variabilité des données répétées d'un sujet ainsi que leur dépendance pour un sujet, ce qui n'est pas possible dans une approche marginale.

A l'origine, les modèles à fragilité ont permis de modéliser une hétérogénéité résultant de variables individuelles non observées. Cette hétérogénéité conduit à une sélection de la population : les sujets les plus fragiles décèdent en premier et la population survivante, plus robuste, est différente de la population d'origine (Vaupel et Yashin, 1985; Aalen, 1994; Aalen et Husebye, 1991; Hougaard, 1995, 2000).

Modèle à fraction guérie

Les modèles de survie supposent classiquement que la totalité de la population étudiée est à risque de connaître l'événement d'intérêt. Cependant, pour beaucoup de maladies, on peut supposer qu'une partie de la population n'est pas à risque de connaître l'événement. Les modèles de mélange prennent en compte le fait que les sujets sont issus de différentes sous-populations sans que l'on puisse savoir à laquelle ils appartiennent. Les modèles à fraction guérie (Cure model) ont été développés pour considérer qu'une partie de la population a un risque nul de connaître l'événement (Boag, 1949; Farewell, 1982). Ce concept a largement été développé dans le contexte du modèle à risques proportionnels (Kuk et Chen, 1992; Peng et Dear, 2000; Sy et Taylor, 2000). Des travaux concernant ce type de modèle ont également été effectués dans une approche non-paramétrique (Maller et Zhou, 1996). Les modèles à fragilité diffèrent des modèles à fraction guérie dans le sens où ils supposent que tous les individus sont à risque de connaître l'événement d'intérêt avec un risque pouvant varier mais non nul (Aalen, 1988; Hougaard et al., 1994).

2.1.2 Modèles multi-états

Le modèle de survie présenté dans la section précédente ne comporte que 2 états : un seul événement est étudié. Bien souvent, les analyses de survie classiques font l'hypothèse d'indépendance du délai de survenue de l'événement d'intérêt et du délai de censure. On parle de censure non-informative. Cette hypothèse est discutable dans beaucoup de situations comme dans le cadre de l'analyse de risques compétitifs (Gail, 1975) où plusieurs événements sont en concurrence et peuvent constituer une censure informative. Par exemple, pour l'étude de la survenue d'un événement, le décès est souvent considéré comme une censure à droite. Cependant, il est parfois difficile de supposer que le temps de décès est indépendant du temps d'événement, particulièrement lorsque la maladie considérée augmente le risque de décès. Il peut donc s'avérer nécessaire d'étudier simultanément la survenue de plusieurs événements.

Définitions et notations

Les modèles multi-états sont une généralisation des modèles de survie. Ils offrent la possibilité d'étudier l'évolution des sujets à travers les différents états d'un processus et de se focaliser sur les risques instantanés au cours du temps de survenue des événements associés à ces états. Un processus multi-états est un processus stochastique $\{R(t), t > 0\}$ prenant ses valeurs dans un espace fini d'états $\mathcal{S} = \{1, ..., K\}$. Le processus R a une distribution initiale $\pi_h(0) = P[R(0) = h]$, h appartenant à $\mathcal{S}$ et génère un historique $\mathcal{X}_t$ (une σ-algèbre) défini comme l'observation du processus dans l'intervalle $[0, t]$. Les notations précédemment introduites dans le cadre de l'analyse de survie simple sont généralisées aux modèles multi-états en fonction de cet historique.

- $\alpha_{kl}(t)$ est la fonction d'intensité de transition

$$\alpha_{kl}(t) = \lim_{dt \to 0} \frac{P_{kl}(R(t + dt) = l | R(t) = k, \mathcal{X}_t)}{dt}$$

  $\alpha_{kl}(t)dt$ est la probabilité d'effectuer une transition de l'état k à l'état l entre t et $t + dt$ pour un sujet, conditionnellement au fait que ce sujet soit dans l'état k à l'instant t

- $\Lambda_{kl}(t)$ est la fonction d'intensité de transition cumulée :

$$\Lambda_{kl}(t) = \int_{T_{\text{orig}}}^{t} \alpha_{kl}(v)dv$$

– $P_{kl}(s,t)$ est la probabilité de transition de l'état k au temps s vers l'état l au temps t qui peut s'écrire comme suit :

$$P_{kl}(s,t) = P\big(R(t) = l | R(s) = k, \mathcal{X}_s\big) \text{ avec } s < t$$

Structures standards de modèles multi-états

Les états peuvent caractériser la bonne santé, un niveau de gravité d'une maladie, une rémission ou le décès. Les états du processus sont souvent définis en fonction des symptômes cliniques présentés par les sujets au cours de leur suivi, ou en fonction des marqueurs biologiques. Un changement d'état est appelé transition. Lorsque le processus peut sortir d'un état, ce dernier est dit transitoire. A l'inverse, un état dont un processus ne peut pas sortir est dit absorbant. Des modèles multi-états complexes avec un nombre important d'états et des formes de transitions variées sont envisageables. Toutefois, ils possèdent des structures qui ne sont que des extensions de structures standards (Hougaard, 1999).

– Modèle à risques compétitifs

Le modèle à risques compétitifs peut être appliqué lorsque plusieurs événements terminaux peuvent se produire (Andersen et al., 1993). Ce modèle, présenté en figure 2.2, permet de prendre en compte plusieurs événements d'intérêt exclusifs, comme par exemple différentes causes de mortalité ou de sortie d'étude. Les événements sont dits en concurrence. Ce modèle possède un état transitoire et plusieurs états absorbants : lorsqu'un événement a eu lieu, les autres ne peuvent plus se produire.

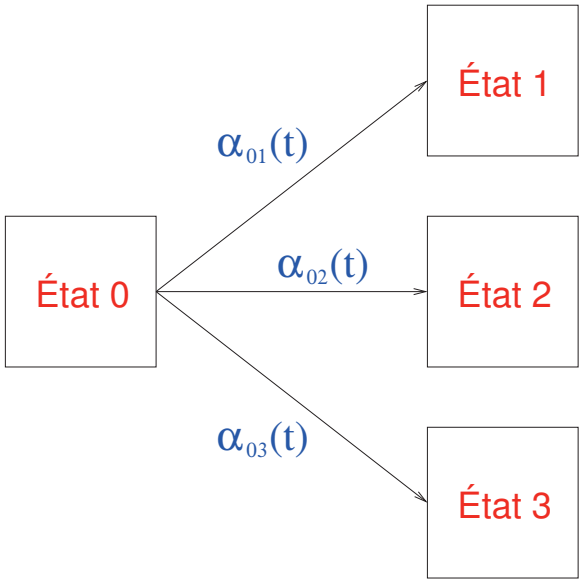

Fig. 2.2 : Structure du modèle multi-états à risques compétitifs

– Modèle à états progressifs

Le modèle à états progressifs, présenté en figure 2.3, permet de prendre en compte des états de transition se succédant avant l'état terminal (Hougaard, 1999). Le principal atout de cette approche par rapport à l'analyse de survie classique est de pouvoir caractériser la progression de la maladie en plus de l'étude du délai de survenue de l'événement terminal. Ce modèle permet également d'étudier l'impact de variables explicatives sur la vitesse de progression de la maladie.

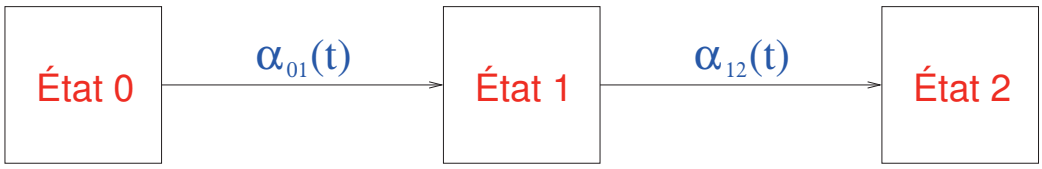

Fig. 2.3 : Structure du modèle multi-états progressifs

– Modèle "Illness-death"

Le modèle "Illness-death" permet d'étudier l'évolution d'un patient atteint d'une maladie irréversible, particulièrement lorsque cette maladie augmente le risque de décès (Therneau et Grambsch, 2000). Le modèle, présenté en figure 2.4, comporte un état absorbant et deux états transitoires : l'état 0 représente l'absence de maladie, l'état 1 correspond à l'état de maladie et l'état 2 représente le décès. Il est en particulier utilisé pour traiter des problèmes de risques semi-compétitifs, c'est-à-

dire des risques d'événements non exclusifs. Il permet notamment de comparer le taux de mortalité chez des sujets sains à celui des sujets malades.

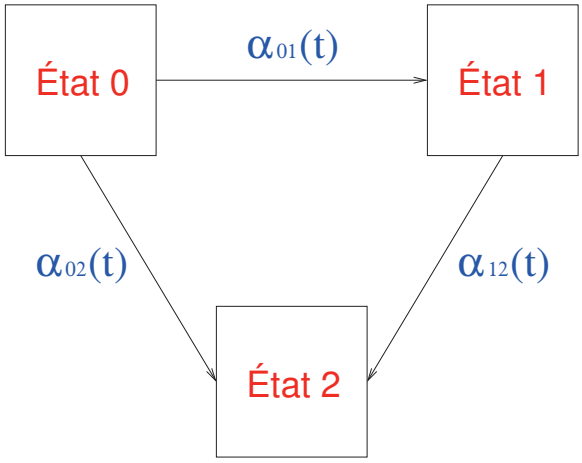

Fig. 2.4 : Structure du modèle Illness-death

Hypothèses markoviennes

Dans la définition du modèle utilisé, le choix de la structure des états (compétitif, progressif ou Illness-death) représente donc un point majeur de la modélisation de l'évolution de la maladie. Beck et Paucker (1983) et Aalen et Johansen (1978) sont parmi les premiers à avoir introduit les modèles de Markov dans l'analyse clinique multi-états. Karlin et Taylor (1975) ont développé des processus markoviens à temps continu et à espaces d'états discrets. L'hypothèse markovienne consiste à supposer les probabilités de transition d'un état à l'autre indépendante de la totalité de l'historique $\mathcal{X}_t$. L'évolution future du processus est supposée dépendre uniquement de l'état du processus au temps courant :

$$P_{kl}(s,t) = P\big(R(t) = l | R(s) = k, \mathcal{X}_t\big) = P\big(R(t) = l | R(s) = k\big)$$

Un modèle multi-états est dit homogène si les intensités de transition ne dépendent pas du temps, c'est-à-dire si $\alpha_{kl}(t) = \alpha_{kl}$. Cela entraîne $P_{kl}(s, s+d) = P_{kl}(t, t+d) = P_{kl}(0, d)$. Sinon, le modèle est dit non homogène, soit $P_{kl}(s,t) = P(R(t) = l | R(s) = k)$.

Dans le domaine du vivant, l'hypothèse markovienne est souvent trop forte. Les modèles markoviens non-homogènes permettent de modéliser une matrice de probabilités dépendantes du temps et donc d'assouplir la restriction faite dans les modèles markovien homogène. Dans le cas où les intensités de transition dépendent non pas de la durée totale

du suivi t, mais du temps écoulé depuis la dernière transition d ($\alpha(d)$), le modèle est dit semi-markovien.

Intérêts des modèles multi-états

Les modèles multi-états permettent d'étudier l'évolution complexe de sujets pouvant connaître plusieurs événements. Par exemple, ils permettent de décrire les risques instantanés α_{kl} de transition entre les états, les probabilités de transition d'un état à l'autre et l'effet de variables explicatives sur les intensités de transition. Il est possible de tenir compte d'une hétérogénéité dans la population, caractérisée par un vecteur de covariables x dans les modèles multi-états. L'un des modèles de régression les plus utilisés pour les processus multi-états est le modèle des intensités proportionnelles $\alpha_{kl}(t) = \alpha^0_{kl}(t)\exp(\beta' x)$. La théorie relative aux modèles multi-états est de plus en plus riche (Andersen, 1988; Kalbfleisch et Lawless, 1989; Hougaard, 1999) et les applications de ces modèles en épidémiologie, plus récentes, sont appelées à se développer (Commenges, 1999) notamment dans le contexte de données incomplètes avec des mécanismes de troncature et de censure. Ils ont très largement été utilisés dans le contexte de l'infection par le VIH (Longini et Clarks, 1989; Frydman, 1992; Gentleman et al., 1994; Aalen et al., 1997; Alioum et al., 1998) ainsi que dans de nombreuses autres maladies chroniques (Kay, 1986; Andersen, 1988; Klein et al., 1993; Marshall et Jones, 1995; Keiding et al., 2001; Saint-Pierre et al., 2003; Putter et al., 2006) comme les pathologies cancéreuses ou l'asthme. Ils s'avèrent extrêmement utiles pour l'estimation d'incidence, de taux de décès, ou d'espérance de vie.

En présence d'événements récurrents, ces modèles permettent de s'intéresser à la survenue d'un nombre aléatoire d'un même événement pour chaque sujet. Toutefois, les états sont tous transitoires et le dernier état ne peut pas être considéré comme absorbant.

Modèles de Markov cachés

Les modèles multi-états supposent que les états sont observables. Si un sujet est observé au temps t, l'état dans lequel il se trouve est supposé connu. Dans certains cas, l'état du sujet à l'instant t n'est pas directement observé, mais une autre variable associée au processus multi-état est observée. Les modèles de Markov cachés (Hidden Markov Model) permettent de modéliser des états progressifs supposés existants dans le développement de

la maladie mais non observables, on parle alors d'états latents. Un processus de Markov caché peut être considéré comme un couple de processus stochastiques $\{R(t), O(t), t > 0\}$ tel que le processus $\{R(t)\}$, appelé processus d'état, soit une chaîne de Markov non-observée et que le processus observé $\{O(t)\}$ soit lié au processus d'état par une fonction probabiliste f telle que $O(t) = f(R(t))$. Les modèles de Markov cachés supposent l'indépendance des observations sachant les états :

$$\forall\ s, t\ \ \{O(s) \perp O(t) | R(s), R(t)\}$$

Cette approche par processus de Markov caché a été utilisée dans différentes applications médicales comme la modélisation de l'infection par le VIH (Guihenneuc-Jouyaux et al., 2000), d'une sclérose en plaque (Altman et Petkau, 2005), la reconnaissance de gènes (Krogh, 1998). L'inclusion d'effets aléatoires permet de tenir compte de la variabilité inter-individuelle (Humphreys, 1998; Seltman, 2002; Altman, 2007) et de lever l'hypothèse d'indépendance des observations sachant les états (Altman, 2007).

2.1.3 Application au vieillissement cognitif

A partir des données de la cohorte Paquid, plusieurs travaux en analyse de survie ont été réalisés dans le contexte du vieillissement cérébral normal et pathologique des personnes âgées. La fonction de risque instantané de survenue d'une démence peut être interprétée comme l'incidence de la démence en fonction de l'âge.

Une première approche a été d'utiliser une méthode à noyaux pour obtenir un estimateur de la fonction de risque lisse à partir de l'estimation obtenue par une version corrigée de la méthode de Turnbull, celle-ci ne fournissant pas directement d'estimation de la fonction de risque (Commenges et al., 1998). La méthode de Turnbull consiste à caractériser un intervalle C (réunion d'intervalles disjoints) en dehors duquel l'estimateur non paramétrique du maximum de vraisemblance de la fonction de survie est constant, puis à chercher parmi les fonctions de survie constantes par morceaux en dehors de C celle qui maximise la vraisemblance.

Joly et al. (1999) ont proposé une seconde approche non-paramétrique pour estimer directement une fonction de risque lisse dans un modèle à risques proportionnels avec données tronquées à gauche et censurées par intervalle. Cette approche, implémentée

dans le programme PHMPL, est basée sur la vraisemblance pénalisée (Joly et al., 1998). L'estimation de l'incidence suggère que les hommes ont un risque de démence plus élevé que les femmes dans les âges les moins avancés et que, après 75 ans, les femmes ont un risque supérieur à celui des hommes. Cette approche par vraisemblance pénalisée de l'estimation de la fonction de risque a par la suite été étendue à des modèles à fragilité pour données groupées (Rondeau et al., 2003) : elle n'a pas montré d'hétérogénéité géographique significative du risque de démence entre les différentes communes de la cohorte Paquid.

Dans le contexte du vieillissement des personnes âgées, l'un des problèmes est la réduction des capacités fonctionnelles, évaluée en terme d'incapacité ou de dépendance. L'étude du processus d'évolution de l'autonomie fonctionnelle des personnes âgées en fonction de la progression et de la récupération à travers différents stades d'incapacité croissante a été réalisée à partir des données de la cohorte Paquid. Un modèle à 5 états a été défini avec 4 états transitoires et réversibles représentant les niveaux d'incapacité et un état absorbant caractérisant le décès (Barberger-Gateau et al., 2004). Le processus d'évolution vers l'incapacité ou le décès est un processus de Markov non-homogène avec des intensités de transitions constantes par morceaux pour lequel Alioum et Commenges (2001) ont développé un progamme appelé MKVPCI. L'étude de l'effet de covariables dans le modèle a permis de clarifier le rôle de certains facteurs individuels sur l'évolution de l'incapacité et la mortalité (Alioum et al., 2004). L'implication de la démence sur le processus d'évolution de l'incapacité montre un impact considérable sur le processus de perte d'autonomie fonctionnelle du sujet âgé. Les sujets déments ont un risque plus élevé d'évolution vers des stades d'incapacités modérées et sévères (Barberger-Gateau et al., 2004).

Un travail sur la prise en compte de la mortalité dans l'étude du vieillissement cognitif a été réalisé d'abord avec un modèle progressif à 3 états pour données tronquées à gauche (Helmer et al., 2001) puis à l'aide d'un modèle "Illness-death" semi-paramétrique estimé par vraisemblance pénalisée. L'objectif était d'étudier l'incidence de la démence chez les sujets âgés en prenant en compte la mortalité de ces sujets (Joly et al., 2002) puisqu'il peut y avoir une sous-estimation de l'incidence en utilisant un modèle de survie simple en raison de la censure par intervalle. Cela a permis de vérifier que la différence observée entre les risques de démence des femmes et des hommes n'était pas liée à un risque de décès différent. Cependant, le croisement entre les risques instantanés de démence, estimé

à 75 ans par l'analyse de survie simple, a lieu un peu plus tardivement, vers 80 ans, et les incidences sont plus élevées que celles obtenues auparavant. Le modèle considéré est un modèle de Markov non-homogène avec l'âge comme temps de base. Deux processus sont modélisés dans ce modèle : le processus d'apparition d'une démence qui est continu mais observé en temps discret et le processus de décès qui est observé en temps continu. La transition de l'état 0 à l'état 1 est censurée par intervalle ou à droite et les transitions vers l'état 2 sont observées ou censurées à droite. Pour la démence, les sujets sont observés à des visites supposées indépendantes de l'état où se trouve le sujet. Ce modèle "Illness-death" permet également d'évaluer le surrisque de décès des sujets déments.

Un modèle de Markov non-homogène (Commenges et Joly, 2004) à 5 états (non-dément vivant à domicile, dément vivant à domicile, non-dément vivant en institution, dément vivant en institution et décédé) a été proposé pour estimer le risque d'entrée en institution des sujets déments et non-déments dans la cohorte Paquid.

D'autres travaux ont également porté sur l'étude du vieillissement cognitif des personnes âgées à partir de données d'autres études épidémiologiques et utilisant des méthodes d'analyse de survie complexe.

Ripatti et al. (2003) ont proposé un modèle "Illness-death" pour étudier conjointement le risque de survenue d'une démence et du décès. La particularité de ce modèle est l'inclusion d'une composante de fragilité permettant de considérer une hétérogénéité latente au sein de la population. Ce modèle est l'extension naturelle des modèles à fragilité au contexte des modèles multi-états.

Salazar et al. (2007) ont étudié le risque de transition vers la démence et/ou le décès à l'aide d'un modèle multi-états. Le modèle mis en oeuvre contient 2 états absorbants (dément, décédé) et 3 états transitoires (sain, MCI amnésique, MCI non amnésique). Le modèle est un modèle de Markov homogène et il permet de modéliser l'évolution progressive d'un sujet âgé vers un stade sévère de détérioration cognitive (démence ou décès) en tenant compte d'éventuelles transitions vers les états MCI (amnésique, non amnésique). En revanche, ce modèle ne permet pas de prendre en considération le surrisque de décès des sujets déments, ces deux événements étant en concurrence. De plus, l'hypothèse d'homogénéité des probabilités de transition entre les états mériterait d'être assouplie.

Van Den Hout et Matthews (2008) ont également étudié le risque de survenue d'une

démence conjointement au risque de survenue du décès à l'aide d'un modèle "Illness-death" avec des intensités de transition entre les états constantes par morceaux : le modèle développé est un modèle markovien non-homogène. Il a comparé l'hypothèse paramétrique d'intensités de transition constantes par morceaux avec des intensités dépendantes de loi de Weibull. Enfin, Van Den Hout et al. (2009) ont étendu cette approche au contexte des processus de Markov caché.

2.2 Analyse de données longitudinales

L'étude de données longitudinales, c'est-à-dire l'analyse de l'évolution de variables mesurées de façon répétée au cours du temps, présente un intérêt majeur en épidémiologie car elle permet d'étudier l'évolution des individus face à une pathologie. Par exemple, l'étude d'un ou plusieurs marqueurs au cours du temps permet de percevoir les variations liées à la dégradation de l'état de santé du patient ou encore une amélioration en réaction à la prise d'un traitement. Les mesures répétées au cours du temps d'un sujet sont souvent fortement corrélées, deux observations chez un même sujet ayant tendance à avoir une valeur plus proche que deux mesures prises chez deux individus distincts. Il est maintenant bien admis que la modélisation temporelle de données répétées au cours du temps doit tenir compte de cette corrélation afin de ne pas conduire à une inférence incorrecte des paramètres du modèle d'évolution.

2.2.1 Modèles mixtes pour données longitudinales

Modèle linéaire mixte

Aujourd'hui, le modèle linéaire mixte, introduit par Harville (1977) et popularisé par Laird et Ware (1982), apparaît comme la méthode de référence pour l'étude d'un marqueur Gaussien au cours du temps. Il permet de prendre en considération la corrélation des observations dans un contexte de données répétées et permet de décrire l'évolution moyenne au cours du temps pour la population ainsi que les évolutions individuelles par l'intermédiaire d'effets aléatoires individuels mesurant l'écart de chacun des individus par rapport à l'évolution moyenne de la population. L'idée sous-jacente au modèle linéaire à effets aléatoires est que la variable réponse suit un modèle de régression linéaire mais avec

des coefficients de régression spécifique à chaque sujet. Dans une population de N sujets, le vecteur des observations du sujet i est noté $Y_i = (Y_{i1}, ..., Y_{in_i})$ où n_i est le nombre de mesures répétées du marqueur pour le sujet i. Le modèle linéaire mixte est défini de la manière suivante :

$$\begin{cases} Y_i = X_i\beta + Z_i u_i + \epsilon_i \\ u_i \sim \mathcal{N}(0, G) \\ \epsilon_i \sim \mathcal{N}(0, \Sigma_i) \\ u_i \perp \epsilon_i, \quad \forall\ i \text{ et } u_i \perp u_j, \quad \forall\ i, j \end{cases}$$

où X_i est une matrice de variables explicatives de dimension $n_i * p$ (incluant notamment le temps) associée au vecteur d'effets fixes β, Z_i est une sous-matrice de X_i de dimension $n_i * q$ associée au vecteur d'effets aléatoires u_i spécifiques à chaque individu (avec $q \leq p$ le nombre d'effets aléatoires). Le vecteur d'erreur ϵ_i est supposé indépendant du vecteur d'effets aléatoires u_i pour chaque sujet. Soit $f_i(Y_i|u_i)$ et $f_i(u_i)$ les fonctions de densités respectives du vecteur des mesures répétées sachant les effets aléatoires et du vecteur d'effets aléatoires. Ces distributions définissent la formulation hiérarchique du modèle linéaire mixte. La densité marginale du vecteur des mesures s'écrit donc :

$$f_i(Y_i) = \int f_i(Y_i|u_i) f_i(u_i) du_i$$

Il est alors facile de montrer que $f_i(Y_i)$ est la densité d'une loi multivariée normale :

$$Y_i \sim \mathcal{N}(X_i\beta, V_i = Z_i G Z_i' + \Sigma_i)$$

Cette formulation marginale du modèle est celle classiquement utilisée pour l'estimation du vecteur des paramètres du modèle $\theta = (\beta, \alpha)$, incluant les paramètres de régression β et le vecteur de paramètres α des matrices de variance-covariance G et Σ_i. Cependant, la formulation marginale du modèle ne prend pas explicitement en considération l'hétérogénéité entre les sujets au travers des effets aléatoires. La matrice de variance-covariance $V_i = Z_i G Z_i' + \Sigma_i$ doit être définie positive. Il est possible, selon la forme de Z_i, que V_i soit définie positive alors que G ne l'est pas. Dans ce cas, le modèle marginal est valide alors que le modèle hiérarchique ne l'est pas. Pour avoir concordance entre la formulation marginale et la formulation hiérarchique, il faut contraindre la matrice G à être définie positive, en reparamétrant le modèle en utilisant la décomposition de Cholesky de G.

Une approche classique pour l'estimation du vecteur de paramètres θ du modèle est la maximisation de la vraisemblance marginale :

$$L(\theta) = \prod_{i=1}^{N} \left(\frac{1}{2\pi}\right)^{n_i/2} |V_i|^{-1/2} \exp\left(-\frac{1}{2}(Y_i - X_i\beta)' V_i^{-1} (Y_i - X_i\beta)\right) \tag{2.1}$$

La log-vraisemblance est plus souvent utilisée pour des raisons pratiques. Conditionnellement aux paramètres de variance α, l'estimateur du maximum de vraisemblance (MLE pour Maximum Likelihood Estimator) du vecteur de paramètres β est donné par Laird et Ware (1982) :

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' \hat{V}_i^{-1} X_i\right)^{-1} \sum_{i=1}^{N} X_i' \hat{V}_i^{-1} Y_i$$

avec $\hat{V}_i = V_i(\hat{\alpha})$. L'estimateur des éléments de la matrice de variance-covariance $\hat{\alpha}$ est connu pour être biaisé dans le cas d'une estimation par maximum de vraisemblance. Patterson et Thompson (1971) ont proposé une approche alternative (REML pour REstrictive Maximum Likelihood). Les différences entre les estimateurs MLE et REML sont très faibles sur de grands échantillons. Les applications développées dans cette thèse portent sur de grands échantillons, les estimateurs utilisés sont ceux du maximum de vraisemblance.

L'estimation des paramètres du modèle par maximum de vraisemblance peut se faire de plusieurs manières. Pour l'approche marginale, Harville (1977) préconise l'utilisation des algorithmes itératifs comme l'algorithme EM (Dempster et al., 1977) ou l'algorithme de Newton-Raphson (Fletcher, 2000). Le principe de ces algorithmes sera développé dans la section 2.4. Une alternative à l'approche par maximum de vraisemblance est l'approche bayésienne. La formulation hiérarchique du modèle mixte à effets aléatoires a été utilisée dans un cadre bayésien par Laird et Ware (1982) ou encore Verbeke et Molenberghs (2000). Un algorithme envisageable dans un cadre bayésien peut être l'algorithme MCMC (Markov Chain Monte Carlo).

Modèle linéaire généralisé à effets mixtes

Les modèles linéaires généralisés, introduits par McCullagh et Nelder (1989), sont une extension des modèles linéaires classiques pour des variables non-gaussiennes. L'introduction d'effets aléatoires dans les modèles linéaires généralisés permet de modéliser la corrélation des données et définit la classe des modèles linéaires généralisés à effets mixtes

(Davidian et Giltinan, 1995; McCulloch et Searle, 2004). La variable réponse Y_{ij} est supposée suivre une distribution de la famille exponentielle, conditionnellement aux vecteurs d'effets aléatoires u_i :

$$f_{Y_i|u_i}(y_i|u_i) = \exp\left\{ \frac{y_i\theta_i - g(\theta_i)}{\tau^2} - c(y_i, \tau) \right\}$$

et l'espérance conditionnelle de y_i est définie par

$$E[y_i|u_i] = \mu_i = \frac{\partial g(\theta_i)}{\partial \theta_i}$$
$$h(\mu_i) = X_i\beta + Z_i u_i$$

où h est appelée fonction de lien associant l'espérance conditionnelle de y_i au prédicteur linéaire. La variance conditionnelle de Y_i s'écrit :

$$Var(Y_i|u_i) = (\tau^{-2})\frac{\partial^2 g(\theta_i)}{\partial \theta_i^2}$$

2.2.2 Modèles de mélange pour données longitudinales

La question de l'hétérogénéité d'une population est un problème statistique longuement abordé dans la littérature (Redner et Walker, 1984; Richardson et Green, 1997; Bohning et Seidel, 2003). Les modèles de mélange constituent une réponse assez intuitive à la prise en compte de données hétérogènes. La population est supposée constituée de Q sous-populations (ou classes latentes) au sein desquelles les observations ont la même distribution (Bohning, 2000).

Formulation

Verbeke et Lesaffre (1996) ont étendu la notion d'hétérogénéité de la population à des profils d'évolution hétérogène en proposant un modèle de mélange pour données longitudinales. Ce modèle est une extension du modèle linéaire mixte proposé par Laird et Ware (1982) (cf. section 2.2.1). Pour introduire le concept de mélange, le modèle linéaire mixte doit être écrit légèrement différemment :

$$Y_i = X_i\beta + Z_i u_i + \epsilon_i$$

où la matrice de variables explicatives Z_i n'est plus une sous-matrice de X_i, mais une matrice distincte et la distribution des effets aléatoires n'est plus d'espérance nulle $u_i \sim \mathcal{N}(\mu, G)$.

L'extension de ce modèle à des données hétérogènes où l'on suppose Q sous-populations d'évolutions distinctes se fait au travers de la distribution des effets aléatoires définie comme un mélange de Q lois normales :

$$u_i \sim \sum_{q=1}^{Q} \pi_q \mathcal{N}(\mu_q, G_q)$$

avec G_q le plus souvent égal à $\omega_q G$ où ω_q est un coefficient de proportionnalité spécifique à la classe q. Chaque profil d'évolution a une probabilité π_q d'existence. La somme des probabilités sur les Q classes d'évolution vaut 1 ($\sum_{q=1}^{Q} \pi_q = 1$). Les paramètres β peuvent aussi être spécifiques à la classe.

D'après Muthén et Shedden (1999), il est possible de définir l'indicatrice c_{iq} qui vaut 1 si le sujet i appartient à la classe q, ce qui donne $P(c_{iq} = 1) = \pi_q$. La fonction de densité du vecteur de données Y_i peut alors s'écrire ainsi :

$$f_i(Y_i) = \sum_{q=1}^{Q} \pi_q f(Y_i | c_{iq} = 1)$$

Conditionnellement à c_{iq}, il est alors possible de déterminer les évolutions moyennes de chacun des profils avec $f(Y_i|c_{iq} = 1)$ qui est une densité multivariée gaussienne d'espérance $E_{iq} = X_i\beta + Z_i\mu_q$ et de variance $V_i = Z_i G_q Z_i' + \sigma^2 I_{n_i}$.

Estimation

L'un des inconvénients majeurs des modèles de mélange vient du nombre de paramètres à estimer qui dépend du nombre Q de classes d'évolution. Des travaux portent sur l'estimation simultanée des paramètres et du nombre de profils d'évolution (Bohning, 1995; Richardson et Green, 1997; Wang et al., 2004) mais les méthodes classiques d'estimation restent difficilement utilisables dans ce contexte. La plupart des approches, en particulier en présence de données répétées, consistent à effectuer 2 étapes : l'estimation des paramètres pour différentes valeurs de Q puis la détermination du nombre optimal de classes. Pour Q fixé, l'algorithme le plus souvent utilisé est l'algorithme EM (Verbeke et Lesaffre, 1996; Muthén et Shedden, 1999; Spiessens et al., 2002), mais Proust et Jacqmin-Gadda (2005) proposent en alternative l'utilisation d'un algorithme de Marquardt amélioré qui est une approche de type Newton-Raphson. Pour déterminer le nombre optimal de profils d'évolution, le Bayesian Information Criterion (BIC) est un critère semblant donner de bons résultats (Hawkins et al., 2001; Zhang et Cheng, 2004).

Certains auteurs ont également proposé un test du type rapport de vraisemblance pour comparer un modèle à Q_0 composantes versus un modèle à Q_1 composantes (Ghosh et Sen, 1985; Chen et Chen, 2001). Cependant, sous les différentes hypothèses de test, les distributions asymptotiques de la statistique du rapport de vraisemblance sont complexes et difficilement calculables autrement que par Bootstrap (McLachlan, 1987; Chen et Chen, 2001). Les résultats asymptotiques usuels ne sont pas valides puisque sous l'hypothèse nulle certains paramètres ne sont pas identifiables, ce qui rend compliquée la mise en oeuvre d'un test.

Classification a posteriori

Après estimation des paramètres par la méthode du maximum de vraisemblance et détermination du nombre optimal de classes, une classification *a posteriori* des sujets peut être réalisée suivant les classes latentes estimées. Pour chaque sujet, il est possible de calculer la probabilité d'appartenance à chacune des classes à l'aide du théorème de Bayes :

$$\hat{\pi}_{iq} = P(c_{iq} = 1|Y_i, \hat{\theta}) = \frac{\hat{\pi_q} f(Y_i|c_{iq} = 1; \hat{\theta})}{\sum_{l=1}^{G} \hat{\pi}_l f(Y_i|c_{iq} = 1; \hat{\theta})}$$

où $f(Y_i|c_{iq} = 1; \hat{\theta})$ est la densité multivariée Gaussienne définie dans le paragraphe précédent prise aux valeurs estimé de $\hat{\theta}$.

Extension

La probabilité d'appartenance aux classes d'évolution peut être décrite en fonction des caractéristiques du sujet par un modèle de régression multinomiale (Muthén et Shedden, 1999) :

$$\pi_{iq} = P(c_{iq}|X_{2i}) = \frac{e^{\xi_{0q} + X'_{2i}\xi_{1q}}}{\sum_{l=1}^{Q} e^{\xi_{0l} + X'_{2i}\xi_{1q}}} \text{ avec } \xi_{01} = 0 \text{ et } \xi_{11} = 0$$

où ξ_{oq} est l'intercept de la classe q et ξ_{1q} le vecteur de paramètres spécifiques à la classe q associé aux vecteurs de variables explicatives X_{2i}. Par la suite, certains auteurs ont proposé de modéliser conjointement un événement sous forme de variable binaire dont la probabilité dépend des classes latentes. Plusieurs travaux (Muthén et Shedden, 1999; Muthén, 2002; Lin et al., 2002a) sont à l'origine de l'approche par classes latentes dans l'étude conjointe d'un marqueur longitudinal et d'un événement clinique, sur laquelle nous reviendrons dans la section 2.3.

2.2.3 Modèles pour données longitudinales multivariées

Modèle mixte pour données multivariées

Avec le modèle linéaire à effets aléatoires dans sa formulation standard, l'évolution d'un seul marqueur longitudinal peut être étudiée. Mais en présence de données multivariées, il se peut que l'on s'intéresse d'une part à l'évolution des marqueurs et d'autre part à la relation entre ces différentes variables. Il est donc nécessaire de modéliser la corrélation de ces marqueurs. L'analyse conjointe de plusieurs marqueurs permet d'augmenter la puissance des analyses et d'étudier la structure de dépendance des marqueurs.

Une première approche consiste à étendre le modèle linéaire mixte. Nous considérons le cas avec deux marqueurs longitudinaux, la méthodologie étant la même dans un cas plus complexe. Pour un modèle bivarié, le vecteur des variables réponses peut être scindé en deux : $Y_i = \begin{bmatrix} Y_{1i} \\ Y_{2i} \end{bmatrix}$ avec Y_{ki} le vecteur des n_{ki} mesures du marqueur k ($k = 1, 2$). Le modèle mixte s'écrit alors :

$$\begin{cases} Y_i = X_i\beta + Z_i u_i + \epsilon_i \\ u_i \sim \mathcal{N}(0, G) \text{ et } G = \begin{bmatrix} G_{11} & G_{12}^T \\ G_{12}^T & G_{22} \end{bmatrix} \\ \epsilon_i \sim \mathcal{N}(0, \Sigma_i) \\ u_i \perp \epsilon_i, \quad \forall\ i \end{cases}$$

et $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, $X_i = \begin{bmatrix} X_{i1} & 0 \\ 0 & X_{i2} \end{bmatrix}$, $u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}$, $Z_i = \begin{bmatrix} Z_{i1} & 0 \\ 0 & Z_{i2} \end{bmatrix}$ et $\epsilon_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix}$ avec X_{ik} une matrice de variables explicatives pour le marqueur k associée au vecteur d'effets fixes β_k et Z_{ik} une matrice de variables explicatives (sous matrice de X_{ik}) associée au vecteur d'effets aléatoires u_{ik}.

La vraisemblance du modèle pour données multivariées est similaire à celle développée dans le cas univarié (2.1). Les algorithmes précédemment évoqués peuvent donc être étendus au contexte multivarié. Shah et al. (1997) proposent un algorithme EM alors que Verbeke et Molenberghs (2000) ou encore Thiébaut et al. (2002) proposent un algorithme de type Newton-Rapson, ayant l'avantage de pouvoir être implémenté dans les logiciels statistiques classiques.

Les marqueurs sont indépendants si les matrices G et Σ_i sont block-diagonales. La structure de corrélation des marqueurs peut être assouplie en introduisant des processus stochastiques corrélés modifiant la structure de Σ_i. Thiébaut et al. (2002) proposent par exemple d'introduire des processus bivariés autorégressifs d'ordre 1. Toutefois, plus le nombre de marqueurs à étudier est important, plus le nombre de paramètres à estimer est grand. Un nombre trop important de paramètres de variance-covariance pour les effets aléatoires peut générer des difficultés numériques (Fieuws et al., 2007).

Modèle multivarié à processus latent

Une autre approche consiste à supposer que les différents marqueurs sont des mesures bruitées d'un même processus latent. L'approche par modèle mixte à variables latentes proposée par Roy et Lin (2000) est une extension du modèle linéaire mixte pour données multivariées. La quantité non-observée U_{ij} pour le sujet i au temps j est explicitement modélisée par un modèle linéaire mixte classique, ce qui permet d'évaluer son évolution :

$$U_{ij} = X_{ij}\alpha + Z_{ij}u_{ij} + \epsilon_{ij}$$

où X_{ij} est un vecteur de variables explicatives associé au vecteur d'effets fixes α, Z_{ij} est un sous-vecteur de variables explicatives associé au vecteur d'effets aléatoires gaussien u_{ij} et les erreurs ϵ_{ij} sont supposées gaussiennes centrées réduites. L'effet d'une variable explicative sur l'évolution du processus latent est mesuré par les paramètres α. Un modèle d'observation fait le lien entre la variable latente et les marqueurs longitudinaux. Chaque mesure d'un marqueur k est une combinaison linéaire de la variable latente au temps de mesure à une erreur de mesure près :

$$Y_{ijk} = \beta_{0k} + U_{ij}\beta_{1k} + b_{ik} + e_{ijk}$$

où b_{ik} est un effet aléatoire spécifique pour chaque marqueur et les erreurs de mesures e_{ijk} sont supposées gaussiennes et indépendantes. Dans ce modèle, chaque marqueur est défini par une combinaison linéaire (β_{0k}, β_{1k}) qui lui est propre et permettant de modéliser son niveau et son échelle, ce qui autorise en conséquence U_{ij} à ne pas avoir de dimension.

Le modèle mixte multivarié peut être utilisé pour caractériser l'évolution globale de cette quantité latente. Toutefois, l'une de ses principales limites est de considérer que les

marqueurs ont la même unité et la même échelle. Le modèle de Roy et Lin (2000) permet de modéliser l'évolution d'une quantité latente en étudiant conjointement des mesures répétées bruitées de cette quantité non-observée ainsi que de s'affranchir des problèmes d'échelle et d'unité des différents marqueurs. Le modèle multivarié permet d'étudier l'effet de covariables sur l'évolution globale de la cognition, mais cela suppose que le paramètre caractérisant l'impact d'une covariable soit commun aux modèles mixtes des différents marqueurs. Les modèles à variables latentes permettent de relâche l'hypothèse d'un effet de covariables commun aux marqueurs. Dans le modèle de Roy et Lin, l'effet global d'une variable explicative est directement obtenu au travers le paramètre α du modèle mixte pour la variable latente. En revanche, l'une des limites les plus importantes du modèle défini par Roy et Lin (2000) est de ne traiter que des marqueurs quantitatifs gaussiens.

Modèle non-linéaire à processus latent pour données longitudinales multivariées

Plus récemment, Proust-Lima et al. (2006) ont étendu l'approche de Roy et Lin (2000) afin de pouvoir étudier plusieurs marqueurs quantitatifs non gaussiens. Le modèle proposé est un modèle non linéaire à processus latent combinant un modèle linéaire mixte pour le processus latent et un modèle non linéaire d'observation pour lier les marqueurs au processus latent. Le processus latent est noté $\Lambda_i(t)$ pour l'individu i, $i = 1, ..., N$, au temps t. La forme générale du modèle est définie à l'aide d'un modèle linéaire mixte :

$$\Lambda_i(t) = X_{1i}(t)'\beta + Z_i(t)'u_i + \sigma_\omega \omega_i(t), \quad t \geq 0$$

où X_{1i} est un p_1-vecteur de variables explicatives potentiellement dépendant du temps. Le $(q+1)$-vecteur $Z_i(t) = (1, t, ..., t^q)$ est un polynôme du temps de degré q, sous-vecteur de $X_{1i}(t)$ et u_i est un vecteur d'effets aléatoires : $u_i \sim \mathcal{N}(0, G)$ avec G une matrice non structurée définie positive, le processus $\omega_i = (\omega_i(t))_{t>0}$ est un mouvement brownien standard permettant des variations locales et un éloignement de la tendance polynomiale globale. Le modèle linéaire mixte prend en considération la corrélation des mesures répétées. Le modèle ne possède pas d'erreur indépendante puisqu'il est supposé représenter la cognition réelle en temps continu.

Pour le sujet i, $Y_{ik} = (Y_{i1k}, ..., Y_{ijk}, ..., Y_{in_{ik}k})'$ est le vecteur des mesures du marqueur k aux temps t_{ijk}, $j = 1, ..., n_i$, $k = 1, ..., K$, $i = 1, ..., N$. Les temps de mesures t_{ijk} peuvent

être différents d'un sujet à l'autre et d'un marqueur à l'autre. Le score Y_{ijk} est relié au processus latent par le modèle suivant :

$$g(Y_{ijk}; \eta_k) = \Lambda_i(t_{ijk}) + \alpha_{ik} + X_{2i}(t_{ijk})'\gamma_k + \epsilon_{ijk}$$

où g est une transformation non linéaire monotone croissante dépendant des paramètres η_k dont les paramètres sont estimés. La fonction de répartition Béta a été retenue pour g car elle est souple et ne dépend que de deux paramètres. Pour un même niveau du processus latent, l'effet aléatoire $\alpha_{ik} \sim \mathcal{N}(0, \sigma^2_{\alpha_k})$ permet une variation intra-individuelle entre les marqueurs. Cela permet une variabilité entre les marqueurs conditionnellement aux valeurs du processus latent. Le vecteur X_{2i} est un vecteur de covariables associés au processus latent Λ. Les erreurs gaussiennes ϵ_{ijk} sont indépendantes de moyenne 0 et de variance σ_{ϵ_k}.

Les principales caractéristiques du modèle développés par Proust-Lima et al. (2006) sont de pouvoir :

- tenir compte de données multivariées non-gaussiennes,
- étudier l'évolution d'un processus latent en temps continu représentant le facteur commun étudié au travers de plusieurs marqueurs,
- évaluer la forme de la transformation liant les marqueurs quantitatifs au processus latent,
- tenir compte de données non équilibrées (nombres et temps de mesures, covariables,...),
- distinguer l'effet de variables explicatives sur l'évolution du processus latent et sur l'évolution des marqueurs observés.
- étudier les propriétés métrologiques des tests psychométriques telles que la sensibilité (caractérisée par curvilinéarité de la transformation non linéaire) et les effets seuils et plafonds des tests psychométriques.

Ce modèle a servi de base au développement d'un modèle conjoint à classes latentes pour données longitudinales multivariées (cf. section 2.2.2) et nous avons étudié l'intérêt de ce modèle pour la prise en compte des sorties d'étude (cf. chapitre 3).

2.2.4 Modèles d'évolution non-linéaire

L'une des hypothèses majeures du modèle linéaire mixte est de supposer que l'évolution du marqueur est de forme linéaire. Le modèle linéaire mixte a été étendu pour modéliser des évolutions plus souples ayant des formes non linéaires. De manière générale, le vecteur des mesures Y_i peut s'écrire ainsi :

$$Y_i = f(X_i, \beta, Z_i, u_i) + \epsilon_i$$

La fonction f peut être définie de manière paramétrique ou non-paramétrique et ϵ_i caractérise les erreurs de mesure.

Dans une approche paramétrique, les fonctions polynomiales offrent un vaste panel de formes de courbes permettant de définir la fonction f pour caractériser l'évolution de Y. Cox (1977) et Seber et Wild (1989) ont décrit en grand nombre de relations paramétriques pouvant être employées. Les modèles dynamiques à base d'équations différentielles sont également de plus en plus utilisés notamment en pharmacocinétique (Gibaldi et Perrier, 1982) ou dans le contexte de l'infection par le VIH (Perelson et al., 1996). Davidian et Giltinan (2003) proposent un état des lieux de ces méthodes dans le contexte de données répétées. L'idée sous-jacente est de définir un modèle d'équations différentielles caractérisant un modèle biologique et ainsi définir la fonction f liant les prédicteurs, ayant une interprétation biologique, à la variable observée (Perelson et al., 1996).

Dans une approche non-paramétrique, les fonctions B-splines sont des fonctions polynomiales par morceaux à support compact qui sont combinées linéairement pour approcher une fonction sur un intervalle. De Boor (1978) décrit leur méthodologie et décline des variations autour des B-splines également présentées par Silverman (1985) ou Ramsay (1988). Les fonctions splines ont largement été utilisées pour décrire l'évolution de mesures répétées (Zhang et al., 1998; Verbyla et al., 1999; Jacqmin-Gadda et al., 2002). Ces fonctions ont l'avantage de fournir une estimation lisse de la fonction à étudier.

Modèle à changement de pente

Parmi les différents modèles d'évolution non-linéaire, nous nous attachons à présenter dans ce paragraphe les modèles à changement de pente, approche utilisée et étendue dans ce travail. Les modèles à changement de pente sont des modèles d'évolution divisée en

plusieurs phases. Soit $Y_i(t)$ la valeur du marqueur du sujet i au temps t. On note le marqueur $Y_{ij} = Y_i(t_{ij})$ pour $j = 1, ..., n_i$, et $i = 1, ..., N$ où N est le nombre de sujets, t_{ij} est le temps à la mesure j du sujet i et Y_i est le vecteur des n_i mesures du sujet i. Les développements qui vont suivre correspondent à une évolution du marqueur Y_{ij} définie à l'aide d'un modèle mixte à effets aléatoires segmenté en 2 parties et supposant une évolution linéaire avant et après le changement de pente τ. On note $f_{Y_i|\tau}$ la densité du modèle d'évolution du marqueur, fonction du temps de changement de pente τ. $f_{Y_i|\tau}$ est une densité multivariée gaussienne avec pour moyenne $E_i = E(Y_i|\tau)$ et pour variance $V_i = Var(Y_i|\tau)$. Le modèle linéaire à changement de pente en deux phases à tendances linéaires s'exprime de la manière suivante :

$$Y_{ij} = \begin{cases} b_{01i} + b_{11i}t_{ij} + \epsilon_{ij} & \text{si } t_{ij} \leq \tau, \\ b_{02i} + b_{12i}t_{ij} + \epsilon_{ij} & \text{si } t_{ij} > \tau. \end{cases}$$

avec $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2_\epsilon)$ et $\epsilon_{ij} \perp \epsilon_{ij'} \forall\ j, j'$.

– Continuité du modèle

Très souvent le marqueur a une évolution que l'on suppose continue au cours du temps, et ce notamment au temps de changement de phase τ donc $b_{01i} + b_{11i}\tau = b_{02i} + b_{12i}\tau$, d'où $b_{20i} = b_{01i} + b_{11i}\tau - b_{12i}\tau$. En réutilisant la formulation de b_{20i}, le modèle peut se réécrire ainsi :

$$Y_{ij} = \begin{cases} b_{01i} + b_{11i}t_{ij} + \epsilon_{ij} & \text{si } t_{ij} \leq \tau, \\ b_{01i} + b_{11i}\tau_i + b_{12i}(t_{ij} - \tau_i) + \epsilon_{ij} & \text{si } t_{ij} > \tau. \end{cases}$$

où $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$.

– Reparamétrisation du modèle

Bacon et Watts (1971) ont noté que cette formulation n'était pas sensible pour la détection de changement de pente et suggère la reparamétrisation suivante qui, de plus, permet aisément d'introduire un lissage de la transition entre les deux phases :

$$\beta_{0i} = b_{01i} + b_{11i}\tau\ , \quad \beta_{1i} = \frac{b_{11i} + b_{12i}}{2}\ , \quad \beta_{2i} = \frac{b_{12i} - b_{11i}}{2}$$

Avec cette reparamétrisation, le paramètre β_{0i} correspond à la valeur moyenne du marqueur au changement de pente, β_{1i} est la moyenne des 2 pentes linaires et β_{2i} est égale à la moitié de la différence des 2 pentes.

Les trajectoires d'évolution individuelle sont spécifiées comme étant la somme d'une trajectoire moyenne et d'une spécificité de l'individu (mesurée comme un écart à la moyenne par un effet aléatoire), on note alors $\beta_{ki} = \mu_k + u_{ki}, k = 0, 1, 2$. Les paramètres du modèle peuvent également dépendre de covariables fixes : $\mu_k = \phi_k + X_{1ki}\alpha_k$ où X_{1ki} est un vecteur de variables explicatives.

Avec la paramétrisation de Bacon et Watts (1971), l'évolution du marqueur Y_{ij} devient alors :

$$
\begin{aligned}
Y_{ij} &= \begin{cases} b_{01i} + b_{11i}t_{ij} + \epsilon_{ij} & \text{si } t_{ij} \leq \tau, \\ b_{01i} + b_{11i}\tau + b_{12i}(t_{ij} - \tau) + \epsilon_{ij} & \text{si } t_{ij} > \tau. \end{cases} \\
&= \begin{cases} b_{01i} + b_{11i}\tau + b_{11i}t_{ij} - b_{11i}\tau + \epsilon_{ij} & \text{si } t_{ij} \leq \tau, \\ b_{01i} + b_{11i}\tau + b_{12i}t_{ij} - b_{12i}\tau + \epsilon_{ij} & \text{si } t_{ij} > \tau. \end{cases} \\
&= \begin{cases} \beta_{0i} + (\beta_{1i} - \beta_{2i})t_{ij} - (\beta_{1i} - \beta_{2i})\tau + \epsilon_{ij} \\ \beta_{0i} + (\beta_{1i} + \beta_{2i})t_{ij} - (\beta_{1i} + \beta_{2i})\tau + \epsilon_{ij} \end{cases} \\
&= \begin{cases} \beta_{0i} + \beta_{1i}(t_{ij} - \tau) - \beta_{2i}(t_{ij} - \tau) + \epsilon_{ij} \\ \beta_{0i} + \beta_{1i}(t_{ij} - \tau) + \beta_{2i}(t_{ij} - \tau) + \epsilon_{ij} \end{cases} \\
&= \beta_{0i} + \beta_{1i}(t_{ij} - \tau) + \beta_{2i}(t_{ij} - \tau)sgn(t_{ij} - \tau) + \epsilon_{ij}
\end{aligned}
$$

où

$$
sgn(z) = \begin{cases} -1 & z < 0, \\ 0 & z = 0, \\ +1 & z > 0. \end{cases}
$$

et $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$.

– Lissage pour la transition entre les deux phases

Seber et Wild (1989) suggèrent une formulation de la même structure mais une transition lisse entre les deux phases. Ils proposent cela en remplaçant la fonction $sgn(.)$ par

une fonction de transition $trn(.)$ ayant les propriétés suivantes :

$$\begin{aligned} &(i) \quad \lim_{z\to\pm\infty}[z\ trn(z) - |z|] = 0 \\ &(ii) \quad \lim_{\gamma\to 0} trn(z/\gamma) = sgn(z) \\ &(iii) \quad trn(0) = sgn(0) = 0 \end{aligned} \tag{2.2}$$

Différentes familles de fonctions peuvent convenir :

$$trn(z/\gamma) = tanh(z/\gamma) = \frac{e^{z/\gamma} - e^{-z/\gamma}}{e^{z/\gamma} + e^{-z/\gamma}} \tag{2.3}$$

$$trn(z/\gamma) = hyp(z;\gamma) = \frac{1}{z}\sqrt{z^2+\gamma} \tag{2.4}$$

La fonction 2.3 répond à l'ensemble des 3 critères requis pour la fonction de transition 2.2 alors que la fonction 2.4 ne remplit que les 2 premières conditions. La première formulation fournit une fonction lissée passant par le point d'intersection des deux phases linéaires alors que ce n'est pas le cas pour la seconde.

En pratique, nous avons $z = t_{ij} - \tau$. Ainsi, on modélise l'évolution d'un marqueur d'après un modèle à changement de pente avec un lissage entre les deux phases :

$$E(Y_{ij}) = \beta_{0i} + \beta_{1i}(t_{ij} - \tau) + \beta_{2i}(t_{ij} - \tau)trn(t_{ij} - \tau;\gamma) \tag{2.5}$$

La matrice de Variance-Covariance de la densité multivariée gaussienne de Y_i sachant τ se décompose de la manière suivante :

$$V_i = V(Y_i) = A_i G A_i' + \sigma_\epsilon^2 I_{n_i} \tag{2.6}$$

où A_i est une matrice $n_i \times 3$ avec en ligne : $[1, (t_{ij}-\tau), (t_{ij}-\tau)\times trn(t_{ij}-\tau)]$ et $G = Var(\beta_i)$.

– Remarques

Le modèle présenté est un modèle linéaire à changement de pente en deux phases à tendances linéaires. D'autres tendances d'évolution peuvent être envisagées. Nous avons par exemple vu que toutes les formulations précédemment introduites étaient valables avec une seconde phase d'évolution quadratique.

Un assouplissement supplémentaire pourrait être d'envisager une évolution en plusieurs phases. Griffiths et Miller (1975) ainsi que Bunke et Schulze (1985) proposent une

formulation du modèle à changement de pente pour D phases d'évolution non-linéaires.

$$E(Y_{ij}) = \frac{1}{2}\left(f_1(t_{ij};\theta_1) + f_D(t_{ij};\theta_2) + \sum_{j=2}^{D}[f_j(t_{ij};\theta_j) - f_{j-1}(t_{ij};\theta_{j-1})]sgn(t_{ij} - \tau_{j-1};\gamma)\right)$$

où τ_j est le temps du j-ième changement de pente, $j = 1, ..., D$.

Leur formulation permet d'intégrer un lissage entre les deux phases d'évolution. Il est également possible de retrouver la structure de la formulation (2.5) à partir de la paramétrisation d'origine pour un modèle d'évolution en 2 phases :

$$E(Y_{ij}) = \frac{1}{2}(f_1(t_{ij}) + f_2(t_{ij}) + [f_2(t_{ij}) - f_1(t_{ij})]trn(t_{ij} - \tau;\gamma))$$

où $f_1(t_{ij}) = b_{01i} + b_{11i}t_{ij}$ si $t_{ij} \leq \tau$

et $f_2(t_{ij}) = b_{01i} + b_{11i}\tau + b_{12i}(t_{ij} - \tau)$ si $t_{ij} > \tau$

Ainsi,

$$E(Y_{ij}) = b_{01i} + b_{11i}\tau + \frac{b_{11i} + b_{12i}}{2}(t_{ij} - \tau) + \frac{b_{12i} - b_{11i}}{2}(t_{ij} - \tau)trn(t_{ij} - \tau;\gamma)$$

On remarque :

$$\begin{cases} b_{01i} + b_{11i}\tau = \beta_{0i} \\ b_{11i} = \beta_{1i} - \beta_{2i} \\ b_{12i} = \beta_{1i} + \beta_{2i} \end{cases} \iff \begin{cases} \beta_{0i} = b_{01i} + b_{11i}\tau \\ \beta_{1i} = \frac{b_{11i}+b_{12i}}{2} \\ \beta_{2i} = \frac{b_{12i}-b_{11i}}{2} \end{cases}$$

On retrouve donc la formulation de l'équation (2.5).

– Formulation de τ

Le modèle précédemment introduit peut être utilisé pour caractériser une évolution pour laquelle la date de changement de pente τ est connue et est identique pour l'ensemble de la population (Thiébaut et al., 2005). Le modèle reste donc linéaire pour les paramètres à estimer. Le changement de pente peut également être inconnu, mais identique pour l'ensemble des sujets, τ est alors un paramètre du modèle qui devient non linéaire (Hall et al., 2003). Enfin, le modèle le plus souple, qui nous intéresse, considère le temps au changement de pente comme étant inconnu et spécifique à chaque individu (Jacqmin-Gadda et al., 2006; Dominicus et al., 2008). Le modèle considéré est donc un modèle à changement de pente aléatoire où le temps au changement de pente τ_i est un effet aléatoire individuel pour lequel il est nécessaire de spécifier une distribution *a priori*.

2.2.5 Données manquantes au cours du suivi longitudinal

Dans toutes les études longitudinales, les données manquantes constituent une limite aux analyses effectuées. Les cohortes prospectives de personnes âgées, notamment celles portant sur le vieillissement cérébral, comportent un nombre important de données manquantes et de sorties d'étude. Lorsque les données sont ignorables, les paramètres de l'évolution de la cognition peuvent être estimés sans biais. Mais les sorties d'étude et les décès sont associés au déclin cognitif. Si la probabilité d'observation dépend des valeurs non observées de la cognition, les données manquantes sont informatives et ne peuvent pas être ignorées. Les méthodes d'analyse doivent tenir compte des mécanismes d'observation de la variable étudiée pour éviter des biais dans les estimations.

Classification des données manquantes

Trois types de données manquantes ont été définis par Little et Rubin (2002) pour lesquels il est nécessaire d'introduire quelques notations. Soit Y_i le vecteur complet des mesures du sujet i, $i = 1, ..., N$. La distribution de Y_i dépend du vecteur de paramètres ψ. Soit $R(t)$ la variable indicatrice d'observation de $Y(t)$ dont la distribution dépend du paramètre ϕ : $R(t) = 1$ si $Y(t)$ est observé et 0 sinon, avec $R_i = \big(R_i(t_{ij})\big)_{j=1,...,n_i}$. Le vecteur Y_i, peut être partitionné en deux jeux (Y_{io}, Y_{im}) où Y_{io} correspond aux données observées et Y_{im} à celles manquantes pour un sujet i.

- Les données sont manquantes complètement aléatoirement (MCAR pour Missing Completely At Random) si la probabilité qu'une donnée soit manquante est indépendante du processus d'intérêt qu'il soit observé ou non.

$$P(R_i = r_i|Y_i, \phi) = P(R_i = r_i|\phi)$$

- Les données sont manquantes aléatoirement (MAR pour Missing At Random) si la probabilité qu'une donnée soit manquante est indépendante des variables non observées du processus d'intérêt, conditionnellement aux variables observées.

$$P(R_i = r_i|Y_i, \phi) = P(R_i = r_i|Y_{io}, \phi)$$

- Enfin, les données sont manquantes non aléatoirement ou informatives (MNAR pour

Missing Not At Random) si la probabilité qu'une donnée soit manquante est dépendante des réponses non observées.

$$P(R_i = r_i|Y_i, \phi) = P(R_i = r_i|Y_{io}, Y_{im}, \phi)$$

La classification définie par Little et Rubin (2002) concerne aussi bien les données manquantes intermittentes que les sorties d'étude. Dans le cadre de ce travail, l'étude des données incomplètes porte essentiellement sur l'analyse des sorties d'étude. Dans l'étude du déclin cognitif, la typologie des données manquantes pour les sorties d'étude peut alors s'interpréter intuitivement de la manière suivante :

- les sorties d'étude sont MCAR si la probabilité qu'un sujet effectue un test psychométrique à une visite au temps t ne dépend pas des scores obtenus avant cette visite ni du score qu'il aurait obtenu à cette visite,
- les sorties d'étude sont MAR si la probabilité qu'un sujet effectue un test psychométrique à une visite au temps t ne dépend pas du score cognitif qu'il aurait obtenu à la visite au temps t mais exclusivement des scores obtenus aux visites précédentes,
- enfin, les sorties d'étude sont informatives ou MNAR si la probabilité qu'un sujet effectue un test psychométrique à une visite au temps t est dépendante du score non observé à cette même visite.

La distinction entre données manquantes aléatoirement (MCAR et MAR) et non aléatoirement (MNAR) est très importante puisque l'inférence statistique par maximum de vraisemblance n'est pas biaisée sous l'hypothèse de données MCAR ou MAR (Rubin, 1976; Kenward et Molenberghs, 1998). En effet, l'objectif de l'analyse est d'estimer sans biais les paramètres ψ de la distribution de Y, à partir des données observées. La vraisemblance des données observées Y_o et R peut s'écrire :

$$L(\psi, \phi|y_o, r) = f(y_o, r|\psi, \phi) = \int f(y_o, y_m|\psi) P(r|y_o, y_m, \phi) dy_m \tag{2.7}$$

Or, si le mécanisme d'observation est MAR, $P(r|y_o, y_m, \phi) = P(r|y_o, \phi)$ et on obtient donc :

$$\begin{aligned} L(\phi, \psi|y_o, r) &= P(r|y_o, \phi) \int f(y_o, y_m|\psi) dy_m \\ &= P(r|y_o, \phi) f(y_o|\psi) \end{aligned}$$

En présence de données MAR, la distribution conjointe des données observées se décompose de la manière suivante :

$$f(y_{io}, r_i) = f(y_{io}; \psi) P(r_i | y_{io}; \phi)$$

Si, de plus, les paramètres du modèle d'observation ϕ et ceux du modèle pour le marqueur ψ sont distincts, les paramètres d'intérêt ψ peuvent donc être estimés sans biais en maximisant la vraisemblance des mesures observées $f(y_o; \phi)$ (Little, 1995; Verbeke et Molenberghs, 2000), les données sont dites ignorables. En revanche la présence de données MNAR, également appelées données manquantes informatives, doit être intégrée dans le développement de la vraisemblance (cf. equation 2.7).

L'estimation sans biais de ϕ requiert l'estimation conjointe de $f(y_o, r)$ et donc des hypothèses paramétriques sur le processus de données manquantes. Différentes approches ont été proposées pour l'analyse de données longitudinales comportant des données manquantes informatives. Comme elles nécessitent toutes des hypothèses invérifiables sur le lien entre processus de données manquantes et processus d'intérêt, la recommandation actuelle est d'effectuer une analyse de sensibilité des résultats obtenus sous l'hypothèse de données manquantes MNAR (Verbeke et al., 2001; Thijs et al., 2002) et de comparer les résultats à ceux obtenus sous l'hypothèse de données manquantes MAR.

Dans la littérature statistique, deux grands types d'approche sont évoqués pour traiter les sorties d'étude informatives : les modèles par mélange de schémas d'observation (Pattern Mixture Model ou PMM) et les modèles de sélection. Il s'agit de corriger les biais dans l'inférence de l'évolution du marqueur Y en considérant le délai jusqu'à la sortie d'étude T conjointement au processus d'intérêt. Les deux approches diffèrent par leur factorisation de la densité conjointe du marqueur et du délai jusqu'à la sortie d'étude :

- Pattern Mixture Model : $f(Y, T) = f(Y|T) f(T)$
- Modèle de sélection : $f(Y, T) = f(Y) f(T|Y)$

Pattern Mixture Model

L'approche par PMM consiste en une modélisation de l'évolution du processus d'intérêt conditionnellement aux profils de sortie d'étude r (Little, 1993; Verbeke et Molenberghs, 2000), c'est-à-dire conditionnellement aux dates de sortie d'étude. L'idée est de considérer une hétérogénéité de l'évolution du processus d'intérêt de la population qui soit liée au

processus de sortie d'étude. La distribution conjointe de Y_i et T_i se décompose alors ainsi :

$$f(Y_i, T_i; \beta, \zeta) = f(Y_i|T_i; \beta) f(T_i; \zeta)$$

Il est alors possible de définir $f(Y_i|T_i; \beta)$ soit en stratifiant sur les profils de sortie d'étude, soit en incluant la date de sortie d'étude comme variable explicative dans le modèle :

– L'analyse stratifiée est effectuée en estimant un modèle à partir de chacun des sous-échantillons définis par la date de sortie d'étude. Ensuite, les coefficients moyens $\widetilde{\beta}$ du modèle longitudinal sont calculés par les moyennes des coefficients estimés sur chaque groupe de sorties pondérées par les probabilités π_r d'appartenance aux groupes de sorties d'étude ($r = 1, ..., R$).

$$\pi_r = \frac{n_r}{n} \qquad r = 1, ..., R$$

où n_r est le nombre de sujets dans le groupe de sortie d'étude r et n est le nombre de sujets total.

Les coefficients moyens s'obtiennent d'après la formule suivante :

$$\widetilde{\beta} = \sum_{r=1}^{R} \hat{\pi}_r \hat{\beta}_r = f(\hat{\theta})$$

où θ inclut tous les paramètres du modèle pour données longitudinales multivariées spécifiques aux strates β_r et les proportions de chaque strate π_r. La variance s'obtient par la "delta-method" :

$$Var(\widetilde{\beta}) = \left(\frac{\delta f}{\delta \theta}\right)^T Var(\hat{\theta}) \left(\frac{\delta f}{\delta \theta}\right)$$

– La seconde approche des modèles par mélange de schéma d'observation est d'intégrer le profil d'observation comme une variable explicative du modèle d'évolution. Il est ainsi possible de regarder spécifiquement l'effet du profil de sortie d'étude sur le processus d'évolution. Le profil de sortie d'étude peut être considéré comme une variable explicative quantitative ou bien comme une variable explicative qualitative à plusieurs modalités. Elle peut donc être injectée dans le modèle sous la forme de variables indicatrices.

Les Pattern Mixture Models sont des modèles de mélange dont les classes d'évolution sont observées et définies à partir des dates de sortie d'étude. L'un des principaux avantages de l'approche par PMM est de proposer une formulation analytique de la vraisemblance des données observées (Y_o, R). Il est donc possible d'utiliser des méthodes classiques d'estimation par maximum de vraisemblance. En revanche, les Pattern Mixture Model ne fournissent pas un cadre d'estimation de l'évolution marginale du processus d'intérêt. Les paramètres obtenus définissent l'évolution conditionnellement aux profils de sorties d'étude. Fitzmaurice et al. (2001) ont proposé une reparamétrisation du modèle pour distinguer la partie marginale de l'évolution du marqueur des différences d'évolution suivant les profils des sorties d'étude. Cela permet d'obtenir directement l'évolution marginale sans biais du marqueur comme dans le modèle de sélection.

Un autre problème majeur dans l'approche par PMM est le nombre élevé de profils de sorties d'étude : le nombre de paramètres à estimer est d'autant plus important. Pour certains profils de sortie d'étude, le nombre de sujets ou le nombre de mesures de Y_i peut être insuffisant pour estimer tous les paramètres du modèle et conduire à des problèmes d'identifiabilité. Dans ce cas, des contraintes supplémentaires sont nécessaires comme la définition de relation paramétrique entre des paramètres associés à des profils d'évolution différents ou bien le regroupement de certains profils d'évolution (Verbeke et Molenberghs, 2000; Thijs et al., 2002).

Les Pattern Mixture Models constituent une approche permettant une prise en compte simple et intuitive du processus de sortie d'étude en considérant une hétérogénéité de la population face au phénomème de sortie d'étude. Pour sa mise en oeuvre, certaines hypothèses concernant le processus de sortie d'étude doivent être faites. Ces hypothèses souvent restrictives ont le mérite d'être clairement identifiées et simples, ce qui rend l'approche par PMM attrayante dans l'optique d'une analyse de sensibilité aux sorties d'étude. En effet, quelle que soit l'approche méthodologique (PMM ou modèle de sélection), des hypothèses paramétriques invérifiables sont faites sur le processus de sorties d'étude. Dans l'approche PMM, les hypothèses sont clairement identifiées et facilement modifiables pour réaliser une analyse de sensibilité.

Modèle de sélection

Le principe de l'approche par modèle de sélection est de modéliser conjointement l'évolution de la variable d'intérêt Y_i et la survenue de la sortie d'étude sachant l'évolution de Y_i. Le modèle de sortie d'étude est spécifié conditionnellement aux valeurs observées et non-observées de la variable réponse (Diggle et Kenward, 1994). La distribution conjointe de Y_i et T_i se décompose alors ainsi :

$$f(Y_i, T_i; \beta, \zeta) = f(Y_i|\beta) f(T_i|Y_i; \zeta)$$

La dépendance entre Y_i et T_i peut également être exprimée au travers des effets aléatoires du modèle d'évolution (Wulfsohn et Tsiatis, 1997; Henderson et al., 2000). La distribution conjointe de Y_i et T_i est alors définie ainsi :

$$f(Y_i, T_i; \beta, \zeta) = \int f(Y_i|u_i, \beta) f(T_i|u_i; \zeta) f(u_i) du_i$$

La formulation des modèles de sélection s'apparente entièrement à celle des modèles conjoints et sera développée plus précisément dans la section 2.3.

Contrairement à l'approche par PMM, l'évolution du processus Y_i est définie marginalement aux sorties d'étude ; on estime donc directement les paramètres d'intérêt. Cependant cette approche est sensible aux hypothèses paramétriques effectuées, notamment en ce qui concerne la distribution de Y_i et la distribution de T_i sachant Y_i.

2.2.6 Application au vieillissement cognitif

Les modèles présentés pour l'analyse de données longitudinales sont des outils intéressants dans la prise en compte les problèmes méthodologiques rencontrés dans le contexte du vieillissement cognitif des personnes âgées. L'évolution cognitive est caractérisée par un ou plusieurs tests psychométriques (cf. section 1.2.1). Les mesures répétées d'un de ces tests sont une mesure bruitée de la cognition qui peut être modélisée à l'aide des modèles à effets aléatoires permettant de prendre en considération la corrélation temporelle de ces mesures.

Par exemple, Jacqmin-Gadda et al. (1997a) ont étudié une transformation non-linéaire des mesures répétées du test psychométrique du MMSE à l'aide d'un modèle linéaire mixte sur les 5 premières années d'étude de la cohorte Paquid pour les sujets non-déments. Cette

transformation permet de s'affranchir de l'hypothèse de normalité faite dans les modèles linéaires mixtes.

Plusieurs travaux ont également porté sur l'utilisation de modèles linéaires mixtes pour données multivariées dans le contexte du vieillissement cognitif des personnes âgées (Sliwinski et al., 2003b). Par exemple, Harvey et al. (2003) développent une approche par modèle longitudinal mixte pour données multivariées appliquée à deux marqueurs psychométriques évaluant la mémoire et les fonctions exécutives. Hall et al. (2001) ont développé un modèle polynomial pour données bivariées avec changement de pente. Les résultats issus de ce modèle, estimé par une méthode bayésienne, permettent de confirmer que le déclin commence plusieurs années avant le diagnostic de démence.

Proust et Jacqmin-Gadda (2005) étendent le travail de Jacqmin-Gadda en étudiant l'évolution du score au MMSE par un modèle mixte polynomial de terme quadratique. Ce modèle permet de modéliser des formes d'évolution non linéaire. Son approche suppose une hétérogénéité des évolutions cognitives de la population modélisée à l'aide d'un mélange de distribution gaussienne. Ce modèle illustre l'intérêt des modèles de mélange dans l'étude du vieillissement cognitif. Proust montre également l'intérêt de l'utilisation d'un algorithme de type Newton-Raphson qui permet un gain de temps de calcul et une amélioration de la qualité des estimations par rapport à l'algorithme EM classiquement utilisé dans les modèles de mélange.

Proust-Lima et al. (2006) proposent un modèle à processus latent pour l'évolution de plusieurs tests psychométriques quantitatifs non gaussiens (MMSE, BVRT, IST, DSSTW). Ce modèle est formellement défini en section 2.2.3. Il combine une dimension de modélisation multivariée puisqu'il permet d'étudier simultanément plusieurs marqueurs. Il relaxe l'hypothèse de normalité sous-jacente au modèle mixte puisqu'une transformation non-linéaire relie les tests psychométriques à un processus latent gaussien, l'évolution de ce processus étant modélisée par un modèle mixte. L'application de ce travail porte sur l'étude de l'évolution de la cognition latente, facteur commun à plusieurs tests. Ce modèle permet de prendre en compte les propriétés métrologiques différentes des tests psychométriques et d'étudier leur sensibilité pour la détection de changement cognitif (Proust-Lima et al., 2007a). Un travail appliqué (Proust-Lima et al., 2008) a également porté sur l'effet de covariables telles que le sexe et le niveau d'éducation d'un sujet sur le déclin cognitif

et a permis de distinguer leurs effets sur les tests psychométriques eux-mêmes.

Plusieurs travaux ont porté sur les modèles mixtes à changement de pente aléatoire (Hall et al., 2000, 2003) pour étudier l'évolution non linéaire du déclin cognitif (Amieva et al., 2005, 2008). Le changement de pente aléatoire identifié, spécifique au sujet, correspond à l'âge à l'accélération de ce déclin. En utilisant une méthode d'estimation bayésienne, Hall et al. (2003) étudient l'évolution du test de Buschke (1973) et trouvent que l'accélération du déclin cognitif des sujets déments survient entre 5 et 8 ans avant le diagnostic de démence. Ils notent cependant que l'âge à l'accélération du déclin peut différer selon le test psychométrique considéré. Dominicus et al. (2008) montrent l'apport des modèles à changement de pente aléatoire dans le contexte du vieillissement cognitif en le comparant à des modèles mixtes classiques du type linéaire ou quadratique.

Une des limites majeures de ces modèles est leur spécification entièrement paramétrique. Des approches à base de splines ont été envisagées comme une alternative pour étudier l'évolution cognitive. Par exemple, Jacqmin-Gadda et al. (2002) proposent un modèle où l'évolution cognitive est une fonction du temps approximée sur une base de splines et estimée par vraisemblance pénalisée.

Une seconde limite porte sur l'existence d'un biais de sélection. En effet, la comparaison des évolutions cognitives normales et pathologiques est réalisée à partir d'échantillons constitués respectivement de sujets non déments au cours du suivi ou déments à une date donnée. Les modèles conjoints apparaissent comme une solution très intéressante pour pallier à ce problème.

2.3 Modèles conjoints

2.3.1 Définition et objectifs

Dans les études de cohortes, l'état de santé d'un sujet est très souvent suivi à l'aide de marqueurs longitudinaux. Le délai de survenue d'un événement (maladie, décès, sortie d'étude) est également d'un intérêt majeur. L'état de santé du patient étant fortement associé au risque de survenue d'un événement clinique d'intérêt, il est nécessaire de les étudier conjointement. L'objectif général des modèles conjoints est de rendre compte du

comportement conjoint de l'évolution d'un marqueur longitudinal quantitatif Y_i et du temps de survenue d'un événement T_i en considérant leur densité conjointe $f(Y_i, T_i)$; cela permet :

- d'étudier l'association entre l'évolution du marqueur quantitatif et la survenue de l'événement,
- de prédire le risque de survenue d'un événement en fonction de l'évolution du marqueur,
- d'étudier un processus quantitatif en limitant les biais liés aux sorties d'étude ; quand l'événement d'intérêt est la sortie d'étude, le modèle conjoint utilisé s'apparente à un modèle de sélection,
- et de comprendre les mécanismes d'implication de facteurs de risque sur l'évolution du processus et/ou sur le risque de survenue de l'événement.

2.3.2 Modèles à effets aléatoires partagés

Définition et notations

Le modèle à effets aléatoires partagés (ou shared random-effect model) est le modèle classiquement utilisé. Son principe est de supposer que les effets aléatoires caractérisant l'évolution d'un marqueur interviennent dans le modèle pour la survenue de l'événement (Wulfsohn et Tsiatis, 1997; Henderson et al., 2000). La densité conjointe de Y_i et T_i peut donc être développée ainsi :

$$f(Y_i, T_i) = \int f(Y_i|u_i) f(T_i|u_i) f(u_i) du_i$$

où u_i est un vecteur d'effets aléatoires définissant le lien entre marqueur et délai de survenue de l'événement.

L'évolution du marqueur Y_i est définie par un modèle linéaire mixte (cf. section 2.2.1). On note Y_{ij} la mesure du sujet i, $i = 1, ..., N$, au temps t_{ij}, $j = 1, ..., n_i$, avec $Y_i' = (Y_i j)_{j=1,...,n_i}$ définie comme suit :

$$Y_i = X_{1i}\beta + Z_i u_i + \epsilon_i$$

où X_{1i} et Z_i sont les matrices de variables explicatives respectivement associées aux effets fixes β et aux effets aléatoires gaussiens $u_i \sim \mathcal{N}(0, G)$.

Le risque de survenue d'un événement est classiquement étudié à l'aide d'un modèle des risques proportionnels (cf. section 2.1.1). On note T^* le délai de survenue de l'événement et $T = min(T^*, C)$ où C est le temps de censure et $\delta = \mathbb{1}_{\{T^*<C\}}$ est la variable indicatrice de l'événement. Le risque instantané d'événement est donc défini ainsi :

$$\alpha(t|X_{2i}; u_i) = \alpha^0(t)\exp(X_{2i}\lambda + g(u_i, t)\zeta)$$

où $\alpha^0(t)$ est le risque de base, X_{2i} est un vecteur de variables explicatives associé au vecteur de paramètres de régression λ et ζ est le vecteur de paramètres associant le marqueur et le risque de survenue de l'événement par l'intermédiaire des effets aléatoires communs. Il est à noter que si ζ est nul, le risque de survenue de l'événement est indépendant de l'évolution du marqueur. La fonction g caractérise la forme de dépendance sur les effets aléatoires. Il s'agit souvent d'une combinaison linéaire des effets aléatoires. Un exemple assez fréquent est d'utiliser la valeur de l'espérance du marqueur au temps courant t, $E(Y_i|u_i) = X_i\beta + Z_iu_i$ tel que le proposent Wulfsohn et Tsiatis (1997) ou encore Law et al. (2002). Une autre idée est de faire dépendre le risque de survenue de l'événement d'une caractéristique de l'évolution du marqueur telle que la pente d'évolution (Yu et al., 2008).

Estimation

Une hypothèse majeure est faite et permet d'écrire la distribution conjointe du marqueur et de l'événement : il s'agit de l'indépendance conditionnelle du marqueur et de l'événement sachant les effets aléatoires u.

La vraisemblance du modèle peut alors s'écrire :

$$L(\theta; Y, T, \delta) = \prod_{i=1}^{N} \int f(Y_i|u_i; \theta)\alpha(T_i|u_i; \theta)^{\delta} S(T_i|u_i; \theta) f(u_i; \theta) du_i \tag{2.8}$$

où $f(Y_i|u_i)$ est une densité multivariée normale d'espérance $E(Y_i|u_i) = X_i\beta + Z_iu_i$ et de variance $\sigma^2 I_{n_i}$. Certaines approches considèrent l'évolution du marqueur comme définie de manière non paramétrique, à l'aide de fonctions splines (Ding et Wang, 2008). Le risque $\alpha(t)$ peut être défini paramétriquement (Henderson et al., 2000) ou non paramétriquement (Wulfsohn et Tsiatis, 1997).

La vraisemblance des modèles à effets aléatoires partagés comporte des intégrales multiples sans solution analytique. La maximisation de la vraisemblance doit donc se faire en

approximant ces intégrales par calcul numérique. Une solution envisageable est l'approximation par quadrature de Gauss. Dans de nombreux travaux, l'algorithme classiquement utilisé est pour la maximisation de la vraisemblance est l'algorithme EM (Wulfsohn et Tsiatis, 1997; Henderson et al., 2000; Lin et al., 2002b). Certains auteurs ont proposé des approches bayésiennes (Xu et Zeger, 2001; Brown et al., 2005; Chi et Ibrahim, 2006).

Extensions

Des extensions de ce modèle ont concerné la modélisation de données longitudinales multivariées conjointement à la survenue d'un événement. Lin et al. (2002b) ont ainsi cherché à caractériser les interactions entre plusieurs marqueurs longitudinaux et l'événement ainsi que les interactions entre les marqueurs à l'aide d'un modèle conjoint pour la survenue d'un temps d'événement et des variables longitudinales multivariées. Le risque de survenue de l'événement est modélisé par un modèle à fragilité semi-paramétrique. Ce risque est également supposé dépendre des valeurs courantes des différents marqueurs ainsi que de leur interaction avec des variables explicatives. Dans le contexte de l'infection par le VIH, Brown et al. (2005) se sont intéressés à l'évolution conjointe de la charge virale et du taux de cellules CD4 en association avec le risque de survenue du stade SIDA ou du décès. Par ailleurs, ils proposent un assouplissement de la modélisation des marqueurs par une méthode non-paramétrique à base de splines. Thiébaut et al. (2005) ont développé un modèle linéaire mixte bivarié pour décrire l'évolution de la charge virale et des cellules CD4 en considérant conjointement le risque de sortie d'étude informative. Xu et Zeger (2001) ont proposé une modélisation conjointe de données longitudinales multivariées dans une approche à processus latent. Ils illustrent l'intérêt de prendre en considération l'information issue de plusieurs marqueurs longitudinaux d'un même processus pour évaluer le risque de survenue d'un évenement. Le travail de Henderson et al. (2002) asseoit d'ailleurs cette notion de validation de marqueurs longitudinaux comme marqueurs de substitution, ces marqueurs pouvant être utilisés pour prédire le risque de d'événement. Par ailleurs, Roy et Lin (2002) ont proposé une extension de leur approche à variables latentes pour étudier conjointement le risque de sortie d'étude.

Quelques travaux ont concerné la modélisation conjointe de plusieurs événements. Chi et Ibrahim (2006) modélisent conjointement plusieurs marqueurs d'évolution indépendants

entre eux et leur association avec deux événements corrélés. Elashoff et al. (2008) étendent le travail proposé par Henderson et al. (2000). L'originalité repose sur le fait qu'ils modélisent la survenue de risques compétitifs conjointement à l'évolution d'un marqueur.

Limites

L'une des limites des modèles de sélection réside dans les difficultés de calcul numérique de ces intégrales (Henderson et al., 2000; Roy et Lin, 2002) auxquelles sont associées des difficultés de convergence et qui sont accentuées lorsqu'on considère conjointement plusieurs marqueurs (cf. section 2.2.3). De plus, le lien entre le processus d'évolution et la survenue d'un événement se fait par l'intermédiaire d'effets aléatoires communs définis par une hypothèse de normalité. Cette association peut être difficile à interpréter.

L'approche par modèle de mélange développée par Muthén et Shedden (1999) fournit une alternative séduisante aux modèles à effets aléatoires partagés. Tout d'abord, comme les modèles de mélange classiques, les fonctions de vraisemblance ont des expressions analytiques. Le calcul du maximum de vraisemblance est donc plus simple. De plus, l'hypothèse de normalité des effets aléatoires est levée en supposant un mélange de distribution. Enfin, cette approche donne des outils d'interprétations plus intuitifs : différents profils de risques pour l'événement sont associés à différents profils d'évolution du marqueur.

2.3.3 Modèle conjoint à classes latentes

Développés plus récemment que les modèles à effets aléatoires partagés, les modèles conjoints à classes latentes ont tout d'abord servi à étudier l'association entre l'évolution d'un marqueur et la survenue d'un événement clinique. Par exemple, Lin et al. (2000) se sont intéressés aux profils d'évolution de la PSA (antigène spécifique de la prostate) et leur association avec la survenue d'un cancer de la prostate. Garre (2008) a étudié la survenue de rejet de greffe conjointement à l'évolution de marqueurs longitudinaux sanguins à l'aide d'un modèle à classes latentes avec changement de pente aléatoire. Les modèles à classes latentes ont aussi été appliquées en psychologie pour étudier des profils d'évolution (Muthén et Shedden, 1999).

Parallèlement à ces travaux, les modèles conjoints à classes latentes ont également été

utilisés dans l'étude de l'association entre le processus d'évolution d'un marqueur et le processus de données manquantes au cours du suivi (Roy, 2003; Lin et al., 2004; Beunckens et al., 2008). Nous reviendrons dans le chapitre 3 sur l'utilisation de cette approche pour traiter des sorties d'étude informatives.

Définition et notations

Le principe des modèles conjoints à classes latentes est de supposer une population hétérogène avec Q sous-populations (classes latentes) correspondant à Q profils d'évolution distincts auxquels sont associés des risques d'événement différents. La variable latente c_{iq} est égale à 1 si le sujet i appartient à la classe d'évolution q et 0 sinon. Alors que, dans le modèle à effets aléatoires partagés, le lien entre les distributions du marqueur et de l'événement se fait par l'intermédiaire d'effets aléatoires communs, ce lien est à présent caractérisé par l'appartenance à une classe latente. La distribution conjointe de Y_i et T_i est alors définie de la manière suivante :

$$f(Y_i, T_i) = \sum_{q=1}^{Q} f(Y_i|c_{iq} = 1) f(T_i|c_{iq} = 1)\pi_{iq}$$

L'appartenance à chacune des classes d'évolution est définie par un modèle logistique multinomial :

$$\pi_{iq} = P(c_{iq} = 1|X_{3i}) = \frac{e^{\xi_{0q} + X'_{3i}\xi_{1q}}}{\sum_{l=1}^{Q} e^{\xi_{0l} + X'_{3i}\xi_{1q}}} \text{ avec } \xi_{01} = 0 \text{ et } \xi_{11} = 0$$

où ξ_{oq} est l'intercept de la classe q et ξ_{1q} le vecteur de paramètres spécifiques à la classe q associé aux vecteurs de variables explicatives X_{3i}.

Classiquement, un modèle linéaire à effets aléatoires avec mélange caractérise l'évolution du marqueur Y_i du sujet i :

$$\begin{cases} Y_i|_{c_{iq}=1} = X_{1i}\beta_q + Z_i u_i + \epsilon_i \\ u_i \sim \mathcal{N}(\mu_q, \omega_q G) \\ \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2) \end{cases}$$

où X_{1i} et Z_i sont les vecteurs de variables explicatives respectivement associés aux effets fixes β et aux effets aléatoires u_i.

Enfin, le risque de survenue de l'événement peut être modélisé de différentes manières. Muthén et Shedden (1999) ou Lin et al. (2000) proposent un modèle logistique définissant

la probabilité de survenue de l'événement :

$$P(T_i = 1|c_{iq} = 1) = \frac{\exp(\gamma_{0q} + X_{2i}\gamma_{1q})}{1 + \exp(\gamma_{0q} + X_{2i}\gamma_{1q})} \text{ avec } \gamma_{01} = 0 \text{ et } \gamma_{11} = 0$$

où γ_{0q} est l'intercept spécifique à chacune des classes et X_{2i} est le vecteur de variables explicatives associé au vecteur de paramètres γ_{1q} potentiellement différent dans chaque classe latente. Une alternative peut être de modéliser la survenue d'un événement à l'aide d'un modèle des risques proportionnels (Lin et al., 2002a; McCulloch et al., 2002) et ainsi étudier conjointement le processus d'évolution et le délai de survenue de l'événement :

$$\alpha(t)|_{c_{iq}=1} = \alpha_q^0(t)\exp(X_{2i}\gamma_{1q})$$

où α_q^0 est le risque instantané de base de l'événement que l'on peut définir comme spécifique à chaque classe latente ou non.

Estimation

L'hypothèse majeure de ce modèle est l'hypothèse d'indépendance du marqueur et de l'événement conditionnellement aux classes latentes. La vraisemblance conjointe du modèle peut alors s'écrire :

$$L(\theta; Y, T) = \prod_{i=1}^{N} \sum_{q=1}^{Q} \pi_{iq} f(Y_i|c_{iq} = 1; \theta) f(T_i|c_{iq} = 1) \tag{2.9}$$

où $f(T_i|c_{iq} = 1)$ est la densité d'un modèle logistique dans l'approche de Lin et al. (2000) ou bien la densité d'un modèle des risques proportionnels dans Lin et al. (2002a) ou McCulloch et al. (2002).

Lorsque les paramètres sont estimés, les probabilités *a posteriori* d'appartenance à chaque classe peuvent être calculées pour chaque sujet en conditionnant sur ses observations Y_i et T_i. Les sujets peuvent ensuite être rattachés à la classe à laquelle ils ont la plus forte probabilité d'appartenir. Le théorème de Bayes permet d'obtenir la probabilité *a posteriori* $\hat{\pi}_{iq}$ à partir de la formule suivante :

$$\begin{aligned}\hat{\pi}_{iq} &= P(c_{iq} = 1|Y_i, T_i, X_i) \\ &= \frac{P(c_{iq} = 1|X_i) f(Y_i, T_i|c_{iq} = 1, X_i)}{\sum_{l=1}^{Q} P(c_{il} = 1|X_i) f(Y_i, T_i|c_{il} = 1; X_i)}\end{aligned}$$

La dépendance entre évolution et survenue de l'événement qui se fait au travers de variables latentes discrètes c_{iq} est l'un des avantages des modèles à classes latentes par rapport aux modèles à effets aléatoires partagées. En effet, la densité conjointe peut alors s'écrire comme dans les modèles de mélange, c'est-à-dire comme une somme sur les classes latentes. Cela évite le calcul numérique d'intégrales multiples sur les effets aléatoires. Une contrepartie est que l'hypothèse d'indépendance conditionnelle de l'évolution du marqueur et du temps de survenue d'événement est difficile à évaluer car les classes latentes ne sont pas directement observées. Bandeen-Roche et al. (1997) suggèrent que cette hypothèse puisse être contrôlée *a posteriori* : cette hypothèse étant considérée comme valide s'il y a indépendance entre marqueur et événement après ajustement ou stratification sur les classes définies *a posteriori*. Toutefois, les approches pour évaluer la classification *a posteriori* étant variées (Lin et al., 2002a, 2004; Roy, 2003; Guo et Amemyia, 2006), il n'existe pas de consensus sur l'évaluation de l'hypothèse d'indépendance conditionnelle. Le développement de méthodes d'évaluation reste un problème majeur des modèles à classes latentes. Dans ce sens, Jacqmin-Gadda et al. (2009) ont proposé un test du score pour l'hypothèse nulle d'indépendance entre le marqueur et l'événement sachant les classes latentes contre l'hypothèse alternative d'un risque d'événement dépendant d'un ou plusieurs effets aléatoires du modèle mixte pour l'évolution du marqueur.

L'estimation des modèles conjoints à classes latentes présente des inconvénients similaires à ceux des modèles de mélange. L'estimation des modèles est réalisée pour un nombre fixe de classes d'évolution. Le nombre de classes Q peut être supposé connu ou bien choisi par l'intermédiaire d'un critère de sélection tel que le BIC.

2.3.4 Application au vieillissement cognitif

Dans la plupart des études épidémiologiques sur le vieillissement cognitif, les analyses effectuées souffrent d'un biais de sélection lié aux sorties d'étude et aux décès (cf. section 1.2.2). Une façon de corriger ces problèmes de sélection est d'étudier conjointement l'évolution cognitive et la survenue d'une démence, d'une sortie d'étude ou du décès. L'utilisation des modèles conjoints dans l'étude du vieillissement cognitif et de son association avec la survenue d'une démence est récente et peu de travaux ont été réalisés jusqu'à

présent dans ce contexte.

Dans le modèle de Proust-Lima et al. (2006), le processus de vieillissement cognitif est supposé homogène alors qu'il existe plusieurs profils d'évolution cognitive. Proust-Lima et al. (2007b) proposent une extension de son modèle qui permet d'étudier conjointement plusieurs marqueurs quantitatifs et une variable binaire à l'aide d'un modèle conjoint non linéaire à classes latentes. Le modèle non linéaire à processus latent a été étendu en un modèle à classes latentes pour étudier l'hétérogénéité de la population. De plus, la probabilité de survenue d'une démence est modélisée conjointement à l'aide d'un modèle logistique tel que Lin et al. (2000) ou Muthén et Shedden (1999) le définissent. Ce modèle permet de décrire les différents profils d'évolution cognitive associés à la survenue d'une démence et de prédire l'événement clinique en fonction du déclin cognitif. Toutefois, ce modèle comporte une limite importante : l'estimation des paramètres se fait sur un échantillon de sujets ayant un diagnostic de démence à ce temps. L'échantillon est donc selectionné, ce qui peut induire des biais. Afin de lever ce problème, une seconde approche (Proust-Lima et al., 2009) a consisté à modéliser conjointement le risque de survenue de l'événement par un modèle de survie. L'évolution est décrite par le modèle non linéaire à classes latentes et le risque de survenue de l'événement par un modèle de survie. Cette formulation s'apparente à celle de Lin et al. (2002a). Ce modèle a été utilisé pour étudier la relation entre l'évolution cognitive et le risque de survenue d'une démence (Proust-Lima et al., 2009). Nous avons utilisé ce modèle non linéaire à classes latentes pour l'analyse conjointe de plusieurs marqueurs quantitatifs et du délai de sortie d'étude afin d'estimer l'évolution cognitive en tenant compte des biais induits par cette censure informative. Ce travail est présenté dans le chapitre 3.

Hashemi et al. (2003) proposent un modèle conjoint pour l'évolution d'un processus latent caractérisant la cognition et pour le risque de survenue d'une démence. Le processus latent est décrit à l'aide d'un modèle mixte. Le marqueur longitudinal utilisé (MMSE) est une mesure avec erreur du processus latent. Contrairement à ce que proposent Proust-Lima et al. (2006), le lien entre l'évolution cognitive latente et le marqueur est linéaire. Ce modèle d'évolution permet de décrire l'évolution cognitive à partir de marqueurs quantitatifs gaussiens. Le risque d'événement est modélisé conjointement et l'originalité de ce travail réside surtout dans le fait que l'événement (démence) est défini par le passage du

processus latent en dessous d'un seuil à estimer. En résumé, le modèle combine un modèle linéaire pour l'évolution d'un processus stochastique et un modèle à seuil pour la survenue d'un événement. .

Le modèle proposé par Hashemi et al. (2003) présente plusieurs inconvénients. L'évolution du processus latent est limitée à une forme linéaire. De plus, il ne permet de considérer que des marqueurs quantitatifs gaussiens. Or, certains tests, tel que le MMSE, présentent une nature proche d'une variable ordinale alors que d'autres sont par construction quantitatifs comme le DSSTW. La cognition latente telle qu'elle est définie par Hashemi et al. (2003) peut, après avoir franchi le seuil de définition de la démence, repasser au dessus de ce seuil, ce qui offre une probabilité non nulle que le processus latent soit au dessus du seuil à une visite donnée alors qu'il est passé en dessous dans l'intervalle entre deux visites. Le travail de Ganiayre et al. (2008) essaie de s'affranchir de ces difficultés. Premièrement, ils proposent un modèle plus souple pour l'évolution de la cognition latente. Le modèle permet une forme d'évolution de type polynomiale et autorise une relation non-linéaire entre le processus latent et les paramètres du modèle. De plus, ce modèle est développé pour des variables ordinales à l'aide des modèles à seuil pour lier chaque test psychométrique au processus latent. Enfin, l'originalité principale de ce travail consiste non pas à étudier conjointement le risque de démence mais le risque de survenue d'un diagnostic de démence, ce qui autorise le modèle à ne pas se préoccuper de ce qui survient entre deux visites. Une contrepartie de ce modèle est qu'il est très coûteux en calcul numérique. Il est donc difficile d'envisager l'étude conjointe de plusieurs tests psychométriques.

Une extension du modèle à changement de pente aléatoire de Hall et al. (2003) proposée par Jacqmin-Gadda et al. (2006) consiste à étudier conjointement l'accélération du déclin cognitif et l'âge de survenue d'une démence. Ce modèle permet à la fois de décrire le déclin cognitif en phase pré-diagnostique de démence, d'estimer l'âge à l'accélération du déclin et l'âge au diagnostic de démence. L'évolution non linéaire du déclin est modélisée par un modèle mixte polynomial par morceaux avec un changement de pente caractérisant l'âge d'accélération du déclin cognitif. La survenue de la démence est modélisée par un modèle log-normal dépendant de l'âge au changement de pente.

Yu et Ghosh (2009) modélisent conjointement l'accélération du déclin cognitif et la survenue de deux événements compétitifs : la démence et le décès sans démence. Les

sujets décédés sans démence ne sont pas à risque de déclin accéléré. Le modèle conjoint proposé combine un modèle mixte à changement de pente aléatoire pour le test cognitif et deux modèles de Weibull pour la survenue de la démence et du décès. L'aspect le plus intéressant de ce modèle est qu'il suppose qu'une fraction de la population a un risque nul de démence et d'accélération du déclin. Cependant, ce modèle ne permet pas de modéliser la censure informative liée au décès puisqu'il suppose l'indépendance entre le temps de survenue du décès et le niveau cognitif. Une seconde limite de ce modèle est de ne pas supposer que le risque de démence augmente une fois que l'accélération du déclin a eu lieu.

2.4 Méthodes d'estimation

Les paramètres des modèles présentés au cours de ce chapitre peuvent être estimés par maximisation de la vraisemblance $L(\theta)$. L'emploi d'algorithme itératif, du type EM ou Newton-Raphson, est recommandé par Harville (1977). Nous présentons le principe de ces deux algorithmes ainsi qu'une extension de type Newton-Raphson proposée par Marquardt (1963), algorithme utilisé dans nos travaux aux chapitres 3 et 4. Toutefois, le calcul du maximum de vraisemblance peut ne pas être aisé si le nombre d'effets aléatoires du modèle est important. Le calcul des intégrales multiples qui en découle est effectué numériquement et rend parfois difficile la maximisation. Une approche totalement différente pourrait être l'utilisation d'une méthode d'estimation bayésienne.

2.4.1 Algorithme EM

Soit y une variable aléatoire observée d'une distribution que l'on essaie de caractériser à l'aide d'un modèle de paramètres θ. Il est courant que y soit un vecteur de données incomplètes et qu'il existe une variable non observée u (les effets aléatoires) dont la connaissance nous donne un vecteur de données complètes (y,u).

Proposé par Dempster et al. (1977), l'algorithme "Expectation-Maximisation" (EM) est une méthode itérative d'estimation en présence de données incomplètes, basée sur la maximisation de l'espérance conditionnelle de la log-vraisemblance des données complètes

(y,u) sachant les données observées y :

$$E(\log(L(\theta; y, u)|\theta^*) = \int \log(L(\theta; y, u))p(u|y; \theta^*)du$$

Pour tout couple (θ, θ^*), si $E(\log(L(\theta; y, u|\theta^*)) \geq E(\log(L(\theta; y, u|\theta))$, alors $\log(L(\theta^*; y)) \geq \log(L(\theta; y))$. Ainsi, la maximisation de $E(\log(L(\theta; y, u|\theta))$ permet d'atteindre un maximum de $\log(L(\theta; y))$.

Dempster et al. (1977) proposent un algorithme itératif où chaque itération k se décompose en 2 étapes :

E : Calcul de l'espérance de la log-vraisemblance complète sachant les données incomplètes y et l'estimation courante du vecteur de paramètres $\theta^{(k)}$

$$E(\log(L(\theta; y, u)|\theta^{(k)})) = \int \log(L(\theta; y, u))p(u|y; \theta^{(k)})du$$

M : Maximisation de cette espérance $E(\log(L(\theta; y, u)|\theta^{(k)}))$ avec

$$\theta^{(k+1)} = \underset{\theta}{\text{Argmax}}\ E(\log(L(\theta; y, u)|\theta^{(k)}))$$

L'algorithme EM offre à la fois une assurance de convergence de la suite $(\theta^{(k)})_{k\geq 1}$ (Dempster et al., 1977) ainsi que de bonnes propriétés asymptotiques des estimateurs (Nityasuddhi et Bohning, 2003). L'algorithme EM doit être exécuté plusieurs fois avec des valeurs initiales $\theta^{(0)}$ différentes pour éviter les maximas locaux et converger vers le maximum global. Chaque itération de cet algorithme est rapide et facilement calculable. Bien que l'algorithme EM conduise les paramètres rapidement dans la région de l'optimum, la vitesse de progression vers l'optimum ralentit à son approche, ce qui rend l'algorithme EM assez lent en convergence, surtout dans le cas de modèle complexe. L'algorithme de Newton-Raphson est souvent préféré pour sa rapidité de convergence (Lindstrom et Bates, 1988). De plus, les critères d'arrêt de cet algorithme sont peu restrictifs. Par ailleurs, l'algorithme EM ne fournit pas une estimation directe de la variance des paramètres qui doit être calculée *a posteriori*.

2.4.2 Algorithme de Newton-Raphson

L'algorithme de Newton-Raphson (Fletcher, 2000) est un algorithme d'optimisation classique pour trouver les racines d'une fonction d'une ou plusieurs dimensions. Il peut être

utilisé pour trouver un maximum local d'une fonction dès lors qu'il s'agit d'un point de stationnarité défini comme la racine de la dérivée de cette fonction et convient donc pour maximiser la fonction de vraisemblance $L(\theta)$. Il s'agit d'un algorithme itératif basé sur l'utilisation du gradient de la fonction $\nabla(\theta^{(k)})$ et de la matrice hessienne $H^{(k)}$. A l'itération k, le vecteur de paramètres θ est mis à jour en utilisant la relation de récurrence qui suit :

$$\theta^{(k+1)} = \theta^{(k)} - \left(H^{(k)}\right)^{-1}\nabla\left(L(\theta^{(k)})\right)$$

Durant le processus itératif de maximisation de la vraisemblance, le calcul des dérivées premières et secondes peut se faire par différences finies. La matrice des dérivées secondes H peut ne pas être toujours définie positive. Plutôt que d'utiliser l'algorithme de Newton-Raphson, certains auteurs recommandent l'utilisation de méthodes quasi-newtoniennes qui ne nécessitent pas le calcul de la matrice inverse des dérivées secondes H^{-1} mais celui d'une matrice $\tilde{H}^{-1}$ calculée à chaque itération qui tend vers H^{-1} au voisinage de l'optimum. Une contrepartie de ce type de méthode est le temps de convergence qui est plus élevé que pour l'algorithme Newton-Raphson simple. Dans le travail présenté, l'algorithme utilisé est l'algorithme de Marquardt (1963) de type Newton-Raphson :

$$\theta^{(k+1)} = \theta^{(k)} - \delta\left(\tilde{H}^{(k)}\right)^{-1}\nabla\left(L(\theta^{(k)})\right)$$

Le paramètre δ est égal à 1 par défaut mais il peut être modifié (par une optimisation unidimensionnelle) pour assurer une amélioration de la vraisemblance à chaque itération et la matrice $\tilde{H}^{(k)} = \left(\tilde{H}_{ij}^{(k)}\right)$ correspond à la matrice hessienne dont la diagonale a été augmentée afin d'obtenir une matrice définie positive, c'est-à-dire

$$\left(\tilde{H}_{ii}^{(k)}\right) = H_{ii}^{(k)} + \lambda\left[(1-\eta)\left|H_{ii}^{(k)}\right| + \eta tr\left(H^{(k)}\right)\right] \quad \text{et} \quad \tilde{H}_{ij}^{(k)} = H_{ji}^{(k)} \ \text{ si } i \neq j$$

λ et η sont fixées à de petites valeurs puis modifiées en cours d'algorithme : ils augmentent lorsque la matrice Hessienne n'est pas définie positive et diminuent dans le cas contraire.

Les algorithmes de type Newton-Raphson ont l'avantage de converger avec des critères d'arrêt stricts et de fournir directement une estimation de la variance des paramètres, ce qui n'est pas le cas de l'algorithme EM. Dans l'algorithme de Marquardt utilisé dans ce travail, la convergence est définie pour une combinaison de 3 critères d'arrêt :

– la somme des écarts au carré des m paramètres du modèle pour 2 itérations succes-

sives doit être inférieure à la valeur seuil ϵ_a

$$\sum_{j=1}^{m}\left(\theta_j^{(k)} - \theta_j^{(k-1)}\right)^2 \leq \epsilon_a$$

– l'écart entre les vraisemblances du modèle pour 2 itérations successives doit être inférieur à la valeur seuil ϵ_b

$$\left| L\left(\theta^{(k)}\right) - L\left(\theta^{(k-1)}\right) \right| \leq \epsilon_b$$

– un critère basé sur le gradient de la log-vraisemblance qui est en pratique le critère principal pour déterminer le maximum :

$$\nabla\left(L(\theta^{(k)})\right)'\left(H^{(k)}\right)^{-1}\nabla\left(L(\theta^{(k)})\right) \leq \epsilon_c$$

2.4.3 Principe bayésien et algorithme MCMC

Dans les méthodes d'estimation bayésienne, le principe essentiel consiste à considérer les paramètres θ comme une variable aléatoire plutôt que comme un vecteur de paramètres inconnus. La distribution *a priori* de θ est notée $\pi(\theta)$. L'inférence se base sur la loi de distribution de θ conditionnelle aux observations Y. Il s'agit de la distribution *a posteriori* qui est notée $\pi(\theta|Y)$. Le théorème de Bayes nous donne :

$$\pi(\theta|Y) = \frac{f(Y|\theta)\pi(\theta)}{\int f(Y|\theta)\pi(\theta)d\theta}$$

où $f(Y|\theta) = L(\theta)$ est la vraisemblance du modèle. La loi *a posteriori*, $\pi(\theta|Y)$, est proportionnelle au produit de la vraisemblance et de la loi *a priori* de θ, $f(Y|\theta)\pi(\theta)$ à une constante de normalisation près $\int f(Y|\theta)\pi(\theta)d\theta$.

Le problème d'estimation des paramètres consiste alors à considérer la distribution *a posteriori* des paramètres. Son évaluation est rendue difficile puisque la fonction de vraisemblance n'a pas d'expression analytique et que la constante de normalisation ne peut pas toujours être calculée. Les algorithmes de Monte-Carlo par Chaînes de Markov (MCMC) ont été développés pour contourner ces problèmes.

Les algorithmes MCMC sont utilisés lorsque la loi $\pi(.)$ n'est pas simulable directement et/ou qu'elle est connue à une constante de normalisation près. Les 2 algorithmes MCMC les plus courants sont la méthode du rééchantillonage de Gibbs (Geman et Geman, 1984)

et l'algorithme de Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970). Ils sont basés sur le même principe : la génération d'une chaîne de Markov $(\theta^{(m)})_m$ de loi stationnaire la loi d'intérêt $\pi(.)$. Dans le travail que nous présentons, nous n'avons pas exploré l'approche bayésienne pour l'estimation des paramètres des modèles, nous renvoyons vers des ouvrages de reférences (Gilks et al., 1996; Robert et Casella, 2004). L'algorithme MCMC nécessite un grand nombre d'itération pour atteindre la convergence et obtenir la distribution empirique des paramètres estimés. Le processus de convergence est extrêmement long, mais il permet d'estimer les paramètres de modèles à plusieurs effets aléatoires et ce, même si la structure de covariance est complexe et les effets aléatoires non gaussiens. Les critères de convergence utilisés dans ce type d'algorithme ne sont pas clairement définis et apparaissent comme peu restrictifs.

2.4.4 Calcul numérique d'intégrale

Les modèles utilisés et développés dans ce travail ont été estimés par maximisation de la vraisemblance dont le calcul analytique est impossible car les intégrales sur les effets aléatoires n'ont pas de solution analytique. La maximisation de la vraisemblance ne peut se faire que par une approximation numérique de ces intégrales. Plusieurs méthodes d'approximation existent ; nous présenterons dans ce paragraphe la quadrature de Gauss-Hermite utilisée pour évaluer les différentes intégrales. Soit la fonction de log-vraisemblance suivante :

$$l(\theta) = \sum_{i=1}^{N} log \int f(Y_i|u_i) f_u(u_i) du_i$$

L'objectif est d'approximer l'intégrale $\int f(Y_i|u_i) f_u(u_i) du_i$ n'ayant pas de solution analytique. La formule générale de Gauss-Hermite d'ordre r est la suivante :

$$\int g(x) \exp(-x^2) dx = \sum_{q=1}^{r} g(\zeta^{(q)}) A^{(q)}$$

où $\zeta^{(q)}$ sont les noeuds de la quadrature et $A^{(q)}$ sont les poids de la quadrature, donnés dans Abramowitz et Stegun (1972).

Avec un effet aléatoire u_i scalaire, si la densité $f_u(u_i)$ est gaussienne d'espérance μ_{u_i} et de variance $\sigma^2_{u_i}$ alors la log-vraisemblance par la méthode de Gauss-Hermite s'écrit

simplement avec un changement de variable :

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{N} \log \int f(Y_i|u_i) \frac{\exp\left(-\frac{(u_i-\mu_{u_i})^2}{2\sigma_{u_i}^2}\right)}{\sigma_{u_i}\sqrt{2\pi}} du_i \\ &= \sum_{i=1}^{N} \log \int f(Y_i|\sigma_{u_i}\sqrt{2}x + \mu_{u_i}) \frac{\exp(-x^2)}{\sigma_{u_i}\sqrt{2\pi}} \sigma_{u_i}\sqrt{2} dx \\ &= \sum_{i=1}^{N} \log \sum_{q=1}^{r} \frac{f(\sigma_{u_i}\sqrt{2}\zeta^{(q)} + \mu_{u_i})}{\sqrt{\pi}} A^{(q)} \end{aligned}$$

Si la densité des effets aléatoires $f_u(u_i)$ n'est pas gaussienne alors la log-vraisemblance peut s'écrire de la manière suivante :

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{N} \log \int f(Y_i|u_i) f_u(u_i) \frac{\exp(-x^2)}{\exp(-x^2)} du_i \\ &= \sum_{i=1}^{N} \log \sum_{q=1}^{r} f(Y_i|\zeta^{(q)}) f_u(\zeta^{(q)}) \exp((\zeta^{(q)})^2) A^{(q)} \end{aligned}$$

Chapitre 3

Comparaison des modèles à classes latentes et des Pattern Mixture Models pour l'analyse des données longitudinales incomplètes

3.1 Introduction

Dans le chapitre précédent, nous avons présenté deux approches permettant de tenir compte des données manquantes informatives dans les études longitudinales : l'approche par Pattern Mixture Model et celle par modèle de sélection. Plus récemment, une troisième approche a été proposée pour traiter des sorties d'étude informatives : celle du modèle à classes latentes (Roy, 2003; Lin et al., 2004; Beunckens et al., 2008). Il s'agit de modéliser conjointement l'évolution et les sorties d'étude par un modèle conjoint à classes latentes. L'idée sous-jacente est que la population est constituée de classes hétérogènes présentant une évolution de la variable d'intérêt et un risque de sortie d'étude propres à chaque classe. Cette approche est fondée sur l'hypothèse d'indépendance entre l'évolution et le risque de sortie d'étude conditionnellement à la classe. Comme pour les Pattern Mixture Models, cette approche suppose une hétérogénéité de la population. Cependant, contrairement aux PMM, les sous-populations (classes) ne sont pas déterminées *a priori* par la date de sortie d'étude mais seulement associées au risque de sortie d'étude. Les classes ne

sont plus observées mais latentes. Les hypothèses caractérisant le lien entre l'évolution cognitive et le mécanisme de sortie d'étude sont différentes selon l'approche utilisée. Les interprétations des analyses obtenues peuvent différer et doivent se faire avec précaution.

Dans ce chapitre, nous développons un travail, effectué dans le contexte du vieillissement cognitif, portant sur le problème des sorties d'étude informatives et ayant fait l'objet d'un article publié dans la revue *The International Journal of Biostatistics*. L'évolution cognitive a été modélisée par l'intermédiaire du modèle non-linéaire pour données longitudinales multivariées à processus latent proposé par Proust-Lima et al. (2006). Ce modèle, schématisé en figure 3.1, permet d'analyser simultanément l'évolution de plusieurs tests psychométriques en supposant que ces tests sont des mesures avec erreur d'une transformation d'un processus latent qui représente le niveau cognitif global. Une transformation non linéaire permet de tenir compte de la non normalité des tests psychométriques.

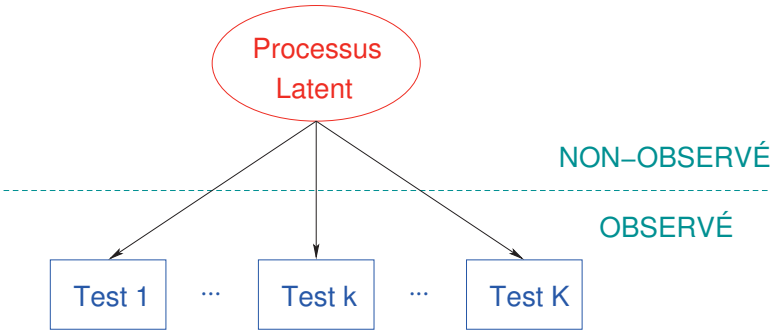

Fig. 3.1 : Modèle non linéaire à processus latent pour données longitudinales multivariées

Nous conduisons une analyse de sensibilité des estimations de l'influence des flavonoïdes sur le déclin cognitif à l'aide de différentes approches statistiques tenant compte des sorties d'étude. Nous avons comparé l'approche par Pattern Mixture Model et deux approches à classes latentes : dans l'approche par PMM, les profils de sorties d'étude sont considérés comme étant des variables explicatives du modèle d'évolution (cf. figure 3.2) ; dans la première approche à classes latentes, la probabilité d'appartenir à une classe latente dépend du profil de sortie d'étude (cf. figure 3.3) et, dans la seconde, le délai de sortie d'étude est modélisé conjointement à l'évolution cognitive et dépend des classes latentes (cf. figure 3.4).

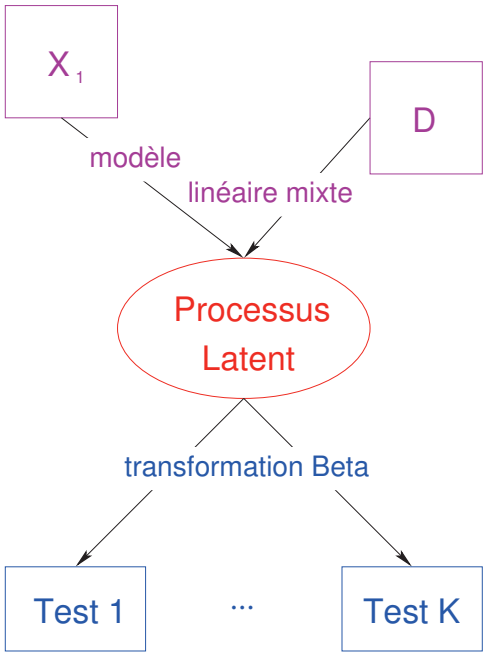

Fig. 3.2 : Pattern Mixture Model pour données longitudinales multivariées avec processus latent

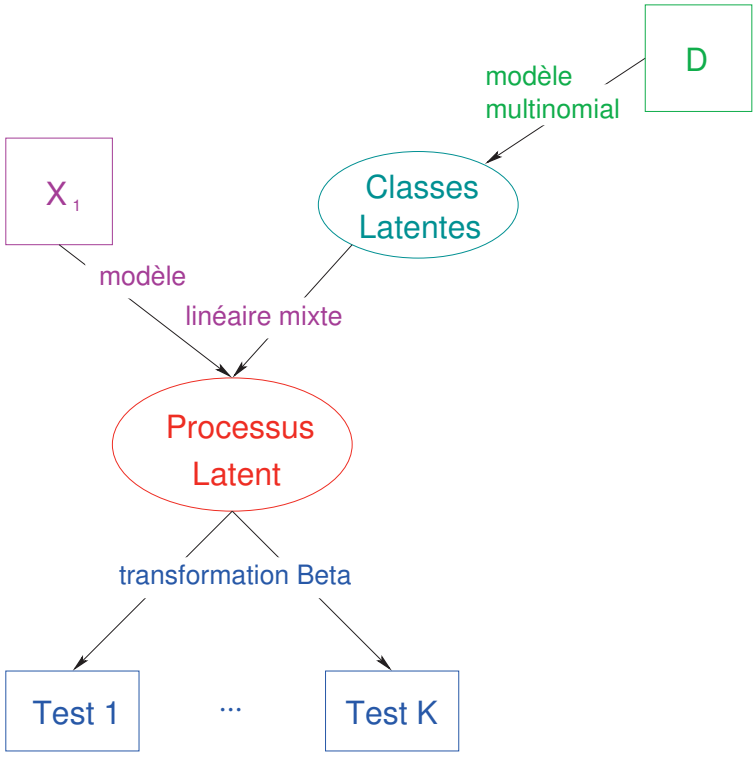

Fig. 3.3 : Modèle simple à classes latentes pour données longitudinales multivariées avec processus latent (D : profil de sortie d'étude)

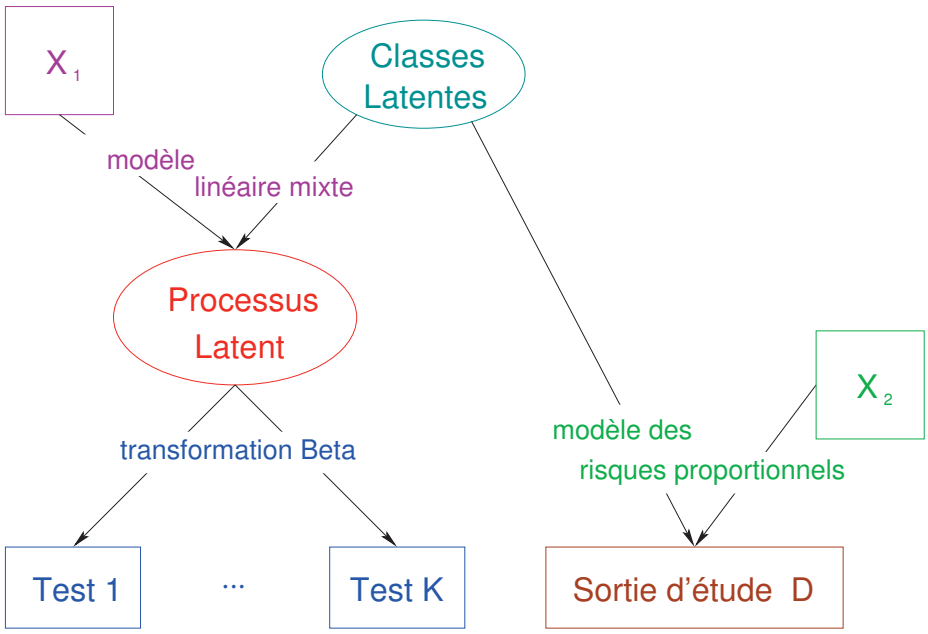

Fig. 3.4 : Modèle conjoint à classes latentes pour données longitudinales multivariées avec processus latent (D : délai de sortie d'étude)

3.2 Article

Résumé de l'article

Les données manquantes, et particulièrement les sorties d'étude, sont fréquentes dans les analyses de données longitudinales. Les estimateurs du maximum de vraisemblance sont consistants quand les données sont manquantes aléatoirement (MAR), mais cette hypothèse étant invérifiable, il est recommandé d'effectuer une analyse de sensibilité sous différentes hypothèses de données manquantes informatives (MNAR). Pour traiter des sorties d'étude informatives, les modèles par schémas d'observation (Pattern Mixture Models) ont été développés. Plus récemment, les modèles à classes latentes (Latent Class Models) ont été proposés comme une alternative aux PMM présentant l'avantage d'assouplir certaines hypothèses. L'objectif de cet article est de comparer les approches par PMM et LCM pour la prise en compte des sorties d'étude. Ce travail a été réalisé dans le contexte du vieillissement cérébral des personnes âgées afin d'étudier l'évolution cognitive mesurée par plusieurs tests psychométriques. Le modèle longitudinal multivarié à processus latent développé par Proust-Lima et al. (2006) a été utilisé pour réaliser une analyse de sensibilité et comparer les estimations sous l'hypothèse MAR avec celles des modèles sous hypothèse MNAR par une approche PMM et deux approches LCM. Dans l'approche par PMM, les profils de sorties d'étude sont inclus comme covariables du modèle longitudinal multivarié. Dans l'approche par LCM simple, les profils de sorties d'étude sont considérés comme des prédicteurs de la probabilité d'appartenance à une classe, alors que dans l'approche par LCM conjoint, le délai de survenue d'une sortie d'étude est modélisé conjointement à l'aide d'un modèle des risques proportionnels dépendant des classes latentes. Nous montrons que les valeurs des paramètres peuvent être différents d'une approche à l'autre et que leurs interprétations sont différentes. Les paramètres obtenus par PMM sont ajustés sur les profils de sorties d'étude, tandis que ceux issus des deux approches par LCM sont ajustés sur les classes latentes. Cette différence est particulièrement mise en évidence dans notre jeu de données puisque les modèles à classes latentes montrent une hétérogénéité sous-jacente de notre population plus grande que pour l'approche par PMM. Pour les modèles à classes latentes, nous suggérons plusieurs analyses complémentaires afin que les difficultés dans l'interprétation des paramètres, quand ces modèles sont utilisés pour traiter des sorties d'étude informatives, soient mieux comprises.

The International Journal of Biostatistics

Volume 4, *Issue* 1 2008 *Article* 14

Pattern Mixture Models and Latent Class Models for the Analysis of Multivariate Longitudinal Data with Informative Dropouts

Etienne Dantan* Cécile Proust-Lima†

Luc Letenneur‡ Helene Jacqmin-Gadda**

*INSERM, U897, Center of Epidemiology and Biostatistics of Bordeaux, etienne.dantan@isped.u-bordeaux2.fr

†Université Victor Segalen Bordeaux II, cecile.proust@isped.u-bordeaux2.fr

‡INSERM, U897, Center of Epidemiology and Biostatistics of Bordeaux, luc.letenneur@isped.u-bordeaux2.fr

**INSERM, U897, Center of Epidemiology and Biostatistics of Bordeaux, helene.jacqmin-gadda@bordeaux.inserm.fr

Pattern Mixture Models and Latent Class Models for the Analysis of Multivariate Longitudinal Data with Informative Dropouts

Etienne Dantan, Cécile Proust-Lima, Luc Letenneur, and Helene Jacqmin-Gadda

Abstract

Missing data and especially dropouts frequently arise in longitudinal data. Maximum likelihood estimates are consistent when data are missing at random (MAR) but, as this assumption is not checkable, pattern mixture models (PMM) have been developed to deal with informative dropout. More recently, latent class models (LCM) have been proposed as a way to relax PMM assumptions. The aim of this paper is to compare PMM and LCM in order to tackle informative dropout in a longitudinal study of cognitive ageing measured by several psychometric tests. Using a multivariate longitudinal model with a latent process, a sensitivity analysis was performed to compare estimates under the MAR assumption, from a PMM and from two LCM. In the PMM, dropout patterns are included as covariates in the multivariate longitudinal model. In the simple LCM, they are predictors of the class membership probabilities while, in the joint LCM, the dropout time is jointly modeled using a proportional hazard model depending on latent classes. We show that parameter interpretation is different in the two kinds of models and thus can lead to different estimated values. PMM parameters are adjusted on the dropout patterns while LCM parameters are adjusted on the latent classes. This difference is highlighted in our data set because the latent classes exhibit much more heterogeneity than dropout patterns. We suggest several complementary analyses to investigate the characteristics of latent classes in order to understand the meaning of the parameters when using LCM to deal with informative dropout.

KEYWORDS: dropout, joint model, latent class model, missing data, mixed model, pattern mixture model

1 Introduction

The main objective of longitudinal studies is to describe change of an outcome over time. Missing responses and especially dropouts frequently hamper the analyses of longitudinal data. It is now well known that maximum likelihood estimators computed from the available data are consistent when missing data are ignorable (Verbeke, 2000), that is when data are Missing At Random (MAR) (Little, 1987) and parameters in the dropout model and in the response model are functionally independent. However, the MAR assumption cannot be tested since it requires that the missingness probability does not depend on the unobserved values of the outcome. When the response process and the missingness process are not independent, data are called Missing Not At Random (MNAR) or informative, and the analysis requires joint modeling of the two processes. To date, three approaches have been proposed: selection models (Diggle, 1994), pattern mixture models (Little, 1993) and latent class models (Roy, 2003). Based on different factorizations of the joint distribution, they rely on strong and uncheckable hypotheses on the missingness process and on the distribution of the unobserved outcome. As a consequence, it is generally recommended to perform a primary MAR analysis followed by a sensitivity analysis under various MNAR assumptions (Verbeke, 2001; Thijs, 2002).

Selection models factor the joint distribution into the marginal distribution of the outcome and the distribution of the missingness probability given the outcome. They are called either outcome-dependent when the dropout probability at time t depends directly on the unobserved outcome, or random-effect dependent (also referred to as shared parameter models) when the dropout probability depends on the random-effects from the mixed model for the outcome (Little, 1995). In the beginning, selection models raised enthusiasm because they allowed direct estimation of the parameters from the marginal distribution of the outcome that are of primary interest. The enthusiasm waned when it was shown that these estimates were very sensitive both to misspecification of the complete distribution of the outcome or to the assumed shape of the dependency between the dropout process and the outcome process (Kenward, 1998). These results increased the attractivity of the main alternative approach, the Pattern Mixture Models.

Pattern Mixture Models (PMM) factor the joint distribution into the marginal distribution of the missingness process and the conditional distribution of the outcome given the missingness pattern. Thus, evolution of the outcome is described conditionally on time of dropout and marginal parameters are not directly available. Simple pattern mixture analyses may be performed

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

by stratifying on the dropout pattern or by including the dropout pattern as covariate in the model. In most cases, some patterns have too few subjects or too few measurements per subject and PMM require specification of constraints to ensure parameter identifiability. These constraints are parametric relationships (most often, equality) between parameters associated with different dropout patterns (Little, 1993; Molenberghs, 1998) or grouping of several patterns. PMM are often preferred to selection models in sensitivity analysis because the constraints are explicit and more meaningful than the assumptions required in selection models, and software is readily available (PMM are estimated with the same software as used under the MAR assumption).

More recently, several papers have proposed to tackle MNAR data in longitudinal studies using Latent Class Models (LCM) (Roy, 2003; Lin, 2004; Beunckens, 2007). As in PMM, the idea underlying LCM is that the population is a mixture of sub-populations with different profiles of outcome evolution. However, in PMM, the sub-populations are known *a priori* since they are defined by the dropout patterns (possibly grouping some patterns) while class membership is unobserved in LCM. It is data-driven assuming only an association with the dropout. LCM have been recommended as an alternative to PMM for data sets with numerous or sparse patterns because it is typically expected that LCM will include fewer latent classes than dropout patterns. This avoids identifiability issues and noisy estimates that are difficult to interpret (Roy, 2003; Lin, 2004). However, the well-defined constraints required for PMM identifiability are replaced in LCM by an assumption of conditional independence between dropout and responses, given the latent classes which can be only partially checked. In Roy (2003), dropout time is included as covariate in the class membership probability (this will be referred to afterwards as simple LCM). In Lin et al. (2004) and Beunckens et al. (2007), the missingness process is modeled jointly with the outcome process through latent classes (referred to as joint LCM). Thus, these joint latent class models may also be considered as random-effect dependent selection models with qualitative random-effects linking the two processes. Nevertheless, they have the notable advantage that the correlation between repeated measures of the outcome and the correlation between the outcome and the missingness process are modeled separately. However, LCM share the drawbacks of mixture models regarding possible local maxima in the likelihood (Redner, 1984) and somewhat unclear interpretation of the latent classes (Bauer, 2003). Although LCM have been presented as a relaxing of PMM, the two approaches have never been compared on real data sets except in Beunckens et al. (2007), who briefly presented estimates with the two approaches but did not discuss the different parameter interpretations.

DOI: 10.2202/1557-4679.1088

In many longitudinal studies, repeated measures of several correlated outcomes are often collected. In some cases, these outcomes are so correlated that they may be viewed as several noisy markers of the same latent variable that is the actual variable of interest. When studying risk factors of cognitive ageing, for instance, association with the latent cognitive process is of greater interest than association with specific psychometric tests used to measure cognition. Recently, a latent process model has been proposed to analyze such multivariate longitudinal data (Proust, 2006) and has been applied to the Paquid cohort study for investigating nutritional risk factors in cognitive ageing under the MAR assumption (Letenneur, 2007). However, prospective cohorts of elderly subjects contain many dropouts that have been found associated with cognitive decline (Jacqmin-Gadda, 1997).

In the present paper, we conducted a sensitivity analysis of these results comparing pattern mixture and two latent class analyses (simple and joint LCM) to deal with dropouts. The latter was performed using latent class extension of the Proust et al. model for jointly modeling multivariate longitudinal data and time-to-event (Proust-Lima, submitted). The main goal of this work is to highlight and discuss the different interpretations of the parameters in these models that may lead to apparently inconsistent results.

In the next section, we present the data set. Section 3 is devoted to a brief description of the latent process model for multivariate longitudinal data (Proust, 2006). The three methods for handling informative dropout are reviewed in section 4. Data analysis results are presented in section 5 and the relative merits of the two approaches are discussed in section 6.

2 Data

Data come from the French prospective cohort study PAQUID on functional and cerebral aging (Letenneur, 1994). This cohort included at baseline 3,777 community dwellers aged 65 and over randomly recruited from electoral rolls in two administrative areas of southwest France. Participants were visited at home by a psychologist for the baseline interview in 1988-1989 and one (V1), three (V3), five (V5), eight (V8), ten (V10) and thirteen years (V13) after the baseline assessment. At each visit, a set of psychometric tests was administered, including evaluation of global mental status by the Mini-Mental State Examination (MMSE) (Folstein, 1975), visual memory by the Benton's Visual Retention Test (BVRT) (Benton, 1965) and verbal fluency by the Isaacs Set Test shortened to 15 seconds (IST) (Isaacs, 1973). Diagnosis of dementia was assessed using DSM-III R criteria (American Psychiatric Association, 1987).

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

The present analysis is based on a sub-sample of 1,640 non-demented subjects who participated in a nutritional study at the three-year follow-up (Letenneur, 2007) and completed at least one of the three psychometric tests at one of the next visits. Evolution of cognitive performance measured by the MMSE, the BVRT and the IST was analyzed over a 10-year period between the three-year follow-up (considered as the baseline visit, i.e. t=0) and the 13-year follow-up (t=10). For this sensitivity analysis, we focused only on educational level (without the first French diploma from primary school versus this diploma or higher level) and sex, two factors whose impact on cognitive decline in the elderly is still debated (Wiederholt, 1993; Elias, 1997; Le Carret, 2003). The analyses were adjusted for age in 4 age-groups [65-70[, [70-75[, [75-80[and [≥80[.

In longitudinal studies, missing data can be monotone when a missing data is not followed by any measurement, or intermittent when a missing data is followed by at least one measurement. Monotone missing data are also called dropouts. We considered that one subject dropped out at time t if none of the three psychometric tests was completed at time t and until the end of the study. Five dropout profiles were defined: dropout at V5, V8, V10 and V13 and no dropout. In this analysis, we studied only the sensitivity to informative dropouts; intermittent missing data were taken to be ignorable. Characteristics of subjects according to dropout time are described in Table 1 as percentages.

3 The latent process model for multivariate longitudinal data

The analysis under the MAR assumption was carried out using the nonlinear latent process model for multivariate longitudinal outcomes proposed by Proust et al. (2006) which is outlined in figure 1 and briefly described below. Let define Y_{ijk} the outcome k measured at time t_{ijk} for subject i with $k = 1, ..., K$, $i = 1, ..., N$ and $j = 1, ..., n_{ik}$. In our application, the outcomes are the 3 psychometric tests (K=3). We assume that a transformation of Y_{ijk} is a measure with error of a continuous latent process $\Lambda_i(t)$ representing the latent cognitive level in our application.

$$h_k(Y_{ijk}; \eta_k) = \Lambda_i(t_{ijk}) + \alpha_{ik} + \epsilon_{ijk} \quad (1)$$

where h_k is a flexible monotonous increasing transformation that depends on test-specific parameters η_k to be estimated and serves as a link function be-

DOI: 10.2202/1557-4679.1088

Table 1: Distribution of characteristics of subjects in Paquid sample according to dropout patterns (%) and p-value for the Chi-square test of independence.

	Visit of dropout					
	V5 (N=236)	V8 (N=256)	V10 (N=173)	V13 (N=261)	No dropout (N=714)	p-value
Intermittent missing data	0.0	1.2	4.6	21.5	12.7	
Diagnosed as demented[2]	0.4	8.2	9.3	7.7	15.4	< 0.001
Men	49.6	43.4	54.3	39.8	37.2	< 0.001
No diploma[3]	33.9	27.3	26.6	25.7	21.3	0.003
Age at V3						< 0.001
65-70	4.7	9.0	7.5	8.1	16.9	
70-75	25.0	21.5	28.3	31.0	42.9	
75-80	19.5	22.3	22.6	24.9	26.9	
$\geq$80	50.8	47.3	41.6	36.0	13.3	

[1] % of subjectswith intermittent missing data among subjects dropped out at each visit (e.g. 4.6% of the 173 subjects dropped out at V10 had missed at least one visit before (at V3 or V5))
[2] before dropout
[3] no first diploma from primary school

tween the measured score and the latent process. We chose the Beta cumulative density function (CDF) for h_k because it depends only on two parameters and offers large flexibility in the shape (Proust, 2006). As a Beta CDF is defined in $[0-1]$, each marker was rescaled to the unit interval. The subject-and-marker-specific random intercept α_{ik} was introduced to allow variability in the individual performance to each psychometric test conditionally on the latent process value. This random intercept and the independent error ϵ_{ijk} followed respectively the independent Gaussian distributions $\mathcal{N}(0, \sigma^2_{\alpha_k})$ and $\mathcal{N}(0, \sigma^2_{\epsilon_k})$.

Change over time of the latent process is described by a linear mixed model:

$$\Lambda_i(t) = X^T_{1i}(t)\beta + Z^T_i(t)u_i \tag{2}$$

where $X_{1i}(t)$ is a vector of possibly time-dependent covariates associated with the vector of fixed-effects β, $Z_i(t)$ is a subvector of $X_{1i}(t)$ and u_i is a vector of subject-specific random-effects with $\mathcal{N}(0, B)$ distribution, where B is an unstructured positive definite matrix. In our application, the latent process evolution is assumed to be a linear function of time.

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

The log-likelihood of the model may be decomposed as the log-likelihood for the transformed outcomes $\tilde{Y}_{ijk} = h_k(Y_{ijk}; \eta_k)$ plus the jacobian $J(y_i, \theta)$ of the transformations:

$$l(\theta) = \sum_{i=1}^{N} f(\tilde{y}_i; \theta) + \sum_{i=1}^{N} ln(J(y_i, \theta)) \quad (3)$$

where $\tilde{y}_i = (\tilde{y}_{i11}, ..., \tilde{y}_{in_{i1}1}, ..., \tilde{y}_{in_{iK}K})^T$.

Maximum likelihood estimators of the whole set of parameters θ are computed using the Marquardt algorithm (Marquardt, 1963), a Newton-Raphson-like algorithm. The estimation procedure was implemented in a Fortran90 program (Proust, 2006). The code for the estimation the latent process model for multivariate longitudinal data and its user's guides files are available at the following url: http://biostat.isped.u-bordeaux2.fr (program NLMULTIMIX).

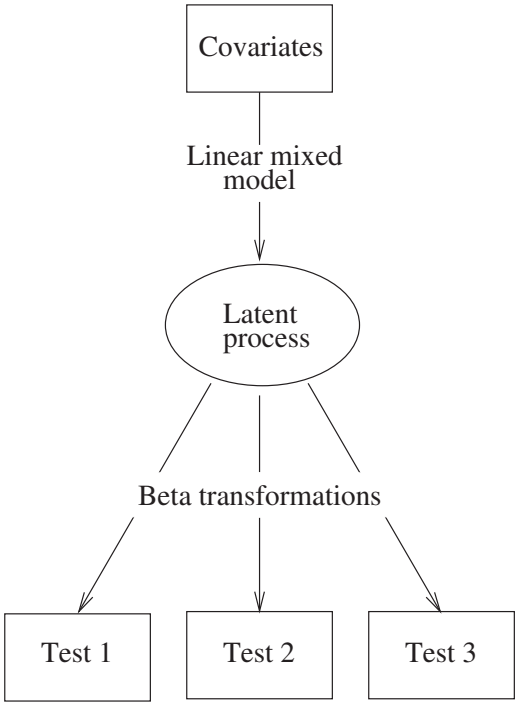

Figure 1: Diagram of the latent process model for multivariate longitudinal data estimated under the MAR assumption.

DOI: 10.2202/1557-4679.1088

4 Methods for handling informative dropout

4.1 Pattern mixture approach

PMM factorizes the joint distribution of the response variable Y and the dropout process D given covariates X as $[Y, D|X] = [Y|D, X][D|X]$. The dropout distribution $[D|X]$ is often estimated as the proportion of each dropout pattern for each combination of covariate values, while the estimation of the conditional distribution $[Y|D, X]$ requires a model for the evolution of Y given dropout patterns D and covariate X, which can be estimated separately by maximum likelihood. With the primary model defined by (1) and (2) and denoting $D_{il} = 1$ if subject i dropped out at visit V_l, a flexible model for $[Y|D, X]$ is:

$$\Lambda_i(t) \mid D_{il} = 1 = X_{1i}^T(t)\beta_l + Z_i^T(t)u_{il} \quad \text{with } u_{il} \sim \mathcal{N}(0, B_l) \tag{4}$$

All the parameters are pattern-specific and estimates are obtained through stratified analyses on the dropout patterns. Some dropout patterns can be gathered if necessary to ensure parameter identifiability. However, this leads to a very large number of parameters that are estimated on subsamples of unequal sizes and unequal numbers of measurements per subject, thereby inducing a large variability. More parsimonious models are obtained by introducing the dropout patterns as covariates in the primary model and assuming that some parameters are common over the patterns. Some authors have given examples of meaningful assumptions to help to specify the constraints (Little, 1993; Kenward, 1998). As these models are nested in the stratified model, model selection may be guided by the likelihood ratio test. In the present work, estimates from the PMM were obtained using the estimation program developed for the latent process model.

4.2 Simple latent class model

The latent class model assumes that the population is divided into G unobserved sub-populations with distinct profiles of evolution for the latent process. Denoting C the latent class variable, the joint distribution $[Y, D|X]$ is decomposed as $[Y, D|X] = \sum_C [Y|C, X][C|D, X][D|X]$. As for PMM, $[D|X]$ may be estimated empirically while parameters from $[Y|C, X][C|D, X]$ are estimated by maximum likelihood conditioning on dropout patterns. Note that interpretation of the regression parameters in the conditional distribution $[Y|C, X]$ becomes unclear if the same covariates are included in the class membership

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

probability $[C|D, X]$. Thus, in the following, we assumed $[C|D, X] = [C|D]$. We return to this assumption in section 4.4.

More specifically, by denoting $c_{ig}(g = 1, ..., G)$ the latent class membership variables, which equal 1 if subject i belongs to latent class g and 0 otherwise, the latent process evolution is defined given the latent class by:

$$\Lambda_i(t) \mid_{c_{ig}=1} = X_{1i}^T(t)\beta_g + Z_i^T(t)u_{ig} \ , \quad t \geq 0 \tag{5}$$

$$u_{ig} \sim \mathcal{N}(0, \omega_g^2 B)$$

Similarly to PMM, more parsimonious models are obtained by assuming that some parameters are common over classes. As in Roy (2003), the class membership probability is defined using a multinomial logistic regression with the dropout pattern included as a categorical variable:

$$\pi_{ig} = P(c_{ig} = 1|D_i) = \frac{e^{\xi_{0g} + D_i^T \xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l} + D_i^T \xi_{1l}}} \tag{6}$$

where ξ_{0g} is the intercept for class g and ξ_{1g} is the vector of class-specific parameters associated with the vector of indicator variables for dropout patterns: $D_i^T = (D_{i5}, D_{i8}, D_{i10}, D_{i13})$ (reference: no dropout). For identifiability, $\xi_{01} = 0$ and $\xi_{11} = 0$.

The log-likelihood of the simple LCM is:

$$l(\theta) = \sum_{i=1}^{N} ln\left(\sum_{g=1}^{G} \pi_{ig} f(\tilde{y}_i|c_{ig} = 1; \theta) \right) + \sum_{i=1}^{N} ln(J(y_i; \theta)) \tag{7}$$

4.3 Joint latent class model

In the joint latent class model, the joint distribution $[Y, D|X]$ is decomposed as $[Y, D|X] = \sum_C [Y|C, X][D|C, X][C|X]$. The dropout process is thus jointly modeled and depends on Y through the latent class variable. Parameters for the 3 distributions are estimated simultaneously by maximizing the joint likelihood. As for the simple LCM, the model for $[Y|C, X]$ is defined by (1) and (5). To facilitate parameter interpretation in $[Y|C, X]$, we do not include any covariate in the class membership probability ($[C|X] = [C]$).

In the present work, we model the time to dropout using a proportional hazard model. Let define δ_i the dropout indicator that equals 1 if subject i dropped out and 0 if subject i was seen at the last visit V13. The dropout time is not observed continuously but only at discrete visit times. We denote T_{oi} the time at the last observation and T_{mi} the time at the first missing value

DOI: 10.2202/1557-4679.1088

(next planned visit). Thus, if $\delta_i = 0$, $T_{oi} = 10$. The risk function may be defined as follows:

$$\lambda(t \mid c_{ig} = 1, X_{2i}; \gamma_{0g}, \gamma_1) = \lambda_0(t) e^{\gamma_{0g} + X_{2i}\gamma_1} \quad \text{with } \gamma_{01} = 0 \tag{8}$$

where γ_{0g} is the logarithm of the relative risk of dropout in latent class g compared to latent class 1 adjusted for covariates X_{2i}. The vector of covariates X_{2i} is associated with the vector of parameters γ_1. In this work, we assume that the impact of X_{2i} on the risk of dropout does not depend on latent classes. A 5-step function was used for the baseline risk function $\lambda_0(t)$.

To make sure that probabilities remain in $[0, 1]$, the class membership probability is defined by:

$$\pi_{ig} = P(c_{ig} = 1) = \frac{e^{\xi_{0g}}}{\sum_{l=1}^{G} e^{\xi_{0l}}} \quad \text{with } \xi_{01} = 0 \tag{9}$$

The individual contribution to the likelihood is decomposed according to the latent classes using the conditional independence of dropout time and outcomes. Then the density $f(y_i|c_{ig} = 1)$ is computed using the Jacobian of the Beta transformations as in (3). Thus, the log-likelihood is obtained by:

$$\begin{aligned} l(\theta) = \sum_{i=1}^{N} ln \Bigg(\sum_{g=1}^{G} \pi_{ig} f(\tilde{y}_i | c_{ig} = 1; \theta) [S(T_{oi}|c_{ig} = 1) - S(T_{mi}|c_{ig} = 1)]^{\delta_i} \\ \times S(T_{oi}|c_{ig} = 1)^{1-\delta_i} \Bigg) + \sum_{i=1}^{N} ln(J(y_i; \theta)) \end{aligned} \tag{10}$$

where S denotes the survival function. For both simple and joint latent class models, parameters were estimated by maximizing (7) or (10) for several numbers of latent classes G by using the Marquardt algorithm (Proust-Lima, submitted; Marquardt, 1963) The optimal number of classes was selected according to the Bayesian Information Criterion (BIC) (Schwartz, 1978). As local maxima are possible with mixture models, each model was estimated by using different sets of initial parameters to ensure convergence to the global maximum. When parameters are estimated, subjects may be classified in the latent classes using the posterior probabilities to belong to each latent class given the data. These probabilities are computed using the Bayes theorem. The estimation procedure was implemented in Fortran90 and codes for methods dealing with informative dropouts can be provided on request.

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

4.4 Parameter interpretation

In PMM and LCM, parameters are estimated from the conditional distribution $[Y|D,X]$ or $[Y|C,X]$. These parameters provide interesting information about the heterogeneity of the outcome distribution, but in most cases their interpretation is different from those in the marginal distribution $[Y|X]$. The latter may be computed by using

$$\left.\begin{array}{rl} \underline{\text{PMM}}: & [Y|X] = \sum_D [Y|D,X][D|X] \\ \underline{\text{Simple LCM}}: & [Y|X] = \sum_C [Y|C,X][C|X] = \sum_C [Y|C,X] \sum_D [C|D][D|X] \\ \underline{\text{Joint LCM}}: & [Y|X] = \sum_C [Y|C,X][C] \end{array}\right\} \quad (11)$$

To emphasize the different interpretations, let assume a conditional model for Y including one covariate, X_{11i}, with a common effect over the groups (defined by dropout patterns in PMM or latent classes in LCM) and one covariate X_{12i} with a group-specific effect. Whatever the kind of model (PMM or LCM), the group-specific expectation of the latent process in group g is:

$$E(\Lambda_i|X_{11i}, X_{12i}, g) = X_{11i}\beta_1 + X_{12i}\beta_{2g}. \quad (12)$$

Formula (12) shows that β_{2g} is a measure of the impact of X_{12i} in each group and β_1 is a measure of the impact of X_{11i} adjusted for the differential effect of X_{12i} across the groups. For instance, if X_{12i} is a vector of 1, β_1 is the effect of X_{11i} adjusted for the group difference on the intercept. The meaning of β_1 depends on the meaning of the groups, so it is different in a PMM where the groups are the dropout patterns and in a LCM where the groups are the latent classes.

To obtain the marginal effect of X_{11i} adjusted for the other covariates, the marginal expectation needs to be computed:

$$E(\Lambda_i|X_{11i}, X_{12i}) = X_{11i}\beta_1 + X_{12i} \sum_g \beta_{2g} P(g|X_{11i}, X_{12i}) \quad (13)$$

and the marginal effect of X_{11i} is:

$$\begin{array}{l} E(\Lambda_i|X_{11i}=1, X_{12i}) - E(\Lambda_i|X_{11i}=0, X_{12i}) = \beta_1 + \\ X_{12i} \times \sum_g \beta_{2g}(P(g|X_{11i}=1, X_{12i}) - P(g|X_{11i}=0, X_{12i})) \end{array} \quad (14)$$

DOI: 10.2202/1557-4679.1088

Thus, β_1 can be interpreted as the marginal effect of X_{11i} only if the group membership probability does not depend on X_{11i}. This probability is $P(D_{ig} = 1|X_i)$ in the PMM, $P(c_{ig} = 1|X_i) = \sum_l Pr(c_{ig} = 1|D_{il} = 1)P(D_{il} = 1|X_i)$ in the simple LCM and $P(c_{ig} = 1|X_i)$ in the joint LCM. As a consequence, even if the estimation of conditional parameters in PMM and LCM does not require the specification of $[D|X]$, additional assumptions about the dropout pattern distribution are required for marginal inference. More specifically, parameters common over classes retain the marginal interpretation only under the strong and generally violated assumption that the dropout probabilities do not depend on X. Furthermore, common parameters in a joint LCM retain the marginal interpretation only when the class membership probability does not depend on X. In the other cases, the difference in the marginal means (13) depends on the value of all the covariates with group-specific effects.

5 Application

5.1 Analysis under the MAR assumption

Parameters from the model for multivariate longitudinal data defined by (1) and (2) including time from the baseline visit (V3), sex, educational level, age and the 3 interactions with time as explanatory variables were estimated under the MAR assumption. Estimates are given in Table 2.

The difference between men and women with regard to initial cognitive level was not significant ($p = 0.07$), but men had a slower cognitive decline than women ($p = 0.01$). Subjects without a diploma had a lower initial level of cognition ($p < 0.01$), but the impact on cognitive evolution was not significant ($p = 0.56$). Older subjects had a lower initial cognitive level ($p < 0.01$) and a sharper cognitive decline ($p < 0.01$).

5.2 Pattern mixture models

Pattern mixture analyses were carried out by putting together subjects dropped out at V5 (with only one measure) and V8 to ensure parameter identifiability. This gave 4 dropout patterns (V5/V8, V10, V13, no dropout). The primary model stratified on the 4 dropout patterns, that is with all parameters specific to the pattern (4×27 parameters), was estimated and the log-likelihood was computed by summing the log-likelihoods of the 4 pattern-specific models. Then, we sought a more parsimonious model including dropout pattern and interactions, with dropout pattern as covariates in the mixed model for the la-

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

Table 2: Estimates from the model for multivariate longitudinal data under the MAR assumption and in the PMM, simple LCM and joint LCM.

	MAR		PMM		Simple LCM		Joint LCM	
	β	se	β	se	β	se	β	se
intercept	0.643	0.009	0.659	0.009				
t[1]	−0.077	0.011	−0.071	0.010				
class 1					0.683	0.009	0.696	0.010
class 1 * t					−0.045	0.009	−0.041	0.010
class 2					0.628	0.010	0.624	0.016
class 2 * t					−0.166	0.034	−0.099	0.021
class 3					0.638	0.019	0.641	0.019
class 3 * t					−0.551	0.004	−0.494	0.044
Men	0.008	0.005	0.012	0.004	0.008*	0.004	0.004*	0.004
Men * t	0.021	0.008	0.024	0.008	0.012*	0.006	0.008*	0.006
No Diploma	−0.104	0.005	−0.103	0.005	−0.096	0.005	−0.096	0.005
No Diploma * t	−0.006*	0.010	−0.006*	0.009	−0.016	0.008	−0.018	0.008
Age[2]								
70-75	−0.014*	0.008	−0.011*	0.007	−0.009*	0.006	−0.011*	0.007
75-80	−0.048	0.008	−0.041	0.008	−0.036	0.007	−0.039	0.007
≥80	−0.098	0.008	−0.078	0.008	−0.073	0.007	−0.084	0.007
Age * t[2]								
70-75 * t	−0.023*	0.012	−0.019*	0.011	−0.014*	0.009	−0.017*	0.009
75-80 * t	−0.067	0.013	−0.058	0.012	−0.044	0.010	−0.049	0.010
≥80 * t	−0.117	0.014	−0.094	0.014	−0.067	0.011	−0.085	0.011
Dropout[3]								
V5/V8			−0.047	0.006				
V10			−0.022	0.007				
V13			−0.023	0.007				
Dropout * t								
V5/V8 * t			−0.081	0.026				
V10 * t			−0.048	0.016				
V13 * t			−0.056	0.012				
Beta parameters	1.173	0.028	1.146	0.029	0.983	0.034	0.985	0.034
	−2.690	0.037	−2.665	0.037	−2.496	0.038	−2.497	0.038
	0.325	0.021	0.306	0.021	0.199	0.026	0.202	0.025
	−2.389	0.025	−2.383	0.025	−2.315	0.024	−2.318	0.024
	0.470	0.024	0.445	0.024	0.320	0.028	0.323	0.028
	−2.377	0.029	−2.361	0.028	−2.264	0.028	−2.266	0.028

[1] t: unit = 10 years
[2] Likelihood Ratio Test with 3 df significant at $\alpha = 0.01$, for each approach
[3] Reference: No dropout
* Non significative at $\alpha = 0.05$

tent process (4) and allowing pattern-specific covariance matrix for the random effects. According to the likelihood ratio statistic, none of the interactions between the dropout patterns and the covariates (age, sex and educational level) was significant, but the random-effect variance was significantly different over dropout patterns. Thus, the best model included dropout pattern as a simple effect and with an interaction with time. Estimates of fixed effect are displayed in Table 2.

DOI: 10.2202/1557-4679.1088

The likelihood ratio test revealed that dropout patterns were significantly associated with cognition ($\chi^2 = 187.4$ for 6 df, $p < 0.01$): early dropout was associated with a lower initial cognitive level and a sharper cognitive decline. Subjects dropped out at V10 and V13 presented similar profiles of cognitive evolution. Regarding the covariates of interest (sex, educational level and age), PMM results were close to those obtained under the MAR assumption, except for the sex effect on initial level which appeared to be more significant: men exhibited a significantly higher mean initial level than women ($p = 0.01$).

5.3 Latent class models

The two LCM were estimated by including the same covariates as under the MAR assumption (age, sex, educational level and their interaction with time) in the mixed model for the latent process (5) and assuming class-specific intercepts and class-specific slopes with time. Thus, the mixed model in the LCM was identical to the mixed model in the PMM estimated above, except that dropout patterns were replaced by latent classes. We successively estimated models with 1, 2, 3 and 4 latent classes. The 3-latent class model was selected according to the BIC.

Table 3 displays parameter estimates of the multinomial logistic model for class membership probabilities defined by (6) for the simple LCM. Dropout patterns were globally significantly associated with the latent classes ($\chi^2 = 118.7$ for 8 df, $p < 0.0001$). Subjects who did not drop out from the study had a very low probability to be in class 2 or 3: the odds (or ratio of the probabilities) of being in class 2 (respectively 3) compared to class 1 was $exp(-2.08) = 0.12$ (respectively $exp(-3.61) = 0.027$). Moreover, the odds ratio of class 2 versus 1 clearly increased as the time of follow-up decreased compared to subjects without dropout. The odds for being in class 3 versus 1 was maximum for subjects dropped out at V8 and V13.

In the joint LCM, dropout profiles were no longer included as covariates, but the risk of dropout was modeled jointly using a proportional hazard model (8) as a function of the latent classes and covariates. Estimates from this proportional hazard model are displayed in Table 4. The risk of dropout was globally associated with the latent classes ($\chi^2 = 72.2$ for 2 df, $p < 0.0001$). More specifically, adjusted for age, sex and educational level, the risk of dropout was higher in class 2 (hazard ratio=2.98, $p < 0.01$) and, to a lesser extent, in class 3 compared to class 1 (hazard ratio=2.34, $p < 0.01$). Men ($p < 0.01$), subjects with low educational level ($p = 0.04$), and older subjects ($p < 0.01$) had a higher risk of dropout.

Estimates from the model for multivariate longitudinal data in the two

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

Table 3: Estimates from the multinomial logistic model for the class membership probabilities in the simple LCM.

Covariates	Estimate	Standard error	p-value
Probability to be in Class 2:			
Intercept	−2.08	0.41	< 0.001
Dropout at V5[1]	∞	−	−
Dropout at V8	2.21	0.45	< 0.001
Dropout at V10	1.74	0.40	< 0.001
Dropout at V13	1.74	0.31	< 0.001
Probability to be in Class 3:			
Intercept	−3.62	0.33	< 0.001
Dropout at V5	−0.75	14.82	0.960
Dropout at V8	2.50	0.46	< 0.001
Dropout at V10	0.93	0.68	0.177
Dropout at V13	1.66	0.40	< 0.001

[1] Reference: No dropout

Table 4: Estimates from the proportional hazard model for dropout in the joint LCM.

	Estimate	Standard error	p-value
Class 2[1]	1.094	0.157	< 0.001
Class 3[1]	0.852	0.239	< 0.001
Men[2]	0.432	0.075	< 0.001
No Diploma[3]	0.174	0.083	0.036
Age[4,5]			
70-75	0.258	0.147	0.079
75-80	0.483	0.151	< 0.001
$\geq$80	1.349	0.148	< 0.001

[1] Reference: Class 1
[2] Reference: Women
[3] Reference: Diploma
[4] Reference: 65-70 years
[5] Likelihood Ratio Test with 3 df significant at $\alpha = 0.001$

DOI: 10.2202/1557-4679.1088

LCM are displayed in Table 2. Regarding covariates of interest, conclusions of the PMM and the two LCM were qualitatively similar for the effect of age and educational level on initial level. However, there were large differences regarding the impact of educational level on cognitive decline: educational level was not associated with cognitive decline in the PMM ($\beta = -0.006, p = 0.49$) as under the MAR assumption ($\beta = -0.006, p = 0.56$), while both LCM revealed that subjects without diploma had a significantly sharper decline ($\beta = -0.016, p = 0.03$ in the simple LCM and $\beta = -0.018, p = 0.02$ in the joint LCM). Conclusions about sex effect also differed according to the analysis. The association between sex and initial cognitive level was not significant (except in the PMM where $\beta = 0.012, p = 0.01$ and the effect was close to significance for simple LCM ($\beta = 0.008, p = 0.006$)). Men exhibited a significantly slower decline under the MAR assumption ($\beta = 0.021, p = 0.01$) and in the PMM analyses ($\beta = 0.024, p < 0.01$) which was marginally significant in the simple LCM ($\beta = 0.012, p = 0.05$) but not significant in the joint LCM ($\beta = 0.008, p = 0.19$). In general, the two LCM led to similar estimates for the association with covariates except that the sex effect tended to be more pronounced in the simple LCM.

5.4 Description of the latent classes

As explained in section 4.4, regression parameters in LCM were adjusted for the latent classes. To investigate the differences between the latent classes and the dropout patterns, Figure 2 shows the mean evolution of the latent cognitive process for each dropout pattern predicted by the PMM and the mean evolution in each latent class estimated from the two LCM. In the PMM, there were only slight differences of evolution between dropout patterns, whereas in the LCM the three classes displayed dramatically different profiles. From both LCM, subjects in class 1 had a high initial cognitive level and a very slow decline over time. This class represented about 62.9% of the sample in the simple LCM and 69.1% in the joint LCM. Subjects in class 2 had a low initial cognitive level but their decline was hardly more pronounced. In the simple and joint LCM respectively, 33.4% and 27.2% of subjects were *a posteriori* classified in class 2. The initial cognitive level in class 3 was close to that in class 2, but these subjects exhibited a dramatic decline over time. With both LCM, 3.7% of subjects belonged to class 3. Figure 2 clearly shows that the latent classes exhibit much more heterogeneity than the dropout patterns.

Comparison of subject characteristics from the 5 dropout patterns (Table 1) with those of subjects *a posteriori* classified in the 3 classes by the two LCM (Table 5) clarifies the different evolution profiles. Chi-square tests

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

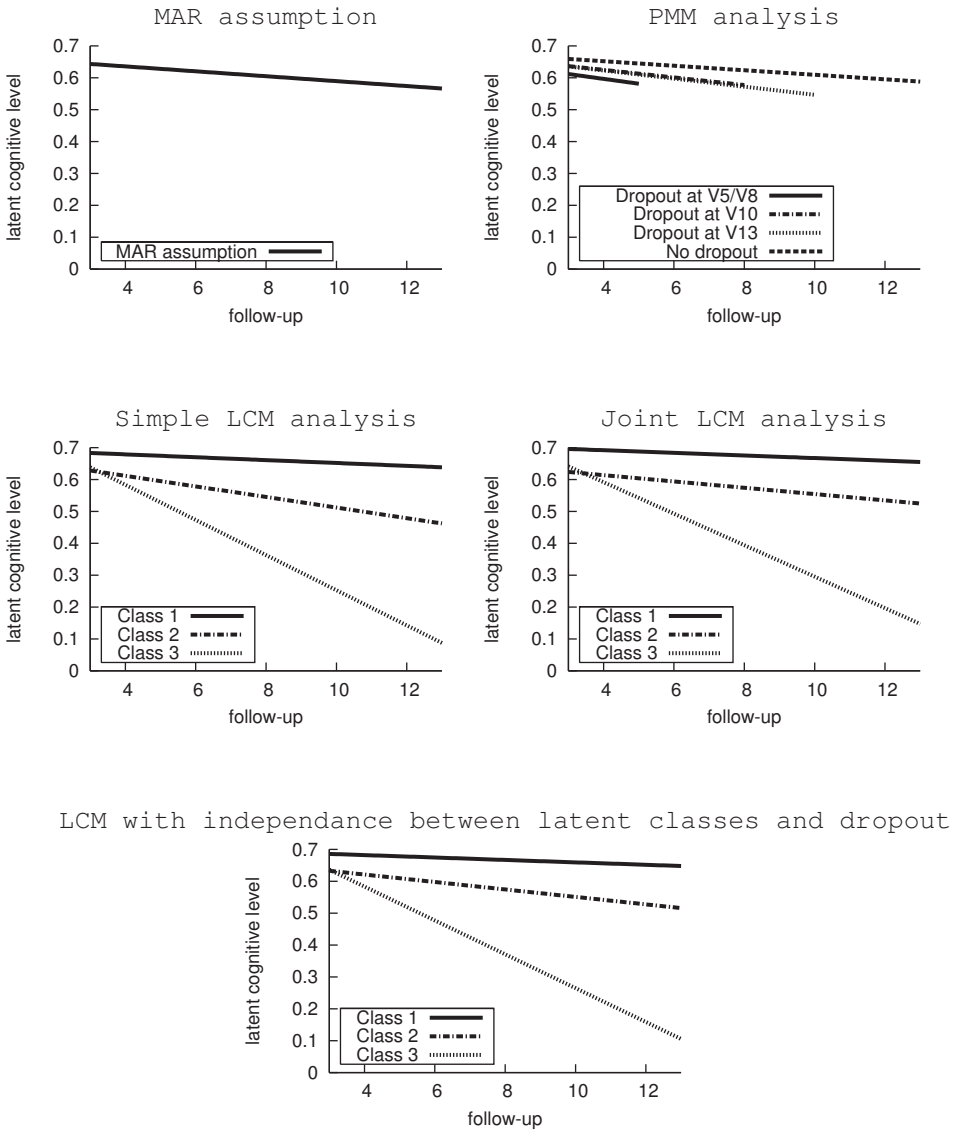

Figure 2: Estimated mean evolution of the latent process under the MAR assumption, given the dropout patterns by the PMM analysis, given the latent classes by the simple LCM, the joint LCM and the LCM with independence between the latent classes and the dropout for women with diploma and aged between 65 and 70 years-old.

DOI: 10.2202/1557-4679.1088

Table 5: Characteristics of the subjects according to the posterior latent classes for the simple LCM and the joint LCM (%).

	Class 1	Class 2	Class 3	p-value
Simple LCM	(N=1032)	(N=548)	(N=60)	
Dropout at				< 0.001
V5	0.0	43.0	0.0	
V8	11.7	21.2	31.7	
V10	10.7	10.6	8.3	
V13	14.5	16.6	33.3	
No dropout	63.1	8.6	26.7	
Demented	7.4	10.6	56.7	< 0.001
Women	58.4	54.7	75.0	0.009
Diploma	76.6	70.3	81.7	< 0.001
Joint LCM	(N=1023)	(N=556)	(N=61)	
Dropout at				< 0.001
V5	5.8	31.8	0.0	
V8	10.5	24.7	18.0	
V10	9.1	13.1	11.5	
V13	14.4	16.7	34.4	
No dropout	60.2	13.7	36.1	
Demented	7.4	10.3	57.4	< 0.001
Women	56.0	58.8	78.7	0.002
Diploma	76.7	70.3	80.3	0.012

showed that dementia diagnosis, sex, educational level and age were significantly associated with dropout patterns ($p < 0.01$ for each covariate). There was more dementia diagnosis among subjects with longer follow-up. Men, subjects without diploma and older subjects appeared to drop out earlier.

The posterior classification obtained with the simple LCM (upper part of Table 5) confirmed that class 1 corresponded to subjects with a long follow-up (63% of subjects did not drop out), while class 2 was associated with the shortest follow-up profile (64.2% dropped out at V5 or V8) and class 3 had a medium follow-up (nobody dropped at V5 but only 26.7% did not drop out). The percentage of positive diagnosis of dementia was relatively low in classes 1 and 2 but was superior to 50% in class 3. The sex distribution was similar for classes 1 and 2, but there was an imbalance in class 3 which included 75% of women. Although the difference was less glaring for diploma distribution, the percentage of qualified subjects was higher in class 3 (81%). Table 5 shows that the latent classes obtained by the two LCM had similar characteristics.

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

In addition, more than 87% of subjects were classified in the same class by the two LCM. Thus, the two latent class analyses highlight heterogeneous profiles of cognitive evolution in our sample that are only partially associated with the dropout patterns but highly associated with dementia diagnosis.

5.5 Conditional independence assumption

The joint latent class model is based on the assumption that cognitive evolution and dropout time are conditionally independent in view of the latent classes (Lin, 2004; Roy, 2007). This assumption was evaluated by estimating the association between cognitive scores and dropout time after adjustment for the posterior latent classes in the longitudinal model. The strength of association when adjusting for posterior classes was markedly reduced ($\chi^2 = 22.6$ (6 df) with adjustment versus $\chi^2 = 94.5$ (6 df) without), though it remained highly significant ($p < 0.001$). This suggested that the conditional independence assumption was not valid.

5.6 LCM assuming independence with dropout

We compared the latent classes obtained from the joint LCM with those obtained with the same model but assuming independence between latent classes and dropout ($\gamma_{0g} = 0,\ \forall g$ in (8)). The estimated cognitive evolutions in the 3 latent classes identified by the two models were close (Figure 2). Moreover, 70% of the subjects were classified in the same class by both models. Thus, the assumption regarding the link between cognitive decline and dropout had little impact on the latent classes in this data set. This result confirms that latent classes reflect the heterogeneity of cognitive evolution rather than the association between dropout and cognition.

5.7 Estimated Beta transformations

As suggested by a referee, the Beta CDF parameters could influence parameters of the longitudinal model. By comparing Beta CDF parameters in table 2, we observed slight differences between each approach. However, these differences had little impact on the shape of the estimated transformations as displayed in figure 3, so it was unlikely that they influenced estimates of regression parameters.

Nevertheless, to be sure that these slight differences did not explain the discrepancies between PMM and LCM due to covariate effects, we re-estimated

DOI: 10.2202/1557-4679.1088

the 3 models by handling informative dropouts, with the Beta CDF parameters set at the values estimated in the MAR analysis. Estimates were close to those obtained in table 2 while the differences between the 3 methods in the estimated sex and educational level effects remained (results not shown). Finally, we computed the estimated marginal evolution of the latent process and each cognitive score in its natural scale for men and women aged between 65 and 70 with high educational level. As the Beta transformation was not linear, we computed $E\{h_k^{-1}(\tilde{Y})|g\}$ for each dropout pattern or each latent class g by simulation, as explained in Proust-Lima et al (2006). Then, the marginal expectations were obtained as the mean of the group-specific expectation weighted by the proportion of each pattern or each class given the covariates (11). Figure 4 shows that the trend is similar for the latent process and for the cognitive scores on their natural scales.

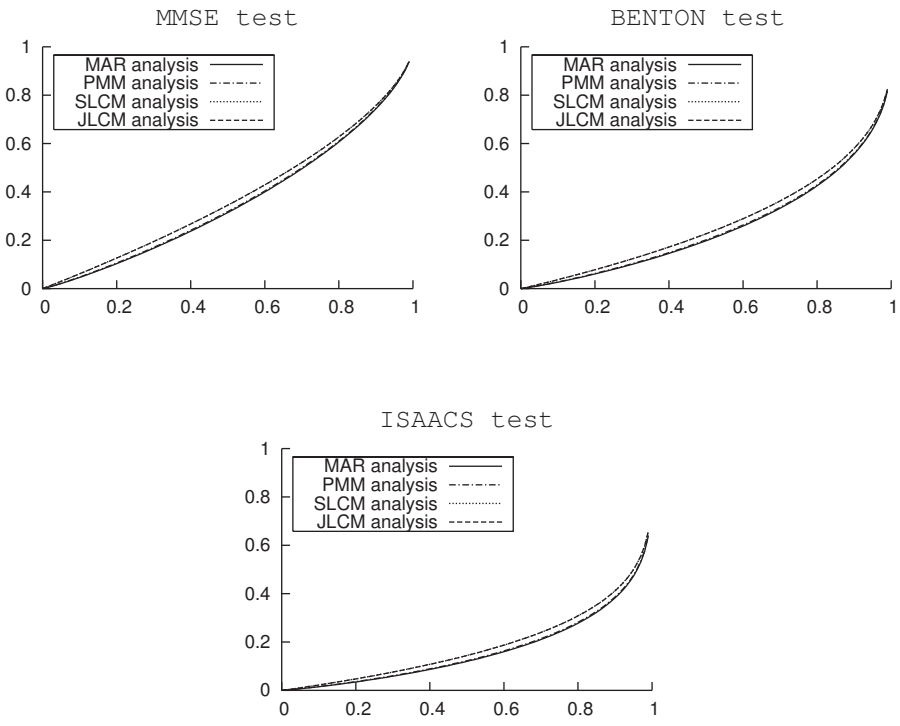

Figure 3: Beta distribution of each psychometric test under the MAR assumption and in the PMM, simple LCM and joint LCM.

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

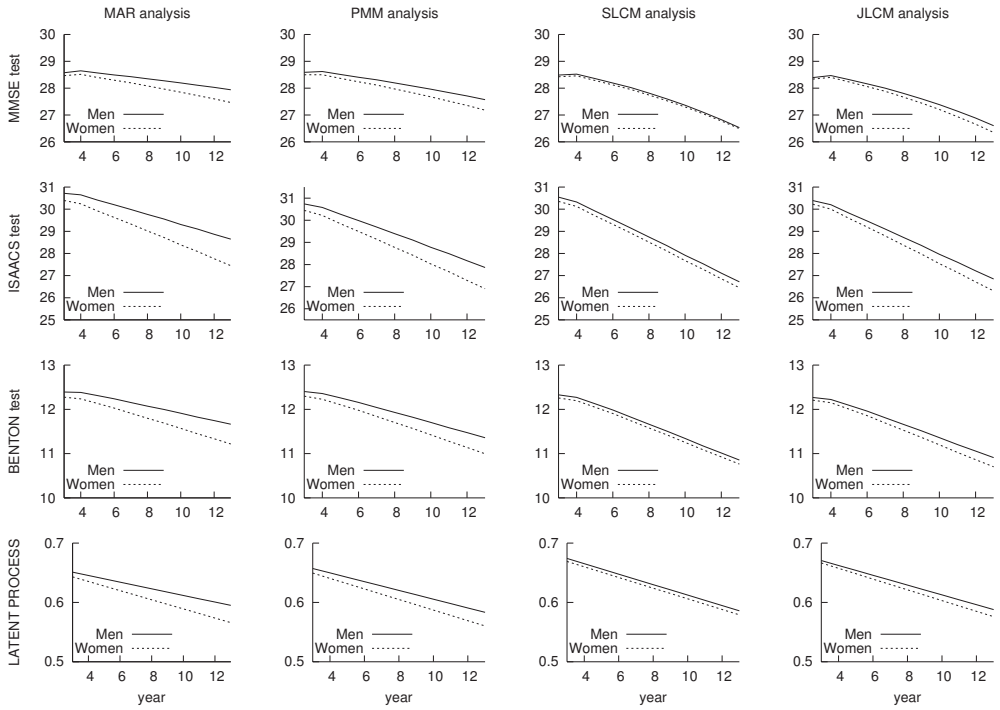

Figure 4: Comparison of men and women evolutions of the 3 psychometric tests and the latent evolution for each approach, for subjects with high educational level and aged between 65 and 70.

6 Discussion

By comparing three methods (PMM and two kinds of LCM) to account for informative dropout in a longitudinal study of cognitive aging, we show that LCM and PMM may lead to different parameter estimates due to different interpretations of the parameters. In the PMM, however the parameters are interpreted with adjustment for the dropout pattern. In the two LCM, the parameters are adjusted for the latent classes, which may reflect more than dropout patterns.

For instance, under the MAR assumption, we found that men tended to have a better initial cognitive level and a slower decline than women, and that these differences were reinforced when adjusting for dropout pattern in the PMM. Indeed, since the dropout rate was higher for men and for subjects with low cognitive level (see Table 1), the difference between men and women who dropped out at the same time tended to be higher than the unadjusted difference.

DOI: 10.2202/1557-4679.1088

On the other hand, as the latent class approaches provide parameters adjusted for the latent classes, it is essential to perform complementary analyses to investigate the characteristics of the latent classes in order to better understand the meaning of the estimated parameters. We suggest several analyses to describe the classes and evaluate the link with dropout patterns. In the Paquid data set, the three classes discriminated subjects who kept a high cognitive level over the 10 years of follow-up and had a low rate of dropout, subjects with a poor initial cognitive level who tended to drop out earlier from the study and, those with pathological decline. The latter group included more than 50% of subjects diagnosed as demented before the end of their follow-up and 75% of women. This result is in agreement with previous findings showing a higher risk of dementia among women in the oldest ages (Commenges, 1998). Thus, it is clear that these 3 classes do not only represent dropout patterns but are the consequence of other sources of heterogeneity in the data. In the joint LCM, when adjusting for these 3 classes and especially on the pathological class, which is highly associated with sex, we found no residual association between sex and cognitive level, suggesting no difference between men and women in normal cognitive aging. However, in the simple LCM, the residual sex effect was borderline significant. This may be due to the fact that the latent classes from the simple LCM are more strongly associated with the dropout patterns and slightly less with sex (see Table 1). Furthermore, when adjusting for the latent classes, subjects without diploma exhibited a higher rate of decline which was not observed in the PMM and MAR analyses.

The slight differences observed between the estimates from the simple LCM and the joint LCM were due to the adjustment for sex, educational level and age in the proportional hazard model of the latter. Without these adjustments, estimates were nearly identical to those from the simple LCM (results not shown). Indeed $\sum_C [Y|C,X][C|D][D] = \sum_C [Y|C,X][D|C][C]$. The minor differences arose from the use of a Cox model instead of a logistic one. Moreover, if covariates X are included both in the class membership probability and in the dropout probability, $\sum_C [Y|C,X][C|D,X][D|X] = \sum_C [Y|C,X][D|C,X][C|X]$. Parameters from the conditional distribution $[Y|C,X]$ have the same meaning (and similar values) in the two LCM, but their interpretation is unclear because X impacts both the class membership and the distribution of Y given the class. However, an interesting advantage of the joint model is that it allows adjustment for covariates when modeling dropout given the class. Indeed, conditionally on the covariates included in the mixed model, if data are missing at random, the risk of dropout does not depend on the outcome, data are missing at random.

It should be underlined that this work is based on data from a large ob-

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

servational cohort study including a representative sample of the elderly population. For instance, the sample includes both demented and non-demented subjects with a wide range of educational levels. This heterogeneity explains the large discrepancy observed between the estimates from the two approaches and highlights the different parameter interpretations. The latter may be partly hidden with more homogeneous data sets, such as those in clinical trials. Implementing investigative methods to deal with informative missing data in observational cohort studies is especially useful since missing data are more frequent in these studies.

In the Paquid cohort, some dropouts are due to death. We did not distinguish death from other sources of dropout. Indeed, separating death from dropout would have greatly increased the number of dropout patterns and created sparse patterns. Moreover, it is unclear how to classify subjects who missed a visit and died thereafter. A proper way to take death into account would be to jointly model time-to-dropout and time-to-death and then compute the estimated evolution given that the subject is alive. However this is beyond the score of this paper.

One could argue that interpretation of conditional parameters is not essential given that the main interest lies in marginal inference. However, in the litterature, parameters from the conditional distribution $[Y|D, X]$ or $[Y|C, X]$ are often considered as retaining a marginal interpretation if they are not group-specific. Section 4.4 shows that this is true only under restrictive assumptions and that, in most cases, computation of marginal parameters is difficult since it depends on the values of the other covariates. Moreover, computation of the variance of these marginal estimates may be untractable. For example, we computed the marginal gender effect on the initial level using formula (13). The marginal estimates were 0.009 for the PMM (close to the MAR estimate), 0.005 for the simple LCM and 0.004 for the joint LCM. The latter was identical to the conditional estimate as the model assumes that the class-membership probabilities do not depend on covariates, but this make the marginal estimate difficult to compare to the two others. To circumvent these problems, interesting hybrid approaches between selection models and PMM (Wilkins, 2006; Wilkins, 2007) or LCM (Roy, 2007) have recently been proposed, but they raise additional estimation problems.

Other important differences between the LCM and PMM concerns the underlying assumption about the missingness process. Indeed, to reach indentifiability in PMM, some *a priori* assumptions must be made. Typically, some dropout patterns with few measurements are assumed to follow a common evolution with other dropout patterns. One advantage of PMM is that several sets of constraints may be used to estimate several PMM in the framework of a

DOI: 10.2202/1557-4679.1088

sensitivity analysis. When averaging parameter estimates over the pattern to obtain marginal parameters, we need the additional assumption that the pre-dropout model remains valid after the dropout time within each pattern. The LCM avoid *a priori* identifiability constraints as the grouping of dropout patterns is data driven, but they rely on the conditional independence assumption between the time-to-dropout and the responses given the latent classes. That is an missing completely at random (MCAR) assumption within each class. This hypothesis of conditional independence against the alternative of residual dependence on the observed responses may be evaluated using the posterior classification, but it is obviously impossible to assess residual dependence on the unobserved response.

To conclude, we recommend the use of PMM rather than LCM to tackle MNAR data because the parameters are more simply interpreted and are easily implemented with existing software. However, to understand differences in parameter estimates between PMM and MAR analysis, the association between dropout patterns and covariates needs to be described. When data encompass many observed patterns or patterns with few measurements per subject, several pattern mixture analyses should be performed with different constraints on the parameters to ensure identifiability. f LCM are preferred in such a case, they require complementary analyses to characterize the classes and to avoid misinterpretation of the parameters. However, the two approaches share a common drawback compared to the selection models because the parameters of the marginal distribution of the response variable cannot be easily obtained. Despite this drawback, PMM are useful to perform sensitivity analyses for incomplete longitudinal data, whereas LCM should be used with caution in this context, particularly with heterogeneous data. However, as illustrated in the Paquid data set, the latent class models remain very interesting to explore heterogeneity in data by highlighting several profiles of evolution and by giving additional insight into the impact of covariates.

References

American Psychiatric Association. *Diagnostic and Statistical Manual of mental disorders, edition III Revised (DSM IIIR)*. Washington DC: American Psychiatric Association, 1987.

Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological methods* 2003; **8**: 338-63.

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

Benton A. *Manuel pour l'application du test de rétention visuelle. Applications cliniques et expérimentales.* Centre de psychologie appliquée. Paris 1965.

Beunckens C, Molenberghs G, Verbeke G, Mallinckrodt C. A latent-class mixture model for incomplete longitudinal gaussian data. *Biometrics* 2007; published online: 30-Jun-2007. doi: 10.1111/j.1541-0420.2007.00837.x

Commenges D, Letenneur L, Joly P, Alioum A, Dartigues JF. Modelling age-specific risk: application to dementia. *Statistics in medicine* 1998; **17**(17): 1973-1988.

Diggle PJ, Kenward MG. Informative dropouts in longitudinal data analysis (with discussion). *Applied statistics* 1994; **43**: 49-93.

Elias MF, Elias PK, D'Agostino RB, Silbershatz H, Wolf PA. Role of age, education, and gender on cognitive performance in the Framingham Heart Study: community-based norms. *Experimental aging research* 1997; **23**(3): 201-35.

Folstein MF, Folstein SE, Mc Hugh PR. Mini-Mental State. A practical method for grading the cognitive state of patients for the clinicians. *Journal of psychiatric Research* 1975; **12**: 189- 198.

Isaacs B, Kennie AT. The Set test as an aid to the detection of dementia in old people. *The British journal of psychiatry: the journal of mental science* 1973; **123**(575): 467-70.

Jacqmin-Gadda H, Fabrigoule C, Commenges D, Dartigues JF. A 5-year longitudinal study of the Mini Mental State Examination in normal aging. *American journal of epidemiology* 1997; **145**: 498-506.

Kenward MG. Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in medicine* 1998; **17**: 2723-2732.

Le Carret N, Lafont S, Mayo W, Fabrigoule C. The effect of education on cognitive performances and its implication for the constitution of the cognitive reserve. *Developmental neuropsychology* 2003; **23**(3): 317-37.

Letenneur L, Commenges D, Dartigues JF, Barberger-Gateau P. Incidence of

DOI: 10.2202/1557-4679.1088

dementia and Alzheimer's disease in elderly community residents of south-western France. *International journal of epidemiology* 1994; **23**: 1256-61.

Letenneur L, Proust-Lima C, Le Gouge A, Dartigues JF, Barberger-Gateau P. Flavonoid intake and cognitive decline over a 10 year period. *American journal of epidemiology* 2007; **165**: 1364-1371.

Lin H, Turnbull BW, Mc Culloch CE, Slate EH. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* 2002,**97**: 53-65.

Lin H, McCulloch CE, Rosenheck RA. Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* 2004; **60**(2): 295-305.

Little RJA, Rubin DB. *Statistical Analysis with Missing Data.* Wiley: New York, 1987.

Little RJA. Pattern-Mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association* 1993; **88**: 125-134.

Little RJA. Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association* 1995; **90**: 1112-1121.

Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM journal of applied mathematics* 1963; **11**: 431-41.

Molenberghs G, Michiels B, Kenward M, Diggle P. Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica* 1998; **52**: 153-161.

Proust C, Jacqmin-Gadda H, Taylor J, Ganiayre J, Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* 2006; **62**: 1014-1024.

Proust-Lima C, Joly P, Dartigues JF, Jacqmin-Gadda H. Joint modelling of multivariate longitudinal outcomes and time-to-event: a nonlinear latent class approach. Submitted for publication.

Redner RA, Walker HF. Mixture densities, maximum likelihood and the EM

The International Journal of Biostatistics, Vol. 4 [2008], Iss. 1, Art. 14

algorithm. *SIAM review. Society for Industrial and Applied Mathematics* 1984; **26**: 195-239.

Roy J. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* 2003; **59**(4): 829-36.

Roy J, Daniels MJ. A General Class of Pattern Mixture Models for Nonignorable Dropout with Many Possible Dropout Times. *Biometrics* 2007; in press.

Schwartz G. Estimating the dimension of a model. *The annals of statistics* 1978; **6**: 461-464.

Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-mixture models. *Biostatistics* 2002; **3**: 245-265.

Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data.* Springer Series in Statistics: New York, 2000.

Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward MG. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics* 2001; **57**(1): 7-14.

Wiederholt WC, Cahn D, Butters NM, Salmon DP, Kritz-Silverstein D, Barrett-Connor E. Effects of age, gender and education on selected neuropsychological tests in an elderly community cohort. *Journal of the American Geriatrics Society* 1993; **41**(6): 639-47.

Wilkins KJ , Fitzmaurice GM . A Hybrid Model for Nonignorable Dropout in Longitudinal Binary Responses *Biometrics* 2006; **62** (1): 168-176.

Wilkins KJ , Fitzmaurice GM . A marginalized pattern-mixture model for longitudinal binary data when nonresponse depends on unobserved responses *Biostatistics* 2007;**8** (2): 297-305.

DOI: 10.2202/1557-4679.1088

Chapitre 4

Modèle conjoint à état latent

4.1 Introduction

Le concept d'événement de démence suppose l'existence d'un temps de survenue de la démence dont on étudie la distribution. Cependant, l'installation de la démence est progressive et c'est une caractéristique essentielle de la physiopathologie du vieillissement cognitif des personnes âgées. L'approche que nous proposons induit une modélisation de l'hétérogénéité de la population face au risque de démence plus dynamique que l'approche par classes latentes utilisée dans le chapitre précédent. En effet, dans le modèle proposé par Proust-Lima et al. (2009), l'hétérogénéité de l'évolution cognitive est modélisée au travers de classes latentes présentant une évolution cognitive et un risque d'événement propre à chaque classe. Le modèle développé dans cette section est un modèle à état latent permettant de modéliser un déclin pré-diagnostique. Le sujet peut changer d'état au cours du temps. De plus, le décès étant associé à un déclin cognitif plus marqué, il est nécessaire de l'introduire dans la modélisation afin de tenir compte d'éventuels biais induits par cette censure informative.

Dans ce chapitre, nous présentons le modèle développé par Jacqmin-Gadda et al. (2006) pour l'étude conjointe de l'accélération du déclin cognitif et de la survenue d'une démence. Bien qu'il comporte certaines limites, il a servi de base au développement de l'approche que nous proposons. Nous définissons un modèle à état latent pour l'étude conjointe de l'évolution d'un processus, la survenue d'une maladie et la survenue du décès. Ce modèle a été développé dans le domaine du vieillissement cognitif et permet de considérer conjointe-

ment l'évolution cognitive et son accélération pré-démentielle, la survenue d'une démence et la survenue d'un décès. Dans un article soumis pour publication, nous présentons le modèle conjoint à état latent, son estimation, sa validation par l'intermédiaire de simulations ainsi qu'une application dans l'étude du vieillissement cognitif. A la suite de cet article, nous proposons une étude de simulation plus complète montrant notamment l'intérêt de la modélisation du décès dans la réduction des biais des paramètres du modèle. Nous présentons également une application sur l'étude de l'impact du niveau d'études complète celle de l'article. Nous terminons ce chapitre par une discussion sur les forces et les limites de ce modèle.

4.2 Génèse du modèle

Modèle à changement de pente aléatoire pour la modélisation conjointe du déclin cognitif et d'une démence

Hall et al. (2003) ont proposé un modèle à changement de pente aléatoire pour décrire le déclin cognitif de sujets déments sans modéliser conjointement le temps de survenue d'une démence. Les paramètres du modèle sont donc estimés avec des données ne contenant que des sujets diagnostiqués déments au cours du suivi. Jacqmin-Gadda et al. (2006) proposent comme extension un modèle conjoint pour l'étude de l'âge de survenue d'une démence et de l'accélération du déclin pré-diagnostique (cf. figure 4.1) afin de pouvoir comparer des déclins normaux et pathologiques en évitant les biais liés à la sélection des sujets dont le statut de démence est connu à une date de point.

Le score cognitif du sujet i à l'âge t, noté $Y_i(t)$ avec $Y_{ij} = Y_i(t_{ij})$, $j = 1, ..., n_i$ et $i = 1, ..., N$, est modélisé par un modèle mixte par morceaux avec une tendance linéaire avant l'accélération du déclin et polynomiale après :

$$Y_{ij} = (\mu_0 + u_{i0}) + (\mu_1 + u_{1i}) \times t_{ij} + \sum_{k=2}^{K} (\mu_k + u_{ki}) \times \{(t_{ij} - \tau_i)^+\}^k + \epsilon_{ij}$$

où $z^+ = 0$ si $z \leq 0$ et $z^+ = z$ si $z > 0$, ϵ_{ij} suit une loi normale $\mathcal{N}(0, \sigma_\epsilon^2)$ et le vecteur d'effets aléatoires $u_i = (u_{0i}, u_{1i}, ..., u_{Ki})'$ suit une loi normale $\mathcal{N}(0, G)$.

L'âge à l'accélération du déclin cognitif τ_i est supposé indépendant des effets aléatoires

u_i et suit une distribution log-normale :

$$\log(\tau_i) \sim \mathcal{N}(Z_{\tau i}\alpha_\tau, \sigma_\tau^2)$$

où $Z_{\tau i}$ est un vecteur de covariables spécifiques au sujet i.

Le modèle pour l'âge de survenue d'une démence T_i est défini de la manière suivante :

$$X_i = \log(T_i) = Z_{xi}\gamma + \eta\log(\tau_i) + \varepsilon_i$$

où $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ et Z_{xi} est un vecteur de covariables.

L'objectif est d'estimer le moment où l'évolution cognitive des sujets en phase de pré-démence se distingue de l'évolution des sujets normaux et d'étudier la forme du déclin cognitif en estimant l'évolution moyenne du marqueur, sachant l'âge à l'accélération du déclin $E(Y_i|\tau_i)$, ou sachant le temps de démence en fonction de variables explicatives $E(Y_i(t)|X_i)$. Ce modèle permet également d'estimer un âge médian de survenue d'une démence sachant l'âge au changement de pente :

$$\text{med}(T_i|\tau_i) = \tau_i^\eta \exp(Z_{xi}\gamma)$$

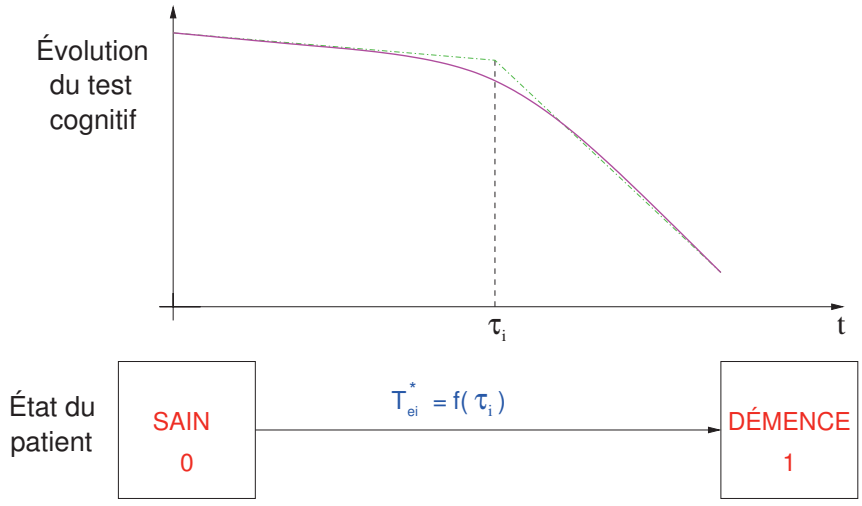

Fig. 4.1 : Modèle conjoint à changement de pente aléatoire pour l'évolution cognitive et l'âge de démence T_{ei}^* (Jacqmin-Gadda et al., 2006)

L'application de ce modèle sur les données de la cohorte Paquid a permis de montrer que l'évolution cognitive avant le changement de pente τ_i ne différait pas selon le niveau d'éducation des sujets et que le déclin des sujets de haut niveau d'études était plus marqué que celui des sujets de faible niveau d'éducation après τ_i. L'âge médian du changement de pente était également significativement différent selon le niveau d'études (90 ans pour les sujets de haut niveau d'études contre 70 ans chez les sujets de faible niveau d'études). Bien que l'âge médian de démence soit plus élevé chez les sujets de haut niveau d'études (94 ans) que chez les sujets de faible niveau (88 ans), l'analyse montre que, sachant l'âge d'accélération du déclin cognitif τ_i, le délai de survenue d'une démence est plus court chez les sujets de haut niveau d'études.

Ce modèle a plusieurs limites. Premièrement, il suppose que le risque de survenue d'une démence n'augmente pas une fois l'accélération du déclin entamé, ce qui semble peu réaliste dans le contexte du déclin cognitif. Enfin, suivant une autre hypothèse, le processus de censure de la démence est supposé non informatif et le processus de données manquantes ignorable lorsqu'il n'est pas dû à la démence. Or, cette hypothèse d'indépendance conditionnelle entre le temps de sortie d'étude et le temps de survenue d'une démence est invalide dans le contexte du vieillissement cognitif, d'où l'intérêt d'un modèle conjoint.

Premier modèle à état latent envisagé pour l'étude conjointe de l'évolution d'un processus, la survenue d'une maladie et la survenue du décès

Dans ce travail, nous proposons de tenir compte de la censure informative induite par le décès, afin de limiter les biais liés au phénomène de sélection de la population par le décès. Nous modélisons conjointement les temps de survenue de la démence, du décès (2 états observés) et de l'accélération du déclin cognitif (état latent) en utilisant un modèle multi-états. Le déclin cognitif est modélisé par un modèle linéaire mixte bi-phasique à changement de pente aléatoire (cf. figure 4.2).

L'évolution cognitive non linéaire $Y_{ij} = Y_i(t_{ij})$, $j = 1, ..., n_i$ et $i = 1, ..., N$, est modélisée par un modèle mixte à changement de pente aléatoire linéaire par morceaux avec un lissage entre les deux phases d'évolution tel que défini en section 2.2.4 avec une densité

multivariée gaussienne d'espérance :

$$E(Y_{ij}) = (\phi_0 + \alpha_0' X_{0i}) + (\phi_1 + \alpha_1' X_{1i})(t_{ij} - \tau) + (\phi_2 + \alpha_2' X_{2i})(t_{ij} - \tau) trn(t_{ij} - \tau; \gamma)$$

et de variance-covariance :

$$V_i = V(Y_i) = A_i G A_i' + \sigma_\epsilon^2 I_{n_i}$$

où G est la matrice de variance-covariance des effets aléatoires ($u_i \sim \mathcal{N}(0, G)$) et σ_ϵ^2 la variance de l'erreur.

L'âge d'entrée dans l'état latent τ_i correspond à l'âge d'accélération du déclin cognitif et est défini par un modèle des risques proportionnels avec une distribution de Weibull comme intensité de transition de base.

A la différence du modèle développé par Jacqmin-Gadda et al. (2006), ce n'est plus l'âge de survenue d'une démence qui est modélisé conjointement, mais le risque de survenue

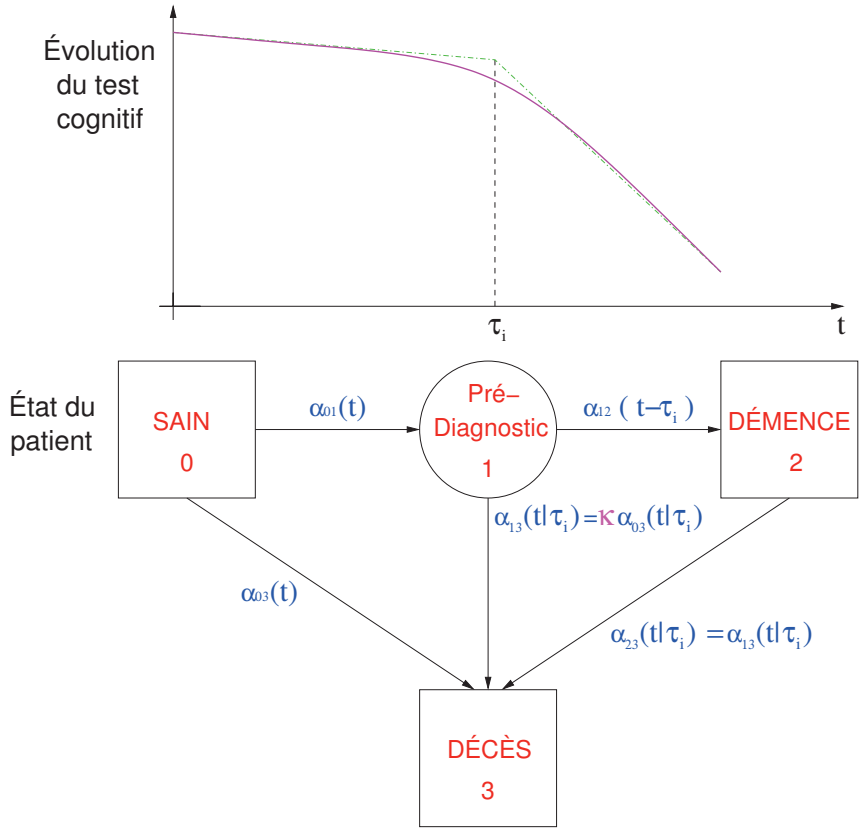

Fig. 4.2 : Modèle conjoint à état latent pour l'accélération du déclin cognitif, la survenue d'une démence et la survenue du décès (1ère version avec proportionnalité du risque de décès dans la seconde phase par l'intermédiaire du coefficient κ)

d'une démence. Le risque de démence est supposé nul avant la phase de déclin accéléré (état 0). Dans la phase de déclin pré-diagnostique (état 1), le risque de démence dépend du délai écoulé depuis l'entrée dans cet état, τ_i. La phase pré-diagnostique est donc un état transitoire obligatoire avant la démence. Le risque de démence est défini par un modèle des risques proportionnels avec une distribution de Weibull comme intensité de transition de base.

Le risque de survenue du décès est défini par un modèle des risques proportionnels avec une fonction exponentielle par morceaux comme intensité de transition de base. Nous considérons qu'il existe un risque de décès plus élevé pour un sujet en deuxième phase d'évolution, c'est-à-dire lorsque l'accélération du déclin cognitif est déjà commencée. L'intensité de transition est définie comme étant proportionnelle au fait d'être dans la seconde phase d'évolution (coefficient κ).

La procédure d'estimation des paramètres de ce modèle a été validée par simulation. En revanche, son utilisation sur les données réelles de la cohorte Paquid s'est avérée impossible. En effet, avec ce modèle, les estimations *a posteriori* des âges d'accélération du déclin cognitif ont montré que les âges de décès se situaient tous après l'accélération du déclin cognitif. Par conséquent, le paramètre mesurant l'excès de risque de décès dans l'état 1 et l'état 2 était non identifiable. Devant ce constat, nous avons cherché à modéliser différemment le risque de décès en adoptant une formulation plus souple. Dans le modèle finalement proposé, le risque de décès à un temps donné dépend du niveau cognitif courant. La formulation des fonctions de transition instantanées α_{01}, α_{12} et α_3 est donc identique car le risque de décès ne depend de τ_i qu'au travers du niveau cognitif courant $E(Y(t)|\tau_i)$ considéré comme une variable quantitative :

$$\alpha_{03}(t) = \alpha_{13}(t) = \alpha_{23}(t) = \alpha_3(t)\exp(\eta E(Y(t)|\tau_i))$$

Ce modèle est plus souple que le précédent car le risque de décès change continuellement avec l'évolution de la cognition et permet donc de s'abstraire de l'hypothèse peu raisonnable selon laquelle le risque de décès est constant dans les états 1 et 2 (ajusté sur l'âge).

4.3 Article

Résumé de l'article

Dans différentes maladies chroniques, l'état de santé d'un patient est suivi par des marqueurs quantitatifs. L'évolution est souvent caractérisée par un processus de dégradation en deux phases : l'évolution peut être normale, puis évoluer vers une forme pathologique précédant le diagnostic de maladie. Nous proposons un modèle conjoint à état latent pour la modélisation conjointe des mesures répétées d'un marqueur quantitatif, d'un temps de survenue de maladie et d'un temps de survenue de décès. En utilisant les données de la cohorte PAQUID, nous avons étudié conjointement l'évolution du déclin cognitif, le risque de démence et le risque de décès. Ce modèle permet d'estimer une évolution moyenne du score cognitif sachant l'âge de démence pour un sujet vivant et dément, l'évolution moyenne du score cognitif pour un sujet non-dément, ainsi que l'âge à l'accélération du déclin cognitif et la durée de la phase pré-démentielle.

Joint model with latent state for longitudinal and multi-state data

Etienne Dantan[1,2], Pierre Joly[1,2], Jean-François Dartigues[1,2] and Hélène Jacqmin-Gadda[1,2]

[1] INSERM, U897, Bordeaux, F-33000, France

[2] Université Victor Segalen Bordeaux 2, ISPED, Bordeaux, F-33000, France

Abstract: In many chronic disease, the patient health status is followed up by quantitative markers. The evolution is often characterised by a two-phase degradation process: the evolution can be normal and then evolves in a pathological form preceding the disease diagnosis. We proposed a joint multi-state model with latent state for the joint modelling of repeated measures of a quantitative marker, time-to-illness and time-to-death. Using data from the PAQUID cohort on cognitive aging of elderly people, we jointly studied the cognitive decline, dementia risk and death risk. We estimated the mean evolution of cognitive scores given age at dementia for subject alive and demented, the mean evolution of cognitive score for subject alive and non-demented, in addition to age at acceleration of the cognitive decline and duration of the pre-dementia phase.

Keywords: Cognitive aging; Joint model; Longitudinal data; Random changepoint; Survival model.

Corresponding author: Etienne Dantan, INSERM U897, ISPED, Université Bordeaux II, case 11, 146 rue Léo Saignat, 33076 Bordeaux cedex, France. Ph: (33) 5 57 57 11 36; Fax: (33) 5 56 24 00 81; e-mail: Etienne.Dantan@isped.u-bordeaux2.fr

1 Introduction

Many chronic diseases, such as cancer, HIV or Alzheimer's disease, present evolutions characterised by a long-term process often beginning before the disease diagnosis. The subjects go successively through several states from normal to severe disease and the disease diagnosis can be considered as one of this state. Moreover, patient health status may be followed up by one or several quantitative markers as, for instance, Prostate Specific Antigen for prostate cancer, CD4 T-cell counts for HIV or cognitive tests for Alzheimer's disease. Multi-state models (Hougaard, 1999) may be used to describe the development of such diseases (Commenges et al., 1999) but they do not allow to describe the time-course of the biomarkers. During the last 10 years, joint models have been developped to study the relationship between a longitudinal response process and a time-to-event (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000; Tsiatis and Davidian, 2004). They have been extensively applied in HIV infection (DeGruttola and Tu, 1994; Brown et al., 2005), in cancer (Law et al., 2002; Pauler and Finkelstein, 2002) and also in Alzheimer's disease (Proust-Lima et al., 2009). Some joint models have been proposed for a longitudinal biomarker and several time-to-events, either for recurrent events (Han et al., 2007) or for competing risks (Elashoff et al., 2008). However, at our knowledge, there is no attempt of combining multi-state model and mixed model for longitudinal data in order to study the links between the biomarker evolution and the risk of transition between the different states of the disease. The mixed Hidden Markov Model (Altman, 2007) combines a mixed model with latent states but it is designed to analyse only repeated measures of the marker and does not handle time-to-event data. The latent states are

defined exclusively by the value of the marker.

Among chronic diseases, Alzheimer's disease is characterised by a very long pre-diagnosis phase with a two-phase degradation process of cognitive functions (Amieva et al., 2008) and large inter-individual variability. The knowledge of the pre-diagnosis phase is a real public health challenge both for the understanding of the pathological process by identifying the successive emergence of clinical symptoms, and for early detection of subjects at high risk of Alzheimer's disease. Indeed, these subjects could be the target population for new treatments since post-diagnosis treatments have exhibited only a modest impact on clinical symptom evolution. Random changepoint mixed models have been used to describe this long-term pre-diagnosis cognitive decline (Hall et al., 2003; Dominicus et al., 2008). Jacqmin-Gadda et al. (2006) proposed a joint model with random changepoint for time-to-dementia and cognitive decline that distinguishs normal and pathological cognitive aging dealing with the right censoring of dementia and thus avoiding selection biases that arise when comparing two groups of subjects, normal and demented at a given time point. However, this model has two main limits.

Firstly, it does not take death into account. In most cohorts on aging, dementia diagnosis is assessed at the follow-up visits while death times are observed exactly. Subjects who died and are dementia free at their last visit are considered as censored at their last visit. Estimates can be biased as risk of death is higher among subjects with poor cognitive decline (Wilson et al., 2003) or among demented subjects (Joly et al., 2002). Very recently, Yu and Ghosh (2009) extended the Jacqmin-Gadda et al's model to model jointly dementia-free death considered as a competing event using a cure model approach. However, this model does not address informative censoring due to death since it assumes independence between time-to-death and cognitive level given covariates. The main difference between the two models is the assumption of a cured fraction with a null risk of dementia. Indeed, a multi-state model is required to take into account the dependence of

death risk on cognitive status.

Secondly, in these two models, the risk of dementia depends on the age at cognitive decline acceleration but it does not increase when the subject enters in the phase of accelerated decline. It is more realistic to assume the risk of dementia is null before the acceleration of the decline and then increases gradually. Indeed, one can considered the phase of accelerated decline before dementia as a latent state of the disease, that could be called pre-dementia or pre-diagnosis phase, and is only indirectly observed through measures of cognition. This transitional state could have a strong link with the syndrome called Mild Cognitive Impairment (MCI) which is defined by a memory impairment without dementia (Petersen et al., 1999) and largely discussed in the litterature (Dubois et al., 2007).

Thus, the objective of this paper is to propose a joint multi-state model with latent state for the joint modelling of repeated measures of a quantitative marker, time-to-illness and time-to-death. The age at the change of the marker evolution trend corresponds to the age at the entrance in the latent state: the pre-diagnosis phase. A mixed random effect model with a random changepoint was used to describe change over time of the marker. A proportional hazard model with age as time scale and marker evolution was defined for the death risk whereas the disease risk depended mainly on the time in the pre-diagnosis state. We introduce this joint model in section 2. The maximum likelihood estimation procedure is presented in section 3 whereas section 4 is devoted to simulations. In section 5, the model is applied to the PAQUID cohort, a french prospective cohort including 3777 subjects aged 65 years and older at baseline and followed up during 15 years with eight assessments of cognition. This analysis allows to estimate the mean trajectories of cognitive scores given age at dementia for subjects alive or given age at death, in addition to age at acceleration of the cognitive decline and duration of the pre-dementia phase

2 Joint Model

We consider a population of N subjects with a marker evolution that can be normal or evolve to a pathological shape preceding the disease diagnosis. The change over time is a continuous process with two phases and a smooth transition (Figure 1). We consider a multi-state model with three observed states (healthy, ill and died) and a latent transitional state called pre-diagnosis state that corresponds to the second phase of marker evolution. We denote τ_i the age of the subject i at entry in the pre-diagnosis phase (state 1).

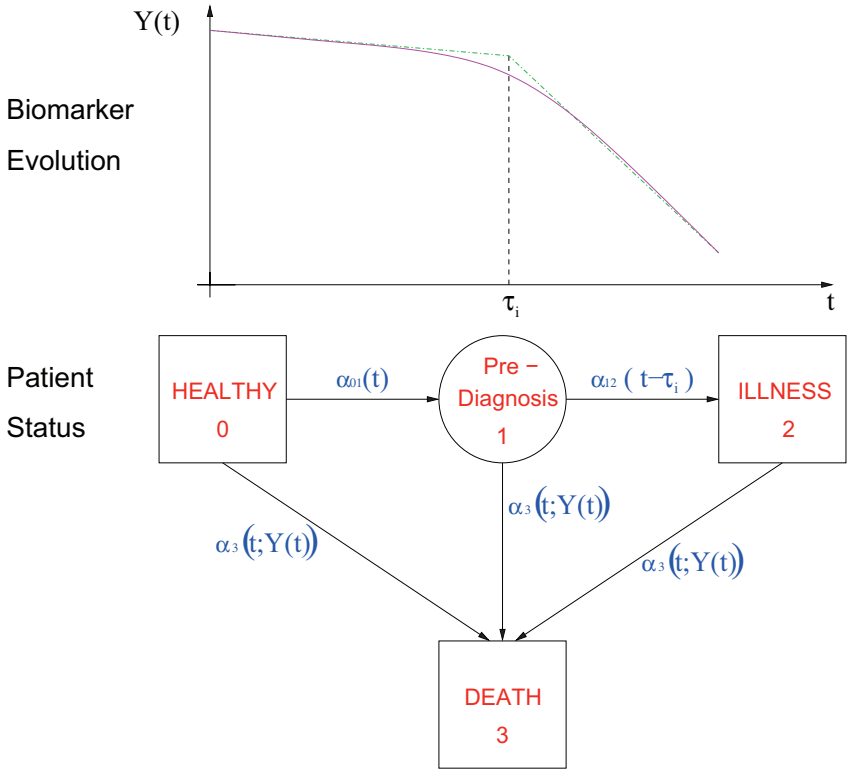

Figure 1: Joint model for the pre-diagnosis phase of disease

2.1 Marker model

Let $Y_i(t)$ be the marker value of subject i at age t. We note $Y_{ij} = Y_i(t_{ij})$ for $j = 1, ..., n_i$, and $i = 1, ..., N$. So, Y_i is the vector of the n_i measurements for subject i. The marker change over time is described by a segmented mixed model with a linear trend before and after the changepoint:

$$Y_{ij} = \begin{cases} b_{01i} + b_{11i}t_{ij} + \epsilon_{ij} & \text{if } t_{ij} \leq \tau_i \\ b_{02i} + b_{12i}t_{ij} + \epsilon_{ij} & \text{if } t_{ij} > \tau_i \end{cases}$$

with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2_\epsilon)$. To insure the continuity at the changepoint τ_i, we impose the constraint $b_{02i} = b_{01i} + b_{11i}\tau_i$. We adopt the parameter transformation recommanded by Bacon and Watts (1971):

$$\beta_{0i} = b_{01i} + b_{11i}\tau_i \qquad \beta_{1i} = \frac{b_{11i} + b_{12i}}{2} \qquad \beta_{2i} = \frac{b_{12i} - b_{11i}}{2}$$

The model can be re-written:

$$Y_{ij} = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)sgn(t_{ij} - \tau_i) + \epsilon_{ij}$$

with the function $sgn(t) = -1$ if $t < 0$ and $sgn(t) = 1$ if $t \geq 0$. Thus, β_{0i} corresponds to the mean marker value at the random changepoint, β_{1i} is the mean of the two slopes and β_{2i} is half the difference between the slopes. These coefficients are modeled as the sum of a mean effect that may depend on fixed covariates and an individual random effect:

$$\beta_{ki} = \phi_k + \alpha'_k X_{ki} + u_{ki} \tag{1}$$

for $k = 0, 1, 2$. The vector $u_i = (u_{0i}, u_{1i}, u_{2i})'$ of random effects is $N(0, G)$ and is independant of τ_i.

Finally, the function $sgn(t)$ is replaced by the function $trn(t; \gamma) = \frac{1}{t}\sqrt{t^2 + \gamma}$ (Seber and Wild, 2003) to insure a smooth transition between the two phases:

$$Y_{ij} = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)trn(t_{ij} - \tau_i; \gamma) + \epsilon_{ij} \tag{2}$$

This trn function allows the continuity of the first and second derivatives at τ_i with respect to all parameters in the mixed model. As $\lim_{\gamma \to 0} trn(t;\gamma) = sgn(t)$ for small γ, the expectation of Y may be approximated by:

$$E(Y_{ij}|\tau_i, u_i) \approx \widetilde{Y}_i(t_{ij}|\tau_i, u_i) = \begin{cases} \beta_{0i} + (\beta_{1i} - \beta_{2i})(t_{ij} - \tau_i) & \text{if } t_{ij} \leq \tau_i \\ \beta_{0i} + (\beta_{1i} + \beta_{2i})(t_{ij} - \tau_i) & \text{if } t_{ij} > \tau_i \end{cases} \tag{3}$$

As we assume a linear-linear trend for $Y(t)$, we choose a very small value for γ ($\gamma = 0.1$ in the following).

In (1), u_{1i} is a random effect on the mean slopes whereas u_{2i} is a random effect on half the difference of the two slopes. We also envisage another formulation below where $u_{s_1 i}$ is directly a random effect on the first slope and $u_{s_2 i}$ on the second slope:

$$\begin{aligned} \beta_{1i} &= \phi_1 + \alpha_1 X_{1i} + \frac{u_{s_1 i}}{2} + \frac{u_{s_2 i}}{2} \\ \beta_{2i} &= \phi_2 + \alpha_2 X_{2i} - \frac{u_{s_1 i}}{2} + \frac{u_{s_2 i}}{2} \end{aligned} \tag{4}$$

Thus, the random effect vector is $u_i = (u_{0i}, u_{s_1 i}, u_{s_2 i})'$.

2.2 Multi-state model

2.2.1 Latent state transition intensity

We assume a proportional hazard model for the transition from the state healthy (state 0) to the latent pre-diagnosis state (state 1):

$$\alpha_{01}(t) = \alpha_{01}^0(t) e^{\theta_\tau' X_\tau} \tag{5}$$

where $\alpha_{01}^0(t)$ is the baseline transition intensity and X_τ a q_τ-vector of covariates associated with a q_τ-vector of parameters θ_τ and the age t as time scale. The cumulative intensity transition to the latent state at time t is:

$$\Lambda_{01}(t) = \int_0^t \alpha_{01}(v) dv$$

2.2.2 Illness transition intensity

The transition intensity from the pre-diagnosis state (state 1) to the illness state (state 2) is assumed to follow a proportional hazard model depending on the time since entry in state 1 (τ_i) as time scale and possibly depending on random effects.

$$\alpha_{12}(t-\tau_i|u_i) = \alpha_{12}^0(t-\tau_i)e^{\theta_e' X_{ei}+\nu' u_i} \tag{6}$$

where $\alpha_{12}^0(t-\tau_i)$ is the baseline transition intensity, X_{ei} a q_e-vector of covariates associated with a q_e-vector of parameters θ_e, and ν a k-vector of parameters associated to random effects u_i. Thus, the cumulative transition intensity from state 1 to state 2 is:

$$\Lambda_{12}(t-\tau_i|u_i) = \int_0^{t-\tau_i} \alpha_{12}(v|u_i)dv$$

As we assume that direct transition from state 0 to state 2 is impossible, the instantaneous illness intensity function is defined as:

$$\alpha_e(t|\tau_i, u_i) = \mathbb{1}_{\{t>\tau_i\}}\alpha_{12}(t-\tau_i|u_i)$$

and the corresponding cumulative transition intensity as:

$$\begin{aligned}\Lambda_e(t|\tau_i, u_i) &= \int_0^t \alpha_e(v|\tau_i, u_i)dv \\ &= \mathbb{1}_{\{t>\tau_i\}}\Lambda_{12}(t-\tau_i|u_i)\end{aligned}$$

2.2.3 Death transition intensity

The death transition intensity is assumed to be independent of the patient state conditionally on the current expected biomarker value $\widetilde{Y}(t|\tau_i, u_i)$. This assumption is discussed in section 6. We assume that a proportional hazard model describes the death transition intensity as a function of age, $\widetilde{Y}(t|\tau_i, u_i)$ and a q_d-vector of covariates X_{di}:

$$\alpha_3(t|\tau_i, u_i) = \alpha_3^0(t)e^{\theta_d' X_{di}+\eta\widetilde{Y}_i(t|\tau_i,u_i)} \tag{7}$$

where $\widetilde{Y}_i(t|\tau_i, u_i)$ is defined in (3). The cumulative transition intensity to death is:

$$\Lambda_3(t|\tau_i, u_i) = \int_0^t \alpha_3(v|\tau_i, u_i)dv$$

3 Estimation

3.1 Log-likelihood

Let define (T_{di}, δ_{di}) where $T_{di} = min(T^*_{di}, C_{di})$ and $\delta_{di} = \mathbb{1}_{\{T^*_{di} < C_{di}\}}$ with T^*_{di} the death age and C_{di} the censoring age for death. Let define (T_{ei}, δ_{ei}) where $T_{ei} = min(T^*_{ei}, C_{ei}, T_{di})$ and $\delta_{ei} = \mathbb{1}_{\{\{T^*_{ei} < C_{ei}\} \& \{T^*_{ei} < T_{di}\}\}}$ with T^*_{ei} the age at the illness diagnosis and C_{ei} the censoring age for this event. Note that the censoring time for illness diagnosis may be different from the censoring time for death since a face to face interview is required for the disease diagnosis while age at death is exactly observed. Parameters are estimated using a maximum likelihood approach under the following additional assumptions:

1. Missing value of the marker before the disease or the death are missing at random
2. Censoring for death is not informative (in practice, this is often an administrative censoring corresponding to the end of the study)
3. Censoring for illness diagnosis when not due to death is not informative

We note θ the vector of all regression parameters in (2), (5), (6), (7) and variance-covariance parameters (G and σ^2_ϵ). The parameter γ of the function trn is fixed by the user to a small value. Given that the three outcomes, the marker Y_i, the illness age T_{ei} and the death age T_{di} are independent given τ_i and u_i, the individual contribution to the likelihood is developped as following:

$$
\begin{aligned}
L_i(\theta) &= L_i(Y_i, T_{ei}, \delta_{ei}, T_{di}, \delta_{di}; \theta) \\
&= \int_{-\infty}^{+\infty} \int_0^{+\infty} \alpha_e(T_{ei}|\tau_i, u_i)^{\delta_{ei}} e^{-\Lambda_e(T_{ei}|\tau_i, u_i)} \alpha_3(T_{di}|\tau_i, u_i)^{\delta_{di}} e^{-\Lambda_3(T_{di}|\tau_i, u_i)} \\
&\qquad \times f_Y(Y_i|\tau_i, u_i) f(\tau_i, u_i) d\tau_i du_i \qquad (8) \\
&= \int_{-\infty}^{+\infty} \int_0^{+\infty} [\mathbb{1}_{\{T_{ei} > \tau_i\}} \alpha_{12}(T_{ei} - \tau_i|u_i)]^{\delta_{ei}} e^{-\mathbb{1}_{\{T_{ei} > \tau_i\}} \Lambda_{12}(T_{ei} - \tau_i|u_i)} \\
&\qquad \times \alpha_3(T_{di}|\tau_i, u_i)^{\delta_{di}} e^{-\Lambda_3(T_{di}|\tau_i, u_i)} f_Y(Y_i|\tau_i, u_i) f_\tau(\tau_i) f_u(u_i) d\tau_i du_i
\end{aligned}
$$

where $f_u(u_i)$ is a multivariate Gaussian density with mean 0 and variance G and $f_\tau(\tau_i)$ is the density for τ_i defined as:

$$f_\tau(\tau_i) = \alpha_{01}(\tau_i)e^{-\Lambda_{01}(\tau_i)}$$

The function $f_Y(Y_i|\tau_i, u_i)$ is a multivariate Gaussian density with mean and variance given by:

$$E(Y_{ij}|\tau_i, u_i) = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)trn(t_{ij} - \tau_i; \gamma)$$

and

$$V_i = Var(Y_i|\tau_i, u_i) = \sigma^2_\epsilon I_{n_i}$$

This likelihood is developped in the Appendix using a multi-state model approach. In particular, it is demonstrated that the terms for possible unobserved transition to illness state between the last visit and death vanished since death only depends on the current marker value. Thus the main bias associated with interval censoring of illness diagnosis is addressed without complicating the likelihood.

Most studies include only subjects alive and not already diagnosed at the begining of the study. This induces left-truncation that must be taken into account in the likelihood by dividing the likelihood by the joint probability to be alive and not ill at the age of entry in the study T_{0i}:

$$l(\theta) = \log \prod_{i=1}^{N} \frac{L_i(Y_i, T_{ei}, \delta_{ei}, T_{di}, \delta_{di}; \theta)}{Pr(T^*_{ei} > T_{0i}, T^*_{di} > T_{0i})} \tag{9}$$

with

$$\begin{aligned} Pr(T^*_{ei} > T_{0i}, T^*_{di} > T_{0i}) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} & f_\tau(\tau_i) f_u(u_i) \\ & \times e^{-\mathbb{1}_{\{T_{0i} > \tau_i\}} \Lambda_{12}(T_{0i} - \tau_i | u_i)} \\ & \times e^{-\Lambda_3(T_{0i}|\tau_i, u_i)} d\tau_i du_i \end{aligned}$$

3.2 Estimation algorithm

We developped a `Fortran` program to estimate θ by a maximum likelihood approach using a Marquardt optimisation algorithm (Marquardt, 1963) which is a Newton-Raphson like

algorithm. As the log-likelihood integrals have not analytical solutions, they are computed by a Gauss-Hermite quadrature using the following approximation:

$$\int_{-\infty}^{+\infty} g(x)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \approx \sum_{q=1}^{r} g\left(\zeta^{(q)}\right) A^{(q)}$$

where r is the number of quadrature points, $\zeta^{(q)}$ are quadrature nodes and $A^{(q)}$ quadrature weights. To constrain some parameters to be positive (σ_ϵ^2 and parameters of baseline intensity functions), we estimated the square-root of these parameters and the Cholesky decomposition of the covariance matrix G.

4 Simulation study

We performed a simulation study to investigate the behaviour of the maximum likelihood estimators. To limit computation time, data were simulated with a model including only a random intercept in addition to the random changepoint τ_i and no covariate. One hundred samples of 500 subjects were simulated. For each subject, we simulated the random intercept u_{0i}, a pre-diagnosis age τ_i following the transition intensity $\alpha_{01}(t)$, the entry age in the study T_{0i} according to a uniform distribution on [65-75], the censoring age for death event $C_{di} = T_{0i} + 25$ which is an administrative censoring corresponding to 25 years of follow-up, the age to death T^*_{di} according to the transition intensity $\alpha_3(t|\tau_i, u_{0i})$, the censoring age for illness event C_{ei} according a uniform distribution on [75-100] and the illness age T^*_{ei} according to the transition intensity $\alpha_{12}(t-\tau_i)$. Then, we defined times of measurement: $t_{i0} = T_{0i}$ and $t_{ij} = min(t_{i(j-1)} + 3, T_{ei})$ while $t_{ij} < T_{ei}$ and $t_{ij} < T_{di}$. We also generated the gaussian error ϵ_i to calculate the vector of measures Y_i.

Parameter values for the mixed model and for the transition intensity function from state 0 to state 1 were chosen from model estimations on the Paquid subjects who have reached the first French primary diploma (Cerificat d'Etude Primaire, CEP).

$$Y_i(t) = (\phi_0 + u_{0i}) + \phi_1(t - \tau_i) + \phi_2\sqrt{(t-\tau_i)^2 + \gamma} + \epsilon_i$$

with $\phi_0 = 10.66$, $\phi_1 = -0.40$, $\phi_2 = -0.33$

$u_{0i} \sim (0, \sigma^2_{u_{0i}} = 1.13^2)$, $\epsilon_i \sim N(0, \sigma^2_\epsilon = 1.55^2)$

and $\gamma = 0.1$

$$\alpha_{01}(t) = \lambda_\tau \kappa_\tau (\kappa_\tau t)^{\lambda_\tau - 1}$$

with $\lambda_\tau = 3.98^2, \kappa_\tau = 0.11^2$

To mimic Paquid data, the death and dementia parameters were adjusted to obtain samples with nearly 60% of death and 15% of dementia diagnosis. The transition intensity from state 1 to state 2 were defined as Weibull distributions and the baseline intensity function for death was a stepwise function with 5 steps.

$$\alpha_{12}(t - \tau_i) = \lambda_e \kappa_e (\kappa_e (t - \tau_i))^{\lambda_e - 1}$$

with $\lambda_e = 1.38^2, \kappa_e = 0.46^2$

$$\alpha_3(t|\tau_i, u_i) = a_l e^{\eta \widetilde{Y}_i(t|\tau_i, u_i)} \text{ if } T_l \leq t < T_{l+1}, l = 1, .., 5$$

with $\eta = -0.17$, $a_1 = 0.15^2$, $a_2 = 0.20^2$,

with $a_3 = 0.30^2$, $a_4 = 0.50^2$ and $a_5 = 0.55^2$

with $T_1 = 65$, $T_2 = 75$, $T_3 = 80$, $T_4 = 85$ and $T_5 = 90$

The numerical integrals were calculated with a Gauss-Hermite quadrature with 20 points. The algorithm did not converge for 6 over the 100 replicates. Results are presented in Table 1. They showed good behaviour of the estimations in term of bias and asymptotic standard error except for an underestimation of the variance of the parameter $\sqrt{\lambda_\tau}$ from the Weibull distribution for the entry age in the pre-diagnosis phase. However, this has a modest impact on the shape of the Weibull distribution (not shown) and on its mean and standard error: the 95% confidence interval for the mean of this Weibull distribution computed with the empirical variances of λ_τ and κ_τ is [84.51-87.34] versus [85.04-86.82] when computed with the asymptotic variances. Maybe an alternative

parametrization could be useful. Simulation results were not improved by increasing the number of quadrature points while considerably increasing the computation time. At the inverse, a lower order of the quadrature did not give satisfying results.

Table 1: Simulated Value (SV), Mean estimate (ME), Relative Bias (RB), Asymptotical Standard Error (ASE) and Empirical Standard Error (ESE) for 100 replications of the joint model of cognitive evolution, dementia and death with a sample of 500 subjects

Parameters	SV	ME	RB(%)	ASE	ESE
ϕ_0	10.66	10.70	0.4	0.125	0.159
ϕ_1	-0.40	-0.37	-6.3	0.026	0.025
ϕ_2	-0.33	-0.31	7.1	0.026	0.025
σ_ϵ	1.55	1.56	0.8	0.028	0.027
$\sqrt{\lambda_e}$	1.38	1.44	4.6	0.121	0.148
$\sqrt{\kappa_e}$	0.46	0.44	-3.3	0.023	0.028
$\sqrt{a_1}$	0.15	0.16	9.6	0.037	0.038
$\sqrt{a_2}$	0.20	0.22	7.6	0.039	0.039
$\sqrt{a_3}$	0.30	0.33	9.4	0.044	0.048
$\sqrt{a_4}$	0.50	0.54	8.8	0.058	0.059
$\sqrt{a_5}$	0.55	0.59	6.6	0.057	0.059
η	-0.17	-0.19	9.5	0.023	0.026
$\sqrt{\lambda_\tau}$	3.98	4.09	2.9	0.072	0.155
$\sqrt{\kappa_\tau}$	0.11	0.11	0.2	0.0001	0.0001
$\sigma_{u_{0i}}$	1.13	1.13	0.3	0.062	0.071

5 Application

5.1 The PAQUID dataset

The PAQUID cohort is a french prospective study on cognitive aging including 3777 subjects aged 65 years and older and living at home at the initial visit. The detailed methodology of the PAQUID study has been previously described (Dartigues et al., 1992). Subjects were initially interviewed at home in 1988 and the 1, 3, 5, 8, 10, 13, and 15 years later. Measurements at the initial visit were excluded because of the first passing effect previously described by Jacqmin-Gadda et al. (1997). The time basis used for the analyses is the age. Dementia diagnosis is assessed at each visit by the investigator psychologist and then confirmed by clinical examination by a neurologist if the subjects were screened positive by the psychologist. The dementia age is computed as the mean between the age at dementia diagnosis and the age at the last visit without dementia. Age of censoring for dementia was the age at the last follow-up visit for the subject. As we are interested in the pre-diagnosis cognitive decline and the rate of missing data for cognitive measures increases after dementia diagnosis, the cognitive measures collected after the diagnosis were excluded. As subjects were not demented at baseline, the left-truncation has to be taken into account. Vital status was collected on every participant (including refusals and dropouts) thanks to the phone contact and, if necessary, death age was obtained from the practionner. Death age is censored at the end of follow-up, that is age at entry plus 15.

The psychometric test considered in this analysis is the Benton Visual Retention Test (Benton, 1965) which evaluates the visual memory with scores range from 0 to 15. We focus on this analysis on subjects with CEP as previous analyses have shown that a 2 phases linear-linear mixed model well fitted the pre-diagnosis cognitive decline in this population (Jacqmin-Gadda et al. 2006; Amieva et al., 2008). The sample includes 2396 subjects with CEP and non demented at the initial visit. The mean age of entry in the

study is 74.48 years and the mean age of dropout whatever the reason is 82.14 years. The mean number of measurements is 2.88. During the follow-up, 365 subjects were diagnosed as demented with a mean dementia age of 85.28 years (SE=5.92) and 1437 subjects died with a mean death age of 84.69 years (SE=7.06).

5.2 Model

The aim of the analysis was to study the caracteristics of the pre-dementia period. Different models based on the joint model presented in section 2 were compared. First, we chose the distributions for the baseline intensity functions. The baseline transition intensity from state 0 to state 1 and from state 1 to state 2 were defined by Weibull distributions. The baseline intensity for death was a stepwise function with 7 steps (every 5 years). Five random effect structures are compared in Table 2. Due to computation time and convergence difficulty, we restrained to models with three random effects, all including the age of entry in the pre-diagnosis phase, τ_i, and a random effect on the score level at τ_i, β_{0i} (model 1). Models 2 to 5 included a third random effect either on the mean slope (model 2), or on the difference between the slopes (model 3), or on the slope in the second phase (model 4 and 5). In model 5, the dementia transition intensity depended on the random slope in the second phase. We retained Model 4 that had the better AIC.

Finally, the best model retained is defined as:

$$
\begin{aligned}
Y_i(t) &= \beta_{0i} + \beta_1(t-\tau_i) + \beta_{2i}\sqrt{(t-\tau_i)^2+\gamma} + \epsilon_i \\
&\quad \text{with } \beta_{0i} = \phi_0 + u_{0i}\ ,\ \ \beta_{1i} = \phi_1 + \frac{u_{s_2 i}}{2} \\
&\quad \text{and } \beta_{2i} = \phi_2 + \frac{u_{s_2 i}}{2} \text{ and } \gamma = 0.1 \\
\alpha_{01}(t) &= \lambda_\tau \kappa_\tau (\kappa_\tau t)^{\lambda_\tau - 1} \\
&\quad \text{with } \kappa_\tau > 0,\ \lambda_\tau > 0 \\
\alpha_{12}(t-\tau_i) &= \lambda_e \kappa_e (\kappa_e (t-\tau_i))^{\lambda_e - 1} \\
&\quad \text{with } \kappa_e > 0,\ \lambda_e > 0 \\
\alpha_3(t|\tau_i, u_i) &= a_l e^{\eta \widetilde{Y}_i(t|\tau_i, u_i)} \text{ if } T_l \leq t < T_{l+1}, l = 1, .., 7 \\
&\quad \text{with } a_l > 0,\ T_1 = 65,\ T_2 = 70,\ T_3 = 75,\ T_4 = 80, \\
&\qquad \text{and } T_5 = 85,\ T_6 = 90 \text{ and } T_7 = 95
\end{aligned}
$$

Table 2: Random effect structure selection using AIC

Model	Random effects	$\nu^{\ddagger}$	Likelihood	Number of parameters	AIC
1	u^*_{0i}	–	−20808.97	17	41651.94
2	u^*_{0i}, u^*_{1i}	–	−20796.06	19	41630.12
3	u^*_{0i}, u^*_{2i}	–	−20799.10	19	41636.20
4	$u^*_{0i}, u^{\dagger}_{s_2 i}$	–	−20777.15	19	41592.30
5	$u^*_{0i}, u^{\dagger}_{s_2 i}$	✓	−20776.48	20	41592.96

* Defined by the equation 1

† Defined by the equation 4

‡ Defined by the equation 6

5.3 Results

Estimates for the joint model 4 are presented in Table 3. The mean Benton score at entrance in the pre-diagnosis state is equal to 10.66 (SE=0.07). During the healthy phase, the Benton score decreases by 0.073 points each year whereas it decreases by 0.82 points each year in the pre-diagnosis phase. The expected age at entry in the prediagnosis phase was 86.38 years (SE=6.70). Figure 2 displays the estimated mean scores $E(Y(t))$ given age.

The expected transition time between the pre-diagnosis state and the diagnosis state is 4.82 years (SE=2.59). The estimated transition intensities from healthy state to pre-dementia state and from pre-dementia state to dementia state are presented in Figure 3. Figure 3 also displays the estimated transition intensity for the death conditionnally on τ_i. The different values used for τ_i are the median (87.26 years) and also 65 years to represent a subject always in the prediagnosis phase during the follow-up and 100 years to represent a subject who remains in the healthy phase.

For comparison, the estimates of the bivariate model for the cognitive scores and dementia assuming independence with death are also presented in table 3. The mean Benton score at entry in the pre-diagnosis state is estimated to 10.77 whereas the slopes were -0.068 points each year in the healthy phase and -0.66 points each year in the pre-diagnosis phase. The expected age at entry in the prediagnosis phase was 86.59 years (SE=6.95) and the estimated transition time between state 1 and the state 2 is 5.36 years (SE=2.79). Thus, the estimates from the bivariate model are close to those of the complete model except an underestimation of the rate of decline in the pre-diagnosis phase and a slight overestimation of the delay time between the state 1 and state 2. This is in agreement with the previously reported steeper cognitive decline in late pre-death period (Wilson et al., 2003).

Table 3: Estimates, standard errors (SE) for the complete model (CM) and the bivariate model (BM) from PAQUID data (N=2396)

Parameters	CM		BM	
	Estimates	SE	Estimates	SE
ϕ_0	10.656	0.073	10.772	0.071
ϕ_1	-0.448	0.022	-0.364	0.020
ϕ_2	-0.375	0.022	-0.296	0.020
$\sqrt{\lambda_e}$	1.419	0.058	1.417	0.060
$\sqrt{\kappa_e}$	0.429	0.011	0.407	0.011
$\sqrt{a_1}$	0.366	0.044	–	–
$\sqrt{a_2}$	0.355	0.035	–	–
$\sqrt{a_3}$	0.443	0.038	–	–
$\sqrt{a_4}$	0.549	0.043	–	–
$\sqrt{a_5}$	0.690	0.049	–	–
$\sqrt{a_6}$	0.845	0.058	–	–
$\sqrt{a_7}$	0.935	0.073	–	–
η	-0.162	0.015	–	–
$\sqrt{\lambda_\tau}$	3.980	0.040	3.908	0.043
$\sqrt{\kappa_\tau}$	0.106	0.0001	0.106	0.0001
σ_1^a	1.292	0.040	1.299	0.039
σ_2^a	0.081	0.046	0.074	0.038
σ_3^a	0.333	0.032	0.355	0.032

[a] Variance-covariance parameters ($\sigma^2_{u_{0i}}$,$\sigma_{u_{02i}}$,$\sigma^2_{u_{1i}}$) were replaced by the corresponding Cholesky parameters (σ_1,σ_2,σ_3)

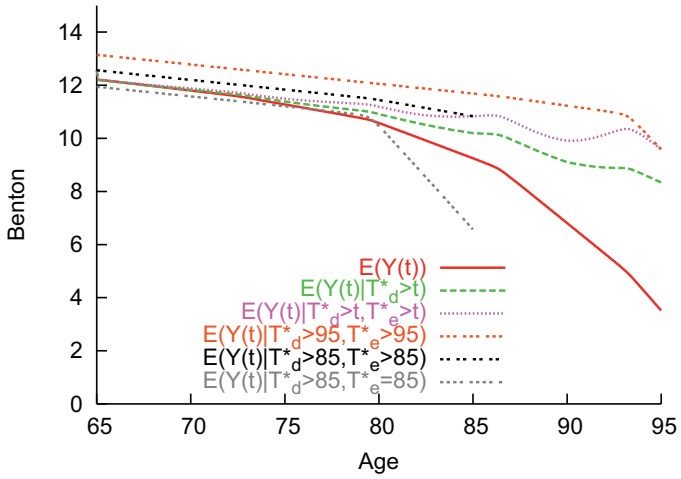

Figure 2: Marginal mean scores given age (thick plain line) and several mean trajectories given information on dementia age and death age

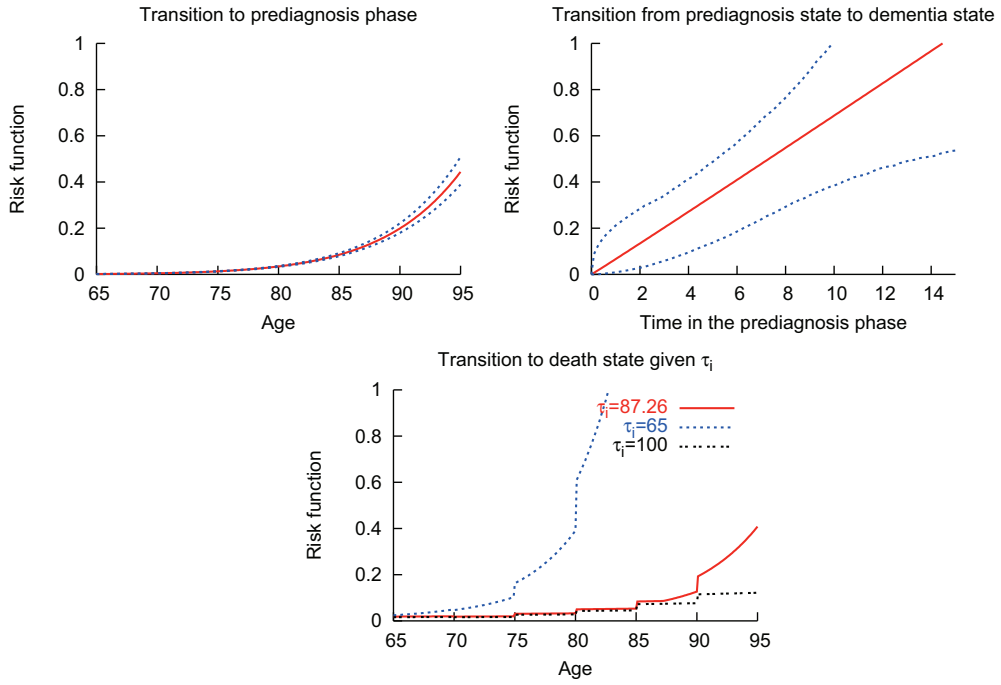

Figure 3: Instantaneous risk functions for the transitions from state 0 to state 1, from state 1 to state 2 and for death conditionally to τ_i values

One of the main advantage of this model is to allow estimation of mean score trajectories conditionnally on death and/or dementia ages. Figure 2 displays the estimated means of $Y(t)$ given that the subject is alive at age t ($E(Y(t)|T_d^* > t)$), and given that the subject is alive and non-demented at age t ($E(Y(t)|T_d^* > t, T_e^* > t)$) obtained with the following formulas:

$$\begin{aligned} E(Y(t)|T_d^* > t) &= \int_{-\infty}^{+\infty}\int_{0}^{+\infty} E(Y(t)|\tau_i, u_i) \\ &\qquad \times f(\tau_i, u_i|T_d^* > t)d\tau_i du_i \\ &= \int_{-\infty}^{+\infty}\int_{0}^{+\infty} E(Y(t)|\tau_i, u_i) \\ &\qquad \frac{e^{-\Lambda_3(t|\tau_i,u_i)}f_\tau(\tau_i)f_u(u_i)}{\int_{-\infty}^{+\infty}\int_{0}^{+\infty} e^{-\Lambda_3(t|\tau_i,u_i)}f_\tau(\tau_i)f_u(u_i)d\tau_i du_i} d\tau_i du_i \end{aligned} \tag{10}$$

$$\begin{aligned} E(Y(t)|T_d^* > t, T_e^* > t) &= \int_{-\infty}^{+\infty}\int_{0}^{+\infty} E(Y(t|\tau_i, u_i)) \\ &\qquad \times f(\tau_i, u_i|T_d^* > t, T_e^* > t)d\tau_i du_i \\ &= \int_{-\infty}^{+\infty}\int_{0}^{+\infty} E(Y(t|\tau_i, u_i))f_\tau(\tau_i)f_u(u_i) \\ &\qquad \times \frac{e^{-\Lambda_3(t|\tau_i,u_i)-\mathbb{1}_{\{T_{ei}>\tau_i\}}\Lambda_{12}(T_{ei}-\tau_i|u_i)}}{P(T_d^* > t, T_e^* > t)} d\tau_i du_i \end{aligned} \tag{11}$$

where

$$\begin{aligned} P(T_d^* > t, T_e^* > t) &= \int_{-\infty}^{+\infty}\int_{0}^{+\infty} e^{-\mathbb{1}_{\{T_{ei}>\tau_i\}}\Lambda_{12}(T_{ei}-\tau_i|u_i)} \\ &\qquad \times e^{-\Lambda_3(t|\tau_i,u_i)}f_\tau(\tau_i)f_u(u_i)d\tau_i du_i \end{aligned}$$

We also computed the expected trajectories for a subject alive and demented at age 85 years ($E(Y(t)|T_d^* > 85, T_e^* = 85)$), and for a subject alive and dementia free at 85 and 95 years ($E(Y(t)|T_d^* > 85, T_e^* > 85)$, $E(Y(t)|T_d^* > 95, T_e^* > 95)$)(Figure 2).

5.4 Goodness-of-fit

We evaluated separately the fit of the model for the three outcomes (cognitive scores, age at dementia and death). First, for non-demented and demented subjects, we compared

the means of the posterior score expectations $\hat{Y}(t) = E(Y(t)|\hat{\tau}_i, \hat{u}_i)$ and the means of observed cognitive scores using 5-year age intervals (Figure 4). The random effect predictions $\hat{\tau}_i, \hat{u}_i$ were computed by the mode of their posterior distribution given the data $Y_i, T_{ei}, \delta_{ei}, T_{di}, \delta_{di}$. Moreover, Figure 5 displays the observed data and individual trajectories estimated by $E(Y(t)|\hat{\tau}_i, \hat{u}_i)$ for 15 selected subjects. These figures show that the model well captures the different evolutions of demented and non-demented subjects and the individual evolution shapes.

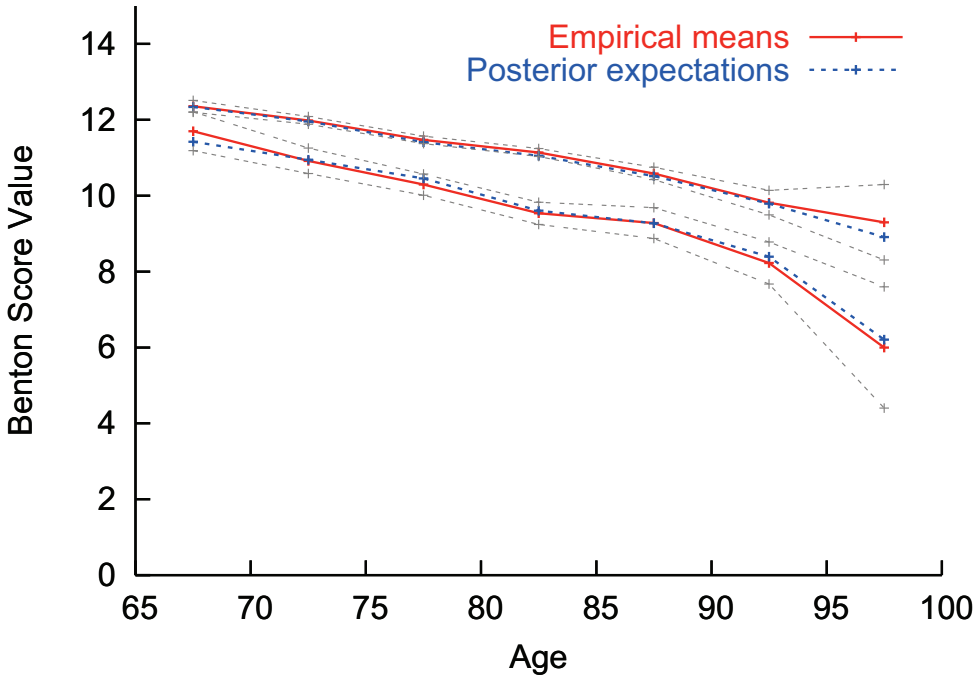

Figure 4: Posterior evolution compared to empirical evolution for non-demented subjects (upper curve) and demented subjects (lower curve)

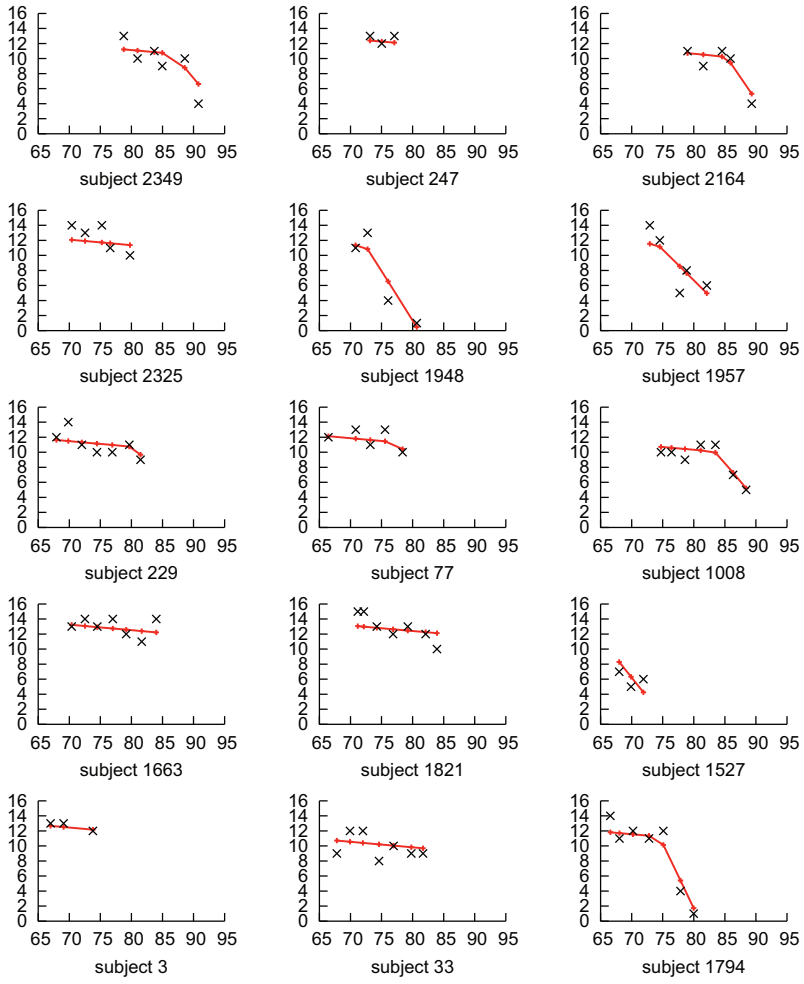

Figure 5: Individual fit of the Benton Score evolution for 15 subjects

To evaluate the fit of death data, we compared the marginal survival function from the joint model with the death survival function estimated by

$$S_d(t) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} e^{-\Lambda_3(t|\tau_i,u_i)} f_\tau(\tau_i) f_u(u_i) d\tau_i du_i \tag{12}$$

with a non parametric Proportional Hazard Model in a Penalized Likelihood approach (PHMPL) (Joly et al., 1999) (Figure 6).

To check the parametric assumption regarding baseline transition intensities, it is not possible to compare the estimates for the risk of dementia obtained from the complete model with those obtained with PHMPL because PHMPL considered died subjects as censored at their last visit. Thus, we compare PHMPL estimates of the dementia survival function with the same function estimated by the bivariate model (Figure 6) where died subjects are treated as censored, using the following formula:

$$\begin{aligned} S_e(t) &= 1 - P(T_e^* < t) \\ &= 1 - \int_{-\infty}^{+\infty} f_u(u_i) \Bigg\{ \int_0^t e^{-\Lambda_{01}(\tau_i)} \alpha_{01}(\tau_i) \\ &\quad \times \int_{\tau_i}^t e^{\mathbb{1}_{\{v>\tau_i\}} \Lambda_{12}(v-\tau_i)} \alpha_{12}(v-\tau_i) dv d\tau_i \Bigg\} du_i \end{aligned} \tag{13}$$

6 Discussion

We have developped a joint model with latent state for a longitudinal process and illness-death data. Such data are frequent in the monitoring of chronic diseases. This model allows to describe the pre-diagnosis phase of disease while limiting bias in parameter estimation. The application emphasized that the model is particularly well designed for the study of the cognitive decline in the pre-dementia phase. It may be viewed as an improvement of the random changepoint model proposed by Jacqmin-Gadda et al. (2006) with the aim to handle informative right censoring due to death and better account for the succession of event in time. In the context of cognitive aging, some authors considered that MCI is a transitional state between healthy and dementia without reversibility (Petersen et

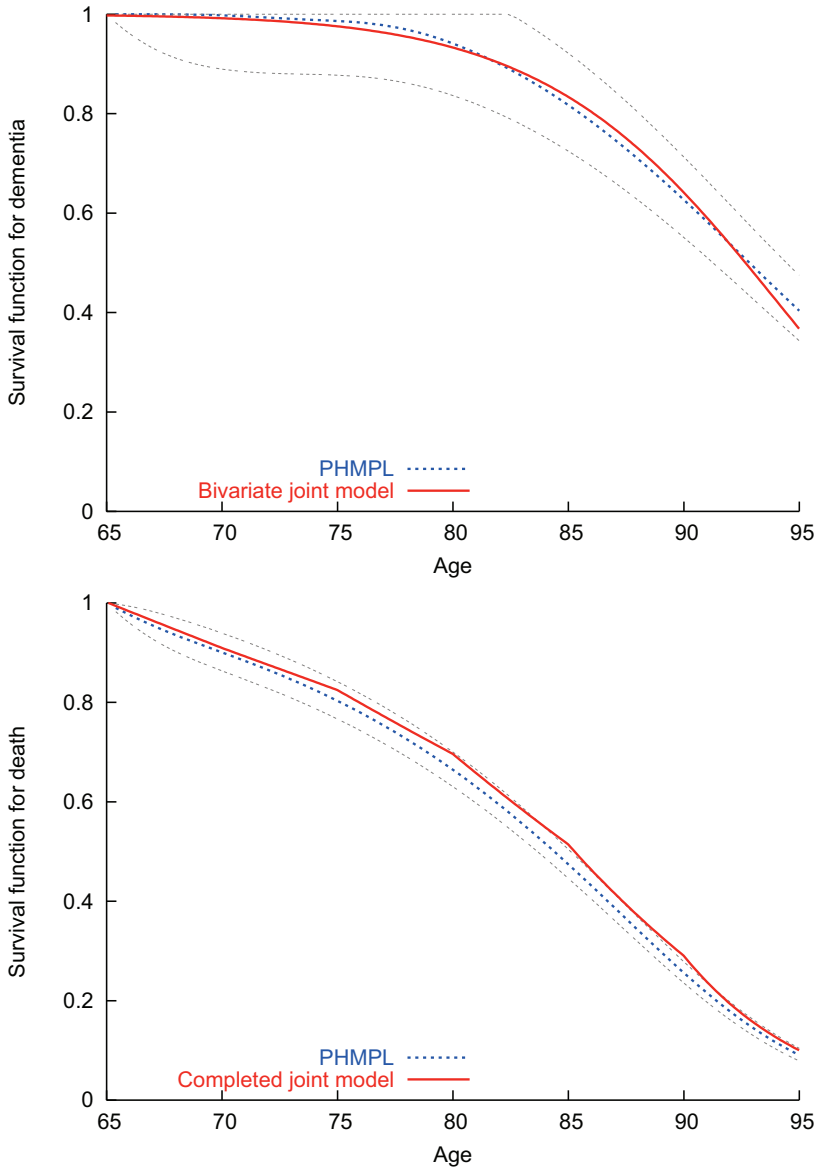

Figure 6: Marginal survival functions for dementia using the bivariate joint model and for death using the complete joint model compared to corresponding PHMPL survival functions

al., 1999) while others thinks it is a risk factor for dementia and may be reversible (Larrieu et al., 2002; Ritchie et al., 2001). Another subject of divergence in the litterature regards the tests and the thresholds used to define the MCI. In this work, the pre-diagnosis phase corresponds to the first concept of MCI and we think it is more sensible to define it by a change in the slope of decline instead of a threshold on a cognitive score as the cognitive levels are highly variable in the population beyond any pathological process.

The application was focused on the population with CEP because our previous work have shown that their pre-diagnosis cognitive evolution exhibited a clear phase of accelerated decline well approximated by a segmented linear-linear model, while the change of slopes was less clear in less educated subjects (see figure 1 in Jacqmin-Gadda et al. (2006)). The estimated age at entry in the accelerated decline phase was later in the previous model (90.3 versus 87.3 years) but the meaning of this phase is different in the two models. Here, this is an obligatory transition state before dementia (with a null risk of dementia before this state) while, in Jacqmin-Gadda et al. (2006) and Yu and Ghosh (2009), dementia may arise before τ_i and the risk of dementia did not increase with the time spend in the second phase. It could be interesting to test the existence of this pre-diagnosis phase of accelerated decline in different populations, for various cognitive tests or according to the type of dementia. However, the development of such a test would require to model more flexibly the dependence between the risk of dementia and the cognitive scores. Comparison with an alternative model allowing direct transition from the healthy state to dementia could also be interesting. The estimation of such extended models could nevertheless be limited by numerical difficulties.

In the proposed joint model, the transition intensities from the 3 transcient states to the death state were identical after adjustement on the current cognitive score. With this model, we demonstrated in the Appendix that the likelihood accounts for possible unobserved transitions to dementia between the last visit and death without any complication.

This is important because this is the most serious bias due to intermittent observations. An alternative model would be to include 3 state-specific transition intensity to death. However, this would seriously complicate the likelihood due to interval censoring of dementia and the model could not be estimated if no death is observed among healthy subjects that seems to be the case in the PAQUID cohort. Moreover, in our opinion, a model with 3 state-specific transition intensities is not more flexible than dependence on the quantitative cognitive score that continuously change with time.

While we have checked the parametric assumptions of the model for the Paquid data, it would be useful to relax some of them. For instance, a cure model allowing a sub-population to have a null risk of dementia could be investigated as in Yu and Ghosh (2009). Alternatively, flexible distribution, such as the semi-nonparametric distribution of Gallant and Nychka (1987), would be interesting for the age at entry in the pre-diagnosis phase since the assumption regarding this distribution can not be evaluated by comparison with observed data. Once again, such improvements are limited by numerical difficulties. Parametric approaches have advantages but require careful evaluation of the fit. We may also envisage a more efficient estimation algorithm.

Bibliography

Altman, R. (2007). Mixed Hidden Markov Models: An extension of the Hidden Markov Modeel to the Longitudinal Data Setting. *Journal of the American Statistical Association* **102,** 201-210.

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., and Dartigues, J. F. (2008). Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Annals of Neurology* **64,** 492-498.

Bacon, D. W., and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58,** 525–534.

Benton, A. L. (1965) Manuel pour l'application du Test de Retention Visuelle. Applications cliniques et expérimentales, 2ème édition française. Paris: Centre de Psychologie Appliquée.

Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible B-spline model for longitudinal biomarkers and Survival. *Biometrics* **61,** 64–73.

Commenges, D., Letenneur, L., Joly, P., Alioum, A., and Dartigues, J. F. Modelling age-specific risk : application to dementia. *Statistics In Medicine* **17,** 1973–1988.

Dartigues, J. F., Gagnon, M., Barberger-Gateau, P., Letenneur, L., Commenges, D., Sauvel, C., Michel, P., and Salamon, R. (1992). The Paquid epdemiological program on brain aging. *Neuroepidemiology* **11,** 14–8.

DeGruttola, V., and Tu, X. M. (1994). Modelling Progression of CD4-Lymphocyte counts and its relationship to survival time. *Biometrics* **50,** 1003–1014.

Dominicus, A., Ripatti, S., Pedersen, N. L., and Palmgren, J. (2008). A random change point model for assessing variability in repeated measures of cognitive function. *Statistics In Medicine* **27,** 5786–5798.

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* **6,** 734–46.

Elashoff, R. M., Li, G., and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64,** 762–771.

Gallant, A. R., and Nychka, D. W. (1987). Semi-Nonparametric Maximum Likelihood Estimation. *Econometrika* **55,** 363–390.

Hall, C. B., Ying, J., Kuo, L., Lipton, R. B. (2003). Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics*

and Data Analysis **42,** 91–109.

Han, J., Slate, E. H., and Peña, E. A. (2007). Parametric latent class joint model for a longitudinal biomarker an recurrent events. *Statistics In Medicine* **26,** 5285–5302.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1,** 465–480.

Hougaard, P. (1999). Multi-state models. A review. *Life Time Data Analysis* **5,** 239–264.

Jacqmin-Gadda, H., Commenges, D., and Dartigues, J. F. (2006). Random change-point model for joint modeling of cognitive decline and dementia. *Biometrics* **62,** 254–260.

Jacqmin-Gadda H, Fabrigoule C, Commenges D and Dartigues JF (1997). A five year longitudinal study of mini-mental state examination in normal aging. *American Journal of Epidemiology* **145,** 498–506.

Joly, P., Commenges, D., Helmer, C., and Letenneur, L. (2002). A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* **3,** 433–43.

Joly, P., Letenneur, L., ALioum, A., and Commenges, D. (1999). PHMPL: a computer program for hazard estimation using a penalised likeihood method with interval-censored and left-truncated data. *Computer Methods and Programs in Biomedicine* **60,** 225–231.

Larrieu, S., Letenneur, L., Orgogozo, J. M. Fabrigoule, C., Amieva, H. Le Carret, N., Barberger-Gateau, P., and Dartigues J. F. (2002). *Neurology* **59,** 1594–1599.

Law, N. J., Taylor, J. M. G., and Sandler, H. (2002). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics* **3,** 547–563.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* **11,** 431–41.

Pauler, D. K., and Finkelstein, D. M. (2002). Predicting time to prostate cancer

recurrence based on joint models fo non-linear longitudinal biomarkers and event time outcomes. *Statistics In Medicine* **21,** 3897–3911.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* **56,** 303–8.

Proust-Lima, C., Joly, P., Dartigues, J. F., and Jacqmin-Gadda, H. (2009). Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics and Data Analysis* **53,** 1142–1154.

Ritchie, K., Artero, S., Touchon, J. (2001) Classification criteria for mild cognitive impairment:a population-based validation study. *Neurology* **56,** 37–42.

Seber, G. A. F., and Wild, C. J. (2003). *Non linear regression,*. Wiley series in probability and mathematical statistics.

Tsiatis, A. A., and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14,** 809-834.

Wilson, R. S., Beckett, L. A., Bienas, J. L., Evans, D. A., and Bennett, D. A. (2003). Terminal decline in cognitive function. *Neurology* **60,** 1782–1787.

Wulfsohn, M. S., and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53,** 330–339.

Yu B. and Ghosh P. (2009). Joint modeling for cognitive trajectory and risk of dementia in the presence of death *Biometrics*, doi:10.1111/j.1541-0420.2009.01261.x

Appendix

Likelihood formulation in a multi-state approach

In section 3, we defined (T_{ei}, δ_{ei}) where $T_{ei} = min(T^*_{ei}, C_{ei})$ and $\delta_{ei} = \mathbb{1}_{\{T^*_{ei} < C_{ei}\}}$ with T^*_{ei} the age at the illness diagnosis and C_{ei} the censoring age for this event. We also defined (T_{di}, δ_{di}) where $T_{di} = min(T^*_{di}, C_{di})$ and $\delta_{di} = \mathbb{1}_{\{T^*_{di} < C_{di}\}}$ with T^*_{di} the death age and C_{di}

the censoring age for death. The individual contribution to the likelihood of subject i may be developped in a multi-state model approach in summing the contribution for all possible trajectories. L_1 is the individual contribution for a subject observed as demented, whereas L_2 is the individual contribution for a subject free of dementia at last visit. The individual contribution for a demented subject is defined by the transition from state 0 to 1 then from state 1 to 2 and eventually transition from state 2 to 3. Given the conditional independence between Y_i, T_{di} and T_{ei} given τ_i and u_i the random effect vector, L_1 may be developped as following:

$$\begin{aligned}
L_1 = \int_{-\infty}^{+\infty} f_u(u) \Bigg\{ \int_0^{T_{ei}} & f_Y(Y_i|\tau,u) e^{-\Lambda_{01}(\tau)-\Lambda_3(\tau|\tau,u)} \\
& \times \alpha_{01}(\tau) e^{-\Lambda_{12}(T_{ei}-\tau|u)-\Lambda_3(T_{ei}|\tau,u)+\Lambda_3(\tau|\tau,u)} \\
& \times \alpha_{12}(T_{ei}-\tau|u) e^{-\Lambda_3(T_{di}|\tau,u)+\Lambda_3(T_{ei}|\tau,u)} \\
& \times \alpha_3(T_{di}|\tau,u)^{\delta_{di}} d\tau \Bigg\} du \\
= \int_{-\infty}^{+\infty} f_u(u) \Bigg\{ \int_0^{T_{ei}} & f_Y(Y_i|\tau,u) e^{-\Lambda_{01}(\tau)} \alpha_{01}(\tau) \\
& \times e^{-\Lambda_{12}(T_{ei}-\tau|u)} \alpha_{12}(T_{ei}-\tau|u) \\
& \times e^{-\Lambda_3(T_{di}|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} d\tau \Bigg\} du \\
= \int_{-\infty}^{+\infty} \int_0^{+\infty} & f_Y(Y_i|\tau,u) f_\tau(\tau) f_u(u) \\
& \times [\mathbb{1}_{\{T_{ei}>\tau\}} \alpha_{12}(T_{ei}-\tau|u)] \\
& \times e^{-\mathbb{1}_{\{T_{ei}>\tau\}} \Lambda_{12}(T_{ei}-\tau|u)} \\
& \times \alpha_3(T_{di}|\tau,u)^{\delta_{di}} e^{-\Lambda_3(T_{di}|\tau,u)} d\tau du
\end{aligned}$$

The individual contribution L_2 for a subject non-observed as demented is the sum of the contribution for several possible trajectories. Such a subject may remain in state 0 until the time of last information on death status, T_{di}, when, according to δ_{di}, the subject transits or not to state 3 (2 last lines in L_2). The subject may transit to state 1 before T_{di} and remains in state 1 until T_{ei} (2 first lines in L_2) and then either he remains in state 1

until T_{di} or he may go to the dementia state between T_{ei} and T_{di} (without being observed) with possible final transition from state 2 to state 3 (5 next lines in L_2).

$$
\begin{aligned}
L_2 = \int_{-\infty}^{+\infty} f_u(u) \Bigg\{ & \int_0^{T_{di}} f_Y(Y_i|\tau,u) e^{-\Lambda_{01}(\tau)-\Lambda_3(\tau|\tau,u)} \alpha_{01}(\tau) \\
& \times e^{-\mathbb{1}_{\{T_{ei}>\tau\}}\Lambda_{12}(T_{ei}-\tau|u)-\Lambda_3(T_{ei}|\tau,u)+\Lambda_3(\tau|\tau,u)} \\
& \times \Bigg[e^{-\Lambda_3(T_{di}|\tau,u)+\Lambda_3(T_{ei}|\tau,u)-\Lambda_{12}(T_{di}-\tau|u)} \\
& \quad \times e^{\mathbb{1}_{\{T_{ei}>\tau\}}\Lambda_{12}(T_{ei}-\tau|u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \\
& \quad + \int_{T_{ei}}^{T_{di}} e^{-\Lambda_3(v|\tau,u)+\Lambda_3(T_{ei}|\tau,u)-\mathbb{1}_{\{v>\tau\}}\Lambda_{12}(v-\tau|u)} \\
& \qquad \times e^{\mathbb{1}_{\{T_{ei}>\tau\}}\Lambda_{12}(T_{ei}-\tau|u)} \mathbb{1}_{\{v>\tau\}} \alpha_{12}(v-\tau|u) \\
& \qquad \times e^{-\Lambda_3(T_{di}|\tau,u)+\Lambda_3(v|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} dv \Bigg] d\tau \\
& + \int_{T_{di}}^{+\infty} f_Y(Y_i|\tau,u) e^{-\Lambda_3(T_{di}|\tau,u)-\Lambda_{01}(\tau)} \\
& \qquad \times \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \alpha_{01}(\tau) d\tau \Bigg\} du
\end{aligned}
$$

As the risk of death is independent from the health status given u_i and τ_i, the following calculation shows that the integral on the age at dementia between T_{ei} and T_{di} vanishs.

Thus, it is not necessary with this model to distinguish these last 2 trajectories:

$$\begin{aligned}
L_2 = \int_{-\infty}^{+\infty} \Bigg\{ \int_0^{T_{di}} & f_Y(Y_i|\tau,u) e^{-\Lambda_{01}(\tau)} \alpha_{01}(\tau) e^{-\Lambda_3(T_{di}|\tau,u)} \\
& \times \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \Bigg[e^{-\Lambda_{12}(T_{di}-\tau|u)} + \\
& \int_{T_{ei}}^{T_{di}} e^{-\mathbb{1}_{\{v>\tau\}}\Lambda_{12}(v-\tau|u)} \mathbb{1}_{\{v>\tau\}} \alpha_{12}(v-\tau|u) dv \Bigg] d\tau \\
& + \int_{T_{di}}^{+\infty} f_Y(Y_i|\tau,u) e^{-\Lambda_3(T_{di}|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \\
& \times e^{-\Lambda_{01}(\tau)} \alpha_{01}(\tau) d\tau \Bigg\} f_u(u) du \\
= \int_{-\infty}^{+\infty} \Bigg\{ \int_0^{T_{di}} & f_Y(Y_i|\tau,u) f_\tau(\tau) e^{-\Lambda_3(T_{di}|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \\
& \times \Big[e^{-\Lambda_{12}(T_{di}-\tau|u)} + e^{-\mathbb{1}_{\{T_{ei}>\tau\}}\Lambda_{12}(T_{ei}-\tau|u)} \\
& - e^{-\Lambda_{12}(T_{di}-\tau|u)} \Big] d\tau \\
& + \int_{T_{di}}^{+\infty} f_Y(Y_i|\tau,u) e^{-\Lambda_3(T_{di}|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \\
& \times f_\tau(\tau) d\tau \Bigg\} f_u(u) du \\
= \int_{-\infty}^{+\infty} \Bigg\{ \int_0^{+\infty} & f_Y(Y_i|\tau,u) e^{-\Lambda_3(T_{di}|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \\
& \times \left[\mathbb{1}_{\{T_{di}>\tau\}} e^{-\mathbb{1}_{\{T_{ei}>\tau\}}\Lambda_{12}(T_{ei}-\tau|u)} + \mathbb{1}_{\{T_{di}\leq\tau\}} \right] \\
& \times f_\tau(\tau) d\tau \Bigg\} f_u(u) du
\end{aligned}$$

As $T_{ei} \leq T_{di}$, we finally obtain:

$$\begin{aligned}
L_2 = \int_{-\infty}^{+\infty} \int_0^{+\infty} & f_Y(Y_i|\tau,u) e^{-\Lambda_3(T_{di}|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \\
& \times e^{-\mathbb{1}_{\{T_{ei}>\tau\}}\Lambda_{12}(T_{ei}-\tau|u)} f_\tau(\tau)\ f_u(u)\ \ d\tau du
\end{aligned}$$

By combining the contribution of demented and censored subjects using the illness indicator δ_{ei}, we obtain the individual contribution to the likelihood L_i (8).

$$\begin{aligned}
L_i = \int_{-\infty}^{+\infty} \int_0^{+\infty} & f_Y(Y_i|\tau,u) e^{-\Lambda_3(T_{di}|\tau,u)} \alpha_3(T_{di}|\tau,u)^{\delta_{di}} \\
& \times \left[\mathbb{1}_{\{T_{ei}>\tau\}} \alpha_{12}(T_{ei}-\tau|u) \right]^{\delta_{ei}} e^{-\mathbb{1}_{\{T_{ei}>\tau\}}\Lambda_{12}(T_{ei}-\tau|u)} \\
& \times f_\tau(\tau)\ f_u(u)\ \ d\tau du
\end{aligned}$$

4.4 Simulations complémentaires

Le modèle proposé, défini de manière paramétrique, nécessite une évaluation des estimateurs du maximum de vraisemblance. L'objet de cette section est double : il s'agit de confirmer les conclusions de l'article en terme de performances d'estimation des paramètres du modèle et d'évaluer l'impact de la prise en compte du décès dans la modélisation. A partir de plusieurs schémas de simulations, nous avons comparé les estimations du modèle trivarié (évolution du score cognitif, démence et décès) et celle du modèle bivarié (évolution du score cognitif et démence).

4.4.1 Méthodologie pour la génération des données

Dans le paragraphe 4.4.2, nous présentons 4 études de simulations construites de manière identique. Chaque étude de simulation a été réalisée avec 100 jeux de données contenant 500 sujets. Ces 4 études sont réalisées de manière analogue à celle présentée dans l'article, mais avec quelques modifications afin de mieux mettre en évidence les performances du modèle.

Simulations des données

Pour l'ensemble des sujets et la totalité des simulations, nous définissons :

- la durée maximum du suivi $d_{\max}$,
- le nombre maximum de mesure $n_{\max}$,

Pour chaque sujet i, nous simulons :

- le vecteur d'effets aléatoires portant sur l'évolution $u_i = (u_{0i}, u_{1i}, u_{2i})'$ suivant une loi normale $\mathcal{N}(0, G)$. Pour des raisons de temps de calculs, nous nous sommes limités à l'inclusion d'un seul effet aléatoire dans le modèle d'évolution, en plus de l'âge au changement de pente. Dans la simulation présentée dans l'article, nous avions choisi un effet aléatoire $u_i = u_{0i}$ sur le niveau cognitif à l'âge d'accélération du déclin τ_i ; ici nous incluons un effet aléatoire sur la pente de la seconde phase d'évolution $u_i = u_{s_2 i}$, avec $G = \sigma^2_{u_{s_2 i}}$,
- l'âge à l'accélération du déclin cognitif τ_i suivant une distribution de Weibull d'in-

tensité de transition

$$\alpha_{01}(t) = \lambda_\tau \kappa_\tau (\kappa_\tau t)^{\lambda_\tau - 1}$$

- l'âge d'entrée dans l'étude T_{0i} selon une distribution uniforme sur l'intervalle $[70 - 80]$,
- l'âge de censure identique pour la démence et le décès correspondant à l'âge à la fin de l'étude $C_{ei} = C_{di} = (T_{0i} + d_{\max})$,
- l'âge de décès T^*_{di} suivant l'intensité de transition suivante :

$$\alpha_3(t|\tau_i, u_i) = a_l \exp(\eta \tilde{Y}_i(t|\tau_i, u_i)) \quad si \quad T_l \leq t < T_{l+1}, l = 1, ...m_3$$

avec une intensité de transition de base exponentielle par morceaux et un risque dépendant de l'espérance du marqueur au temps courant t, $\tilde{Y}_i(t|\tau_i, u_i)$

- le délai de survenue d'une démence $delai_i$ depuis l'entrée dans l'état latent τ_i suivant une distribution de Weibull d'intensité de transition

$$\alpha_{12}(t - \tau_i) = \lambda_e \kappa_e (\kappa_e (t - \tau_i)^{\lambda_e - 1}$$

ce qui permet de définir un âge de démence T^*_{ei} tel que $T^*_{ei} = \tau_i + delai_i$

- les temps d'observation ou de censure de la démence et du décès ainsi que les indicateurs correspondants tels que

$$T_{di} = \min(T^*_{di}, C_{di}) \quad , \quad \delta_{di} = \mathbb{1}_{\{T^*_{di} \leq C_{di}\}}$$

et

$$T_{ei} = \min(T^*_{ei}, C_{e_i}, T_{di}) \quad , \quad \delta_{ei} = \mathbb{1}_{\{\{T^*_{ei} \leq C_{ei}\} \& \{T^*_{ei} \leq T_{di}\}\}}$$

- le vecteur des temps de mesures : $t_{0i} = T_{0i}$ et $t_{ij} = \min(t_{i(j-1)} + pas, T_{ei})$ tant que $t_{ij} < T_{ei}$ et $t_{ij} < T_{di}$, où le paramètre $pas = \frac{d_{\max}}{n_{\max}}$ définit l'intervalle de temps régulier entre deux visites successives,
- l'erreur gaussienne ϵ_i suivant une loi normale $\mathcal{N}(0, \sigma^2_\epsilon)$,
- et le vecteur de mesures

$$Y_i(t) = \phi_0 + (\phi_1 + \frac{u_{s2i}}{2})(t - \tau_i) + (\phi_2 + \frac{u_{s2i}}{2})\sqrt{(t - \tau_i)^2 + \gamma} + \epsilon_i$$

Modifications du schéma de simulation par rapport à celui de l'article

Par rapport au schéma de simulation utilisé dans l'article, deux modifications ont été effectuées. Nous avons inclus un effet aléatoire sur la seconde phase de l'évolution, au lieu d'un effet aléatoire sur le niveau cognitif à l'âge d'accélération du déclin. Le risque de décès est donc d'autant plus fort que la pente du déclin dans la phase pré-diagnostique est marquée. Nous avons également supprimé le phénomène de sortie d'étude indépendant du décès en supposant la censure pour l'événement démence égale à la censure pour le décès, ainsi $T_{ei} = \min(T_{ei}^*, T_{di})$. La démence est exclusivement censurée par le décès ou la durée totale du suivi et il n'y a donc pas d'autres phénomènes de sortie d'étude.

Quatre schémas d'études de simulations

Les 4 schémas d'études de simulations présentés dans la section suivante varient en fonction de la durée du suivi maximum et du nombre maximum de mesures potentiellement observables pour un sujet, ce qui induit des variations sur les intervalles de temps entre 2 mesures. Nous avons comparé les schémas de simulations suivants :

- Simulation 1 : $d_{\max} = 15$ ans, $n_{\max} = 6$ et $pas = 3$ ans,
- Simulation 2 : $d_{\max} = 15$ ans, $n_{\max} = 4$ et $pas = 5$ ans,
- Simulation 3 : $d_{\max} = 25$ ans, $n_{\max} = 6$ et $pas = 5$ ans,
- Simulation 4 : $d_{\max} = 25$ ans, $n_{\max} = 4$ et $pas = 8.33$ ans.

Les valeurs des paramètres pour le modèle mixte définissant l'évolution cognitive ont été choisies à partir d'une étude préliminaire sur les données de la cohorte Paquid. Pour les 4 schémas d'étude, elles sont identiques aux valeurs de l'article. Les paramètres pour les distributions des événements démence, décès et accélération du déclin sont identiques pour les études de simulations ayant une même durée de suivi, c'est-à-dire les simulations 1 et 2 d'une part et les simulations 3 et 4 d'autre part. Cela permet de mesurer l'impact de la fréquence des mesures sur les biais de sélection liés au décès (cf. section 4.4.2). En revanche, nous avons choisi des paramétrisations différentes pour les échantillons ayant une durée de suivi distincte, ceci afin de conserver un taux de démence et un taux de décès similaires en fin de suivi. Les paramètres ont été définis afin d'avoir environ 15% de déments et près de 60% de décès dans tous les cas, ce qui correspond à la structure de la cohorte Paquid.

De plus, les durées de suivi différentes nécessitent également une adaptation du nombre d'intervalles pour la fonction de risque exponentielle par morceaux de l'intensité de transition du décès afin de conserver l'identifiabilité du modèle. Avec 25 ans de suivi, nous avons défini 5 intervalles de bornes $T_1 = 70$, $T_2 = 80$, $T_3 = 85$, $T_4 = 90$, et $T_5 = 95$. Avec 15 ans de suivi, nous avons défini 4 intervalles de bornes $T_1 = 70$, $T_2 = 80$, $T_3 = 85$ et $T_4 = 90$. Nous n'avons pas simulé d'effet de covariable dans le modèle.

4.4.2 Résultats de 4 études de simulations

Pour chaque échantillon, nous avons estimé le modèle trivarié incluant l'évolution cognitive, le risque de démence et le risque de décès ainsi que le modèle bivarié incluant uniquement l'évolution cognitive et le risque de démence. Les estimations des paramètres des modèles de chaque étude de simulations sont présentées dans les tableaux 4.2, 4.3, 4.4 et 4.5.

Comme certains paramètres du modèle ne sont pas interprétables directement, nous présentons également les estimations de l'espérance de l'âge d'entrée dans la phase prédiagnostique $E_{01}(t)$ (cf. tableau 4.6), l'espérance du délai de survenue d'une démence $E_{12}(t-\tau_i)$ (cf. tableau 4.7) ou encore les pentes moyennes d'évolution linéaire dans chacune des phases, avant et après τ_i (cf. tableau 4.8) ainsi que leurs variances asymptotiques, calculées par "delta-method", et leurs variances empiriques respectives. Tel que nous le définissons dans l'article, la pente moyenne p_1 dans la première phase d'évolution est égale à $\phi_1 - \phi_2$ alors que la pente moyenne p_2 dans la seconde phase d'évolution est égale à $\phi_1 + \phi_2$. Pour l'âge à l'accélération du déclin et le délai de survenue d'une démence, l'espérance s'obtient par la formule de l'espérance d'une distribution de Weibull :

$$E(t) = \frac{1}{\kappa}\Gamma\left(\frac{\lambda+1}{\lambda}\right)$$

où Γ est la fonction de distribution Gamma et κ et λ sont les paramètres d'une distribution de Weibull.

Convergence des modèles

L'algorithme d'estimation des paramètres a un taux de convergence tout à fait acceptable, de l'ordre de 95% pour les 2 types de modèles (cf. tableau 4.1).

Tab. 4.1 : Taux de convergence des 4 études de simulations

	Modèle trivarié	Modèle bivarié
Simulation 1	94%	100%
Simulation 2	97%	98%
Simulation 3	95%	97%
Simulation 4	95%	93%

Estimations du modèle trivarié

Pour le modèle trivarié, chacun des 4 schémas de simulations révèle un bon comportement de la procédure d'estimation en terme de biais et de variance asymptotique, excepté une sous-estimation de la variance des paramètres de la distribution de Weibull pour l'âge à l'accélération du déclin ($\sqrt{\lambda_\tau}$ et $\sqrt{\kappa_\tau}$) (cf. tableaux 4.2, 4.3, 4.4 et 4.5) : l'écart-type asymptotique de ces 2 paramètres est deux à trois fois plus faible que l'écart-type empirique. Le tableau 4.6 montre que l'estimateur de l'espérance de l'âge à l'accélération du déclin $E_{01}(t)$ est non biaisé. En revanche, la sous-estimation de la variance de $\sqrt{\lambda_\tau}$ et $\sqrt{\kappa_\tau}$ se répercute sur l'estimation de la variance de l'espérance. Toutefois, on observe que l'intervalle de confiance de l'espérance $E_{01}(t)$ calculé avec la variance empirique montre une précision tout à fait acceptable (au maximum de plus ou moins 2 ans). Les biais concernant l'espérance du délai de survenue d'une démence $E_{12}(t - \tau_i)$ ($\sqrt{\lambda_e}$ et $\sqrt{\kappa_e}$) semblent mineurs (cf. tableau 4.7). Enfin, les estimations des pentes d'évolution p_1 et p_2 sont également correctes avec un biais légèrement supérieur systématiquement pour la pente de la seconde phase d'évolution (cf. tableau 4.8) pour laquelle nous avons beaucoup moins d'information.

Tab. 4.2 : - Simulation 1 - Valeurs Simulées (VS), Estimations Moyennes (EM), Biais Relatifs en % (BR), Ecarts-types Asymptotiques (ETA) et Ecarts-types Empiriques (ETE) pour 100 jeux de données simulés du modèle conjoint trivarié de l'évolution cognitive, de la démence et du décès et du modèle conjoint bivarié de l'évolution cognitive et de la démence : 15 ans de suivi, 6 mesures maximum et intervalle de temps de 3 ans entre 2 mesures (pour N=500 sujets)

	Modèle trivarié				Modèle bivarié			
VS	EM	BR	ETA	ETE	EM	BR	ETA	ETE
$\phi_0 = 10.660$	10.727	0.6	0.0950	0.1340	10.688	0.3	0.1004	0.1350
$\phi_1 = -0.400$	-0.380	-4.9	0.0263	0.0330	-0.358	-10.5	0.0267	0.0351
$\phi_2 = -0.330$	-0.313	-5.1	0.0267	0.0321	-0.291	-11.8	0.0266	0.0343
$\sigma_\epsilon = 1.550$	1.555	0.4	0.0255	0.0264	1.561	0.7	0.0257	0.0262
$\sqrt{\lambda_e} = 1.540$	1.606	4.3	0.1264	0.1828	1.600	3.9	0.1350	0.2046
$\sqrt{\kappa_e} = 0.380$	0.371	-2.3	0.0133	0.0165	0.367	-3.4	0.0136	0.0171
$\sqrt{a_1} = 0.350$	0.358	2.3	0.0514	0.0544	–	–	–	–
$\sqrt{a_2} = 0.400$	0.401	0.4	0.0498	0.0511	–	–	–	–
$\sqrt{a_3} = 0.850$	0.864	1.6	0.0895	0.0992	–	–	–	–
$\sqrt{a_4} = 0.900$	0.911	1.2	0.1137	0.1274	–	–	–	–
$\eta = -0.180$	-0.181	0.7	0.0223	0.0228	–	–	–	–
$\sqrt{\lambda_\tau} = 3.978$	4.080	2.6	0.0837	0.1539	4.022	1.1	0.0902	0.1684
$\sqrt{\kappa_\tau} = 0.106$	0.106	0.3	0.0002	0.0006	0.106	0.0	0.0003	0.0006
$\sigma_{u_{s_2 i}} = 0.333$	0.321	-3.7	0.0438	0.0491	0.304	-8.8	0.0443	0.0504

Tab. 4.3 : - Simulation 2 - Valeurs Simulées (VS), Estimations Moyennes (EM), Biais Relatifs en % (BR), Ecarts-types Asymptotiques (ETA) et Ecarts-types Empiriques (ETE) pour 100 jeux de données simulés du modèle conjoint trivarié de l'évolution cognitive, de la démence et du décès et du modèle conjoint bivarié de l'évolution cognitive et de la démence : 15 ans de suivi, 4 mesures maximum et intervalle de temps de 5 ans entre 2 mesures (pour N=500 sujets)

	Modèle trivarié				Modèle bivarié			
VS	EM	BR	ETA	ETE	EM	BR	ETA	ETE
$\phi_0 = 10.660$	10.696	0.3	0.1173	0.1427	10.644	−0.1	0.1257	0.1387
$\phi_1 = -0.400$	−0.393	−1.8	0.0296	0.0401	−0.353	−11.7	0.0286	0.0357
$\phi_2 = -0.330$	−0.323	−2.1	0.0296	0.0373	−0.284	−13.9	0.0287	0.0341
$\sigma_\epsilon = 1.550$	1.556	0.4	0.0326	0.0340	1.568	1.1	0.0332	0.0363
$\sqrt{\lambda_e} = 1.540$	1.581	2.6	0.1485	0.2269	1.596	3.7	0.1622	0.2250
$\sqrt{\kappa_e} = 0.380$	0.376	−1.1	0.0162	0.0176	0.368	−3.2	0.0161	0.0173
$\sqrt{a_1} = 0.350$	0.375	7.0	0.0555	0.0703	–	–	–	–
$\sqrt{a_2} = 0.400$	0.419	4.7	0.0533	0.0662	–	–	–	–
$\sqrt{a_3} = 0.850$	0.896	5.4	0.0959	0.1210	–	–	–	–
$\sqrt{a_4} = 0.900$	0.955	6.1	0.1216	0.1438	–	–	–	–
$\eta = -0.180$	−0.190	5.8	0.0232	0.0278	–	–	–	–
$\sqrt{\lambda_\tau} = 3.978$	4.067	2.2	0.0908	0.1860	3.977	0.0	0.0996	0.1791
$\sqrt{\kappa_\tau} = 0.106$	0.106	0.2	0.0002	0.0005	0.106	−0.2	0.0003	0.0006
$\sigma_{u_{s_2 i}} = 0.333$	0.327	−1.8	0.0480	0.0505	0.304	−8.8	0.0484	0.0554

Tab. 4.4 : - Simulation 3 - Valeurs Simulées (VS), Estimations Moyennes (EM), Biais Relatifs en % (BR), Ecarts-types Asymptotiques (ETA) et Ecarts-types Empiriques (ETE) pour 100 jeux de données simulés du modèle conjoint trivarié de l'évolution cognitive, de la démence et du décès et du modèle conjoint bivarié de l'évolution cognitive et de la démence : 25 ans de suivi, 6 mesures maximum et intervalle de temps de 5 ans entre 2 mesures (pour N=500 sujets)

	Modèle trivarié				Modèle bivarié			
VS	EM	BR	ETA	ETE	EM	BR	ETA	ETE
$\phi_0 = 10.660$	10.696	0.3	0.0925	0.1089	10.688	0.3	0.0936	0.1090
$\phi_1 = -0.400$	-0.381	-4.7	0.0227	0.0267	-0.340	-14.9	0.0214	0.0271
$\phi_2 = -0.330$	-0.313	-5.2	0.0228	0.0260	-0.273	-17.3	0.0217	0.0272
$\sigma_\epsilon = 1.550$	1.552	0.1	0.0268	0.0308	1.560	0.6	0.0271	0.0318
$\sqrt{\lambda_e} = 1.240$	1.290	4.0	0.1076	0.1134	1.322	6.6	0.1107	0.1262
$\sqrt{\kappa_e} = 0.320$	0.316	-1.2	0.0158	0.0182	0.309	-3.3	0.0143	0.0168
$\sqrt{a_1} = 0.400$	0.410	2.6	0.0590	0.0630	–	–	–	–
$\sqrt{a_2} = 0.450$	0.458	1.7	0.0602	0.0619	–	–	–	–
$\sqrt{a_3} = 0.500$	0.509	1.8	0.0599	0.0631	–	–	–	–
$\sqrt{a_4} = 0.600$	0.614	2.4	0.0627	0.0657	–	–	–	–
$\sqrt{a_5} = 0.650$	0.664	2.2	0.0598	0.0667	–	–	–	–
$\eta = -0.200$	-0.203	1.4	0.0204	0.0198	–	–	–	–
$\sqrt{\lambda_\tau} = 3.970$	4.032	1.6	0.0527	0.1219	4.020	1.3	0.0619	0.1173
$\sqrt{\kappa_\tau} = 0.102$	0.102	0.1	0.0001	0.0003	0.102	-0.1	0.0002	0.0004
$\sigma_{u_{s2i}} = 0.333$	0.326	-1.9	0.0397	0.0448	0.303	-9.0	0.0375	0.0396

Tab. 4.5 : - Simulation 4 - Valeurs Simulées (VS), Estimations Moyennes (EM), Biais Relatifs en % (BR), Ecarts-types Asymptotiques (ETA) et Ecarts-types Empiriques (ETE) pour 100 jeux de données simulés du modèle conjoint trivarié de l'évolution cognitive, de la démence et du décès et du modèle conjoint bivarié de l'évolution cognitive et de la démence : 25 ans de suivi, 4 mesures maximum et intervalle de temps de 8.33 ans entre 2 mesures (pour N=500 sujets)

	Modèle trivarié				Modèle bivarié			
VS	EM	BR	ETA	ETE	EM	BR	ETA	ETE
$\phi_0 = 10.660$	10.678	0.2	0.1125	0.1398	10.658	0.0	0.1162	0.1463
$\phi_1 = -0.400$	−0.388	−3.0	0.0268	0.0356	−0.330	−17.5	0.0227	0.0292
$\phi_2 = -0.330$	−0.318	−3.7	0.0270	0.0342	−0.261	−21.0	0.0233	0.0285
$\sigma_\epsilon = 1.550$	1.550	0.0	0.0336	0.0393	1.562	0.8	0.0343	0.0387
$\sqrt{\lambda_e} = 1.240$	1.301	4.9	0.1276	0.1415	1.333	7.5	0.1394	0.1465
$\sqrt{\kappa_e} = 0.320$	0.319	−0.2	0.0188	0.0192	0.307	−4.2	0.0173	0.0179
$\sqrt{a_1} = 0.400$	0.397	−0.7	0.0593	0.0651	–	–	–	–
$\sqrt{a_2} = 0.450$	0.453	0.6	0.0615	0.0682	–	–	–	–
$\sqrt{a_3} = 0.500$	0.497	−0.6	0.0606	0.0637	–	–	–	–
$\sqrt{a_4} = 0.600$	0.601	0.2	0.0632	0.0735	–	–	–	–
$\sqrt{a_5} = 0.650$	0.650	0.0	0.0597	0.0640	–	–	–	–
$\eta = -0.200$	−0.200	−0.2	0.0215	0.0243	–	–	–	–
$\sqrt{\lambda_\tau} = 3.970$	4.034	1.6	0.0634	0.1238	4.012	1.1	0.0766	0.1484
$\sqrt{\kappa_\tau} = 0.102$	0.102	0.1	0.0002	0.0004	0.102	−0.1	0.0002	0.0004
$\sigma_{u_{s2i}} = 0.333$	0.321	−3.7	0.0448	0.0529	0.286	−14.2	0.0399	0.0432

Tab. 4.6 : Valeurs Simulées de l'espérance de l'âge à l'entrée dans la phase pré-diagnostique (VS), Valeurs Estimées de l'âge à l'accélération du déclin (VE), Biais Relatifs (BR) (%), Ecarts-Types Asymptotiques (ETA) et Ecarts-Types Empiriques (ETE) pour les 4 études de simulations issues du modèle trivarié et bivarié

$E_{01}(t)$		Modèle trivarié				Modèle bivarié			
	VS	VE	BR	ETA*	ETE*	VE	BR	ETA*	ETE*
Simulation 1	86.09	85.67	−0.5	0.316	0.990	86.19	0.1	0.405	0.998
Simulation 2	86.09	85.87	−0.2	0.346	0.950	86.50	0.5	0.461	1.084
Simulation 3	92.96	92.89	−0.1	0.235	0.588	93.13	0.2	0.263	0.722
Simulation 4	92.96	92.93	0.0	0.265	0.704	93.26	0.3	0.329	0.764

*estimé par "delta-method"

Tab. 4.7 : Valeurs Simulées de l'espérance du délai de survenue d'une démence (VS), Valeurs Estimées du délai de survenue d'une démence (VE), Biais Relatifs (BR) (%), Ecarts-Types Asymptotiques (ETA) et Ecarts-Types Empiriques (ETE) pour les 4 études de simulations issues du modèle trivarié et bivarié

$E_{12}(t-\tau_i)$		Modèle trivarié				Modèle bivarié			
	VS	VE	BR	ETA*	ETE*	VE	BR	ETA*	ETE*
Simulation 1	6.14	6.50	6.0	0.468	0.611	6.66	8.6	0.497	0.656
Simulation 2	6.14	6.35	3.4	0.553	0.612	6.64	8.2	0.588	0.644
Simulation 3	8.79	9.10	3.6	1.056	1.214	9.46	7.6	0.959	1.127
Simulation 4	8.79	8.93	1.6	1.205	1.189	9.65	9.8	1.247	1.200

*estimé par "delta-method"

Tab. 4.8 : Valeurs Simulées des pentes dans les 2 phases d'évolution (VS), Valeurs Estimées des pentes dans les 2 phases d'évolution (VE), Biais Relatifs (BR) (%), Ecarts-Types Asymptotiques (ETA) et Ecarts-Types Empiriques (ETE) pour les 4 études de simulations issues du modèle trivarié et bivarié

$p_1 = -0.07$	Modèle trivarié				Modèle bivarié			
$p_2 = -0.73$	VE	BR	ETA*	ETE*	VE	BR	ETA*	ETE*
Simulation 1								
p_1	−0.067	−3.7	0.008	0.010	−0.067	−4.2	0.008	0.010
p_2	−0.693	−5.0	0.052	0.064	−0.649	−11.1	0.053	0.069
Simulation 2								
p_1	−0.070	−0.6	0.010	0.010	−0.069	−1.4	0.010	0.009
p_2	−0.716	−1.9	0.058	0.077	−0.638	−12.7	0.056	0.069
Simulation 3								
p_1	−0.068	−2.1	0.006	0.006	−0.068	−3.3	0.006	0.006
p_2	−0.694	−4.9	0.045	0.052	−0.613	−16.0	0.043	0.054
Simulation 4								
p_1	−0.070	0.0	0.007	0.007	−0.069	−1.1	0.007	0.008
p_2	−0.706	−3.3	0.053	0.069	−0.591	−19.1	0.045	0.057

*estimé par "delta-method"

Estimations du modèle bivarié

Si l'on s'intéresse aux estimations du modèle bivarié pour comprendre l'impact de la prise en compte de la censure informative induite par le décès, nous constatons que dans chacune des 4 études de simulations, les paramètres liés aux pentes de l'évolution ϕ_1 et ϕ_2 sont largement sous-estimés dans le modèle bivarié (cf. tableaux 4.2, 4.3, 4.4 et 4.5). Ces biais se traduisent par une nette sous-estimation de la pente p_2 de la seconde phase d'évolution (cf. tableau 4.8). Le paramètre d'écart-type de l'effet aléatoire $\sigma_{u_{s_2 i}}$ est également sous-estimé. L'estimation des paramètres de la distribution de Weibull pour l'âge à l'accélération du déclin ($\sqrt{\lambda_\tau}$ et $\sqrt{\kappa_\tau}$) ne semble pas perturbée par la non prise en compte du décès, tout comme l'espérance de l'âge d'entrée dans la phase pré-diagnostique $E_{01}(t)$ (cf. tableau 4.6). Ceci s'explique par le fait que le modèle mixte n'inclut pas d'effets aléatoires sur la pente dans la première phase. Comme le lien entre cognition et décès passe par les effets aléatoires, le décès n'induit donc pas de censure informative dans la première phase. En revanche, les biais sur les paramètres de la distribution du délai de survenue de la démence sont légèrement plus importants pour le modèle bivarié comparé au modèle trivarié. Ils se traduisent par un biais sur l'espérance du délai de survenue d'une démence $E_{12}(t - \tau_i)$ qui peut atteindre près de 10% (cf. tableau 4.7).

Réflexions sur les différents schémas de simulations

La comparaison directe de l'estimation des paramètres du modèle de l'approche trivariée et l'approche bivariée met en évidence des biais lorsque le décès n'est pas modélisé conjointement. L'impact de ces biais est difficile à évaluer, il est plus intéressant d'étudier des paramètres interprétables. En particulier, la pente dans la seconde phase d'évolution est sous-estimée et, dans une moindre mesure, le délai de survenue d'une démence surestimé. Par ailleurs, nous observons que le schéma de simulation des données influe sur la qualité des estimations.

Comme nous ne simulons pas de censure pour la démence, la seule source de données manquantes au cours du suivi longitudinal est liée à la survenue du décès. Pour mesurer l'impact des données manquantes liées au décès, nous avons calculé le pourcentage de scores cognitifs manquants induits par le décès. Plus le suivi est long, plus la proportion de données manquantes induites par le décès est élevée. Le pourcentage de données

manquantes dans la première phase d'évolution est plus faible (entre 10% et 15%) comparativement à celui de la seconde phase (entre 40% et 55%). Les tableaux 4.8 et 4.9 montrent une association nette entre le pourcentage de données manquantes induites par le décès et le biais sur la seconde pente d'évolution lorsque le décès n'est pas modélisé conjointement.

Pour les 4 schémas de simulations, nous avons calculé le pourcentage de déments non diagnostiqués (car décédés avant la visite suivant la date de démence) parmi l'ensemble des sujets déments. Ce pourcentage augmente quand le délai entre les visites augmente :

- Simulation 1, pas =3 ans : 18%
- Simulation 2, pas =5 ans : 29%
- Simulation 3, pas =5 ans : 22%
- Simulation 4, pas =8.33 ans : 36%

Plus le pourcentage de déments non diagnostiqués augmente, plus la différence est importante entre les biais sur l'espérance du délai de survenue d'une démence $E_{12}(t - \tau_i)$ dans le modèle trivarié et dans le modèle bivarié (cf. tableau 4.7) :

- Simulation 1 : 6.0% contre 8.6%
- Simulation 2 : 3.4% contre 8.2%
- Simulation 3 : 3.6% contre 7.6%
- Simulation 4 : 1.6% contre 9.8%

Toutefois, l'impact des déments non diagnostiqués semble moindre que celui du pourcentage de scores cognitifs manquants. En effet, le biais sur l'espérance du délai de survenue d'une démence $E_{12}(t - \tau_i)$ ne dépasse pas 10%.

Tab. 4.9 : Nombre moyen de mesures avant τ_i, Nombre moyen de mesures après τ_i, Nombre moyen de mesures des 500 sujets des 100 jeux de données simulés

	Nombre moyen de mesures avant τ_i	Nombre moyen de mesures après τ_i	Nombre moyen de mesures
Simulation 1 ($d_{\max}$=15 ans, $n_{\max} = 6$ et *pas* =3 ans)			
-avec le décès simulé	3.65	0.87	4.52
-sans le décès simulé	4.12	1.46	5.58
→ Données manquantes induites par le décès	11.31 %	40.69 %	19.01 %
Simulation 2 ($d_{\max}$=15 ans, $n_{\max} = 4$ et *pas* =5 ans)			
-avec le décès simulé	2.41	0.61	3.02
-sans le décès simulé	2.72	1.09	3.81
→ Données manquantes induites par le décès	11.39 %	44.16 %	20.74 %
Simulation 3 ($d_{\max}$=25 ans, $n_{\max} = 6$ et *pas* =5 ans)			
-avec le décès simulé	3.47	0.75	4.22
-sans le décès simulé	4.09	1.53	5.62
→ Données manquantes induites par le décès	15.05 %	51.13 %	24.88 %
Simulation 4 ($d_{\max}$=25 ans, $n_{\max} = 4$ et *pas* =8.33 ans)			
-avec le décès simulé	2.37	0.52	2.88
-sans le décès simulé	2.76	1.10	3.86
→ Données manquantes induites par le décès	14.20 %	53.03 %	25.27 %

4.5 Application à la cohorte Paquid : impact du niveau d'éducation

Afin d'explorer le rôle du niveau d'éducation dans le processus de dégradation cognitive, nous avons voulu compléter l'application présentée dans l'article. Dans cet objectif, nous avons réalisé une analyse stratifiée et une analyse ajustée sur le niveau d'éducation. Dans une première partie, nous présentons l'analyse effectuée exclusivement sur les sujets de bas niveau d'études, c'est-à-dire les sujets sans certificat d'étude primaire (CEP-) ; les résultats ont été comparés avec ceux de l'article portant sur les sujets de haut niveau d'études (CEP+). Dans un second temps, nous avons réalisé une analyse sur l'échantillon total avec le niveau d'éducation comme variable d'ajustement. Le tableau 4.10 présente un descriptif des échantillons d'analyse issus de la cohorte de personnes âgées Paquid. Les critères de sélection des échantillons d'étude sont les mêmes que dans l'article. Le score cognitif étudié est le score de Benton dont les mesures sont effectuées aux visites 1, 3, 5, 8, 10, 13 et 15 ans après l'inclusion.

4.5.1 Analyse stratifiée : échantillon des sujets de bas niveau d'études

Pour comparer les résultats de l'application aux sujets de haut niveau d'études (sujets ayant le certificat d'étude primaire : CEP+) à ceux de l'application aux sujets de faible niveau d'études (sujets n'ayant pas le certificat d'étude primaire : CEP-), nous avons utilisé exactement le même modèle que dans l'article. L'évolution du score de Benton est modélisée par un modèle à changement de pente aléatoire avec 2 phases d'évolution linéaires. Les intensités de transition vers l'état pré-diagnostique et l'état dément sont définies avec une intensité de base de Weibull alors que le risque de décès est défini par une distribution exponentielle par morceaux. L'analyse porte sur 1279 sujets de bas niveau d'études. Le descriptif de cet échantillon (cf. tableau 4.10) montre une proportion légèrement plus élevée de sujets déments et de sujets décédés que pour les sujets de haut niveau d'études. Le nombre moyen de mesures au cours du suivi de ces sujets est plus faible.

Tab. 4.10 : Descriptif des échantillons étudiés : sujets de haut niveau d'études (CEP+), sujets de bas niveau d'études (CEP-) et échantillon total

	CEP+	CEP-	Echantillon total
Nombre de sujets	2396	1279	3675
Age moyen d'entrée	74.48	76.72	75.26
Age moyen de sortie	82.14	83.01	82.44
Nombre moyen de mesures	2.88	1.77	2.49
Nombre de déments	365	296	661
% de déments	15.2%	23.1%	18.0%
Age moyen de démence	85.28	85.39	85.32
Nombre de décès	1437	893	2330
% de décès	59.9%	69.8%	63.4
Age moyen de décès	84.69	86.03	85.21

Interprétations des paramètres estimés

Les paramètres estimés des modèles trivariés et bivariés sont présentés dans le tableau 4.11. Les estimations de l'espérance de l'âge d'entrée dans la phase pré-diagnostique $E(\tau_i)$, du délai de survenue d'une démence $E(T^*_{ei} - \tau_i)$ et des pentes d'évolution linéaire p_1 et p_2, obtenus à partir du modèle trivarié, sont présentés dans le tableau 4.12. Pour obtenir les intervalles de confiance pour ces différentes fonctions d'intérêt, nous utilisons une technique de Bootstrap paramétrique (Efron et Tibshirani, 1993), 1993). Par exemple, pour l'espérance $E_{01}(t)$ qui est une fonction complexe de λ_τ et κ_τ, à partir de leur distribution asymptotique estimée, nous générons 2000 réalisations des paramètres λ_τ et κ_τ. Pour chaque réplication, nous calculons la fonction d'intérêt et nous ordonnons les 2000 valeurs obtenues. Les bornes inférieures et supérieures de l'intervalle de confiance sont respectivement les $2.5^{\text{ième}}$ et $97.5^{\text{ième}}$ percentiles empiriques.

Pour les sujets CEP-, le score de Benton à l'entrée dans l'état latent ($\phi_0 = 8.69$, ET=0.14) est plus faible que pour les sujets CEP+ ($\phi_0 = 10.66$, ET=0.07). Le déclin cognitif pour les sujets CEP- semble légèrement plus marqué dans la phase d'évolution

normale ($p_1 = -0.102$ pour CEP- vs $p_1 = -0.073$ pour CEP+) mais plus doux dans la phase pré-diagnostique ($p_2 = -0.520$ pour CEP- vs $p_1 = -0.824$ pour CEP+). L'espérance de l'âge d'entrée dans la phase pré-diagnostique montre que l'accélération du déclin cognitif survient plus précocément chez les sujets CEP- (82.40 ans) par rapport aux sujets CEP+ (86.38 ans). Le délai de survenue d'une démence est légèrement plus long pour les sujets CEP- (5.17 ans) que pour les sujets CEP+ (4.82 ans).

Analyse bivariée

Lorsque le décès est supposé indépendant de l'évolution cognitive et de la démence (modèle bivarié) (cf. tableau 4.11), le score moyen de Benton à l'entrée en phase pré-diagnostique est estimé à 9.00 tandis que les pentes sont de -0.08 dans la première phase et -0.38 dans la seconde. L'âge à l'entrée dans l'état latent est estimé à 81.59 ans et le délai de survenue d'une démence à 6.44 ans. On retrouve donc les résultats observés dans les simulations : une sous-estimation du déclin dans la seconde phase d'évolution et, dans une proportion moindre, une légère sur-estimation du délai de survenue d'une démence lorsque le décès n'est pas modélisé. En revanche l'estimation de l'âge d'entrée dans la phase pré-diagnostique de démence ne semble pas modifiée.

Adéquation du modèle aux données

Pour évaluer l'adéquation de nos hypothèses paramétriques aux données de l'échantillon des sujets de bas niveau d'études, nous avons réalisé exactement les mêmes analyses que celles présentées dans l'article. Pour les sujets déments et non-déments, nous avons comparé les moyennes du score estimé *a posteriori* $\hat{Y}(t) = E(Y(t)|\hat{\tau}_i, \hat{u}_i)$ et les moyennes du score cognitif observé sur des intervalles d'âge de 5 ans (cf. figure 4.3). Le modèle évalue correctement les évolutions des sujets déments et non-déments.

En ce qui concerne l'événement décès, la survie marginale et l'intensité de transition vers le décès obtenue à l'aide du modèle conjoint trivarié sont comparées aux fonctions correspondantes obtenues à l'aide d'un modèle des risques proportionnels non paramétrique dans une approche de vraisemblance pénalisée (PHMPL) (Joly et al., 1999) (cf. figure 4.4). Pour des raisons identiques à celles évoquées dans l'article, nous comparons la

Tab. 4.11 : Valeurs Estimées (VE), Ecart-types (ET) des paramètres pour le modèle trivarié et le modèle bivarié à partir des sujets de bas niveau d'études de la cohorte PAQUID, obtenus dans l'analyse stratifiée (N=1279)

Paramètres	Modèle trivarié		Modèle bivarié	
	VE	ET	VE	ET
ϕ_0	8.692	0.1399	8.996	0.1716
ϕ_1	−0.311	0.0236	−0.232	0.0211
ϕ_2	−0.209	0.0261	−0.147	0.0225
σ_ϵ	1.836	0.0371	1.801	0.0357
$\sqrt{\lambda_e}$	1.636	0.1043	1.723	0.1707
$\sqrt{\kappa_e}$	0.414	0.0143	0.372	0.0199
$\sqrt{a_1}$	0.302	0.0773	–	–
$\sqrt{a_2}$	0.505	0.0700	–	–
$\sqrt{a_3}$	0.584	0.0702	–	–
$\sqrt{a_4}$	0.704	0.0766	–	–
$\sqrt{a_5}$	0.846	0.0861	–	–
$\sqrt{a_6}$	1.005	0.1009	–	–
$\sqrt{a_7}$	1.286	0.1393	–	–
η	−0.248	0.0271	–	–
$\sqrt{\lambda_\tau}$	3.614	0.0742	3.376	0.1128
$\sqrt{\kappa_\tau}$	0.108	0.0002	0.108	0.0005
σ_1^a	1.518	0.0837	1.564	0.0865
σ_2^a	0.211	0.0450	0.117	0.0468
σ_3^a	0.221	0.0368	0.194	0.0396

[a] Paramètres de Variance-Covariance ($\sigma^2_{u_{0i}}$,$\sigma_{u_{0s_2i}}$,$\sigma^2_{u_{s_2i}}$) remplacés par les paramètres correspondants obtenus par décomposé de Cholesky (σ_1,σ_2,σ_3)

Tab. 4.12 : Valeurs Estimées (VE) des espérances du délai de survenue d'une démence $T^*_{ei} - \tau_i$, de l'âge à l'accélération du déclin cognitif τ_i, des pentes d'évolution linéaire dans les deux phases p_1 et p_2 ainsi que leur intervalle de confiance respectif à 95% (b_{inf},b_{sup}) obtenus par le modèle stratifié sur le niveau d'études (N=3675)

	Bas niveau d'études			Haut Niveau d'études		
Paramètres	VE	b_{inf}	b_{sup}	VE	b_{inf}	b_{sup}
$E(T^*_{ei} - \tau_i)$	5.174	4.755	5.665	4.819	4.486	5.211
$E(\tau_i)$	82.400	81.881	82.830	86.383	86.123	86.671
p_1	−0.102	−0.130	−0.075	−0.073	−0.110	−0.038
p_2	−0.520	−0.633	−0.411	−0.824	−0.924	−0.728

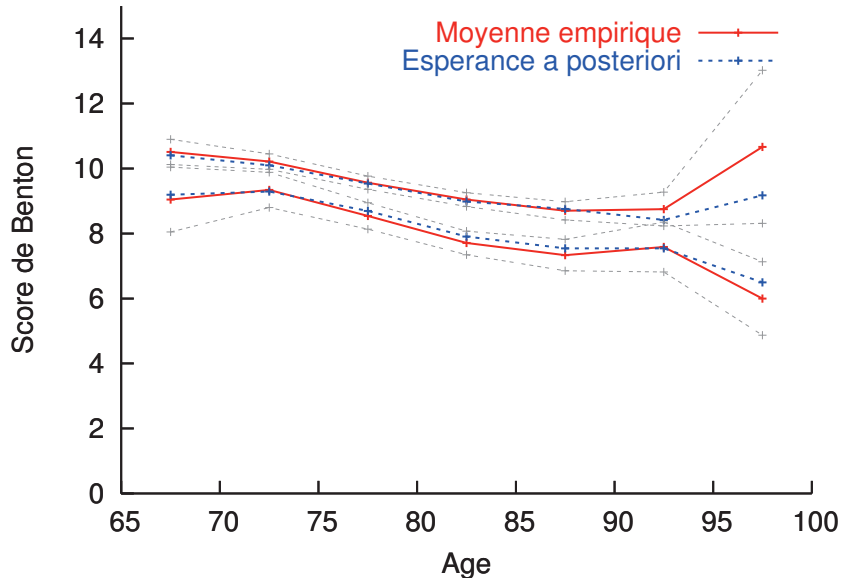

Fig. 4.3 : Evolution *a posteriori* comparée à l'évolution empirique pour les sujets non-déments (courbe supérieure) et les sujets déments (courbe inférieure) chez les sujets de bas niveau d'études

survie marginale et l'intensité de transition pour la démence obtenue à l'aide du modèle bivarié avec les fonctions correspondantes issue de PHMPL (cf. figure 4.5). Nous pouvons raisonnablement affirmer que le modèle choisi pour le risque de décès est suffisamment souple et qu'il permet d'ajuster correctement l'intensité de transition vers le décès. En revanche, les hypothèses paramétriques effectuées sur le risque de démence semblent moins bien ajustées au risque obtenu par PHMPL.

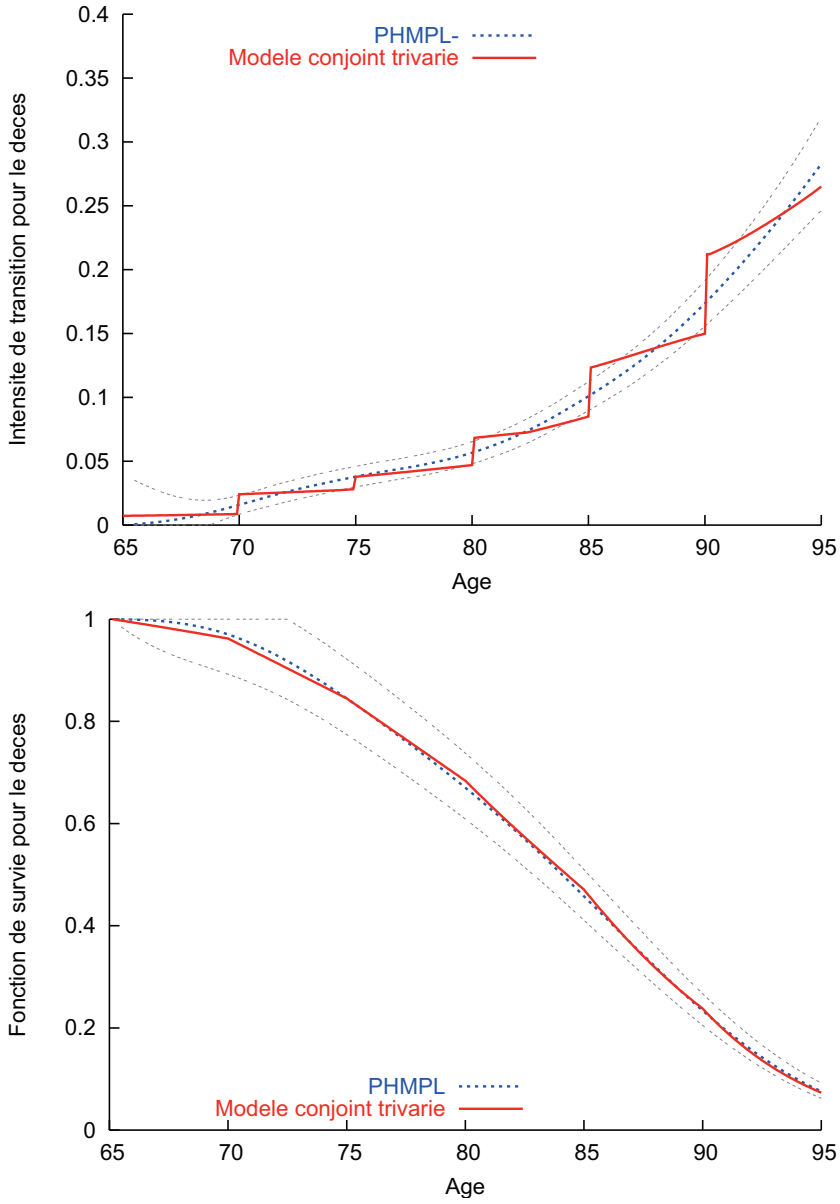

Fig. 4.4 : Fonction de survie et intensité de transition marginales pour le décès obtenues à l'aide du modèle trivarié, comparées aux fonctions correspondantes obtenues par PHMPL

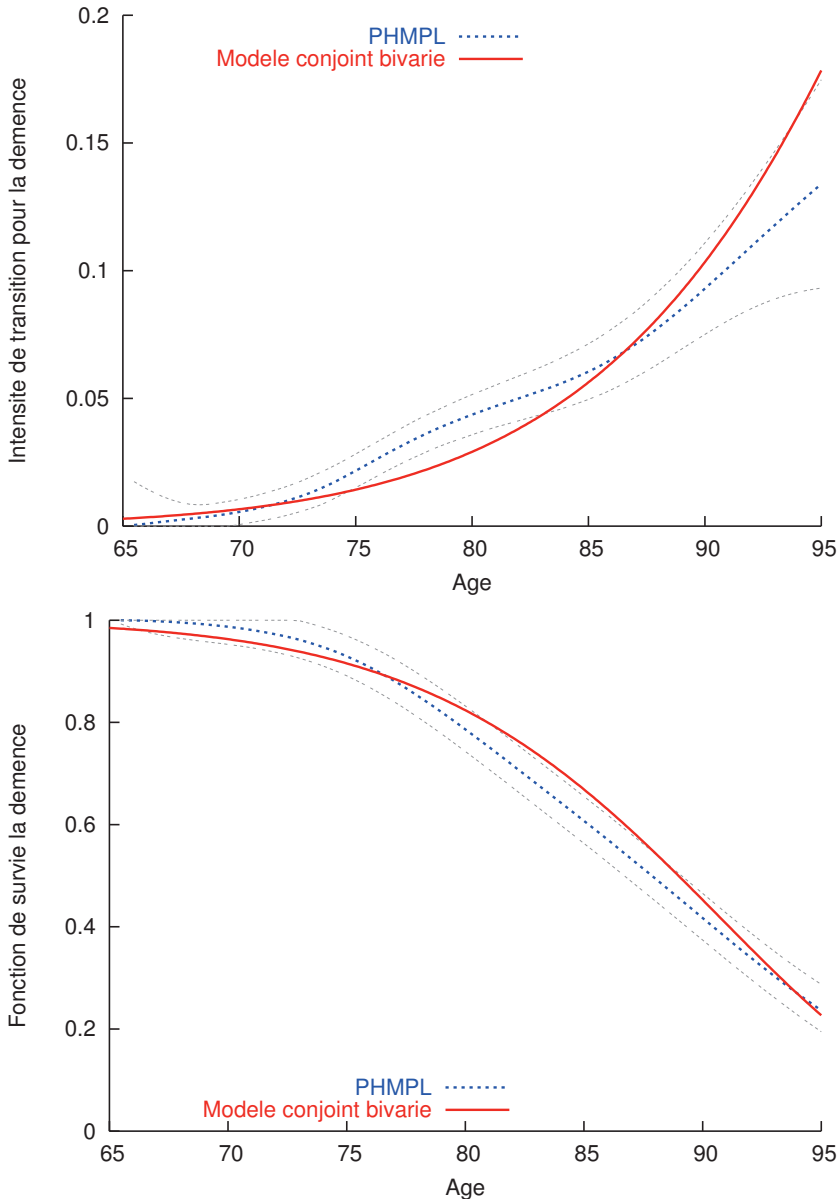

Fig. 4.5 : Fonction de survie et intensité de transition marginales pour la démence obtenues à l'aide du modèle bivarié, comparées aux fonctions correspondantes obtenues par PHMPL

4.5.2 Analyse ajustée

L'inconvénient de cette analyse stratifiée est de ne fournir aucun test concernant l'effet du niveau d'études sur les différents compartiments du modèle conjoint (Evolution, Transitions vers l'état pré-diagnostique, vers la démence et vers le décès). Dans cet objectif, nous avons réalisé l'analyse ajustée avec la variable binaire suivante : CEP=1 si le sujet est CEP+ et 0 sinon.

Définition du modèle

La variable CEP est introduite dans les 3 sous-modèles. Dans le modèle mixte, elle intervient sur le niveau cognitif à l'âge d'entrée dans la phase pré-diagnostique, la pente moyenne et la différence de pente :

$$\begin{aligned} Y_i(t) =&(\phi_0 + \alpha_0 CEP + u_{i0}) + (\phi_1 + \alpha_1 CEP + \frac{u_{s_2 i}}{2})(t - \tau_i) \\ &+ (\phi_2 + \alpha_2 CEP + \frac{u_{s_2 i}}{2}) \times \sqrt{(t - \tau_i)^2 + \gamma} + \epsilon_i \end{aligned}$$

avec $u_{s_2 i} \sim \mathcal{N}(0, \sigma^2_{u_{s_2 i}})$ et $\epsilon_i \sim \mathcal{N}(0, \sigma^2_\epsilon)$

De même, la variable CEP est introduite dans le modèle pour l'entrée dans la phase pré-diagnostique, le risque de démence et le risque de décès, à l'aide de modèles des risques proportionnels (risque de base de Weibull pour les transitions de l'état 0 à 1 et de l'état 1 à 2 et risque de base exponentiel par morceaux pour la transition vers l'état 3) :

$$\alpha_{01}(t) = \lambda_\tau \kappa_\tau (\kappa_\tau t)^{\lambda_\tau - 1} \exp(\theta_\tau CEP) \text{ avec } \lambda_\tau > 0 \text{ et } \kappa_\tau > 0$$

$$\alpha_{12}(t - \tau_i) = \lambda_e \kappa_e (\kappa_e (t - \tau_i))^{\lambda_e - 1} \exp(\theta_e CEP) \text{ avec } \lambda_e > 0 \text{ et } \kappa_e > 0$$

$$\alpha_3(t|\tau_i, u_i) = a_l \exp(\eta \tilde{Y}_i(t|\tau_i, u_i) + \theta_d CEP) \quad \text{si } T_l \leq t < T_{l+1}, l = 1, ..., 7$$

avec $a_l > 0$ et $T_1 = 65$, $T_2 = 70$, $T_3 = 75$, $T_4 = 80$, $T_5 = 85$, $T_6 = 90$ et $T_7 = 95$

Estimation des paramètres

Les paramètres du modèle sont présentés dans le tableau 4.13. Le niveau d'études (CEP) n'est pas significativement associé au risque de transition vers l'état latent ($\theta_\tau = -0.08$, $IC_{95\%}$=[-0.22 ;0.07]). L'espérance de l'âge d'entrée dans l'état latent est estimée à 84.65 ans pour les sujets CEP- et à 85.07 ans pour les sujets CEP+ (cf. tableau 4.14). Par contre, le CEP est significativement associé au risque de démence depuis l'état latent

($\theta_e = -1.11$, $IC_{95\%}$=[-2.00 ;-0.22]). L'espérance du délai de survenue d'une démence est plus court chez les sujets CEP- (4.08 ans) que chez les sujets CEP+ (6.52 ans). Le risque de décès est significativement plus élevé chez les sujets CEP- que chez les sujets CEP+ ($\theta_d = 0.23$, $IC_{95\%}$=[0.13 ;0.34]). Le niveau cognitif à l'âge d'entrée dans la phase pré-diagnostique est significativement plus élevé chez les sujets CEP+ que chez les sujets CEP- ($\alpha_0 = 2.34$, $IC_{95\%}$=[1.81 ;2.87]). La non significativité des paramètres α_1 et α_2 est plus difficilement interprétable et il est nécessaire d'étudier cet effet directement sur les pentes d'évolution :

- $p_1(CEP+) - p_1(CEP-)$= 0.0352 ($IC_{95\%}$=[0.01 ;0.06])
- $p_2(CEP+) - p_2(CEP-)$= 0.0358 ($IC_{95\%}$=[0.01 ;0.06])

Tab. 4.13 : Valeurs Estimées (VE), Ecarts-types (ET) et Intervalles de confiance à 95 % des paramètres pour le modèle trivarié obtenues dans l'analyse ajsuté (N=3675)

Paramètres	VE	ET	b_{inf}	b_{sup}
ϕ_0	8.5554	0.1120	8.336	8.775
ϕ_1	−0.3720	0.0175	−0.406	−0.338
ϕ_2	−0.2683	0.0158	−0.299	−0.237
α_0	2.3412	0.2718	1.808	2.874
α_1	0.0355	0.0535	−0.069	0.140
α_2	0.0003	0.0020	−0.004	0.004
σ_ϵ	1.6033	0.0155	1.573	1.634
$\sqrt{\lambda_e}$	1.5393	0.0751	1.392	1.686
$\sqrt{\kappa_e}$	0.4660	0.0173	0.432	0.500
θ_e	−1.1102	0.4560	−2.004	−0.216
$\sqrt{a_1}$	0.3181	0.0428	0.234	0.402
$\sqrt{a_2}$	0.3458	0.0403	0.267	0.425
$\sqrt{a_3}$	0.4243	0.0420	0.342	0.507
$\sqrt{a_4}$	0.5243	0.0446	0.437	0.612
$\sqrt{a_5}$	0.6532	0.0481	0.559	0.748
$\sqrt{a_6}$	0.7930	0.0540	0.687	0.899
$\sqrt{a_7}$	0.9372	0.0653	0.809	1.065
θ_d	0.2338	0.0539	0.128	0.339
η	−0.1740	0.0195	−0.212	−0.136
$\sqrt{\lambda_\tau}$	3.8955	0.0374	3.822	3.969
$\sqrt{\kappa_\tau}$	0.1068	0.0001	0.107	0.107
θ_τ	−0.0764	0.0733	−0.220	0.067
σ_1^a	1.3613	0.0370	1.289	1.434
σ_2^a	0.1431	0.0299	0.085	0.202
σ_3^a	0.2581	0.0258	0.207	0.309

[a] Paramètres de Variance-Covariance ($\sigma^2_{u_{0i}}$,$\sigma_{u_{0s_2i}}$,$\sigma^2_{u_{s_2i}}$) remplacés par les paramètres correspondants obtenus par décomposé de Cholesky (σ_1,σ_2,σ_3)

Tab. 4.14 : Valeurs Estimées (VE) des espérances du délai de survenue d'une démence $T^*_{ei} - \tau_i$, de l'âge à l'accélération du déclin cognitif τ_i, des pentes d'évolution linéaire dans les deux phases p_1 et p_2 ainsi que leur intervalle de confiance respectif à 95% (b_{inf},b_{sup}) obtenus par le modèle ajusté sur le niveau d'études (N=3675)

	Bas niveau d'études			Haut Niveau d'études		
Paramètres	VE	b_{inf}	b_{sup}	VE	b_{inf}	b_{sup}
$E(T^*_{ei} - \tau_i)$	4.081	3.715	4.472	6.521	2.962	23.808
$E(\tau_i)$	84.647	84.413	84.889	85.074	82.474	87.773
$E(p_1)$	−0.104	−0.118	−0.089	−0.068	−0.150	0.017
$E(p_2)$	−0.640	−0.709	−0.569	−0.604	−0.671	−0.535

4.5.3 Comparaison des 2 approches

Interprétations différentes de l'effet du niveau d'études selon le modèle considéré

Les résultats issus de l'approche par modélisation ajustée sur le niveau d'études sont surprenants dans le sens où ils sont contradictoires avec ceux du modèle proposé par Jacqmin-Gadda et al. (2006) et ceux du modèle stratifié. En effet, ces deux approches suggèrent que l'âge d'entrée dans la phase de déclin pré-diagnostique est plus élevé pour les sujets CEP+ (86.4 ans versus 82.4 ans chez les sujets CEP-) alors que la différence n'est pas significative dans le modèle ajusté. De plus, les résultats montrent que le délai de survenue d'une démence dans la phase pré-diagnostique est plus courte chez les sujets CEP- (4.1 ans) par rapport aux sujets CEP+ (6.5 ans) dans le modèle ajusté alors qu'on trouve l'inverse dans le modèle stratifié (5.2 ans chez les sujets CEP- versus 4.8 ans chez les sujets CEP-). Enfin, le modèle ajusté révèle une accélération du déclin cognitif plus forte chez les sujets de bas niveau d'études (-0.64 points par an) par rapport aux sujets de haut niveau d'études (-0.6 points par an) dans la seconde phase d'évolution alors que c'est l'inverse pour le modèle stratifié (-0.8 points pour les sujets de haut niveau d'études contre -0.5 points par an chez les bas niveau).

Choix du modèle

Le calcul du critère d'information bayésien $\big(\text{BIC} = -2\text{log(vraisemblance)} + n_p\text{log}(n_s)\big)$ révèle une meilleure vraisemblance du modèle stratifié par rapport au modèle ajusté, rapporté au nombre de paramètres estimés, n_p, et au nombre de sujets, n_s dans chacune des deux approches :

- $\text{BIC}_{\text{stratifié}} = -2 \times (-20777.15 - 9082.13) + 38\text{log}(3675) = 60030.51$
- $\text{BIC}_{\text{ajusté}} = -2 \times (-29939.57) + 25\text{log}(3675) = 60084.37$

Hypothèse de proportionnalité des risques

Les intensités de transition estimées depuis l'état sain vers l'état pré-diagnostique (état latent) pour les sujets de bas et de haut niveau d'études, issues de l'analyse stratifiée, sont représentées sur le graphique 4.6. Nous constatons tout d'abord que le risque de transition de l'état sain vers l'état pré-diagnostique est toujours plus élevé chez les sujets de bas niveau d'études que chez ceux de haut niveau. Le graphique nous montre également que l'hypothèse de proportionnalité des risques dans le modèle ajusté n'est pas vérifiée. En effet, nous constatons que le rapport entre les intensités de transition des sujets de bas et de haut niveau d'études n'est pas constant au cours du temps.

Les intensités de transition depuis l'état pré-diagnostique vers l'état démence estimées dans l'analyse stratifiée pour les sujets CEP- et CEP+ sont présentées sur la figure 4.7. L'interprétation des paramètres issus du modèle stratifié révèle une espérance du délai de survenue d'une démence plus long chez les sujets de bas niveau d'études par rapport à ceux de haut niveau. Toutefois, les paramètres ne permettent pas de saisir immédiatement le résultat original suivant : l'étude des intensités de transition dans chacun des deux échantillons montre que le risque de transition vers la démence est plus élevé chez les sujets de haut niveau d'études durant les 5 annnées suivant l'accélération du déclin cognitif alors que ce risque devient plus fort chez les sujets de bas nivau d'étude au-delà de ces 5 ans. L'hypothèse de proportionnalité des risques faite dans le modèle ajusté est donc clairement inadaptée.

Association risque de décès-cognition identique selon le CEP

Le modèle ajusté dans la formulation actuelle fait l'hypothèse d'une association entre

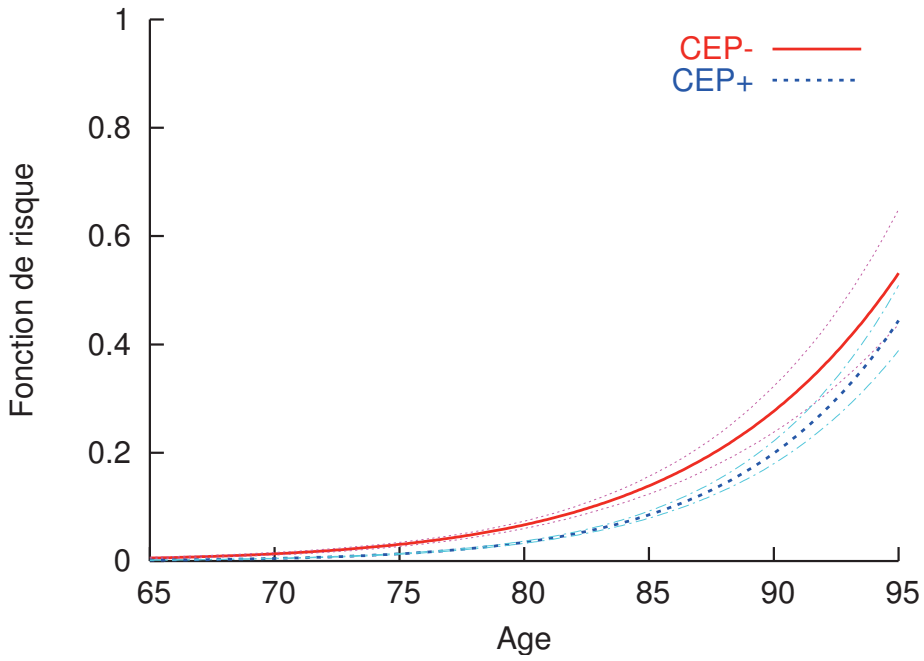

Fig. 4.6 : Intensités de transition de l'état sain vers l'état latent, obtenues par l'analyse stratifiée pour les sujets de haut et bas niveau d'études ainsi que leur intervalle de confiance respectif

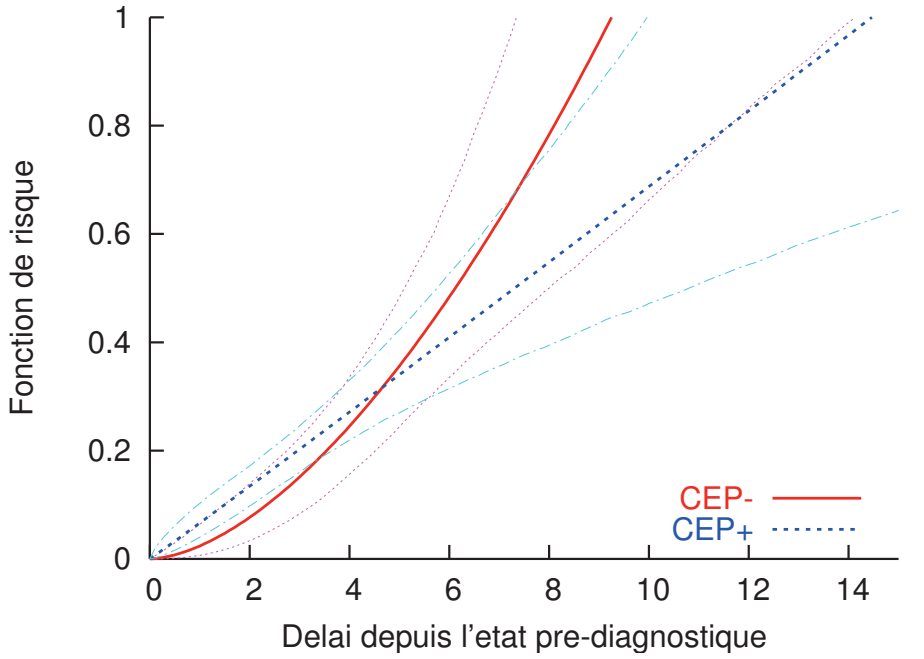

Fig. 4.7 : Intensités de transition de l'état latent vers l'état démence, obtenues par l'analyse stratifiée pour les sujets de haut et bas niveau d'études ainsi que leur intervalle de confiance respectif

niveau cognitif et décès identique quel que soit le niveau d'études. L'analyse stratifiée montre que cette hypothèse n'est pas vérifiée. Au regard de l'estimation du paramètre η dans l'analyse stratifiée, nous pouvons constater que l'association du niveau cognitif courant sur le risque de décès est différente selon le niveau d'études : $\eta = -0.25$ ($IC_{95\%}$=[-0.30 ;-0.19]) chez les sujets CEP+ contre $\eta = -0.17$ ($IC_{95\%}$=[-0.19 ;-0.13]) chez les sujets CEP-.

4.5.4 Synthèse de l'application

Les incohérences mises en évidence entre les analyses stratifiées et ajustées viennent probablement du non respect de l'hypothèse de proportionnalité des risques pour la transition vers la démence et du non respect de l'hypothèse d'un effet identique de la cognition sur le risque de décès quel que soit le niveau d'études. La minimisation du critère de sélection BIC nous amène alors à retenir l'analyse stratifiée sur le niveau d'études plutôt que l'analyse ajustée. Les principaux résultats épidémiologiques qui en sont issus sont les suivants :

- l'âge d'entrée dans la phase pré-diagnostique est plus précoce pour les sujets CEP- (82.40 ans) que pour les sujets CEP+ (86.38 ans) ;
- le délai de survenue d'une démence depuis l'entrée dans la phase pré-diagnostique est plus long chez les sujets CEP- (5.17 ans) que pour les sujets CEP+ (4.82 ans) ;
- le risque de démence chez les sujets CEP+ est plus élevé que chez les sujets CEP- durant les 5 années suivant l'entrée dans la phase pré-diagnostique, cette relation semble s'inverser au-delà de 5 ans (cf. figure 4.7) ; le risque de démence étant défini de manière paramétrique, il faudrait confirmer ce résultat à l'aide de modèles plus souples ;
- ajusté sur le niveau cognitif courant, les sujets CEP- ont un risque de décès plus élevé que les sujets CEP+ : le risque relatif de décès pour une perte de 5 points de Benton est de 3.5 tandis qu'il est de 2.3 pour les sujets CEP+.

Les scores moyens des sujets CEP- sont toujours plus faibles que ceux des sujets CEP+ mais le déclin dans la seconde phase d'évolution est plus prononcée chez les sujets CEP+

que chez les sujets CEP- (cf. Evolutions marginales $E(Y(t))$ des 2 sous-populations en figure 4.8).

Le modèle développé permet de décrire des évolutions conditionnelles aux âges de démence et/ou de décès. Sur la figure 4.8, pour les deux niveaux d'études, nous illustrons l'évolution moyenne d'un sujet vivant au temps t ($E(Y(t)|T_d^* > t)$) et l'évolution moyenne d'un sujet vivant et non dément au temps t ($E(Y(t)|T_d^* > t, T_e^* > t)$). Ce graphe illustre le fait que le phénomène de sélection par le décès apparaît plus marqué chez les sujets de bas niveau d'études. La différence entre les courbes $E(Y(t))$ et $E(Y(t)|T_d^* > t)$ est plus prononcée chez les sujets CEP- que chez les sujets CEP+. Pour les sujets CEP-, l'espérance du score cognitif sachant que le sujet est en vie ne décline pratiquement pas. La population des sujets CEP- survivants est extrêmement sélectionnée sur le plan cognitif. On observe d'ailleurs que le conditionnement supplémentaire sur l'absence de démence modifie peu l'évolution moyenne chez les sujets CEP- alors que l'évolution moyenne d'un sujet non dément et vivant chez les sujets CEP+ est nettement supérieure à l'évolution moyenne d'un sujet vivant chez les CEP+.

Pour les deux sous-populations, nous présentons sur la figure 4.9 l'évolution d'un sujet vivant à un âge donné, celle d'un sujet vivant et non dément à un âge donné et celle d'un sujet vivant et dément à un âge donné. Ce graphe confirme le déclin cognitif plus marqué des sujets CEP+ dans la phase pré-diagnostique.

A l'aide du modèle proposé, l'étude de l'impact du niveau d'études sur le vieillissement cognitif confirme le concept de capacité de réserve cognitive des sujets de haut niveau d'études (Stern et al., 1994). L'interprétation des résultats épidémiologiques présentés va dans le même sens que celle issue de l'analyse de Jacqmin-Gadda et al. (2006).

Nous retrouvons un âge à l'accélération du déclin plus précoce chez les sujets CEP- par rapport aux sujets CEP+. L'évolution moyenne du score de Benton est plus marquée dans la seconde phase d'évolution pour les sujets CEP+ que pour les sujets CEP-. Le délai de survenue d'une démence est plus court pour les sujets CEP+ que pour les sujets CEP-. Toutefois, l'ordre de grandeur de ce délai est différent de celui obtenu par Jacqmin-Gadda. Nous émettons deux hypothèses qui pourraient contraindre notre modèle à estimer un délai de survenue d'une démence plus court que dans l'approche de Jacqmin-Gadda

et al. (2006). Tout d'abord, nous avons supposé une évolution linéaire dans la seconde phase alors que Jacqmin-Gadda et al. (2006) proposent une évolution de forme cubique. Par ailleurs, nous supposons un risque nul de démence avant l'entrée dans l'état latent alors que Jacqmin-Gadda et al. (2006) autorisent une survenue d'une démence avant l'accélération.

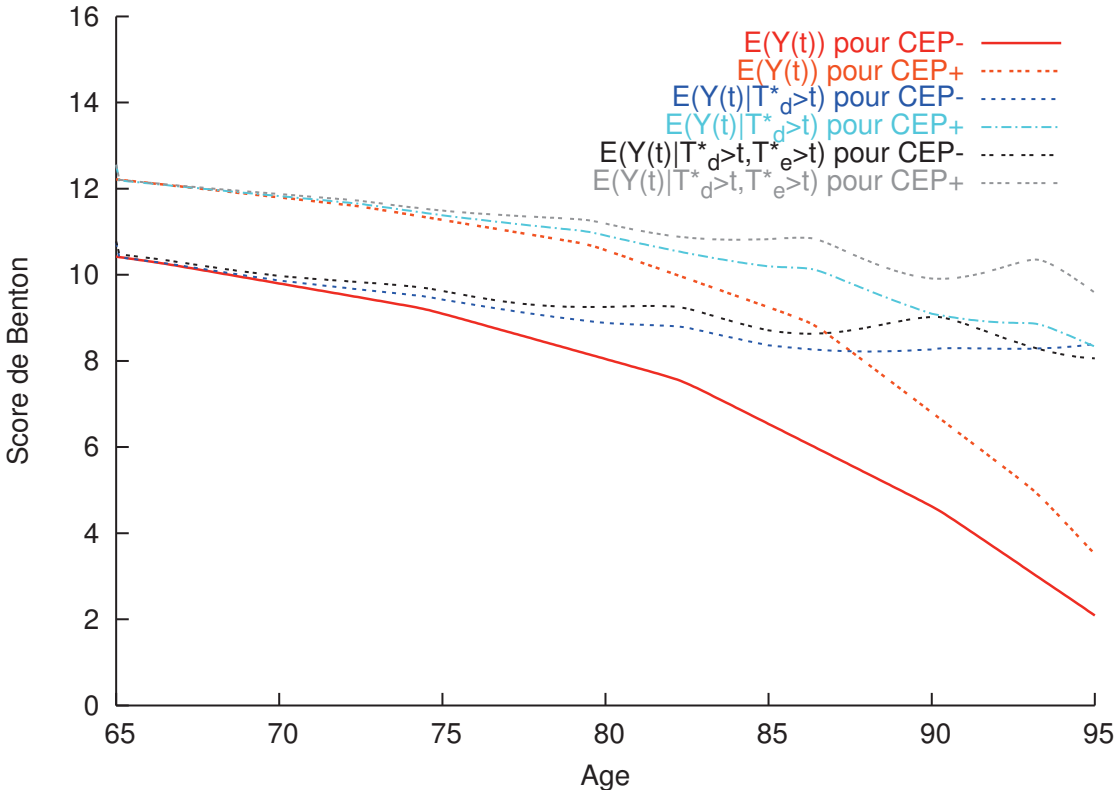

Fig. 4.8 : Evolution marginale du score de Benton et plusieurs évolutions estimées sachant une information sur la démence et/ ou le décès, stratifiées sur le CEP

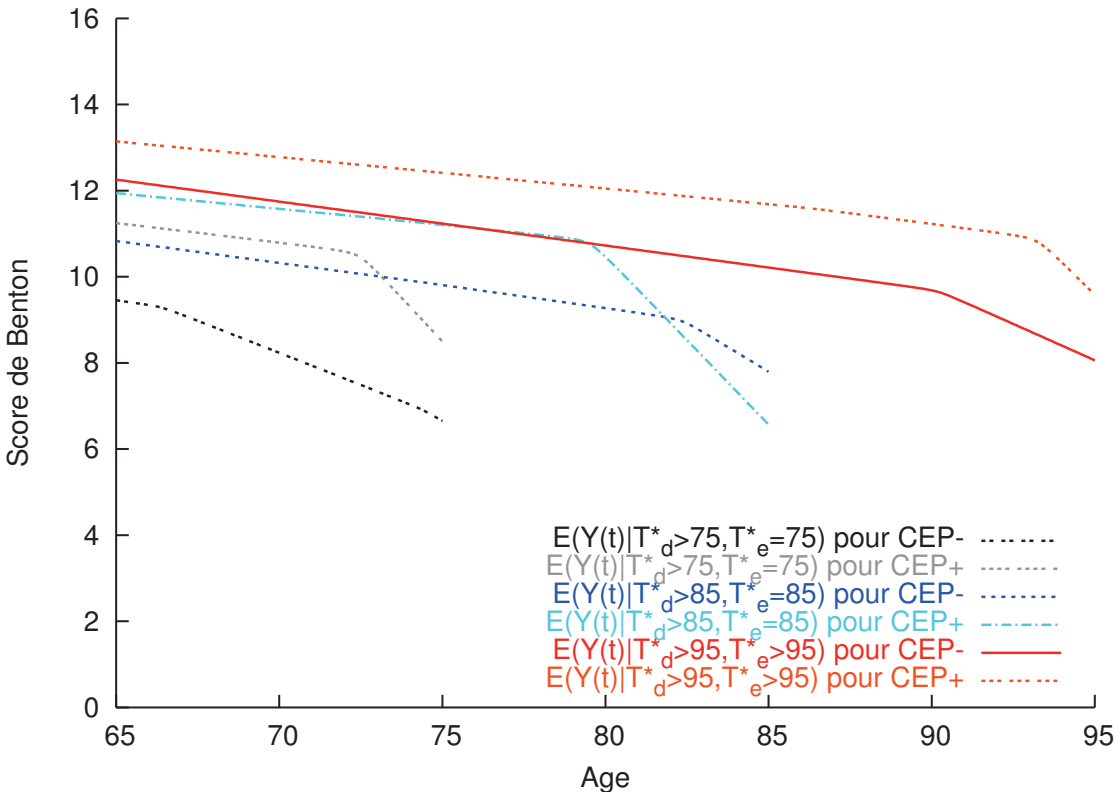

Fig. 4.9 : Evolutions conditionnelles d'un sujet vivant à un âge donné, d'un sujet vivant et non dément à un âge donné et d'un sujet vivant et dément à un âge donné pour les 2 niveaux d'études

4.6 Conclusion

Développé dans le contexte du vieillissement cérébral des personnes âgées, le modèle proposé combine un modèle linéaire mixte bi-phasique pour l'évolution d'un processus quantitatif gaussien et un modèle multi-états avec 3 états observés (sain, dément et décédé) et un état latent non-observé (pré-diagnostique). Il permet de considérer conjointement l'évolution d'un processus et le risque de survenue de deux événements semi-compétitifs. Ce modèle présente plusieurs forces qui en font une approche très intéressante pour l'étude du vieillissement cognitif des personnes âgées. Néanmoins, il possède quelques limites qu'il serait nécessaire de traiter pour affiner la modélisation.

4.6.1 Forces de l'approche

La modélisation de l'accélération du déclin dans la phase pré-démentielle est un élément essentiel pour la compréhension de l'histoire naturelle du vieillissement cognitif. Amieva et al. (2005, 2008) ont décrit une évolution cognitive en deux phases (déclin normal puis accéléré) à l'aide d'un modèle semi-paramétrique. Mais leur approche ne permet pas d'identifier l'âge à l'accélération du déclin. Les modèles à changement de pente aléatoire sont une solution qui a été mise en avant : Hall et al. (2000) ou encore Dominicus et al. (2008) définissent l'accélération du déclin cognitif exclusivement comme un changement de pente aléatoire. Toutefois, une modélisation conjointe semble plus pertinente pour considérer les associations entre l'évolution cognitive, le risque de démence et le risque de décès et ainsi éviter des biais. En effet, il a été montré qu'un déficit accru chez les sujets déments peut être observé avant le diagnostic de démence et que le décès est souvent précédé d'un déclin cognitif marqué.

Le modèle conjoint proposé à deux phases d'évolution permet d'estimer l'âge d'accélération du déclin et la durée de la phase pré-diagnostique ainsi que de tester l'impact de covariables sur ces différentes phases ($1^{\text{ère}}$ et $2^{\text{ème}}$ pentes, âge à l'accélération du déclin, durée de la phase pré-diagnostique).

L'un des intérêts majeurs de notre approche est de considérer l'existence d'un état pré-diagnostique non observé dont l'âge d'entrée correspond au changement de phase du processus d'évolution. Jacqmin-Gadda et al. (2006) supposent que la démence peut

survenir avant l'accélération du déclin (changement de pente) avec un risque de survenue d'une démence dépendant de l'âge au changement de pente. La modélisation du risque de démence que nous proposons est plus dynamique. Le risque de démence est modifié lorsque le sujet entre dans la phase pré-diagnostique. Plus précisément, nous avons choisi un modèle où le risque de démence est nul avant l'entrée dans l'état pré-diagnostique et augmente après en fonction du délai passé dans cette phase. L'état latent est donc considéré comme obligatoire avant la survenue d'une démence. Cet état latent peut être interprété comme une phase de Mild Cognitive Impairment à condition de considérer ce dernier comme irréversible et évolutif.

La modélisation conjointe du décès est un apport important de notre approche par rapport à celles déjà existantes pour modéliser le déclin cognitif (Hall et al., 2000, 2003; Jacqmin-Gadda et al., 2006; Dominicus et al., 2008). D'une part, l'étude de simulations montre l'intérêt de cette modélisation conjointe pour corriger les biais dans l'estimation des paramètres du modèle. D'autre part, la modélisation conjointe du décès fournit des outils de description originaux comme la description des évolutions de sujets vivants. Nous considérons que la modélisation du risque de décès est souple dans la mesure où il varie continuellement en fonction du niveau cognitif et non pas par étape en fonction d'un seuil sur la pente de déclin.

4.6.2 Limites de l'approche et perspectives

Hypothèses paramétriques

Une première limite concerne les hypothèses paramétriques du modèle. Bien que nous ayons montré l'adéquation de certaines d'entre elles, d'autres mériteraient d'être reconsidérées. Les fonctions de risque de base pourraient être modélisées par des approches non paramétriques à base de splines. La contrepartie serait un alourdissement de la méthode d'estimation. C'est pourquoi une approche semi-nonparamétrique (Gallant et Nychka, 1987) serait probablement un bon compromis, en particulier, en ce qui concerne l'intensité de transition vers l'état pré-diagnostique pour laquelle nous n'avons aucune observation, ce qui rend l'étude d'adéquation d'une transition paramétrique délicate.

L'hypothèse d'indépendance entre l'âge d'entrée dans l'état pré-diagnostique τ_i et les

effets aléatoires de l'évolution u_i n'est pas prouvée. Mais de précédentes analyses ont montré qu'une dépendance entre ces deux entités entrainaient des problèmes de convergence (Jacqmin-Gadda et al., 2006).

L'application du modèle ajusté aux données de la cohorte Paquid montre que les hypothèses de risques proportionnels ne sont pas adéquates pour étudier l'influence du niveau d'études. Des modifications peu coûteuses peuvent être effectuées telles que la stratification du risque de démence selon le niveau d'études ou l'ajout d'une interaction entre niveau cognitif courant et niveau d'études pour le risque de décès. Ces deux améliorations permettraient notamment de reconsidérer le modèle ajusté et offriraient ainsi la possibilité de tester l'impact du niveau d'études sur les différentes phases du processus de vieillissement cognitif.

Aspect numérique

Une autre faiblesse de ce modèle porte sur la complexité de sa vraisemblance. En effet, l'intérêt de cette approche conjointe réside dans le lien entre les différents sous-modèles. Or, dans notre approche, ce lien se fait par l'intermédiaire de l'âge d'entrée dans la phase pré-diagnostique ainsi que par les effets aléatoires définissant l'évolution cognitive. Les effets aléatoires génèrent des intégrales multiples dans la fonction de vraisemblance que l'on calcule numériquement. Le calcul du maximum de vraisemblance est donc extrêmement coûteux en temps et conduit en pratique à limiter le nombre d'effets aléatoires individuels.

Nous avons évalué les propriétés des estimations du maximum de vraisemblance obtenues par l'algorithme de Marquardt. Nous avons préféré un algorithme de type Newton-Raphson pour ses qualités en terme de robustesse (peu de problème de convergence, critères d'arrêt stricts), de temps de calculs et de disponibilité immédiate des variances des paramètres, en comparaison à des méthodes bayésiennes notamment. Toutefois, au vu de la complexité de la vraisemblance et des temps de calculs observés, il pourrait être intéressant de se tourner vers d'autres algorithmes d'optimisation comme l'algorithme MCMC qui autoriserait une plus grande marge de manoeuvre dans l'ajout d'effets aléatoires pour assouplir la modélisation.

Approfondissement de l'étude du biais lié au décès

L'étude de simulations que nous avons réalisée porte notamment sur l'aptitude de notre approche à corriger les biais liés à la censure informative dûe au décès. Il serait intéressant de prolonger cette étude afin d'évaluer l'importance de ces biais pour les paramètres de régression mesurant l'association avec des covariables.

Par ailleurs, les résultats montrent que le biais sur l'espérance du délai de survenue d'une démence est relativement faible (environ 10%) même lorsque l'échantillon simulé comporte une proportion importante de cas de démences non observés (environ 35% de cas non observés car le décès est survenu avant la visite suivant l'apparition de la démence). Notre hypothèse est que ce biais est faible en raison de la modélisation conjointe de l'évolution d'un test cognitif. Les mesures répétées du test apportent de l'information sur le risque de démence du patient entre sa dernière visite et son décès. Il serait intéressant de comparer ce biais à celui que l'on observerait pour un simple modèle de survie pour la démence.

Estimation des variances des fonctions d'intérêt

Dans les simulations effectuées, nous avons calculé les variances des fonctions d'intérêt (espérances de l'âge à l'entrée dans l'état latent, du délai de survenue d'une démence et des pentes d'évolution) à l'aide de la "delta-method". La construction d'intervalles de confiance issus du calcul de ces variances repose sur une hypothèse de normalité des estimateurs de ces fonctions d'intérêt qui n'est pas prouvée. Dans l'application, nous avons donc opté pour l'estimation d'intervalles de confiance par une technique de Bootstrap paramétrique. Celle-ci fournit une variance bootstrapée de ces fonctions qui semble plus faible que celles calculées par "delta-method". Une variance trop importante des paramètres définissant la fonction d'intérêt peut rendre non valide l'utilisation de la "delta-method" qui repose sur une approximation effectuée au voisinage des paramètres. La méthode par Bootstrap paramétrique pour l'estimation d'intervalles de confiance nous paraît plus fiable car elle ne repose ni sur une hypothèse de normalité, ni sur une approximation. Pour confirmer cette hypothèse, nous pourrions conduire une étude de simulations complémentaires pour comparer la "delta-method" et le Bootstrap paramétrique pour le calcul des intervalles de confiance des fonctions d'intérêt à partir de différentes valeurs des variances des para-

mètres de départ.

Test de l'existence de l'état latent

La formulation actuelle du modèle suppose l'existence d'une phase de déclin accélérée pour la totalité des sujets. Les sujets peuvent toutefois sortir de l'étude ou décéder avant la survenue de cet état pré-diagnostique. Or, selon les caractéristiques des sujets ou selon la dimension cognitive étudiée, l'hypothèse d'une accélération systématique du déclin est peut-être trop forte. Il est réaliste de penser que certains sujets ne sont pas à risque de transiter vers l'état pré-diagnostique ou que certaines dimensions cognitives ne présentent pas d'accélération du déclin avant la démence. Il serait donc très intéressant d'un point de vue épidémiologique de pouvoir tester l'existence de cet état pré-diagnostique. Cependant, cela requiert une flexibilité plus importante du modèle avec une dépendance du risque de démence sur l'évolution des tests psychométriques qui ne passe pas uniquement au travers de l'âge d'entrée dans la phase pré-diagnostique. Cela se traduirait par une complexification du modèle, limitée par des difficultés d'ordre numérique.

Autres applications

L'approche proposée permet de traiter conjointement l'évolution d'un processus caractérisé par des mesures répétées d'un marqueur quantitatif et la survenue de deux événements semi-compétitifs. Le travail que nous avons effectué montre l'intérêt de cette approche pour la modélisation du déclin cognitif. Elle pourrait être appliquée à d'autres pathologies. Par exemple, dans la surveillance du cancer de la prostate, on constate une diminution marquée du marqueur de PSA (Prostate Specific Antigen) après l'arrêt d'un traitement, suivie d'une stabilisation ou d'une augmentation progressive caractérisant une possible rechute. Le modèle proposé pourrait servir à étudier cette évolution en deux phases et le risque de rechute en modélisant conjointement l'évolution bi-phasique, le risque de rechute et le risque de décès. Il serait alors possible d'estimer l'influence de la pente dans la seconde phase sur le risque de rechute tout en tenant compte de la survenue du décès. Dans l'infection par le VIH, le modèle proposé pourrait permettre d'estimer la réaugmentation de la charge virale précédant le passage au stade SIDA.

Chapitre 5

Discussion générale

Au cours de cette recherche, nous avons réalisé deux travaux appliqués à l'étude du déclin cognitif des personnes âgées. Le premier concerne l'utilisation de méthodes permettant de tenir compte des sorties d'étude informatives dans la modélisation de l'évolution d'un processus longitudinal. Le second porte sur la modélisation de l'accélération d'un déclin pré-diagnostique et la survenue de deux événements concurrentiels. Ces deux approches présentent des concepts différents de modélisation, la notion de classes latentes pouvant être opposée à celle d'état latent. Néanmoins, elles visent toutes les deux la prise en compte de données incomplètes dans les études longitudinales et apparaissent pertinentes dans la modélisation de l'histoire naturelle du vieillissement cognitif. En conclusion à ce travail de thèse, nous souhaitons revenir sur les points de convergence de ces deux approches.

5.1 Modélisation des données incomplètes

Les travaux présentés dans les chapitres 3 et 4 portent tous les deux sur le traitement de données incomplètes. Dans le premier travail, nous nous sommes intéressés à des méthodes pour traiter des données manquantes dans l'analyse d'un processus longitudinal. Les approches considérées ne distinguent pas les différentes causes de sorties d'étude. L'approche par PMM se révèle simple d'utilisation et d'interprétation, ce qui en fait une approche intéressante pour la réalisation d'une analyse de sensibilité. En revanche, l'approche par classes latentes montre que l'interprétation des paramètres doit se faire avec précaution et

des analyses complémentaires sont nécessaires pour bien comprendre l'ajustement effectué sur les classes latentes.

Dans le second travail, nous nous intéressons aux sorties d'études dont la cause est connue. Nous attachons à évaluer l'impact du décès sur le processus d'évolution longitudinal. En raison de la collecte en temps discrets des diagnostics de démence, le décès est une source de données manquantes informatives dans le contexte du vieillissement cognitif puisqu'un décès peut survenir avant la visite de diagnostic de démence. Nous mettons également en oeuvre plusieurs développements dans le modèle proposé pour tenir compte des données de survie incomplètes (troncature à gauche, censure à droite et par intervalle).

5.2 Modèles à variables latentes

Pour modéliser l'histoire naturelle des maladies chroniques telle que la maladie d'Alzheimer, les modèles à variables latentes (processus latent, classes latentes, état latent) permettent de considérer des aspects non observables (évolution cognitive, hétérogénéité des évolutions, accélération du déclin) et améliorent ainsi la compréhension des phénomènes sous-jacents. Le modèle de Proust-Lima et al. (2006) permet de modéliser simultanément l'évolution de plusieurs marqueurs en les considérant comme différentes mesures imparfaites d'un même processus latent. Dans les deux modèles étudiés aux chapitres 3 et 4, les classes latentes et l'état latent permettent de prendre en compte l'hétérogénéité de l'évolution cognitive.

Toutefois, une différence fondamentale existe entre les deux approches. Le modèle à classes latentes tient compte de l'hétérogénéité de la population de manière statique. Les profils d'évolution peuvent être différents d'une classe à l'autre, ce qui permet de mettre en évidence des évolutions variées associées à des risques différents de survenue de l'événement. Mais au sein d'une classe latente, les profils d'évolution sont figés, seuls les effets aléatoires permettent une variabilité inter-individuelle. L'approche par état latent apparaît plus dynamique. Les sujets changent de profils d'évolution au cours du temps avec une accélération du déclin plus ou moins tardive.

Néanmoins, d'un point de vue épidémiologique, il est peu probable que l'ensemble des sujets âgés soient à risque de connaître une accélération du déclin précédant le diagnostic

de démence. L'accélération du déclin cognitif peut en effet ne concerner qu'une partie de la population des personnes âgées, l'autre partie ayant un déclin lent et régulier. Une perspective qui pourrait servir de point de convergence entre les deux approches serait de considérer une hétérogénéité sous-jacente à la population lorsqu'on étudie l'accélération pré-démentielle du déclin. Une solution envisageable serait alors de combiner le modèle conjoint à état latent proposé et un modèle à fraction de risque guérie (Cure model) qui suppose qu'une partie de la population ait un risque nul de démence et d'accélération du déclin.

D'un point de vue théorique, les modèles conjoints offrent des perspectives très vastes de modélisation et donc de compréhension de l'histoire naturelle. Mais une limite pratique, déjà évoquée, est le calcul du maximum de vraisemblance qui peut être délicat dans ces modèles complexes. Toutes les perspectives de développements des approches étudiées allongeront le temps d'estimation des paramètres sous réserve que la procédure d'optimisation converge. Il apparaît nécessaire de penser à un algorithme d'estimation alternatif permettant un gain de temps sans détérioration de la qualité des estimations. Les futurs développements de ce type d'approches ne peuvent se faire sans une réflexion importante concernant les difficultés numériques et les temps de calculs.

5.3 Pertinence pour la modélisation du vieillissement cognitif

Les travaux réalisés ont été motivés par l'étude du déclin cognitif des personnes âgées. Les deux approches présentées permettent d'étudier des aspects de l'histoire naturelle du vieillissement cognitif qui sont mal connus.

L'intérêt de l'approche par classes latentes est de modéliser l'hétérogénéité non observée des évolutions cognitives au sein de la population des personnes âgées. L'évolution de marqueurs psychométriques caractérise l'évolution cognitive latente. Différents profils d'évolution peuvent être mis en évidence et sont associés à des risques différents de survenue de la démence (Proust-Lima et al., 2009) ou de survenue d'une sortie d'étude (Dantan et al., 2008).

Dans l'étude du vieillissement, plusieurs événements comme la survenue de la démence et le décès peuvent être fortement liés à l'évolution d'un processus et, par ailleurs, intervenir de manière concurrentielle. Leur modélisation conjointe permet donc de mieux comprendre leur dépendance mutuelle. Le modèle conjoint à état latent développé dans le chapitre 4 permet d'évaluer l'association entre l'évolution cognitive et la survenue d'une démence en tenant compte de la censure informative liée au décès. Il permet d'identifier l'âge d'accélération du déclin cognitif chez les sujets déments et de décrire les évolutions de sujets normaux, de sujets pré-déments ou déments.

5.4 Outils pronostiques pour le diagnostic précoce

La notion d'outils pronostiques intéresse au plus haut point les épidémiologistes ainsi que les autorités de santé. Il serait en effet utile de pouvoir prédire les risques de survenue d'événements cliniques pour pouvoir diagnostiquer précocément la survenue d'une démence. L'identification des sujets à haut risque de démence est nécessaire pour une prise en charge adaptée et une mise en place précoce de traitements afin d'agir sur le processus de dégradation des fonctions cognitives. Le diagnostic précoce est d'autant plus important que la recherche médicale sur le développement de nouveaux traitements montre des signes prometteurs.

Les modèles conjoints sont une voie de développement de ces outils. En effet, la quantité et la diversité des informations prises en compte pour la modélisation sont plus importantes que pour les modèles de survie simple, les outils de prédiction en seraient donc d'autant plus fins. Le principe est d'utiliser l'information à disposition (mesures répétées de marqueurs quantitatifs) jusqu'à un temps t pour calculer la probabilité de survenue d'un événement entre t et $t + s$. Tout ajout de nouvelles mesures répétées permet un réajustement de la probabilité individuelle. Cependant, peu de travaux de développement d'outils pronostiques ont été réalisés à partir de modèles conjoints. Taylor et al. (2005) ont utilisé un modèle conjoint à effets aléatoires partagés alors que Proust-Lima et Taylor (2009) ont développé un outil pronostique à partir d'un modèle conjoint à classes latentes. Ces deux modèles ont été appliqués dans le contexte du cancer de la prostate pour faire de la prédiction de rechute de cancer en fonction des mesures répétées de l'antigène spécifique

de la prostate.

Cependant la principale difficulté méthodologique n'est pas tant le développement de l'outil de prédiction, une fois le modèle estimé, que sa validation. En effet, la prédiction individuelle de survenue de l'événement d'intérêt découle directement des paramètres estimés. L'évaluation de cet outil en terme de pouvoir discriminant (sensibilité et spécificité) et pouvoir prédictif est un aspect important dans la recherche méthodologique.

5.5 Conclusion générale

Dans ce travail de recherche, nous nous sommes intéressés à la modélisation de l'évolution cognitive et son association avec la survenue d'une démence, avec comme objectif la limitation des biais dans l'estimation dus à des données longitudinales et des données de survie incomplètes. Malgré des limites liées aux hypothèses paramétriques et aux difficultés numériques, cette recherche représente une contribution utile à la compréhension d'un processus d'évolution complexe tel que le vieillissement cognitif. Ces modèles sont utiles pour décrire l'histoire naturelle de la pathologie, comprendre le rôle de facteurs de risque et pour développer des outils de diagnostic précoce. Ces approches statistiques ont également toute leur place dans la modélisation d'autres maladies chroniques. A partir d'outils de modélisation statistiques existants, notre travail aura donc été de proposer des nouveaux modèles statistiques pour contribuer à l'avancée de ce champ de recherche.

Bibliographie

Aalen OO (1975). Non parametric inference for a family of counting processes. *The Annals of Statistics*, **6,** 534–545.

Aalen OO et Johansen S (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations.*Scandinavian Journal of Statistics*, **5,** 141–50.

Aalen OO (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, **7,** 1121–1137.

Aalen O et Husebye E (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine*, **10,** 1227–1240.

Aalen OO (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, **3,** 227–243.

Aalen O, Farewell VT, Angelis D, Day N et Gill N (1997). A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment : Application to AIDS prediction in England and Wales. *Statistics in Medicine*, **16,** 2191–2210.

Abramowitz M et Stegun IA (1972). *Handbook of Mathematical Functions*, Applied Mathematics Series, n°55, National Bureau of Standards.

Alioum A, Leroy VV, Commenges D, Dabis F et Salamon R (1998). Effect of gender, age, transmission category, and antiretroviral therapy on the progression of human immunodeficiency virus infection using multistate Markov models. Groupe d'épidémiologie clinique du SIDA en Aquitaine. *Epidemiology*, **9,** 605–612.

Alioum A et Commenges D (2001). MKVPCI : a computer program for Markov models

with piecewise constant intensities and covariates. *Computer Methods and Programs in Biomedicine*, **64,** 109–119.

Alioum A, Pérès K, Verret C, Regnault A et Barberger-Gateau P (2004). Determinants of Progression and Recovery through States of Disability : use of a Markov Model with Piecewise Constant Intensities and Covariates. In : V. Antonov, C. Huber, M. Nikulin and V. Polischook (eds), *Longevity, aging and degradation models*. St Petersburg, 2004, St Petersburg State Politechnical University, pp5–14.

Altman RM et Petkau AJ (2005). Application of Hidden Markov Models to Multiple Sclerosis Lesion Count Data. *Statistics in Medicine*, **24,** 2335–2344.

Altman R (2007). Mixed Hidden Markov Models : an extension of the Hidden Markov Model to the Longitudinal Data Setting. *Journal of the American Statistical Association* **102,** DOI :10.1198/016214506000001086

Amieva H, Jacqmin-Gadda H, Orgogozo JM, Le Carret N, Helmer C, Letenneur L, Barberger-Gateau P, Fabrigoule C et Dartigues JF (2005). The 9-year cognitive decline before dementia of the Alzheimer type : a prospective population-based study. *Brain*, **128,** 1093–1101.

Amieva H, Le Goff M, Millet X, Orgogozo JM, Pérès K, Barberger-Gateau P, Jacqmin-Gadda H et Dartigues JF (2008). Prodromal Alzheimer's disease : successive emergence of the clinical symptoms. *Annals of Neurology*, **64,** 492–498.

Andersen PK (1988). Multistate models ni survival analysis : a study of nephropathy and mortality in diabetes. *Statistics in Medicine*, **6,** 939–944.

Andersen PK, Borgan, Ø, Gill RD et Keiding N (1993). *Statistical Model Based on Counting Processes*. New-York : Springer-Verlag.

Bäckman L, Small B et Fratiglioni L (2001). Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain*, **58,** 853–858.

Bacon DW et Watts DG (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, **58,** 525–534.

Bandeen-Roche K, Migliorett D, Zeger SL et Rathouz PJ (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, **92,** 1375–1386.

Barberger-Gateau P, Fabrigoule C, Rouch I, Letenneur L et Dartigues JF (1999). Neuropsychological correlates of self-reported performance in instumental activities of daily living and prediction of dementia. *Journal of Gerontology series B : Psychological sciences and social sciences*, **54,** 293–303.

Barberger-Gateau P, Alioum A, Pérès K, Regnault A, Fabrigoule C, Nikulin M et Dartigues JF (2004). The contribution of Dementia to the Disablement Process and Modifying Factors. *Dementia and Geriatrics Cognitive Disorders*, **18,** 330–337.

Beck RJ et Paucker SG (1983). The Markov process in medical prognosis. *Medical Decision Making*, **3,** 419–458.

Benton A (1965). *Manuel pour l'applicatoin du Test de Rétention Visuelle. Applications cliniques et expérimentales*. Centres de Psychologies Appliquée, Paris, 2ème édition française.

Beunckens C, Molenberghs G, Verbeke G et Mallinckrodt C (2008). A latent-class mixture model for incomplete longitudinal gaussian data. *Biometrics*, **64,** 96–105.

Bird TD, Schellenberg GD, Wijsman EM et Martin GM (1989). Evidence for etiologic heterogeneity in Alzheimer's disease. *Neurobiology of Aging*, **10,** 432–434.

Boag JW (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society, Series B Methodological*, **11,** 15–53.

Bohning D (1995). A review of reliable maximum likelihood algorithms for semiparametric mixture models. *Journal of Statistical Planning and Inference*, **47,** 5–28.

Bohning D (2000). *Computer-assisted analysis of mixtures and applications : meta-analysis, disease mapping and others*. Chapman & Hall/CRC, Boca Raton, FL.

Bohning D et Seidel W (2003). Editorial : recent developments in mixture models. *Computational Statistics and Data Analysis*, **41,** 349–357.

Bosworth HB, Schaie KW et Willis SL (1999). Cognitive and sociodemographic risk factors for mortality in the Seattle Longitudinal Study. *Journal of Gerontology : Psychological Sciences*, **54,** P273–P282.

Brooks JO et Yesavage JA (1995). Identification of fast and slow decliners in Alzheimer disease : a different approach. *Alzheimer Disease and Association Disorders*, **9,** S19–25.

Brown ER, Ibrahim JG et DeGruttola V (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, **61,** 64–73.

Bunke H et Schulze U (1985). Approximation of change points in regression models. *In* T.Pukkila and S.Puntanen (Eds.), *Proceedings of the First International Tampere Seminar on Linear Statistical Models and Their Applications*, pp.161–171. Dept. of Math.Sci/Stat., University of Tampere, Finland.

Buschke H (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior*, **12,** 543–550.

Chen H et Chen J (2001). Large sample distribution of the likelihood ration test for normal mixtures. *Statistics and Probability Letters*, **52,** 125–133.

Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky ST et Ganguli M (2001). Patterns of cognitive decline in presymptomatic Alzheimer disease : a prospective community study. *Archives of general psychiatry*, **58,** 853–8.

Chi Y et Tseng C (2002). Comparison of Several Relative Risk Estimators with Interval-Censored Data. *Biometrical Journal*, **44,** 197–212.

Chi YY et Ibrahim JG (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, **62,** 432–445.

Clayton DG (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65,** 141–151.

Clayton D et Cuzyck J (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society Series A*, **148,** 82–117.

Commenges D, Letenneur L, Joly P, Alioum A et Dartigues JF (1998). Modelling age-specific risk : application to dementia. *Statistics in Medicine*, **17,** 1973–1988.

Commenges D (1999). Multi-state models in epidemiology. *Lifetime data analysis*, **5,** 309–321.

Commenges D (2002). Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research*, **11,** 162–182.

Commenges D et Joly P (2004). Multi-state model for Dementia, Institutionalization, and Death. *Communication in Statistics, Theory and Methods*, **33,** 1315–1326.

Cowppli-Bony P, Dartigues JF et Orgogozo JM (2006). Vascular risk factors and Alzheimer disease risk : epidemiological studies review. *Psychologie & Neuropsychiatrie du Viellissement*, **4,** 47–60.

Cox DR (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B Methodological*, **34,** 187–220.

Cox DR (1977). Nonlinear models, residuals and transformations. *Statistics : A Journal of Theoretical and Applied Statistics*, **8,** 3–22.

Cox DR and Oakes D (1984). *Analysis of survival data*. Chapman Hall, London.

Dantan E, Proust-Lima C, Letenneur L et Jacqmin-Gadda H (2008). Pattern Mixture Models and Latent Class Models for the Analysis of Multivariate Longitudinal Data with Informative Dropouts. *The International Journal of Biostatistics*, **4,** Iss 1, Article 14.

Dartigues JF, Gagnon M, Letenneur L, Barbeger-Gateau P, Commenges D, Evaldre M et Salamon R (1992). Principal lifetime occupation and cognitive impairment in a french elderly cohort (Paquid). *American Journal of Epidemiology*, **135,** 981–988.

Dartigues JF, Commenges D, Letenneur L, Barberger-Gateau P, Gilleron V, Fabrigoule C, Mazaux JM, Orgogozo JM et Salamon R (1997). Cognitive predictors of dementia in elderly community residents. *Neuroepidemiology*, **16,** 29–39.

Davidian M et Giltinan DM (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC, Boca Raton, FL.

Davidian M et Giltinan DM (2003). Nonlinear models for repeated measurement data : an overview and update. *Journal of Agricultural, Biological and Environmental Statistics*, **8,** 387–419.

De Boor C (1978). *A practical guide to splines*. Springer, New York.

De la Torre JC (2004). Is Alzheimer's disease a neurodegenerative or a vascular disorder ? data, dogma, and dialectics. *Lancet Neurology*, **3,** 184–90.

Dempster A, Laird N et Rubin D (1977). Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B Methodological*, **39,** 1–38.

Diggle PJ et Kenward MG (1994). Informative dropout in longitudinal data analysis. *Applied statistics*, **43,** 49–93.

Ding J et Wang JL (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, **64,** 546–556.

Dominicus A, Ripatti S, Pedersen NL et Palmgren J (2008). A random change point model for assessing variability in repeated measures of cognitive function. *Statistics in Medicine*, **27,** 5786–5798.

Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'brien J, Pasquier F, Robert P, Rossor M., Salloway S, Stern Y, Visser PJ et Scheltens P (2007). Research criteria for the diagnosis of Alzheimer's disease : revising the NINCDS-ADRDA criteria. *Lancet Neurol.*, **6,** 734–46.

Efron B et Tibshirani RJ (1993). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability vol.57. New York : Chapman and Hall.

Elashoff RM, Li G et Li N (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, **64,** 762–771.

Fabrigoule C, Letenneur L, Dartigues JF, Zarrouk M, Commenges D et Barberger-Gateau P (1995). Social and leisure activities and risk of dementia : a prospective longitudinal study. *Journal of the American Geriatrics Society*, **43,** 485–90.

Farewell VT (1982). The use of mixture models for the analysis of survival data with long-tems survivors. *Biometrics*, **38,** 1041–1046.

Farrer L, Cupples L, Haines J, Hyman B, Kukull W, Mayeux R, Myers R, Pericak-VanceM, Risch N et Van Duijn C (1997). Effects of age, sex and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *Journal of the American Medical Association*, **278,** 1349–56.

Fieuws S, Verbeke G et Molenberghs G (2007). Random-effects models for multivariate repeated measures. *Statistical Methods in Medical Research*, **16,** 387–397.

Fitzmaurice GM, Laird NM and Shneyer L (2001). An alternative parametrization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. *Statistics In Medicine*, **20,** 1009–21.

Flaten TP (2001). Aluminium as a risk factor in Alzheimer's disease, with emphasis on drinking water. *Brain Research Bulletin*, **55,** 187–196.

Fletcher R (2000). *Practical methods of optimization second ed.* Wiley and Sons, New York.

Flicker C, Ferris SH et Reisberg B (1991). Mild cognitive impairment in the elderly : predictors of dementia. *Neurology*, **41,** 1006–9.

Folstein MF, Folstein S et McHugh PR (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, **12,** 189–98.

Fratiglioni L, Launer LJ, Andersen K, Breteler MM, Copeland JR, Dartigues JF, Lobo A, Martinez-Lage J, Soininen H et Hofman A (2000). Incidence of dementia and major subtypes in Europe ; A collaborative study of population-based cohorts. Neurologic Diseases in tje Elderly Research Group. *Neurology*, **54,** S10–5.

Frydman H (1992). A nonparametric estimation procedure for periodically observed three-state Markov process, with pplication to AIDS. *Journal of Royal Statistical Society, Series B Methodological*, **54,** 853–866.

Gail M (1975). A review and critique of some models used in competing risks analysis. *Biometrics*, **32,** 209–222.

Gallant AR et Nychka DW (1987). Semi-Nonparametric Maximum Likelihood Estimation. *Econometrika*, **55,** 363–390.

Ganguli M, Seaberg EC, Ratcliff GG, Belle SH et DeKosky ST (1996). Cognitive stability over 2-years in a rural elderly population : the MoVIES project. *Neuroepidemiology*, **15,** 42–50.

Ganguli M, Dodge HH, Shen C et DeKosky ST (2004). Mild Cognitive Impairment, amnestic type : An epidemiologic study. *Neurology* **63,** 115–121.

Ganiayre J, Commenges D et Letenneur L (2008). A latent process model for dementia and psychometric tests. *Lifetime Data Analysis*, **14,** 115–133.

Garre FG (2008). A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society, Series A*, **171,** 299–308.

Geman S et Geman D (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6,** 721–741.

Gentleman RC, Lawless JF, Lindsey JC et Yan P (1994). Multi-state markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine*, **13,** 805–821.

Ghosh J et Sen P (1985). On the asymptotic performance of the log-likelihood ratio statistics for the mixture model and related results. In Cam, L.L. et Olshen , R., editors, *Proceedings of the Berkeley Conference in honor of J. Neyman and J. Kiefer*, Belmont, CA. Wadsmorth.

Gibaldi M et Perrier D (1982). *Pharmacokinetics (2nd edition)*. New York : Dekker.

Gilks W, Richardson S et Spiegelhalter D (1996). *Markov Chain Monte Carlo Methods in Practice*. Chapman & Hall : New York.

Griffiths D et Miller A (1975). Letter to the Editor. *Technometrics*, **17,** 281.

Guihenneuc-Jouyaux C, Richardson S et Longini IM (2000). Modeling Markers of disease Progression by a Hidden Markov Model Process : Application to Characterizing CD4 Cell Decline. *Biometrics*, **56,** 733–741.

Grober E, Lipton R, Hall C et Crystal H (2000). Memory impairment on free and cued selective reminding predicts dementia. *Neurology*, **54,** 827–832.

Guo J, Wall M et Amemyia Y (2006). Latent class regression on latent factors. *Biostatistics*, **7,** 145–63.

Hall CB, Lipton RB, Sliwinski M et Stewart WF (2000). A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's diseasee. *Statistics in Medicine*, **19,** 1555–1566.

Hall CB, Ying J, Kuo L, Sliwinski M, Buschke H, Katz M et Lipton RB (2001). Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Statistics in Medicine*, **20,** 3695–3714.

Hall CB, Ying J, Kuo L et Lipton RO (2003). Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics & Data Analysis*, **42,** 91–109.

Harvey D, Beckett L et Mungas D (2003). Multivariate modeling of two associated cognitive outcomes in a longitudinal study. *Journal of Alzheimer's disease*, **5,** 357–365.

Harville D (1977). Maximum Likelihood Approaches to Variance Component Estimation and Related Problems. *Journal of the American Statistical Association*, **72,** 320–339.

Hashemi R, Jacqmin-Gadda H et Commenges D (2003). A latent process for joint modeling of events and marker. *Lifetime Data Analysis*, **9,** 331–343.

Hastings W (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, **57,** 97–109.

Hawkins DS, Allen DM et Stromberg AJ (2001). Determining the number of components in mixtures of linear models. *Computational Statistics and Data Analysis*, **38,** 15–48.

Helmer C, Joly P, Commenges D et Dartigues JF (2001). Mortality with dementia : Results from a French Prospective Community-based cohort. *American Journal of Epidemiology*, **154,** 642–8.

Helmer C, Peuchant E, Letenneur L, Bourdel-Marchasson I, Larrieu S, Dartigues JF, Dubourg L, Thomas MJ et Barberger-Gateau P (2003). Association between antioxydant nutritional indicators and the incidence of dementia : results from the PAQUID prospective cohort study. *European Journal of Clinical Nutrition*, **57,** 1555–1561.

Henderson R, Diggle P et Dobson A (2000). A joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1,** 465–80.

Henderson R, Diggle P et Dobson A (2002). Identification and efficacy of longitudinal markers for surivival. *Biostatistics*, **33,** 33–50.

Hougaard P, Myglegard P et Borch-Johnsen K (1994). Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics*, **50,** 1178–1188.

Hougaard P (1995). Frailty models for survival data. *Lifetime data analysis*, **1,** 255–273.

Hougaard P (1999). Multi-state models : a review. *Lifetime data analysis*, **5,** 239–264.

Hougaard P (2000). *Analysis of Multivariate Survival Data*. Springer.

Howieson DB, Dame A, Camicioli R, Sexton G, Payami H et Kaye JA (1997). Cognitive markers preceding Alzheimer's dementia in the healthy oldest old. *Journal of the American Geriatrics Society*, **45,**584–9.

Humphreys K (1998). The Latent Markov Chain With Multivariate Random Effects. *Sociological Methods & Research*, **26,** 269–299.

Isaacs B et Kennie AT (1973). The Set test as an aid to the detection of dementia in old people. *The British Journal of Psychiatry*, **123,** 467–70.

Jacqmin-Gadda H, Fabrigoule C, Commenges D et Dartigues JF (1997a). A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *American Journal of Epidemiology*, **145,** 498–506.

Jacqmin-Gadda H, Fabrigoule C, Commenges D et Dartigues JF (1997b). Longitudinal study of cognitive aging in non-demented elderly subject. *Revue d'Epidémiologie et de Santé Publique*, **45,** 363–72.

Jacqmin-Gadda H, Joly P, Commenges D, Binquet C et Chêne G (2002). Penalized likelihood approach to estimate a smooth mean curve on longitudinal data. *Statistics in Medicine*, **21,** 2391–2402.

Jacqmin-Gadda H, Commenges D et Dartigues JF (2006). Random changepoint model for joint modeling of cognitive decline and dementia. *Biometrics*, **62,** 254–260.

Jacqmin-Gadda H, Proust-Lima C, Taylor JMG et Commenges D (2009). A score test for conditional independence between longitudinal outcome and time-to-event given the classes in the joint latent class model. *Biometrics*, DOI :10.1111/j.1541-0420.2009.01234.x

Jicha GA, Parisi JE, Dickson DW, Johnson K, Cha R, Ivnik RJ, Tangalos EG, Boeve BF, Knopman DS, Braak H et Petersen RC (2006). Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Archives of Neurology*, **63,** 674–681.

Joly P, Commenges D et Letenneur L (1998). A penalized likelihood approach for arbitrarily censored and truncated data : application to age-specificincidence of dementia. *Biometrics*, **54,** 185–194.

Joly P et Commenges D (1999). A penalized likelihood approach for a progressive three-state model with censored and truncated data : application to AIDS. *Biometrics*, **55,** 887–890.

Joly P, Letenneur L, Alioum A et Commenges D (1999). PHMPL : a computer program for hazard estimation using a penalized likelihood method with interval-censored and left-truncated data. *Computer Methods and Programs in Biomedicine*, **60,** 225–231.

Joly P, Commenges D, Helmer C et Letenneur L (2002). A penalized likelihood approach for an illness-death model with interval-censored data : application to age-specific incidence of dementia. *Biostatistics*, **3,** 433–443.

Joly P, Durand C, Helmer C et Commenges D (2009). Estimating life expectancy of dementd and institutionalized subjects from interval-censored observations of a Multi-State model. *Statistical Modelling*, In press.

Joseph L, Wolfson DB, Belisle P Brooks JO, Mortimer JA, Tinklenberg JR et Yesavage JA (1999). Taking account of between-patient variability when modeling decline Alzheimer's disease. *American Journal of Epidemiology*, **149,** 963–73.

Kalbfleisch JD et Prentice RL (1980). *The Statistical Analysis of Failure Time Data*. Wiley , New York.

Kalbfleisch JD et Lawless JF (1989). Inferences based on retrospective ascertainment : An analysis of the data on transfused-related AIDS. *Journal of the American Statistical Association*, **84,** 360–372.

Kaplan EL et Meier P (1958). Non parametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53,** 457–481.

Karlin S et Taylor HM (1975). *A first course in stochastic processes*. Chapter 4, Academic press, second edition.

Karp A, Kareholt I, Qiu C, Bellander T, Winblad B et Fratiglioni L (2004). Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease. *American Journal of Epidemiology*, **159,** 175–83.

Katzman R (1993). Education and the prevalence of dementia and Alzheimer's disease. *Neurology*, **43,** 13–20.

Kay R (1986). A Markov models for analyzing cancer markers and disease states in survival analysis. *Biometrics*, **42,** 855–865.

Keiding N, Klein JP et Horowitz MM (2001). Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine*, **20,** 1871–1885.

Kenward M et Molenberghs G (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, **13,** 236–247.

Klein JP, Keiding N et Copelan EA (1993). Plotting summary predictions in multistate survival models - probabilities of relapse and death in remission for bone-marrow transplantation. *Statistics in Medicine*, **12,** 2315–2332.

Klein JP et Moeschberger, ML (1997). *Survival analysis : techniques for censored and truncated data*. Springer-Verlag, New York.

Korsgaard IR et Andersen AH (1998). The additive genetic gamma frailty model. *Scandinavian Journal of Statistics*, **25,** 255–269.

Krogh A (1998). An introduction to hidden Markov Model for Biological Sequences. *Computational Methods in Molecular Biology*, eds S.L. Salzberg, D.B. Searls, and S. Kasif, Amsterdam : Elsevier, pp 45–63.

Kuk AYC et Chen CH (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79,** 531–541.

Laird NM et Ware JH (1982). Random-effects models for longitudinal data. *Biometrics*, **38,** 963–974.

Lagakos S (1979). General right-censoring and its impact on the analysis of survival data. *Biometrics*, **35,** 139–156.

Lagakos SW, Barraj LM et De Gruttola V (1988). Nonparametric analysis of truncated survival data with application to AIDS. *Biometrika*, **75,** 515–523.

Larrieu S, Letenneur L, Orgogozo JM, Fabrigoule C, Amieva H, Le Carret N, Barberger-Gateau P et Dartigues JF (2002). Incidence and outcome of mild cognitive impairment in a population-based prospective cohort *Neurology*, **59,** 1594–1599.

Larrieu S, Letenneur L, Helmer C, Dartigues JF et Barberger-Gateau P (2004). Nutritional factors and risk of incident dementia in the PAQUID longitudinal cohort. *Journal of Nutrition, Health & Aging*, **8,** 150–4.

Law NJ, Taylor JM et Sandler H (2002). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics*, **3,** 547–63.

Lawless JF (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

Lee EW, Wei LJ and Amato DA (1992). *Cox-type regression analysis for large numbers of small groups of correlated failure time observations*. Survival Analysis : State of the arts, JP Klein, PK Goel Eds, 237–247.

Letenneur L, Commenges D, Dartigues JF et Barbeger-Gateau P (1994). Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *International Journal of Epidemiology*, **23,** 1256–61.

Letenneur L, Gilleron V, Commenges D, Helmer C, Orgogozo J et Dartigues JF (1999). Are sex and educational level independent predictors of dementia and Alzheimer's disease ? Incidence data from the PAQUID project. *Journal of neurology, neurosurgery and psychiatry*, **66,** 177–183.

Liang KY et Zeger SL (1986). Longitudinal data analysis using generalized linear models. *Biometrics*, **73,** 13–22.

Lin DY et Wei LJ (1984). The robust inference for the Cox proportional hazards models. *Journal of the American Statistical Association*, **84,** 1074–1078.

Lin H, McCulloch CE, Turnbull BW, Slate EH et Clark LC (2000). A latent class mixed model for analysing biomarker trajectories with irregularity scheduled observations. *Statistics in Medicine*, **19,** 1303–18.

Lin H, McCulloch CE, Turnbull BW et Slate EH (2002a). Latent class models for joint analysis of longitudinal biomarker and event process data : application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, **97,** 53–65.

Lin H, McCulloch CE et Mayne ST (2002b). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, **21,** 2369–2382.

Lin H, McCulloch CE et Rosenbeck RA (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies *Biometrics*, **60,** 295–305.

Lindstrom MJ et Bates DM (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, **83,** 1014–1022.

Linn RT, Wolf PA, Bachman DL, Knoefel JE, Cobb JL, Belanger AJ, Kaplan EF et D'Agostino RB (1995). The 'Preclinical phase' of probable Alzheimer disease. *Archives of Neurology*, **52,** 485–490.

Little RJA (1993). Pattern Mixture models for multivariate incomplete data. *Journal of American Statistical Association*, **88,** 125–34.

Little RJA (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of American Statistical Association*, **90,** 112–21.

Little TD, Lindenberger U et Maier H (2000). Selectivity and generalizability in longitudinal research : On the effects of continuers and dropouts. In TD Little, KU Schanabel and Baumert (Eds.), *Modelling longitudinal and multilevel data : Practical issues, applied approaches and specific examples*, (pp. 187–20). Mahwah, NJ : Erlbaum.

Little RJA et Rubin DB (2002). *Statistical analysis with missing data (2nd edition)*. Wiley series in probability and statistics, New York.

Longini IR et Clark WS (1989). Statistical analysis of the stages of HIV infection using Markov model. *Statistics in Medicine*, **8,** 831–843.

Luchsinger JA et Mayeux R (2004a). Dietary factors and Alzheimer's disease. *Lancet Neurology*, **3,** 579–87.

Luchsinger JA et Mayeux R (2004b). Cardiovascular risk factors and Alzheimer's disease. *Current Atherosclerosis Reports*, **6,** 261–266.

Maller R et Zhou X (1996). *Survival analysis of long-term survivors*. Wiley, New York.

Marquardt DW (1963). An algorithm for least squares estimation of nonlinear parameters. *SIAM journal on Applied Mathematics*, **11,** 431–41.

Marshall G et Jones RH (1995). Multi-state models and diabetic retinopathy. *Statistics in Medicine*, **14,** 1975–1983.

Masur DM, Sliwinski M, Lipton RB, Blau MD et Crystal HA (1994). Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. *Neurologuy*, **44,** 1427–1432.

McCullagh P et Nelder JA (1989). *Generalized Linear Models*, New York : Chapman & Hall.

McCulloch, CE, Lin H, Slate EH et Turnbull BW (2002). Discovering subpopulation structure with latent class mixed models. *Statistics in Medicine*, **16,** 1587–601.

McCulloch CE et Searle SR (2004). *Generalized, linear, mixed models.* A Wiley-Interscience publication, New York.

McLachlan G (1987). On bootstrapping the likelihood ratio test statistic for the number of components in normal mixture. *Applied Statistics*, **36,** 318–324.

Metropolis N, Rosenbluth A, Rosenbluth M, Teller A et Teller E (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physisics*, **21,** 1087–1092.

Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH et Berg L (2001). Mild Cognitive Impairment represents early-stage Alzheimer disease. *Archives of Neurology*, **58,** 397–405.

Muthén B et Shedden K (1999). Finite mixture moeeling with mixture outcomes using the EM algorithm. *Biometrics*, **55,** 463–9.

Muthén B (2002). Beyond SEM : General latent variable modeling. *Behaviormetrika*, **29,** 81–117.

Nelson W (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14,** 945–965.

Nielsen GG, Gill RD, Andersen PK et Sorensen TIA (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, **19,** 25–43.

Nityasuddhi D et Bohning D (2003). Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances. *Computational Statistics and Data Analysis*, **41,** 591–601.

Odell P, Anderser PK et D'Agostino R (1992) Maximum likelihood Estimation for Interval-Censored Data Using a Weibull-Based Accelerated Failure Time Model. *Biometrics*, **48,** 951–959.

OPEPS (2006). Rapport sur la maladie d'Alzheimer et les maladies apparentées. Office parlementaire d'évaluation des politiques de santé.

Ownby RL, Crocco E, Acevedo A, John V et Loewenstein D (2006). Depression and risk for Alzheimer disease : systematic review, meta-analysis and metaregression analysis. *Archives of General Psychiatry*, **63,**530–8.

Palmer K, Wang HX, Backman L, Winblad B et Fratiglioni L (2002). Differential evolution of cognitive impairment in nondemented older persons : results from the Kungsholmen Project. *American Journal of Psychiatry*, **159,** 436–442.

Patterson HD et Thompson R (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58,** 545–554.

Peng Y et Dear KB (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, **56,** 237–243.

Perelson AS, Neumann AU, Markowitz M, Leonard JM et Ho DD (1996). HIV-1 Dynamics in Vivo : Virion Clearance Rate, Infected Cell Life-Span and Viral Generation Time. *Science*, **271,** 1582–1586.

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG et Kokmen E (1999). Mild cognitive impairment : clinical characterization and outcome. *Archives of Neurology*, **56,** 303–8.

Peto R (1973). Experimental survival curves for interval-censored. *Applied Statistics*, **22,** 86–91.

Proust C et Jacqmin-Gadda H (2005). Estimation of linear mixed models with a mixture of distribution for the random effects. *Computer Methods and Programs in Biomedicine*, **78,** 165–173.

Proust-Lima C, Jacqmin-Gadda H, Taylor JMG, Ganiayre J et Commenges D (2006). A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*, **62,** 1014–1024.

Proust-Lima C, Amieva H, Dartigues JF et Jacqmin-Gadda H (2007a). Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. *American Journal of Epidemiology*, **165,** 344–350.

Proust-Lima C, Letenneur L et Jacqmin-Gadda H (2007b). A non lineaire latent class model for joint analysis of multivariate longitudinal data and a binary outcome. *Statistics in Medicine*, **26,** 2229–45.

Proust-Lima C, Amieva H, Letenneur L, Orgogozo JM, Jacqmin-Gadda H et Dartigues JF (2008). Gender and education impact on brain aging : a general cognitive factor approach. *Psychology and Aging*, **23,** 608–20.

Proust-Lima C, Joly P, Dartigues JF et Jacqmin-Gadda H (2009). Joint modelling of multivariate longitudinal outcomes and a time-to-event : A non linear latent class approach. *Computational Statistics & Data Analysis*, **53,** 1142–1154.

Proust-Lima C et Taylor JMG (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA : a joint modeling approach. *Biostatistics*, **10,** 535–549.

Putter H, Van Der Hage J, De Bock GH, Elgalta R et Van de Velde CJH (2006). Estimation and Prediction in a Multi-state Model for Breast Cancer. *Biometrical Journal*, **48,** 366–380.

Rabbitt P, Diggle P, Holland F et McInnes L (2004). Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. *Journal of Gerontology*, **59B,** 84–97.

Rabbitt P, Lunn M et Wong D (2005). Neglect of dropout underestimates effect of death in longitudinal studies. *Journal of Gerontology : Psychological Sciences*, **60B,** P106–P109.

Rabbitt P, Lunn M et Wong D (2008). Death, dropout, and longitudinal measurements of cognitive change in old age. *Journal of Gerontology*, **63B,** 271–278.

Ramaroson H, Helmer C, Barberger-Gateau P, Letenneur L et Dartigues JF (2003). Prévalence de la démence et de la maladie d'Alzheimer chez les personnes de 75 ans et plus : données réactualisées de la cohorte Paquid. *Revue neurologique*, **159,** 405–411.

Ramsay J (1988). Monotone regression splines in action. *Statistical Science*, **3,** 425–461.

Redner RA et Walker HF (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, **26,** 195–239.

Richardson S et Green PJ (1997). On bayesian analysis mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B Methodological*, **59,** 731–792.

Ringman JM et Cummings JL (2006). Current and emerging pharmocological treatment options for dementia. *Behavioural Neurology*, **17,** 5–16.

Ripatti S, Gatz L, Pedersen NL et Palmgren J (2003). Three-State Frailty Model for Age at Onset of Dementia and Death in Swedish Twins. *Genetic Epidemiology*, **24,** 139–149.

Ritchie K, Artero S et Touchon J (2001). Classification criteria for mild cognitive impairment :a population-based validation study. *Neurology*, **56,** 37–42.

Robert C et Casella G (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York, seconde édition.

Rondeau V, Commenges D, Jacqmin-Gadda H et Dartigues JF (2000). Relation between aluminium concentrations in drinking water and Alzheimer's disease : an 8-year follow-up study. *American Journal of Epidemiology*, **152,** 59–66.

Rondeau V, Commenges D et Joly P (2003). Maximum penalized likelihood estimation in frailty models. *Lifetime Data Analysis*, **9,** 139–153.

Roy J et Lin X (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics*, **56,** 1047–1054.

Roy J et Lin X (2002). Annalysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates : changes in methadone treatment practices. *Journal of the American Statistical Association*, **97,** 40–52.

Roy J (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, **59,** 829–36.

Rubin DB (1976). Inference and missing data. *Biometrika*, **63,** 581–592.

Saint-Pierre P, Combescure C, Daures JP et Godard P (2003). The analysis of asthma control under a Markov assumption with use of covariates *Statistics in Medicine*, **22,** 3755–3770.

Salazar JC, Schmitt FA, Yu L, Mendiondo MM et Kryscio RJ (2007). Shared random effects analysis of multi-state Markov models : application to a longitudinal study of transition to dementia. *Statistics in Medicine*, **26,** 568–580.

Salthouse TA (1996). The processing-speed theory of adult age differences in cognition. *Psychological review*, **103,** 403–28.

Satten G et Sternberg M (1999). Fitting Semi-Markov Models to interval-Censored Data with Unknown Initiation Times. *Biometrics*, **55,** 507–513.

Scarmeas N, Albert SM, Manly JJ et Stern Y (2006). Education and rates if cognitive delicne in incident Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, **77,** 308–16.

Schmand B, Waltra G, Lindeboom J, Teunisse S et Jonker C (2000). Early detection of Alzheimer's disease using the Cambridge Cognitive Examination (CAMCOG). *Psychological Medicine*, **30,** 619–627.

Seber GAF et Wild CJ (1989). *Nonlinear regression*. Wiley series in probability and mathematical statistics.

Seltman HJ (2002). Hidden Markov Models for Analysis of Biological Rhythm Data. *Case Studies in Bayesian Statistics*, Vol.5, Springer-Verlag, pp. 397–405.

Shah A, Laird NM et Schoenfeld D (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association*, **92,** 775–779.

Silverman BW (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B Methodological*, **47,** 1–52.

Sliwinski MJ, Lipton RB, Buschke H et Stewart W (1996). The effects of preclinical dementia on estimates of normal cognitive functionning in aging *Journal of Gerontology series B : Psychological sciences and social sciences*, **51,** 217–25.

Sliwinski MJ, Hofer SM, Hall C, Buschke H et Lipton RB (2003a). Modelling memory decline in older adults : the importance of preclinical dementia, attrition, and chronological age. *Psychological and Aging*, **18,** 658–71.

Sliwinski M, Hofer S et Hall C (2003b). Correlated and coupled cognitive change in order adults with or without preclinical dementia. *Psychology and Aging*, **18,** 672–683.

Small BJ et Bäckman L (1997). Cognitive correlates of mortality : Evidence from a population-based sample of very old adults. *Psychology and aging*, **12,** 309–313.

Small BJ, Fratiglioni L, Viitanen MD, Winblad B et Bäckman L (2000). The course if cognitive impairment in preclinical Alzheimer disease. *Archives of Neurology*, **57,** 839–844.

Small BJ, Fratiglioni L, Von Strauss E et Bâckman L (2003). Terminal decline and cognitive performance in very old age : Does cause of death matter ?. *Psychology and aging*, **18,** 193–202.

Spiessens B, Verbeke G et Komárek A (2002). A SAS-macro for the classification of longitudinal profiles using mixtures of normal distributions in nonlinear and generalized linear model. http ://www.med.kuleuven.ac.be/biostat/research/software.htm

Stern Y, Gurland B, Tatemichi TK, Tang MX Wilder D et Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease (1994). *Journal of American Medical Association*, **271,** 1004–10.

Storandt L, Grant EA, Miller JP et Morris JC (2006). Progression in mild cognitive impairment (MCI) and preMCI : a comparison of diagnostic criteria. *Neurology*, **67,** 467–473.

Sun J (2001). Variance estimation of a survival function for interval-censored survival data. *Statistics in medicine*, **20,** 1249–1257.

Sy JP et Taylor JMG (2000). Estimation in a cox proportional hazards cured models. *Biometrics*, **56,** 227–236.

Taylor JMG, Yu M et Sandler HM (2005). Individualized predictions of disease progression following radiation therapyfor prostate cancer. *Journal of clinical oncology*, **4,** 815–825.

Therneau TM et Grambsch PM (2000). *Modeling Survival Data : Extending the Cox Model*. Springer, Statistics for Biology and Health.

Thiébaut R, Jacqmin-Gadda H, Chène G, Leport C et Commenges D (2002). Bivariate linear models using a SAS proc mixed. *Computer Methods and Programs in Biomedicine*, **69,** 249–256.

Thiébaut R, Jacqmin-Gadda H, Babiker A, Commenges D et la Collaboration CASCADE (2005). Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection. *Statistics in Medicine*, **24,** 65–82.

Thijs H, Molenberghs G, Michiels B, Verbeke G et Curran D (2002). Strategies to fit pattern-mixture models. *Biostatistics*, **3,** 245–265.

Tierney MC, Yao C, Kiss A et McDowell I (2005). Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. *Neurology*, **64,** 1853–1859.

Tombaugh TN et McIntyre NJ (1992). The Mini-Mental State Examination : A comprehensive review. *Journal of the American Geriatric Society*, **40,** 935–992.

Turnbull BW (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of Royal Statistical Society, Series B Methodological*, **38,** 290–295.

Valdois S, Joanette Y, Poissant A, Ska B et Dehaut F (1990). Heterogeneity in the cognitive profile of normal elderly. *Journal of Clinical and Experimental Neuropsychology*, **12,** 587–96.

Van Den Hout A et Matthews FE (2008). Multi-state analysis of cognitive ability data : A piecewise-constant model and a Weibull model. *Statistics in Medicine*, **27,** 5440–5455.

Van Den Hout A, Jagger C et Matthews FE (2009). Estimating life expectancy in health and ill health by using hidden Markov model. *Journal of the Royal Statistical Society, Series C Applied Statistics*, **58,** 449–465.

Vaupel JW et Yashin AI (1985). Heterogeneity's ruses : Somes surprising effect of selection on population. *Journal of the American Statistical Association*, **39,** 176–185.

Verbeke G et Lesaffre E (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91,** 217–221.

Verbeke G et Molenberghs G (2000). *Linear mixed models for longitudinal data*. Springer Series in Statistics : New York.

Verbeke G, Molenberghs G, Thijs H, Lesaffre E et Kenward MG (2001). Sensitivity analysis for nonrandom dropout a local influence approach. *Biometrics*, **57,** 7–14.

Verbyla AP, Cullis BR, Kenward M et Welham SJ (1999). The analysis of designed experiments and longitudinal data using smoothing splines. *Applied Statistics*, **48,** 269–312 (with discussion).

Visser PJ, Kester A, Jolles J et Verheis F (2006). Ten-year risk if dementia in subjects with mild cognitive impairment. *Neurology*, **67,** 1201–1207.

Wang HX, Luo B, Zhang QB et Wei S (2004). Estimation for the number of component in a mixture model using stepwise split-and -merge EM algorithm. *Pattern Recognition Letters*, **25,** 1799–1809.

Wechsler D (1981). *Wechsler Adult Intellligence Scale - Revised*. The Psychological Corporation : New York.

Wei LJ, Lin DY et Weissfeld L (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of te American Statistical Association*, **84,** 1065–1073.

Wilson RS, Beckett LA, Bienas JL, Evans DA et Bennett DA (2003). Terminal decline in cognitive function. *Neurology*, **60,** 1782–1787.

Woodroofe M (1975). Estimating a distribution function with truncated data. *Annals of Statistics*, **13,** 163–177.

Wulfsohn MS et Tsiatis AA (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, **53,** 330–9.

Xu J et Zeger SL (2001). The evaluation of multiple surrogate endpoints. *Biometrics*, **57,** 81–7.

Yashin AI, Vaupel JW and Iachine IA (1995). Correlated individual frailty : an advantageous approach to survival analysis of bivariate data. *Mathematical Population studies*, **5,** 145–159.

Yu M, Taylor JMG et Sandler HM (2008). Individual prediction in prostate cancer studies using a joint longitudinal-survival-cure model. *Journal of the American Statistical Association*, **103,** DOI :10.1198/016214507000000400

Yu B et Ghosh P (2009). Joint modeling for cognitive trajectory and risk of dementia in the presence of death. *Biometrics*, DOI :10.1111/j.1541-0420.2009.01261.x

Zhang D, Lin X, Raz J et Sowers M (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93,** 341–348.

Zhang MH et Cheng QS (2004). Determine the number of components in a mixture model by the extended KS test. *Pattern Recognition Letters*, **25,** 211–216.

Liste des tableaux

Table des figures

Liste des publications et communications scientifiques

Publications

Dantan E, Proust-Lima C, Letenneur L et Jacqmin-Gadda H (2008). Pattern Mixture Models and Latent Class Models for the Analysis of Multivariate Longitudinal Data with Informative Dropouts. *The International Journal of Biostatistics*, **4,** Iss 1, Article 14.

Dantan E, Joly P, Dartigues JF et Jacqmin-Gadda H (2009). Joint model with latent state for longitudinal and multi-state data. Article soumis.

Communications orales

Dantan E, Proust-Lima C, Jacqmin-Gadda H (2007). Modèles par mélange de schémas d'observation et modèles à classes latentes pour le traitement des sorties d'étude informative. *Journée de la Statistique*, Angers (France)

Dantan E, Proust-Lima C, Jacqmin-Gadda H (2007). Pattern Mixture Models and Latent Class Models for the Analysis of Multivariate Longitudinal Data with informative Dropouts. *International Society for Clinical Biostatistics*, Alexandroupolis (Grèce)

Dantan E, Proust-Lima C, Jacqmin-Gadda H (2007). Modèles par mélange de schémas d'observation et modèles à classes latentes pour le traitement des sorties d'étude inform-

tive. *Groupe De Recherche Statistiques et santé*, Paris (France)

Dantan E et Jacqmin-Gadda H (2009). Joint model with latent state for the pre-diagnosis phase of dementia. *International Society for Clinical Biostatistics*, Prague (République tchèque)