



HAL
open science

**Modèles mixtes à structure latente pour
donnéeslongitudinales multivariées
hétérogènes :application à l'étude du vieillissement
cognitif etde la démence**

Cecile Proust-Lima

► **To cite this version:**

Cecile Proust-Lima. Modèles mixtes à structure latente pour donnéeslongitudinales multivariées hétérogènes :application à l'étude du vieillissement cognitif etde la démence. Sciences du Vivant [q-bio]. Université Victor Segalen - Bordeaux II, 2006. Français. NNT : . tel-00192386

HAL Id: tel-00192386

<https://theses.hal.science/tel-00192386>

Submitted on 27 Nov 2007

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

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

Université Victor Segalen Bordeaux 2

Année 2006

Thèse n° 1373

THESE

pour le

DOCTORAT DE L'UNIVERSITE BORDEAUX 2

Mention : Sciences Biologiques et Médicales

Option : Epidémiologie et Intervention en Santé Publique

Présentée et soutenue publiquement

Le 19 décembre 2006

Par

Cécile PROUST-LIMA

Née le 27 janvier 1979 à Poitiers

**Modèles mixtes à structure latente pour données
longitudinales multivariées hétérogènes :
application à l'étude du vieillissement cognitif et
de la démence**

Membres du jury

Monsieur Jean-Louis Foulley, Directeur de recherche, Jouy-en-Josas	Président
Monsieur Michel Chavance, Directeur de recherche, Villejuif	Rapporteur
Monsieur Christian Lavergne, Professeur, Montpellier	Rapporteur
Monsieur Jean-François Dartigues, Professeur, Bordeaux	Membre invité
Madame Hélène Jacquemin-Gadda, Chargée de recherche, Bordeaux	Directrice de thèse

A mon papa parti trop vite

Remerciements

A Monsieur Jean-Louis Foulley

Je vous remercie d'avoir accepté de présider ce jury. Vous êtes le premier à m'avoir enseigné les modèles mixtes et le traitement des données corrélées. Peut-être n'aurais-je pas pris cette voie sans vous. Trouvez ici l'expression de ma gratitude.

A Monsieur Michel Chavance

Vous m'avez fait un grand honneur en acceptant de juger ce travail. Votre grande expérience en modélisation des données longitudinales et des données incomplètes m'apporteront beaucoup. Je vous remercie sincèrement d'avoir accepté d'être le rapporteur de ma thèse.

A Monsieur Christian Lavergne

Je suis très honorée que vous ayez accepté de juger ce travail. Votre grande connaissance des modèles de mélange et votre regard sur les méthodes d'estimation m'apporteront assurément un éclairage nouveau sur mon travail. Je vous remercie vivement d'avoir accepté d'être le rapporteur de ma thèse.

A Monsieur Jean-François Dartigues

Vous me faites un grand honneur et un grand bonheur en participant à ce jury. J'ai été ravie et très touchée de collaborer avec vous ces dernières années. Vous m'avez soutenue et avez été des plus motivants voyant toujours l'intérêt épidémiologique des méthodes que nous développons. Votre grand enthousiasme et votre soutien m'ont beaucoup aidée à réaliser ce travail. Je vous en suis profondément reconnaissante.

A Madame Hélène Jacqmin-Gadda

Je ne pourrai jamais assez te remercier pour tout ce que tu m'as apporté. Tu m'as pris sous ton aile il y a maintenant près de 5 ans, m'as accordé toute ta confiance et m'as appris peut-être tout ce que je sais aujourd'hui. Ce fut un réel bonheur de travailler avec toi. Ta rigueur, ta vivacité d'esprit et ta gentillesse m'ont beaucoup apporté et motivé. Pendant les coups durs et les moments de doute, tu as su aussi trouver les bons mots et ça m'a énormément touchée. Merci infiniment pour tout ça et le reste.

Un immense merci à tous ceux que j'ai eu le plaisir de croiser à l'ISPED

A l'équipe biostat dans son ensemble. Merci à vous tous de m'avoir accueillie. Ça a été un plaisir de travailler dans une équipe aussi unie et motivée. Un merci tout particulier à Daniel pour tes idées et ton regard critique, et à Rodolphe pour tes conseils. Merci aussi Guillaume pour les mille et un sauvetages d'ordi ... 4614 à la rescousse.

A l'équipe PAQUID. Ce fut un plaisir de découvrir cette petite famille et notamment ses deux mamans Marie-Hélène et Annick. Je remercie aussi chaleureusement Luc, pour tes jeux de mots drôles ... ou pas ! Merci surtout pour ton soutien et tes idées (l'apoe, mais bien sûr).

A tous les enseignants de l'ISPED qui m'ont appris un bien joli métier. A Pierre d'abord, mon tuteur préféré, merci pour tes conseils, ton expérience et ta grande disponibilité. Merci à Alioum pour la confiance que tu m'as accordée pendant ces trois années, j'en ai été très touchée. Merci aussi à l'équipe de P1, Marianne, Sylvie et Pascale pour m'avoir fait partager votre passion.

A messieurs Roger Salamon et Rachid Salmi pour ce qu'est l'ISPED, cette grande famille où il fait si bon travailler dans la joie et la bonne humeur.

Au bureau 45 et tous ses habitants plus ou moins permanents. A Ben, Seb, Sophie, Jérémie, Sandy, Julien, Elisabeth, Reza, Julia et Marianne, et ceux que j'oublie. Mais surtout, aux discussions sans fin avec Sophie, aux fous rires devant les net-sends et aux powerpoints des reines du craquage, Sophie et Sandy ... quelle bonne époque !

Au ex-bureau 30A, à Delphine, Renaud, Charlotte, Eric, Djoana et tous ceux de passage. Votre bureau était une deuxième maison pour moi. Merci pour votre bonne humeur, les nombreuses rigolades et les goûters improvisés devant un gateau, une brioche, des chocolats apportés par la gourmande Charlotte.

Aux copines de piscine, Anne, Valérie, Gaëlle et Delphine pour ces sorties ludiques mais néanmoins sportives. Merci pour tous ces moments de détente.

A l'ISPED dans son ensemble, à tous ceux et celles que j'ai croisés dans un couloir, la cafétéria, le salon où causent les filles ...

A Barbara pour les longues discussions à l'accueil lorsque j'avais un fax à envoyer, à Nadia pour sa gentillesse et sa douceur, à Marie-Pierre pour sa patience et sa disponibilité si précieuses.

Aux conversations inattendues le midi ou en plein après-midi. Et surtout surtout, au rire de Marthe-Aline, à cet anti-dépresseur, ce baume de joie de vivre.

Merci à tous et à toutes.

Et biensûr des énormes merci à tous ceux que j'aime et qui sont là quoi qu'il arrive

A mes trois très chers amis sots. Merci pour ces soirées délurées, ces voyages improvisés et ces moments de franche rigolade. Merci Eric pour ta joie de vivre et ton petit décalage. Garde ce côté fleur bleue devenu légendaire, il te va si bien. Un gros gros merci au plus sot parmi les sots, Reno. Merci d'avoir été là avec ta douce folie. Ton amitié compte énormément pour moi. Et ton âme de chercheur et ta rigueur scientifique sont aussi des exemples que j'essaie tant bien que mal de suivre. Quand va t-on enfin travailler ensemble? Et bien sûr ma chère et tendre Sophie, la pétillante et pleine de vie Sophie. Quel bonheur de t'avoir rencontrée. De collègues, nous sommes devenues amies. Merci pour toutes ces années et les prochaines.

A Charlotte, sa joie de vivre et sa bonne humeur. J'ai pris beaucoup de plaisir à discuter "stats" avec toi. Tu es vive d'esprit et intuitive, deux grandes qualités. Merci pour ton amitié toutes ces années et merci en particulier d'avoir été là aussi aux moments difficiles.

A ma soeur de promo, Gaëlle. J'ai été ravie de partager ces années avec toi, du DEA aux derniers mois de rédaction en passant par les après-midi où la motivation n'était vraisemblablement pas au rendez-vous. Merci pour tout. On se retrouvera pour le champagne!

A Delphine et ses mille et une histoires savoureuses. Merci pour ces breaks autour de la piscine dès les beaux jours arrivés et pour tous les récits épiques des nuits d'Aglaé ou de l'achat de la maison qui m'ont bien changé les idées.

A Audrey et Karine. Merci à toutes les deux pour votre amitié, pour nos soirées filles où nous étions loin, bien loin des préoccupations professionnelles. Et merci Audrey pour ce petit amour Céleste que tu m'as fait connaître et qui me rappelle qu'il n'y a vraiment pas que le travail dans la vie.

A ma très très chère Karine. Très loin de Bordeaux et du travail, tu sais toujours me changer les idées pendant nos papotages sans fin. Merci d'être là, merci d'être toi.

A ma petit soeur Hélène et ses nombreux encouragements. J'espère que tu trouveras ta voie comme j'ai trouvé la mienne. A mon grand frère, Olivier. Merci pour toutes ces très bonnes soirées autour du "gros tube" ou à refaire le monde à Margonnet.

A mes parents Bernard et Sylvie, amoureux des mathématiques et de l'enseignement qui m'avez donné envie d'aller là où je suis aujourd'hui. Merci à tous les deux pour vos encouragements, votre soutien et votre amour. Merci pour ce que vous représentez et représenterez toujours pour moi, même absents.

Et parce qu'on garde toujours le meilleur pour la fin. Mille millions de mercis à mon petit mari chéri, Bruno, pour sa patience et sa tendresse. Tout au long de ces années, tu m'as supportée, réconfortée et câlinée même pendant mes pires moments d'angoisse et de mauvaise humeur. Tu es un amour, mon amour. Merci d'être là avec moi.

Table des matières

1	Introduction	12
1.1	Epidémiologie du vieillissement cognitif	12
1.1.1	Facteurs de risque potentiels de démence	13
1.1.2	Histoire naturelle du vieillissement cognitif	14
1.1.3	Questions ouvertes sur le vieillissement cognitif et la démence	15
1.2	Problèmes méthodologiques liés à l'étude du vieillissement	17
1.2.1	Multiplicité des mesures de cognition	17
1.2.2	Hétérogénéité des déclin cognitifs	18
1.2.3	Association entre déclin cognitif et démence	19
1.2.4	Sortie d'étude informative	19
1.3	Objectif et plan du mémoire	21
2	Etat des connaissances	23
2.1	Le modèle linéaire mixte	23
2.2	Modèles pour données longitudinales multivariées	25
2.2.1	Modèles linéaires mixtes pour données longitudinales multi- variées	25
2.2.2	Modèles à variables latentes pour données longitudinales mul- tivariables	31
2.3	Modélisation des données hétérogènes	37
2.3.1	Définition d'un modèle de mélange	37

2.3.2	Estimation d'un modèle de mélange	38
2.3.3	Modèles de mélanges pour données longitudinales	43
2.4	Modélisation conjointe de marqueurs longitudinaux et d'un événement	51
2.4.1	Modèles à effets aléatoires partagés	52
2.4.2	Modèles conjoints à classes latentes	56
2.4.3	Cas particulier : modélisation de la sortie d'étude informative	61
3	Modèle nonlinéaire à processus latent pour des données longitudi-	
	nales multivariées	64
3.1	Article méthodologique	65
3.2	Etudes de simulation	77
3.2.1	Méthodologie des études de simulation	77
3.2.2	Première étude : paramétrisation de la fonction Beta	78
3.2.3	Deuxième étude : situation proche du jeu de données PAQUID	81
3.3	Applications épidémiologiques	83
3.3.1	Propriétés métrologiques des tests psychométriques	83
3.3.2	Effet du sexe et du niveau d'études sur les mesures de déclin cognitif	87
3.4	Discussion	90
4	Modèle nonlinéaire à classes latentes pour l'analyse conjointe de	
	plusieurs marqueurs quantitatifs et d'une variable binaire	93
5	Modèle nonlinéaire à classes latentes pour l'analyse conjointe de	
	plusieurs marqueurs quantitatifs et d'un temps d'événement	112
5.1	Formulation du modèle conjoint paramétrique	113
5.1.1	Modèle pour les données longitudinales	113
5.1.2	Modèle de survie conjoint	114
5.2	Estimation du modèle	117
5.2.1	Vraisemblance	117

5.2.2	Troncature à gauche	118
5.2.3	Classification et probabilité <i>a posteriori</i> de l'événement	118
5.3	Modélisation semi-paramétrique de la survie	120
5.3.1	Risque d'événement modélisé par des fonctions splines	121
5.3.2	Vraisemblance pénalisée	122
5.3.3	Choix du critère de lissage	123
5.4	Analyse de l'adéquation	124
5.4.1	Qualité de classification	124
5.4.2	Prédictions dans l'échelle des marqueurs	125
5.4.3	Courbes de survie prédites	125
5.5	Etude de simulation	126
5.5.1	Génération des données	126
5.5.2	Résultats de l'étude de simulation	128
5.6	Application : profils de déclin cognitif associés au risque de démence	128
5.6.1	Objectif de l'étude	128
5.6.2	Les données d'application	130
5.6.3	Spécification du modèle conjoint	130
5.6.4	Résultats	132
5.6.5	Adéquation du modèle	140
6	Discussion et perspectives	149
6.1	Forces de l'approche	149
6.1.1	Modélisation de plusieurs marqueurs non Gaussiens	149
6.1.2	Hétérogénéité des évolutions et association avec la survenue d'un événement	151
6.1.3	Prise en compte des variables explicatives	152
6.1.4	Outil de diagnostic de démence	154
6.2	Modèle à classes latentes et sortie d'étude informative	155
6.3	Limites de l'approche et développements futurs	156

<i>TABLE DES MATIÈRES</i>	11
6.3.1 Adéquation du modèle à classes latentes	156
6.3.2 Transformations nonlinéaires des marqueurs	161
6.3.3 Multiplicité des événements	162
6.4 Conclusion	163
7 Bibliographie	164
8 Annexes	180
8.1 Liste des publications et communications scientifiques	180
8.2 Articles scientifiques	182

Chapitre 1

Introduction

Avec l'allongement de la durée de vie, l'étude du vieillissement cognitif et de la survenue d'une démence a suscité un intérêt croissant chez les épidémiologistes. Mais l'étude du vieillissement pose certains problèmes méthodologiques qui nécessitent des développements statistiques. Le présent travail a pour objet de répondre à des problèmes actuels de recherche sur le vieillissement cognitif par le développement de nouvelles méthodes statistiques pour l'analyse de données longitudinales.

1.1 Epidémiologie du vieillissement cognitif

Le nombre de personnes atteintes de troubles cognitifs liés à l'âge ne cesse d'augmenter avec le vieillissement de la population. Ces troubles 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. La plupart du temps, ces troubles restent légers, liés à un processus de vieillissement cognitif normal mais lorsque le déclin cognitif devient plus important, ils peuvent donner lieu à une démence.

La démence est définie par une altération de la mémoire et d'au moins une autre fonction cognitive avec des répercussions sur la vie quotidienne du malade. La maladie d'Alzheimer représente la démence la plus fréquente chez le sujet âgé. En France, on estime aujourd'hui 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 à près de 225 000 le nombre de nouveaux cas chaque année (OPEPS, 2006). Après 75 ans, la prévalence de la maladie d'Alzheimer et des syndromes apparentés a été estimée à 13.2% chez les hommes et 20.5% chez les femmes, soit une prévalence totale de 17.8% après 75 ans (Ramaroson et al., 2003).

1.1.1 Facteurs de risque potentiels de démence

Les premiers travaux épidémiologiques sur la démence réalisés à partir de données de cohortes prospectives ont consisté, outre à estimer l'incidence et la prévalence de la démence (Letenneur et al., 1994; Lobo et al., 2000; Ramaroson et al., 2003), à évaluer l'impact de facteurs biologiques, environnementaux ou socio-démographiques sur le risque de démence et de maladie d'Alzheimer. Il a ainsi été montré qu'un âge avancé (Letenneur et al., 1994), le sexe féminin (Fratiglioni et al., 2000), un bas niveau d'études (Letenneur et al., 1999) et des facteurs génétiques (génotype de l'apolipoprotéine E - Farrer et al., 1997) représentaient des facteurs de risques de la maladie. Depuis, d'autres facteurs de risque de survenue de démence ont aussi été suspectés comme les facteurs sociaux (peu d'activités de loisir ou sociales - Fabrigoule et al., 1995), les facteurs de risque vasculaires (l'hyper-tension artérielle, l'hypercholestérolémie, le diabète ou le statut tabagique - Cowppli-Bony et al., 2006; de la Torre, 2004) et des facteurs nutritionnels (faible consommation de poisson ou de vitamines - Larrieu et al., 2004; Luchsinger et Mayeux, 2004).

Aujourd'hui, la recherche de facteurs protecteurs d'une démence sur lesquels il est possible d'agir préventivement représente un réel enjeu de santé publique et une part importante des travaux épidémiologiques. Cependant, la démence étant un processus de dégradation progressif et continu, 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. En

effet, le diagnostic de démence ne représente qu'un point arbitraire dans l'histoire de la maladie (Morris et al., 1999) et il est donc primordial d'étudier le déclin des fonctions cognitives au cours du temps en plus du risque de démence, afin de mieux comprendre l'histoire naturelle de la maladie et d'évaluer les mécanismes faisant intervenir les facteurs de risque.

1.1.2 Histoire naturelle du vieillissement cognitif

La démence s'inscrit dans un processus continu de détérioration cognitive et on sait aujourd'hui que le symptôme prédictif le plus sensible d'une démence est le déficit cognitif. Des études pronostiques ont d'abord montré que, 3 à 6 ans avant le diagnostic de démence, le niveau cognitif des sujets déments était plus faible que celui des sujets non-déments (Backman et al., 2001 ; Dartigues et al., 1997). Des études longitudinales ont ensuite permis de mettre en évidence que ce déficit cognitif accru chez les sujets déments venait d'un déclin cognitif pré-diagnostique.

En effet, il a été montré que plusieurs années avant leur diagnostic clinique de démence, les sujets déments souffraient déjà d'un déclin pathologique de leurs performances cognitives (Hall et al., 2000 ; Joseph et al., 1999). Ce déclin d'abord assez lent s'accélérerait entre 5 et 8 ans avant le diagnostic clinique de démence, la date du changement de pente variant selon les études et le type de fonction cognitive étudiée (Hall et al., 2000 ; Hall et al., 2001 ; Hall et al., 2003). Des travaux plus récents ont aussi permis de montrer que la forme du déclin cognitif précédant le diagnostic clinique de démence variait suivant le niveau d'éducation du patient (Jacqmin-Gadda et al., 2006 ; Amieva et al., 2005 ; Scarmeas et al., 2006), les sujets de haut niveau d'études ayant une accélération plus brutale de leur déclin cognitif dans les dernières années précédant le diagnostic que les sujets de bas niveau d'études.

La distinction entre déclin cognitif normal et déclin cognitif pathologique est déterminante pour comprendre le processus de dégradation pathologique vers la démence et la recherche d'outils permettant de détecter plus précocément une démence.

Mais, pour l'instant, les études menées sur le vieillissement cognitif normal ont abouti à des résultats assez variables. En effet, l'existence d'un déclin des fonctions cognitives varie suivant la mesure de cognition étudiée (Jacqmin-Gadda et al., 1997a; Ganguli et al., 1996) et notamment suivant qu'elle possède ou non une composante de vitesse (Jacqmin-Gadda et al., 1997b; Salthouse, 1996).

1.1.3 Questions ouvertes sur le vieillissement cognitif et la démence

Aujourd'hui, de nombreuses questions de recherche restent encore sans réponse concernant à la fois la description du vieillissement cognitif, qu'il soit normal ou pathologique, et les mécanismes d'implication des facteurs de risque d'un déclin cognitif et d'une démence.

Tout d'abord, l'histoire naturelle du vieillissement cognitif reste encore 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; Hall et al., 2001) mais sa forme et la manière dont il se distingue du déclin cognitif normal ne sont pas encore clairement établies. De nouveaux approfondissements sont nécessaires afin de mieux comprendre le processus de vieillissement cognitif normal et notamment son lien avec le déclin cognitif menant à une démence.

Ensuite, avec l'arrivée de traitements permettant d'améliorer les performances cognitives des patients déments (Ringman et Cummings, 2006), diagnostiquer précocément une démence est devenu primordial afin d'initier les traitements le plus tôt possible dans le processus de dégradation pré-déméntiel. Mais pour cela, il est nécessaire de développer des méthodes permettant de distinguer au plus tôt un déclin pathologique associé à une démence d'un déclin cognitif s'inscrivant dans un processus de vieillissement normal.

Dans la perspective de mettre en place des campagnes de prévention auprès du

grand public, il est aussi nécessaire de préciser les mécanismes d'action de facteurs de risque tels que les facteurs vasculaires, nutritionnels ou environnementaux sur le déclin cognitif et sur le diagnostic d'une démence.

Enfin, parmi les facteurs de risque de déclin cognitif et de démence sénile, les mécanismes d'action du sexe et du niveau d'éducation restent encore assez flous. D'après plusieurs études d'incidence, les sujets ayant un bas niveau d'études semblent avoir un risque accru de démence par rapport aux sujets à haut niveau d'études (Letenneur et al., 1999 ; Stern et al., 1994). Il a été montré depuis que la forme du déclin cognitif pouvait varier suivant le niveau d'éducation, faisant émerger l'hypothèse d'une capacité de réserve retardant le retentissement clinique de la maladie chez les sujets à haut niveau d'éducation (Amieva et al., 2005 ; Stern et al., 1994). En effet, les sujets à haut niveau d'éducation garderaient un niveau cognitif stable plus longtemps que les sujets à bas niveau d'études mais auraient un déclin plus rapide de leurs fonctions cognitives juste avant le diagnostic de démence (Jacqmin-Gadda et al., 2006) et après le diagnostic (Scarmeas et al., 2006 ; Mendiondo et al., 2000). Ce phénomène est encore discuté aujourd'hui et, en général, les mécanismes d'action possibles du niveau d'éducation sur le déclin cognitif normal ou pathologique doivent encore être approfondis.

L'effet du sexe sur la démence et le déclin cognitif est aussi controversé. Plusieurs études ont montré que les incidences de démence chez les hommes et les femmes sont différentes, les hommes ayant un risque accru de démence avant 80 ans et cette tendance s'inversant après 80 ans (Letenneur et al., 1999 ; Andersen et al., 1999). Cependant, d'autres études n'ont pas retrouvé ce résultat (Barnes et al., 2003) et n'ont pas non plus trouvé de différence entre le déclin cognitif des hommes et des femmes (Barnes et al., 2003 ; Mendiondo et al., 2000).

1.2 Problèmes méthodologiques liés à l'étude du vieillissement

L'étude de l'histoire naturelle de la démence et de ses facteurs de risque nécessitent l'utilisation de méthodes statistiques pour l'analyse de données longitudinales. De plus, elle suscite des problèmes méthodologiques, la plupart du temps liés au besoin de réduire les biais des analyses et de prendre en compte toute l'information disponible.

1.2.1 Multiplicité des mesures de cognition

La cognition est la quantité d'intérêt dans l'étude du vieillissement cognitif. Cependant, elle n'est pas directement observée : il s'agit d'une quantité latente. De plus, elle est constituée de plusieurs dimensions telles que la mémoire (épisode, de travail, sémantique, court terme, long terme, etc), le langage, l'attention, l'abstraction et les fonctions exécutives (raisonnement logique, inhibition, etc). Dans les études épidémiologiques, la cognition est donc mesurée par plusieurs échelles, les tests psychométriques, qui constituent les marqueurs de la maladie.

Un test psychométrique est une échelle ordinale ou quantitative discrète qui mesure une ou plusieurs dimensions de la cognition. Le test psychométrique le plus connu est le Mini Mental State Examination (MMSE) introduit par Folstein et al. (1975). C'est un score composite constitué de 30 items évaluant principalement la mémoire mais aussi l'orientation dans le temps et l'espace, le langage et les capacités visuo-constructives. Les tests psychométriques n'ont pas les mêmes propriétés métrologiques et mesurent des atteintes cognitives avec une sensibilité différente (Galasko et al., 2000 ; Yesavage et Brooks, 1991 ; Morris et al., 1999). Choisir un de ces tests comme mesure de la cognition est arbitraire et peut faire varier les résultats trouvés (Amieva et al., 2005 ; Jacqmin-Gadda et al., 1997b ; Morris et al., 1999). Par

contre, ils peuvent être considérés comme des mesures bruitées de la même quantité latente : la cognition.

Aujourd'hui, il apparaît nécessaire, d'une part d'étudier conjointement plusieurs tests psychométriques pour mesurer l'impact des facteurs de risque sur le niveau cognitif global plutôt que sur un test en particulier et augmenter la quantité d'information, et d'autre part d'explorer les caractéristiques de chaque test psychométrique et ses limites.

1.2.2 Hétérogénéité des déclin cognitifs

On sait que les sujets en phase pré-diagnostique de démence souffrent déjà d'un déclin pathologique de leurs performances cognitives plusieurs années avant le diagnostic clinique de démence (Hall et al., 2000 ; Jacqmin-Gadda et al., 2006). Ainsi, le déclin cognitif d'un échantillon représentatif de sujets âgés est hétérogène puisqu'il peut inclure à la fois des sujets normaux et des sujets en phase pré-diagnostique de démence.

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

Pour mieux comprendre l'histoire naturelle de la démence et distinguer un processus de vieillissement normal d'un processus de vieillissement pathologique pouvant mener à une démence, il est donc nécessaire de prendre en compte l'hétérogénéité des déclin cognitifs.

1.2.3 Association entre déclin cognitif et démence

L'étude du vieillissement cognitif normal et l'étude du déclin cognitif pendant la phase pré-diagnostique de démence sont souvent réalisées sur des échantillons différents en sélectionnant soit des sujets normaux à une date donnée, soit des sujets déments à une date donnée. Cependant ce type d'analyse peut souffrir de biais de sélection.

En effet, dans une population de sujets non diagnostiqués déments à une date donnée, il peut exister une part non négligeable de sujets en phase pré-diagnostique de démence qui peut induire une sur-estimation du déclin cognitif lié au processus de vieillissement normal et une augmentation de la variance des estimations (Sliwinski et al., 1996 ; Sliwinski et al., 2003). Il s'agit d'un problème de censure à droite de la démence (Jacqmin-Gadda et al., 2006).

D'autre part, lorsque l'on sélectionne un échantillon de sujets déments ou de sujets non-déments à une date donnée, on ne conserve que les sujets vus à cette date afin de connaître leur statut démentiel. On exclut donc les sujets sortis de l'étude prématurément. Si le processus de sortie d'étude est associé au déclin cognitif, cela peut entraîner un biais dans les analyses (Jacqmin-Gadda et al., 1997a).

Une façon de corriger ces problèmes de sélection est d'étudier conjointement l'évolution cognitive et la survenue d'une démence ou d'une sortie d'étude. Cette modélisation conjointe permet aussi de proposer des outils de diagnostic basés sur les mesures répétées du marqueur (Henderson et al., 2002 ; Lin et al., 2002a), de distinguer l'effet de facteurs de risque sur le déclin cognitif et sur le risque de démence et d'envisager plusieurs modélisations de l'association entre le déclin cognitif et la survenue de la démence (Ganiayre et al., 2006 ; Jacqmin-Gadda et al., 2006).

1.2.4 Sortie d'étude informative

Les études de cohortes qui consistent à suivre des patients sur une longue durée souffrent souvent d'une part importante de données manquantes. Ces données man-

quantes sont dues à des sujets qui refusent de continuer dans la cohorte ou qui sont perdus de vue par les investigateurs. Lorsqu'un sujet disparaît de l'étude définitivement, on parle de sortie d'étude. Dans les cohortes sur le vieillissement cognitif, le processus de sortie d'étude est particulièrement important puisqu'en plus du processus usuel, une part non négligeable de sujets décède au cours du suivi.

Les processus de sortie d'étude ont beaucoup été étudiés et une classification des sorties d'étude et des données manquantes en général a été établie (Little et Rubin, 1987) :

- La sortie d'étude est dite complètement aléatoire si la probabilité de non-réponse est indépendante de la variable étudiée, qu'elle soit observée ou non. Dans le vieillissement cognitif, cela impliquerait que le fait qu'un sujet remplisse ou ne remplisse pas un test psychométrique à un temps t ne dépende pas des scores obtenus précédemment ni de celui qu'il aurait obtenu au temps t .
- La sortie d'étude est dite aléatoire si la probabilité de non-réponse ne dépend que des valeurs observées précédemment sur la variable étudiée. Dans le vieillissement cognitif, ce serait le cas si le fait qu'un sujet remplisse ou ne remplisse pas un test psychométrique à un temps t ne dépendait que des scores obtenus précédemment (pas de son score au temps t).
- La sortie d'étude est dite non-ignorable ou informative lorsque la probabilité de réponse dépend des valeurs non-observées de la variable étudiée. Dans le vieillissement cognitif, cela signifierait que le fait qu'un sujet réponde ou ne réponde pas à un test psychométrique à un temps t serait dépendant du score non observé à ce même temps t .

Dans les deux premiers cas, les estimateurs du maximum de vraisemblance des modèles pour données longitudinales ne sont pas biaisés. Par contre, dans le dernier cas, la probabilité de non-réponse doit être modélisée conjointement à l'évolution de la variable d'intérêt afin de corriger les biais sur les estimations (Little, 1995).

Une étude précédente du déclin cognitif sur la cohorte de personnes âgées PA-

QUID suggère que les sorties d'étude ne sont pas ignorables car elles semblent dépendre du score cognitif manquant (Jacqmin-Gadda et al., 1999 ; Jacqmin-Gadda et al., 1997a). Cependant, dans cette analyse, l'impact des sorties d'étude sur les estimations obtenues était assez faible.

1.3 Objectif et plan du mémoire

Face à ces questionnements épidémiologiques et aux problèmes statistiques qu'ils soulèvent, l'objectif de notre travail a été de proposer des modèles statistiques permettant de prendre en compte :

- l'information issue de plusieurs marqueurs de l'évolution de la maladie,
- l'hétérogénéité des déclin cognitifs,
- l'association entre les marqueurs de déclin cognitif et la survenue de l'événement "démence"

dans la perspective de :

- décrire l'évolution des fonctions cognitives chez les personnes âgées afin de mieux comprendre le processus de vieillissement cognitif, qu'il soit normal ou qu'il mène à une démence sénile,
- d'approfondir l'étude des facteurs de risque de déclin cognitif et de survenue de démence.

Dans le chapitre II, nous proposons une revue de la littérature traitant des trois points statistiques développés dans ce travail. D'une part, nous décrivons les modèles pour l'analyse conjointe de données longitudinales multivariées. Nous exposons ensuite la méthodologie des modèles à classes latentes (ou modèles de mélange) permettant de prendre en compte une structure hétérogène des données et enfin, nous présentons les modèles statistiques développés pour étudier conjointement l'évolution d'un marqueur quantitatif et la survenue d'un événement.

Le chapitre III présente un modèle nonlinéaire à processus latent pour étudier conjointement plusieurs marqueurs longitudinaux dans le cas où ils sont des mesures d'une même quantité sous-jacente. Une originalité de ce travail est que le modèle permet de traiter des marqueurs quantitatifs longitudinaux pour lesquels l'hypothèse de normalité n'est pas valide. Dans l'étude du vieillissement cognitif, ce travail permet à la fois de mettre en évidence des propriétés métrologiques des tests psychométriques utilisés pour mesurer les changements cognitifs et de distinguer l'effet des variables explicatives sur le déclin cognitif latent de leur effet spécifique sur les outils de mesure.

Dans le chapitre IV, nous proposons une approche pour modéliser conjointement plusieurs marqueurs longitudinaux corrélés et un événement clinique par un modèle à classes latentes. Ce modèle conjoint combine le modèle non-linéaire à processus latent présenté précédemment pour décrire l'évolution des marqueurs et un modèle logistique pour prédire la probabilité de l'événement en fonction des classes latentes mises en évidence. Ce modèle appliqué à l'étude du vieillissement cognitif et à la survenue d'une démence permet à la fois de décrire les profils d'évolution cognitive associés à la survenue d'une démence et de proposer un outil de diagnostic d'une démence en utilisant l'information sur le déclin cognitif.

Le chapitre V présente une extension du modèle conjoint à classes latentes présenté au chapitre précédent pour prendre en compte non plus la probabilité de démence à un temps donné mais le risque de démence au cours du temps par l'introduction d'un modèle de survie à risques proportionnels. Ce modèle appliqué à l'étude du vieillissement cognitif et à la survenue d'une démence permet de décrire les profils de déclin associés au risque de démence et de proposer des outils de diagnostic.

Dans le chapitre VI, nous discutons les méthodes proposées d'un point de vue statistique en insistant sur leurs forces et leurs faiblesses. Nous replaçons aussi ce travail dans le cadre de la recherche épidémiologique actuelle sur le vieillissement cognitif et la démence sénile.

Chapitre 2

Etat des connaissances

Ce chapitre présente un état des lieux des recherches statistiques réalisées dans la modélisation des données longitudinales multivariées, la prise en compte de données hétérogènes et la modélisation conjointe de marqueurs longitudinaux et de la survenue d'un événement. La plupart des méthodes décrites dans ce chapitre étant des extensions du modèle linéaire mixte classique, nous commençons par présenter ce modèle.

2.1 Le modèle linéaire mixte

Le modèle linéaire mixte introduit par Laird et Ware (1982) est la méthode la plus utilisée pour étudier l'évolution d'un marqueur au cours du temps. Il consiste à prendre en compte la corrélation entre les mesures répétées des sujets dans l'estimation des paramètres de population, c'est à dire de l'évolution moyenne au cours du temps et de l'association avec des variables explicatives. Dans une population de N sujets, considérons le vecteur de réponses $Y_i = (Y_{i1}, \dots, Y_{in_i})$ correspondant aux n_i mesures répétées du marqueur pour le sujet i . Le modèle linéaire mixte s'écrit de la manière suivante :

$$Y_i = X_i\beta + Z_iu_i + \epsilon_i \quad (2.1)$$

où X_i est la matrice de taille $n_i \times p$ contenant les variables explicatives (et notamment le temps) et associée au vecteur d'effets fixes β . Z_{ij} est une sous-matrice de X_i de taille $n_i \times q$ associée aux effets aléatoires u_i propres à chaque sujet et ϵ_i est le vecteur d'erreurs. Les erreurs ϵ_i sont indépendantes des effets aléatoires u_i . Les effets aléatoires u_i sont distribués suivant une loi normale d'espérance $\mathbf{0}$ et de matrice de variance-covariance D : $u_i \sim \mathcal{N}(\mathbf{0}, D)$. Les erreurs ϵ_{ij} contenues dans le vecteur ϵ_i sont normales centrées de variance Σ_i . La plupart du temps, les erreurs de mesures sont indépendantes entre elles et de même variance ($\Sigma_i = \sigma^2 I_{n_i}$) mais on peut aussi les supposer corrélées entre elles afin de tenir compte de la corrélation entre les mesures répétées d'un même sujet. Une structure de corrélation classique est la structure auto-régressive à l'ordre p dans laquelle l'erreur de mesure au temps t_{ij} dépend des p précédentes erreurs de mesures.

Le modèle linéaire mixte peut s'écrire de manière marginale :

$$Y_i \sim \mathcal{N}(X_i \beta, V_i = Z_i D Z_i^T + \Sigma_i) \quad (2.2)$$

C'est cette écriture marginale qui est utilisée pour estimer le vecteur de paramètres du modèle θ contenant le vecteur d'effets fixes β , les éléments de D et les éléments de Σ_i . En effet, à partir de l'expression (2.2) la vraisemblance du modèle s'écrit :

$$\mathcal{L}(\theta) = \prod_{i=1}^N \left(\frac{1}{2\pi} \right)^{n_i/2} |V_i|^{-1/2} \exp \left(-\frac{1}{2} (Y_i - X_i \beta)^T V_i^{-1} (Y_i - X_i \beta) \right) \quad (2.3)$$

Les estimateurs du maximum de vraisemblance des paramètres θ sont obtenus en maximisant le logarithme de la vraisemblance (2.3) par une méthode itérative. Deux algorithmes sont très utilisés : l'algorithme EM (Dempster et al., 1978) et l'algorithme de Newton-Raphson (Fletcher, 2000, chapitre 3). L'algorithme de Newton-Raphson est souvent préféré pour sa rapidité de convergence comparée à celle de l'EM (Lindstrom et Bates, 1988). Aujourd'hui, le modèle linéaire mixte est devenu la méthode classique pour étudier les changements au cours du temps des variables quantitatives et des procédures d'estimation sont disponibles dans tous les logiciels de statistiques, la plus connue étant la procédure MIXED sous SAS.

Plusieurs travaux ont testé la robustesse de ce modèle à des écart aux hypothèses. Ainsi, Verbeke et Lesaffre (1997) ont montré que les estimateurs des effets fixes étaient consistants dans le cas où les effets aléatoires n'étaient pas distribués normalement et Jacqmin-Gadda et al. (2006) ont montré que les estimateurs des effets fixes étaient robustes dans le cas d'erreurs non Gaussiennes ou hétéroscédastiques mais ne l'étaient plus dans le cas d'erreurs corrélées entre elles.

2.2 Modèles pour données longitudinales multivariées

Le modèle linéaire mixte permet d'analyser les mesures répétées d'un seul marqueur longitudinal. Or, il est très fréquent de disposer de plusieurs marqueurs longitudinaux et de vouloir les analyser conjointement. En effet, analyser conjointement plusieurs marqueurs permet d'une part de mieux comprendre le lien qui existe entre eux et d'autre part d'exploiter leur structure de dépendance pour augmenter la puissance des analyses, notamment lorsqu'on étudie l'effet d'une variable explicative sur l'évolution des marqueurs. Nous allons présenter maintenant deux types d'approches pour étudier conjointement plusieurs marqueurs longitudinaux, l'extension des modèles linéaires mixtes aux données multivariées et les modèles à variable latente ou processus latent.

2.2.1 Modèles linéaires mixtes pour données longitudinales multivariées

Dans un souci de clarté, nous considérons uniquement le cas bivarié, c'est à dire avec deux marqueurs longitudinaux, mais la méthodologie est la même pour un nombre plus important de marqueurs.

Présentation du modèle linéaire mixte bivarié

Pour chaque sujet i , on dispose des vecteurs d'observation $Y_i^1 = (Y_{i1}^1, \dots, Y_{in_{i1}}^1)$ et $Y_i^2 = (Y_{i1}^2, \dots, Y_{in_{i2}}^2)$. Le modèle linéaire mixte défini en (2.1) s'écrit pour chaque marqueur $k = 1, 2$:

$$Y_i^k = X_i^k \beta^k + Z_i^k u_i^k + \epsilon_i^k \quad (2.4)$$

où les matrices X_i^k et Z_i^k sont des matrices de variables explicatives associées respectivement au vecteur d'effets fixes β^k et au vecteur d'effets aléatoires u_i^k ; ϵ_i^k est le vecteur d'erreurs. En considérant le vecteur de réponses conjoint $Y_i = (Y_i^{1T}, Y_i^{2T})^T$, le modèle linéaire mixte bivarié est défini de la manière suivante :

$$Y_i = \begin{pmatrix} Y_i^1 \\ Y_i^2 \end{pmatrix} = \begin{pmatrix} X_i^1 & 0 \\ 0 & X_i^2 \end{pmatrix} \begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix} + \begin{pmatrix} Z_i^1 & 0 \\ 0 & Z_i^2 \end{pmatrix} \begin{pmatrix} u_i^1 \\ u_i^2 \end{pmatrix} + \begin{pmatrix} \epsilon_i^1 \\ \epsilon_i^2 \end{pmatrix} \quad (2.5)$$

Par cette écriture du modèle bivarié, on retrouve l'expression d'un modèle linéaire mixte classique $Y_i = X_i \beta + Z_i u_i + \epsilon_i$ dans lequel :

$$X_i = \begin{pmatrix} X_i^1 & 0 \\ 0 & X_i^2 \end{pmatrix}, Z_i = \begin{pmatrix} Z_i^1 & 0 \\ 0 & Z_i^2 \end{pmatrix}, \beta = \begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix}, u_i = \begin{pmatrix} u_i^1 \\ u_i^2 \end{pmatrix} \text{ et } \epsilon_i = \begin{pmatrix} \epsilon_i^1 \\ \epsilon_i^2 \end{pmatrix}$$

Estimation du modèle

La vraisemblance de ce modèle s'écrit de manière analogue à celle d'un modèle linéaire mixte simple donnée en (2.3). L'estimation des paramètres peut donc être réalisée par le même type de procédure de maximisation de la vraisemblance que dans le cas univarié. Dans le cadre multivarié, plusieurs algorithmes ont été proposés : un algorithme EM (Shah et al., 1997), un algorithme de Fisher-scoring (Sy et Taylor, 1997) ou un algorithme de type Newton-Raphson (Thiébaud et al., 2002). Une approche Bayésienne a aussi été proposée (Matsuyama et Ohashi, 1997). En utilisant l'analogie avec le modèle linéaire mixte univarié, ce modèle peut aussi être implémenté dans les logiciels statistiques classiques (Thiébaud et al., 2002).

Prise en compte de la corrélation entre les marqueurs

La particularité du modèle linéaire mixte multivarié est qu'il permet d'étudier spécifiquement la structure de corrélation des marqueurs. En effet, dans le cas bivarié, les effets aléatoires u_i sont normaux d'espérance nulle et de matrice de variance-covariance $D = \begin{pmatrix} D^1 & D^{12} \\ D^{12T} & D^2 \end{pmatrix}$. Suivant cette décomposition et dans le cas où les erreurs de mesure sont indépendantes entre elles (Σ_i diagonale), les matrices de variance-covariance D^1 et D^2 capturent les structures de corrélation propres à chaque marqueur alors que la matrice de variance-covariance D^{12} capture la corrélation qui existe entre les deux marqueurs longitudinaux. En particulier, si les marqueurs sont indépendants, $D^{12} = \mathbf{0}$. C'est donc en évaluant la matrice de variance-covariance D^{12} que l'on peut d'une part évaluer s'il existe une corrélation entre les marqueurs et d'autre part tenir compte de cette corrélation pour améliorer la description des marqueurs et l'évaluation de l'association avec des variables explicatives.

La prise en compte de la corrélation entre les marqueurs peut aussi être assouplie en introduisant dans le modèle (2.5) des structures de corrélation entre les erreurs de mesures (dans la matrice Σ_i), le plus souvent à travers des processus stochastiques corrélés. Sy et Taylor (1997) ont ainsi introduit des processus d'Ornstein-Uhlenbeck intégrés dans le modèle (2.5) et Thiébaud et al. (2002), des processus autorégressifs d'ordre 1. Ces processus permettent à la fois d'assouplir la trajectoire de chaque marqueur par l'introduction d'une corrélation entre les mesures répétées d'un même marqueur et d'assouplir la relation entre les marqueurs par l'introduction d'une corrélation entre les processus stochastiques, c'est à dire entre les mesures répétées des deux marqueurs.

Applications des modèles linéaires mixtes multivariés

Plusieurs études ont consisté à explorer la structure de corrélation entre des marqueurs longitudinaux. Dans le vieillissement cognitif, ce type de modèle a permis de mettre en évidence une très forte corrélation entre des mesures de différentes dimensions de la cognition (Sliwinski et al., 2003; Anstey et al., 2003; Beckett et al., 2004) et de prendre en compte ce lien entre plusieurs mesures de cognition pour comparer leurs formes de déclin (Hall et al., 2001; Harvey et al., 2003). Dans l'étude du Virus de l'Immuno-déficience Humaine (VIH), l'analyse conjointe de la charge virale et des CD4 (Thiébaud et al., 2003) ou bien des CD4 et des microglobulines β_2 (Sy et Taylor, 1997) ont permis de comprendre comment les évolutions de ces marqueurs dépendaient les unes des autres sans faire d'hypothèse *a priori* sur le sens des dépendances.

Prendre en compte la corrélation entre des marqueurs permet non seulement de comprendre le lien qui existe entre eux ou de comparer leurs évolutions mais aussi d'améliorer les estimations des effets fixes du modèle, notamment de préciser l'association avec des variables explicatives et de corriger des biais. Ainsi, dans un essai clinique d'efficacité, Todem et al. (2006) ont modélisé un marqueur de tolérance au traitement conjointement au marqueur d'efficacité, marqueur principal de l'étude. Cela leur a permis d'obtenir une estimation plus précise de l'efficacité du traitement. Dans un autre essai clinique, Matsuyama et Ohashi (1997) ont utilisé la modélisation conjointe de plusieurs marqueurs pour prendre en compte la sortie d'étude informative. La sortie d'étude étant liée au niveau d'un marqueur secondaire, ils ont modélisé conjointement l'évolution de ce marqueur et celle du marqueur principal pour évaluer l'effet d'un traitement. Par cette modélisation conjointe, ils ont respecté l'hypothèse de données manquantes aléatoires (car ne dépendant que de données observées) et les estimateurs du maximum de vraisemblance étaient donc non biaisés.

Extensions du modèle linéaire mixte multivarié

Des extensions du modèle linéaire mixte multivarié ont été proposées pour prendre en compte des données non Gaussiennes. Ainsi Shelton et al. (2004) ont proposé une macro SAS permettant d'estimer des modèles mixtes pour plusieurs marqueurs longitudinaux binaires. Chaque marqueur est modélisé par une régression logistique à intercept aléatoire, les intercepts aléatoires étant corrélées entre elles. Todem et al. (2006) ont proposé un modèle mixte pour étudier conjointement plusieurs marqueurs longitudinaux ordinaux. Les auteurs associent une variable latente continue à chaque marqueur ordinal par un modèle à seuils et modélisent les variables latentes sous-jacentes suivant le modèle linéaire mixte pour données multivariées défini en (2.5). Hall et Clutter (2004) suivis de Marshall et al. (2006) ont aussi proposé des modèles non linéaires mixtes multivariés. Dans ces approches, le principe est le même que dans un modèle linéaire multivarié classique ; les effets aléatoires, intervenant de manière non linéaire, capturent la corrélation entre les marqueurs.

Limites de l'approche

Les modèles mixtes multivariés ont reçu un intérêt croissant ces dernières années. En effet, ils offrent un cadre souple pour étudier conjointement plusieurs marqueurs et permettent à la fois d'étudier spécifiquement la structure de corrélation des marqueurs et de la prendre en compte pour augmenter la puissance des analyses et corriger des possibles biais. Cependant, cette approche a certaines limites.

Tout d'abord, en spécifiant un modèle différent pour chaque marqueur (2.4), le nombre de paramètres à estimer devient vite très grand. Si cela ne pose pas de problème en ce qui concerne les effets fixes, le grand nombre de paramètres de variance-covariance des effets aléatoires peut engendrer des problèmes numériques (O'Brien et Fitzmaurice, 2005). Les applications se limitent donc souvent à peu de marqueurs et/ou font des hypothèses sur la structure de corrélation des effets

aléatoires afin de réduire le nombre de paramètres.

Ensuite, Fieuw et Verbeke (2004) ont montré que les estimateurs de la corrélation entre les effets aléatoires reliant deux marqueurs dans un modèle mixte bivarié pouvaient être biaisés lorsque la structure de variance-covariance des erreurs n'était pas bien spécifiée. Lorsque l'objectif est d'étudier spécifiquement la structure de corrélation entre plusieurs marqueurs, les auteurs recommandent donc d'explorer la structure de corrélation des erreurs en même temps que la structure de corrélation des effets aléatoires et d'interpréter les résultats avec prudence. Néanmoins, lorsque l'objectif est de gagner de la puissance dans l'estimation des effets fixes, une mauvaise spécification des erreurs n'engendre pas de biais.

Le modèle mixte multivarié peut être utilisé pour modéliser conjointement plusieurs mesures d'une même quantité sous-jacente afin d'augmenter la puissance dans l'estimation de l'effet global d'une variable explicative sur tous les marqueurs. Dans le modèle bivarié (2.5), cela consiste à considérer que le paramètre associé à la variable explicative est commun dans β^1 et β^2 . Le problème majeur est que, la plupart du temps, les marqueurs utilisés, même s'ils mesurent la même quantité sous-jacente, n'ont pas la même unité ni la même échelle. Or, considérer un paramètre commun à tous les marqueurs n'est possible que si l'effet a la même amplitude et la même interprétation pour tous les marqueurs, c'est à dire si les marqueurs ont la même échelle et la même unité (O'Brien et Fitzmaurice, 2005). Pour étudier l'effet global d'une variable explicative sur des marqueurs ayant des dimensions différentes, Gray et Brookmeyer (1998) ont proposé un modèle de régression marginale estimé par une approche par GEE (Generalized Equation Estimation). L'idée est de paramétrer l'effet de la variable explicative pour qu'il soit sans dimension (indépendant des unités des marqueurs) et qu'il puisse varier suivant l'échelle de temps de la trajectoire de chaque marqueur. Gray et Brookmeyer (2000) ont ensuite étendu ce modèle pour prendre en compte des données de plusieurs types : ordinales, binaires ou quantitatives. Cette approche qui consiste à étudier la composante commune de plusieurs marqueurs en s'affranchissant des problèmes d'unité ou de différents types de

données se rapproche beaucoup des modèles à variable latente.

2.2.2 Modèles à variables latentes pour données longitudinales multivariées

Deux approches basées sur les variables latentes ont été proposées dans la littérature pour étudier conjointement des marqueurs longitudinaux lorsqu'ils sont des mesures bruitées d'une ou plusieurs quantités latentes : les modèles à équations structurelles et les modèles mixtes à variable latente. Dans cette section, nous présentons ces deux approches et discutons leurs différences, leurs avantages et leurs limites.

Les modèles à équations structurelles

L'utilisation d'une variable latente pour modéliser une quantité non observée mais mesurée par plusieurs marqueurs est née dans le domaine de la psychologie. Ce type d'approche nommé SEM pour Structural Equation Modelling (Muthén, 2002) consiste à relier des marqueurs observés à des quantités latentes dans le *modèle d'observation* puis à expliquer la relation entre les quantités latentes par des autorégressions dans le *modèle structurel*.

Le *modèle d'observation* permet de définir le nombre de variables latentes ainsi que la manière dont elle sont liées aux données observées. Chaque marqueur observé est une combinaison linéaire avec erreur d'une partie des variables latentes (Sánchez et al., 2005). Le *modèle structurel* consiste à régresser les variables latentes les unes sur les autres. Classiquement, le modèle structurel est une régression linéaire des variables latentes sur les autres mais, ces dernières années, beaucoup de travaux ont permis de considérer des relations plus complexes, notamment en introduisant des produits de variables latentes ou des polynômes de variables latentes (Song et Lee, 2004 ; Wall et Amemiya, 2000). Une grande partie des développements récents en SEM s'est focalisée sur l'assouplissement du modèle structurel. En effet, en psychologie, le grand intérêt des SEM réside dans la possibilité d'étudier le lien entre

des variables latentes. Nettement moins de travaux ont donc porté sur l'assouplissement de la relation entre les variables latentes et les données observées. Néanmoins, comme il est très fréquent en psychologie d'avoir des données de types différents et notamment des données binaires ou ordinales, des travaux ont permis d'étendre les modèles classiques à la prise en compte de plusieurs types de données (Dunson, 2003 ; Lee et Song, 2004 ; Rabe-Hesketh et al., 2004). Cela se fait par l'introduction de fonctions de lien dans le modèle d'observation, de la même manière que dans les modèles linéaires généralisés.

Jusqu'à récemment, l'approche par équations structurelles s'était limitée aux données transversales mais quelques travaux ont depuis permis d'étendre ces modèles aux données groupées (Song et Lee, 2004) puis aux données répétées (Dunson, 2003 ; Rabe-Hesketh et al., 2004 ; Skrondal et Rabe-Hesketh, 2004). Dunson (2003) a ainsi proposé un modèle issu de la théorie des modèles à équations structurelles pour modéliser l'évolution d'une quantité non-observée mais mesurée par plusieurs marqueurs de nature différente. Dans le cas de données continues, chaque mesure répétée d'un marqueur est une combinaison linéaire de q variables latentes $\xi_{ij} = (\xi_{ij1}, \dots, \xi_{ijq})^T$ définies à chaque temps de mesure j et pour chaque sujet i :

$$Y_{ijk} = Z_{ijk}^T \gamma_k + \xi_{ij}^T \lambda_{jk} + \zeta_{ik} \quad (2.6)$$

où Z_{ijk}^T est un vecteur de variables explicatives associé au vecteur d'effets fixes propres à chaque marqueur γ_k , λ_{jk} est le vecteur de facteurs ("loading factors") reliant la mesure Y_{ijk} aux vecteur de variables latentes ξ_{ij} et ζ_{ik} est une intercept aléatoire spécifique à chaque sujet et à chaque marqueur pour assouplir la relation. Les facteurs λ_{jk} modélisent le lien entre les données observées et les variables latentes. Dans la majorité des cas, ils ne sont pas estimés pour des raisons d'identifiabilité, mais plutôt définis *a priori*.

La corrélation entre les données répétées des sujets est prise en compte dans le modèle structurel qui relie les q variables latentes $(\xi_{ijl}, l = 1, q)$ au temps j aux q

variables latentes aux temps précédents $j' = 1, \dots, j - 1$:

$$\xi_{ijl} = x_{ij}^T \nu_l + \sum_{j'=1}^{j-1} (\xi_{ij'j}^T - Z_{ij'k}^T - x_{ij'}^T \nu) \phi_{jj'} + \delta_{ijl} \quad (2.7)$$

où x_{ij} est un vecteur de variables explicatives associé au vecteur de paramètres ν_l pour la variables latente l , $\phi_{jj'}$ est un vecteur de paramètres de transition et δ_{ijl} est une erreur indépendante Gaussienne centrée réduite.

Ce modèle, proposé dans une approche Bayésienne, inclut des effets de variables explicatives à la fois dans le modèle structurel permettant ainsi de faire de l'inférence à chaque temps de mesure (ν_l) mais aussi des variables explicatives dans le modèle d'observation, permettant d'étudier l'effet spécifique de variables explicatives sur chaque marqueur (γ_k). En prenant en compte la corrélation entre les mesures répétées des marqueurs par le biais des variables latentes, le modèle proposé par Dunson (2003) fait le lien entre le modèle à équations structurelles et le modèle à effets aléatoires incluant une variable latente tel que Roy et Lin (2000) l'ont proposé.

Le modèle mixte multivarié à variable latente

Roy et Lin (2000) ont proposé un modèle linéaire mixte pour données multivariées permettant de modéliser conjointement plusieurs marqueurs dans le cas où les marqueurs sont des mesures de la même quantité non-observée. A la différence des modèles linéaires mixtes multivariés présentés dans au paragraphe 2.2.1, la quantité non-observée est modélisée à part entière. Il s'agit d'une variable latente U_{ij} (i pour le sujet et j pour le temps de mesure) définie suivant un modèle linéaire mixte classique (2.1) :

$$U_{ij} = X_{ij}^T \alpha + Z_{ij}^T a_i + \epsilon_{ij} \quad (2.8)$$

où X_{ij} et Z_{ij} sont respectivement un vecteur de variables explicatives et un sous-vecteur de variables explicatives incluant le temps associé au vecteur d'effets fixes α et au vecteur d'effets aléatoires Gaussiens a_i . Les erreurs ϵ_{ij} sont indépendantes normales centrées réduites. Le modèle permet ainsi d'estimer l'évolution de la variable

latente et d'évaluer les déterminants de cette évolution. Comme dans l'approche par équations structurelles, un modèle d'observation relie la variable latente aux marqueurs observés. Chaque mesure d'un même marqueur k est une combinaison linéaire (β_{0k}, β_{1k}) de la variable latente à ce temps de mesure plus des erreurs :

$$Y_{ijk} = \beta_{0k} + U_{ij}\beta_{1k} + b_{ik} + e_{ijk} \quad (2.9)$$

où b_{ik} est une intercept aléatoire spécifique à chaque marqueur permettant d'assouplir la relation linéaire avec la variable latente et e_{ijk} sont des erreurs de mesure Gaussiennes indépendantes. Dans ce modèle, chaque marqueur a sa propre combinaison linéaire (β_{0k}, β_{1k}) qui permet de modéliser son niveau et son échelle. La variable latente U_{ij} n'a donc pas de dimension.

Ce modèle estimé par un algorithme EM classique permet d'estimer l'effet global d'une variable explicative à travers les paramètres α . En effet, les paramètres α ont une interprétation globale puisqu'ils représentent les effets des variables explicatives sur la variable latente sous-jacente à tous les marqueurs après prise en compte de leurs unités différentes.

Différences et similitudes entre les deux approches

Par l'introduction de variables latentes représentant la quantité non-observée, les deux approches de Dunson (2003) et de Roy et Lin (2000) permettent de décrire l'évolution de la quantité non observée en s'affranchissant du problème d'unités différentes des marqueurs. Ce problème soulevé par O'Brien et Fitzmaurice (2005) était une des limites principales des modèles linéaires mixtes multivariés. Ainsi, les deux approches permettent d'estimer l'effet global d'une variable explicative sur tous les marqueurs même s'ils n'ont pas la même échelle, le niveau et l'échelle de chaque marqueur étant modélisés dans le modèle d'observation. Le modèle proposé par Dunson (2003) permet même d'utiliser simultanément des données de plusieurs types (ordinales, binaires, continues) alors que le modèle proposé par Roy et Lin

(2000) ne traite que les marqueurs quantitatifs Gaussiens.

Cependant, une limite du modèle de Dunson (2003) est que l'effet global de la variable explicative n'est pas directement estimé dans le modèle. En effet, Dunson (2003) définit une variable latente différente à chaque temps de mesure, et relie les variables latentes à différents temps de mesure entre elles. L'effet d'une variable explicative sur les variables latentes est donc conditionnel aux variables latentes aux temps précédents. Cela rend compliquée l'inférence sur l'effet marginal d'une variable explicative alors que, dans l'approche de Roy et Lin (2000), l'évolution de la variable latente étant modélisée suivant un modèle mixte, l'effet global d'une variable explicative est obtenu directement.

La manière de modéliser la dépendance sérielle des données par un modèle de régression entre variables latentes constitue la particularité des modèles à équations structurelles. Cependant, dans le cadre de données répétées, cela pose certains problèmes. Outre la difficulté d'obtenir l'effet marginal d'une variable explicative, cette approche n'est utilisable qu'avec des données équilibrées, c'est à dire des temps de mesure égaux pour tous les sujets et tous les marqueurs, ce qui restreint son applicabilité. Pour pallier le problème de données équilibrées tout en conservant la flexibilité des modèles à équations structurelles, Rabe-Hesketh et al. (2004) ont proposé une nouvelle classe de modèles : les modèles linéaires généralisés latents et mixtes (GLLAMM pour Generalized Linear Latent and Mixed Modeling). Cette classe de modèles généralise à la fois le modèle mixte à variable latente proposé par Roy et Lin (2000) et le modèle longitudinal à équations structurelles proposé par Dunson (2003).

Cependant, même dans l'approche de Rabe-Hesketh et al. (2004), deux points particuliers ne sont pas traités. Bien que les marqueurs soient mesurés à des temps discrets, la quantité d'intérêt sous-jacente n'a pas de raison d'être définie seulement aux temps d'observation. Cette quantité est un processus défini en temps continu dont on n'observe des mesures qu'en certains temps discrets. Ensuite, bien que Rabe-Hesketh et al. (2004) et Dunson (2003) traitent plusieurs types de données, lorsque

les données sont quantitatives, l'hypothèse de normalité est faite comme dans un modèle linéaire mixte classique. Or il est très courant que des marqueurs quantitatifs ne respectent pas l'hypothèse de normalité, notamment dans le vieillissement cognitif.

Dans ce travail de thèse, nous souhaitons étudier conjointement plusieurs marqueurs quantitatifs non Gaussiens d'un même processus latent. Nous développerons donc pour cela des modèles mixtes à processus latent dans lesquels les marqueurs quantitatifs ne seront pas forcément distribués normalement.

2.3 Modélisation des données hétérogènes

Traiter les mélanges de population est une question courante en statistique et a fait l'objet de beaucoup de travaux ces dernières décennies (Bohning et Seidel, 2003; Richardson et Green, 1997; Redner et Walker, 1984). Cet intérêt pour les modèles de mélange vient en grande partie de leur construction intuitive et de leurs nombreuses applications dans des domaines variés de la statistique. Bien que ces modèles posent encore des problèmes méthodologiques et fassent l'objet de travaux théoriques, ils sont beaucoup utilisés en médecine, en psychologie ou en génétique avec des structures de modèle assez complexes incluant notamment des données longitudinales. Dans cette section, nous présentons les développements et problèmes méthodologiques des mélanges de population. Nous décrivons ensuite l'extension de ces modèles à la prise en compte des données longitudinales, ce qui correspond à notre problématique dans l'étude du vieillissement cognitif.

2.3.1 Définition d'un modèle de mélange

Le problème de mélange de populations ou de données hétérogènes survient lorsqu'une ou plusieurs caractéristiques de la population étudiée ne sont pas observées. La population est alors constituée de sous-populations au sein desquelles les observations ont la même distribution (Bohning, 2000).

Lorsque l'on est en présence d'un mélange de G sous-populations non-observées, il n'est pas correct de supposer que toutes les observations de la population sont identiquement distribuées suivant $f(x|\lambda)$. Chaque sous-population a sa propre distribution $f_g(x|\lambda_g)$, $g = 1, \dots, G$. Cependant l'appartenance d'une observation x_i , $i = 1, \dots, N$ à une certaine sous-population ou classe latente est inconnue; il s'agit d'une variable latente c_{ig} qui vaut 1 si le sujet i appartient à la classe latente g et 0 sinon. En notant $\pi_g = P(c_{ig} = 1)$ la probabilité de chaque classe latente g , la distribution conjointe d'un sujet et de sa classe latente est $f(x, c_g) = \pi_g f_g(x|\lambda_g)$ et

la distribution marginale des observations est alors de la forme :

$$f(x|\lambda) = \sum_{g=1}^G \pi_g f_g(x|\lambda_g) \quad (2.10)$$

Les distributions f_g envisagées sont des lois binomiales, des lois de Poisson ou des lois exponentielles mais le plus fréquemment, les modèles de mélange font intervenir des mélanges de lois normales et les principaux développements méthodologiques ont été proposés dans ce cas.

Si la construction d'un tel modèle dans l'optique de capturer l'hétérogénéité non observée dans une population paraît simple et naturelle, son traitement statistique pose certains problèmes et a donné lieu à de nombreux travaux concernant tout particulièrement l'estimation des paramètres et notamment l'estimation du nombre de classes latentes G . Nous allons maintenant revenir sur ces développements.

2.3.2 Estimation d'un modèle de mélange

Une des grandes difficultés dans l'estimation d'un modèle de mélange est que le nombre de paramètres à estimer, à savoir $\lambda = (\lambda_g)_{g=1,\dots,G}$ et $\pi = (\pi_g)_{g=1,\dots,G-1}$, dépend du nombre G de classes latentes considéré. Il est donc difficile d'utiliser des méthodes classiques d'estimation pour estimer simultanément les paramètres λ , π et le nombre de classes latentes G . La plupart des méthodes envisagées se sont donc focalisées sur une estimation en deux temps, avec une estimation des paramètres à nombre de composantes fixe et une détermination du nombre optimal de composantes par des tests statistiques ou des critères de sélection de type AIC.

Estimation d'un modèle de mélange à nombre de composantes fixé

Lorsque le nombre de composantes dans un mélange de population est connu, les méthodes d'estimation des paramètres sont le plus souvent basées sur la maximisation de la vraisemblance. Plusieurs algorithmes de type Newton (Fletcher, 2000,

chapitre 3) ont été proposés donnant de bons résultats (Haughton, 1997) mais l'algorithme le plus utilisé est de loin l'algorithme EM (Dempster et al., 1977; Redner et Walker, 1984). Il est facile à implémenter et propose une manière intuitive de traiter les données manquantes que sont les variables d'appartenance aux classes.

L'idée de cet algorithme est de considérer la log-vraisemblance des données complètes c'est à dire la log-vraisemblance conjointe des observations x et des variables latentes d'appartenance $c = (c_g)_{g=1,G}$:

$$\mathcal{L}(\theta|x, c) = \sum_{i=1}^N \sum_{g=1}^G c_{ig} (\ln(f(x_i|\lambda_g)) + \ln(\pi_g)) \quad (2.11)$$

Les estimateurs du maximum de vraisemblance $\hat{\theta} = (\hat{\pi}, \hat{\lambda})$ sont ensuite obtenus en répétant les étapes E et M suivantes. L'étape E (expectation) consiste à prédire l'espérance de la log-vraisemblance complète à partir des estimations courantes. Par linéarité de l'expression, cela revient à prédire la probabilité que chaque sujet a d'appartenir à une classe latente soit, à l'itération $k + 1$:

$$\hat{\pi}_{ig}^{(k+1)} = E(c_{ig}|x_i, \hat{\theta}^{(k)}) = \frac{\hat{\pi}_g^{(k)} f(x_i|\hat{\lambda}_g^{(k)})}{\sum_{l=1}^G \hat{\pi}_l^{(k)} f(x_i|\hat{\lambda}_l^{(k)})} \quad (2.12)$$

L'étape M (maximisation) consiste ensuite à obtenir les estimations courantes $\hat{\theta}^{(k+1)}$ qui maximisent l'espérance de la log-vraisemblance à partir des estimations de l'étape précédente. Cette maximisation est faite en deux temps, d'une part la mise à jour des probabilités $\hat{\pi}_g^{(k+1)} = \sum_{i=1}^N \hat{\pi}_{ig}^{(k+1)} / N$ est directement issue de l'étape précédente et les autres estimations courantes $\hat{\lambda}_g^{(k+1)}$ en sont déduites.

L'implémentation de cet algorithme est particulièrement facile à mettre en oeuvre, offre une assurance de convergence (Dempster et al., 1977) et les estimateurs ont de bonnes propriétés asymptotiques (Nityasuddhi et Bohning, 2003). C'est ce qui lui a valu sa grande notoriété dans le domaine du mélange de distributions. Il est en particulier préféré aux algorithmes de type Newton car il a de meilleurs propriétés de convergence globale (Redner et Walker, 1984). Néanmoins, il est aussi connu pour avoir une convergence très lente. Il a donc été proposé de combiner les performances

de l'algorithme EM en début d'optimisation avec la convergence quadratique d'un algorithme de type Newton-Raphson afin d'accélérer la convergence vers le maximum (Redner et Walker, 1984). Cela étant, avec la puissance des ordinateurs actuels, ce problème de rapidité s'est révélé secondaire dans des problèmes simples de mélange. Nous verrons que ce n'est plus le cas pour l'extension de ces modèles aux données longitudinales.

L'estimation des modèles de mélange à nombre de composantes fixé par le maximum de vraisemblance est cependant compliquée par l'existence de maxima locaux. Pour pallier cette multimodalité, des méthodes basées sur l'exploration de l'espace des paramètres avant l'optimisation finale ont été proposées : l'idée est d'effectuer quelques itérations d'un algorithme de type EM à plusieurs endroits de l'espace des paramètres, de comparer l'avancement de chacun en terme de vraisemblance et d'utiliser les valeurs des paramètres obtenues dans le meilleur cas comme valeurs initiales de l'estimation finale réalisée par le même type d'algorithme (Biernacki et al., 2003 ; Karlis et Xekalaki, 2003 ; Berchtold, 2004). Plus généralement, quel que soit l'algorithme utilisé, l'exploration de l'espace des paramètres par des grilles de valeurs initiales permet généralement de trouver le maximum global.

Méthodes d'estimation pour un nombre de composantes inconnu

Pour éviter ces problèmes de multimodalité et d'estimation en deux temps des paramètres d'un modèle de mélange, quelques méthodes ont été proposées pour estimer simultanément les paramètres λ , π et le nombre de classes latentes G . Richardson et Green (1997) ont par exemple proposé une approche bayésienne. L'idée de l'algorithme est de pouvoir passer d'un espace de paramètres à l'autre par des sauts, permettant ainsi au cours de la procédure d'estimation de varier le nombre possible de classes latentes. Ceci se fait par des étapes de division ou de fusion de classes. Récemment, cette idée de division et fusion de classes a été reprise dans une méthode d'estimation basée sur le maximum de vraisemblance (Wang et al., 2004). Il s'agit d'itérer deux étapes jusqu'à convergence : (1) l'estimation du modèle

de mélange avec un nombre fixe de composantes par un algorithme EM et (2) la division des classes trop hétérogènes et la fusion des classes trop proches par des critères.

Parallèlement à ces approches, Bohning (1995) a proposé un algorithme permettant d'obtenir en deux temps l'estimateur non-paramétrique du maximum de vraisemblance s'il existe. Dans un premier temps, à partir d'une grille fictive de classes $(\lambda_g)_{g=1,G}$ où G est pris très grand, il estime les probabilités $(\pi_g)_{g=1,G}$ associées et ne sélectionne que celles qui sont positives, obtenant ainsi un estimateur du nombre de classes latentes \hat{G} . La deuxième étape consiste à estimer les paramètres du modèle pour le nombre estimé de composantes \hat{G} par un algorithme EM classique. Schlattmann (2003) a depuis amélioré l'estimation du nombre de composantes dans cette méthode par une technique de Bootstrap.

Ces méthodes demandent beaucoup de calcul numérique et le plus souvent, le nombre de composantes est considéré comme un paramètre qu'il faut traiter à part. Ce sont les raisons pour lesquelles, dans la plupart des applications des modèles de mélange, l'estimation des paramètres à nombre de composantes fixé et la sélection du nombre de composantes *a posteriori* ont été préférées.

Sélection du nombre de composantes

Le test du nombre de composantes dans un modèle de mélange a fait l'objet d'un grand nombre de travaux. En effet, la théorie du rapport de vraisemblance pour tester un modèle à G_0 composantes *versus* un modèle à G_1 composantes ne satisfait pas les conditions de régularité usuelles et les résultats asymptotiques classiques ne s'appliquent pas : la distribution asymptotique du rapport de vraisemblance n'est pas un χ^2 . Le problème vient du fait que l'hypothèse nulle est en bordure de l'espace des paramètres et que sous l'hypothèse nulle, le modèle n'est pas identifiable. En effet, considérons un mélange simple de deux distributions à un paramètre. La densité de

probabilité du mélange s'écrit :

$$\pi f(x, \theta_1) + (1 - \pi)f(x, \theta_2) \quad (2.13)$$

L'hypothèse nulle selon laquelle la population est homogène avec une densité $f(x, \theta)$ peut s'écrire sous la forme $\mathcal{H}_0 : \pi = 0$ et θ_2 n'est alors pas identifiable, ou de façon équivalente : $\mathcal{H}_0 : \theta_1 = \theta_2$ et π n'est donc pas identifiable.

Dans le cas de la comparaison d'un modèle de mélange à deux composantes *vs.* une composante, la distribution asymptotique de la statistique du rapport de vraisemblance (SRV) a cependant été calculée pour des lois normales avec un seul paramètre de moyenne (Ghosh et Sen, 1985) et deux paramètres de moyenne (Chen et Chen, 2001). Mais ces distributions asymptotiques dépendent entièrement de la spécification du modèle et sont difficilement calculables autrement que par Bootstrap (McLachlan, 1987 ; Chen et Chen, 2001). Chen et al. (2001) ont proposé un test basé sur une vraisemblance pénalisée offrant de bonnes propriétés asymptotiques, le problème majeur étant que la pénalité doit être spécifiée différemment suivant le modèle de mélange étudié et notamment suivant le nombre de composantes et le nombre de paramètres inconnus (Chen et Kalbfleisch, 2005). D'autres tests statistiques ou critères ont été proposés, basés sur la distance L^2 (Charnigo et Sun, 2004 ; Miloslavsky et van der Laan, 2003), sur une mesure de distorsion (Sugar et James, 2003), sur une extension du test de Kolmogorov-Smirnov (Zhang et Cheng, 2004) ou sur le facteur bayésien (Moreno et Liseo, 2003) mais aucune méthode ne s'est avérée meilleure que les autres ni plus facile à utiliser. Lo et al. (2001) ont proposé un test de type SRV avec de bonnes propriétés asymptotiques mais les conditions requises pour l'application de leur test ne sont pas satisfaites la plupart du temps (Jeffries, 2003). Le développement d'un test de type SRV qui pourrait s'appliquer dans tous les cas de mélange quels que soient la distribution envisagée, le nombre de composantes et la complexité du modèle nécessite encore des recherches, la plupart des travaux ci-dessus étant limités pour l'instant à des mélanges de lois avec un seul paramètre inconnu.

Hawkins et al. (2001) ont comparé une vingtaine de critères de sélection proposés dans la littérature et ont montré que le Bayesian Information Criterion (BIC) (Schwartz, 1978) donnait de très bons résultats. Les tests et critères proposés depuis et comparés au BIC ont confirmé ces conclusions (Zhang et Cheng, 2004; Miloslavsky et van der Laan, 2003; Bauer et Curran, 2003a). Ses bonnes propriétés et son écriture simple quel que soit le modèle considéré (et notamment le modèle incluant des données longitudinales) en ont fait depuis plusieurs années le critère classique pour déterminer le nombre de composantes dans les modèles de mélange.

2.3.3 Modèles de mélanges pour données longitudinales

Verbeke et Lesaffre (1996) ont proposé d'étendre les modèles de mélange aux données longitudinales. Ils ont ainsi étendu l'idée de populations hétérogènes à des profils d'évolution hétérogènes.

Définition

Le modèle de mélange pour données longitudinales peut être vu comme une extension du modèle linéaire mixte proposé par Laird et Ware (1982) et présenté dans la première partie de ce chapitre. Cependant, pour simplifier la présentation du modèle de mélange pour données longitudinales, nous partons d'une écriture du modèle de Laird et Ware (1982) légèrement différente de celle présentée précédemment :

$$Y_i = X_i\beta + Z_iu_i + \epsilon_i \quad (2.14)$$

où la matrice de variables explicative Z_i n'est plus une sous-matrice de X_i mais une matrice distincte. Les effets fixes (vecteur μ) des variables explicatives incluses dans Z_i sont spécifiés directement dans la distribution des effets aléatoires $u_i : u_i \sim \mathcal{N}(\mu, D)$. Les deux définitions (2.14) et (2.1) sont équivalentes.

Dans le cas de données hétérogènes, on fait l'hypothèse que les données sont issues de G populations ayant des profils d'évolution différents. L'hétérogénéité est

alors prise en compte dans la distribution des effets aléatoires définie comme un mélange de G lois normales : $u_i \sim \sum_{g=1}^G \pi_g \mathcal{N}(\mu_g, D_g)$. Chaque sous-population a son propre profil d'évolution moyen défini par $Z\mu_g$ et a une probabilité π_g ; la somme de ces probabilités sur les G classes valant 1 ($\sum_{g=1}^G \pi_g = 1$). Dans les modèles de mélange pour données longitudinales, la matrice de variance-covariance des effets aléatoires est la plupart du temps supposée invariante selon les classes de mélange ($D_g = D, \forall g$) car cela permet d'éviter des problèmes numériques dans la procédure d'estimation (Bauer et Curran, 2003a ; Legler et al., 2004 ; Verbeke et Molenberghs, 2000). En effet, dans le cas où la matrice de variance-covariance est spécifique à chaque classe, la log-vraisemblance peut se retrouver infiniment grande au bord de l'espace des paramètres ce qui induit une non-identifiabilité du modèle. Ce problème n'est pas spécifique aux modèles de mélange pour données longitudinales (Redner et Walker, 1984). Il a été étudié en univarié dans le cas d'un mélange de deux lois normales et il semble que l'algorithme EM, grâce à son caractère local ne soit pas trop attiré vers le point de singularité de l'espace des paramètres où la log-vraisemblance est infinie comparé à d'autres algorithmes aux propriétés de convergence globale (Nityasuddhi et Bohning, 2003).

A l'origine, ce modèle pour données longitudinales hétérogènes a été développé pour prendre en compte et tester un écart à l'hypothèse de normalité des effets aléatoires (Verbeke et Lesaffre, 1996 ; Verbeke et Lesaffre, 1997), mais il est depuis largement utilisé et étendu pour traiter des données hétérogènes issues de plusieurs sous-populations non observées (Bauer et Curran, 2003a) : dans des essais cliniques pour étudier la non-compliance (Pauker et Laird, 2000) ou la réponse au traitement (Xu et Hedeker, 2001), en génétique pour faire de la classification de gènes (Celeux et al., 2005) ou en psychologie pour étudier les profils de comportement (Muthén et Shedden, 1999 ; Muthén et al., 2002).

Ces dernières années, la plupart des développements sur les modèles de mélange pour données longitudinales ont été réalisés par l'équipe de Muthén (Muthén et Shedden, 1999 ; Muthén, 2001 ; Muthén et al., 2002 ; Muthén, 2004) dans l'op-

tique de décrire des sous-populations non-observées. Muthén et Shedden (1999) ont ainsi proposé d'étendre la première version du modèle de Verbeke et Lesaffre (1996) afin de prédire la probabilité d'appartenir à une classe latente en fonction des caractéristiques du sujet. En définissant la variable latente d'appartenance aux classes c_i qui vaut g si le sujet i appartient à la classe g , ils ont exprimé la probabilité d'appartenir à la classe g par un modèle multinomial :

$$\pi_{ig} = P(c_i = g | X_{2i}) = \frac{e^{\xi_{0g} + X_{2i}^T \xi_{1g}}}{\sum_{l=1}^G e^{\xi_{0l} + X_{2i}^T \xi_{1l}}} \quad (2.15)$$

où ξ_{0g} représente l'intercept de la classe g et ξ_{1g} le vecteur de paramètres spécifiques à la classe g associé aux variables explicatives X_{2i} . Pour des raisons d'identifiabilité, $\xi_{0g} = 0$ et $\xi_{1g} = \mathbf{0}$.

Muthén et Shedden (1999) ont aussi proposé de prendre en compte, conjointement aux profils hétérogènes d'évolution, un événement clinique sous forme de variable binaire prédite par l'appartenance aux classes latentes. Par cette modélisation conjointe, Muthén et Shedden (1999) suivis d'autres travaux (Lin et al., 2000 ; Lin et al., 2002a ; Muthén et al., 2002) ont introduit l'approche par classes latentes dans l'étude conjointe d'un marqueur longitudinal et d'un événement clinique. Nous reviendrons sur ce développement particulier des modèles à classes latentes conjoints dans la dernière section de ce chapitre.

Estimation des paramètres

Comme dans les modèles de mélange simples, l'estimation des paramètres dans les modèles de mélange pour données longitudinales pose certains problèmes. A notre connaissance, tous les travaux réalisés dans ce domaine se sont basés sur une estimation à nombre fixé de composantes étant donné la complexité engendrée par le modèle mixte seul. Le plus souvent, une estimation par maximum de vraisemblance via l'algorithme EM présenté en sous-section 2.3.2 a été utilisée (Verbeke et Lesaffre, 1996 ; Muthén et Shedden, 1999 ; Xu et Hedeker, 2001 ; Muthén et al., 2002). Dans son logiciel M+, Muthén (Muthén et Muthén, 2001) a développé un algorithme

EM pour traiter les modèles de mélange pour données longitudinales dans le cas où les temps de mesure sont identiques pour tous les sujets. Komarek (2001) ont développé une macro SAS gratuite utilisant l'algorithme EM et la procédure MIXED puis Spiessens et al. (2002) l'ont étendue aux modèles mixtes généralisés en utilisant la procédure NLMIXED.

Dans un modèle de mélange pour données longitudinales, l'étape M de l'algorithme EM nécessite l'estimation des modèles linéaires mixtes classiques obtenus conditionnellement aux classes latentes. Cette estimation doit donc se faire par un algorithme itératif de type Newton-Raphson. L'algorithme EM étant connu pour sa lenteur, l'estimation d'un modèle linéaire mixte dans la phase M rend la procédure d'estimation globale encore plus lourde. Dans un travail préparatoire au présent travail, nous avons donc proposé une alternative à l'algorithme EM basée sur une maximisation directe de la log-vraisemblance observée par un algorithme de Marquardt légèrement amélioré (Proust et Jacqmin-Gadda, 2005 en annexe). En effet, la densité du vecteur d'observation y_i s'écrit :

$$f_i(y_i) = \sum_{g=1}^G \pi_g f(y_i | c_i = g) = \sum_{g=1}^G \pi_g \phi_{ig}(y_i) \quad (2.16)$$

où ϕ_{ig} est la densité multivariée Gaussienne d'espérance $E_{ig} = X_i\beta + Z_i\mu_g$ et de variance $V_i = Z_i D Z_i^T + \sigma^2 I_{n_i}$. De là, la log-vraisemblance des données observées dans un modèle de mélange pour données longitudinales se décompose suivant les classes latentes de la manière suivante :

$$\mathcal{L}(\mathcal{Y}; \theta) = \sum_{i=1}^N \ln(f_i(y_i)) = \sum_{i=1}^N \ln\left(\sum_{g=1}^G \pi_g \phi_{ig}(y_i)\right) \quad (2.17)$$

Cette log-vraisemblance peut être maximisée directement comme décrit dans Proust et Jacqmin-Gadda (2005, en annexe) par un algorithme de Marquardt. L'algorithme de Marquardt est plus robuste que l'algorithme de Newton-Raphson car il consiste à gonfler la diagonale de la matrice Hessienne lorsqu'elle n'est pas définie positive. De plus, une recherche unidimensionnelle sur le pas utilisé dans la mise à jour permet d'assurer l'amélioration systématique de la vraisemblance à chaque itération.

L'estimation d'un modèle de mélange par un algorithme de Newton-Raphson a été critiquée (Redner et Walker, 1984 ; Verbeke et Molenberghs, 2000). Malgré sa vitesse de convergence quadratique (donc plus rapide que l'algorithme EM), cet algorithme a tendance à avoir plus de problèmes de convergence que l'algorithme EM. Dans notre travail préparatoire, nous avons donc comparé les performances de notre algorithme de Marquardt amélioré avec celles de l'algorithme EM sur deux échantillons. Les deux algorithmes ont convergé vers le maximum global un nombre équivalent de fois mais l'algorithme de Marquardt convergeait beaucoup plus rapidement (10 minutes *vs.* 2 heures pour un modèle avec une fonction quadratique du temps et 2 classes de mélange estimé sur un échantillon de 1392 sujets).

Nous avons vu précédemment que les modèles de mélange patissaient d'une multimodalité et devaient donc être estimés avec plusieurs jeux de valeurs initiales pour s'assurer de la convergence vers le maximum global. D'autre part, un modèle de mélange doit être estimé pour toute une série de nombres de composantes différents afin de sélectionner le nombre optimal de classes. Par conséquent, étant donné la nécessité de multiplier les estimations du modèle de mélange dans une analyse, il est primordial d'utiliser un algorithme rapide.

Classification *a posteriori*

Une des particularités des modèles de mélange est qu'ils permettent de réaliser une classification *a posteriori* des sujets suivant les classes latentes estimées dans le modèle. Après l'estimation des paramètres du modèle et la détermination du nombre optimal de classes, il est possible de calculer *a posteriori* la probabilité que chaque sujet appartienne à chaque classe latente $\hat{\pi}_{ig}$, $g = 1, G$, par le théorème de Bayes :

$$\hat{\pi}_{ig} = P(c_i = g | Y_i, \hat{\theta}) = \frac{\hat{\pi}_g \phi_{ig}(y_i; \hat{\theta})}{\sum_{l=1}^G \hat{\pi}_l \phi_{il}(y_i; \hat{\theta})} \quad (2.18)$$

où $\phi_{il}(y_i; \hat{\theta})$ est la densité multivariée Gaussienne définie au paragraphe précédent prise aux valeurs estimées des paramètres $\hat{\theta}$. Pour chaque sujet, on dispose alors de

ses probabilités *a posteriori* d'appartenir à chacune des classes latentes. En affectant à chaque sujet, la classe latente à laquelle il a la plus grande probabilité d'appartenir, on obtient une classification *a posteriori* des sujets suivant les classes latentes estimées dans le modèle. Cette classification peut ensuite être utilisée pour évaluer l'association entre les classes latentes des profils d'évolution et les caractéristiques des sujets (Verbeke et Lesaffre, 1996 ; Xu et Hedeker, 2001).

Adéquation du modèle

Avec l'intérêt croissant qu'ont reçu les modèles de mélange pour données longitudinales et notamment d'un point de vue appliqué (Bauer et Curran, 2003a), le test d'adéquation est devenu un point crucial du développement de ce type de modèle. En effet, s'agissant de modèles paramétriques complexes, il est nécessaire d'évaluer leurs hypothèses sous-jacentes.

Comme pour les modèles de mélange classiques, le nombre de composantes dans le modèle de mélange pour données longitudinales constitue l'hypothèse la plus importante. Dans la plupart des travaux sur les modèles de mélange pour données longitudinales (Verbeke et Lesaffre, 1996 ; Muthén et Shedden, 1999 ; Muthén et al., 2002 ; Bauer et Curran, 2003a), l'évaluation du nombre de composantes est faite par le Bayesian Information Criterion (BIC) (Schwartz, 1978) qui consiste à pénaliser la déviance du modèle par le nombre de paramètres multiplié par le logarithme du nombre de sujets dans l'échantillon. Ce critère est préféré au critère d'Akaike (AIC) qui pénalise la déviance par deux fois le nombre de paramètres et a tendance à surestimer le nombre de composantes (Hawkins et al., 2001 ; Wang et al., 2005).

Cependant, si le BIC permet de déterminer le nombre optimal de classes latentes sous l'hypothèse de mélange de population, il ne permet pas d'évaluer l'adéquation du modèle obtenu (Muthén, 2004). D'autres travaux ont donc consisté à explorer l'adéquation d'un modèle de mélange en confrontant les prédictions du modèle final aux données observées. Verbeke et Molenberghs (2000) ont ainsi proposé d'évaluer si la distribution marginale des prédictions était bien le mélange de distributions

normales supposé dans le modèle. Cette méthode est basée sur le principe que si le modèle est correct, la transformation des données observées par la fonction de répartition estimée du modèle doit être uniforme. Muthén (2004) a proposé, dans le cas de données équilibrées et sans données manquantes, de comparer à chaque temps de mesure les indicateurs d'aplatissement (Kurtosis) et d'asymétrie (Skewness) des prédictions obtenus par simulations avec ceux des données observées. Wang et al. (2004) ont de leur côté utilisé des méthodes basées sur le calcul des résidus spécifiques à chaque classe. En utilisant une technique d'affectation aléatoire des sujets aux classes latentes (Bandein-Roche et al., 1997) à partir des probabilités *a posteriori* d'appartenance aux classes, ils calculent les résidus prédits dans chaque classe latente. Cette approche leur permet à la fois d'évaluer le nombre de composantes, la mauvaise spécification de la trajectoire moyenne ou la structure de covariance du modèle. Cependant, tout comme la méthode de Muthén (2004), cette technique est développée uniquement dans le cas de données équilibrées sans données manquantes. De manière générale en modélisation, une approche simple pour évaluer l'adéquation d'un modèle aux données consiste à comparer la courbe d'évolution prédite à la courbe d'évolution observée. La particularité dans les modèles de mélange est qu'il existe une partie latente liée à l'appartenance de chaque sujet aux classes latentes. Muthén (2001) a donc proposé de pondérer les vecteurs d'évolution prédite et observée par la probabilité *a posteriori* que chaque sujet a d'appartenir à chaque classe latente. Cette astuce permet de comparer pour chaque classe latente, les évolutions prédite et observée.

Évaluer la précision de la classification *a posteriori* obtenue par le modèle de mélange peut aussi donner des indications sur l'adéquation du modèle. Selon Muthén (2001), une classification *a posteriori* est bonne (c'est à dire bien discriminante) lorsque, pour les sujets classés dans une certaine classe *a posteriori*, la moyenne de leurs probabilités d'appartenir à cette classe est proche de 1 alors que les moyennes de leurs probabilités d'appartenir aux autres classes sont proches de 0.

Enfin, Bauer et Curran (2003a) ont initié une discussion sur la difficulté de distin-

guer par un modèle de mélange une structure hétérogène des données d'une structure homogène mais non normale des données. Toutes les méthodes proposées ci-dessus permettent d'évaluer si le modèle s'adapte bien aux données mais ne permettent pas de distinguer un mélange de lois normales qui exprime une structure hétérogène d'un mélange de lois normales qui approxime une loi plus complexe. La discussion qui a suivi ce travail (Rindskopf, 2003; Muthén, 2003; Cudeck et Henly, 2003; Bauer et Curran, 2003b) montre la difficulté de l'interprétation des modèles statistiques et notamment des composantes dans un modèle de mélange. Lorsque l'on estime un modèle de mélange, il faut toujours garder en tête la possibilité que les composantes obtenues ne traduisent qu'une approximation d'une distribution homogène.

Dans ce travail de thèse, nous n'étudierons pas spécifiquement les modèles de mélange pour données longitudinales mais plutôt leur extension à l'étude conjointe d'un marqueur et d'un événement présentée dans la section suivante. Cependant, les méthodes, limites et recommandations données ici seront toujours valables.

2.4 Modélisation conjointe de marqueurs longitudinaux et d'un événement

Il est très fréquent dans les études de cohortes ou les essais thérapeutiques de collecter à la fois des mesures répétées de marqueurs et un délai jusqu'à un événement. Les marqueurs mesurent la plupart du temps la progression de la maladie et le délai mesure le laps de temps jusqu'au diagnostic de la maladie, jusqu'au décès ou bien jusqu'à une sortie d'étude.

Les mesures répétées des marqueurs et le délai jusqu'à l'événement ont souvent une association et il peut être intéressant, voire nécessaire, de les étudier conjointement dans le but de :

- prendre en compte l'événement pour faire de l'inférence sur les mesures longitudinales du marqueur. C'est le cas dans les études longitudinales lorsque l'on souhaite par exemple prendre en compte la sortie d'étude informative pour obtenir des estimateurs sans biais de l'évolution du marqueur.
- utiliser les mesures répétées du marqueur pour pronostiquer l'événement. C'est le cas dans les études de survie lorsque l'on souhaite par exemple évaluer si un marqueur longitudinal est un bon marqueur de substitution d'un événement clinique.
- comprendre le mécanisme de l'association entre l'évolution du marqueur et la survenue de l'événement.

La modélisation simultanée de la distribution conjointe $f(Y, T)$ des deux processus que sont les mesures répétées d'un marqueur Y et le délai jusqu'à un événement T permet de répondre à ces objectifs. De manière usuelle, le modèle à effets aléatoires partagés est utilisé. Les effets aléatoires caractérisant l'évolution du marqueur sont inclus en tant que variables explicatives dans le modèle du risque d'événement et représentent l'unique lien entre le marqueur et le délai d'événement. Récemment, une nouvelle approche a été proposée basée sur des classes latentes associées à la

fois à l'évolution du marqueur et au risque d'événement.

Dans cette section, nous présentons ces deux types de modélisation et discutons leurs avantages et leurs limites. Enfin, nous présentons le cas particulier de la prise en compte de la sortie d'étude informative dans les études longitudinales qui a fait l'objet de nombreux développements en modélisation conjointe.

2.4.1 Modèles à effets aléatoires partagés

Le modèle à effets aléatoires partagés associe l'évolution du marqueur et la survenue de l'événement à travers des effets aléatoires communs (Wulfsohn et Tsiatis, 1997; Henderson et al., 2000). L'hypothèse est que l'évolution du marqueur et le délai d'événement sont indépendants conditionnellement aux effets aléatoires :

$$f(Y_i, T_i) = \int f(Y_i | u_i) f(T_i | u_i) f(u_i) du_i \quad (2.19)$$

Les effets aléatoires u_i représentent donc l'unique lien qui existe entre les deux processus.

Spécification du modèle

Plus précisément, l'évolution du marqueur est modélisée par le modèle linéaire mixte présenté en section 2.1. Pour le sujet i , $i = 1, \dots, N$ et la mesure répétée j , $j = 1, \dots, n_i$, la mesure Y_{ij} du marqueur est définie de la manière suivante :

$$Y_{ij} = X_{ij}\beta + Z_{ij}u_i + \epsilon_{ij} \quad (2.20)$$

où X_{ij} et Z_{ij} sont les vecteurs de variables explicatives associés aux effets fixes β et aux effets aléatoires Gaussiens u_i .

Le risque d'événement $\lambda(t)$ est le plus souvent défini par un modèle à risques proportionnels dans lequel des fonctions des effets aléatoires $g(u_i, t)$ sont introduites

en tant que variables explicatives :

$$\lambda(t|x_i; u_i) = \lambda_0(t) \exp(x_i\gamma + g(u_i, t)\delta) \quad (2.21)$$

où $\lambda_0(t)$ est le risque instantané de base (paramétrique ou non), x_i est le vecteur de variables explicatives associé au vecteur d'effets fixes γ , et δ est le vecteur de paramètres reliant le marqueur au risque instantané d'événement. En particulier, si $\delta = \mathbf{0}$ alors le marqueur et l'événement sont indépendants l'un de l'autre.

Le plus souvent, les fonctions des effets aléatoires sont des combinaisons linéaires des effets aléatoires traduisant l'association entre l'événement et une caractéristique de l'évolution du marqueur telle que la valeur courante ou la pente courante du marqueur (Yu et al., 2004). Un exemple consiste à introduire dans le modèle de survie la valeur réelle du marqueur : $g(u_i, t) = E(Y_i | u_i) = X_{ij}\beta + Z_{ij}u_i$ (Wulfsohn et Tsiatis, 1997 ; Law et al., 2002). Henderson et al. (2000) ont aussi généralisé cette approche en introduisant des processus Gaussiens corrélés dans le modèle longitudinal et dans le modèle de survie.

Estimation des paramètres

Dans le modèle à effets aléatoires partagés, l'hypothèse d'indépendance du marqueur et de l'événement conditionnellement aux effets aléatoires permet d'écrire la distribution conjointe du marqueur et de l'événement :

$$\begin{aligned} f(Y, T) &= \int_u f(Y, T|u) f(u) du \\ &= \int_u f(Y|u) f(T|u) f(u) du \end{aligned} \quad (2.22)$$

En notant E_i l'indicateur binaire d'événement, la vraisemblance du modèle s'écrit en utilisant cette décomposition (Hogan et Laird, 1997) :

$$L(\theta; (Y, T)) = \prod_{i=1}^N \int_{u_i} f(Y_i|u_i; \theta) \lambda(T_i|u_i; \theta)^{E_i} S(T_i | u_i; \theta) f(u_i; \theta) du_i \quad (2.23)$$

où $f(Y_i|u_i)$ est une densité multivariée normale d'espérance $E(Y_i | u_i) = X_{ij}\beta + Z_{ij}u_i$ et de variance $\sigma^2 I_{n_i}$, $f(u_i)$ est une densité normale d'espérance nulle et de variance

D et le risque instantané $\lambda(t)$ est estimé de manière paramétrique ou non (Wulfsohn et Tsiatis, 1997 ; Henderson et al., 2000).

La vraisemblance du modèle à effets aléatoires partagés inclut une intégrale multiple sur les effets aléatoires sans solution analytique. La maximisation de la vraisemblance nécessite donc l'approximation d'intégrales réalisée le plus souvent par quadrature Gaussienne. Ceci complique nettement l'estimation des modèles à effets aléatoires partagés et notamment restreint le nombre d'effets aléatoires possibles. Dans la plupart des travaux, les paramètres du maximum de vraisemblance sont estimés par un algorithme EM qui consiste à maximiser l'espérance de la vraisemblance complète, les effets aléatoires étant considérés comme des données non-observées (Wulfsohn et Tsiatis, 1997 ; Henderson et al., 2000 ; Lin et al., 2002b). Cependant, cette formulation nécessite aussi l'évaluation numérique d'intégrales sur les effets aléatoires. Récemment, des approches Bayésiennes ont aussi été proposées (Xu et Zeger, 2001 ; Chi et Ibrahim, 2006).

Extension du modèle à effets aléatoires partagés

La principale extension des modèles à effets aléatoires partagés a consisté à prendre en compte conjointement plusieurs marqueurs d'évolution soit par des modèles mixtes multivariés présentés en section 2.2.1 soit par des modèles mixtes à variables latentes présentés en section 2.2.2. Ainsi, Thiébaud et al. (2005) ont combiné un modèle linéaire mixte multivarié pour décrire l'évolution de la charge virale et des CD4 en prenant en compte conjointement la sortie d'étude informative. Brown et al. (2005) ont quant à eux étudié l'évolution conjointe des deux marqueurs et leur association avec la survenue du stade SIDA ou du décès. Dans ce travail, ils ont aussi assoupli la modélisation des marqueurs en introduisant un modèle linéaire mixte bivarié non paramétrique basé sur des splines cubiques. Lin et al. (2002b) ont aussi modélisé plusieurs marqueurs longitudinaux conjointement à l'événement pour caractériser à la fois les interactions entre les marqueurs et l'événement, et les

interactions directes entre les marqueurs.

En utilisant des modèles mixtes à variables latentes, Xu et Zeger (2001) ont évalué l'intérêt d'utiliser plusieurs marqueurs longitudinaux plutôt qu'un seul comme marqueurs de substitution d'un événement final dans un essai clinique sur la schizophrénie. Ils ont pour cela introduit en variable explicative du modèle de survie les variables latentes sous-jacentes aux marqueurs. Roy et Lin (2002) ont quant à eux étendu leur modèle mixte à variable latente proposé en 2000 pour prendre en compte la sortie d'étude informative, la variable latente sous-jacente à tous les marqueurs étant prédictrice de la sortie d'étude. Enfin, Chi et Ibrahim (2006) ont modélisé conjointement plusieurs marqueurs d'évolution indépendants entre eux par des modèles mixtes et leur association avec deux événements corrélés. Les variables latentes sous-jacentes aux marqueurs étaient entrées en variables explicatives dans les deux modèles de survie, la corrélation entre les événements étant définie par un paramètre de fragilité commun.

Limites de l'approche

La principale limite des modèles à effets aléatoires partagés réside dans la difficulté de calculer les intégrales multiples sur les effets aléatoires (Henderson et al., 2000). L'extension à l'étude de plusieurs marqueurs corrélés renforce le problème puisque comme nous l'avons vu en section 2.2.1, la corrélation entre les marqueurs est prise en compte par des effets aléatoires corrélés. En considérant que tous les marqueurs sont des mesures d'une même variable latente, le nombre d'effets aléatoires est réduit mais, même dans ce cas, Roy et Lin (2002) évoquent les difficultés d'estimation liées aux intégrales numériques sur les effets aléatoires.

Une deuxième limite de ces modèles vient des hypothèses paramétriques. En effet, l'association entre l'événement et les marqueurs est entièrement spécifiée et il est difficile d'évaluer l'impact de cette hypothèse sur les estimations du modèle. En particulier, des études de sensibilité sont nécessaires (Hogan et Laird, 1997; Fitzmaurice et al., 2001).

Une troisième limite concerne l'interprétation assez complexe de l'association entre le marqueur et l'événement. En effet, cette association s'exprime le plus souvent sous forme de corrélation entre les effets aléatoires du marqueur et le risque d'événement. En plus, la corrélation entre les mesures répétées du marqueur et la corrélation entre l'événement et le marqueur ne sont pas différenciées (Roy, 2003).

Enfin, lorsqu'on étudie l'association entre l'évolution d'un marqueur et la survenue d'un événement, on peut s'attendre à avoir au moins deux types d'évolutions différentes suivant que l'événement a lieu ou pas. Or, dans un modèle à effets aléatoires partagés, on fait l'hypothèse que l'évolution du marqueur est homogène : les effets aléatoires qui prédisent l'événement sont supposés Gaussiens.

Les hypothèses paramétriques difficiles à vérifier, les difficultés numériques et l'interprétation complexe de l'association entre les deux processus dans les modèles à effets aléatoires partagés ont conduit à considérer une autre approche pour modéliser conjointement des marqueurs et un événement : les modèles conjoints à classes latentes.

2.4.2 Modèles conjoints à classes latentes

Ces modèles sont des extensions des modèles de mélange pour données longitudinales présentés en section 2.3.3. L'hypothèse est qu'il existe dans la population un nombre inconnu de sous-populations homogènes à la fois pour l'événement et le marqueur. Chaque sous-population peut avoir sa propre distribution du marqueur et sa propre distribution de l'événement.

Spécification du modèle

Faisons l'hypothèse qu'il existe G classes latentes dans la population. La variable latente c_i décrit l'appartenance aux classes latentes : $c_i = g$ si le sujet i appartient à la classe g . L'appartenance à chaque classe, c'est à dire $\pi_{ig} = P(c_i = g)$, peut être

décrite par le modèle multinomial comme défini en (2.15). Conditionnellement à la classe latente g , le marqueur Y_i est supposé suivre un modèle linéaire mixte :

$$Y_i|_{c_i=g} = X_i\beta + W_i\alpha_g + Z_iu_{ig} + \epsilon_{ig} \quad (2.24)$$

où X_i est la matrice de variables explicatives indépendante de la classe et associée au vecteur d'effets fixes β communs à toutes les classes. La matrice W_i contient des variables explicatives distinctes de X_i et est associée au vecteur d'effets fixes α_g différents dans chaque classe du mélange. La matrice Z_i est une sous-matrice de W_i associée au vecteur d'effets aléatoires u_{ig} distribué suivant une $\mathcal{N}(\mathbf{0}, D)$. Les erreurs ϵ_{ig} sont indépendantes et distribuées suivant une $\mathcal{N}(\mathbf{0}, \sigma^2 I_{n_i})$.

Par rapport à un modèle de mélange pour données longitudinales classique, l'événement est maintenant aussi prédit par les classes de mélange. Muthén et Shedden (1999) suivi de Lin et al. (2000) ont défini la probabilité d'événement T_i dans chaque classe de mélange g par un modèle logistique :

$$P(T_i = 1|_{c_i=g}) = \frac{\exp(\delta_{0g} + x_i\delta_{1g})}{1 + \exp(\delta_{0g} + x_i\delta_{1g})} \quad (2.25)$$

où δ_{0g} est l'intercept spécifique à chaque classe et x_i est le vecteur de variables explicatives associé au vecteur de paramètres δ_{1g} potentiellement différent dans chaque classe latente.

Lin et al. (2002a) et McCulloch et al. (2002) ont ensuite étendu ce modèle pour prendre en compte le délai jusqu'à l'événement plutôt que la probabilité d'événement à un temps donné. Le risque d'événement est défini par un modèle à risques proportionnels :

$$\lambda(t)|_{c_i=g} = \lambda_{0g}(t)\exp(x_i\delta_{1g}) \quad (2.26)$$

où $\lambda_{0g}(t)$ est le risque instantané de base de l'événement spécifique ou non à chaque classe latente.

Dans le modèle conjoint à classes latentes, la variable de classes c_i représente le seul lien entre l'évolution du marqueur et l'événement. Ce modèle est donc similaire aux modèles à effets aléatoires partagés sauf que dans le cas présent l'effet aléatoire

est discret.

Estimation des paramètres

La variable latente représentant le seul lien entre les distributions du marqueur et de l'événement, la distribution conjointe de Y_i et T_i s'écrit de la manière suivante :

$$f(Y_i, T_i) = \sum_g f(Y_i|c_i = g)f(T_i|c_i = g)f(c_i = g) \quad (2.27)$$

La contribution de chaque sujet à la vraisemblance peut donc s'écrire comme une somme sur les classes latentes des contributions spécifiques à chaque classe, et la vraisemblance a la forme suivante :

$$L(\phi; Y_i, T_i) = \prod_{i=1}^N \sum_{g=1}^G \pi_{ig} f(Y_i|c_i = g; \phi) f(T_i|c_i = g; \phi) \quad (2.28)$$

où $f(Y_i|c_i = g; \phi)$ est une distribution multivariée Gaussienne d'espérance $X_i\beta + W_i\alpha_g$ et de variance covariance $Z_i D Z_i^T + \sigma^2 I_{n_i}$, et $f(T_i|c_i = g; \phi)$ est la distribution du modèle logistique défini en (2.25) ou du modèle de survie défini en (2.26).

Cette vraisemblance ayant une expression analytique, les estimateurs du maximum de vraisemblance peuvent être obtenus par des méthodes itératives. Les mêmes difficultés que pour le modèle de mélange pour données longitudinales sont rencontrées, en particulier les problèmes de multimodalité et d'estimation à G fixé. L'algorithme EM est généralement préféré pour l'estimation des paramètres et le critère BIC est utilisé pour sélectionner le nombre optimal de classes latentes (Muthén et Shedden, 1999 ; Lin et al., 2000 ; Lin et al., 2002a).

Applications

Muthén et Shedden (1999) ont proposé un modèle conjoint à classes latentes et l'ont appliqué dans plusieurs études en psychologie. Ils ont ainsi étudié l'association

entre l'évolution de la consommation d'alcool entre 18 et 25 ans et la probabilité de dépendance à l'alcool à 30 ans (Muthén et Shedden, 1999). Une autre étude a porté sur l'impact d'une intervention pédagogique à la fois sur l'évolution du comportement agressif au cours du temps et sur des événements tels que l'exclusion de l'école ou une arrestation (Muthén et al., 2002). A travers cette dernière application, un intérêt majeur des modèles à classes latentes est souligné : ce modèle permet d'évaluer l'impact de variables explicatives sur l'évolution du marqueur, sur l'événement lui même ou bien sur l'association entre les deux. Dans l'étude de l'incidence du cancer de la prostate, plusieurs travaux ont permis de mettre en évidence les différents profils d'évolution de la PSA et son association avec la survenue d'un cancer de la prostate. La PSA, l'antigène spécifique de la prostate, est le biomarqueur reconnu dans l'étude du cancer de la prostate. Lin et al. (2000) ont d'abord montré grâce à un modèle à classes latente qu'il existait 4 profils différents d'évolution associés à des risques plus ou moins accrus de cancer plusieurs années plus tard et Lin et al. (2002a) ont confirmé ces résultats et proposé un outil de détection précoce d'un cancer à partir des mesures répétées de PSA.

Avantages et limites du modèle conjoint à classes latentes

Le modèle à classes latentes est pour l'instant beaucoup moins fréquemment utilisé que le modèle à effets aléatoires partagés pour étudier conjointement un marqueur longitudinal et un événement. Pourtant, les modèles conjoints à classes latentes offrent certains avantages par rapport au modèles à effets aléatoires partagés.

Tout d'abord, la variable latente partagée dans un modèle à classes latentes étant discrète, la densité conjointe s'écrit comme une somme sur les classes latentes et non une intégrale sur les effets aléatoires. Les effets aléatoires interviennent uniquement dans le modèle mixte qui dans le cas linéaire a une vraisemblance analytique. La vraisemblance du modèle conjoint à classes latentes a donc la plupart du temps une expression analytique et est facilement maximisable. Deuxièmement, dans le

modèle à classes latentes, la corrélation entre le marqueur et l'événement (par les classes latentes) et la corrélation entre les mesures répétées du marqueur (par les effets aléatoires) sont bien distinctes alors qu'elles sont confondues dans le modèle à effets aléatoires partagés classique. De plus, la représentation de profils et courbes de survie dans chaque classe latente simplifie l'interprétation des résultats. Enfin, lorsqu'on étudie l'association entre l'évolution d'un marqueur et la survenue d'un événement, il est intuitif de faire une hypothèse d'hétérogénéité dans la population, hypothèse que le modèle à classes latentes prend bien en compte.

Les modèles conjoints à classes latentes peuvent donc paraître préférables aux modèles de sélection. Cependant, ils présentent deux inconvénients. L'hypothèse d'indépendance conditionnellement aux classes latentes est au moins aussi forte que l'hypothèse d'indépendance conditionnellement aux effets aléatoires dans le modèle à effets aléatoires partagés. En plus, dans un modèle à effets aléatoires partagés, il est possible de spécifier plusieurs types de combinaisons linéaires des effets aléatoires dans le modèle de survie, ce qui rend la modélisation de la dépendance plus souple. Mais en contrepartie, le modèle conjoint à classes latentes permet de tester des effets de variables explicatives spécifiques aux classes latentes. Le deuxième inconvénient est que, les modèles à classes latentes étant une extension des modèles de mélange pour données longitudinales, ils peuvent présenter des maxima locaux et doivent être estimés pour un nombre de classes fixé, le nombre optimal de classes latentes étant choisi par un critère de sélection.

Dans le vieillissement cognitif, comme il existe une hétérogénéité des déclin, nous avons préféré utiliser les modèles conjoints à classes latentes plutôt que des modèles à effets aléatoires partagés pour étudier conjointement l'évolution cognitive hétérogène et la survenue d'une démence dans ce travail de thèse.

2.4.3 Cas particulier : modélisation de la sortie d'étude informative

Une des applications qui a largement motivé le développement de modèles pour décrire conjointement l'évolution d'un marqueur et la survenue d'un événement est le traitement de la sortie d'étude informative dans les analyses longitudinales. L'objectif est de corriger les biais dans l'inférence de l'évolution du marqueur Y en prenant en compte le délai jusqu'à la sortie d'étude T . Pour cela, deux approches ont été envisagées (Hogan et Laird, 1997 ; Little, 1995) :

- le modèle de sélection : $f(Y, T) = f(Y)f(T|Y)$
- le modèle par mélange de profils d'observations : $f(Y, T) = f(Y|T)f(T)$

Les modèles de sélection pour la sortie d'étude

Le modèle de sélection consiste à modéliser la distribution jointe du marqueur Y_i et du temps d'événement T_i pour un sujet i ($i = 1, \dots, N$) soit en faisant l'hypothèse que la distribution de T_i dépend directement de Y_i ("outcome-dependent model") :

$$f(Y_i, T_i; \theta, \psi) = f(Y_i; \theta)f(T_i|Y_i; \psi) \quad (2.29)$$

Soit en supposant qu'elle dépend des effets aléatoires du modèle mixte pour Y_i ("random-effects-dependent model") :

$$f(Y_i, T_i; \theta, \psi) = \int f(Y_i | u_i; \theta)f(T_i | u_i; \psi)f(u_i)du_i \quad (2.30)$$

La deuxième formulation, correspondant à un modèle à effets aléatoires partagés présenté en section 2.4.1, est la formulation la plus courante.

Le modèle de sélection est particulièrement intéressant pour prendre en compte la sortie d'étude car il permet de décrire directement et sans biais l'évolution marginale du marqueur caractérisée par le vecteur θ (Roy et Lin, 2002 ; Little, 1995). En effet, les paramètres décrivant l'évolution du marqueur marginalement par rapport

à la sortie d'étude sont estimés directement. Cependant, l'intégration numérique sur les Y_i manquants dans (2.29) ou sur les effets aléatoires dans (2.30) complique l'estimation de ces modèles.

Les modèles par mélange de profils d'observations

Cette méthode consiste à modéliser la distribution jointe du marqueur Y_i et du temps d'événement T_i pour un sujet i ($i = 1, \dots, N$) en considérant la distribution conditionnelle de Y_i sachant T_i :

$$f(Y_i, T_i; \phi, \zeta) = f(Y_i | T_i; \phi) f(T_i; \zeta) \quad (2.31)$$

Plus précisément, $f(Y_i | T_i)$ est défini soit en incluant la date de sortie d'étude comme variable explicative dans le modèle pour Y_i , soit en stratifiant sur les profils d'observation (ou dates de sortie d'étude).

L'idée de conditionner l'analyse sur les profils de sortie d'étude permet une prise en compte simple et intuitive du processus de sortie d'étude et évite notamment de faire des hypothèses sur sa distribution. De plus, la vraisemblance du modèle de mélange a une écriture analytique qui permet donc d'utiliser des méthodes classiques d'estimation par le maximum de vraisemblance.

Cependant, le modèle par mélange de profils d'observations ne permet pas d'obtenir directement les paramètres d'intérêt, c'est à dire ceux caractérisant l'évolution marginale du marqueur et notamment l'effet marginal de variables explicatives sur le marqueur. Pour pallier cette limite, 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 de sortie d'étude. Cela permet d'obtenir directement l'évolution marginale sans biais du marqueur comme dans le modèle de sélection.

Enfin, un problème majeur dans les modèles par mélange de profils observés est qu'il peut exister un nombre important de profils et que pour certains profils,

l'information sur Y_i peut être insuffisante pour estimer le modèle. Dans ce cas, des contraintes supplémentaires sont nécessaires. Pour pallier ces problèmes, Roy (2003) a proposé un modèle dans lequel le nombre de classes dans le mélange est réduit ; les classes ne sont plus observées mais latentes. Au sein de chaque classe latente, la distribution du marqueur est la même et l'appartenance à la classe latente est déterminée en fonction du temps de sortie d'étude. Ce modèle peut être vu comme un modèle de mélange pour données longitudinales dans lequel le temps jusqu'à l'événement prédit l'appartenance à un profil d'évolution. Depuis, Lin et al. (2004) ont proposé une approche par classes latentes pour prendre en compte les données manquantes intermittentes. Les modèles conjoints à classes latentes constituent en effet une approche intermédiaire entre les modèles par mélange de profils d'observations et les modèles à effets aléatoires partagés.

Chapitre 3

Modèle nonlinéaire à processus latent pour des données longitudinales multivariées

Dans ce chapitre, nous proposons un modèle nonlinéaire à processus latent pour étudier conjointement plusieurs tests psychométriques quantitatifs non Gaussiens. Le modèle statistique, son estimation ainsi qu'une illustration dans l'étude du vieillissement cognitif ont fait l'objet d'un article publié dans la revue *Biometrics* et présenté dans la section suivante. Nous détaillons ensuite deux études de simulation ayant pour but d'évaluer les performances du modèle et décrivons enfin deux travaux d'application basés sur cette méthodologie. Le premier a consisté à évaluer les propriétés métrologiques de quatre tests psychométriques et le deuxième a consisté à distinguer l'association du sexe et du niveau d'éducation avec le déclin cognitif global de leur association avec les tests psychométriques. Enfin, nous discutons les limites de l'approche et les améliorations possibles.

3.1 Article méthodologique

Résumé de l'article

La cognition n'est pas une quantité directement observée, elle est mesurée par plusieurs tests psychométriques quantitatifs. Ces tests psychométriques ont des distributions différentes, souvent non Gaussiennes, certains tests souffrant en particulier d'un effet plafond ou plancher. Etudier l'évolution cognitive et l'effet de variables explicatives en utilisant l'information d'un seul test peut donc aboutir à des résultats différents. Dans cet article, nous proposons d'étudier l'évolution du facteur commun à plusieurs tests psychométriques, ce facteur commun pouvant être vu comme le processus cognitif sous-jacent aux tests psychométriques. Le facteur commun est défini en temps continu et mesuré à des temps discrets par les tests psychométriques. Son évolution suit un modèle linéaire mixte incluant des variables explicatives dépendantes du temps et un mouvement Brownien. Pour prendre en compte les propriétés métrologiques différentes des tests psychométriques, nous considérons que chaque test psychométrique est une transformation nonlinéaire souple paramétrée du facteur commun. L'estimation des paramètres du modèle contenus à la fois dans le modèle linéaire mixte et dans les transformations nonlinéaires est obtenue en maximisant la vraisemblance observée du modèle. Nous proposons deux méthodes graphiques pour évaluer l'adéquation du modèle et une illustration sur un échantillon de la cohorte française PAQUID.

A Nonlinear Model with Latent Process for Cognitive Evolution Using Multivariate Longitudinal Data

Cécile Proust,^{1,*} Hélène Jacqmin-Gadda,¹ Jeremy M. G. Taylor,² Julien Ganiayre,¹ and Daniel Commenges¹

¹INSERM E0338, Université de Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France

²Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, Michigan 48109, U.S.A. **email*: Cecile.Proust@isped.u-bordeaux2.fr

SUMMARY. Cognition is not directly measurable. It is assessed using psychometric tests, which can be viewed as quantitative measures of cognition with error. The aim of this article is to propose a model to describe the evolution in continuous time of unobserved cognition in the elderly and assess the impact of covariates directly on it. The latent cognitive process is defined using a linear mixed model including a Brownian motion and time-dependent covariates. The observed psychometric tests are considered as the results of parameterized nonlinear transformations of the latent cognitive process at discrete occasions. Estimation of the parameters contained both in the transformations and in the linear mixed model is achieved by maximizing the observed likelihood and graphical methods are performed to assess the goodness of fit of the model. The method is applied to data from PAQUID, a French prospective cohort study of ageing.

KEY WORDS: Cognitive ageing; Mixed model; Multiple outcomes; Random effects.

1. Introduction

In cognitive ageing studies, cognition is generally evaluated through a battery of psychometric tests, which are quantitative measures of various dimensions of cognition. Describing cognitive evolution and assessing the impact of covariates on this evolution is an interesting approach to help us understand the process of cognitive ageing. As the various psychometric tests are highly correlated, multivariate longitudinal analyses of several psychometric tests are often performed using multivariate linear mixed models (Hall et al., 2001; Harvey, Beckett, and Mungas, 2003; Sliwinski, Hofer, and Hall, 2003). These models highlight both the differences in the shapes of evolution for each dimension and the strong correlation between the dimensions.

The idea of a latent cognitive process explaining the cognitive decline in the elderly is hypothesized in neuropsychology. This latent cognitive process can be viewed as a common cognitive factor across all the psychometric tests (Salthouse et al., 1996; Fabrigoule et al., 1998) and is supposed to be a better predictor of dementia and cognitive decline. As a consequence, it would be of substantial interest to focus the analysis on this latent process by describing its evolution and evaluating the impact of covariates directly on it.

In a cross-sectional framework, Sammel and Ryan (1996) proposed a latent variable model in which covariates could affect directly the latent variable, and the multiple outcomes were assumed to be measures of the underlying latent variable with error. In a longitudinal framework, Gray and Brookmeyer (1998) proposed a marginal regression model, with estimation via generalized estimating equations, to

assess an overall treatment effect on several continuous and repeated outcomes. Roy and Lin (2000) also extended the linear latent variable model of Sammel and Ryan (1996) to repeated multivariate data. In practice, the assumption of a linear relationship between the outcomes and a Gaussian latent variable is frequently too strong, because the psychometric tests often have non-Gaussian distributions due to different metrological properties and different behaviors with ageing (Hall et al., 2001; Amieva et al., 2005). For instance, some tests may be more sensitive to changes at high levels of cognition than at low levels of cognition, while others may have the same sensitivity at high and low levels of cognition. Thus, we propose to introduce parameterized flexible nonlinear transformations to link the quantitative tests with the latent process. The latent process is defined in continuous time by a linear mixed model including a Brownian motion, and nonlinear transformations of the psychometric tests are noisy measures of the latent process at discrete occasions, the shapes of the estimated nonlinear transformations giving information on the metrological properties of each test.

This extension of mixed models to latent variable models is related to structural equation models (SEM), mainly developed in psychometrics, because in both approaches the quantity of interest cannot be measured directly and is evaluated instead by a set of outcomes or items (Muthén, 2002; Dunson, 2003; Rabe-Hesketh, Skrondal, and Pickles, 2004). Thus the formulation of the model has two components, a measurement model which links the latent variables with the observations and a structural model which explains the latent variable structure. In the last decade, there have been

major improvements in SEM (Sánchez et al., 2005). These include (i) to handle clustered or repeated data (Longford and Muthén, 1992; Dunson, 2003; Rabe-Hesketh et al., 2004; Skrondal and Rabe-Hesketh, 2004; Song and Lee, 2004), (ii) to allow mixture of count, ordinal, and dichotomous outcomes (Dunson, 2003; Lee and Song, 2004; Rabe-Hesketh et al., 2004), (iii) to relax linearity of the relationship between the latent variables by using nonlinear structural models (Jöreskog and Yang, 1996; Arminger and Muthén, 1998; Wall and Amemiya, 2000; Lee and Song, 2004; Song and Lee, 2004), and (iv) to relax linearity between the continuous responses and the latent variables (Yalcin and Amemiya, 2001).

Our modeling approach differs in a number of ways. First, we focus on the change over time of a single common latent process, while the main interest of SEM lies in the relationship between several latent variables. Moreover, when dealing with quantitative outcomes, SEM generally assumes a Gaussian or a Poisson distribution for the outcomes. Except for threshold models for ordinal data (Dunson, 2003; Lee and Song, 2004; Rabe-Hesketh et al., 2004), when nonlinear transformations link the latent variables and the outcomes, they do not depend on parameters to be estimated. As threshold models are not appropriate for quantitative scores with many possible values, we estimate the shape of the transformations by using parameterized nonlinear functions. Finally, our model includes a continuous-time latent process; this gives a description of the evolution of the latent cognitive level for all times in the range of the observations and furthermore, it can easily handle data where the number and times of the observations are different for each subject and for each outcome.

Nonlinearity in SEM either in the structural model or in the relationship between observed outcomes and latent variables requires the development of suitable estimation methods. For models including products of latent variables, Jöreskog and Yang (1996) proposed a frequentist approach based on the maximization of the likelihood, while Arminger and Muthén (1998) proposed a Bayesian approach using a Markov chain Monte Carlo (MCMC) algorithm. For models with nonlinear relationships between the responses and the latent variables, Yalcin and Amemiya (2001) proposed to compute a quadratic approximation of the nonlinear transformations, and then maximized the approximate likelihood. In contrast, to handle the nonlinear relationships between the responses and the latent process, we propose to maximize the exact likelihood of the observed data, which is a product of the likelihood of the transformed data (the transformed data are multivariate Gaussian in our model) and the Jacobian of the nonlinear transformations.

The main characteristics of our methodology can be summarized as follows:

- (a) it can be applied to multivariate longitudinal non-Gaussian quantitative outcomes;
- (b) it can study the evolution of a continuous-time latent process representing the common factor across all the outcomes;
- (c) it can estimate the shape of the transformations linking the quantitative outcomes and the underlying latent process;
- (d) it can handle any type of unbalanced data (number and time of measurements, covariates, ...) and missing at random data;
- (e) it can estimate impact of covariates on both the latent process and the observed outcomes.

The next section focuses on the formulation of the model for the latent process and the outcomes on the parameterized nonlinear transformations. Section 3 is devoted to maximum likelihood estimation (MLE). In Section 4 we discuss goodness of fit and Section 5 focuses on an application of the method to data from the French prospective cohort study PAQUID (Letenneur et al., 1994).

2. Methodology

2.1 The Latent Process: Structural Model

Consider the continuous-time latent process $\Lambda_i = (\Lambda_i(t))_{t \geq 0}$ representing the common cognitive factor for individual i with $i = 1, \dots, N$. Λ_i is defined at time $t, t \in \mathbb{R}^+$ according to a linear mixed model,

$$\Lambda_i(t) = X_{1i}(t)^T \beta + Z_i(t)^T u_i + \sigma_w w_i(t), \quad t \geq 0, \quad (1)$$

where $X_{1i}(t)$ is the q_1 vector of time-dependent covariates associated with the vector of fixed effects β . The $(p+1)$ vector $Z_i(t) = (1, t, \dots, t^p)^T$ is a time polynomial of degree p (or any vector of functions of time) and the vector of random effects at subject level $u_i \sim N(\mu, D)$, where D is an unstructured positive definite matrix. The process $w_i = (w_i(t))_{t \geq 0}$ is a standard Brownian motion; $w_i(t)$ models local variation and departure from the polynomial trend while the random effects account for the variability of the trend across the subjects. No independent error is added because this latent process is assumed to represent the actual cognition in continuous time. Note that the linearity in β or in the covariates is not crucial. Any function of time could be included in the model, because the model is still linear in the random effects, to ensure the normality of the latent process. Moreover, the Brownian motion also adds flexibility to the parametric function of time.

2.2 The Measurement Model

Now consider K quantitative outcomes. Each outcome could be an individual psychometric test, or the sum of scores from an itemized test. For subject i and outcome k , we observe the n_{ik} vector of measurements $y_{ik} = (y_{i1k}, \dots, y_{ij_k k}, \dots, y_{in_{ik} k})^T$, where y_{ijk} is the score of subject i at occasion j for test k . The number and times of measurements may be completely different for each subject and each outcome. In the spirit of latent growth curve modeling (Muthén, 2002) and SEM (Yalcin and Amemiya, 2001), we assume that this measurement y_{ijk} is related to the latent process at time t_{ijk} through the following flexible model,

$$g_k(y_{ijk}; \eta_k) = \tilde{y}_{ijk} = \Lambda_i(t_{ijk}) + \alpha_{ik} + X_{2i}(t_{ijk})^T \gamma_k + \epsilon_{ijk}, \quad (2)$$

where the function g_k comes from a family of nonlinear transformations \mathcal{G} depending on the vector of parameters η_k , which will be estimated; the random effects α_{ik} are independently distributed according to an $N(0, \sigma_{\alpha_k}^2)$ distribution; the vectors $X_{2i}(t_{ijk})$ and γ_k are, respectively, a q_2 vector of time-dependent covariates and the associated vector of contrasts

for the test k ; ϵ_{ijk} are independent Gaussian errors with mean 0 and variance $\sigma_{\epsilon_k}^2$.

As in Dunson (2003), the random effect α_{ik} accounts for the fact that for a same value of the latent process, two subjects can score differently in the cognitive domain associated with psychometric test k . The contrasts γ_k make the relationship between the outcomes and the latent process more flexible by allowing some covariates to be differently associated with the various outcomes. The sum of the contrasts over the K tests for a given covariate equals 0. Thus, parameters β in (1) capture the mean association with the covariates contained both in $X_{1i}(t)$ and $X_{2i}(t)$, while parameters γ_k in (2) capture the variability of the association for each test around this mean value.

2.3 The Choice of the Family of Functions \mathcal{G}

For all the outcomes, the transformations $g_k(y; \eta_k)$ come from the same family of functions \mathcal{G} . The choice of the family is a key aspect of the model; it determines the flexibility of the link between the joint outcomes with various behaviors and the underlying latent process. The transformations must be monotonic and increasing functions of y and depend on few parameters to make the estimation of the model easier. So, the choice of the family \mathcal{G} is a compromise between flexibility and parsimony.

The first transformation considered here is the beta cumulative distribution function (CDF), which can take very different shapes, including concave, convex, and sigmoid, according to the parameters, as illustrated in Figure 1. It is defined for $y \in [0, 1]$, $\eta_{1k} > 0$, and $\eta_{2k} > 0$ by

$$g_k(y; \eta_{1k}, \eta_{2k}) = \int_0^y \frac{x^{\eta_{1k}-1}(1-x)^{\eta_{2k}-1}}{B(\eta_{1k}, \eta_{2k})} dx. \quad (3)$$

As the beta CDF is defined in $[0, 1]$, for each psychometric test, a preliminary step consists of rescaling the tests to the unit interval.

The main drawback of this transformation is its computational complexity. As a consequence, simpler transformations have also been considered to compare the fits of the models: the linear transformation, the logit transformation combined with a linear transformation, and the Weibull cumulative distribution function (details in the Appendix). When using a linear transformation, the model is a multivariate linear mixed model similar to Roy and Lin (2000) or Rabe-Hesketh et al. (2004), with an additional Brownian motion term. In that case, constraints have to be added to make the model identifiable: we assume the intercept μ_0 equals 0 and the variance of the random intercept u_{0i} equals 1. In contrast, when using a CDF, the requirement that $g_k(y)$ is in $[0, 1]$ avoids additional constraints on the latent process.

3. Estimation

Parameter estimation is achieved using maximum likelihood techniques assuming that missing data are missing at random. A nonstandard aspect of the model is the presence of parameters both in the nonlinear transformation g_k of the outcome and in the model for the transformed response $\tilde{y}_i = (\tilde{y}_{i11}, \dots, \tilde{y}_{in_{i1}}, \dots, \tilde{y}_{ijk}, \dots, \tilde{y}_{i1K}, \dots, \tilde{y}_{in_{iK}})^T$, where $\tilde{y}_{ijk} = g_k(y_{ijk})$. The log likelihood of interest is the log likelihood of the outcomes in their natural scale, and thus includes the Jacobian of the transformations g_k . It is given by

$$L(y; \theta) = L(\tilde{y}; \theta) + \ln(J(y; \theta)) = \sum_{i=1}^N L(\tilde{y}_i; \theta) + \sum_{i=1}^N \ln(J(y_i; \theta)), \quad (4)$$

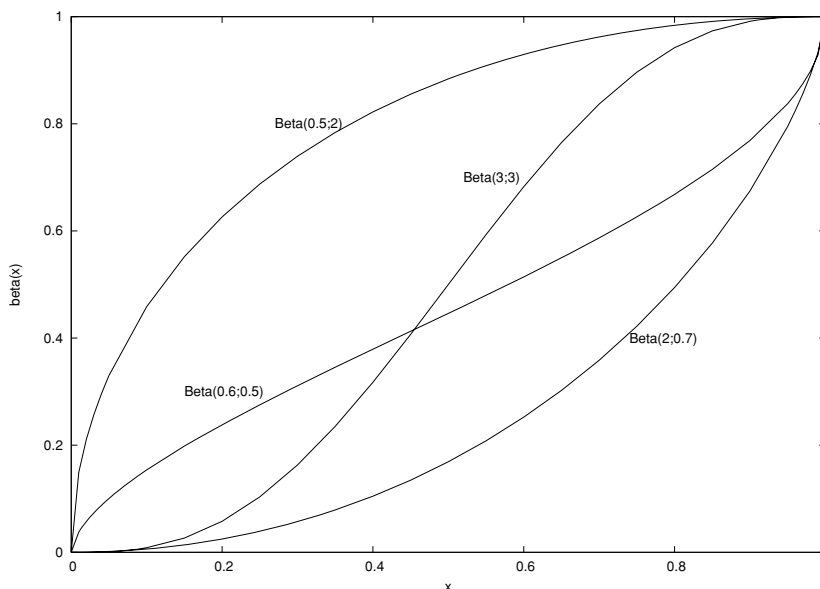


Figure 1. Examples of beta transformations for various pairs of parameter values.

where θ is the complete vector of parameters containing the transformation parameters $\eta'_k = (\eta_{1k}, \eta_{2k}), k = 1, \dots, K$, the fixed parameters $\mu, \beta, \gamma_1, \dots, \gamma_K$, and the variance-covariance parameters $\text{vec}(D), \sigma_w, \sigma_{\alpha_1}, \dots, \sigma_{\alpha_K}, \sigma_{\epsilon_1}, \dots, \sigma_{\epsilon_K}$. $J(y; \theta)$ is the Jacobian of the transformation given the data and the vector of parameters θ . For the beta transformation, the Jacobian is defined by

$$J(y; \theta) = \prod_{k=1}^K \prod_{j=1}^{n_{ik}} \frac{y_{ijk}^{\eta_{1k}-1} (1-y_{ijk})^{\eta_{2k}-1}}{B(\eta_{1k}, \eta_{2k})}. \quad (5)$$

Formulae of the Jacobian for the other potential transformations are given in the Appendix.

$L(\tilde{y}_i; \theta)$ is the log likelihood of the transformed data for subject i . Let $Z_i^k = (Z(t_{i1k}), \dots, Z(t_{in_{ik}k}))^T$ be the $n_{ik} \times (p+1)$ matrix of time polynomials for subject i and test k ; $X_{1i}^k = (X_{1i}(t_{i1k}), \dots, X_{1i}(t_{in_{ik}k}))^T$ and $X_{2i}^k = (X_{2i}(t_{i1k}), \dots, X_{2i}(t_{in_{ik}k}))^T$ are, respectively, the $n_{ik} \times q_1$ matrix of time-dependent covariates for the latent process and $n_{ik} \times q_2$ matrix of time-dependent covariates for the psychometric tests. Let I_n be the identity matrix of size n , and J_n , the matrix of size n where all the elements equal 1. Then, the density of \tilde{y}_i is a multivariate Gaussian density of size $n_i = \sum_{k=1}^K n_{ik}$ with mean $E_i = (E_{i1}^T, \dots, E_{iK}^T)^T$ and covariance matrix V_i given by

$$E_{ik} = Z_i^k \mu + X_{1i}^k \beta + X_{2i}^k \gamma_k \quad (6)$$

$$V_i = \begin{pmatrix} Z_i^1 \\ \vdots \\ Z_i^K \end{pmatrix} D \begin{pmatrix} Z_i^{1T} & \dots & Z_i^{KT} \end{pmatrix} + V_w + \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_K \end{pmatrix},$$

$$\text{with } \Sigma_k = \sigma_{\alpha_k}^2 J_{n_{ik}} + \sigma_{\epsilon_k}^2 I_{n_{ik}} \quad (7)$$

and V_w the covariance matrix for the Brownian process with argument $\sigma_w^2 (\min(t_l, t_m))$ for $(l, m) \in [1, n_i]^2$. The contribution of subject i to the log likelihood of the transformed data $L(\tilde{y}_i; \theta)$ is the logarithm of this multivariate density taken at the observation values. The log likelihood (4) has a closed form (except for the computation of the beta CDFs for which standard routines are available) and is maximized using a modified Marquardt algorithm (Marquardt, 1963), which is a Newton-Raphson-like algorithm. The vector of parameters θ is updated until convergence using

$$\theta^{(l+1)} = \theta^{(l)} - \delta (\tilde{H}^{(l)})^{-1} \nabla (L(y; \theta^{(l)})). \quad (8)$$

The step δ equals 1 by default but can be modified to ensure that the likelihood is improved at each iteration. The matrix \tilde{H} is a diagonal-inflated Hessian to ensure positive definiteness. $\nabla(L(y; \theta^{(l)}))$ is the gradient of the log likelihood (4) at iteration l . First and second derivatives are computed by finite differences. The program is written in Fortran90 and is available on the web site <http://www.isped.u-bordeaux2.fr>. This algorithm is less computationally demanding than alternative Monte Carlo approaches such as in Arminger and Muthén (1998), who proposed a Bayesian approach for latent variable models with nonlinear relationships between the latent variables. Nevertheless, it is computationally intensive and, for example, with a sample of 563 subjects (8227 observations) and a model with 36 parameters (the final model in

the application), the CPU time is around 15 minutes using a Bi-Xeon 3.06 GHz 1024 MB RAM.

Moreover, after convergence, standard error estimates of the parameter estimates are directly obtained using the inverse of the Hessian. A bootstrap method using 200 resamples of the N subjects is also performed for obtaining standard errors of $g_k(y, \hat{\eta}_k)$, where y is in the range of the psychometric test k .

4. Assessment of the Fit

An unsolved question in mixed modeling is the assessment of the goodness of fit. In this work, we propose two approaches to evaluate the adequacy of the model, a residual-based approach and a prediction-based approach. The residual-based approach consists of evaluating the Gaussian distribution of the standardized marginal residuals $\hat{\epsilon}_i$ given by

$$\hat{\epsilon}_i = U_i (\tilde{y}_i - \hat{E}_i), \quad (9)$$

where U_i is the upper triangular matrix of the Cholesky transformation of V_i^{-1} and $\hat{E}_i = E_{\hat{\theta}}(\tilde{y}_i)$ is obtained by replacing the parameters by their MLE in (6). A normal quantile plot with the 95% confidence bands computed using the Kendall and Stuart formula (Kendall and Stuart, 1977, p. 251) is then displayed to evaluate whether the empirical distribution of the standardized residuals $\hat{\epsilon}_{ijk}$ is close to the theoretical $N(0, 1)$ distribution.

To evaluate the fit of the data on the natural scale of the tests, we plot the observed mean evolution of each test versus the estimated marginal mean evolution or the conditional mean evolution, which includes random effects estimates. The marginal estimated means $E_{\hat{\theta}}(g_k^{-1}(\tilde{y}_{ijk}))$ and the conditional estimated means $E_{\hat{\theta}}(g_k^{-1}(\tilde{y}_{ijk}) | \hat{u}_i, \hat{\alpha}_{ik}, \hat{w}_i)$ are computed by numerical integration of $g_k^{-1}(\tilde{y}_{ijk})$ over the marginal distribution of \tilde{y}_{ijk} , $N(E_{ik}(\hat{\theta}); V_i(\hat{\theta}))$, or over the conditional distribution $N(E_{ik}(\hat{\theta}) + \hat{W}_{ijk}; \hat{\sigma}_k I_{n_{ik}})$. Here the marginal expectation and variance of \tilde{y}_{ijk} is given by (6) and (7) and $\hat{W}_{ijk} = Z_i(t_{ijk})^T \hat{u}_i + \hat{w}_i(t_{ijk}) + \hat{\alpha}_{ik}$ is the empirical Bayes estimate of the subject-specific deviation from the model.

5. Application: Cognitive Evolution in the Elderly

5.1 The Data

The aim of this analysis is to describe the decline with age of the global cognitive ability measured by several psychometric tests and to evaluate the association of covariates, especially Apolipoprotein E (apoE) genotype, with the latent cognitive process. Indeed, the presence of one or two $\epsilon 4$ alleles of apoE is associated with a higher risk of Alzheimer's disease (Farrer et al., 1997) but it is not well established whether the $\epsilon 4$ allele is more generally associated with cognitive ageing (Winnock et al., 2002).

The data came from the French prospective cohort study PAQUID, initiated in 1988 to study normal and pathological ageing (Letenneur et al., 1994). Subjects included in the cohort were 65 years and older at the initial visit and were followed six times with intervals of 2 or 3 years. At each visit, a battery of psychometric tests was completed and an evaluation of whether the person satisfied the criteria for a diagnosis of dementia was carried out. Measurements at the initial visit were excluded because of a first passing effect (Jacquemin-Gadda et al., 1997). In the analysis, we included subjects who

were free of dementia at the first follow-up and with at least one measurement for each of four ($K = 4$) psychometric tests during the follow-up.

The four tests considered are the Mini Mental State Examination ($k = 1$), the Isaacs Set Test ($k = 2$), the Benton Visual Retention Test ($k = 3$), and the Digit Symbol Substitution Test of Wechsler ($k = 4$). The Mini Mental State Examination (MMSE) evaluates various dimensions of cognition (memory, calculation, orientation in time and space, language, and word registration); it ranges from 0 to 30 and the distribution is strongly skewed to left with a ceiling effect. The Isaacs Set Test (IST) shortened at 15 seconds evaluates verbal fluency accounting for the speed of execution: subjects have to give a list of words (with a maximum of 10 words) in four semantic categories. It ranges from 0 to 40 and the distribution is close to a Gaussian distribution with a little heavier left tail. The Benton Visual Retention Test (BVRT) evaluates visual memory: subjects have to recognize 15 geometric figures among four proposals. It ranges from 0 to 15 and the distribution is skewed to left but the ceiling effect is less strong than for the MMSE. The Digit Symbol Substitution Test of Wechsler (DSSTW) evaluates attention: given a table of correspondence between symbols and numbers, subjects have to translate a sequence of 90 numbers into the right sequence of symbols. In the sample, it ranges from 0 to 76 and the distribution is approximately Gaussian. For the four tests, low values indicate a more severe impairment. In the analysis, rescaled scores computed as the value of the test plus 0.5 divided by 1 plus the range of the observed values produced values in the open interval (0, 1) and were considered as continuous. For the DSSTW, the observed range was 76 while the maximum possible value was 90. An additional analysis performed using 90 instead of 76 for rescaling led to nearly identical results. More generally, we think it is better to use the observed range for rescaling to avoid interpreting the relationship between the score and the latent process on an unobserved range of values.

The apoE genotype was collected on a subsample of the PAQUID cohort, so the sample used in the analysis consisted of 563 subjects having between 1 and 6 measurements per test (median = 4). The covariates included in the analysis were gender, educational level (graduated from primary school vs. lower level), and the apoE genotype ($\epsilon 4$ carrier vs. $\epsilon 4$ noncarrier). The time scale was the age minus 65 years per 10 years ($t = \frac{\text{age} - 65}{10}$).

5.2 Comparison of the Fit for the Various Families of Transformations

We first assumed that the latent cognition was a quadratic function of time without covariates in expression (1) and without any contrast in expression (2). Using this model, we compared the fit for the beta CDF, the linear transformation, the combination of a linear transformation and the logit transformation, and the Weibull CDF. According to the Akaike information criterion (AIC) (see Table 1), the beta transformation gave a markedly better fit.

5.3 Estimations of the Model with the Beta Transformation

Using the beta transformation, the best fitting model included a quadratic function of time with three random coefficients and the three covariates (educational level, gender, and apoE

Table 1

Fit of the data for various transformations in the model without covariates and a quadratic function of time

Family of transformation	Number of parameters	Log likelihood	AIC
Linear transformation	20	-21584.1	43208.2
Beta CDF	22	-20387.1	40818.2
Logit + linear transformation	20	-20876.4	41792.8
Weibull CDF	22	-20654.7	41353.4

genotype) in the model for the latent process. As it was suspected that ability in visual memory, verbal fluency, and attention could be differently associated with gender and educational level, we also included contrasts between tests for these covariates. Interactions between apoE genotype and time variables were also included in the latent process. Interactions between gender and time and between educational level and time were not found to be significant and did not confound the association between apoE and cognitive evolution. Thus they were excluded from the final model. Estimates of the fixed effect parameters in the final model are presented in Table 2.

Table 2

Estimates of the fixed effect parameters in the best model with the beta transformation (log likelihood = -19715.55; number of parameters = 36; AIC = 39503.1)

Parameter	Estimate	SE
Intercept: μ_0	0.538	0.013
Linear slope: μ_1	-0.0044	0.0098
Quadratic slope: μ_2	-0.0291	0.0040
Gender ^a	-0.0062	0.0071
Education ^b	0.111	0.0088
apoE ^c	0.0070	0.0096
apoE ^c \times t^2	-0.0103	0.0033
Contrasts on gender ^a ($p = 0.027^d$)		
On MMSE ($k = 1$)	-0.0095	0.0052
On IST ($k = 2$)	-0.0052	0.0062
On BVRT ($k = 3$)	0.0148	0.0052
On DSSTW ($k = 4$)	-0.0001	0.0047
Contrasts on education ^b ($p = 0.136^d$)		
On MMSE ($k = 1$)	-0.0117	0.0061
On IST ($k = 2$)	0.0108	0.0070
On BVRT ($k = 3$)	-0.0044	0.0062
On DSSTW ($k = 4$)	0.0053	0.0058
η_{11} (MMSE)	1.409	0.097
η_{21} (MMSE)	0.401	0.018
η_{12} (IST)	0.952	0.064
η_{22} (IST)	0.697	0.041
η_{13} (BVRT)	0.887	0.062
η_{23} (BVRT)	0.569	0.032
η_{14} (DSSTW)	0.477	0.027
η_{24} (DSSTW)	0.838	0.057

^aReference: female.

^bReference: not graduated from primary school.

^cReference: $\epsilon 4$ noncarrier.

^dLikelihood ratio test for the contrast variables (χ^2 with 3 degrees of freedom).

The test-specific random effects α_{ik} dramatically improved the fit (574 increase of the log likelihood for four additional parameters), which means that for a same value of latent cognition, subjects score differently in cognitive domains associated with the psychometric tests. Accounting for the within-subject variability with a Brownian motion was also relevant since it increased the log likelihood of 13.8.

Gender was not significantly associated with the mean common factor level, while subjects who graduated from primary school had a significantly better mean common factor level. Inclusion of contrasts between tests for gender improved significantly the fit of the model, showing that gender does not have the same impact on each psychometric test: men tend to perform better on the BVRT than women, while the trend is reversed for the other tests. Contrasts between tests for educational level are not significant, which suggests that the effect of educational level does not differ from test to test.

The apoE genotype was only included in the latent process evolution (equation (1)) because the hypothesis to evaluate was an association between the $\epsilon 4$ allele and the decline of latent cognitive performance. We had no hypothesis regard-

ing a link with a specific psychometric measure. We found no association between the $\epsilon 4$ allele and the mean level of the common factor at age 65 years but found a strong association ($p = 0.0018$) with the change over time of the common factor: $\epsilon 4$ carriers have a steeper decline than $\epsilon 4$ noncarriers as shown in Figure 4a. The model including both the interactions $\text{apoE} \times t$ and $\text{apoE} \times t^2$ had exactly the same likelihood as the model including only $\text{apoE} \times t^2$. Thus we retained the latter.

Figure 2 displays the estimated beta transformations for the four tests with the 95% pointwise confidence interval computed using a bootstrap method. The four estimated transformations are very different: the curve is convex for the MMSE and the BVRT, concave for the DSSTW, and close to linear for the IST. Moreover, the BVRT and the MMSE scores cover, respectively, only 80% and 88% of the latent process range while the DSSTW covers around 95% and the IST covers almost the entire range.

These results suggest that the MMSE and the BVRT are not appropriate to identify small changes in cognition among subjects with a high cognitive level, because the maximum scores of these tests are reached for a value of the latent

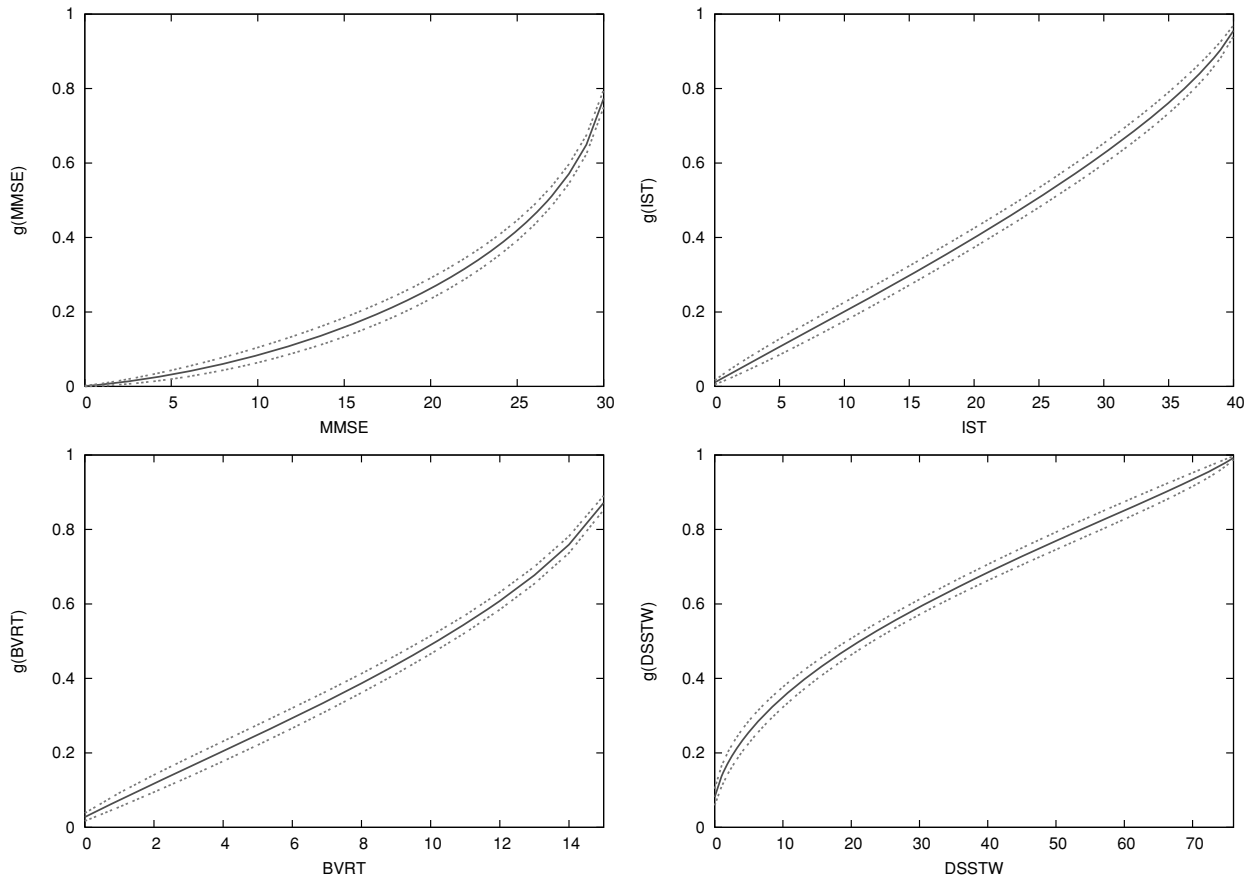


Figure 2. Estimated beta transformation for each test (solid line) and the 95% pointwise confidence interval (dashed line, obtained by bootstrap).

process lower than its maximum. These estimated curves highlight the ceiling effect of the two tests. More generally, the nonlinear shape of the MMSE reveals that a decline in the MMSE should be interpreted by taking the initial level into account: one point lost from a score above 25 represents a more substantial decrease of cognition (about 0.06) than one point lost from a score under 15 (about 0.01). For the DSSTW, the curve is close to linearity above a score of 10 but one point lost under a score of 10 represents a more substantial decrease of the latent cognition. Subjects with a latent cognition lower than 0.1 tend to score 0 on the DSSTW, probably because they do not even understand the instructions. In contrast, the IST appears to be useful to evaluate cognition in a heterogeneous population including high-level and impaired subjects, because it is close to linearity on almost the entire range of the latent cognition.

5.4 Assessment of the Fit

Figure 3 contains the normal quantile plots of the standardized marginal residuals defined in (9) for each of the four psychometric tests. The normality assumption of the residuals seems to be well satisfied for each of the four psychometric

tests. In contrast, when using a linear transformation, normal quantile plots showed a poor agreement with the normal assumption (results not displayed).

Figure 4b shows for each of the four tests the estimated marginal and conditional mean evolutions with age compared with the observed mean evolution and its 95% confidence limits; the sample size used to compute each mean is also given. The conditional estimated means, which include random effect estimates, are very close to the observed means for every test, showing a good fit of the model. However, the marginal estimated means, which include only fixed effects, are outside the 95% confidence interval of the observed means for the IST and BVRT at older ages and for the DSSTW in most cases. These differences may be explained by the rate of missing data, which is very low for the MMSE, higher for the IST and BVRT particularly among the oldest participants, and much higher for the DSSTW at all ages. Indeed, during the interview, the tests were always completed in the same order (MMSE, BVRT, IST, DSSTW) and recommendations were given to the interviewers to avoid missing data for the MMSE, because it is used for the screening of dementia. Hence almost all subjects completed the MMSE but subjects with a poor

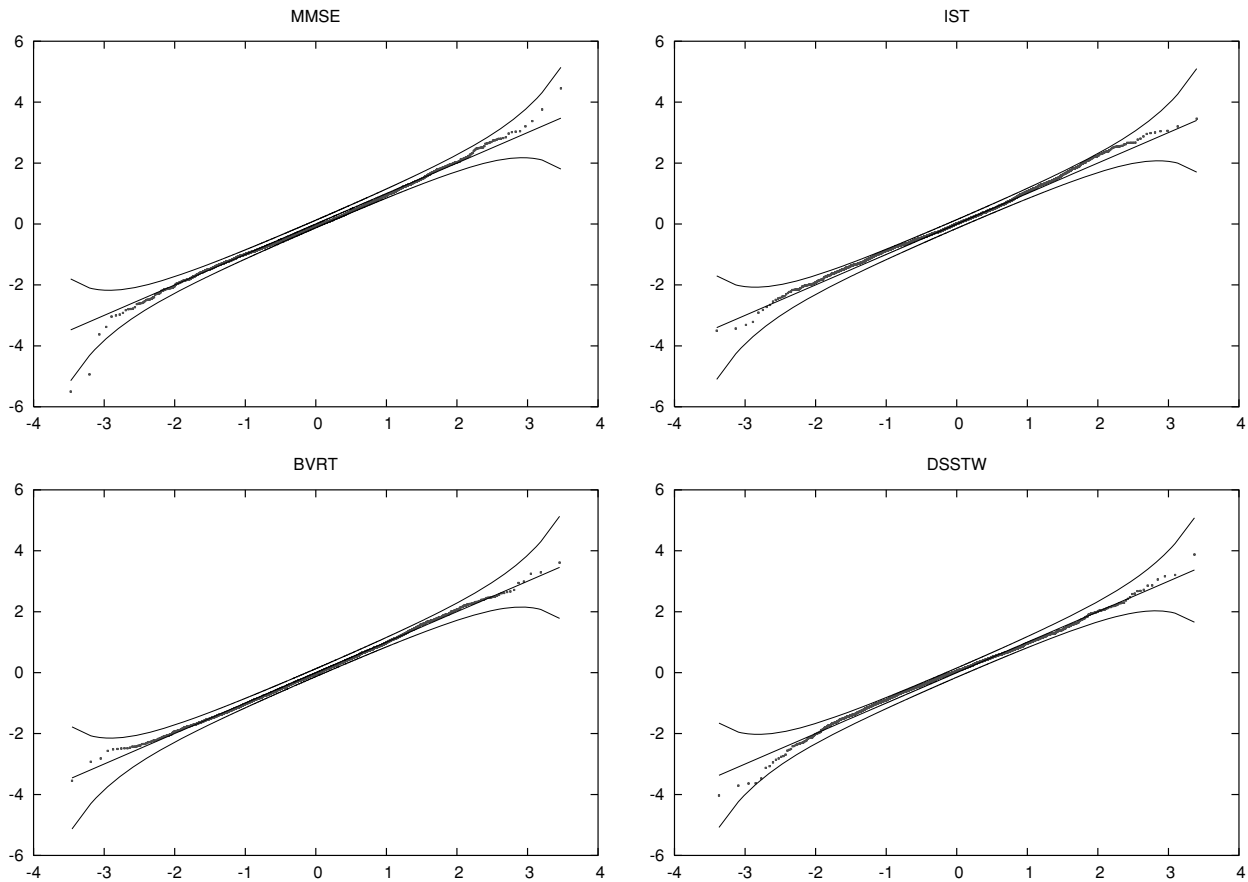


Figure 3. Normal quantile plot of the standardized marginal residuals for each test (solid lines = “y = x” reference line and 95% confidence interval).

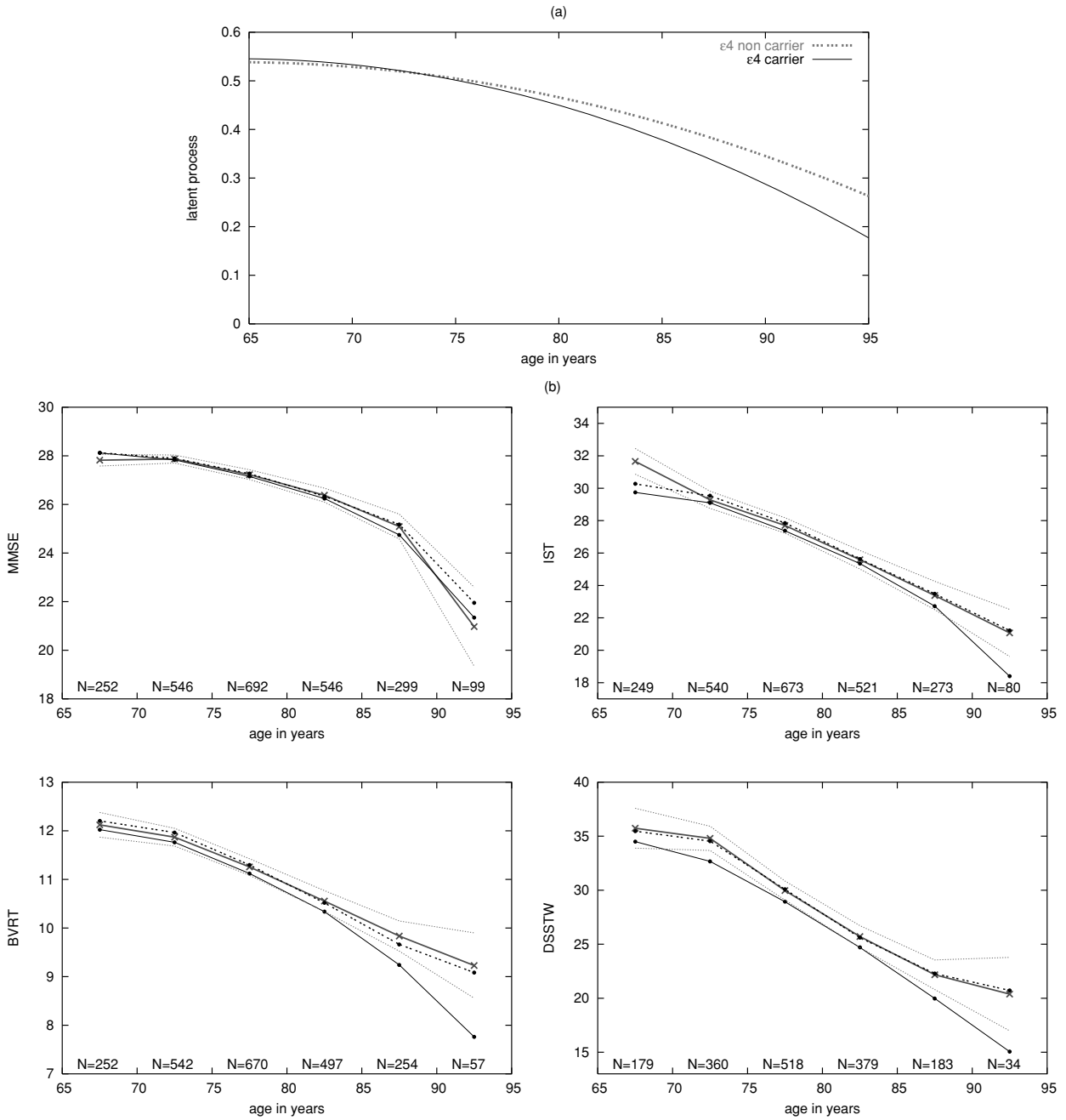


Figure 4. (a) Predicted mean evolution for the latent process for $\epsilon 4$ carriers and for $\epsilon 4$ noncarriers. (b) Estimated and observed mean evolutions for each test with the number of subjects used for the computation of each observed mean (solid line with crosses = observed mean evolution; solid line with dots = estimated marginal mean evolution; dashed line with dots = estimated subject-specific mean evolution; dashed line = 95% confidence interval of the observed mean).

cognitive level tended to refuse the other tests and particularly the DSSTW, which is more difficult. Missing data are thus associated with random effects. For instance, the mean of subject-specific deviations \hat{W}_{ijk} is -0.0037 for subjects aged

90 or more who completed the MMSE ($N = 99$), and 0.028, 0.042, and 0.056, respectively, for those who completed the IST ($N = 80$), BVRT ($N = 57$), and DSSTW ($N = 34$). The impact of missing data on conditional and marginal estimates

has previously been discussed by Molenberghs and Verbeke (2001).

5.5 Multivariate Model versus Univariate Models

For each test, the univariate model detected an association between apoE genotype and cognition with a larger p -value (p -value from the likelihood ratio test for apoE $\times t^2$ parameter: $p = 0.0043$ for MMSE, $p = 0.018$ for IST, $p = 0.020$ for BVRT, $p = 0.054$ for the DSSTW) than for the multivariate model ($p = 0.0018$). By using a multivariate model compared to four univariate models, we had a gain of power in assessing the association between apoE genotype and cognition. Moreover, note that interpretation of the association with the latent process and with each psychometric test is different.

The gain in efficiency can also be evaluated by comparing the AIC from the multivariate model and the AIC computed by pooling the likelihoods from the four univariate models with the total number of parameters in these four models. In our case, even if we added in the multivariate model the constraint that apoE had a common effect on the four tests, the AIC from the multivariate model was markedly better (39503.1 vs. 40203.8 for the four univariate models).

6. Discussion

We proposed a nonlinear model for multivariate longitudinal non-Gaussian quantitative outcomes when the outcomes are indirect measures of a common underlying continuous-time process. Such data are very frequent in psychometrics, but the methodology has many other potential areas of application, as, for instance, the study of the course of chronic illnesses evaluated by several biological markers.

In this work, psychometric tests are analyzed by considering their sum scores as quantitative variables. This is the most frequent way to consider psychometric tests in gerontology: the summary scores are used to evaluate cognitive level and risk of dementia (Hall et al., 2001; Sliwinski et al., 2003; Amieva et al., 2005). From a neuropsychological perspective, the alternative approach, which consists of analyzing item responses using SEM or item response model (Skrondal and Rabe-Hesketh, 2004, Chapter 3), could be useful if the objective was to understand the underlying components of the tests. This methodology is interesting when the tests consist of a limited number of items evaluating different cognitive domains (such as the MMSE) or exhibiting different levels of difficulty (such as the BVRT). On the contrary, this methodology would be difficult to apply to the IST score, which is the number of words cited by the subjects (except for considering the four subscores for each semantic category) and the DSSTW score, which is the count of symbols correctly assigned to a sequence of numbers.

Given that the summary scores are quantitative discrete variables, we could either consider them as continuous variables or as ordinal variables. However, threshold models for ordinal data require estimation of one threshold for each possible value of the scores, which would be very challenging for multivariate modeling of scores with so many different values. In our application including four tests, this would have implied estimation of more than 150 additional parameters. Thus, we decided to analyze the scores as continuous variables

and to use nonlinear transformations depending on a limited number of parameters as link functions between the Gaussian latent process and the observed outcomes. Various link functions have been considered, but we found that the beta CDF was flexible enough with only two parameters. With many fewer parameters than threshold models, the estimated curves provide interesting information on the relationship between evolution of the latent cognitive level and evolution of the observed scores. Moreover, goodness-of-fit analyses show that the beta transforms of the four test scores fitted well a Gaussian distribution. Nevertheless, if necessary for other applications, it would be easy to include a different family of continuous transformation for each test. This model could also be extended to allow a mixture of continuous, binary, and ordinal outcomes with few categories as in Dunson (2000), Dunson (2003), or Rabe-Hesketh et al. (2004).

Another asset of this model is the way of accounting for covariate effects. Using fixed contrasts, we were able to distinguish between the association with the latent process and the differential association with the various psychometric tests. Being able to compare covariate effects over the tests is also an advantage of multivariate modeling. With a large number of tests it may also be possible and advantageous to have the effects of the covariates on the tests be random rather than fixed.

We assumed the data to be missing at random and thus ignorable using a maximum likelihood approach. Even if we have shown that missing data were associated with random effects, this does not preclude missing data from being ignorable. Indeed, if missing data at the last three tests depend on the observed MMSE score or on the observed evolution of the MMSE, it induces a dependency on the random effects, but the missing data are ignorable because missing values may be predicted using observed data. This is an advantage of multivariate modeling, in that by using more observed information it is more robust to missing data. Nevertheless, as it is not excluded that missing data are informative, it could be useful to jointly model time to dropout using a shared random effect model as in Roy and Lin (2002). However, this would increase complexity of the estimation process and estimates would depend on uncheckable parametric assumptions. Another useful extension would be to jointly model dementia, defining dementia diagnosis as the time at which the latent process first reaches an estimated threshold (Hashemi, Jacqmin-Gadda, and Commenges, 2003).

ACKNOWLEDGEMENTS

The authors thank Jean-François Dartigues, Luc Letenneur, and the PAQUID research team for providing the data.

REFERENCES

- Amieva, H., Jacqmin-Gadda, H., Orgogozo, J., LeCarret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C., and Dartigues, J. (2005). The 9 year cognitive decline before dementia of the Alzheimer type: A prospective population-based study. *Brain* **128**, 1093–1101.
- Arminger, G. and Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler

- and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271–300.
- Dunson, D. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* **62**, 355–366.
- Dunson, D. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association* **98**, 555–563.
- Fabrigoule, C., Rouch, I., Taberly, A., Letenneur, L., Commenges, D., Mazaux, J.-M., Orgogozo, J.-M., and Dartigues, J.-F. (1998). Cognitive process in preclinical phase of dementia. *Brain* **121**, 135–141.
- Farrer, L., Cupples, L., Haines, J., Hyman, B., Kukull, W., Mayeux, R., Myers, R., Pericak-Vance, M., Risch, N., and 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**, 1349–1356.
- Gray, S. M. and Brookmeyer, R. (1998). Estimating a treatment effect from multidimensional longitudinal data. *Biometrics* **54**, 976–988.
- Hall, C., Ying, J., Kuo, L., Sliwinski, M., Buschke, H., Katz, M., and Lipton, R. (2001). Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Statistics in Medicine* **20**, 3695–3714.
- Harvey, D., Beckett, L., and Mungas, D. (2003). Multivariate modeling of two associated cognitive outcomes in a longitudinal study. *Journal of Alzheimer's Disease* **5**, 357–365.
- Hashemi, R., Jacqmin-Gadda, H., and Commenges, D. (2003). A latent process model for joint modeling of events and marker. *Lifetime Data Analysis* **9**, 331–343.
- Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., and Dartigues, J.-F. (1997). A 5-year longitudinal study of the Mini Mental State Examination in normal aging. *American Journal of Epidemiology* **145**, 498–506.
- Jöreskog, K. and Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In *Advanced Structural Equation Modeling: Issues and Techniques*, G. A. Marcoulides and R. E. Schumacker (eds), 57–88. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Volume 1. New York: MacMillan Publishing.
- Lee, S. and Song, X. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology* **57**, 131–150.
- Letenneur, L., Commenges, D., Dartigues, J.-F., and Barberger-Gateau, P. (1994). Incidence of dementia and Alzheimer's disease in the elderly community residents of south-western France. *International Journal of Epidemiology* **23**, 577–590.
- Longford, N. and Muthén, B. (1992). Factor analysis for clustered observations. *Psychometrika* **57**, 581–597.
- Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* **11**, 431–441.
- Molenberghs, G. and Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling* **1**, 235–269.
- Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika* **29**, 81–117.
- Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika* **69**, 167–190.
- Roy, J. and Lin, X. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* **56**, 1047–1054.
- Roy, J. and Lin, X. (2002). Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: Changes in methadone treatment practices. *Journal of the American Statistical Association* **97**, 40–52.
- Salthouse, T., Hancock, H., Meinz, E., and Hambrick, D. (1996). Interrelations of age, visual acuity, and cognitive functioning. *Journal of Gerontology: Psychological Sciences* **51B**, 317–330.
- Sammel, M. and Ryan, L. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 650–663.
- Sánchez, B. N., Budtz-Jorgensen, E., Ryan, L. M., and Hu, H. (2005). Structural equation models: A review with applications to environmental epidemiology. *Journal of the American Statistical Association* **100**, 1443–1455.
- Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modelling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, Florida: Chapman & Hall/CRC.
- Sliwinski, M., Hofer, S., and Hall, C. (2003). Correlated and coupled cognitive change in older adults with or without preclinical dementia. *Psychology and Aging* **18**, 672–683.
- Song, X. and Lee, S. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **57**, 29–52.
- Wall, M. and Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association* **95**, 929–940.
- Winnock, M., Letenneur, L., Jacqmin-Gadda, H., Dallongeville, J., Amouyel, P., and Dartigues, J. (2002). Longitudinal analysis of the effect of apolipoprotein E $\epsilon 4$ and education on cognitive performance in elderly subjects: The PAQUID study. *Journal of Neurology, Neurosurgery and Psychiatry* **72**, 794–797.
- Yalcin, I. and Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science* **16**, 275–294.

Received January 2005. Revised December 2005.

Accepted January 2006.

APPENDIX

Details on the Transformations

For subject i and test k , we give the expressions of the function and its Jacobian for the linear transformation, the combination of a linear and a logit transformation, and the Weibull CDF.

The linear transformation is defined for $y \in \mathbb{R}$, $\eta_{1k} \in \mathbb{R}$, and $\eta_{2k} \in \mathbb{R}^*$ as

$$g_k(y; \eta_{1k}, \eta_{2k}) = \frac{y - \eta_{1k}}{\eta_{2k}}$$

$$J(y_i; \theta) = \prod_{k=1}^K \frac{1}{\eta_{2k}}$$

The combination of a linear and a logit transformation is defined for $y \in (0, 1)$, $\eta_{1k} \in (0, 1)$, and $\eta_{2k} \in (0, 1)$ as

$$g_k(y; \eta_{1k}, \eta_{2k}) = \frac{\ln\left(\frac{y}{1-y}\right) - \ln\left(\frac{\eta_{1k}}{1-\eta_{1k}}\right)}{\ln\left(\frac{\eta_{2k}}{1-\eta_{2k}}\right)}$$

$$J(y_i; \theta) = \prod_{k=1}^K \prod_{j=1}^{n_{ik}} \frac{1}{\ln\left(\frac{\eta_{2k}}{1-\eta_{2k}}\right) y_{ijk} (1-y_{ijk})}$$

The Weibull CDF is defined for $y \in (0, \infty)$, $\eta_{1k} \in (0, \infty)$, and $\eta_{2k} \in (0, \infty)$ as

$$g(y; \eta_{1k}, \eta_{2k}) = 1 - \exp\left(-\left(\frac{y}{\eta_{1k}}\right)^{\eta_{2k}}\right)$$

$$J(y_i; \theta) = \prod_{k=1}^K \prod_{j=1}^{n_{ik}} \frac{\eta_{2k} y^{\eta_{2k}-1}}{\eta_{1k}^{\eta_{2k}}} \exp\left(-\left(\frac{y}{\eta_{1k}}\right)^{\eta_{2k}}\right)$$

3.2 Etudes de simulation

Dans cet article, nous n'avons pas présenté d'étude de simulation pour évaluer les performances du modèle. L'adéquation du modèle aux données a été vérifiée par deux méthodes graphiques : l'évaluation de la distribution des résidus suivant une loi normale et la comparaison des évolutions prédite et observée pour chaque test. Ces deux méthodes ont mis en évidence une très bonne adéquation du modèle aux données. Néanmoins, le modèle proposé étant paramétrique et assez complexe, nous dédions cette section à deux études de simulation ayant pour objectif d'évaluer la qualité des estimateurs du maximum de vraisemblance des paramètres du modèle et notamment d'évaluer si la fonction de répartition Beta utilisée pour lier les tests psychométriques au processus latent est une transformation satisfaisante. La première étude de simulation consiste à comparer deux paramétrisations de la fonction de répartition Beta et la deuxième étude consiste à évaluer si le choix d'une fonction de répartition Beta n'entraîne pas de biais important dans l'estimation des paramètres du facteur commun. Avant de présenter les résultats des deux études, nous exposons la méthodologie employée pour simuler les données.

3.2.1 Méthodologie des études de simulation

Les deux études de simulation sont construites de la même manière. Il s'agit de simuler les scores aux K tests psychométriques dans un échantillon de N sujets. Les scores Y_{ijk} sont simulés à partir des temps de mesure t_{ijk} , i pour le sujet, j pour la mesure et k pour le test, de la manière suivante :

Pour chaque sujet i ,

1. on simule les effets aléatoires $u_i = (u_{0i}, u_{1i}, u_{2i})^T$ à partir d'une loi normale $\mathcal{N}(\mu, D)$;
2. pour chaque test k , on simule l'intercept aléatoire spécifique α_{ik} à partir de la loi $\mathcal{N}(0, \sigma_{\alpha_k}^2)$;
3. pour chaque mesure j de chaque test k , on définit ensuite la valeur du prédicteur

linéaire : $\tilde{Y}_{ijk} = \Lambda_i(t_{ijk}) + \alpha_{ik} + \epsilon_{ijk} = u_{0i} + u_{1i}t_{ijk} + u_{2i}t_{ijk}^2 + \alpha_{ik} + \epsilon_{ijk}$, où ϵ_{ijk} est simulé à partir de la loi $\mathcal{N}(0, \sigma_{\epsilon_k}^2)$;

4. tant qu'au moins un des \tilde{Y}_{ijk} n'est pas dans l'intervalle $[0, 1]$, les étapes 1 à 3 sont réitérées ;
5. si toutes les valeurs des \tilde{Y}_{ijk} sont dans l'intervalle $[0, 1]$, on calcule pour chaque triplet (i, j, k) le score au test $Y_{ijk} = g_k^{-1}(\tilde{Y}_{ijk}; \eta_{1k}, \eta_{2k})$, où $g_k(\cdot; \eta_{1k}, \eta_{2k})$ est la fonction de répartition Beta avec les paramètres canoniques $(\eta_{1k}; \eta_{2k})$.

Ce schéma de simulation met en évidence le point faible de notre approche. La fonction de répartition Beta est définie de $[0, 1]$ dans $[0, 1]$ alors que le processus latent Gaussien est défini dans \mathbb{R} . La deuxième étude de simulation a pour objectif d'évaluer le biais engendré par la troncature nécessaire des données Gaussiennes à $[0; 1]$.

3.2.2 Première étude : paramétrisation de la fonction Beta

La fonction de répartition Beta (donnée en équation (3) de l'article) est définie en fonction de deux paramètres η_1 et η_2 définis positifs. Pour estimer ces paramètres avec l'algorithme de Marquardt, nous avons considéré deux paramétrisations différentes :

- P1 : des transformations directes des paramètres classiques η_1 ($\eta_1 > 0$) et η_2 ($\eta_2 > 0$) : ν_1 et ν_2 tels que $\eta_1 = \nu_1^2$ et $\eta_2 = \nu_2^2$
- P2 : des transformations des paramètres de moyenne m ($0 < m < 1$) et variance v ($v > 0$) de la fonction Beta : κ_1 et κ_2 tels que $m = \frac{e^{\kappa_1}}{1 + e^{\kappa_1}}$ et $v = \exp(\kappa_2)$.

Les deux paramétrisations sont équivalentes, les paramètres η_1 et η_2 étant reliés aux paramètres de moyenne et de variance de la manière suivante : $m = \frac{\eta_1}{\eta_1 + \eta_2}$ et $v = \frac{\eta_1 \eta_2}{(\eta_1 + \eta_2)^2 (\eta_1 + \eta_2 + 1)}$. Pour évaluer et comparer les deux paramétrisations, nous avons simulé 100 échantillons de 100 sujets. Nous n'avons pas utilisé d'échantillon similaire aux données de la cohorte PAQUID mais avons choisi des valeurs des pa-

ramètres de façon à réduire la probabilité que \tilde{Y}_{ijk} soit en dehors de $[0, 1]$, l'objectif de cette première étude de simulation n'étant que le choix de la paramétrisation. Chaque sujet a 11 mesures répétées pour chacun des deux marqueurs considérés, les temps de mesure étant les mêmes pour tous les sujets : 0, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7 et 3.0. Les \tilde{Y}_{ijk} sont simulés suivant un modèle linéaire mixte incluant une intercept aléatoire de loi $\mathcal{N}(\mu_0 = 0.6, \sigma_0^2 = 0.009)$ et une pente aléatoire de loi $\mathcal{N}(\mu_1 = -0.08, \sigma_1^2 = 0.001)$, l'intercept et la pente aléatoires sont indépendantes entre elles. Les transformations Beta sont paramétrées par $(\eta_1^{(1)} = 0.49; \eta_2^{(1)} = 0.81)$ et $(\eta_1^{(2)} = 0.81; \eta_2^{(2)} = 0.64)$ et les erreurs sont indépendantes, d'écart-type $\sigma_\epsilon^{(k)} = 0.03$ pour chaque marqueur ($k = 1, 2$). Nous n'avons pas inclus d'intercept aléatoire spécifique à chaque test. Les tableaux 3.1(A) et 3.1(B) donnent les estimations moyennes, les biais relatifs, les écart-types empiriques, les écart-types asymptotiques moyens et les taux de couverture des 100 simulations pour les deux paramétrisations P1 et P2.

En utilisant la première paramétrisation, les écart-types asymptotiques moyens sont nettement plus grand que les écart-types empiriques alors qu'en utilisant la deuxième paramétrisation, les deux grandeurs sont équivalentes. Ce problème de surestimation de l'écart-type avec la première paramétrisation semble venir d'un mauvais conditionnement de la matrice Hessienne. Le conditionnement de la matrice Hessienne est défini par le rapport entre sa plus grande et sa plus petite valeur propre. Pour la première paramétrisation, la moyenne des conditionnements sur les 100 simulations est de 85219.5 (minimum de 11167.7 et maximum de 1074156.0) alors qu'elle est de 26531.0 pour la deuxième paramétrisation (minimum de 10082.3 et maximum de 65705.3). La matrice Hessienne a tendance à être très mal conditionnée lorsqu'on utilise la paramétrisation classique ce qui engendre une mauvaise estimation de l'écart-type asymptotique. Cette première étude de simulation montre donc que la fonction de répartition Beta doit être paramétrée en utilisant sa moyenne et sa variance plutôt qu'en fonction de ses paramètres canoniques. Dans la suite de ce travail, nous utiliserons donc la deuxième paramétrisation P2 de la fonction de

TAB. 3.1 – Estimation moyenne, biais relatif, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture des 100 simulations pour (A) la paramétrisation P1 et (B) la paramétrisation P2.

(A) Paramétrisation P1

paramètre	valeur simulée	estimation moyenne	biais relatif	ET empirique	ET asymptotique moyen	taux de couverture
$\nu_1^{(1)}$	0.700	0.693	-0.010	0.0226	0.0472	1.00
$\nu_2^{(1)}$	0.900	0.889	-0.012	0.0486	0.106	1.00
$\nu_1^{(2)}$	0.900	0.889	-0.013	0.0348	0.0736	1.00
$\nu_2^{(2)}$	0.800	0.791	-0.011	0.0387	0.0829	1.00
μ_0	0.600	0.599	-0.002	0.0234	0.0358	0.99
μ_1	-0.080	-0.079	-0.014	0.0488	0.0941	1.00
σ_0	0.095	0.092	-0.035	0.0071	0.0120	1.00
σ_1	0.032	0.031	-0.031	0.0025	0.0042	1.00
$\sigma_\epsilon^{(1)}$	0.030	0.030	-0.015	0.0015	0.0034	1.00
$\sigma_\epsilon^{(2)}$	0.030	0.030	-0.013	0.0017	0.0033	0.99

(B) Paramétrisation P2

paramètre	valeur simulée	estimation moyenne	biais relatif	ET empirique	ET asymptotique moyen	taux de couverture
$\kappa_1^{(1)}$	-0.251	-0.244	-0.027	0.0596	0.0582	0.94
$\kappa_2^{(1)}$	-2.358	-2.335	-0.010	0.0669	0.0631	0.91
$\kappa_1^{(2)}$	0.118	0.119	0.009	0.0517	0.0512	0.96
$\kappa_2^{(2)}$	-2.383	-2.361	-0.009	0.0598	0.0579	0.92
μ_0	0.600	0.598	-0.004	0.0235	0.0217	0.97
μ_1	-0.080	-0.078	-0.023	0.0055	0.0053	0.96
σ_0	0.095	0.091	-0.039	0.0089	0.0082	0.90
σ_1	0.032	0.030	-0.041	0.0026	0.0028	0.93
$\sigma_\epsilon^{(1)}$	0.030	0.029	-0.021	0.0019	0.0017	0.95
$\sigma_\epsilon^{(2)}$	0.030	0.029	-0.021	0.0019	0.0017	0.93

répartition Beta. De plus, cette simulation a permis de montrer que dans le cas où l'échantillon n'inclut qu'un nombre négligeable de valeurs hors de l'intervalle $[0, 1]$ pour le processus latent, le modèle à processus latent donne de très bons résultats en termes de biais et de taux de couverture.

3.2.3 Deuxième étude : situation proche du jeu de données PAQUID

La deuxième étude de simulation a pour objectif d'évaluer si notre modélisation par la fonction de répartition Beta n'engendre pas des biais importants dans l'estimation des paramètres à cause de la troncature des données Gaussiennes à $[0; 1]$. Pour évaluer ce biais, nous avons simulé des données similaires à celles de l'échantillon de 563 sujets issu de la cohorte PAQUID et analysé dans l'article méthodologique. Ainsi, nous pouvons évaluer l'importance du biais pour les données réelles utilisées dans l'application.

Les échantillons simulés sont constitués des 563 sujets de l'échantillon de PAQUID et incluent tous les temps de mesure observés. A partir de ces temps de mesure et des valeurs des paramètres estimées sur l'échantillon de PAQUID, nous avons généré les mesures aux quatre marqueurs longitudinaux. Le modèle pour le facteur commun incluait une intercept, une pente et un terme quadratique aléatoires corrélés. Par rapport à l'illustration présentée dans l'article, ce modèle n'incluait pas de processus Brownien, de variables explicatives et d'intercept aléatoire spécifique à chaque test.

Le tableau 3.2 résume les résultats des 200 simulations effectuées. A chaque simulation, les étapes 1 à 3 ont été relancées pour 11.6 sujets en moyenne à cause d'au moins une valeur inférieure à 0 et pour 2.3 sujets en moyenne à cause d'au moins une valeur supérieure à 1. Pour cet échantillon issu de PAQUID, un nombre non négligeable de valeurs du processus latent sont donc inférieures à 0, généralement

TAB. 3.2 – Estimation moyenne, biais relatif, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture des 200 simulations de l'échantillon réel de 563 sujets

paramètre	valeur simulée	estimation moyenne	biais relatif	ET empirique	ET asymptotique moyen	taux de couverture
$\kappa_1^{(1)}$	1.232	1.150	-0.067	0.060	0.061	0.78
$\kappa_2^{(1)}$	-2.789	-2.666	-0.044	0.074	0.078	0.68
$\kappa_1^{(2)}$	0.331	0.289	-0.129	0.041	0.041	0.87
$\kappa_2^{(2)}$	-2.375	-2.306	-0.029	0.041	0.043	0.63
$\kappa_1^{(3)}$	0.459	0.407	-0.113	0.046	0.045	0.82
$\kappa_2^{(3)}$	-2.334	-2.260	-0.032	0.042	0.044	0.63
$\kappa_1^{(4)}$	-0.539	-0.549	0.018	0.035	0.036	0.96
$\kappa_2^{(4)}$	-2.280	-2.232	-0.021	0.027	0.032	0.72
μ_0	0.623	0.625	0.0036	0.011	0.011	0.97
μ_1	-0.010	-0.012	0.130	0.0094	0.0092	0.95
μ_2	-0.289	-0.261	-0.098	0.038	0.037	0.87
$chol(D)_{11}^*$	0.092	0.084	-0.083	0.0073	0.0067	0.77
$chol(D)_{12}^*$	-0.053	-0.047	-0.119	0.0159	0.0149	0.94
$chol(D)_{22}^*$	0.104	0.096	-0.081	0.0084	0.0082	0.85
$chol(D)_{13}^*$	0.159	0.131	-0.179	0.074	0.069	0.95
$chol(D)_{23}^*$	-0.401	-0.366	-0.099	0.044	0.041	0.87
$chol(D)_{33}^*$	0.07	0.187	-0.096	0.013	0.013	0.72
$\sigma_\epsilon^{(1)}$	0.084	0.079	-0.054	0.0029	0.0031	0.71
$\sigma_\epsilon^{(2)}$	0.084	0.079	-0.059	0.0032	0.0033	0.70
$\sigma_\epsilon^{(3)}$	0.097	0.091	-0.060	0.0033	0.0036	0.65
$\sigma_\epsilon^{(4)}$	0.035	0.033	-0.055	0.0013	0.0015	0.76

* $chol(D)$ représente la transformée de Cholesky de la matrice D de variance-covariance des effets aléatoires.

pour des temps d'observation élevés. Malgré cela, les estimations de l'évolution du facteur commun sont correctes, avec des biais relatifs assez faibles (au maximum 13% pour la pente). Les paramètres des fonction Beta et des erreurs de mesure sont aussi légèrement biaisés (biais inférieurs à 13%). On peut noter que, comme l'échantillon contient beaucoup d'information, les écart-types asymptotiques sont très petits et le léger biais sur les paramètres suffit à donner des taux de couverture assez mauvais. La même simulation réalisée sur un sous-échantillon de 100 sujets et présentée dans le tableau 3.3 a donné des résultats nettement meilleurs en terme de taux de couverture.

Les légers biais trouvés sur les estimateurs de certains paramètres restent négligeables d'un point de vue clinique. On peut donc conclure de ces études de simulations que le modèle à processus latent incluant une fonction de répartition Beta paramétrée par sa moyenne et sa variance donne de bons résultats.

3.3 Applications épidémiologiques

Cette approche à processus latent pour étudier conjointement plusieurs tests psychométriques a donné lieu à deux travaux appliqués dans l'étude du vieillissement cognitif. Le premier a consisté à étudier le lien entre les tests psychométriques et le facteur commun représentant le niveau cognitif latent. La deuxième application a permis de distinguer l'effet de variables explicatives sur le déclin cognitif de leur effet sur les tests psychométriques eux-mêmes. Nous présentons dans cette section ces deux travaux réalisés en collaboration avec l'équipe vieillissement de l'unité INSERM 593.

3.3.1 Propriétés métrologiques des tests psychométriques

Cette première application du modèle à processus latent a fait l'objet d'une publication dans l'*American Journal of Epidemiology*. L'article complet intitulé **Sensitivity of four psychometric tests to measure cognitive changes in brain**

TAB. 3.3 – Estimation moyenne, biais relatif, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture des 200 simulations de l'échantillon réel de 100 sujets

paramètre	valeur simulée	estimation moyenne	biais relatif	ET empirique	ET asymptotique moyen	taux de couverture
$\kappa_1^{(1)}$	1.232	1.127	-0.085	0.187	0.175	0.95
$\kappa_2^{(1)}$	-2.789	-2.646	-0.051	0.213	0.211	0.89
$\kappa_1^{(2)}$	0.331	0.27	-0.169	0.126	0.118	0.98
$\kappa_2^{(2)}$	-2.375	-2.293	-0.034	0.102	0.111	0.93
$\kappa_1^{(3)}$	0.459	0.393	-0.143	0.141	0.129	0.95
$\kappa_2^{(3)}$	-2.334	-2.249	-0.037	0.116	0.116	0.87
$\kappa_1^{(4)}$	-0.539	-0.551	0.022	0.102	0.097	0.98
$\kappa_2^{(4)}$	-2.280	-2.225	-0.024	0.071	0.079	0.92
μ_0	0.623	0.624	0.002	0.0300	0.0294	0.96
μ_1	-0.010	-0.105	0.008	0.022	0.021	0.95
μ_2	-0.289	-0.261	-0.096	0.092	0.090	0.92
$chol(D)_{11}^*$	0.092	0.082	-0.108	0.017	0.016	0.87
$chol(D)_{12}^*$	-0.053	-0.045	-0.159	0.038	0.033	0.91
$chol(D)_{22}^*$	0.104	0.087	-0.165	0.023	0.022	0.91
$chol(D)_{13}^*$	0.159	0.124	-0.22	0.181	0.158	0.87
$chol(D)_{23}^*$	-0.401	-0.317	-0.219	0.133	0.142	0.91
$chol(D)_{33}^*$	0.07	0.158	-0.237	0.055	0.076	0.82
$\sigma_\epsilon^{(1)}$	0.084	0.078	-0.073	0.0078	0.0082	0.91
$\sigma_\epsilon^{(2)}$	0.084	0.078	-0.076	0.0080	0.0085	0.93
$\sigma_\epsilon^{(3)}$	0.097	0.089	-0.077	0.0094	0.0096	0.89
$\sigma_\epsilon^{(4)}$	0.035	0.033	-0.070	0.0039	0.0038	0.87

* $chol(D)$ représente la transformée de Cholesky de la matrice D de variance-covariance des effets aléatoires.

aging population-based studies est présenté en annexe. Dans ce paragraphe, nous décrivons l'objectif de l'application et les principaux résultats obtenus.

Dans les études épidémiologiques, la cognition est évaluée par des tests psychométriques qui mesurent les changements cognitifs avec une sensibilité et une échelle différentes. Utiliser une batterie complète de tests psychométriques peut se révéler difficile à mettre en oeuvre car d'une part, les longues séances de tests sont coûteuses et d'autre part, les sujets souffrant d'un déficit cognitif ont plus tendance à refuser de répondre à de longues séances de tests que les sujets sans déficit cognitif. Il est donc souvent judicieux de sélectionner un faible nombre de tests ou même un seul test psychométrique pour étudier les changements cognitifs dans une étude en population. Un bon test psychométrique doit être court et facile à utiliser mais il doit surtout être capable de détecter n'importe quel changement cognitif dans l'intervalle de cognition observé dans la population.

L'objectif de ce travail a été de comparer la sensibilité aux changements cognitifs de quatre tests psychométriques largement utilisés dans les études en population par le biais du modèle à processus latent. En effet, en estimant les transformations nonlinéaires qui relient les quatre tests psychométriques à leur facteur commun, nous sommes en mesure de comparer leurs propriétés métrologiques sur l'intervalle de cognition observé dans l'échantillon.

Nous avons sélectionné pour cela un échantillon de 2252 sujets à partir de la cohorte PAQUID. Les sujets inclus étaient non-déments au début du suivi et avaient au moins une mesure à chaque test durant les 12 ans de suivi. Les quatre tests utilisés étaient :

- le Mini Mental State Examination (MMSE) (Folstein et al., 1975) qui évalue plusieurs dimensions de la cognition. C'est le test le plus utilisé pour étudier les changements cognitifs en population générale ;
- le test de rétention visuelle de Benton (BVRT) (Benton, 1965) qui mesure la mémoire visuelle, la mémoire épisodique et la mémoire de travail ;
- le set test d'Isaacs tronqué à 15 secondes (IST15) (Isaacs et Kennie, 1973),

- un test de fluence verbale qui mesure la mémoire sémantique, les processus exécutifs, la mémoire de travail et l'accès au langage ;
- le test de substitution des symboles (DSST) (Wechsler, 1981) qui mesure le raisonnement logique, la mémoire à court terme et la praxie (capacité à assembler des éléments pour construire une figure).

A partir des estimations du modèle à processus latent incluant une fonction quadratique du temps pour l'évolution du facteur commun, nous avons mis en évidence plusieurs propriétés pour les quatre tests. Tout d'abord, le MMSE et le BVRT souffrent d'un effet plafond et ne sont donc pas des tests satisfaisants pour étudier les changements cognitifs parmi les sujets ayant un haut niveau cognitif. Cependant, ces deux tests ayant une bonne sensibilité au changement cognitif dans des niveaux assez bas de la cognition, ils peuvent être utilisés pour étudier la cognition dans des populations de sujets atteints de déficits cognitifs légers. Au contraire, le DSST a une bonne sensibilité aux changements cognitifs dans les niveaux élevés de la cognition et peut être utilisé pour évaluer des changements cognitifs chez des sujets ayant un haut niveau cognitif. Néanmoins, comme le DSST souffre d'un effet plancher sans doute dû au fait que les sujets ayant un déficit cognitif important ne peuvent pas comprendre les instructions compliquées du test, il ne peut pas être utilisé dans des populations de sujets avec un déficit cognitif sévère. Ces trois tests psychométriques sont donc appropriés pour évaluer les changements cognitifs dans des populations particulières, soit chez des sujets de haut niveau cognitif pour le DSST, soit chez des sujets atteints d'un déficit cognitif léger pour le BVRT et le MMSE. Contrairement à ces trois tests, l'IST15 a une bonne sensibilité sur tout l'intervalle de cognition et un point perdu à un niveau cognitif élevé représente la même perte cognitive qu'un point perdu à un niveau cognitif bas.

A travers cette application, nous avons mis en évidence que, parmi les quatre tests psychométriques considérés, l'IST était le test le plus approprié pour étudier les changements cognitifs dans des populations hétérogènes, incluant à la fois des

sujets avec un déficit cognitif sévère et des sujets avec un haut niveau cognitif. Plus généralement, en utilisant le modèle à processus latent et en comparant les transformations nonlinéaires estimées sur une batterie de tests, nous sommes en mesure d'obtenir des informations pour sélectionner le ou les meilleurs tests psychométriques suivant la nature de la population étudiée et l'objectif de l'étude.

3.3.2 Effet du sexe et du niveau d'études sur les mesures de déclin cognitif

La deuxième application du modèle mixte à processus latent a consisté à explorer l'effet de deux variables explicatives sur la dégradation cognitive. Ce travail a donné lieu à un article actuellement soumis intitulé **Impact of gender and education on cognitive decline with age : a longitudinal approach** et présenté en annexe.

Nous avons vu dans le paragraphe 1.1.3 que l'effet du sexe sur le déclin cognitif était encore discuté, l'association entre sexe et déclin cognitif n'étant pas systématiquement retrouvée dans les études de cohorte. Parallèlement à ces travaux, des études en neuropsychologie ont mis en évidence que les hommes et les femmes ne réussissaient pas de la même manière les tests psychométriques. Les femmes obtiennent de meilleurs résultats que les hommes dans les tests incluant une composante verbale alors que les hommes obtiennent de meilleurs résultats dans les tests évaluant la dimension visuospatiale (Reite et al., 1993 ; Wiederholt et al., 1993). On peut donc se demander si les résultats ambigus obtenus sur l'association entre sexe et déclin cognitif ne seraient pas uniquement liés à la mesure de cognition choisie pour évaluer les changements cognitifs. De même, plusieurs études ont mis en évidence que les sujets à haut niveau d'éducation avaient un niveau cognitif plus élevé que les sujets à faible niveau d'éducation, et ce, en étudiant différents tests psychométriques et en ajustant sur l'âge (Elias et al., 1997). Comme le niveau d'éducation peut agir sur la capacité à passer des tests, les sujets à faible niveau d'éducation ayant souvent plus de difficultés à gérer les conditions de test, on peut

se demander si l'écart de performances observé entre les sujets à haut et bas niveau d'éducation ne serait pas en partie lié à une capacité différentielle à passer les tests psychométriques.

D'une manière générale, il est souvent difficile de distinguer l'effet d'une variable explicative sur le processus de dégradation cognitive de son effet sur l'outil de mesure de cognition. Dans l'approche à processus latent, nous sommes en mesure de distinguer les effets globaux de variables explicatives sur le facteur commun considéré comme l'indicateur global de cognition de leurs effets sur chaque test psychométrique. Ceci nous permet donc d'évaluer en particulier dans quelle mesure le sexe et le niveau d'études sont associés au déclin cognitif général et à chaque test psychométrique.

A partir de la cohorte PAQUID, nous avons sélectionné tous les sujets ayant au moins une mesure aux quatre tests psychométriques (MMSE, IST15, BVRT et DSST), quel que soit leur processus de vieillissement cognitif, et notamment quel que soit leur statut de démence. L'échantillon était constitué de 2228 sujets. Le niveau d'éducation spécifié en deux classes (a obtenu le certificat d'études primaires ou non) et le sexe ont été introduits dans le modèle à processus latent à la fois dans le modèle mixte pour le facteur commun et dans les modèles d'observation de chaque test. Un ajustement sur les facteurs vasculaires a aussi été pris en compte pour éviter des problèmes de confusion. Le modèle final incluait :

- dans le modèle structurel : une fonction quadratique de l'âge (AGE), un mouvement Brownien ($w_i(t)$) ainsi que les effets du sexe (S), du niveau d'éducation (CEP) et des facteurs vasculaires ($(VASC_l)_{l=1, p_{VASC}}$) à la fois sur le niveau initial (niveau à 65 ans) du facteur commun et en interaction avec l'âge :

$$\begin{aligned} \Lambda(t) = & u_{0i} + u_{1i}AGE + u_{2i}AGE^2 + S \times (\beta_0 + \beta_1AGE + \beta_2AGE^2) \\ & + CEP \times (\beta_3 + \beta_4AGE + \beta_5AGE^2) \\ & + \sum_{l=1}^{p_{VASC}} VASC_l \times (\beta_{6+3(l-1)} + \beta_{7+3(l-1)}AGE + \beta_{8+3(l-1)}AGE^2) \quad (3.1) \end{aligned}$$

- dans le modèle d'observation : des intercepts et pentes spécifiques à chaque test et des contrastes sur le sexe et le niveau d'études ainsi que l'intercept aléatoire spécifique à chaque test (α_{ik}) et l'erreur de mesure (ϵ_{ijk}) :

$$h_k(y_{ijk}, \eta_k) = \Lambda(t_{ijk}) + \gamma_{0k} + \gamma_{1k} \text{AGE} + \gamma_{2k} \text{AGE}^2 + S \times (\gamma_{3k} + \gamma_{4k} \text{AGE} + \gamma_{5k} \text{AGE}^2) \\ + \text{CEP} \times (\gamma_{6k} + \gamma_{7k} \text{age} + \gamma_{8k} \text{AGE}^2) + \alpha_{ik} + \epsilon_{ijk} \quad (3.2)$$

avec $\sum_{k=1}^K \gamma_{jk} = 0, \forall j$

Par rapport au modèle présenté dans l'article méthodologique en section 3.1, nous avons ajouté l'intercept et la pente spécifiques à chaque test dans le modèle d'observation. En effet, lorsqu'on ne les introduit pas dans le modèle, on fait l'hypothèse qu'il n'y a pas de différence de niveau ou de pente entre les tests dans la classe de référence de la variable explicative alors que dans les autres classes, les différences sont modélisées par les contrastes. Cette hypothèse est bien évidemment trop forte et peut induire des interprétations erronées lors de la comparaison d'effet des variables explicatives sur les tests. Nous avons donc ajouté au modèle une intercept et pente spécifiques à chaque test.

Grâce à la modélisation conjointe des quatre tests, nous avons confirmé que le niveau d'éducation était associé au déclin cognitif global : les sujets ayant un haut niveau d'éducation ont un niveau cognitif moyen plus élevé mais ont aussi un déclin cognitif dans le grand âge un peu plus important que les sujets de faible niveau d'éducation. Nous avons aussi trouvé des différences entre les tests. Notamment, l'écart entre le score d'un sujet à haut niveau d'études et le score d'un sujet à bas niveau d'études diminue avec l'âge pour l'IST15 et le DSST et reste le même pour le MMSE et le BVRT. Ce résultat peut être expliqué par la présence d'une composante de vitesse dans l'IST15 et le DSST. En effet, il a été montré que la plupart des changements cognitifs liés à l'âge étaient dus à une baisse de la vitesse d'exécution (Salthouse, 1996) et que la baisse de vitesse d'exécution était un des premiers déficits observés dans le vieillissement. Au vu de ces résultats, on peut se

demander si le déclin cognitif plus important dans le grand âge chez les sujets à haut niveau d'éducation que chez les sujets à faible niveau d'éducation viendrait du fait que la perte de vitesse d'exécution liée à l'âge arriverait plus tardivement chez les sujets à haut niveau d'éducation que chez les sujets à bas niveau d'éducation. Concernant le sexe, nous avons montré qu'après prise en compte des facteurs vasculaires et notamment du statut tabagique, les femmes avaient un déclin cognitif global avec l'âge légèrement plus important que celui des hommes. Comme nous étudions le processus de vieillissement cognitif quel qu'il soit, il est possible que ce résultat soit expliqué par la plus grande incidence de démence chez les femmes que chez les hommes dans les âges avancés (Letenneur et al., 1999). Dans cette analyse conjointe, nous avons aussi confirmé que, quel que soit le niveau cognitif global, les femmes réussissent mieux l'IST (le test incluant une composante verbale) que les hommes à n'importe quel âge alors que les hommes réussissent mieux le BVRT (évaluant les capacités visuospatiales) que les femmes à n'importe quel âge.

3.4 Discussion

Nous avons présenté dans ce chapitre un modèle paramétrique pour étudier l'évolution conjointe de plusieurs marqueurs quantitatifs non Gaussiens. Ce modèle peut être vu comme une extension du travail de Roy et Lin (2000) à deux niveaux : nous prenons en compte des marqueurs quantitatifs non Gaussiens par le biais de transformations nonlinéaires estimées, et nous incluons dans le modèle mixte un processus Brownien pour assouplir la forme paramétrique de l'évolution moyenne en fonction du temps.

L'originalité principale de ce travail réside dans les transformations nonlinéaires paramétrées qui permettent de modéliser la relation entre les marqueurs et le processus latent. Dans l'article, nous avons comparé plusieurs familles de fonctions nonlinéaires paramétriques et avons conclu que les fonctions de répartition Beta

constituaient un bon compromis entre une grande flexibilité des formes estimées et un faible nombre de paramètres à estimer par marqueur. Le faible nombre de paramètres nécessaires pour estimer les transformations est déterminant dans le choix de la famille de transformations. En effet, si chaque transformation nécessite beaucoup de paramètres, il sera numériquement difficile d'inclure plusieurs marqueurs dans le modèle alors qu'un intérêt de cette approche est justement de multiplier les marqueurs d'évolution pour d'une part les comparer entre eux et d'autre part augmenter l'information utilisée pour estimer l'effet de variables explicatives.

L'inconvénient majeur de cette approche basée sur les fonctions de répartition Beta est que la transformation Beta impose que le processus latent soit dans l'intervalle $[0, 1]$ alors qu'il peut théoriquement être à valeurs dans \mathbb{R} . Nous avons vu dans les simulations que ce problème engendrait des biais sur les estimations des effets fixes. Cela étant, ces biais restent légers d'un point de vue clinique. De manière générale, les deux études de simulation montrent que les estimateurs des paramètres du modèle sont assez bons.

Dans un travail non présenté ici, nous avons testé des transformations entre le processus latent et les tests psychométriques basées sur des splines. Ces transformations à valeurs dans \mathbb{R} ont l'avantage de pallier le problème rencontré avec les transformations Beta. Cependant, les transformations par splines estimées sur les données de PAQUID se sont révélées très proches de celles obtenues par les fonctions de répartition Beta et l'utilisation des splines a engendré des problèmes numériques importants liés au grand nombre de paramètres dans le modèle. En effet, les transformations basées sur des splines nécessitaient 7 paramètres de splines (pour 5 noeuds) et un paramètre de niveau par test alors que la fonction de répartition Beta ne nécessitait que 2 paramètres par test. Une autre méthode permettant de pallier le problème lié aux transformations Beta consiste à considérer les tests psychométriques comme des tests ordinaux et à utiliser des modèles à seuils pour lier chaque test psychométrique au processus latent. Cela a été réalisé dans un travail récent (Ganiayre et al., 2006) mais cette approche est aussi extrêmement coûteuse

numériquement. En pratique, il est presque impossible d'étudier plusieurs tests conjointement. En plus, chaque test est supposé ordinal alors que certains tests comme le DSST sont purement quantitatifs par construction. Il apparaît difficile de trouver une alternative aux transformations Beta offrant une grande flexibilité des transformations entre tests et processus latent et n'entraînant pas de problèmes numériques. Malgré les légers biais induits par l'utilisation de la transformation Beta, cette approche reste donc un compromis intéressant.

A travers les deux applications présentées dans ce chapitre, nous avons illustré les intérêts multiples du modèle à processus latent. Le modèle permet en effet de comparer les tests psychométriques par rapport au processus latent et notamment de mettre en exergue leurs propriétés métrologiques. Il permet aussi de différencier l'effet de variables explicatives sur l'évolution cognitive globale de leur effet sur les différents tests psychométriques. Enfin, lorsque l'objectif est d'étudier uniquement l'effet global d'une variable explicative, comme l'effet du traitement dans un essai thérapeutique, ce modèle conjoint a aussi l'avantage d'augmenter l'information utilisée et donc d'augmenter la puissance des analyses. Cela étant, nous avons fait l'hypothèse dans ces travaux que les données manquantes étaient ignorables alors qu'il a été montré que la sortie d'étude dans PAQUID était liée au niveau cognitif et donc informative (Jacqmin-Gadda et al., 1997a ; Jacqmin-Gadda et al., 1999). Plus généralement, nous avons fait l'hypothèse que le processus de vieillissement dans la population était homogène alors qu'il existe plusieurs sous-populations caractérisées par le processus de sortie d'étude ou bien le risque de démence. C'est pourquoi, dans la suite de ce travail, nous allons étendre le modèle à processus latent pour l'étude conjointe de plusieurs marqueurs à la prise en compte d'un événement clinique.

Chapitre 4

Modèle nonlinéaire à classes latentes pour l'analyse conjointe de plusieurs marqueurs quantitatifs et d'une variable binaire

L'objectif de ce travail était de développer un modèle conjoint pour explorer l'association entre l'évolution de plusieurs marqueurs longitudinaux quantitatifs qui mesurent la même quantité non-observée et la probabilité d'un événement. Le développement de ce modèle a été motivé par l'étude conjointe de l'évolution cognitive mesurée par plusieurs tests psychométriques et le risque de démence. Ce travail a fait l'objet d'une publication dans *Statistics in Medicine*.

Nous proposons un modèle nonlinéaire conjoint à classes latentes. Les profils latents d'évolution cognitive sont modélisés par une extension du modèle à processus latent présenté précédemment mais prenant en compte une structure hétérogène des données par un mélange de lois normales pour la distribution des effets aléatoires.

La probabilité de démence est modélisée par une régression logistique spécifique à chaque classe latente. L'intérêt de ce modèle est mis en évidence à travers une application sur les données de la cohorte PAQUID. Le modèle permet de décrire les profils latents d'évolution cognitive associés à la survenue d'une démence et d'évaluer l'association avec des variables explicatives qui peuvent agir à plusieurs niveaux : sur le déclin cognitif, sur la probabilité d'appartenir à une classe latente ou encore sur la probabilité d'événement. Le modèle permet aussi de développer un outil de diagnostic précoce de la démence utilisable quelle que soit l'information disponible sur les tests psychométriques.

STATISTICS IN MEDICINE
Statist. Med. (in press)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2659



A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome

Cécile Proust-Lima^{1,2,*}, Luc Letenneur^{2,3} and Hélène Jacqmin-Gadda^{1,2}

¹*Institut National de la Santé et de la Recherche Médicale, Equipe de Biostatistique E0338, 33076 Bordeaux, France*

²*Université Victor Segalen, 33076 Bordeaux, France*

³*Institut National de la Santé et de la Recherche Médicale, Unité 593, 33076 Bordeaux, France*

SUMMARY

We consider a joint model for exploring association between several correlated longitudinal markers and a clinical event. A nonlinear growth mixture model exhibits the different latent classes of evolution of the latent quantity underlying the correlated longitudinal markers and a logistic regression models the probability of occurrence of the clinical event according to the latent classes. By introducing a flexible nonlinear transformation including parameters to be estimated between each marker and the latent process, the model also deals with non-Gaussian continuous markers. Through an application on cognitive ageing, the two advantages of the model are underlined: (1) the latent profiles of evolution associated with the clinical event are described including covariate effects in the longitudinal model but also in the probability of class membership and in the probability of occurrence of the event, and (2) a diagnostic and a prognostic tools are derived from the model for early detection of the clinical event using any available information about the longitudinal markers. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: cognitive ageing; dementia; early detection; growth mixture model; joint modelling

1. INTRODUCTION

In medical research, it is often of great interest to explore the association between evolution of longitudinal quantitative outcomes and a clinical event in order to both describe the evolution of markers during the course of a chronic disease, and to perform early detection of the disease using data on longitudinal evolution of the quantitative markers. In the screening of prostate cancer, modelling association between prostate-specific antigen (PSA) and prostate cancer showed that a

*Correspondence to: Cécile Proust-Lima, INSERM E0338, ISPED, Université de Bordeaux2, 146 rue Léo Saignat, 33076 Bordeaux cedex, France.

†E-mail: cecile.proust@isped.u-bordeaux2.fr

C. PROUST-LIMA, L. LETENNEUR AND H. JACQMIN-GADDA

significant increase of PSA could be predictive of a cancer some years later [1]. In the treatment of Alzheimer's disease, it is now admitted that the earlier the treatment administrated, the better the evolution of the disease. The joint study of longitudinal markers of cognition and diagnosis of dementia could hence be useful for initiating the treatment earlier.

For studying the association between evolution of quantitative outcomes and a clinical event either in a descriptive perspective or a prognostic perspective, two kinds of joint models have been proposed: shared random-effect models and latent class models. A shared random-effect model [2, 3] consists in modelling the repeated quantitative outcome with a mixed model and including the individual random coefficients as covariates in the model for the event. In cognitive ageing context, Jacqmin-Gadda *et al.* [4] proposed a shared random-effect model with random change points in the specific aim of estimating a change point in the cognitive decline toward dementia. In contrast, a latent class model [1, 5, 6] assumes that the population is made of various subpopulations with different longitudinal evolutions modelled by a latent class variable; the risk of event depends on the longitudinal outcome only through the latent classes. Latent class models have some advantages over shared random-effect models. Indeed, in the latter, the assumption that the random-effects come from a common Gaussian distribution is quite unrealistic when the population consists of several subpopulations (at least ill and not ill subjects). Moreover, by exhibiting profiles of evolution associated with the clinical event, latent class models are simpler to interpret compared with the shared random-effect models which estimate correlations between the event and the random-effects. In particular, latent class models are an attractive tool for clinicians because profiles of evolution are easily drawn for each class, the impact of covariates on the probability of each profile are evaluated and the probability of the event in each latent class is estimated. At last, shared-random-effect models are computationally more demanding due to the numerical integrations across the random-effects required in the computation of the likelihood. In some latent class models, integrals across the random-effects may have a closed form and the dependency between the two parts of the model requires only a sum across the values of the latent class variable.

Latent class models for joint analysis of a longitudinal outcome and an event have already been developed. Muthén and Shedden [6] studied the relation between the shape of heavy drinking trajectory in the 18–25 year age range and the probability of alcohol dependency at age 30. Lin *et al.* [5] extended the model to irregularly spaced longitudinal readings and unequal number of measures for each subject in the context of the prediction of prostate cancer according to PSA profiles. Finally, Lin *et al.* [1] proposed to model the risk of prostate cancer using a semi-parametric survival model instead of a logistic model.

In medical research, various markers are often collected repeatedly and it can be of interest to make use of the information from all the markers [7–9]. In many cases, the markers are highly correlated and may be viewed as various measures of a common underlying quantity which is modelled using a latent process. When the markers are continuous, they are modelled using the Gaussian assumption [8–10] although in many applications, they have a distribution far from a Gaussian distribution. For instance, in neuropsychology, the latent cognitive level is measured by several correlated psychometric tests which often exhibit a non-Gaussian continuous distribution. Recently, Ganiayre *et al.* [11] extended the latent process approach of Hashemi *et al.* [12] to model jointly ordinal psychometric tests and dementia assuming dementia was defined as the crossing of an estimated level by the latent process. However this approach does not take into account subpopulation structure and is numerically untractable for multiple longitudinal outcomes. Proust *et al.* [13] proposed a nonlinear model with a latent

A NONLINEAR LATENT CLASS MODEL

process to analyse multivariate and non-Gaussian longitudinal data by using flexible parametrized nonlinear transformations to model the relationship between the longitudinal outcomes and the latent process.

In this paper, our objective is to extend the joint model developed by Lin *et al.* [5] by using the nonlinear multivariate approach described in Reference [13] in order to propose a latent class model for the joint analysis of non-Gaussian multivariate longitudinal data and a binary outcome. This work was mostly motivated by the study of cognitive ageing. We aimed at: (i) describing profiles of cognitive evolution associated with dementia, and (ii) developing a tool for early diagnostic making use of information from repeated measures of several psychometric tests.

In the following section, we define the joint nonlinear latent class model. In Section 3, we compute the log-likelihood of the joint model and the posterior probabilities stemmed from the mixture model. We then propose a diagnostic and a prognostic tool for detecting the event. Section 4 focuses on an application to data from the French prospective cohort study PAQUID [14]. We conclude in Section 5 with a discussion.

2. METHODOLOGY

2.1. Nonlinear model for multivariate longitudinal data

Consider K correlated continuous outcomes. Each outcome k , $k = 1, \dots, K$ is measured on each subject i , $i = 1, \dots, N$ at n_{ik} occasions. For outcome k and subject i , the vector of measurements is $y_{ik} = (y_{i1k}, \dots, y_{ij_kk}, \dots, y_{in_{ik}k})$, j representing the occasion. The times of measurements denoted t_{ijk} may be different for each subject and each outcome.

Consider now a latent process $\Lambda_i = (\Lambda_i(t))_{t \geq 0}$ defined in continuous time for each subject i and representing the quantity underlying the K outcomes. We assume that the measurement y_{ijk} is related to the latent process at time t_{ijk} through a flexible monotone increasing transformation h_k depending on an outcome-specific vector of parameters η_k to be estimated. This measurement model is specified as follows:

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

where α_{ik} are random intercepts independently distributed according to a $N(0, \sigma_{\alpha_k}^2)$ distribution and ε_{ijk} are independent Gaussian errors with mean zero and variance $\sigma_{\varepsilon_k}^2$.

The random coefficient α_{ik} introduces variability between the markers conditionally on the value of the latent process. For example, in ageing context, α_{ik} takes into account the fact that for a same value of the latent cognition, subjects can score differently in cognitive domains associated with the psychometric tests.

For the transformation h_k , we chose the Beta cumulative distribution function (CDF) which depends only on two parameters $\eta_k = (\eta_{1k}, \eta_{2k})^T$ and offers a large flexibility in the shapes [13]. As the Beta CDF is defined in $[0, 1]$, a preliminary step consists in rescaling the tests to the unit interval.

2.2. Mixture model for the latent process

We assume the population consists of G subpopulations represented by G latent classes. Within each latent class g , $g = 1, G$, the latent process follows an homogeneous linear mixed model. By

C. PROUST-LIMA, L. LETENNEUR AND H. JACQMIN-GADDA

defining c_{ig} , the latent variable which equals one if subject i belongs to the latent class g and zero otherwise (the sum of c_{ig} over the G classes equals one), the latent process evolution for a subject i given that he belongs to latent class g is written as follows:

$$\Lambda_i(t)|_{c_{ig}=1} = Z(t)^T u_{ig} + X_{1i}(t)^T \beta + X_{2i}(t)^T \gamma_g, \quad t \geq 0 \tag{2}$$

where the $(p + 1)$ -vector $Z(t)^T = (1, t, \dots, t^p)$ is a time polynomial of degree p and the random-effect vector u_{ig} is distributed according to the class-specific Gaussian distribution $N(\mu_g, \omega_g^2 B)$ with $\omega_1 = 1$. The vectors $X_{1i}(t)$ and $X_{2i}(t)$ are, respectively, a q_1 -vector and a q_2 -vector of possibly time-dependent covariates; $X_{1i}(t)$ is associated with the vector of fixed effects β which is common for all the classes while $X_{2i}(t)$ is associated with the class-specific vector of fixed effects γ_g . For ensuring the model identifiability, a covariate cannot be included both in $X_{1i}(t)$ and $X_{2i}(t)$.

We included a class-specific variance-covariance matrix for the random effects $\omega_g^2 B$ with $\omega_1 = 1$. Thus, even if the structure of the variance-covariance matrix in each class $g, g > 1$, is the same as in the first class where the variance-covariance matrix is B , the global variability in class $g, g > 1$, can differ from the variability in the first class thanks to the proportional term ω_g^2 .

Note that $\Lambda_i(t)$ defined in equation (2) could also include a stochastic process such as a Brownian motion shared over the latent classes as proposed in Reference [13]. However models including both a mixture distribution and a Brownian motion can lead to numerical problems. We thus not included the Brownian motion in the mixture model formulation.

2.3. Model for the probability of belonging to class g

Each subject i has a probability π_{ig} of belonging to class g with $\sum_{g=1}^G \pi_{ig} = 1$. This class membership probability is modelled using a multinomial logit regression [6] including covariates:

$$\pi_{ig} = P(c_{ig} = 1 | X_{3i}) = \frac{e^{\xi_{0g} + X_{3i}^T \xi_{1g}}}{1 + \sum_{j=2}^G e^{\xi_{0j} + X_{3i}^T \xi_{1j}}} \quad \forall g = 1, G \tag{3}$$

where ξ_{0g} is the intercept for class g and ξ_{1g} is the q_3 -vector of class-specific parameters associated with the q_3 -vector of time-independent covariates X_{3i} . For identifiability, $\xi_{01} = 0$ and $\xi_{11} = 0$. Thus, each element of ξ_{1g} is the log odds-ratio for the probability of belonging to class g compared to class 1 for a unit increase of the corresponding covariate.

2.4. Model for the probability of the clinical event

Consider the binary variable D_i (D_i equals 1 if the event occurs for i and 0 otherwise). We model the probability of D_i conditionally on the latent class variables $(c_{ig})_{g=1,G}$. It means that D_i has a different probability to occur in each latent class g . The probability of D_i conditionally on g can also differ according to a q_4 -vector of covariates X_{4i} as follows:

$$P(D_i = 1 | X_{4i}, c_{ig} = 1) = \frac{e^{\delta_{0g} + X_{4i}^T \delta_{1g}}}{1 + e^{\delta_{0g} + X_{4i}^T \delta_{1g}}} \tag{4}$$

where δ_{0g} is the log-odds for the occurrence of D_i in the latent class g and δ_{1gl} , the l th element of δ_{1g} , is the log odds-ratio for the occurrence of D_i in latent class g for a unit increase of X_{4il} , the l th covariate of X_{4i} .

A NONLINEAR LATENT CLASS MODEL

2.5. Covariates interpretation

A richness of the model is that it distinguishes three possible ways of including covariates with possible overlap between time-independent covariates included in X_1 or X_2 and in X_3 and X_4 . However, as each way has a specific interpretation, inclusion of covariates must be done in accordance with the clinical hypothesis and the objective of the analysis. Covariates included in the latent process model (2) are assumed to impact only the mean evolution of the latent process either through a common effect over the classes (X_1) or through a class-specific effect (X_2). Covariates are included in the logistic model for the event (X_4) for evaluating their impact on the event risk after adjustment on the shape of the latent process evolution. In contrast, when a covariate is associated with the class membership probability (X_3), it means that the covariate has an effect on both the evolution of the latent process (on the mean and the random-effects variance) and on the event risk. Indeed, the two marginal distributions $f(\Lambda_i(t))$ and $P(D_i = 1)$ are given by

$$f(\Lambda_i(t)) = \sum_{g=1}^G \pi_{ig} f(\Lambda_i(t) | c_{ig} = 1) \quad (5)$$

$$P(D_i = 1) = \sum_{g=1}^G \pi_{ig} P(D_i = 1 | c_{ig} = 1) \quad (6)$$

3. ESTIMATION

3.1. Maximum likelihood estimators

For a given number of classes G , parameter estimation is achieved using maximum likelihood techniques assuming that missing data are missing at random. The vector of parameters θ contains the transformation parameters (η_{1k}, η_{2k}) for $k = 1, \dots, K$, the fixed effects $\beta^T, \mu_1^T, \dots, \mu_G^T, \gamma_1^T, \dots, \gamma_G^T$ in the mixture model, $\xi_{02}, \dots, \xi_{0G}, \xi_{12}^T, \dots, \xi_{1G}^T$ in the probability of belonging to the latent classes, $\delta_{01}, \dots, \delta_{0G}, \delta_{11}^T, \dots, \delta_{1G}^T$ in the probability of occurrence of the event and the variance-covariance parameters $\text{vec}(U), \omega_2, \dots, \omega_G, \sigma_{\alpha_1}, \dots, \sigma_{\alpha_K}, \sigma_{\varepsilon_1}, \dots, \sigma_{\varepsilon_K}$ where U is the Cholesky transformation of B , the variance-covariance matrix of the random-effects for the first class and ensures that B is a positive-definite matrix. In the following, we will note $y_i = (y_{i1}^T, \dots, y_{iK}^T)^T$ and $n_i = \sum_{k=1}^K n_{ik}$.

In a mixture model, the individual contribution to the likelihood can be decomposed according to the latent classes [15]. Then, using the conditional independence assumption between y_i and D_i given the latent classes, we obtain:

$$\begin{aligned} f(y_i, D_i) &= \sum_{g=1}^G P(c_{ig} = 1) f(y_i, D_i | c_{ig} = 1) \\ &= \sum_{g=1}^G P(c_{ig} = 1) f(y_i | c_{ig} = 1) P(D_i | c_{ig} = 1) \end{aligned} \quad (7)$$

The probability of belonging to latent class g , $P(c_{ig} = 1)$ is given by the multinomial logit regression (3); the probability of the event given the latent class g , $P(D_i|c_{ig} = 1)$ is given by the logistic model (4). The density $f(y_i|c_{ig} = 1)$ of the longitudinal outcomes in their natural scale can be computed from the density of the transformed variable $\tilde{y}_i = (\tilde{y}_{i1}^T, \dots, \tilde{y}_{iK}^T)^T$ with $\tilde{y}_{ijk} = h_k(y_{ijk})$. Indeed, as \tilde{y}_i follows a linear mixed model given the latent class g , the density of $f(y_i|c_{ig} = 1)$ can be written as the product of the multivariate Gaussian density of $\tilde{y}_i|c_{ig} = 1$ and the Jacobian of the transformations h_k as in Reference [13]:

$$\begin{aligned} f(y_i|c_{ig} = 1; \theta) &= f(\tilde{y}_i|c_{ig} = 1; \theta)J(y_i; \theta) \\ &= \phi_g(\tilde{y}_i; \theta)J(y_i; \theta) \end{aligned} \tag{8}$$

where ϕ_g is a multivariate Gaussian density with class-specific mean vector $E_{ig} = (E_{i1g}^T, \dots, E_{iKg}^T, \dots, E_{iKg}^T)^T$ and class-specific variance-covariance matrix V_{ig} given by

$$E_{ikg} = Z_i^k \mu_g + X_{1i}^k \beta + X_{2i}^k \gamma_k \tag{9}$$

$$V_{ig} = \begin{pmatrix} Z_i^1 \\ \vdots \\ Z_i^K \end{pmatrix} \omega_g^2 B(Z_i^{1T} \dots Z_i^{KT}) + \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_K \end{pmatrix} \tag{10}$$

$$\text{with } \Sigma_k = \sigma_{\alpha_k}^2 \mathbf{1}_{n_{ik}} \mathbf{1}_{n_{ik}}^T + \sigma_{\epsilon_k}^2 I_{n_{ik}}$$

where $Z_i^k = (Z(t_{i1k}), \dots, Z(t_{in_{ik}k}))^T$ is the $n_{ik} \times (p + 1)$ -matrix of time polynomials for subject i and test k ; $X_{1i}^k = (X_{1i}(t_{i1k}), \dots, X_{1i}(t_{in_{ik}k}))^T$ and $X_{2i}^k = (X_{2i}(t_{i1k}), \dots, X_{2i}(t_{in_{ik}k}))^T$ are, respectively, the $n_{ik} \times q_1$ -matrix and $n_{ik} \times q_2$ -matrix of time-dependent covariates with a common effect across the classes or a class-specific effect, and I_n and $\mathbf{1}_n$ are, respectively, the identity matrix of size n and the n -vector of 1's.

The term $J(y; \theta)$ is the Jacobian of the Beta CDF of the longitudinal outcomes y at the parameter vector value θ :

$$J(y_i; \theta) = \prod_{k=1}^K \prod_{j=1}^{n_{ik}} \frac{y_{ijk}^{\eta_{1k}-1} (1 - y_{ijk})^{\eta_{2k}-1}}{B(\eta_{1k}, \eta_{2k})} \tag{11}$$

where $B(\eta_{1k}, \eta_{2k})$ is the complete Beta function with parameters η_{1k} and η_{2k} .

At last, we obtain the following closed form for the log-likelihood of the joint model:

$$\begin{aligned} L(y; \theta) &= \sum_{i=1}^N \ln \left(\sum_{g=1}^G \frac{e^{\xi_{0g} + X_{3i}^T \xi_{1g}}}{\sum_{j=1}^G e^{\xi_{0j} + X_{3i}^T \xi_{1j}}} \times \phi_g(\tilde{y}_i; \theta) \times \frac{[e^{\delta_{0g} + X_{4i}^T \delta_{1g}}]^{D_i}}{1 + e^{\delta_{0g} + X_{4i}^T \delta_{1g}}} \right) \\ &\quad + \sum_{i=1}^N \ln(J(y_i; \theta)) \end{aligned} \tag{12}$$

A NONLINEAR LATENT CLASS MODEL

In spite of the nonlinear structure of the model, the log-likelihood has a closed-form and can be maximized using standard maximization algorithms as described in next section.

3.2. Optimization algorithm

The log-likelihood (12) is maximized using a modified Marquardt algorithm [16]. This is a Newton-Raphson like algorithm where the diagonal of the Hessian matrix at iteration l , $H^{(l)}$, is inflated to obtain a positive definite matrix: $H^{*(l)} = (H_{ij}^{*(l)})$ with $H_{ii}^{*(l)} = H_{ii}^{(l)} + \lambda[(1-v)|H_{ii}^{(l)}| + v\text{tr}(H)]$ and $H_{ij}^{*(l)} = H_{ij}^{(l)}$ if $i \neq j$. Initial values for λ and v are $\lambda = 0.01$ and $v = 0.01$. They are reduced when H^* is positive definite and increased if not. The estimates $\theta^{(l)}$ are then updated to $\theta^{(l+1)}$ using the current modified Hessian $H^{*(l)}$ and the current gradient of the parameters $g(\theta^{(l)})$ according to the formula:

$$\theta^{(l+1)} = \theta^{(l)} - \kappa H^{*(l)-1} g(\theta^{(l)}) \quad (13)$$

where, if necessary, a linesearch for κ ensures that the log-likelihood is improved at each iteration. The convergence is reached when the three following convergence criteria are satisfied: $\sum_{j=1}^m (\theta_j^{(l)} - \theta_j^{(l-1)})^2 \leq 10^{-4}$, $|L^{(l)} - L^{(l-1)}| \leq 10^{-4}$ and $g(\theta^{(l)})^T H^{(l)-1} g(\theta^{(l)}) \leq 10^{-5}$. First and second derivatives are computed by finite differences. Standard errors of the elements of B are computed by the Δ -method while standard errors of the other parameters are directly computed from the inverse of the observed Hessian matrix.

A mixture model is estimated with a fixed number of components G . To choose the optimal number of components, estimated models with different values of G are compared using the Bayesian information criterion (BIC) [17]. Due to the possible multimodality of the likelihood, each model is estimated using several sets of initial values to ensure convergence to the global maximum. Moreover, due to the complexity of the model, simple models are first estimated in order to provide adequate initial values for more complicated models.

3.3. Posterior classification

Posterior classification of the subjects in the different latent classes can be achieved in a mixture model using posterior conditional probabilities. These posterior conditional probabilities $\hat{\pi}_{ig}^{y,D}$ to belong to class g given the observations (y_i, D_i) are computed using the Bayes theorem as follows:

$$\hat{\pi}_{ig}^{y,D} = P(c_{ig} = 1 | y_i, D_i; \hat{\theta}) = \frac{P(c_{ig} = 1) f((y_i, D_i) | c_{ig} = 1; \hat{\theta})}{\sum_{l=1}^G P(c_{il} = 1) f((y_i, D_i) | c_{il} = 1; \hat{\theta})} \quad (14)$$

where the denominator is the contribution to the likelihood of subject i at the optimum. Each subject is then classified in the profile or latent class in which he has the largest probability $\hat{\pi}_{ig}^{y,D}$ of belonging.

In our model, information is of two kinds: information from repeated measures of the markers y_i and information from the event occurrence D_i . In a prognostic or diagnostic perspective, it could be of interest to consider the posterior probabilities $\hat{\pi}_{ig}^y$ of belonging to class g given only the longitudinal observations y_i . These probabilities are computed as follows:

$$\hat{\pi}_{ig}^y = P(c_{ig} = 1 | y_i; \hat{\theta}) = \frac{P(c_{ig} = 1) f(y_i | c_{ig} = 1; \hat{\theta})}{\sum_{l=1}^G P(c_{il} = 1) f(y_i | c_{il} = 1; \hat{\theta})} \quad (15)$$

C. PROUST-LIMA, L. LETENNEUR AND H. JACQMIN-GADDA

3.4. Diagnostic and prognostic tools

To propose a diagnostic and a prognostic tool based on repeated marker measures, we compute from the estimated parameter vector $\hat{\theta}$ the probability of occurrence of the event conditionally on the longitudinal observations as follows:

$$\begin{aligned} P(D_i = 1|y_i; \hat{\theta}) &= \sum_{g=1}^G P(D_i = 1|c_{ig} = 1; \hat{\theta})P(c_{ig} = 1|y_i; \hat{\theta}) \\ &= \sum_{g=1}^G \frac{e^{\delta_{0g} + X_{4i}^T \delta_{1g}}}{1 + e^{\delta_{0g} + X_{4i}^T \delta_{1g}}} \times \hat{\pi}_{ig}^y \end{aligned} \quad (16)$$

For each threshold between 0 and 1, we classify as demented the subjects with $P(D_i = 1|y_i; \hat{\theta})$ above the threshold and compute sensitivity and specificity of the procedure. The ROC curve may then be drawn and the area under the ROC curve (AUC) may be computed to evaluate the performances of the model in detecting the event.

4. APPLICATION

4.1. The PAQUID cohort study

The data come from the French prospective cohort study PAQUID, initiated in 1988 to study normal and pathological ageing [14]. Subjects included in the cohort were 65 years and older at the initial visit and were followed six times with a visit at one (T1), 3 (T3), 5 (T5), 8 (T8), 10 (T10) and 13 (T13) years after the initial visit (T0). At each visit, a battery of psychometric tests was completed and a two phase screening procedure was carried out for the diagnosis of dementia. Subjects who met DSM-III-R criteria [18] A, B and C (impairment of memory and at least one other cognitive function and interference with daily living) or those presenting a decline of three points or more on the Mini Mental State Examination [19] scale since the previous visit were seen by a neurologist who made the clinical diagnosis. Measurements at the initial visit were excluded from the analysis because of a first passing effect [20].

The aim of the analysis was both to describe the profiles of cognitive evolution up to T13 in a subpopulation of subjects free of dementia until T10, and to propose a diagnostic and a prognostic tool of dementia using repeated measures of psychometric tests. Three psychometric tests were considered ($K = 3$): the Isaacs Set Test (IST) [21] shortened at 15s which evaluates verbal fluency, the Benton Visual Retention Test (BVRT) [22] which evaluates visual memory and the Digit Symbol Substitution Test of Wechsler (DSSTW) [23] which evaluates attention and logical reasoning. For the three tests, low values indicate a more severe impairment.

The sample included 834 subjects who were visited at T13, were visited and free of dementia at T10, and completed each of the three psychometric tests at least once between T1 and T13. The median number of measures was 4 for the DSSTW and 5 for the IST and the BVRT. Among the 834 subjects, 114 had a positive diagnosis of dementia at T13-visit. The time variable was the exact time in decade between the age at the T13-visit and the age at the current visit. Thus it ranged from about -1.2 to 0 . The polynomial function of time chosen for the evolution of the latent process included a random intercept, a random slope and a random quadratic slope. The

A NONLINEAR LATENT CLASS MODEL

Table I. Likelihood, BIC and AUC of the adjusted model for various numbers of latent classes.

Number of classes (G)	Number of parameters	Log-likelihood	BIC	AUC
1	26	-29337.87	58850.61	
2	33	-29227.74	58677.44	0.855
3	40	-29216.26	58701.57	0.860
4	47	-29206.93	58730.00	0.862

covariates included in the model were educational level (graduated from primary school *versus* not graduated), gender and a dichotomous variable for age with a cut-off at the median (less than 83 years old at T13-visit *versus* more than 83 years old at T13-visit). The three covariates were tested in the three parts of the model: in the longitudinal model either as a common effect across classes (X_1) or as a class-specific effect (X_2), in the model for the probability of class membership (X_3) and in the model for the probability of occurrence of the event (X_4).

4.2. The nonlinear latent class model

The model adjusted on the three covariates was estimated for a number of latent classes G varying from 1 to 4. Whatever the number of classes, the best adjustment on covariates included age, educational level and gender in association with the mean latent cognitive level but not with the time variables. These associations with the trajectories were common across the classes. Educational level was found associated with the probability of dementia at T13 after adjustment on the latent classes, and for the model with 2, 3 and 4 classes, age was also associated with the probability of latent class membership.

Using the BIC [17] for selecting the model with the optimal number of latent classes, the model with two classes was retained (see Table I). Fixed-effect estimates are presented in Table II and the two predicted evolutions of the latent process ($Z(t)\hat{\mu}_g$, $g = 1, 2$) are displayed in Figure 1.

The first class trajectory is quite linear with a slight decline until T13-visit. As seen in Table II, the associated probability of dementia is almost null ($\exp(-2.815)/(1 + \exp(-2.815)) = 0.057$). This trajectory seems to represent the mean cognitive evolution in normal ageing. In contrast, the mean cognitive level in the second class is always lower than in the first class and the decline is sharper and nonlinear. The associated probability of dementia is very high ($\exp(2.715)/(1 + \exp(2.715)) = 0.938$). Thus this trajectory seems to represent the mean decline in a pre-diagnostic phase of dementia. In the following, we will call class 1 'slight decline' class and class 2 'marked decline' class.

Whatever the class trajectory, subjects who graduated from primary school had a significantly better mean cognitive level, subjects older than 83 years had a significantly lower cognitive level than younger subjects and men had a significantly higher mean cognitive level. After adjustment on the latent classes, educational level was found associated with the probability of dementia at T13: adjusted on the shape of the cognitive decline, subjects not graduated from primary school had a higher risk of dementia at T13 (OR = 5.464, CI_{95 per cent} = [1.107; 26.967]) than subjects graduated from primary school. This effect of education on the risk of dementia was not significantly different according to the classes. Finally, age was found to be associated with the class membership probability. The odds-ratio for being in the marked decline class rather than in

C. PROUST-LIMA, L. LETENNEUR AND H. JACQMIN-GADDA

Table II. Estimations of the fixed effects in the final nonlinear mixture model with two latent classes.

Parameter	Estimate	Standard-error
Class 1		
Intercept	0.511	0.015
Linear slope	-0.0484	0.0063
Quadratic slope	0.0100	0.0048
Class 2		
Intercept	0.376	0.021
Linear slope	-0.259	0.024
Quadratic slope	-0.106	0.017
Covariates		
Gender*	0.0121	0.0055
Educational level [†]	0.108	0.008
Age [‡]	-0.0364	0.0057
Probability of dementia at T13		
Intercept class 1 (δ_{01})	-2.815	0.453
Intercept class 2 (δ_{02})	2.715	0.916
Educational level [†]	-1.698	0.815
Probability of belonging to class 2		
Intercept	-2.188	0.208
Age [‡]	0.889	0.231

*Reference: female.

[†]Reference: no diploma.[‡]Reference: younger than 83 years old at T13-visit.

the slight decline class was 2.433 (CI_{95 per cent} = [1.547; 3.826]) for subjects above 83 years old compared to younger people. Older subjects were more prone to experience both a sharp cognitive decline and a dementia.

The class-specific scale parameter in the variance-covariance matrix for the random-effects was estimated at $\omega_2 = 1.056$ (SE = 0.105). Thus, the variability was not significantly different for the two classes.

4.3. Posterior classification

According to the posterior class-membership probabilities given all the observed data $\hat{\pi}_{ig}^{y,D}$, 719 subjects were classified in the slight decline class including 708 subjects without dementia at T13 and 115 subjects were classified in the marked decline class including 103 subjects diagnosed demented at T13. This almost perfect discrimination underlines that we were able to discriminate with this model the evolution of pre-demented and normal subjects. Nevertheless, a good discrimination was expected since information about dementia was used to compute the posterior probabilities.

Thus, classification obtained using only information from the longitudinal outcomes, that is to say $\hat{\pi}_{ig}^y$, is displayed in Table III: 749 subjects were classified in the slight decline class including 685 subjects without dementia at T13 and 85 subjects were classified in the marked decline class

A NONLINEAR LATENT CLASS MODEL

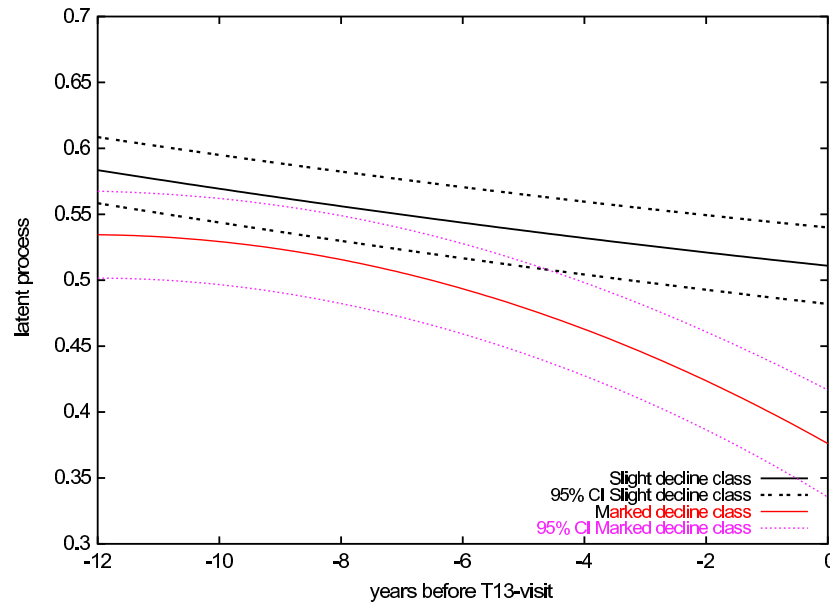


Figure 1. Predicted trajectories for the two classes of the adjusted nonlinear latent class model for a female without diploma and younger than 83 years old at T13 (plain lines = mean predicted evolution; dashed lines = 95 per cent confidence bands).

Table III. Posterior classification from $\hat{\pi}_{ig}^y$ and dementia diagnosis at T13.

Dementia diagnosis at T13	Classification		Total
	Slight decline class	Marked decline class	
Positive	64	50	114
Negative	685	35	720
Total	749	85	834

including 50 subjects with a positive diagnosis of dementia at T13. As expected with the significant effect of age on class membership, we found that subjects classified in the marked decline class were older than those classified in the slight decline class (67 per cent were older than 83 years at the end of the follow-up *versus* 44 per cent in the slight decline class). In contrast, the two posterior classes did not differ according to gender (42 per cent of men in the marked decline class *versus* 38 per cent in the slight decline class) and educational level (78 per cent with a high educational level *versus* 81 per cent).

According to this classification, we also computed in Table IV the means of the posterior probabilities of belonging to each class l over the n_g subjects classified in class g ($(1/n_g) \sum_{i=1}^{n_g} \hat{\pi}_{il}^y$, $(g, l) \in \{1, 2\}^2$), high diagonal terms indicating a good classification quality [24]. The mean of the posterior probabilities to be in the slight decline class was 0.920 for subjects classified in this class showing a very high discrimination and an unambiguous class affectation. In contrast, the mean posterior probability to be in the marked decline class was 0.753 for subjects classified

C. PROUST-LIMA, L. LETENNEUR AND H. JACQMIN-GADDA

Table IV. Posterior classification table: mean of the posterior probabilities of belonging to each class l over the n_g subjects classified in class g ($(1/n_g)\sum_{i=1}^{n_g} \hat{\pi}_{il}^y, (g, l) \in \{1, 2\}^2$).

Class $g \setminus$ class l	Slight decline class	Marked decline class
Slight decline class	0.920	0.080
Marked decline class	0.247	0.753

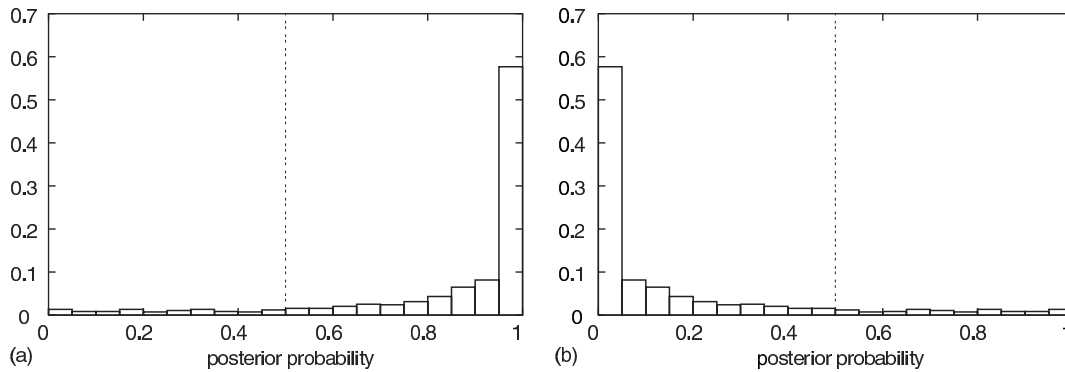


Figure 2. Empirical distribution of the posterior probabilities of being in the slight decline class (a) and in the marked decline class (b).

in this class revealing a more ambiguous affectation. Indeed, subjects classified in the marked decline class had a non-negligible mean posterior probability of being in the slight decline class ($(1/n_2)\sum_{i=1}^{n_2} \hat{\pi}_{i1}^y = 0.247$). The histogram of the posterior probabilities displayed in Figure 2 emphasizes the more ambiguous affectation to the marked decline class.

4.4. Adequation of the model

To evaluate the fit of the data, we compared for each test and each class, the weighted mean trajectory of the observed scores with the weighted mean trajectory of the predicted values. For each observation, that is each triplet (i, j, k) , the marginal predicted value in the natural scale of the outcome, that is the estimate of $E(h_k^{-1}(\tilde{y}_{ijk})|c_{ig} = 1)$ was obtained using a numerical integration of $h_k^{-1}(\tilde{y}_{ijk})$ using the estimated Gaussian density $\phi_g(\tilde{y}_{ijk}; \hat{\theta})$ defined by (9) and (10). Then, at each visit, the mean of the available marginal predictions weighted by the posterior probabilities was computed and compared with the corresponding weighted mean of the observed values. The results are displayed in Figure 3. For the two classes, the fit is good whatever the psychometric test indicating a good adequacy of the model.

4.5. Diagnosis of dementia

To evaluate the performances of the model to diagnose dementia using only information about the psychometric tests, we computed the posterior probabilities of dementia $P(D_i = 1|y_i; \hat{\theta})$ using (16) and built the ROC curve represented in Figure 4. The area under the curve was 0.855. We

A NONLINEAR LATENT CLASS MODEL

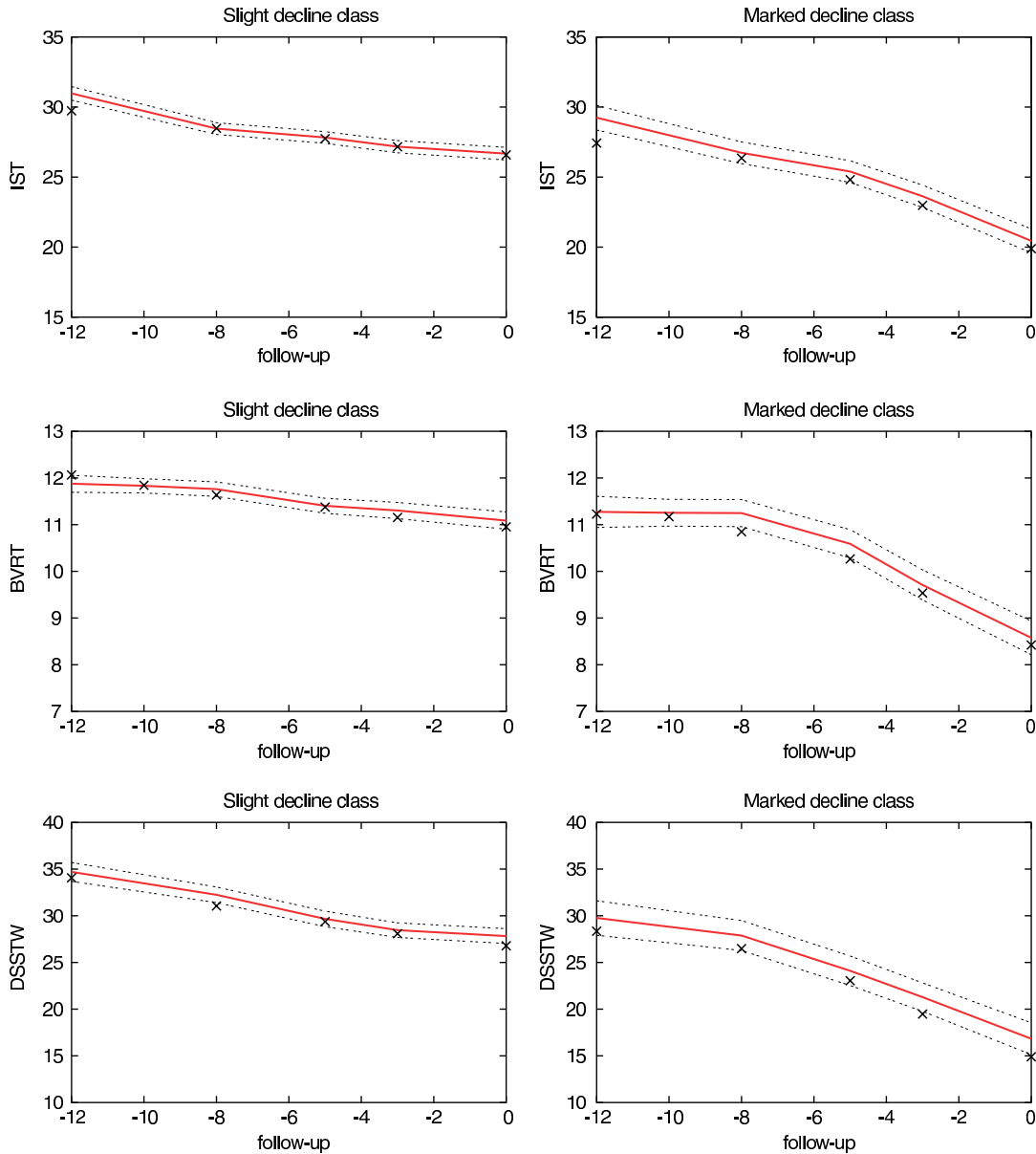


Figure 3. Marginal predicted weighted mean evolution (crosses) *versus* observed weighted mean evolution (plain line) with its 95 per cent confidence bands (dashed lines) for each class and each test in their natural scale according to the visits.

also performed a leave-one-out procedure for computing the area under the curve. It consisted in leaving iteratively each subject i out, estimating the parameters $\hat{\theta}_{(-i)}$ without i and computing $P(D_i = 1|y_i; \hat{\theta}_{(-i)})$. This procedure lead to a very close AUC (0.852).

C. PROUST-LIMA, L. LETENNEUR AND H. JACQMIN-GADDA

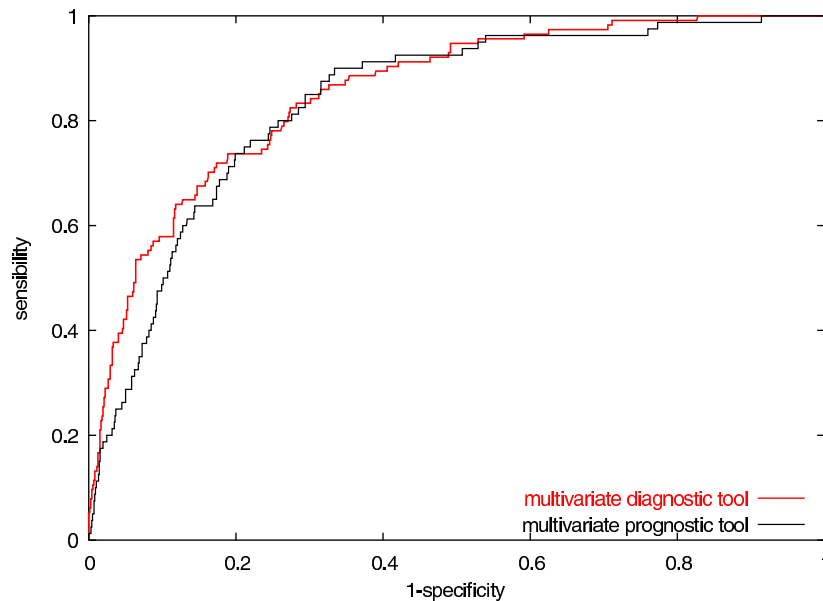


Figure 4. ROC curve of the posterior probabilities of dementia from the multivariate nonlinear latent class model for the diagnosis analysis ($N = 834$) in grey plain line and for the prognosis analysis ($N = 1059$) in black plain line.

One advantage of this diagnostic tool for dementia is that it can be used whatever the number and the times of measurements of the cognitive tests in the past 12 years and even if the subject has measurements for only one test. We compared the diagnostic ability of the estimated multivariate model (Table II) according to the number of psychometric tests used for the computation of the posterior probabilities of dementia.

Table V clearly shows that the AUC increases with the number of psychometric tests used for computing the posterior probability of dementia, underlining the interest of a flexible model which can handle multivariate data.

We compared the performance of our model with a simpler method for diagnosing dementia according to cognitive ability, that is a logistic regression including six explanatory variables: age, educational level, gender and cognitive test scores for the IST, the BVRT and the DSSTW at the last visit T13. As the logistic regression requires no missing data for all the explanatory variables included in the model, the model was estimated on a reduced sample of 647 subjects having a measure of each test at the T13-visit. On this reduced sample, the AUC from the logistic regression was 0.852 which is nearly as good as for our multivariate longitudinal model. However, 187 subjects (22 per cent of the sample) with incomplete measures were excluded from this analysis including a high proportion of demented subjects (58 demented were excluded among the 114 demented of the initial sample). By using longitudinal and multivariate cognitive assessment, our model allows to compute a probability of dementia whatever the available information on the three tests unlike the logistic regression. It is of particular interest in epidemiological studies in which subjects in pre-dementia phase do not often complete all the tests.

A NONLINEAR LATENT CLASS MODEL

Table V. Areas under the curve (AUC) obtained from the multivariate nonlinear mixture model in a diagnosis analysis (probabilities of dementia at T13 computed using cognitive measures between T1 and T13) or in a prognosis analysis (probabilities of dementia at T10 computed using cognitive measures between T1 and T8).

Psychometric tests used to compute probabilities of dementia	AUC (diagnostic at T13)	AUC (prognostic at T10 given T1–T8)
IST	0.775	0.840
BVRT	0.796	0.788
DSSTW	0.808	0.802
IST + BVRT	0.831	0.841
IST + DSSTW	0.831	0.832
BVRT + DSSTW	0.842	0.814
IST + BVRT + DSSTW	0.855	0.837

4.6. Prognosis of dementia

Even if the model was built in a diagnostic perspective, it can be used to compute probability of dementia at time T knowing at least one measure from IST, BVRT or DSSTW in the last 12 years. Thus, it can be used as a prognostic tool. We explored the ability of our model to predict dementia 2 years after the last cognitive measurements. More specifically, we tried to predict dementia at T10 using cognitive measurements between T1 and T8 in the PAQUID study. In this way, for each of the 1059 subjects with at least 1 measure at each test between T1 and T8 and with a negative diagnosis of dementia at T8, we computed the probability of dementia at T10 knowing measures collected between T1 and T8 and using the vector of parameters estimated on the previous sample (Table II). As for the previous analysis on diagnosis of dementia, we compared the results according to the number of psychometric tests used to compute the probability of dementia. We imposed that each subject had at least one measure at each test to compare the AUC according to the number of tests on samples of the same size. The results are presented in Table V.

Whatever the number of psychometric tests used for computing the probability of dementia at 2 years, the AUC was quite high. The ROC curve when using the three psychometric tests is displayed in Figure 4. For the BVRT and the DSSTW, we found again that the AUC for the prognosis increased when using more information. However, in this sample, the prognostic ability of IST was very good and did not increase when other tests were added.

5. DISCUSSION

We proposed a nonlinear latent class model for multivariate longitudinal outcomes and a binary variable. In ageing context, by modelling jointly cognitive trajectory using three psychometric tests and dementia diagnosis, we distinguished the cognitive decline in a normal cognitive ageing process and the cognitive decline in a pre-diagnosis phase of dementia. From the estimates of the model, we proposed a diagnostic and prognostic tool and compared, according to this tool, the ability of our model to predict dementia when varying the quantity of available information on cognitive ability. Thus, we underlined that using several cognitive tests increased the power of the diagnostic and prognostic tool.

C. PROUST-LIMA, L. LETENNEUR AND H. JACQMIN-GADDA

A key aspect of the flexibility of our model lies in the nonlinear transformations which allow to model continuous outcomes with a distribution far from a Gaussian distribution. Beta CDFs were chosen because firstly, they are flexible functions with only two parameters per marker and secondly, they gave a very good fit of the cognitive test scores in our elderly sample as it was previously shown [13]. The interest of these estimated transformations was extensively discussed in Reference [13] showing in particular that they give interesting results about the metrological properties of each psychometric test.

In a perspective of proposing diagnostic and prognostic tools, one could wonder whether it is sensible to select the optimal number of classes using the BIC criterion which tends to retain a parsimonious model. One could prefer selecting the optimal number of classes using a less conservative criterion like the AIC for example, or stopping the process when the AUC of the model is not significantly improved. In our application on dementia diagnosis, the AUC from the three and four class model were just a little improved so we retained only two latent classes of evolution, the first one representing an almost normal cognitive ageing process and the second one the cognitive decline in a pre-diagnosis phase of dementia. By stratifying on dementia status and estimating the two separated multivariate models, we could have probably obtained quite the same evolutions but we would not have been able to compute a diagnostic tool of dementia.

Our joint model is an extension of the classic mixture models [25, 26]. We no longer study the main heterogeneity in the population, we rather investigate the unobserved subpopulation structure in association with the clinical event. Such joint models including a mixture model for longitudinal data and a logistic regression had already been proposed by Lin *et al.* [5] for studying one longitudinal outcome with a Gaussian assumption. They had then improved their model by replacing the logistic regression by a survival model [1]. In this paper, we focused on a multivariate approach allowing flexible relationships between the observed markers and the latent process. We only considered a logistic regression for the clinical event but we showed that even without a time-to-event model, we were able to propose a prognostic tool in addition to the diagnostic tool. However, as a model for a binary event does not account for time-to-event, parameters must be estimated on a selected sample for which the dementia status is known at a given time. To avoid this selection in the estimation step and include cases where the event arises at any time, the model should be next extended for including a proportional hazard model for the time-to-event instead of the logistic regression.

REFERENCES

1. 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.
2. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**:465–480.
3. Whulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.
4. Jacqmin-Gadda H, Commenges D, Dartigues J-F. Random changepoint model for joint modeling of cognitive decline and dementia. *Biometrics* 2006; **62**:254–260.
5. Lin H, Mc Culloch CE, Turnbull BW, Slate EH, Clark LC. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* 2000; **19**:1303–1318.
6. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 1999; **50**:463–469.
7. Thiébaud R, Jacqmin-Gadda H, Babiker A, Commenges D. The CASCADE collaboration. Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+

A NONLINEAR LATENT CLASS MODEL

- cell count and HIV RNA viral load in response to treatment of HIV infection. *Statistics in Medicine* 2005; **24**:65–82.
8. Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. *Biometrics* 2001; **57**:81–87.
 9. Roy J, Lin X. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* 2000; **56**:1047–1054.
 10. Rabe-Hesketh S, Skrondal A, Pickles A. Generalized multilevel structural equation modeling. *Psychometrika* 2004; **69**:167–190.
 11. Ganiayre J, Commenges D, Letenneur L. A latent process model for dementia and psychometric tests. *Technical Report*, 2006.
 12. Hashemi R, Jacqmin-Gadda H, Commenges D. A latent process model for joint modeling of events and marker. *Lifetime Data Analysis* 2003; **9**:331–343.
 13. Proust C, Jacqmin-Gadda H, Taylor JMG, Ganiayre J, Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* 2006, online publication on 26 April 2006. doi: 10.1111/j.1541-0420.2006.00573.x
 14. Letenneur L, Commenges D, Dartigues J-F, Barberger-Gateau P. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *International Journal of Epidemiology* 1994; **23**:1256–1261.
 15. Proust C, Jacqmin-Gadda H. Estimation of linear mixed models with a mixture of distribution for the random-effects. *Computer Methods and Programs in Biomedicine* 2005; **78**:165–173.
 16. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics* 1963; **11**:431–441.
 17. Schwartz G. Estimating the dimension of a model. *The Annals of Statistics* 1978; **6**:461–464.
 18. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (DSM-III-R, 3rd edition revised). Washington, DC, 1987.
 19. Folstein MF, Folstein SE, McHugh PR. Mini Mental State. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975; **12**:189–198.
 20. Jacqmin-Gadda H, Fabrigoule C, Commenges D, Dartigues J-F. A 5-year longitudinal study of the Mini Mental State Examination in normal aging. *American Journal of Epidemiology* 1997; **145**:498–506.
 21. Isaacs B, Kennie AT. The Set Test as an aid to the detection of dementia in old people. *British Journal of Psychiatry* 1973; **123**:467–470.
 22. Benton AL. *Manuel pour l'application du Test de Réention Visuelle. Applications cliniques et expérimentales* (2ème édition française). Centre de Psychologie appliquée, Paris, 1965.
 23. Wechsler D. *WAIS-R Manual*. Psychological Corporation: New York, 1981.
 24. Muthén B, Brown CH, Masyn K, Jo B, Khoo S-T, Yang C-C, Wang C-P, Kellam SG, Carlin JB, Liao J. General growth mixture modeling for randomized preventive intervention. *Biostatistics* 2002; **3**:459–475.
 25. Legler JM, Davis WW, Potosky AL, Hoffman RM. Latent variable modelling of recovery trajectories: sexual function following radical prostatectomy. *Statistics in Medicine* 2004; **23**:2875–2893.
 26. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data* (Chapter 12). Springer Series in Statistics, 2000.

Chapitre 5

Modèle nonlinéaire à classes latentes pour l'analyse conjointe de plusieurs marqueurs quantitatifs et d'un temps d'événement

Par la modélisation conjointe de plusieurs marqueurs longitudinaux et d'une variable binaire, le modèle présenté au chapitre précédent permet de décrire les différents profils d'évolution cognitive associés à la survenue d'une démence et de proposer un outil de diagnostic de démence basé sur la trajectoire cognitive du sujet. Cependant, ce modèle a une limite importante. En utilisant un modèle de régression logistique pour prédire l'événement clinique, nous étudions la probabilité de l'événement à un temps donné. L'estimation des paramètres se fait donc sur un échantillon de sujets sélectionnés : tous les sujets ont eu un diagnostic de démence au temps considéré. Cela peut impliquer des biais sur les estimations des paramètres et par conséquent des biais sur l'outil de diagnostic proposé. Pour pallier

ce problème de sélection, il est nécessaire d'étudier la survenue de l'événement de manière dynamique. C'est ce que nous proposons dans ce chapitre en modélisant le risque de survenue de l'événement par un modèle de survie paramétrique différent dans chaque classe latente.

Nous détaillons la méthodologie du modèle et l'estimation des paramètres. Nous proposons ensuite une extension semi-paramétrique du modèle avec une modélisation de la fonction de risque basée sur des splines puis une étude de simulation afin d'évaluer la qualité des estimateurs du maximum de vraisemblance. Enfin, après avoir proposé des méthodes pour évaluer l'adéquation aux données, nous appliquons ce modèle dans le cadre de l'étude du vieillissement. Cette application consiste à décrire les profils de déclin cognitif en fonction de l'âge associés à la survenue d'une démence et évaluer la probabilité de démence prédite par les trajectoires des tests psychométriques.

5.1 Formulation du modèle conjoint paramétrique

Le modèle nonlinéaire à classes latentes détaillé dans ce chapitre est une extension directe du modèle présenté dans le chapitre précédent. En particulier, le modèle pour les données longitudinales est le même. Néanmoins, pour plus de clarté, nous rappelons la forme du modèle pour les données longitudinales avant de définir le modèle pour la fonction de risque.

5.1.1 Modèle pour les données longitudinales

Pour chaque sujet i , $i = 1, \dots, N$ et chaque marqueur k , $k = 1, \dots, K$, n_{ik} mesures sont collectées. Nous disposons donc des données y_{ijk} mesurées aux temps t_{ijk} , $j = 1, \dots, n_{ik}$. Ces mesures sont liées au processus latent $(\Lambda_i(t))_{t \geq 0}$ qui représente la cognition latente en temps continu par des fonctions de répartition Beta, notées h_k

et paramétrées par η_k :

$$\tilde{y}_{ijk} = h_k(y_{ijk}; \eta_k) = \Lambda_i(t_{ijk}) + \alpha_{ik} + \epsilon_{ijk} \quad (5.1)$$

où \tilde{y}_{ijk} est la variable transformée par h_k , α_{ik} est une intercept aléatoire distribuée suivant une loi normale $\mathcal{N}(0, \sigma_{\alpha_k}^2)$ et ϵ_{ijk} sont des erreurs de mesure indépendantes distribuées suivant une loi normale $\mathcal{N}(0, \sigma_{\epsilon_k}^2)$.

Le modèle à classes latentes est basé sur l'hypothèse qu'il existe G profils d'évolution du processus latent. Les G variables indicatrices latentes $(c_{ig})_{g=1,G}$ définissent l'appartenance aux classes latentes, la probabilité d'appartenir à chaque classe étant modélisée par une régression multinomiale dépendante de variables explicatives X_{3i} . Enfin, l'évolution du processus latent dans la classe g est définie suivant un modèle linéaire mixte :

$$\Lambda_i(t) |_{c_{ig}=1} = Z(t)^T u_{ig} + X_{1i}(t)^T \beta + X_{2i}(t)^T \gamma_g, \quad t \geq 0 \quad (5.2)$$

où $Z(t)$ est le vecteur de variables dépendantes du temps associé au vecteur d'effets aléatoires spécifiques à la classe u_{ig} , distribués suivant une loi multivariée normale $\mathcal{N}(\mu_g, \omega_g^2 B)$ avec $\omega_1 = 1$. Les vecteurs de variables explicatives $X_{1i}(t)$ et $X_{2i}(t)$ sont respectivement associés au vecteur d'effets fixes β communs aux classes et au vecteur d'effets fixes γ_g spécifiques aux classes. Aucune variable explicative ne peut être présente simultanément dans les deux vecteurs $X_{1i}(t)$ et $X_{2i}(t)$ pour des raisons d'identifiabilité.

5.1.2 Modèle de survie conjoint

Quelques notations

Dans les études épidémiologiques, le délai de survenue d'événement T_i^* n'est généralement pas observé pour tous les sujets, certains n'ayant pas subi l'événement à la fin de la fenêtre d'observation. Il s'agit du phénomène de censure à droite qu'il

faut prendre en compte dans la modélisation de l'événement. On définit donc la variable binaire d'événement E_i qui vaut 1 si le sujet i a subi l'événement dans la fenêtre d'étude et 0 sinon. Le délai de survie observé est $T_i = \min(T_i^*, C_i)$ où C_i est le temps de censure : $T_i = T_i^*$ si $E_i = 1$ et $T_i = C_i$ si $E_i = 0$.

Formulation générale du modèle de survie

Nous proposons dans ce travail de modéliser le délai jusqu'à l'événement T_i^* par un modèle à risques proportionnels. Ainsi, dans chaque classe latente, la fonction de risque est définie de la façon suivante :

$$\lambda(t \mid c_{ig} = 1, X_{4i}; \zeta_g, \delta_g) = \lambda_{0g}(t; \zeta_g) e^{X_{4i}\delta_g} \quad (5.3)$$

où X_{4i} est le vecteur de variables explicatives associé au vecteur de paramètres δ_g qui peut être différent dans chaque classe et $\lambda_{0g}(t; \zeta_g)$ est le risque instantané de l'événement spécifique à la classe g dans les modalités de référence des variables explicatives (risque de base). La fonction de survie associée est :

$$\begin{aligned} S(t \mid c_{ig} = 1, X_{4i}; \zeta_g, \delta_g) &= \exp\left(-\int_0^t \lambda(u \mid c_{ig} = 1, X_{4i}; \zeta_g, \delta_g) du\right) \\ &= S_{0g}(t; \zeta_g)^{\exp(X_{4i}\delta_g)} \end{aligned} \quad (5.4)$$

où $S_{0g}(t; \zeta_g) = \exp\left(-\int_0^t \lambda_{0g}(u; \zeta_g)\right)$ est la fonction de survie dans les modalités de référence des variables explicatives.

Simplifications du modèle

Nous avons d'abord considéré un modèle de survie stratifié sur les classes latentes c'est à dire des fonctions de risque de base $\lambda_{0g}(t; \zeta_g)$ spécifiques aux classes latentes. Cependant, d'une part cette stratification peut induire un nombre important de paramètres lorsque le nombre de classes latentes considéré est grand et, d'autre part

cela ne permet pas de quantifier simplement l'augmentation de risque d'une classe à l'autre. Nous avons donc envisagé une deuxième spécification des fonctions de risque : le risque d'événement $\lambda_0(t; \zeta)$ est défini dans la classe de référence puis le risque dans les autres classes est proportionnel à celui de la classe de référence : $\lambda_{0g}(t) = \lambda_0(t; \zeta)e^{\delta_{0g}}$ avec $\delta_{01} = 0$ si la classe 1 est la classe de référence. Ainsi, $e^{\delta_{0g}}$ est le risque relatif d'événement pour la classe g par rapport à la classe de référence.

Ensuite, que le modèle soit stratifié ou non, nous avons choisi une modélisation paramétrique de la fonction de risque. En effet, étant donné la complexité du modèle conjoint et notamment le nombre important de paramètres induit par les tests psychométriques et les composantes latentes, nous avons préféré utiliser des lois simples définies par peu de paramètres pour modéliser le risque. Quelques lois statistiques sont fréquemment utilisées en modèles de survie, la loi exponentielle, la loi de Weibull, la loi Gamma, le modèle log-normal ou le modèle log-logistique. Nous avons opté dans ce travail pour la loi de Weibull définie comme suit :

$$\lambda_{0g}(t; (\zeta_{1g}, \zeta_{2g})) = \zeta_{1g}\zeta_{2g}(\zeta_{1g}t)^{\zeta_{2g}-1} \quad (5.5)$$

Et la fonction de survie Weibull associée :

$$S_{0g}(t; (\zeta_{1g}, \zeta_{2g})) = e^{-(\zeta_{1g}t)^{\zeta_{2g}}} \quad (5.6)$$

Enfin, nous avons défini dans le modèle de survie (5.3), des paramètres δ_g associés aux variables explicatives potentiellement différents d'une classe latente à l'autre. En effet, suivant l'objectif de l'application, on voudra :

- soit les considérer différents dans chaque classe, e^{δ_g} représentant alors le risque relatif d'événement dans la classe g pour une augmentation d'une unité de la variable explicative correspondante après ajustement sur les classes latentes
- soit les considérer communs aux classes, e^{δ} représentant dans ce cas le risque relatif d'événement pour une augmentation d'une unité de la variable explicative correspondante quelle que soit la classe latente d'appartenance, après ajustement sur les classes latentes.

5.2 Estimation du modèle

5.2.1 Vraisemblance

Notons θ , l'ensemble des paramètres du modèle conjoint. En tenant compte de l'hypothèse d'indépendance conditionnelle entre les marqueurs longitudinaux et le délai de survie, et en utilisant la décomposition suivant les classes latentes, la contribution individuelle L_i à la vraisemblance du modèle conjoint peut s'écrire de la manière suivante :

$$L_i(\theta) = \sum_{g=1}^G P(c_{ig} = 1; \theta) f(y_i | c_{ig} = 1; \theta) \lambda(T_i | c_{ig} = 1; \theta)^{E_i} S(T_i | c_{ig} = 1; \theta) \quad (5.7)$$

où $P(c_{ig} = 1; \theta) = \pi_{ig}$ est défini suivant le modèle multinomial : $\pi_{ig} = \frac{e^{\xi_{0g} + X_{3i}\xi_{1g}}}{\sum_{l=1}^G e^{\xi_{0l} + X_{3i}\xi_{1l}}}$, et le risque $\lambda(T_i | c_{ig} = 1; \theta)$ ainsi que la survie $S(T_i | c_{ig} = 1; \theta)$ sont définis selon le modèle de Weibull.

La densité des marqueurs longitudinaux dans la classe latente g , $f(y_i | c_{ig} = 1; \theta)$, est le produit de la densité multivariée normale $\phi_g(\tilde{y}_i; \theta)$ de la variable transformée \tilde{y}_i et du jacobien des transformations des marqueurs $J(y_i; \theta)$:

$$\begin{aligned} f(y_i | c_{ig} = 1; \theta) &= f(\tilde{y}_i | c_{ig}=1; \theta) J(y_i; \theta) \\ &= \phi_g(\tilde{y}_i; \theta) J(y_i; \theta) \end{aligned} \quad (5.8)$$

De là, la log-vraisemblance $\mathcal{L}(\theta)$ du modèle conjoint incluant un modèle de survie pour l'événement a l'écriture analytique suivante :

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{i=1}^N \ln \left(\sum_{g=1}^G \pi_{ig} \times \phi_g(\tilde{y}_i; \theta) \times (\zeta_{1g} \zeta_{2g} (\zeta_{1g} T_i)^{\zeta_{2g}-1} e^{X_{4i}\delta_g})^{E_i} e^{-(\zeta_{1g} T_i)^{\zeta_{2g}} e^{X_{4i}\delta_g}} \right) \\ &\quad - \sum_{i=1}^N \ln (J(y_i; \theta)) \end{aligned} \quad (5.9)$$

Les estimateurs du maximum de vraisemblance $\hat{\theta}$ sont alors obtenus en maximisant cette expression à l'aide de l'algorithme de Marquardt présenté précédemment.

5.2.2 Troncature à gauche

Il est très fréquent dans une étude de cohorte sur le vieillissement que les sujets soient inclus à la condition qu'ils n'aient pas subi l'événement avant l'âge d'entrée dans l'étude. Pour tenir compte de cette troncature à gauche, les paramètres doivent être estimés par maximisation de la vraisemblance conditionnelle à la troncature. La contribution individuelle à la vraisemblance L_i doit donc être divisée par la probabilité de survie jusqu'à l'inclusion dans l'étude T_{0i} :

$$L_i^{(2)}(\theta) = \frac{L_i(\theta)}{S(T_{0i}; \theta)} \quad (5.10)$$

où $S(T_{0i}; \theta)$ est la fonction de survie marginale en T_{0i} définie comme suit :

$$S(T_{0i}; \theta) = \sum_{g=1}^G P(c_{ig} = 1) f(T_i \geq T_{0i} \mid c_{ig} = 1; \theta) = \sum_{g=1}^G \pi_{ig} S_{0g}(T_{0i}; \theta)^{\exp(X_{4i} \delta_g)} \quad (5.11)$$

En cas de données tronquées à gauche, la log-vraisemblance à maximiser devient donc :

$$\mathcal{L}^{(2)}(\theta) = \mathcal{L}(\theta) - \sum_{i=1}^N \ln \left(\sum_{g=1}^G \pi_{ig} e^{-(\zeta_{1g} T_{0i}) \zeta_{2g} e^{X_{4i} \delta_g}} \right) \quad (5.12)$$

5.2.3 Classification et probabilité *a posteriori* de l'événement

A partir des estimations $\hat{\theta}$, nous pouvons calculer les probabilités *a posteriori* d'appartenance aux classes latentes sachant les données :

- la probabilité $\hat{\pi}_{ig}^{y_i, T}$ d'appartenir à chaque classe latente g sachant à la fois les mesures répétées aux tests y_i , les variables explicatives x_i et le délai jusqu'à l'événement ou jusqu'à la sortie d'étude T_i :

$$\begin{aligned} \hat{\pi}_{ig}^{y_i, T} &= P(c_{ig} = 1 \mid y_i, (T_i, E_i), x_i; \hat{\theta}) \\ &= \frac{P(c_{ig} = 1 \mid x_i; \hat{\theta}) f(y_i, (T_i, E_i) \mid c_{ig} = 1, x_i; \hat{\theta})}{\sum_{l=1}^G P(c_{il} = 1 \mid x_i; \hat{\theta}) f(y_i, (T_i, E_i) \mid c_{il} = 1, x_i; \hat{\theta})} \end{aligned} \quad (5.13)$$

- la probabilité $\hat{\pi}_{ig}^y$ d'appartenir à chaque classe latente g sachant les mesures répétées aux tests y_i et les variables explicatives x_i définie comme suit :

$$\hat{\pi}_{ig}^y = P(c_{ig} = 1 \mid y_i, x_i; \hat{\theta}) = \frac{P(c_{ig} = 1 \mid x_i; \hat{\theta})f(y_i \mid c_{ig} = 1, x_i; \hat{\theta})}{\sum_{l=1}^G P(c_{il} = 1 \mid x_i; \hat{\theta})f(y_i \mid c_{il} = 1, x_i; \hat{\theta})} \quad (5.14)$$

A partir de l'une et l'autre des probabilités *a posteriori*, il est possible de classer les sujets dans la classe à laquelle ils ont la plus grande probabilité d'appartenir. La première classification, conditionnelle aux observations sur les tests psychométriques et le délai d'événement, permet de vérifier la bonne adéquation du modèle aux données alors que la classification utilisant uniquement l'information sur les tests psychométriques permet d'évaluer la qualité de discrimination des tests psychométriques et est utilisée pour calculer la probabilité d'événement en fonction de l'évolution des tests.

De la même manière que dans le chapitre précédent, nous pouvons aussi prédire, à partir du modèle conjoint incluant un modèle de survie, la probabilité de survenue de démence en utilisant les mesures longitudinales de cognition y_i et des variables explicatives x_i . Ainsi, à partir des paramètres estimés $\hat{\theta}$, la probabilité de survenue de démence avant le temps t est calculée à partir de l'information sur les tests avant le temps t , $(y_{is})_{s < t}$:

$$\begin{aligned} P(T_i \leq t \mid (y_{is})_{s < t}, x_i; \hat{\theta}) &= \sum_{g=1}^G P(T_i \leq t \mid c_{ig} = 1, x_i; \hat{\theta})P(c_{ig} = 1 \mid (y_{is})_{s < t}, x_i; \hat{\theta}) \\ &= \sum_{g=1}^G (1 - S(t \mid c_{ig} = 1; \hat{\theta})) \times \hat{\pi}_{ig}^{yt} \\ &= \sum_{g=1}^G (1 - e^{-(\zeta_{1g}t)^{\zeta_{2g}} e^{x_i \delta_g}}) \times \hat{\pi}_{ig}^{yt} \end{aligned} \quad (5.15)$$

La probabilité de survenue de démence avant le temps t donnée en (5.15) constitue un outil de détection de démence que l'on peut calculer en tout t à partir de

l'information sur les tests psychométriques avant t .

L'évaluation des capacités prédictives d'un modèle de survie est compliquée par le fait que la réponse observée est un temps de survie censuré (Heagerty et Zheng, 2005). En effet, la sensibilité, la spécificité, la courbe ROC et l'AUC ne sont calculables que pour des temps fixés. De plus, différentes définitions de la sensibilité et de la spécificité ont été proposées en se basant soit sur la survie jusqu'en t , soit sur la probabilité de connaître l'événement en t . Pour évaluer notre outil de détection de démence, nous proposons de construire la courbe ROC associée à la probabilité de démence avant t (définition cumulée/dynamique suivant Heagerty et Zheng, 2005). Parmi tous les sujets ayant connu l'événement avant t et les sujets étant encore à risque d'événement à t , nous calculons, à partir de l'information sur les tests antérieure à t , $(y_{is})_{s < t}$, leur probabilité de démence avant t , $P(T_i^* \leq t \mid (y_{is}), s < t)$. Puis, nous classons en déments les sujets pour lesquels cette probabilité est au dessus d'un seuil ε , nous permettant ainsi de calculer la spécificité et la sensibilité de l'outil de diagnostic en fonction de ε . En faisant varier ε dans l'intervalle $[0, 1]$, nous pouvons tracer la courbe ROC associée à l'outil de détection en t et calculer l'aire sous la courbe ROC en t ($AUC(t)$) qui permet d'évaluer la qualité de l'outil de diagnostic.

5.3 Modélisation semi-paramétrique de la survie

Bien que nous ayons privilégié dans ce chapitre une modélisation paramétrique du délai jusqu'à l'événement pour des raisons de complexité numérique, nous avons aussi envisagé une modélisation plus souple à l'aide de fonctions splines pour modéliser la fonction de risque. Nous présentons dans cette section le modèle de survie basé sur les fonctions splines et décrivons la méthode d'estimation utilisant une vraisemblance pénalisée des observations.

5.3.1 Risque d'événement modélisé par des fonctions splines

Les B-splines sont des fonctions polynomiales par morceaux que l'on combine de manière linéaire pour approcher une fonction inconnue sur un intervalle. Leur méthodologie et leurs applications ont été décrites par De Boor (1978). Dans ce travail, nous utilisons des versions normalisées des B-splines, appelées M-splines. Les M-splines sont particulièrement bien adaptés pour modéliser la fonction de risque. En effet, ce sont des fonctions positives ou nulles qui peuvent approcher le risque $\lambda_0(t)$ et leurs primitives, les I-splines, sont des fonctions monotones croissantes adaptées pour approcher le risque cumulé $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ qui est utilisé dans le calcul de la fonction de survie. Les M-splines et I-splines ont été décrits dans plusieurs travaux (Ramsay, 1988 ; Joly, 1996).

Notons t_1, \dots, t_T , la séquence de noeuds des splines et k l'ordre des splines. Les M-splines et les I-splines d'ordre k sont respectivement des polynômes de degré $k - 1$ et de degré k . A partir de la séquence des T noeuds, m fonctions splines M_i et I_i ($i = 1, m$) sont définies par des équations de récurrence détaillées dans Ramsay (1988) et associées à un coefficient ζ_i à estimer, le nombre de paramètres m à estimer étant $T + k - 2$. Les fonctions de risque $\lambda(\cdot)$ et $\Lambda(\cdot)$ sont ensuite définies comme une combinaison linéaire des m fonctions splines :

$$\lambda(\cdot) = \sum_{i=1}^m \zeta_i M_i \quad (5.16)$$

$$\Lambda(\cdot) = \sum_{i=1}^m \zeta_i I_i \quad \text{avec } \zeta_i \geq 0 \quad (5.17)$$

Une fonction estimée sur une base de splines peut avoir une grande variabilité locale, surtout lorsque le nombre de noeuds est important. En épidémiologie, la fonction de risque n'a généralement pas de raison d'avoir des variations locales importantes. Une manière de limiter les variations locales est de pénaliser la vraisemblance du modèle pour privilégier les fonctions de risque lisses. La pénalisation est basée sur la dérivée seconde de la fonction à lisser. Comme nous utilisons une

pénalisation par la dérivée seconde de la fonction de risque, nous avons opté dans ce travail pour des M-splines d'ordre 4, c'est à dire des polynômes de degré 3 pour décrire la fonction de risque.

5.3.2 Vraisemblance pénalisée

La log-vraisemblance pénalisée est la somme de la log-vraisemblance classique des observations dans le modèle et d'une pénalité qui est d'autant plus grande que la fonction à estimer a des variations locales importantes. Les estimateurs des paramètres obtenus en maximisant la log-vraisemblance pénalisée représentent donc le meilleur compromis entre adéquation aux données et lissage de la courbe à estimer.

Dans notre travail, la fonction à lisser étant la fonction de risque, nous pénalisons la log-vraisemblance par la norme L^2 de la dérivée seconde du risque $\lambda_0(\cdot)$. La log-vraisemblance pénalisée a donc la forme suivante :

$$\mathcal{L}_p(\theta; \kappa) = \mathcal{L}(\theta) + \kappa \int_t \lambda_0''(u; \theta)^2 du \quad (5.18)$$

où κ est le paramètre de lissage qui contrôle le poids donné à la pénalisation par rapport à la log-vraisemblance. L'intégrale $\int_t \lambda_0''(u; \theta)^2 du$ calculée sur le support des temps d'observation est la norme L^2 de la dérivée seconde de la fonction de risque. La log-vraisemblance $\mathcal{L}(\theta)$ a la même forme que celle présentée précédemment en équations (5.9) ou (5.12) à la différence près que le risque $\lambda(T_i; \theta)$ et les fonctions de survie $S(T_i; \theta)$ et $S(T_{0i}; \theta)$ sont définis à partir de fonctions splines par (5.16) et (5.17).

Pour réduire le nombre de paramètres à estimer, nous n'avons pas considéré le modèle de survie stratifié sur les classes latentes dans l'approche semi-paramétrique. Nous estimons la fonction de risque dans la classe de référence et supposons que les risques dans les autres classes sont proportionnels.

5.3.3 Choix du critère de lissage

L'estimation du paramètre de lissage κ représente la difficulté majeure de cette approche par rapport au modèle. On peut choisir de fixer le paramètre de lissage κ , auquel cas la log-vraisemblance pénalisée peut être maximisée de manière classique par l'algorithme de Marquardt. Cependant, choisir κ de manière subjective peut se révéler délicat car son ordre de grandeur dépend beaucoup des données. Il peut donc être plus intéressant d'estimer le paramètre de lissage afin d'obtenir une méthode automatique d'estimation du modèle par vraisemblance pénalisée. La validation croisée permet d'estimer le paramètre κ . Elle consiste à optimiser suivant κ , la somme des contributions individuelles à la log-vraisemblance \mathcal{L}_i de chaque sujet i à partir des estimateurs, $\hat{\theta}_{-i}$, obtenus en excluant le sujet i :

$$V(\kappa) = -\frac{1}{N} \sum_{i=1}^N \mathcal{L}_i(\hat{\theta}_{-i}, \kappa) \quad (5.19)$$

Pour chaque valeur de κ , cette méthode nécessite l'estimation d'autant de modèles que de sujets dans l'échantillon. Elle est donc extrêmement coûteuse en temps de calcul. Une approximation de $V(\kappa)$ a été proposée par O'Sullivan (1988) et étendue aux données tronquées et censurées par Joly (1996). Il s'agit d'approcher les paramètres $\hat{\theta}_{-i}$ à partir de l'estimateur obtenu sur l'échantillon entier $\hat{\theta}$ afin d'éviter l'estimation des N modèles par valeur de κ . Cette technique consiste à approcher la quantité $V(\kappa)$ par un développement du premier ordre de $\mathcal{L}_i(\hat{\theta}_{-i}, \kappa)$ autour de $\mathcal{L}(\hat{\theta}, \kappa)$, la log-vraisemblance à l'optimum pour l'ensemble de l'échantillon :

$$\bar{V}(\kappa) = -\frac{1}{N} \left(\mathcal{L}(\hat{\theta}, \kappa) + \text{trace} \left(\left[\hat{H} - 2\kappa\Omega \right]^{-1} \hat{H} \right) \right) \quad (5.20)$$

où \hat{H} est la matrice Hessienne à l'optimum pour l'ensemble de l'échantillon et Ω est la matrice des dérivées secondes par rapport aux paramètres du terme de pénalisation.

Cette méthode basée sur l'optimisation de $\bar{V}(\kappa)$ nécessite uniquement l'estimation du modèle sur l'échantillon entier pour chaque valeur de κ . L'estimateur $\hat{\kappa}$ est

ensuite obtenu en minimisant $\bar{V}(\kappa)$, soit à l'aide d'une grille de valeurs pour κ soit par un algorithme de minimisation. Cette méthode a été implémentée dans le programme `phmp1` (Joly et al., 1999) pour l'estimation de la fonction de risque modélisée par des splines dans le cas de données tronquées et censurées.

Dans notre approche, estimer le paramètre de lissage en utilisant cette technique nécessiterait l'estimation du modèle conjoint sur toute une grille de valeurs, ce qui n'est pas envisageable étant donné la complexité numérique de notre modèle et la taille de notre échantillon d'application. Nous avons donc opté pour une technique intermédiaire de sélection du paramètre de lissage : nous utilisons le programme `phmp1` sur le modèle de survie sans classe latente et sans données longitudinales. Cela nous donne un ordre de grandeur pour le paramètre κ que nous pouvons utiliser dans le modèle conjoint. Nous pouvons ensuite ajuster la valeur de κ si nécessaire au vu des fonctions de risque estimées.

5.4 Analyse de l'adéquation

Les modèles à classes latentes étant des modèles paramétriques complexes, nous détaillons dans cette section trois méthodes permettant d'évaluer l'adéquation du modèle. La première technique consiste à évaluer la qualité de la classification des sujets dans les classes latentes. La deuxième consiste à confronter les prédictions du modèle mixte aux données observées et la troisième méthode vise à évaluer l'ajustement du modèle de survie.

5.4.1 Qualité de classification

Quelques méthodes ont été proposées pour tester la qualité de la classification dans un modèle à classes latentes (Muthén et al., 2002 ; Lin et al., 2004). En effet, une bonne classification a posteriori indique une bonne adéquation du modèle aux données. Ces méthodes sont basées sur le principe qu'une classification est d'autant

meilleure que les probabilités *a posteriori* d'appartenir aux classes sont proches de 1 et de 0. Muthén et al. (2002) ont ainsi proposé de calculer, pour les sujets classés dans une classe *a posteriori*, la moyenne des probabilités *a posteriori* d'y appartenir (elle doit être la plus proche de 1 possible) et les moyennes des probabilités d'appartenir aux autres classes (elles doivent être les plus faibles possibles). De façon plus générale, certains auteurs (Lin et al., 2004) proposent d'examiner la distribution des probabilités *a posteriori* plutôt que leur moyenne pour évaluer la qualité de classification.

5.4.2 Prédiction dans l'échelle des marqueurs

Pour évaluer si le modèle s'ajuste bien aux données, nous proposons aussi de comparer l'évolution prédite pour chaque marqueur et chaque classe latente à l'évolution observée, en distinguant l'évolution prédite marginale de l'évolution prédite conditionnellement aux effets aléatoires individuels. Dans notre approche, les prédictions issues du modèle mixte ne sont pas directement obtenues dans l'échelle des tests mais elles sont calculées par intégration numérique comme décrit dans le chapitre III (page 69). Les prédictions sont calculées conditionnellement aux classes latentes puis la moyenne sur l'ensemble des sujets d'une classe d'âge est calculée en pondérant par les probabilités *a posteriori* $\hat{\pi}_{ig}^y$ d'appartenir à la classe latente g . Ainsi, nous pouvons calculer dans chaque tranche d'âge, la moyenne pondérée des prédictions pour chaque test et chaque classe latente. Nous obtenons ainsi l'évolution moyenne prédite par test psychométrique et par classe latente que nous pouvons comparer à l'évolution moyenne des observations pondérée par les $(\hat{\pi}_{ig}^y)_{g=1,G}$.

5.4.3 Courbes de survie prédites

Pour évaluer l'adéquation des modèles de survie, nous avons envisagé deux analyses *a posteriori*. La première consiste à évaluer l'hypothèse de risque proportionnel entre les classes latentes par le biais de la courbe de survie prédite spécifique à chaque

classe latente :

$$\hat{S}(t | c = g, X) = e^{-(\zeta_{1g}t)^{\zeta_{2g}} e^{\delta X}} \quad (5.21)$$

Cette courbe est comparée à la courbe de survie issue d'un modèle semi-paramétrique incluant X en variables explicatives et estimé sur le sous-échantillon des sujets classés *a posteriori* dans la classe g .

La deuxième méthode consiste à évaluer si, marginalement aux classes latentes, la fonction de risque paramétrique choisie s'ajuste bien aux données. Pour cela, on calcule la courbe de survie marginale prédite :

$$\hat{S}(t | X) = \sum_{g=1}^G \hat{S}(t | c = g, X) P(c = g | X) \quad (5.22)$$

Cette courbe estimée est comparée à la courbe de survie issue d'un modèle semi-paramétrique estimé sur toutes les données et incluant X en variables explicatives.

Quelle que soit l'analyse, les courbes de survie sont comparées pour chaque combinaison de variables explicatives X .

5.5 Etude de simulation

Cette section est dédiée à une étude de simulation que nous avons réalisée afin d'évaluer la qualité des estimateurs du maximum de vraisemblance du modèle conjoint paramétrique. Nous présentons d'abord la méthode utilisée pour simuler les échantillons puis exposons les résultats de l'étude.

5.5.1 Génération des données

Dans cette étude, nous considérons un modèle à deux classes latentes et simulons deux marqueurs quantitatifs longitudinaux. Le modèle de simulation est le suivant :

$$\begin{aligned} \Lambda_i(t) |_{c_{ig}=1} &= \mu_{0g} + u_{0i} + (\mu_{1g} + u_{1i}) \times t \\ \text{avec } u_i &= \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 = 0.009 & 0 \\ 0 & \sigma_1^2 = 0.001 \end{pmatrix} \right) \\ \text{et } \mu_{01} &= 0.6, \quad \mu_{02} = 0.55, \quad \mu_{11} = 0, \quad \mu_{12} = -0.1 \end{aligned} \tag{5.23}$$

$$\begin{aligned} \lambda(t) |_{c_{ig}=1} &= \zeta_{1g} \zeta_{2g} (\zeta_{1g} t)^{\zeta_{2g}-1} \\ \text{avec } \zeta_{11} &= 0.0225, \quad \zeta_{12} = 0.0576, \quad \zeta_{21} = 3.24, \quad \zeta_{22} = 2.25 \end{aligned} \tag{5.24}$$

$$\begin{aligned} h_k(Y_{ijk}; (\eta_1^{(k)}, \eta_2^{(k)})) &= \Lambda_i(t_{ijk}) + \epsilon_{ijk} \\ \text{avec } \eta_1^{(1)} &= 0.49, \quad \eta_2^{(1)} = 0.81, \quad \eta_1^{(2)} = 0.81, \quad \eta_2^{(2)} = 0.64 \\ \text{et } \epsilon_{ijk} &\sim_{iid} \mathcal{N}(0, \sigma_\epsilon^{2(k)} = 0.0009), \quad k = 1, 2 \end{aligned} \tag{5.25}$$

$$P(c_{i1} = 1) = \pi_1 = 0.75 \quad \text{et} \quad P(c_{i2} = 1) = 1 - \pi_1 = 0.25 \tag{5.26}$$

Les échantillons simulés sont constitués de 200 sujets et pour chaque sujet, la simulation des données se fait en trois temps. Nous commençons par simuler l'appartenance à la classe latente $(c_{ig})_{g=1,2}$ suivant une distribution de Bernoulli de probabilité π_1 , le temps d'entrée dans l'étude T_{0i} suivant une distribution uniforme sur $[0, 10]$, le temps d'événement T_i^* suivant la loi de Weibull de la classe d'appartenance et le temps de censure C_i suivant une loi uniforme sur $[T_{0i}, 30]$. On peut ainsi construire le couple (T_i, E_i) .

Ensuite, les temps de mesure des marqueurs sont définis par $t_{i1k} = T_{0i}$ et $t_{ijk} = \min(t_{ij-1k} + 3; T_i)$ tant que $t_{ij-1k} < T_i$. Le nombre et les temps de mesure sont les mêmes pour les deux marqueurs d'un même sujet mais ils diffèrent d'un sujet à l'autre.

Enfin, la troisième étape consiste à simuler les mesures répétées des 2 marqueurs longitudinaux y_{i1} et y_{i2} à partir des temps de mesure comme décrit dans la section 3.2.1 du chapitre III.

5.5.2 Résultats de l'étude de simulation

Sur les 200 simulations réalisées, les sujets avaient en moyenne 5.53 mesures répétées dans la première classe et 4.14 dans la deuxième classe, à risque plus élevé d'événement. Les sujets entraient dans l'étude en moyenne au temps 5.01 dans la première classe et 4.98 dans la deuxième classe et sortaient de l'étude en moyenne au temps 17.09 dans la première classe et 12.86 dans la deuxième classe. Enfin, 7.3% connaissaient l'événement dans la première classe contre 56.2% dans la deuxième classe, ce qui représentait 19.4% d'événement dans l'échantillon. Le tableau 5.1 donne les résultats des 200 simulations en termes d'estimations moyennes, biais et taux de couverture.

La probabilité π_1 ainsi que les paramètres des risques d'événement dans chaque classe latente $(\zeta_{1g}, \zeta_{2g})_{g=1,2}$ sont correctement estimés. Le biais moyen ne dépasse pas 5% et les taux de couverture sont proches de 95%. Par cette étude de simulation, il semble donc que le modèle nonlinéaire conjoint à classes latentes incluant un modèle de survie à entrée retardée donne de bons résultats. Néanmoins, il est à noter que nous avons choisi ici des paramètres d'évolution des marqueurs n'induisant pas de troncatures aux bornes 0 et 1 des lois Beta. On peut imaginer que, de la même manière que dans le modèle non conjoint, des paramètres induisant des troncatures puissent donner de moins bons résultats sur les estimations des paramètres longitudinaux.

5.6 Application : profils de déclin cognitif associés au risque de démence

5.6.1 Objectif de l'étude

Cette application avait pour objectif d'étudier l'association entre l'évolution cognitive représentée par le facteur commun à trois tests psychométriques (IST15,

TAB. 5.1 – Estimation moyenne, biais relatif, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture des 200 simulations de modèle conjoint avec modèle de survie réalisées sur des échantillons de 200 sujets.

paramètre	valeur simulée	estimation moyenne	biais relatif	ET empirique	ET asymptotique moyen	taux de couverture
π_1	1.099	1.052	-0.042	0.242	0.239	0.92
χ_{11}	0.15	0.147	-0.022	0.022	0.019	0.91
χ_{12}	1.8	1.876	0.042	0.529	0.371	0.95
χ_{21}	0.24	0.240	<0.001	0.013	0.012	0.92
χ_{22}	1.5	1.569	0.046	0.185	0.169	0.93
μ_{01}	0.60	0.601	0.002	0.024	0.023	0.96
μ_{02}	0.55	0.553	0.005	0.026	0.028	0.97
μ_{11}	0	0.0001	—	0.004	0.004	0.93
μ_{12}	-0.100	0.097	-0.032	0.012	0.011	0.94
$\kappa_1^{(1)}$	-0.251	0.263	0.047	0.081	0.074	0.95
$\kappa_2^{(1)}$	-2.358	-2.328	0.013	0.077	0.072	0.94
$\kappa_1^{(2)}$	0.118	0.098	-0.166	0.093	0.080	0.96
$\kappa_2^{(2)}$	-2.383	-0.235	-0.014	0.090	0.080	0.94
σ_0	0.095	0.091	-0.043	0.010	0.009	0.91
σ_1	0.032	0.030	-0.042	0.004	0.004	0.91
$\sigma_\epsilon^{(1)}$	0.030	0.029	-0.033	0.003	0.002	0.95
$\sigma_\epsilon^{(2)}$	0.030	0.029	-0.035	0.003	0.002	0.92

BVRT et DSST) et le risque de démence en fonction de l'âge.

5.6.2 Les données d'application

A partir des 3777 sujets de la cohorte PAQUID, nous avons sélectionné les sujets non déments et vivant en Gironde à l'inclusion (T0). Pour les sujets ayant eu un diagnostic de démence pendant le suivi, leurs mesures aux tests psychométriques collectées après le diagnostic ont été éliminées. Enfin, l'échantillon inclut uniquement les sujets ayant au moins une mesure répétée à chacun des trois tests psychométriques pendant le suivi, c'est à dire entre les visites T0 et T13. Nous avons ainsi sélectionné un échantillon de 2383 sujets incluant 355 déments incidents et 2028 sujets non-déments au dernier suivi.

Cet échantillon est constitué de 1382 (58%) femmes et de 692 (29%) sujets sans CEP. L'âge à l'entrée dans l'étude et l'âge à la dernière visite sont décrits dans le tableau 5.2. L'âge moyen au dernier suivi pour les sujets non-déments est plus faible que pour les sujets déments car 16.8% des sujets non-déments n'ont été vus qu'à la visite T0 (voir tableau 5.3).

5.6.3 Spécification du modèle conjoint

Nous avons considéré une évolution quadratique de l'âge pour le facteur commun afin de prendre en compte des déclin cognitifs non linéaires. Le modèle inclut les deux variables explicatives usuelles : le sexe et le niveau d'éducation en deux classes (obtention du certificat d'études primaires (CEP) ou non). Elles sont incluses uniquement dans les probabilités d'appartenir aux classes latentes et dans le risque de démence après ajustement sur les classes afin de décrire les profils de déclin uniquement en fonction de l'âge. Néanmoins, comme nous avons conservé les mesures des tests psychométriques en T0, le modèle longitudinal comporte un effet propre à la

TAB. 5.2 – Age d’entrée dans l’étude et âge de sortie de l’étude en fonction de la survenue d’une démence incidente dans l’échantillon PAQUID (2383 sujets)

	Effectif	Moyenne	Ecart-type
Age à l’entrée dans l’étude :			
Déments	355	76.18	5.82
Non déments	2028	73.88	6.37
Total	2383	74.21	6.35
Age de démence ou de censure :			
Déments	355	85.04	5.77
Non déments	2028	80.25	6.83
Total	2383	80.97	6.90

TAB. 5.3 – Suivi de sortie d’étude suivant le statut démentiel dans l’échantillon de 2383 sujets

Dernier suivi	Nombre de Déments	Nombre de non-déments	Total
0	0	340	340
1	10	234	244
3	45	236	281
5	41	268	309
8	45	181	226
10	100	198	298
13	114	571	685
Total	355	2028	2383

première visite qui permet de prendre en compte l'effet de primo-passation des tests. En effet, une amélioration générale aux tests psychométriques a été observée dans PAQUID entre les visites T0 et T1 (Jacqmin-Gadda et al., 1997a), cette amélioration pouvant être expliquée par le stress de la première évaluation cognitive à T0.

Le modèle conjoint considéré ici peut donc être résumé de la manière suivante :

$$\Lambda_i(t) |_{c_{ig}=1} = \mu_{0g} + u_{0gi} + (\mu_{1g} + u_{1gi}) \times \text{AGE}_i + (\mu_{2g} + u_{2gi}) \times \text{AGE}_i^2 + \beta \mathbf{I}_{t=T_{0i}}$$

$$\lambda(t) |_{c_{ig}=1} = \lambda_{0g}(t) e^{\text{SEXE}_i \delta_1 + \text{CEP}_i \delta_2}$$

$$h_k(Y_{ijk}; (\eta_1^{(k)}, \eta_2^{(k)})) = \Lambda_i(t_{ijk}) + \epsilon_{ijk}$$

$$P(c_{ig} = 1) = \frac{e^{\xi_{0g} + \text{SEXE}_i \xi_{1g} + \text{CEP}_i \xi_{2g}}}{\sum_{l=1}^G e^{\xi_{0l} + \text{SEXE}_i \xi_{1l} + \text{CEP}_i \xi_{2l}}}$$

où $u_{ig} = (u_{0gi}, u_{1gi}, u_{2gi})^T \sim \mathcal{N}(\mathbf{0}, \omega^2 B)$ et $\lambda_{0g}(t)$ est défini suivant un modèle de Weibull.

5.6.4 Résultats

Analyse préliminaire : choix du modèle de survie

Cette analyse préliminaire avait pour objectif de sélectionner la meilleure spécification du modèle de survie : soit un modèle stratifié sur les classes, soit un modèle à risques proportionnels entre les classes. Nous avons pour cela estimé, pour chacune des deux spécifications, un modèle conjoint légèrement simplifié (sans variable explicative et avec $\omega_g = 1, \forall g$) et pour un nombre de classes latentes variant de 1 à 5. Le tableau 5.4 présente la log-vraisemblance et les critères AIC et BIC pour les différents modèles estimés.

Le modèle conjoint avec un risque de démence stratifié sur les classes latentes, bien que plus souple que le modèle avec des risques de démence proportionnels entre

TAB. 5.4 – Adéquation du modèle conjoint non ajusté suivant la spécification du risque de démence et le nombre de composantes G (nombre de paramètre p , log-vraisemblance \mathcal{L} et les critères AIC et BIC)

G	Modèle à risques proportionnels				Modèle stratifié			
	p	\mathcal{L}	AIC	BIC	p	\mathcal{L}	AIC	BIC
2	28	-67712.01	135480.0	135641.8	29	-67711.6	135481.3	135648.8
3	33	-67630.72	135327.4	135518.1	35	-67629.5	135329.0	135531.2
4	38	-67596.62	135269.2	135488.7	41	-67596.2	135274.4	135511.3
5	43	-67579.45	135244.9	135493.3	47	-67573.0	135240.1	135511.5

les classes, améliore très peu l'adéquation aux données. Les vraisemblances des deux modèles sont très proches quel que soit le nombre de classes latentes, et l'AIC et le BIC du modèle à risques proportionnels sont toujours meilleurs que ceux du modèle stratifié correspondant (à l'exception de l'AIC pour le modèle à 5 classes latentes). De plus, le critère du BIC, conduit dans les deux cas à retenir un modèle à 4 classes.

Dans ce modèle à 4 classes latentes, nous avons tenté de comparer les risques obtenus par le modèle de Weibull à ceux obtenus avec une fonction de risque décrite par des splines. Le modèle utilisant des splines était néanmoins trop compliqué et l'estimation n'a pas abouti au bout de 500 iterations.

Description du modèle conjoint avec variables explicatives

Après avoir choisi un modèle à risques proportionnels pour le risque de démence, nous avons estimé le modèle conjoint complet pour un nombre de classes latentes variant de 1 à 6 et, d'après le critère du BIC, le modèle à 5 classes a été retenu (voir tableau 5.5). Les évolutions prédites dans les 5 classes latentes sont représentées en figure 5.1.

On observe un déclin des fonctions cognitives dans toutes les classes latentes mais

TAB. 5.5 – Adéquation du modèle conjoint ajusté avec le risque de démence proportionnel d’une classe à l’autre (nombre de classes latentes G , nombre de paramètre p , log-vraisemblance \mathcal{L} , critère du BIC et aire sous la courbe ROC à plusieurs âges (AUC(t)) ; pour chaque âge, N est le nombre total de sujets utilisés et n le nombre de sujets déments.

G	p	\mathcal{L}	BIC	AUC					
				70 ans (N=619) (n=48)	75 ans (N=868) (n=111)	80 ans (N=872) (n=179)	85 ans (N=636) (n=246)	90 ans (N=437) (n=306)	95 ans (N=369) (n=344)
2	34	-67390.49	135045.38	0.710	0.747	0.751	0.743	0.796	0.808
3	42	-67262.12	134850.83	0.755	0.786	0.817	0.843	0.865	0.849
4	50	-67186.32	134761.43	0.703	0.765	0.809	0.844	0.883	0.885
5	58	-67141.29	134733.60	0.655	0.763	0.819	0.847	0.896	0.910
6	66	-67115.15	134743.53	0.656	0.754	0.824	0.849	0.899	0.912

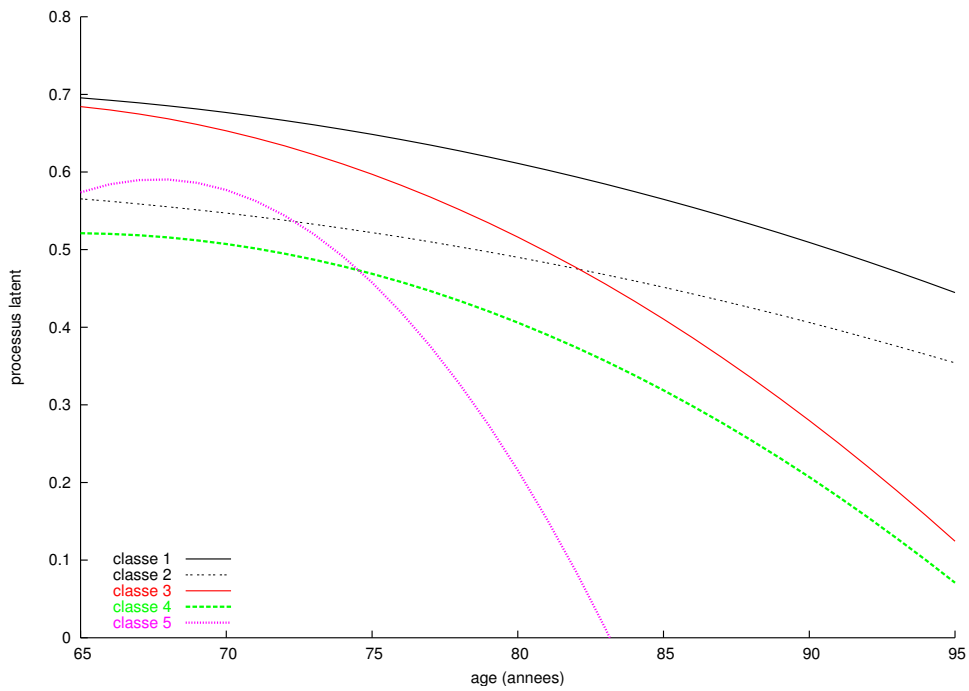


FIG. 5.1 – Evolutions prédites du processus latent dans les 5 classes latentes du modèle conjoint.

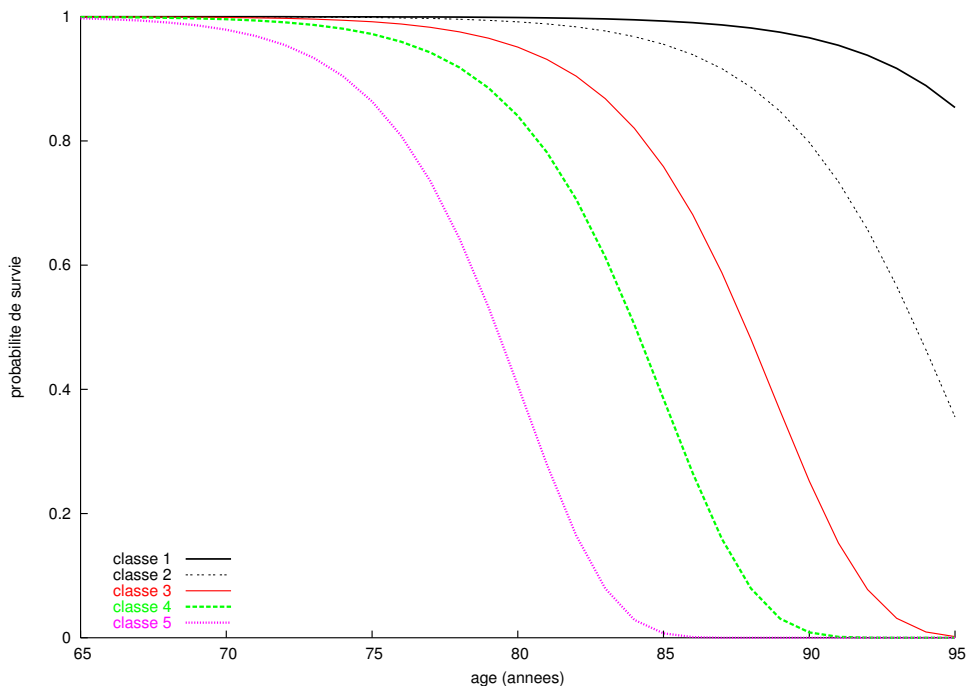


FIG. 5.2 – Fonctions de survie prédites dans les 5 classes latentes du modèle conjoint.

avec des niveaux initiaux et des formes et intensités de déclin différentes. Ainsi, les classes 1 et 3 partent d'un niveau initial à 65 ans très élevé mais la classe 3 a un déclin beaucoup plus marqué avec l'âge que la classe 1. Avec un niveau initial à 65 ans plus faible, on observe la même différence pour les classes 2 et 4 : la classe 4 a un déclin avec l'âge beaucoup plus prononcé. Enfin, la classe 5 est caractérisée par un effondrement de la cognition, le niveau minimum de cognition étant atteint autour de 83 ans. Par rapport à la classe 1 prise en référence, la classe 2 a une variabilité plus faible d'un facteur $\omega_2^2 = 0.822^2 = 0.675$ ($p = 0.0066$) alors que la variabilité dans la classe 5 est nettement plus grande que dans la classe 1 ($\omega_5^2 = 1.768^2 = 3.124$, $p = 0.0009$). La variance des effets aléatoires dans les classes 3 et 4 n'est pas différente de celle dans la classe 1 ($\omega_3^2 = 0.968^2 = 0.938$, $p = 0.383$ et $\omega_4^2 = 0.985^2 = 0.970$, $p = 0.395$).

Les fonctions de survie associées aux classes latentes (figure 5.2) montrent que les risques de démence sont très différents suivant les classes latentes. Les classes 1 et 2 ont des risques de démence assez faible avec des probabilités de démence à 85 ans de moins de 1% et de 4.4% respectivement. Les classes 3 et 4 sont ensuite associées à des risques plus importants de démence, les probabilités de démence étant respectivement de 24% et 62% à 85 ans et atteignant 1 à 91 et 95 ans. Enfin, la classe 5 est associée au plus grand risque de démence avec une probabilité de démence qui atteint 1 dès 85 ans.

D'après les estimations du modèle présentées dans le tableau 5.6, ces classes latentes sont uniquement expliquées par le niveau d'études, le sexe n'étant pas prédicteur de l'appartenance aux classes latentes. Ainsi, les sujets n'ayant pas le CEP ont un risque extrêmement élevé d'être dans les classes 2, 4 et 5 plutôt que dans la classe 1. En fait, l'obtention du CEP discrimine presque parfaitement les classes latentes. Les classes 1 et 3 sont des classes de sujets ayant le CEP, la classe 1 étant associée à un risque très faible de démence et la classe 3 à un risque plus

TAB. 5.6 – Estimations (écart-type) du modèle conjoint ajusté à 5 classes latentes.

Paramètre	Estimations communes	Estimations spécifiques aux classes latentes				
		1	2	3	4	5
<u>Probabilité d'appartenir aux classes latentes :</u>						
Intercept		0	4.338	-10.944	3.714	2.104
			(0.942)	(88.284)	(0.947)	(0.975)
Sexe [†] ($p = 0.084$)		0	-0.991	-0.155	-0.503	-0.245
			(0.307)	(0.230)	(0.392)	(0.315)
CEP [‡] ($p < 0.0001$)		0	-5.535	9.932	-7.163	-4.435
			(0.983)	(88.286)	(1.272)	(0.960)
<u>Risque de démence :</u>						
$\sqrt{\zeta_1}$	0.099 (0.001)					
$\sqrt{\zeta_2}$	5.302 (0.284)					
δ_{0g}		1	1.879	3.686	4.927	6.572
			(0.337)	(0.337)	(0.418)	(0.401)
Sexe [†]	-0.129 (0.145)					
CEP [‡]	0.787 (0.269)					
<u>Modèle longitudinal :</u>						
Intercept		0.696	0.565	0.684	0.521	0.574
		(0.0096)	(0.012)	(0.016)	(0.018)	(0.031)
$\hat{\text{Age}}$		-0.0289	-0.0300	-0.0374	-0.0034	0.129
		(0.0075)	(0.012)	(0.0223)	(0.0197)	(0.057)
$\hat{\text{Age}}^2$		-0.018	-0.0135	-0.0497	-0.0489	-0.245
		(0.0026)	(0.0039)	(0.111)	(0.0083)	(0.032)
Première visite T0	-0.017 (0.001)					

[†] Femmes en référence

[‡] Sujets n'ayant pas le CEP en référence

élevé de démence. Par contre, les classes 2, 4 et 5 associées à des risques croissants de démence sont principalement constituées de sujets n'ayant pas obtenu le CEP.

Après ajustement sur les classes latentes, le CEP a un effet significatif sur le risque de démence. Si l'hypothèse d'indépendance conditionnelle entre démence et évolution des marqueurs est respectée, c'est à dire si l'ajustement sur les classes latentes dans le modèle de survie (5.3) permet de prendre en compte totalement la dépendance entre risque de démence et l'évolution cognitive, on peut interpréter ce résultat comme suit : à niveau cognitif égal, les sujets ayant le CEP ont un risque plus élevé d'être diagnostiqués déments que les sujets n'ayant pas le CEP.

Classification et probabilité *a posteriori* de démence

Quelle que soit la classification *a posteriori*, la proportion de sujets déments dans chaque classe latente est très variable (tableau 5.7) et suit à peu près le gradient des fonctions de survie estimées (figure 5.2). Le tableau 5.8 confirme que l'obtention du CEP discrimine très bien les classes latentes : les sujets classés dans les classes 1 et 3 ayant tous le CEP et la plupart des sujets classés dans les classes 2 et 4 n'ayant pas le CEP. Peu de différence est observée dans les classes suivant le sexe, excepté pour la classe 2 qui inclut une part importante de femmes comparée aux autres classes.

Pour évaluer la probabilité de démence estimée par le modèle comme outil de détection de la démence, nous avons calculé l'AUC associée à la courbe ROC pour différents âges (70, 75, 80, 85, 90 et 95 ans) et pour les modèles de 2 à 6 classes latentes (tableau 5.5). Les AUC varient avec l'âge auquel elles sont calculées. A partir de 80 ans, l'AUC augmente avec le nombre de classes. Par contre, avant 80 ans, l'AUC du modèle à 5 classes est légèrement moins bon que celles des modèles à 3 et 4 classes mais il faut noter que le nombre de déments est très faible dans ces

TAB. 5.7 – Répartition des sujets déments et non-déments dans chaque classe latente selon la classification *a posteriori* utilisée : soit conditionnellement aux observations sur les tests psychométriques et la survenue de démence, soit conditionnellement aux observations sur les tests psychométriques uniquement.

Classification	Répartition	Classe latente				
		1	2	3	4	5
Tests + Démence	Déments	42 (3.3%)	62 (9.5%)	106 (53.8%)	93 (49.7%)	52 (71.2%)
	Non déments	1232 (96.7%)	590 (90.5%)	91 (46.2%)	94 (50.3%)	21 (28.8%)
	Total	1274	652	197	187	73
Tests uniquement	Déments	82 (6.4%)	91 (14.1%)	78 (39.4%)	74 (39.2%)	30 (51.7%)
	Non déments	1209 (93.6%)	556 (85.9%)	120 (60.6%)	115 (60.8%)	28 (48.3%)
	Total	1291	647	198	189	58

TAB. 5.8 – Répartition des hommes/femmes et des sujets avec/sans CEP dans la classification *a posteriori* sachant les observations sur les tests psychométriques.

Variable		Classe latente				
		1	2	3	4	5
Femmes	N	663	475	113	101	30
	%	51.4	73.4	57.1	53.4	51.7
CEP	N	1291	162	198	1	39
	%	100	25.0	100	0.5	67.2
Total		1291	647	198	189	58

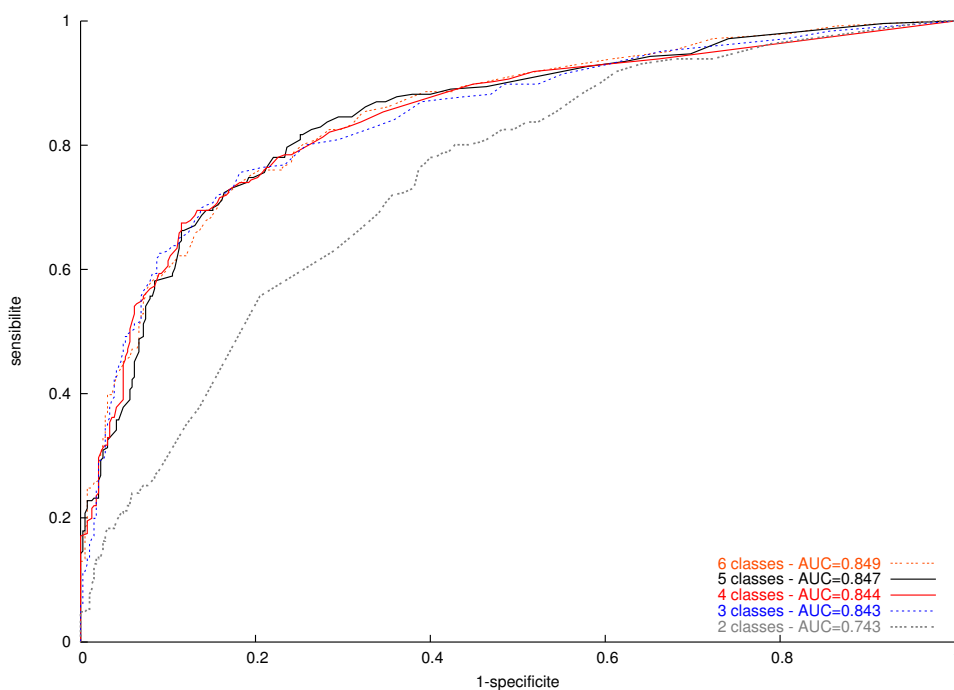


FIG. 5.3 – Courbe ROC associée à la probabilité de démence à 85 ans pour les modèles de 2 à 6 classes latentes.

classes d'âge (tableau 5.5). On peut noter que le modèle à 6 classes donne des AUC très similaires au modèle à 5 classes.

5.6.5 Adéquation du modèle

Nous avons évalué l'adéquation du modèle conjoint final incluant les 5 classes de mélange suivant les méthodes présentées en section 5.4.

Qualité de classification

Le tableau 5.9 présente, pour chaque classe *a posteriori*, les moyennes des probabilités d'appartenir à cette classe et aux 4 autres. Quelle que soit la classe latente, la moyenne des probabilités d'appartenir à la classe dans laquelle les sujets sont classés se situe entre 0.68 et 0.78, c'est à dire très nettement au dessus des moyennes dans les autres classes. Néanmoins, les termes non-diagonaux indiquent que les sujets de

TAB. 5.9 – Moyennes des probabilités *a posteriori* d'appartenir à chaque classe latente suivant la classification finale

Classe d'appartenance	nombre de sujets	Moyenne des probabilités d'appartenir à chaque classe				
		1	2	3	4	5
1	73	0.779	0.083	0.126	0.003	0.009
2	187	0.073	0.723	0.040	0.140	0.024
3	197	0.154	0.106	0.681	0.038	0.021
4	652	0.001	0.204	0.002	0.727	0.066
5	1274	0.012	0.044	0.074	0.097	0.773

la classe 1 avaient une probabilité non négligeable d'être dans la classe 3 (moyenne de 0.126) et inversement (moyenne de 0.154). De la même manière, les sujets dans la classe 2 avaient une probabilité non négligeable d'être dans la classe 4 (moyenne de 0.140) et inversement (moyenne de 0.204). Par contre, les sujets de la classe 5 avaient des probabilités très faibles d'appartenir aux autres classes. A travers cette analyse, nous voyons que la classification *a posteriori* est claire dans l'ensemble bien que la discrimination entre les classes 1 et 3 d'une part et les classes 2 et 4 d'autre part soit plus ambiguë.

Évolutions prédites des tests psychométriques

A partir du modèle final, nous avons calculé la prédiction marginale et conditionnelle de chaque observation comme décrit en section 5.4.2, puis nous avons regroupé les observations par tranche d'âge pour obtenir des courbes d'évolution prédites pour chaque test et chaque classe latente. Les 15 courbes (5 classes \times 3 tests) des évolutions prédites marginales sont représentées en figure 5.4 et les 15 courbes des évolutions prédites spécifiques aux sujets sont représentées en figure 5.5.

Si les évolutions prédites marginales sont très proches des évolutions observées

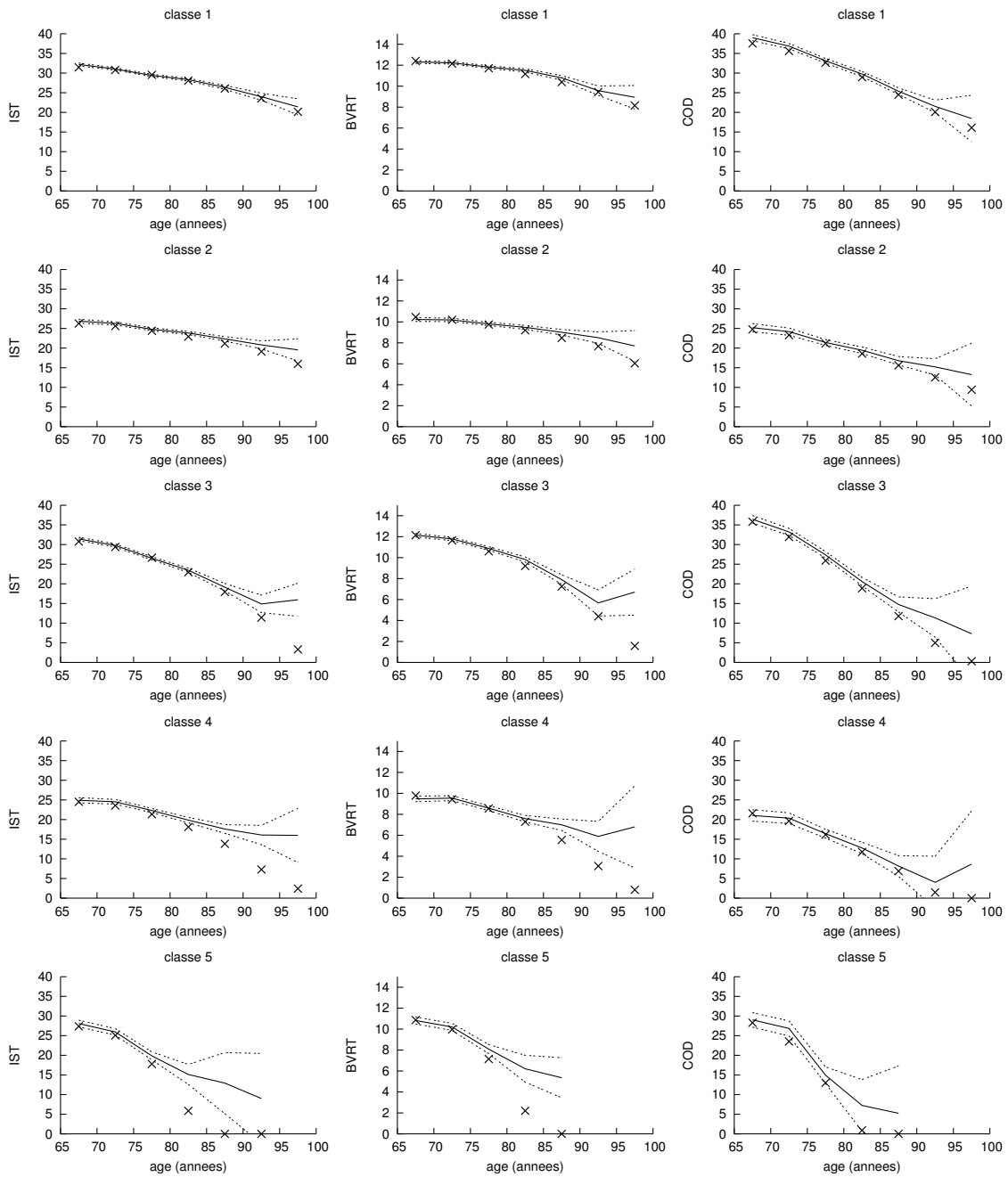


FIG. 5.4 – Evolution moyenne prédite marginale (×) et évolution observée pondérée (en trait plein) avec son intervalle de confiance (en pointillés) pour chaque classe latente et chaque test psychométrique.

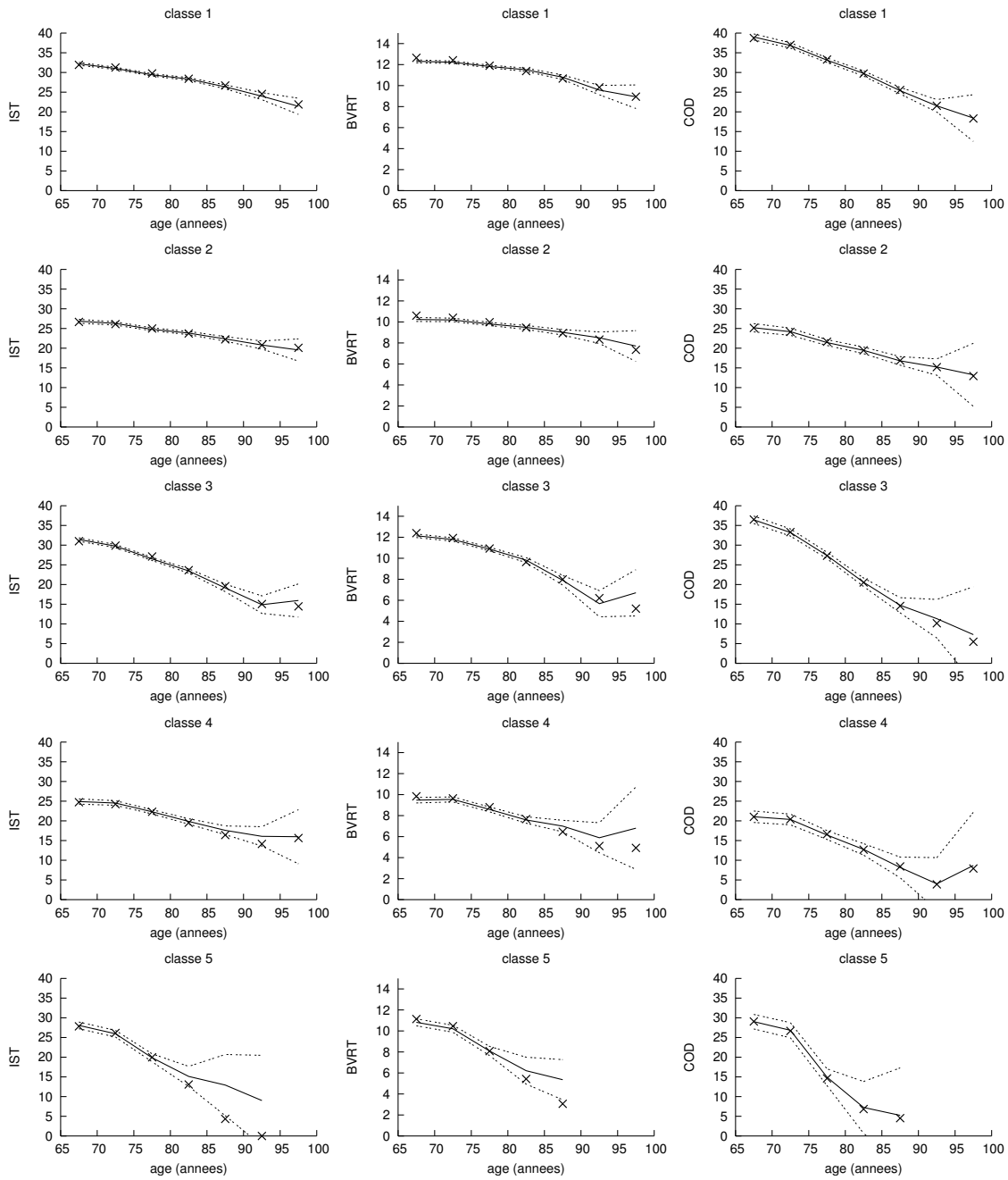


FIG. 5.5 – Evolution moyenne prédite conditionnellement aux effets aléatoires individuels (×) et évolution observée pondérée (en trait plein) avec son intervalle de confiance (en pointillés) pour chaque classe latente et chaque test psychométrique.

pour les classes 1 et 2, elles s'écartent nettement des évolutions observées à partir de 85 ans pour les classes 3, 4 et 5, c'est à dire les classes associées à un risque plus élevé de démence. Ces écarts peuvent être expliqués par les données manquantes. En effet, on sait que les sujets atteints cognitivement ont plus tendance à refuser de passer les tests psychométriques que les autres sujets et de là, les sujets restant dans l'étude ont tendance à avoir un meilleur niveau cognitif que ceux qui en sont sortis. Les évolutions prédites conditionnelles qui incluent les prédictions des effets aléatoires s'ajustent beaucoup mieux aux données observées. Cela signifie que la sortie d'étude est liée à l'évolution des tests mais cela n'implique pas nécessairement qu'elle est non-ignorable. En effet, elle est associée aux effets aléatoires qui sont prédits en fonction des observations. La comparaison des estimations et des observations sur les données disponibles ne permet pas de savoir si les données manquantes sont ignorables mais elle montre un bon ajustement du modèle aux données observées.

Courbes de survie prédites

A partir des estimations issues du modèle final, nous avons calculé deux types de courbes de survie. Nous avons d'abord calculé la courbe de survie prédite marginalement aux classes latentes et l'avons comparée à la courbe de survie estimée de façon semi-paramétrique sur toutes les données par le programme `phmp1` (Joly et al., 1999). La figure 5.6 représente ces deux courbes de survie pour chacune des 4 combinaisons de variables explicatives : femmes sans CEP, femmes avec CEP, hommes sans CEP et hommes avec CEP. Pour chacune des combinaisons, les deux courbes sont très proches indiquant un bon ajustement de la fonction de risque Weibull aux données.

Nous avons ensuite calculé la courbe de survie spécifique à chaque classe latente pour chacune des 4 combinaisons de variables explicatives et l'avons comparée à la courbe de survie prédite par un modèle semi-paramétrique (modèle de Cox) incluant les deux variables explicatives et estimé sur le groupe de sujets classés *a posteriori*

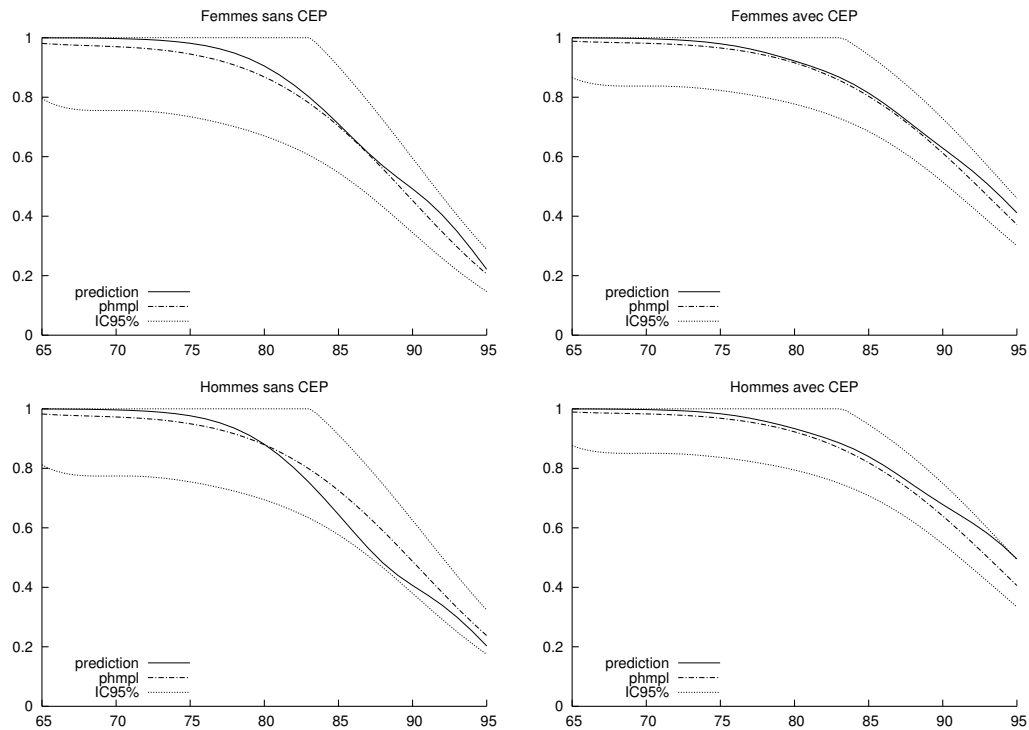


FIG. 5.6 – Courbe de survie prédite marginalement aux classes latentes et courbe de survie estimée semi-paramétriquement par le programme `phmp1` sur tout l'échantillon avec son intervalle de confiance à 95% pour chaque combinaison de variables explicatives (sexe et CEP).

dans la classe considérée. Les figures 5.7 et 5.8 donnent respectivement les courbes de survie chez les femmes et chez les hommes suivant la classe latente et le niveau d'études. Comme certaines classes n'incluent aucun sujet ayant le CEP (classe 4) ou aucun sujet n'ayant pas le CEP (classes 1 et 3), nous n'avons pas représenté les courbes de survie associées à ces combinaisons dans les classes correspondantes. A part un léger écart chez les hommes sans CEP dans la classe 2 sans doute lié au très faible nombre d'évènements (4 ou 2 suivant le CEP), ces courbes de survie indiquent un bon ajustement par classe et combinaison de variable explicative. Ceci montre que l'hypothèse de proportionnalité des risques entre les classes latentes est valide.

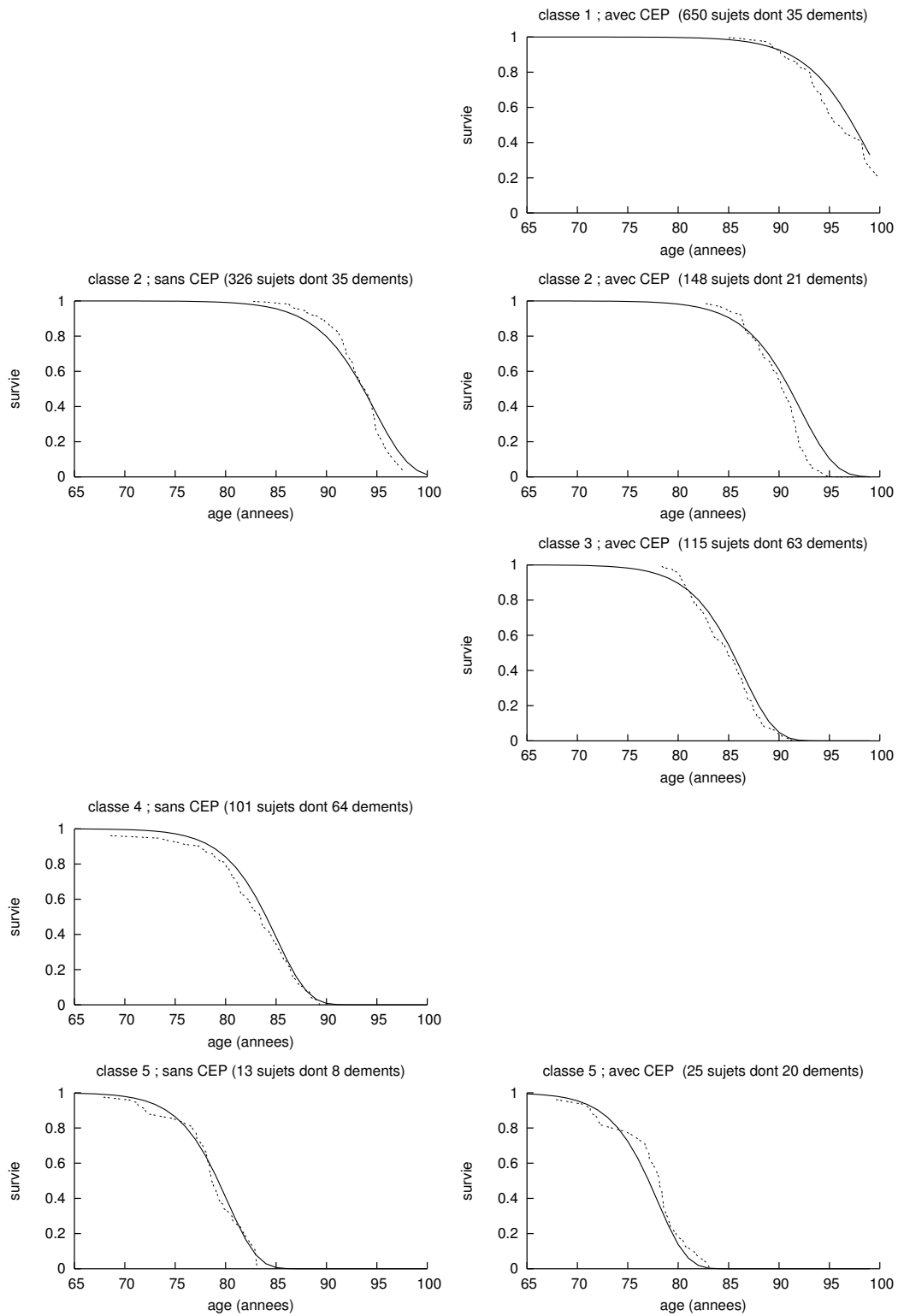


FIG. 5.7 – Courbe de survie prédite dans chacune des 5 classes latentes (en trait plein) et courbe de survie estimée par un modèle de Cox parmi les sujets classés *a posteriori* dans la classe latente correspondante (en pointillés) chez les femmes, suivant leur niveau d'études.

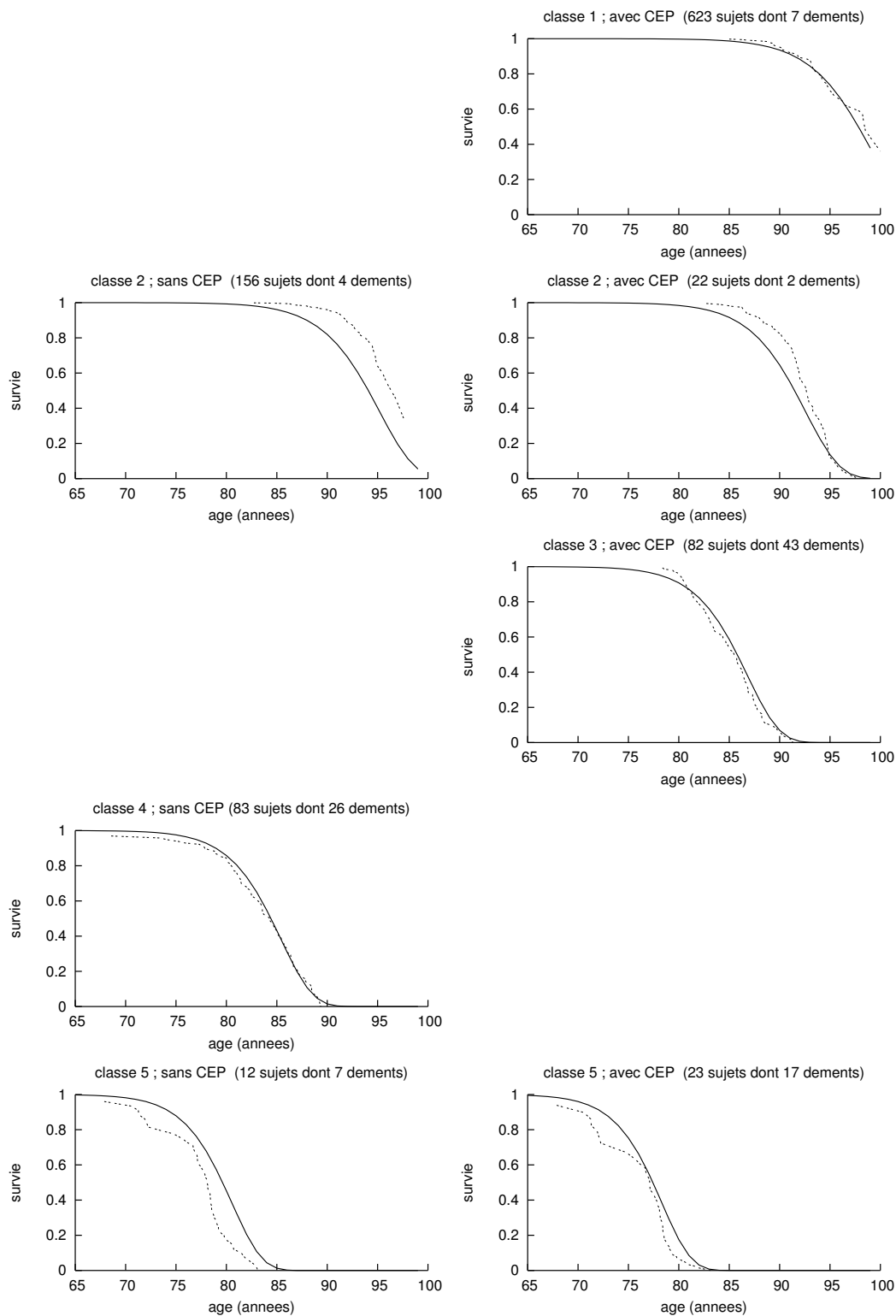


FIG. 5.8 – Courbe de survie prédite dans chacune des 5 classes latentes (en trait plein) et courbe de survie estimée par un modèle de Cox parmi les sujets classés *a posteriori* dans la classe latente correspondante (en pointillés) chez les hommes, suivant leur niveau d'études.

Chapitre 6

Discussion et perspectives

Nous avons développé dans ce travail de thèse une méthodologie traitant à la fois des données longitudinales multivariées, des données hétérogènes et des événements cliniques. Nous avons évalué la performance des modèles proposés par le biais d'études de simulation et d'études d'adéquation aux données réelles mais avons aussi mis en avant leur intérêt épidémiologique au travers d'applications sur l'étude du vieillissement cognitif et de la survenue de démence. Dans ce chapitre, nous souhaitons revenir sur les points forts de cette approche mais aussi sur ses limites et les développements futurs.

6.1 Forces de l'approche

6.1.1 Modélisation de plusieurs marqueurs non Gaussiens

Comme nous l'avons vu dans le chapitre II, la modélisation conjointe de plusieurs marqueurs a reçu un intérêt croissant ces dernières années. En effet, prendre en compte conjointement plusieurs marqueurs permet d'augmenter la puissance de l'étude et éventuellement de réduire les biais si l'échantillon comporte des données manquantes. Dans l'étude du vieillissement cognitif, certains sujets refusent de répondre à des tests psychométriques et ce refus peut être informatif car lié au niveau

cognitif réel du sujet. Lorsqu'à une visite, un sujet ne répond pas à tous les tests psychométriques, modéliser conjointement plusieurs tests permet d'éviter la perte d'information sur le niveau cognitif du sujet à cette visite.

Un deuxième avantage à modéliser conjointement plusieurs marqueurs est qu'aucun des marqueurs n'est privilégié, l'évolution décrite étant celle du processus latent sous-jacent à tous les marqueurs. Dans l'étude du vieillissement cognitif, bien qu'ils soient tous des mesures de la cognition, les tests psychométriques ont des spécificités qui peuvent influencer sur les résultats obtenus. Nous avons en particulier montré, dans la deuxième application du chapitre III, des différences de réussite aux tests suivant le sexe, et des évolutions au cours du temps différentes suivant le niveau d'études. Par la modélisation de leur facteur commun, nous sommes en mesure de décrire le processus cognitif latent et les spécificités des tests psychométriques. Bien que l'idée d'un processus latent représentant la cognition et sous-jacent aux tests psychométriques soit née il y a une dizaine d'années en neuropsychologie et ait reçu depuis un intérêt croissant (Fabrigoule et al., 1998 ; Sliwinski et al., 2003), à notre connaissance, aucune approche n'avait pour l'instant été proposée pour modéliser l'évolution de plusieurs tests psychométriques et de leur facteur commun.

Une particularité dans l'étude du vieillissement est que les tests psychométriques, bien que considérés comme quantitatifs, sont souvent bornés et peuvent pâtir d'effets plafond et plancher (Morris et al., 1999). Jusqu'à présent, ce type de données avait été analysé soit par des méthodes classiques pour données quantitatives sous l'hypothèse de normalité (Joseph et al., 1999 ; Hall et al., 2001 ; Jacqmin-Gadda et al., 2006) soit par des modèles à seuils pour données ordinales (Ganiayre et al., 2006), ces derniers modèles prenant en compte parfaitement les effets plafond ou plancher des tests psychométriques mais présentant des problèmes numériques importants. Nous avons donc proposé dans ce travail une alternative basée sur des transformations non-linéaires des tests psychométriques. Outre donner un cadre souple pour la modélisation conjointe de plusieurs marqueurs, ces transformations estimées nous ont permis de décrire les propriétés métriques des tests psychométriques dans

la première application du chapitre III, et notamment de recommander l'utilisation de l'IST plutôt que du MMSE ou du BVRT dans le cas où, dans une population hétérogène, un seul test psychométrique peut être réalisé.

6.1.2 Hétérogénéité des évolutions et association avec la survenue d'un événement

Verbeke et Lesaffre (1996) ont montré que, dans le cas d'une population hétérogène, faire l'hypothèse d'une distribution normale des effets aléatoires dans un modèle mixte pouvait entraîner des biais sur les estimations, et qu'il était donc préférable de prendre en compte cette hétérogénéité par des modèles de mélange à effets aléatoires. Dans le vieillissement cognitif, il est acquis qu'il existe des sous-populations ayant des évolutions différentes, ces sous-populations correspondant par exemple aux sujets ayant un vieillissement normal et aux sujets ayant un vieillissement pathologique (Sliwinski et al., 1996). Dans un travail précédent ayant pour objectif de proposer un algorithme d'estimation des modèles de mélange pour données longitudinales (Proust et Jacqmin-Gadda, 2005), nous avons étudié l'hétérogénéité des déclin d'un test psychométrique. Cette approche mettait en évidence un petit groupe d'"outliers" et la classification n'était que faiblement associée à la démence. Dans notre travail de thèse, nous avons donc préféré modéliser l'hétérogénéité du déclin cognitif en prenant explicitement en compte son association avec la démence.

Par rapport aux modèles conjoints à effets aléatoires partagés, les modèles conjoints à classes latentes ont une interprétation plus intuitive et sont plus simples à estimer numériquement. Ils sont en effet basés sur l'hypothèse qu'il existe un certain nombre de sous-populations homogènes, chacune ayant sa propre évolution et son propre risque d'événement, et que chaque sujet a une probabilité différente d'appartenir à chaque classe latente.

6.1.3 Prise en compte des variables explicatives

Les modèles que nous avons proposés dans ce travail offrent un cadre souple pour tester l'effet de variables explicatives sur le vieillissement cognitif. En effet, il est possible de tester l'impact de variables explicatives sur l'évolution du processus latent ou sur les outils de mesure de la cognition mais aussi d'introduire des variables explicatives dans les différents sous-modèles du modèle conjoint, permettant ainsi de mieux comprendre le mode d'action de facteurs de risque de déclin cognitif et de démence.

Effet sur le processus latent et sur les marqueurs

Comme nous l'avons vu dans le chapitre II, la modélisation conjointe de plusieurs marqueurs permet d'augmenter la puissance des analyses pour tester l'effet de variables explicatives (Gray et Brookmeyer, 1998). Dans notre approche, nous pouvons certes utiliser l'information sur tous les marqueurs pour évaluer l'effet de variables explicatives sur le processus latent mais nous pouvons aussi distinguer des effets différents sur chaque marqueur. Ce dernier point a été très peu développé dans la littérature. Gray et Brookmeyer (1998) ont pris en compte un effet potentiellement différent du traitement pour chaque marqueur en faisant varier l'échelle de temps de chaque marqueur. Récemment, Guo et Bandeen-Roche (2005) ont aussi pris en compte une hétérogénéité de l'effet du traitement autour de l'effet global par le biais d'un effet aléatoire binaire. Ils ont ainsi pu distinguer deux groupes de marqueurs et quantifier l'effet du traitement dans le groupe de marqueurs très sensibles au traitement et dans le groupe de marqueurs moins sensibles au traitement. Dans notre approche, par l'introduction de contrastes sur les marqueurs, nous sommes en mesure de tester et de quantifier l'hétérogénéité de l'effet du traitement sur les marqueurs.

Variables explicatives dans le modèle conjoint à classes latentes

Dans les modèles conjoints à classes latentes présentés aux chapitres IV et V, l'effet de variables explicatives peut être testé à plusieurs niveaux du modèle :

1. sur la probabilité d'appartenir aux classes latentes. Cela revient à se demander si la variable explicative est prédictrice d'une forme d'évolution et d'un risque spécifique d'événement ;
2. sur l'évolution des marqueurs de manière commune aux classes (modèle linéaire mixte) : on teste ainsi si, quelle que soit la classe d'évolution des marqueurs, il existe une différence d'évolution suivant la variable explicative ;
3. sur l'évolution des marqueurs de manière différente dans chaque classe. Cela revient à tester si, à même classe latente, c'est à dire même risque d'événement, il existe des différences d'évolution suivant la variable explicative ;
4. sur le risque d'événement après ajustement sur les classes latentes. Dans ce cas, l'hypothèse est : ajusté sur la classe latente, c'est à dire sur le profil d'évolution du marqueur, la variable explicative reste associée à l'événement lui-même.

Les multiples façons d'introduire une variable explicative dans le modèle conjoint à classes latentes lui confèrent un grand intérêt en épidémiologie puisque ce modèle permet d'évaluer différentes hypothèses concernant le mode d'action des facteurs de risque. Néanmoins, il est important lorsque l'on introduit des variables explicatives dans le modèle conjoint à classes latentes de le faire de manière réfléchie et non de rentrer systématiquement les variables explicatives à tous les niveaux. Par exemple, un variable incluse dans le modèle mixte (en 2 ou 3) n'influence que l'espérance du marqueur alors qu'une variable incluse dans la probabilité d'appartenir aux classes latentes (en 1) influence l'espérance et la variance du marqueur. Par ailleurs, l'interprétation des classes latentes est différente suivant que les variables explicatives ont été incluses dans le modèle mixte (en 2 ou 3) ou dans la probabilité d'appar-

tenir aux classes latentes (en 1). En effet, dans le premier cas, les classes latentes représentent l'hétérogénéité résiduelle après prise en compte de l'hétérogénéité associée aux variables explicatives alors que dans le deuxième cas, les classes latentes représentent l'hétérogénéité expliquée, au moins partiellement, par les variables explicatives.

Dans la littérature, aucune précaution n'a été prise quant à l'introduction de variables explicatives aux différents niveaux du modèle (Lin et al., 2000; Muthén et al., 2002). Nous recommandons pourtant de ne pas introduire systématiquement les variables explicatives à tous les niveaux du modèle car, outre des problèmes d'identifiabilité, cela peut nuire à l'interprétabilité des résultats. Il est préférable de choisir à quel niveau introduire les variables explicatives suivant l'objectif de l'étude.

6.1.4 Outil de diagnostic de démence

Grâce aux modèles conjoints à classes latentes présentés dans les chapitres IV et V, nous avons été en mesure de proposer des outils de diagnostic de démence calculés à partir de l'information sur l'évolution cognitive. Ce type d'outil avait déjà été proposé dans le cas d'un seul marqueur d'évolution Gaussien par Lin et al. (2002) et Taylor et al. (2005) pour détecter respectivement un cancer de la prostate ou une rechute de cancer de la prostate à partir des mesures répétées de PSA. Par rapport à leur travaux, comme nous avons utilisé plusieurs tests psychométriques pour décrire l'évolution cognitive, l'outil de détection de démence que nous avons proposé peut être utilisé quel que soit le ou les tests mesurés pour un sujet. Cela nous donne donc un outil de diagnostic plus souple.

6.2 Modèle à classes latentes et sortie d'étude informative

Les modèles conjoints à classes latentes ont été récemment utilisés pour prendre en compte la sortie d'étude (Beunckens et al., 2006) ou les données manquantes intermittentes (Lin et al., 2004). L'approche par classes latentes est en effet une méthode intermédiaire entre les modèles de sélection et les modèles par mélange de profils d'observations utilisés dans le traitement des données longitudinales incomplètes informatives. En faisant l'hypothèse qu'il existe un petit nombre de groupes présentant une évolution et un risque de sortie d'étude spécifiques, les modèles à classes latentes évitent à la fois les problèmes numériques du modèle de sélection liés à l'intégration numérique sur les effets aléatoires ou sur les réponses manquantes, et les problèmes numériques rencontrés avec les modèles par mélange de profils d'observations lorsqu'il existe de nombreux profils de sortie d'étude correspondant à de petits effectifs et pour lesquels le nombre de mesures est insuffisant pour estimer l'évolution du marqueur.

Bien que non présenté dans ce document, nous avons utilisé un modèle conjoint à classes latentes pour modéliser conjointement l'évolution cognitive et la sortie d'étude dans la cohorte PAQUID. Les conclusions de cette analyse étaient très différentes de celles de l'analyse par mélange de profils d'observation. De plus, en analysant les résultats, nous avons repéré que les classes latentes mises en évidence par ce modèle conjoint n'étaient que faiblement associées à la sortie d'étude. Notamment, des classes d'évolution très similaires étaient mises en évidence lorsque la sortie d'étude n'était pas modélisée conjointement, ces classes latentes représentant l'hétérogénéité principale des déclin cognitifs plutôt que l'hétérogénéité directement associée à la sortie d'étude. Les paramètres du modèle d'évolution des marqueurs dans ce modèle conjoint ne pouvaient donc pas s'interpréter simplement comme des paramètres ajustés sur la sortie d'étude mais plutôt comme des paramètres estimés après prise en compte de la principale source d'hétérogénéité de l'évolution

(les classes étant elles mêmes très associées aux variables explicatives telles que le sexe ou le niveau d'éducation).

Ainsi, bien que les modèles à classes latentes aient été utilisés pour traiter la sortie d'étude informative, la pertinence d'une telle analyse et notamment l'interprétation des résultats obtenus ne sont pas clairement établies. Il nous semble qu'une telle analyse doit être complétée par des analyses de l'adéquation du modèle aux données afin d'éviter une interprétation abusive des résultats, en particulier dans les grands jeux de données présentant de multiples sources d'hétérogénéité.

6.3 Limites de l'approche et développements futurs

Bien que le modèle conjoint à classes latentes pour données multivariées que nous avons proposé dans ce travail présente certains attraits, il possède aussi des limites sur lesquelles nous allons revenir maintenant.

6.3.1 Adéquation du modèle à classes latentes

Le cas particulier de la sortie d'étude informative a pointé un aspect très important dans l'application des modèles conjoints à classes latentes. Il existe des hypothèses fortes dans le modèle qui nécessitent d'être évaluées. Nous avons présenté dans les chapitres IV et V plusieurs méthodes pour évaluer l'adéquation du modèle. Nous avons d'abord évalué la qualité de la classification *a posteriori*, montrant ainsi une bonne discrimination des classes. Nous avons ensuite évalué la qualité d'ajustement des évolutions prédites par le modèle mixte conditionnellement aux classes latentes par rapport aux évolutions observées, et avons évalué l'hypothèse de proportionalité des risques entre les classes dans le modèle de survie. Ces analyses ont mis en évidence une bonne adéquation du modèle aux données. Enfin, le modèle conjoint

à classes latentes permettant de proposer un outil de détection de l'événement, nous avons aussi évalué les capacités prédictrices du modèle et l'avons comparé à des modèles plus simples dans le chapitre IV. Par contre, dans le chapitre V, l'outil diagnostique étant issu d'un modèle de survie, nous avons uniquement pu calculer les courbes ROC et leurs AUC à des temps donnés. Évaluer les capacités prédictrices du modèle nécessite en effet le développement d'un résumé des AUC pour tous les temps. Dans les modèles de survie, Heagerty et Zheng (2005) ont proposé une mesure globale de concordance calculée comme l'intégrale sur la fenêtre d'observation des AUC dépendantes du temps mais leur mesure n'est pas adaptée à notre cas de figure. En effet, le critère diagnostique en t issu de notre modèle est dépendant du temps puisque le prédicteur linéaire du modèle de survie spécifique au groupe doit être pondéré par $P(c_{ig} = 1 \mid \hat{\theta}, Y_{is}, s < t)$ comme présenté en (5.15). Proposer une mesure résumée des AUC dépendantes du temps pour un critère diagnostique dépendant du temps nécessite donc des développements supplémentaires.

Dans ce travail, nous n'avons pas examiné l'hypothèse majeure d'indépendance entre les marqueurs et l'événement conditionnellement aux classes latentes. Même si cette hypothèse d'indépendance conditionnelle existe dans tout modèle à effets aléatoires partagés, elle peut être plus forte ici car l'effet aléatoire est discret, c'est à dire que la dépendance est régie uniquement par un faible nombre de classes au lieu de combinaisons linéaires des effets aléatoires dans les autres modèles.

L'hypothèse d'indépendance conditionnelle ne peut pas être strictement contrôlée puisque les classes latentes ne sont pas observées. En s'appuyant sur les travaux de Bandeen-Roche et al. (1997), cette hypothèse est généralement vérifiée en utilisant la classification *a posteriori* : l'hypothèse d'indépendance conditionnelle est considérée comme valide s'il y a indépendance entre marqueur et événement après ajustement ou stratification sur les classes définies *a posteriori* (Lin et al., 2002a; Lin et al., 2004; Guo et al., 2006; Roy, 2003). Les différentes analyses publiées diffèrent par 3 aspects.

Tout d'abord, la classification a posteriori peut être définie suivant la probabilité *a posteriori* maximum ou être simulée suivant cette probabilité (Roy, 2003). Ces stratégies ne tiennent pas compte de l'incertitude sur la classification et il est donc préférable d'utiliser une analyse pondérée par les probabilités *a posteriori* en suivant la méthode proposée par Lin et al. (2002a) et Lin et al. (2004) mais une telle méthode est plus difficile à mettre en oeuvre. Ils proposent en effet d'évaluer si, *a posteriori*, après stratification sur les classes, il reste une dépendance entre le délai d'événement et la trajectoire des marqueurs. Ils font pour cela G copies des données, correspondant aux G classes latentes, et pondèrent chaque donnée par la probabilité *a posteriori* d'appartenir à la classe conditionnellement au marqueur et à l'événement. Puis, ils estiment un modèle de survie pour le délai d'événement, stratifié sur la classe latente, en incluant en variable explicative une ou plusieurs fonctions des marqueurs en plus des variables explicatives habituelles. L'hypothèse d'indépendance conditionnelle est alors rejetée si les paramètres associés aux marqueurs sont significativement différents de 0.

Une autre différence est qu'après classification, certains auteurs réalisent une analyse sur l'ensemble de l'échantillon en ajustant sur les classes (Lin et al., 2002a; Lin et al., 2004) tandis que d'autres réalisent une analyse séparément dans chaque classe (Guo et al., 2006; Roy, 2003).

Enfin la principale difficulté réside dans le choix de la fonction des marqueurs à introduire comme variable explicative dans le modèle de survie. Quand cela est possible, il est souhaitable d'utiliser les valeurs observées (Guo et al., 2006; Roy, 2003). Lin et al. (2004) utilisent la dernière valeur observée du marqueur mais cela est difficile dans notre application en raison du taux important de données manquantes et donc de l'irrégularité des mesures. Une autre stratégie consiste à utiliser l'espérance du marqueur (ou dans notre application, du processus latent) estimé à partir du modèle conjoint à classes latentes. Lin et al. (2002a) introduisent ainsi l'espérance marginale $E(Y_i(t)) = \sum_{g=1}^G \hat{\pi}_{ig}^{y,T} E(Y_i(t)|c_{ig} = 1)$ dans le modèle de survie ajusté sur les classes *a posteriori*. Ce choix est cependant discutable, et il nous semble que

l'espérance incluant les effets aléatoires ou l'espérance conditionnellement à la classe pourraient être préférées.

Le développement de méthodes d'évaluation de l'hypothèse d'indépendance conditionnelle et plus généralement de l'adéquation du modèle constitue donc l'un des principaux objectifs de nos travaux dans la suite de cette thèse.

Malgré ces questionnements, nous avons testé la méthode proposée par Lin et al. (2002a) dans l'application présentée en section 5.6, et les premières analyses ont suggéré l'existence d'une dépendance résiduelle entre l'événement et l'espérance du processus latent. Étant donnée la taille des échantillons utilisés dans nos applications et donc la puissance des analyses, il n'est pas surprenant qu'un test statistique rejette l'hypothèse d'indépendance conditionnelle. Sur de grands échantillons, les méthodes graphiques sont souvent préférées pour évaluer l'adéquation d'un modèle. C'est pourquoi, nous nous orientons aujourd'hui vers une analyse graphique des résidus R_i des marqueurs transformés et calculés conditionnellement à l'événement. Dans notre modèle qui inclut un processus latent, ces résidus sont $R_i = h_k(y_i; \hat{\eta}_k) - E(h_k(y_i; \hat{\eta}_k) | T_i, E_i)$. L'examen du nuage de points des résidus R_i en fonction de (T_i, E_i) devrait permettre d'évaluer l'hypothèse d'indépendance conditionnelle.

Dans notre application, il n'est pas inconcevable que l'hypothèse d'indépendance conditionnelle ne soit pas valide. En effet, étant donnée la taille de notre échantillon, il paraît peu vraisemblable d'éliminer toute corrélation entre la cognition et la démence par quelques classes latentes. Une amélioration de notre modèle pourrait consister à inclure dans le modèle de survie une fonction des effets aléatoires en plus de la classe latente. Ainsi, outre tester l'hypothèse d'indépendance conditionnelle en testant le paramètre associé aux effets aléatoires à 0, nous pourrions améliorer la prise en compte de la dépendance entre les marqueurs et le délai d'événement. Cependant, ce type de modèle étant une combinaison d'un modèle à classes latentes et d'un modèle à effets aléatoires partagés, son estimation serait difficile. Il cumule en

effet les difficultés liées à l'intégration numérique sur les effets aléatoires du modèle à effets aléatoires partagés et les difficultés liées à la multimodalité possible d'un modèle à classes latentes.

Le rejet de l'hypothèse d'indépendance conditionnelle dans notre application sur le vieillissement cognitif pourrait aussi s'expliquer par la nature statique du modèle. En effet, dans un modèle conjoint à classes latentes, les classes sont fixées au cours du temps. Or plusieurs études ont montré que les sujets en phase pré-diagnostique de démence "décrochaient" de l'évolution cognitive normale plusieurs années avant le diagnostic de démence (Jacqmin-Gadda et al., 2006 ; Hall et al., 2000). Suivant ces résultats, tous les sujets devraient appartenir initialement à la même classe latente, c'est à dire la classe représentant le vieillissement normal, et les sujets en phase pré-diagnostique de démence devraient changer de classe latente au cours du temps, pour passer dans la classe associée à un vieillissement pathologique. Le modèle conjoint à classes latentes que nous avons développés ne permet pas de prendre en compte de tels changements de classes au cours du temps. Néanmoins, lorsque l'on étudie l'évolution cognitive sur une période courte (12 ans par exemple dans l'application du chapitre IV), un faible nombre de classes fixes peut suffire à décrire cette évolution au cours du temps. Par contre, lorsqu'on étudie l'évolution sur 30 ou 40 ans en fonction de l'âge, il est plus difficile de prendre en compte l'aspect dynamique avec un petit nombre de classes fixes.

Les changements de classes au cours du temps peuvent être modélisés par un processus latent discret à faibles nombres d'états. Un tel modèle rentrerait dans le cadre des modèles à états cachés avec la contrainte supplémentaire d'assurer la continuité de l'évolution cognitive lors de la transition entre deux états.

6.3.2 Transformations nonlinéaires des marqueurs

Une force de notre approche réside dans les transformations non linéaires estimées entre les marqueurs quantitatifs non Gaussiens et le processus latent. Nous avons pour l'instant utilisé la famille des fonctions de répartition Beta qui sont apparues comme un bon compromis entre la variabilité possible des formes de la fonction et leur relative simplicité d'estimation. Cependant, nous avons vu dans le chapitre III par des études de simulation que ces fonctions présentaient une limite importante liée à leur espace d'arrivée $[0, 1]$ incompatible avec la définition dans \mathbb{R} du processus Gaussien. Ceci engendre en particulier des effets de bord qui tendent à réduire la variabilité des erreurs. Même si les biais engendrés sont faibles, il serait intéressant de trouver une alternative aux fonctions de répartition Beta. La famille de fonctions devrait être constituée de fonctions croissantes définies de $[a, b]$ dans \mathbb{R} et devrait offrir des formes de transformations variées incluant des formes concaves, convexes, sigmoïdes et linéaires. A ce jour, nous n'avons pas trouvé de famille de fonctions paramétriques à valeurs dans \mathbb{R} répondant à ces contraintes. En effet, les fonctions d'une même famille ont souvent soit toutes une forme concave, soit toutes une forme convexe alors que nos applications requièrent au moins deux types de formes pour modéliser le MMSE ayant un effet plafond et pas d'effet plancher et le DSST n'ayant pas d'effet plafond mais un effet plancher.

Les fonctions I-splines (à valeurs dans \mathbb{R}^+) associées à une intercept afin qu'elles soient à valeurs dans \mathbb{R} répondent à ces exigences mais elles posent aussi de sérieux problèmes numériques. En effet, ces transformations requièrent un nombre important de paramètres et une estimation par vraisemblance pénalisée doit être envisagée comme vu en section 5.3.2. De plus, les formes de transformations étant très différentes d'un test à l'autre, il est souhaitable d'utiliser un paramètre de lissage différent pour chaque test. Cette approche semi-paramétrique nous semble donc difficile à utiliser en pratique mais peut cependant être utile dans une étape de validation pour évaluer graphiquement l'adéquation des transformations Beta.

Dans l'optique d'augmenter l'information utilisée pour décrire le processus latent et de généraliser notre approche, il serait aussi intéressant de pouvoir inclure dans le modèle non seulement des marqueurs quantitatifs non Gaussiens mais aussi des marqueurs qualitatifs ordinaux. Nous pourrions ainsi considérer des tests psychométriques comportant peu de modalités. Ce type d'approche a été proposé, notamment dans le cadre des modèles à équations structurelles (Dunson, 2000 ; Rabe-Hesketh et al., 2004). Dans notre cas, il s'agirait de combiner un modèle à seuils (Ganiayre et al., 2006 ; Liu et Hedeker, 2006) pour les données ordinales et le modèle proposé au chapitre III pour les données quantitatives.

6.3.3 Multiplicité des événements

Dans notre approche, nous n'avons pour l'instant pris en compte qu'un seul événement conjointement aux marqueurs longitudinaux. Or, dans l'étude du vieillissement, plusieurs événements peuvent intervenir de manière concurrentielle comme la survenue d'une démence (avec des types différents) et le décès, ou de manière non concurrentielle comme la survenue d'une démence et l'entrée en institution. Modéliser conjointement plusieurs événements pourrait permettre de mieux comprendre leur dépendance mutuelle et leur association avec le déclin cognitif. A partir de notre approche, il est possible de considérer plusieurs événements conjointement en ajoutant un modèle de survie pour chaque événement prédit par la classe d'appartenance et envisager une corrélation entre les événements par un paramètre de fragilité commun comme proposé récemment dans le cadre de modèles à effets aléatoires partagés (Chi et Ibrahim, 2006).

6.4 Conclusion

Dans ce travail, nous avons proposé une nouvelle approche pour modéliser le déclin des fonctions cognitives à partir de plusieurs tests psychométriques et son association avec la survenue d'une démence. Malgré les limites abordées dans ce chapitre, cette approche est utile à la fois par sa souplesse concernant notamment la prise en compte de variables explicatives et la modélisation de plusieurs marqueurs non Gaussiens et par sa relative simplicité numérique. Nous avons en particulier montré à travers plusieurs applications son intérêt dans l'étude du vieillissement cognitif et la survenue d'une démence. Cependant, des développements sont encore nécessaires pour évaluer les hypothèses du modèle et en premier lieu l'hypothèse d'indépendance conditionnelle. De plus, il serait intéressant par la suite d'étendre notre approche à des marqueurs de nature différente, à la modélisation conjointe de plusieurs événements mais surtout à la prise en compte de l'aspect dynamique de l'hétérogénéité par des modèles à états cachés.

Chapitre 7

Bibliographie

- Amieva, H., Jacqmin-Gadda, H., Orgogozo, J. M., Le Carret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C. et Dartigues, J. F. (2005). The 9 year cognitive decline before dementia of the Alzheimer type : a prospective population-based study. *Brain* **128**, 1093–101.
- Andersen, K., Launer, L. J., Dewey, M. E., Letenneur, L., Ott, A., Copeland, J. R., Dartigues, J. F., Kragh-Sorensen, P., Baldereschi, M., Brayne, C., Lobo, A., Martinez-Lage, J. M., Stijnen, T. et Hofman, A. (1999). Gender differences in the incidence of AD and vascular dementia : The EURODEM studies. EURODEM Incidence Research Group. *Neurology* **53**, 1992–7.
- Anstey, K. J., Hofer, S. M. et A., L. M. (2003). A latent growth curve analysis of late-life sensory and cognitive function over 8 years : evidence for specific and common factors underlying change. *Psychology and aging* **18**, 714–726.
- Backman, L., Small, B. J. et Fratiglioni, L. (2001). Stability of the preclinical episodic memory deficit in Alzheimer’s disease. *Brain* **124**, 96–102.
- Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L. et Rathouz, P. J. (1997). latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association* **92**, 1375–1386.
- Barnes, L. L., Wilson, R. S., Schneider, J. A., Bienias, J. L., Evans, D. A. et Bennett, D. A. (2003). Gender, cognitive decline, and risk of AD in older persons.

- Neurology* **60**, 1777–81.
- Bauer, D. J. et Curran, P. J. (2003a). Distributional assumptions of growth mixture models : implications for overextraction of latent trajectory classes. *Psychol Methods* **8**, 338–63.
- Bauer, D. J. et Curran, P. J. (2003b). Overextraction of latent trajectory classes : much ado about nothing ? Reply to Rindskopf (2003), Muthén (2003) and Cudeck and Henly (2003). *Psychol Methods* **8**, 384–393.
- Beckett, L. A., Tancredi, D. J. et Wilson, R. S. (2004). Multivariate longitudinal models for complex change processes. *Stat Med* **23**, 231–9.
- Benton, A. (1965). *Manuel pour l'application du Test de Rétention Visuelle. Applications cliniques et expérimentales*. Centre de Psychologie appliquée, Paris, 2ème édition française edition.
- Berchtold, A. (2004). Optimization of mixture models : comparison of different strategies. *Computational statistics* **19**, 385–406.
- Beunckens, C., Molenberghs, G., Verbeke, G. et Mallinckrodt, C. (2006). A latent-class mixture model for incomplete longitudinal gaussian data. Center for statistics, Hasselt University, Belgium.
- Biernacki, C., Celeux, G. et Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis* **41**, 561–575.
- Bohning, D. (1995). A review of reliable maximum likelihood algorithms for semi-parametric 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.
- Brooks, J. O. et Yesavage, J. A. (1995). Identification of fast and slow decliners in Alzheimer disease : a different approach. *Alzheimer Dis Assoc Disord* **9 Suppl**

- 1, S19–25.
- Brown, E. R., Ibrahim, J. G. et DeGruttola, V. (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics* **61**, 64–73.
- Celeux, G., Martin, O. et Lavergne, C. (2005). Mixture of linear mixed models : application to repeated data clustering. *Statistical modelling* **5**, 243–267.
- Charnigo, R. et Sun, J. (2004). Testing homogeneity in a mixture distribution via the L2 distance between competing models. *Journal of the American Statistical Association* **99**, 488–498.
- Chen, H. et Chen, J. (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Statistics and Probability Letters* **52**, 125–133.
- Chen, H., Chen, J. et Kalbfleish, J. D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *J R Statist Soc B* **63**, 19–29.
- Chen, J. et Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter. *Journal of Statistical Planning and Inference IISA 2002 DeKalb Conference* **129**, 93–107.
- Chi, Y.-Y. et Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* **62**, 432–445.
- Cowppli-Bony, P., Dartigues, J. F. et Orgogozo, J. M. (2006). [vascular risk factors and Alzheimer disease risk : epidemiological studies review]. *Psychol Neuropsychiatr Vieil* **4**, 47–60.
- Cudeck, R. et Henly, S. J. (2003). A realistic perspective on pattern representation in growth data : comment on Bauer and Curran (2003). *Psychol Methods* **8**, 378–83 ; discussion 384–93.
- Dartigues, J. F., Commenges, D., Letenneur, D., Barberger-Gateau, P., Gilleron, V., Fabrigoule, C., Mazaux, J. M., Orgogozo, J. M. et Salamon, R. (1997). Cognitive predictors of dementia in elderly community residents. *Neuroepidemiology* **16**, 29–39.
- De Boor, C. (1978). *A practical guide to splines*. Springer, New-York.
- de la Torre, J. C. (2004). Is alzheimer’s disease a neurodegenerative or a vascular

- disorder? data, dogma, and dialectics. *Lancet Neurol* **3**, 184–90.
- Dempster, A., Laird, N. et Rubin, D. (1977). Maximum likelihood from incomplete data via EM algorithm (with discussion). *J R Statist Soc B* **39**, 1–38.
- Dempster, A., Laird, N. et Rubin, D. (1978). Maximum likelihood from incomplete data via em algorithm (with discussion). *J R Statist Soc B* **39**, 1–38.
- Dunson, D. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* **62**, 355–66.
- Dunson, D. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association* **98**, 555–63.
- Elias, M. F., Elias, P. K., D’Agostino, R. B., Silbershatz, H. et Wolf, P. A. (1997). Role of age, education, and gender on cognitive performance in the framingham heart study : community-based norms. *Exp Aging Res* **23**, 201–35.
- Fabrigoule, C., Letenneur, L., Dartigues, J. F., Zarrouk, M., Commenges, D. et Barberger-Gateau, P. (1995). Social and leisure activities and risk of dementia : a prospective longitudinal study. *J Am Geriatr Soc* **43**, 485–90.
- Fabrigoule, C., Rouch, I., Taberly, A., Letenneur, L., Commenges, D., Mazaux, J.-M., Orgogozo, J.-M. et Dartigues, J.-F. (1998). Cognitive process in preclinical phase of dementia. *Brain* **121**, 135–141.
- Farrer, L., Cupples, L., Haines, J., Hyman, B., Kukull, W., Mayeux, R., Myers, R., Pericak-Vance, M., 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. *JAMA* **278**, 1349–56.
- Fieuws, S. et Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles : pitfalls of the random-effects approach. *Stat Med* **23**, 3093–104.
- Fitzmaurice, G. M., Laird, N. M. et Shneyer, L. (2001). An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. *Stat Med* **20**, 1009–21.
- Fletcher, R. (2000). *Practical methods of optimization, second ed.* John Wiley and

- Sons, New-York.
- Folstein, M. F., Folstein, S. E. et McHugh, P. R. (1975). "mini-Mental State". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–98.
- Fratiglioni, L., Launer, L. J., Andersen, K., Breteler, M. M., Copeland, J. R., Dartigues, J. F., 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 the Elderly Research Group. *Neurology* **54**, S10–5.
- Galasko, D. R., Gould, R. L., Abramson, I. S. et Salmon, D. P. (2000). Measuring cognitive change in a cohort of patients with Alzheimer's disease. *Stat Med* **19**, 1421–32.
- Ganguli, M., Seaberg, E. C., Ratcliff, G. G., Belle, S. H. et DeKosky, S. T. (1996). Cognitive stability over 2 years in a rural elderly population : the MoVIES project. *Neuroepidemiology* **15**, 42–50.
- Ganiayre, J., Commenges, D. et Letenneur, L. (2006). A latent process model for dementia and psychometric tests. Technical report, Equipe de Biostatistique INSERM E0338, Université Victor Segalen, Bordeaux, France.
- 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. Wadsworth.
- Gray, S. M. et Brookmeyer, R. (1998). Estimating a treatment effect from multidimensional longitudinal data. *Biometrics* **54**, 976–988.
- Gray, S. M. et Brookmeyer, R. (2000). Multidimensional longitudinal data : estimating a treatment effect from continuous, discrete, or time-to-event response variables. *Journal of the American Statistical Association* **95**, 396–406.
- Guo, H. et Bandeen-Roche, K. (2005). Modeling differentiated treatment effects for multiple outcomes data. Technical report, Johns Hopkins

- University, Dept. of Biostatistics Working Papers. Working Paper 91.,
<http://www.bepress.com/jhubiostat/paper91>.
- Guo, J., Wall, M. et Amemyia, Y. (2006). Latent class regression on latent factors. *Biostatistics* **7**, 145–63.
- Hall, C. B., Lipton, R. B., Sliwinski, M. et Stewart, W. F. (2000). A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Stat Med* **19**, 1555–66.
- Hall, C. B., Ying, J., Kuo, L., Sliwinski, M., Buschke, H., Katz, M. et Lipton, R. B. (2001). Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Stat Med* **20**, 3695–714.
- Hall, C. B., Ying, J. et 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.
- Hall, D. B. et Clutter, M. (2004). Multivariate multilevel nonlinear mixed effects models for timber yield predictions. *Biometrics* **60**, 16–24.
- Harvey, D., Laurel, L. et Mungas, D. (2003). Multivariate modeling of two associated cognitive outcomes in a longitudinal study. *Journal of Alzheimer's disease* **5**, 357–65.
- Haughton, D. (1997). Packages for estimating finite mixtures : a review. *the American statistician* **51**, 194–205.
- Hawkins, D. S., Allen, D. M. et Stromberg, A. J. (2001). Determining the number of components in mixtures of linear models. *Computational Statistics and Data Analysis* **38**, 15–48.
- Heagerty, P. et Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105.
- Henderson, R., Diggle, P. et Dobson, A. (2000). 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 survival. *Biostatistics* **3**, 33–50.

- Hogan, J. W. et Laird, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Stat Med* **16**, 259–72.
- Isaacs, B. et Kenne, A. T. (1973). The Set test as an aid to the detection of dementia in old people. *Br J Psychiatry* **123**, 467–70.
- Jacqmin-Gadda, H., Commenges, D. et Dartigues, J. (1999). [analysis of longitudinal Gaussian data with missing data on the response variable]. *Rev Epidemiol Sante Publique* **47**, 525–34.
- Jacqmin-Gadda, H., Commenges, D. et Dartigues, J. F. (2006). Random change point model for joint modeling of cognitive decline and dementia. *Biometrics* **62**, 254–60.
- Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D. et Dartigues, J. F. (1997a). A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *Am J Epidemiol* **145**, 498–506.
- Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D. et Dartigues, J. F. (1997b). [longitudinal study of cognitive aging in non-demented elderly subjects]. *Rev Epidemiol Sante Publique* **45**, 363–72.
- Jeffries, N. O. (2003). A note on 'Testing the number of components in a normal mixture'. *Biometrika* **90**, 991–994.
- Joly, P. (1996). *Estimation de la fonction de risque dans un contexte général de troncature et de censure : application à l'estimation de l'incidence de la démence*. PhD thesis, Université de Bordeaux II.
- 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.
- Joseph, L., Wolfson, D. B., Belisle, P., Brooks, J. O., r., Mortimer, J. A., Tinklenberg, J. R. et Yesavage, J. A. (1999). Taking account of between-patient variability when modeling decline in Alzheimer's disease. *Am J Epidemiol* **149**, 963–73.
- Karlis, D. et Xekalaki, E. (2003). Choosing initial values for the EM algorithm for

- finite mixtures. *Computational Statistics and Data Analysis* **41**, 577–590.
- Komarek, A. (2001). A SAS-macro for linear mixed models with a finite normal mixture as random-effects distribution. Technical report, Biostatistical Center, Catholic University of Leuven, Belgium, <http://www.med.kuleuven.ac.be/biostat/research/software.htm>.
- Laird, N. et Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–74.
- Larrieu, S., Letenneur, L., Helmer, C., Dartigues, J. F. et Barberger-Gateau, P. (2004). Nutritional factors and risk of incident dementia in the PAQUID longitudinal cohort. *J Nutr Health Aging* **8**, 150–4.
- Law, N. J., Taylor, J. M. 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.
- Lee, S. et Song, X. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology* **57**, 131–150.
- Legler, J. M., Davis, W. W., Potosky, A. L. et Hoffman, R. M. (2004). Latent variable modelling of recovery trajectories : sexual function following radical prostatectomy. *Stat Med* **23**, 2875–93.
- Letenneur, L., Commenges, D., Dartigues, J. F. et Barberger-Gateau, P. (1994). Incidence of dementia and Alzheimer’s disease in elderly community residents of south-western France. *Int J Epidemiol* **23**, 1256–61.
- Letenneur, L., Gilleron, V., Commenges, D., Helmer, C., Orgogozo, J. et Dartigues, J. (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.
- Lin, H., McCulloch, C. E. et Mayne, S. T. (2002). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat Med* **21**, 2369–82.

- Lin, H., McCulloch, C. E. et Rosenheck, R. A. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* **60**, 295–305.
- Lin, H., McCulloch, C. E., Turnbull, B. W., Slate, E. H. et Clark, L. C. (2000). A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Stat Med* **19**, 1303–18.
- Lin, H., Turnbull, B. W., Mc Culloch, C. E. et Slate, E. H. (2002). 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.
- Lindstrom, M. J. et Bates, D. M. (1988). Newton-raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* **83**, 1014–1022.
- Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.
- Little, R. J. et Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley and Sons, New-York.
- Liu, L. C. et Hedeker, D. (2006). A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics* **62**, 261–8. Journal Article.
- Lo, Y., Mendell, N. R. et Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika* **88**, 767–778.
- Lobo, A., Launer, L. J., Fratiglioni, L., Andersen, K., Di Carlo, A., Breteler, M. M., Copeland, J. R., Dartigues, J. F., Jagger, C., Martinez-Lage, J., Soininen, H. et Hofman, A. (2000). Prevalence of dementia and major subtypes in Europe : A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology* **54**, S4–9.
- Luchsinger, J. A. et Mayeux, R. (2004). Dietary factors and alzheimer’s disease. *Lancet Neurol* **3**, 579–87.
- Marshall, G., De la Cruz-Mesia, R., Baron, A. E., Rutledge, J. H. et Zerbe, G. O.

- (2006). Non-linear random effects model for multivariate responses with missing data. *Stat Med* **25**, 2817–2830.
- Matsuyama, Y. et Ohashi, Y. (1997). Mixed models for bivariate response repeated measures data using gibbs sampling. *Stat Med* **16**, 1587–601.
- McCulloch, C. E., Lin, H., Slate, E. H. et Turnbull, B. W. (2002). Discovering subpopulation structure with latent class mixed models. *Stat Med* **21**, 417–29.
- McLachlan, G. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a nomrla mixture. *Appl Statist* **36**, 318–324.
- Mendondo, M. S., Ashford, J. W., Kryscio, R. J. et A., S. F. (2000). Modelling Mini Mental State Examination changes in Alzheimer’s disease. *statistics in medicine* **19**, 1607–1616.
- Miloslavsky, M. et van der Laan, M. J. (2003). Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Computational Statistics and Data Analysis* **41**, 413–428.
- Moreno, E. et Liseo, B. (2003). A default Bayesian test for the number of components in a mixture. *Journal of Statistical Planning and Inference* **111**, 129–142.
- Morris, M. C., Evans, D. A., Hebert, L. E. et Bienias, J. L. (1999). Methodological issues in the study of cognitive decline. *Am J Epidemiol* **149**, 789–93.
- Muthén, B. (2001). *Latent variable mixture modeling*. In G.A. Marcoulides and R.E. Schmacker (eds.), *New developmnets and techniques in Structural equation modeling (pp 1-33)*. Lawrence Erlbaum Associates.
- Muthén, B. (2002). Beyond SEM : General latent variable modeling. *Behaviormetrika* **29**, 81–117.
- Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling : comment on Bauer and Curran (2003). *Psychol Methods* **8**, 369–77 ; discussion 384–93.
- Muthén, B. (2004). *Latent variable analysis : Growth mixture modeling and related techniques for longitudinal data*. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences (pp. 345-368)*. Sage Publications, Newbury

- Park, C.A.
- Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S. G., Carlin, J. B. et Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* **3**, 459–75.
- Muthén, B. et Muthén, L. (2001). *Mplus user's guide*. Muthén and Muthén, Los Angeles, CA.
- Muthén, B. et Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55**, 463–9.
- 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.
- O'Brien, L. M. et Fitzmaurice, G. M. (2005). Regression models for the analysis of longitudinal gaussian data from multiple sources. *Stat Med* **24**, 1725–44. Journal Article.
- OPEPS (2006). Rapport sur la maladie d'Alzheimer et les maladies apparentées. Office parlementaire d'évaluation des politiques de santé.
- O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and statistical computing* **9**, 363–379.
- Pauler, D. K. et Laird, N. M. (2000). A mixture model for longitudinal data with application to assessment of noncompliance. *Biometrics* **56**, 464–72.
- 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.
- Rabe-Hesketh, S., Skrondal, A. et Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika* **69**, 167–90.
- Ramaroson, H., Helmer, C., Barberger-Gateau, P., Letenneur, L. et Dartigues, J. F. (2003). [prevalence of dementia and Alzheimer's disease among subjects aged 75 years or over : updated results of the PAQUID cohort]. *Rev Neurol (Paris)* **159**,

- 405–11.
- Ramsay, J. (1988). Monotone regression splines in action. *Statistical science* **3**, 425–461.
- Redner, R. A. et Walker, Homer, F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM review* **26**, 195–239.
- Reite, M., Cullum, C. M., Stocker, J., Teale, P. et Kozora, E. (1993). Neuropsychological test performance and meg-based brain lateralization : sex differences. *Brain Res Bull* **32**, 325–8. Journal Article.
- Richardson, S. et Green, P. J. (1997). On bayesian analysis of mixtures with an unknown number of components. *J R Statist Soc B* **59**, 731–792.
- Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychol Methods* **8**, 364–8; discussion 384–93.
- Ringman, J. et Cummings, J. (2006). Current and emerging pharmacological treatment options for dementia. *Behavioural Neurology* **17**, 5–16.
- Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* **59**, 829–36.
- 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). Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates : changes in methadone treatment practices. *Journal of the American Statistical Association* **97**, 40–52.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol Rev* **103**, 403–28.
- Sánchez, B. N., Budtz-Jorgensen, E., Ryan, L. M. et Hu, H. (2005). Structural equation models : a review with applications to environmental epidemiology. *Journal of the American Statistical Association* **100**, 1443–1455.
- Scarmeas, N., Albert, S. M., Manly, J. J. et Stern, Y. (2006). Education and rates of cognitive decline in incident Alzheimer’s disease. *J Neurol Neurosurg Psychiatry* **77**, 308–16.

- Schlattmann, P. (2003). Estimating the number of components in a finite mixture model : special case of homogeneity. *Computational Statistics and Data Analysis* **41**, 441–451.
- Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.
- Shah, A., Laird, N. M. et Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* **92**, 775–779.
- Shelton, B. J., Gilbert, G. H., Liu, B. et Fisher, M. (2004). A sas macro for the analysis of multivariate longitudinal binary outcomes. *Computer Methods and Programs in Biomedicine* **76**, 163–175.
- Skrondal, A. et Rabe-Hesketh, S. (2004). *Generalized latent variable modelling : multilevel, longitudinal and structural equation models*. Chapman & hall/CRC.
- Sliwinski, M., Hofer, S. et Hall, C. (2003). Correlated and coupled cognitive change in older adults with or without preclinical demantia. *Psychology ang aging* **18**, 672–83.
- Sliwinski, M., Lipton, R. B., Buschke, H. et Stewart, W. (1996). The effects of preclinical dementia on estimates of normal cognitive functioning in aging. *J Gerontol B Psychol Sci Soc Sci* **51**, P217–25.
- Sliwinski, M. J., Hofer, S. M., Hall, C., Buschke, H. et Lipton, R. B. (2003). Modeling memory decline in older adults : the importance of preclinical dementia, attrition, and chronological age. *Psychol Aging* **18**, 658–71.
- Song, X. et Lee, S. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **57**, 29–52.
- Spießens, B., Verbeke, G. et Komarek, A. (2002). A SAS-macro for the classification of longitudinal profiles using mixtures of normal distributions in nonlinear and generalized linear models. Technical report, Biostatistical Center, Catholic University of Leuven, Belgium,

- <http://www.med.kuleuven.ac.be/biostat/research/software.htm>.
- Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D. et Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *Jama* **271**, 1004–10.
- Sugar, C. A. et James, G. M. (2003). Finding the number of clusters in a dataset : an information-theoretic approach. *Journal of the American Statistical Association* **98**, 750–763.
- Sy, J. P. et Taylor, J. M. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics* **53**, 542–555.
- Taylor, J., Yu, M. et Sandler, H. (2005). Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of Clinical Oncology* **23**, 816–825.
- Thiébaud, R., Jacqmin-Gadda, H., Babiker, A. et Commenges, D. (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. *Stat Med* **24**, 65–82.
- Thiébaud, R., Jacqmin-Gadda, H., Chène, G., Leport, C. et Commenges, D. (2002). Bivariate linear mixed models using sas proc mixed. *Computer Methods and Programs in Biomedicine* **69**, 249–256.
- Thiébaud, R., Jacqmin-Gadda, H., Leport, C., Katlama, C., Costagliola, D., Le Moing, V., Meyer, L., Morlat, P., Chène, G. et the APROCO study group (2003). Bivariate longitudinal model for the analysis of the evolution of HIV RNA and CD4 cell count in HIV infection taking into account left censoring of HIV RNA measures. *J Biopharm Stat* **13**, 271–82.
- Todem, D., Kim, K. et Lesaffre, E. (2006). Latent-variable models for longitudinal data with bivariate ordinal outcomes. *Stat Med*.
- Valdois, S., Joannette, Y., Poissant, A., Ska, B. et Dehaut, F. (1990). Heterogeneity in the cognitive profile of normal elderly. *J Clin Exp Neuropsychol* **12**, 587–96.
- 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 Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541–556.
- Verbeke, G. et Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer, New-York.
- Wall, M. et Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association* **95**, 929–40.
- Wang, C.-P., Brown, C. H. et Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models : examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association* **100**, 1054–1076.
- Wang, H. x., Luo, B., Zhang, Q. b. et Wei, S. (2004). Estimation for the number of components in a mixture model using stepwise split-and-merge EM algorithm. *Pattern Recognition Letters* **25**, 1799–1809.
- Wechsler, D. (1981). *WAIS-R manual*. Psychological Corporation, New York.
- Wiederholt, W. C., Cahn, D., Butters, N. M., Salmon, D. P., Kritz-Silverstein, D. et Barrett-Connor, E. (1993). Effects of age, gender and education on selected neuropsychological tests in an elderly community cohort. *J Am Geriatr Soc* **41**, 639–47.
- Wulfsohn, M. S. et Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–9.
- Xu, J. et Zeger, S. L. (2001). The evaluation of multiple surrogate endpoints. *Biometrics* **57**, 81–7.
- Xu, W. et Hedeker, D. (2001). A random-effects mixture model for classifying treatment response in longitudinal clinical trials. *J Biopharm Stat* **11**, 253–73.
- Yesavage, J. A. et Brooks, J. O. (1991). On the importance of longitudinal research in Alzheimer's disease. *J Am Geriatr Soc* **39**, 942–4.

- Yu, M., Taylor, J. et Sandler, H. M. (2004). Individual prediction in prostate cancer studies using a joint longitudinal-survival-cure model. Technical Report 24, University of Michigan, department of Biostatistics.
- Zhang, M.-H. et Cheng, Q.-S. (2004). Determine the number of components in a mixture model by the extended KS test. *Pattern Recognition Letters* **25**, 211–216.

Chapitre 8

Annexes

8.1 Liste des publications et communications scientifiques

Publications :

Proust, C., Jacqmin-Gadda, H., Taylor, J.M.G., Ganiayre, J. et Commenges, D. (2006). A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* **62** :1014-24

Proust-Lima, C., Letenneur, L., et Jacqmin-Gadda, H. (2006). A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome. *Statistics in Medecine*, publication en ligne le 10 août 2006.doi : 10.1002/sim.2659

Proust-Lima, C., Amieva, H., Dartigues, J.-F. et Jacqmin-Gadda, H. (2006). Sensitivity of four psychometric tests to measure cognitive changes in brain aging population-based studies. *American Journal of Epidemiology*, publication en ligne le 14 novembre 2006. doi :10.1093/aje/kwk017

Proust-Lima, C., Amieva, H., Jacqmin-Gadda, H., Letenneur, L., Orgogozo, J.-M. et Dartigues, J.-F. (2006). Impact of gender and education on cognitive decline with age : a longitudinal approach. Article soumis

Communications orales :

Proust-Lima, C., et Jacqmin-Gadda, H. (2006). A nonlinear latent class model for joint analysis of multiple longitudinal outcomes and a clinical event : cognitive ageing context. *Statistical Latent Variable Models in the Health Sciences*, Pérouse (Italie)

Proust-Lima, C., et Jacqmin-Gadda, H. (2006). A nonlinear latent class model for joint analysis of multiple longitudinal outcomes and a clinical event : cognitive ageing context. *International Workshop on Statistical Modelling*, Galway (Irlande)

Proust, C., et Jacqmin-Gadda, H. (2004). Modèle de mélange à effets aléatoires pour données longitudinales multivariées appliqué à l'étude du vieillissement cognitif après 65 ans. *Journées de la Statistique*, Montpellier (France)

Proust, C., et Jacqmin-Gadda, H. (2004). A growth mixture model for multiple outcomes : application to cognitive aging. *Statistics in Health Sciences*, Nantes (France)

Proust, C., Jacqmin-Gadda, H., Taylor, J.M.G., et Commenges, D. (2005). A nonlinear model with latent process using multivariate longitudinal data : application to cognitive aging. *International Biometrics Society British Regional Conference*, Leicester (Angleterre)

Proust, C., Jacqmin-Gadda, H., Taylor, J.M.G., et Commenges, D. (2005). Un modèle non linéaire avec processus latent pour des données longitudinales multivariées appliqué à l'évolution cognitive chez les personnes âgées. *Journées de la Statistique*,

Pau (France)

Communications affichées :

Proust, C., et Jacqmin-Gadda, H. (2004). A growth mixture model for multivariate outcomes : application to cognitive aging. *International Workshop on Statistical Modelling*, Florence (Italie)

8.2 Articles scientifiques

Sont présentés dans cette section, les trois articles scientifiques résumés dans le rapport de thèse :

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., Amieva, H., Dartigues, J.-F. et Jacqmin-Gadda, H. (2006). Sensitivity of four psychometric tests to measure cognitive changes in brain aging population-based studies. *American Journal of Epidemiology*, sous presse

Proust-Lima, C., Amieva, H., Jacqmin-Gadda, H., Letenneur, L., Orgogozo, J.-M. et Dartigues, J.-F. (2006). Impact of gender and education on cognitive decline with age : a longitudinal approach. *Neurobiology of aging*, soumis



ELSEVIER

Computer Methods and
Programs in Biomedicinewww.intl.elsevierhealth.com/journals/cmpb

Estimation of linear mixed models with a mixture of distribution for the random effects

Cécile Proust*, Hélène Jacqmin-Gadda

*Institut National de la Santé et de la Recherche Médicale, Equipe de biostatistique E0338,
Université de Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France*

Received 23 December 2003; received in revised form 8 October 2004; accepted 8 December 2004

KEYWORDS

Heterogeneous mixed model;
Mixture model;
Longitudinal data;
Newton-Raphson-like algorithm;
Cognitive ageing

Summary The aim of this paper is to propose an algorithm to estimate linear mixed model when random effect distribution is a mixture of Gaussians. This heterogeneous linear mixed model relaxes the classical Gaussian assumption for the random effects and, when used for longitudinal data, can highlight distinct patterns of evolution. The observed likelihood is maximized using a Marquardt algorithm instead of the EM algorithm which is frequently used for mixture models. Indeed, the EM algorithm is computationally expensive and does not provide good convergence criteria nor direct estimates of the variance of the parameters. The proposed method also allows to classify subjects according to the estimated profiles by computing posterior probabilities of belonging to each component. The use of heterogeneous linear mixed model is illustrated through a study of the different patterns of cognitive evolution in the elderly. HETMIXLIN is a free Fortran90 program available on the web site: <http://www.isped.u-bordeaux2.fr>.

© 2005 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Many longitudinal studies consist in assessing the changes over the time of a marker measured repeatedly on each participant. These analyses are generally performed with mixed models [1], which allow to take into account the within-subject correlation and the variability of the marker course between the subjects. However, such a model is based on the strong assumption that the random effects

are sampled from a single multivariate Gaussian distribution which means that the marker course is homogeneous among all the subjects.

To assess this assumption, Verbeke and Lesaffre [2] have proposed a mixed model with a mixture of multivariate Gaussians on the random effects. This heterogeneous linear mixed model allows to relax the normality assumption for the random effects and also allows to highlight distinct evolutions of the marker and classify the subjects according to these different patterns of evolution.

In the Verbeke and Lesaffre's work as in more recent papers [3,4], the mixed models with a mixture on the random effects distribution were estimated using an EM algorithm [5]. For instance,

* Corresponding author. Tel.: +33 5 57 57 11 36;
fax: +33 5 56 24 00 81.

E-mail address: cecile.proust@isped.u-bordeaux2.fr
(C. Proust).

Spießens and Verbeke [3] recently proposed a free SAS-macro (HETNLMIXED) using the EM algorithm and the NL MIXED procedure for the optimization in the M-step. This SAS-macro is an extension of the SAS-macro HETMIXED which was developed earlier for estimating heterogeneous linear mixed models using the MIXED procedure [6]. To our knowledge, HETNLMIXED and its first version HETMIXED are the only free available programs developed for estimating heterogeneous mixed models. The first version HETMIXED was proved to be very slow and limited to small samples due to very large matrices handling and prohibitive computation; it will not be expanded in this work. HETNLMIXED was developed to reduce these computational problems and to allow estimation of both linear and generalized linear models. However, in the linear case, this SAS-macro has the drawback of computing numerically an integral across the random effects while it has a closed form, and thus the macro is limited to a small number of random effects. We have also observed convergence problems when using the macro with large samples except for very simple models.

Moreover, the EM algorithm, which is used in these macros, has some general drawbacks. In particular, it does not have any good convergence criteria; the convergence is only built on a lack of progression of the likelihood or the parameter estimates [7]. Furthermore, the convergence is slow [8] and the EM algorithm does not provide direct estimates of the variance of the parameters. In the particular case of an heterogeneous mixed model, the M-step also requires the estimation of an homogeneous mixed model which is computationally expensive.

Therefore, the first aim of this paper is to propose a program for estimating more general heterogeneous linear mixed models suitable for large samples. The proposed program HETMIXLIN is written in Fortran90 and uses a direct maximization of the likelihood via a Marquardt optimization algorithm. The second objective of this paper is to illustrate the use of heterogeneous linear mixed model through a study of the different patterns of evolution in cognitive ageing.

2. Computational methods and theory

2.1. The heterogeneous linear mixed model

Let $Y_i = (Y_{i1}, \dots, Y_{in_i})$ be the response vector for the n_i measurements of the subject i with $i = 1, \dots, N$. The linear mixed model [1] for the response vector Y_i is defined as:

$$Y_i = X_i \beta + Z_i u_i + \epsilon_i \quad (1)$$

X_i is a $n_i \times p$ design matrix for the p -vector of fixed effects β , and Z_i is a $n_i \times q$ design matrix associated to the q -vector of random effects u_i which represents the subject specific regression coefficients. The errors ϵ_i are assumed to be normally distributed with mean zero and covariance matrix $\sigma^2 I_{n_i}$, and are assumed to be independent from the vector of random effects u_i .

In an homogeneous mixed model [1], u_i is normally distributed with mean μ and covariance matrix D , i.e.

$$u_i \sim N(\mu, D) \quad (2)$$

In the heterogeneous mixed model [2-4], u_i is assumed to follow a mixture of G multivariate Gaussians with different means $(\mu_g)_{g=1,G}$ and a common covariance matrix D , i.e.

$$u_i \sim \sum_{g=1}^G \pi_g N(\mu_g, D) \quad (3)$$

Each component g of the mixture has a probability π_g and the $(\pi_g)_{g=1,G}$ verify the following conditions:

$$0 \leq \pi_g \leq 1 \forall g = 1, G \text{ and } \sum_{g=1}^G \pi_g = 1 \quad (4)$$

In this work, we propose a slightly more general formulation of the model described in (1) in which the effect of some covariates may depend on the components of mixture and some of the random effects may have a common mean whatever the component of mixture. Thus, the X_i design matrix is split in X_{1i} associated with the vector β of fixed effects which are common to all the components and X_{2i} associated with the vectors δ_g of fixed effects which are specific to the components. The Z_i design matrix is also split in Z_{1i} associated with the vector v_i of random effects following a single Gaussian distribution and Z_{2i} associated with the vector u_i of random effects following a mixture of Gaussian distributions. The model is then written as:

$$Y_i = X_{1i} \beta + \sum_{g=1}^G \pi_g X_{2i} \delta_g + Z_{1i} v_i + Z_{2i} u_i + \epsilon_i \quad (5)$$

where $v_i \sim N(0, D_v)$ and $u_i \sim \sum_{g=1}^G \pi_g N(\mu_g, D_u)$; given the component g , the conditional distribution of the vector $\begin{pmatrix} v_i \\ u_i \end{pmatrix}$ is $N\left(\begin{pmatrix} 0 \\ \mu_g \end{pmatrix}, D\right)$ with

$$D = \begin{pmatrix} D_v & D_{vu} \\ D_{uv} & D_u \end{pmatrix}.$$

2.2. Likelihood

Following the previous works [3,4], we define w_{ig} the unobserved variable indicating if the subject i belongs to the component g . We have $P(w_{ig} = 1) = \pi_g$. The density for the vector \mathbf{y}_i can then be written as:

$$f_i(\mathbf{y}_i) = \sum_{g=1}^G \pi_g f(\mathbf{y}_i | w_{ig} = 1) \quad (6)$$

Given w_{ig} , \mathbf{y}_i follows a linear mixed model, and the density $f(\mathbf{y}_i | w_{ig} = 1)$ denoted by ϕ_{ig} is the multivariate Gaussian density with mean E_{ig} and covariance matrix V_i given by:

$$E_{ig} = E(Y_i | w_{ig} = 1) = X_{1i}\beta + X_{2i}\delta_g + Z_{2i}\mu_g \quad \text{and} \\ V_i = \text{var}(Y_i | w_{ig} = 1) = Z_i D Z_i' + \sigma^2 I_{n_i} \quad (7)$$

Let now θ be the vector of the m parameters of the model. θ contains ψ with $\psi' = (\beta', (\delta_g)_{g=1,G}', (\mu_g)_{g=1,G}', \text{Vec}(D)', \sigma^2)$ and π the vector of the $G-1$ first component probabilities $(\pi_g)_{g=1,G-1}$. Note that π_G is entirely determined by π as $1 - \sum_{g=1}^{G-1} \pi_g$. $\text{Vec}(D)$ represents the vector of the upper triangular elements of D . The estimates of θ are obtained as the vector $\hat{\theta}$ that maximizes the observed log-likelihood:

$$L(\mathbf{Y}; \theta) = \sum_{i=1}^N \ln(f_i(\mathbf{y}_i)) = \sum_{i=1}^N \ln \left(\sum_{g=1}^G \pi_g \phi_{ig}(\mathbf{y}_i) \right) \\ = \sum_{i=1}^N -\frac{n_i}{2} \ln(2\pi) - \frac{1}{2} \ln(|V_i|) \\ + \ln \left(\sum_{g=1}^G \pi_g e^{-\frac{1}{2}(\mathbf{y}_i - E_{ig})' V_i^{-1} (\mathbf{y}_i - E_{ig})} \right) \quad (8)$$

2.3. Estimation procedure

We propose to maximize directly the observed log-likelihood (8) using a modified Marquardt optimization algorithm [9], a Newton–Raphson-like algorithm [10]. The diagonal of the Hessian at iteration k , $H^{(k)}$, is inflated to obtain a positive definite matrix as: $H^{*(k)} = (H_{ij}^{*(k)})$ with $H_{ii}^{*(k)} = H_{ii}^{(k)} + \lambda[(1-\eta)|H_{ii}^{(k)}| + \eta \text{tr}(H^{(k)})]$ and $H_{ij}^{*(k)} = H_{ij}^{(k)}$ if $i \neq j$. Initial values for λ and η are $\lambda = 0.01$ and $\eta = 0.01$. They are reduced when H^* is positive definite and increased if not. The estimates $\theta^{(k)}$ are then updated to $\theta^{(k+1)}$ using the current modified

Hessian $H^{*(k)}$ and the current gradient of the parameters $g(\theta^{(k)})$ according to the formula:

$$\theta^{(k+1)} = \theta^{(k)} - \alpha H^{*(k)-1} g(\theta^{(k)}) \quad (9)$$

where if necessary, α which equals 1 by default is modified to ensure that the log-likelihood is improved at each iteration.

To ensure that the covariance matrix D is positive, we maximize the log-likelihood on the non-zero elements of U , the Cholesky factor of D (i.e. $U'U = D$) [7]. Furthermore, to deal with the constraints on π (4), we use the transformed parameters $(\gamma_g)_{g=1,G-1}$ with :

$$\gamma_g = \ln \left(\frac{\pi_g}{\pi_G} \right) \quad (10)$$

Standard errors of the elements of D and $(\pi_g)_{g=1,G-1}$ are computed by the Δ -method [11], while standard errors of the other parameters are directly computed using the inverse of the observed Hessian matrix.

The convergence is reached when the three following convergence criteria are satisfied: $\sum_{j=1}^m (\theta_j^{(k)} - \theta_j^{(k-1)})^2 \leq \epsilon_a$, $|L^{(k)} - L^{(k-1)}| \leq \epsilon_b$ and $g(\theta^{(k)})' H^{(k)-1} g(\theta^{(k)}) \leq \epsilon_d$. The default values are $\epsilon_a = 10^{-5}$, $\epsilon_b = 10^{-5}$ and $\epsilon_d = 10^{-8}$.

As the log-likelihood of a mixture model may have several maxima [8], we use a grid of initial values to find the global maximum. The multimodality of the log-likelihood in mixture models has been often discussed and some authors proposed different strategies to choose the set of initial values [12]. However, none of them seems to be optimal in a general way. We have observed, in our experience, that the results were mainly sensitive to initial values of $(\pi_g)_{g=1,G-1}$ and $(\mu_g)_{g=1,G}$ and less sensitive to the other parameters ($\text{Vec}(U)$, β and σ) for which estimates of the homogeneous mixed models were good initial values.

A mixture model is estimated with a fixed number of components G , otherwise the number of parameters in the model is unknown. To choose the right number of components, one has to estimate models with different values for G and select the best model according to a test or a criterion. Some works favor a bootstrap approach to approximate the asymptotic distribution of the likelihood ratio test between models with different number of components [13] but this approach is very heavy in particular for mixture models with random effects. Criteria such as Akaike's Information Criterion (AIC) [14] or Bayesian Information Criterion (BIC) [15] are often preferred. We use these selection criteria to select the optimal number of components.

2.4. A posteriori classification

After parameter estimation, mixture models allow to classify subjects according to the G components. The classification is based on the posterior probabilities $(\pi_{ig})_{g=1,G}$ that the subject i follows each of the G components. Using $\hat{\theta} = (\hat{\psi}', \hat{\pi}')'$, these probabilities are obtained by the Bayes theorem [2-4] as:

$$\hat{\pi}_{ig} = P(w_{ig} = 1 | Y_i, \hat{\theta}) = \frac{\hat{\pi}_g \phi_{ig}(\hat{\psi}, Y_i)}{\sum_{g=1}^G \hat{\pi}_g \phi_{ig}(\hat{\psi}, Y_i)} \quad (11)$$

We then assign to each subject i the component to which he has the highest probability $(\pi_{ig})_{g=1,G}$ to belong.

3. Program description

The program requires two distinct input files: the data file described in Appendix A and the parameter file named HETMIXLIN.inf which contains the information needed for the estimation of the model: the names of the data file and output files, the number of subjects, the description of the model (number of components G , dimension of the random effects, covariates X_1 , X_2 , Z_1 and Z_2 and covariance structure of D) and the initial values of the parameters. An example of the parameter file is given in Appendix B.

The main output file gives the final log-likelihood, the AIC, the BIC, the convergence criteria, the number of iterations and the parameter estimates with the standard errors, the Wald statistics and the 95% confidence interval. The number of subjects classified in each component is also given.

Finally, another output file contains the posterior probabilities for each subject to belong to each class and the final class membership.

4. Applications

4.1. The height of schoolgirls

We consider the sample of 20 preadolescent schoolgirls introduced by Goldstein [16]. Verbeke and Lesaffre [2] and Komárek et al. [6] modelled the growth curves of their height according to age from 6 to 10. They showed in the homogeneous mixed model that the height course of girls differed significantly according to the category of height of their mother (small, medium and tall). Thus, they used the heterogeneous linear mixed model without introducing the height of the mother in the model to

try to highlight clusters with distinct growth curves among the girls. In this work, we compare the results obtained using our program to those obtained with the HETNLMIXED SAS-macro which uses the EM algorithm. The model is written as:

$$\text{Height}_{ij} = \mathbf{u}_{0i} + \mathbf{u}_{1i} \times \text{age}_{ij} + \epsilon_{ij} \quad (12)$$

where $\mathbf{u}_i = (\mathbf{u}_{0i}, \mathbf{u}_{1i})' \sim \sum_{g=1}^G \pi_g N(\mu_g, D)$ with $\mu_g = (\mu_{0g}, \mu_{1g})'$ and $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$.

We fitted the heterogeneous linear mixed model for two and three components. An extract of the data file and the parameter file for the model with two components are presented in Appendices A and B. The results for the model with two components of mixture obtained with our program HETMIXLIN and the SAS-macro HETNLMIXED are shown in Table 1. The estimates obtained using the two methods are the same but a difference is observed in the standard error estimates; the standard error estimates from HETNLMIXED seem bigger than those from HETMIXLIN. This difference in the standard error estimates from the two algorithms was also observed in the homogeneous case comparing HETNLMIXED, the MIXED procedure, the NLMIXED procedure and HETMIXLIN program. In the three latter programs, standard errors are estimated by the inverse of the Hessian matrix which estimates the Fisher Information matrix [11] and led to the same standard error estimates. By contrast, HETNLMIXED uses an approximation of the Louis's method based on the product of the expectations of the gradient of the complete likelihood [17], the Louis's method [18] being itself an approximation of the observed Hessian matrix. This method appeared to overestimate standard errors in this small sample. However, in our experience, this approximation of the observed Hessian matrix seemed to be improved when the sample size increased. For instance, using a linear mixed model estimated on the 1392 subjects of the PAQUID sample from Section 4.2, the discrepancy was lower.

As the convergence of the two algorithms depends on the choice of the initial values, we fitted the model by the two approaches with the same grid of 32 sets of initial values. The sets differed on all kinds of parameters. HETNLMIXED provided the global maximum 11 times out of the 32 tries and our program found the global maximum 23 times. HETMIXLIN was also faster than the HETNLMIXED SAS-macro (several seconds compared with at least several minutes using a Bi-Xeon 3.06 GHz 1024 MB RAM).

The results for the model with three components of mixture are also shown in Table 1, but we cannot compare our results with those given

Table 1 Estimates and standard errors of the heterogeneous linear mixed model with two components of mixture for the height of schoolgirls using HETMIXLIN (the proposed direct maximization using a Marquardt algorithm) and HETNLMIXED (Spiessens et al. SAS-macro using an EM algorithm) and estimates of the heterogeneous linear mixed model with three components of mixture using HETMIXLIN

Parameter	HETNLMIXED ($G = 2$)		HETMIXLIN ($G = 2$)		HETMIXLIN ($G = 3$)	
	Estimate	S.E. ^a	Estimate	S.E. ^b	Estimate	S.E. ^b
π_1	0.68	0.14	0.68	0.12	0.50	0.18
π_2	0.32		0.32		0.30	0.11
π_3					0.20	
μ_{01}	82.8	1.12	82.8	0.91	84.2	1.18
μ_{11}	5.38	0.091	5.38	0.086	5.32	0.10
μ_{02}	81.9	2.01	81.9	1.52	81.7	1.12
μ_{12}	6.44	0.18	6.44	0.15	6.47	0.12
μ_{03}					79.4	2.19
μ_{13}					5.60	0.20
$\text{var}(\mathbf{u}_{0i})$	6.47	4.94	6.47	3.13	3.50	2.39
$\text{cov}(\mathbf{u}_{0i}, \mathbf{u}_{1i})$	0.13	0.40	0.13	0.35	0.32	0.13
$\text{var}(\mathbf{u}_{1i})$	0.034	0.056	0.034	0.030	0.030	0.024
σ	0.69	0.10	0.69	0.063	0.68	0.06
-2L			166.67		165.94	
AIC			351.35		355.87	
BIC			360.32		367.82	

^a Standard errors obtained using Louis's method.

^b Standard errors obtained using the inverse of the Hessian matrix and the Δ -method for the component probabilities and the variance parameters.

by HETNLMIXED since it converges toward a non-positive definite \hat{D} matrix. Indeed, HETNLMIXED uses the NLMIXED procedure which does not constrain D to be positive definite.

4.2. Cognitive decline in the elderly

The second example illustrates the use of heterogeneous linear mixed models with our estimation method on a large data set as it can be encountered in epidemiological studies. The aim of this analysis is to describe, in a cohort of elderly subjects, the heterogeneity of the evolution of the Mini Mental State Examination (MMSE), the most important psychometric test to evaluate dementia and cognitive impairment, and to compare the classification of subjects stemmed from the mixture model with the dementia diagnosis. The MMSE score ranges from 0 to a maximum of 30 points.

Data come from the French prospective cohort study PAQUID initiated in 1988 to study normal and pathological ageing [19]. The cohort includes 3777 subjects of 65 years and older who lived at home in Southwestern France at baseline. Subjects were interviewed at baseline and were seen again 1 (T1), 3 (T3), 5 (T5), 8 (T8) and 10 (T10) years after

the baseline visit (T0). At each visit, a battery of psychometric tests was completed and a diagnosis of dementia was carried out. In this analysis, we excluded data from T0 because of a learning effect previously described for the cognitive tests between T0 and T1 [20]. We studied the evolution of the MMSE between T1 and T8 for subjects free of dementia till T5 and compared the estimated classification with the dementia diagnosis at T8 and then with the health status at T10. We excluded subjects not seen at T8 to ensure that we had their diagnosis at this visit. This leads to a sample of 1392 subjects having between 1 and 4 measures of the MMSE between T1 and T8.

The model is a quadratic function of time adjusted on covariates associated with cognitive evolution in order to exclude heterogeneity introduced by known factors. The time (t_{ij} for subject i at visit j) is the negative time between the measurement and the visit at T8 (time is zero for diagnosis time at T8). We model the square root of the number of errors to satisfy the normality assumption of the error terms. The model is written as:

$$\begin{aligned}
 Y_{ij} &= \sqrt{30 - \text{MMSE}_{ij}} \\
 &= \beta X_{ij} + \mathbf{u}_{0i} + \mathbf{u}_{1i} t_{ij} + \mathbf{u}_{2i} t_{ij}^2 + \epsilon_{ij}
 \end{aligned} \tag{13}$$

Table 2 Estimates of the homogeneous linear mixed model and the heterogeneous linear mixed model with two components of mixture obtained with HETMIXLIN program for the MMSE evolution adjusted on age, occupation, educational level, living place and interaction with time for age and educational level

Parameter	Homogeneous model ($G = 1$)		Heterogeneous model ($G = 2$)	
	Estimate	S.E.	Estimate	S.E.
π_1	1		0.98	0.005
π_2			0.02	
μ_{01}	-1.92	0.32	-1.51	0.31
μ_{11}	-0.57	0.16	-0.42	0.16
μ_{21}	-0.042	0.018	-0.028	0.024
μ_{02}			0.89	0.36
μ_{12}			0.46	0.19
μ_{22}			0.055	0.028
$\text{var}(\mathbf{u}_{0i})$	0.43	0.025	0.32	0.024
$\text{cov}(\mathbf{u}_{0i}, \mathbf{u}_{1i})$	0.074	0.010	0.034	0.0095
$\text{var}(\mathbf{u}_{1i})$	0.066	0.014	0.029	0.013
$\text{cov}(\mathbf{u}_{0i}, \mathbf{u}_{2i})$	0.042	0.0073	0.028	0.0069
$\text{cov}(\mathbf{u}_{1i}, \mathbf{u}_{2i})$	0.052	0.010	0.038	0.010
$\text{var}(\mathbf{u}_{2i})$	0.073	0.016	0.060	0.016
σ	0.45	0.009	0.45	0.009
-L	4662.9		4609.5	
AIC	9367.7		9269.0	
BIC	9477.8		9400.0	

where X_{ij} is the vector of covariates for subject i at visit j including age, occupation, educational level, living place and interactions with time for age and educational level; $\mathbf{u}_i = (\mathbf{u}_{0i}, \mathbf{u}_{1i}, \mathbf{u}_{2i})' \sim \sum_{g=1}^G \pi_g N(\mu_g, \mathbf{D})$ with $\mu_g = (\mu_{0g}, \mu_{1g}, \mu_{2g})'$ and $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$.

We fitted the heterogeneous linear mixed model with two components of mixture. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were largely improved compared with the homogeneous mixed model ($\Delta\text{AIC} = 98.7$; $\Delta\text{BIC} = 77.8$). Table 2 displays the parameter estimates for the homogeneous and the heterogeneous linear mixed models obtained with HETMIXLIN. The heterogeneous linear mixed model distinguished two different MMSE courses (Fig. 1). First, a large class including 98% of the sample follows a linear evolution with a slight decline of 0.42 points ($\sigma = 0.15$) per year. The second class including 2% of the sample follows a non-linear evolution which speeds to fell down from the second follow-up to the end.

Then, we tried to evaluate if the cognitive profiles highlighted by the model were associated with dementia diagnosis. Among the 1392 subjects, 26

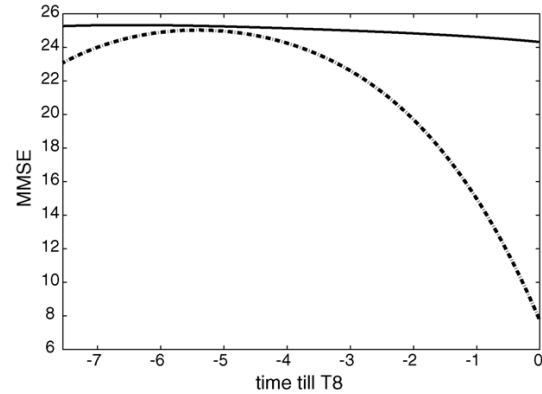


Fig. 1 Mean curves of the MMSE between T1 and T8 for the heterogeneous linear mixed model with two components of mixture for a 70-year-old worker subject with no education and living in Dordogne. Solid line: first component with a probability of 98%; dashed line: second component with a probability of 2%.

(1.9%) are classified in the second component with the non-linear decline (Table 3). Among them, 21 have a positive dementia diagnosis at T8. The predictive positive value (81%) and the specificity (99.6%) of this classification are high but the sensitivity is poor (31%): 47 subjects among the 68 subjects diagnosed as demented at T8 are not detected by the model. These subjects are significantly less disabled ($p < 0.0001$) than the other demented subjects (23% of disabled people versus 76%) and have a significantly lower educational level ($p = 0.001$): 51% have no education or no diploma from primary school versus 9.5% in the demented group detected in the model. All of the five subjects without dementia but classified in the declining group are disabled, two die in the 2 years and one has a positive dementia diagnosis at T10.

The association between cognitive ageing and educational level is an important issue [21]: educa-

Table 3 Relationship between classification stemmed from the heterogeneous linear mixed model with two components and dementia diagnosis at T8

Dementia diagnosis at T8	Classification		Total
	Linear class	Non-linear class	
Positive	47	21	68
Negative	1319	5	1324
Total	1366	26	1392

tional level could have a different effect on normal cognitive ageing and on the decline before dementia. As HETMIXLIN allows to specify distinct parameters for the covariates per component, we fitted the model where the interactions between educational level and time (educational level $\times t$ and educational level $\times t^2$) were different according to the components. The log-likelihood was not improved enough for significance ($\Delta(-2L) = 3.4$; $p = 0.17$). Thus, the association between educational level and cognitive evolution appears to be similar in the two subpopulations. Moreover, the discrimination of demented people was not improved with this model.

Due to limitations of HETNLMIXED explained in Section 1, we did not compare our approach with this program on the PAQUID data set. However, as Newton–Raphson algorithms have been criticized on their global convergence behavior compared with the EM algorithm [8], we compared the convergence performances of our program with those of an EM algorithm we developed in Fortran90. Thus, the comparison was free from the limitations due to SAS environment and to the use of the NLMIXED procedure. The EM algorithm we developed uses a Marquardt optimization in the M-step (convergence criteria: $\epsilon_a = 10^{-2}$, $\epsilon_b = 10^{-2}$ and $\epsilon_d = 10^{-3}$) and the global convergence is reached when two successive calculations of the likelihood differ less than 10^{-8} . This algorithm was tested using the schoolgirls data. It was faster and converged more often than HETNLMIXED: the global maximum was reached 22 times out of the 32 tries with a mean computational time around 30s (versus 11 times out of the 32 tries in at least several minutes for HETNLMIXED). On the PAQUID data set, HETMIXLIN and the EM algorithm we implemented led to the same parameter estimates. Among the 15 sets of initial values, the two programs provided the global maximum an equivalent number of times (9 times for HETMIXLIN versus 10 times for the EM algorithm) but HETMIXLIN was much faster: the CPU time was less than 10min for HETMIXLIN and more than 2h for the EM algorithm.

5. Availability of the program and hardware specification

The program HETMIXLIN is written in Fortran90 and all the subroutines needed in the program are provided. The Fortran source code HETMIXLIN.f, an example of HETMIXLIN.inf, a documentation HETMIXLIN.pdf and the example data file for the school-

girls are available at no charge on the web site: <http://www.isped.u-bordeaux2.fr>.

Two versions are provided on the web site: one for Unix and one for Windows. The version for Windows includes an executable file (a DOS application) and does not need any Fortran90 compiler whereas the version for Unix needs to be compiled. The Unix version has been tested using an Intel Fortran Compiler for Linux Version 7 or 8, a Fortran90 Compaq compiler for Alpha and a Forte Developer 6 update 2 on Solaris SPARC. Examples of the compilation command are given in the documentation HETMIXLIN.pdf.

6. Conclusion

We proposed in this paper a Newton–Raphson-like algorithm to estimate heterogeneous linear mixed models. The main advantages of Newton–Raphson-like algorithms are the speed of convergence, the availability of good convergence criteria based on the derivatives of the likelihood and direct estimates of the variance of the parameters via the inverse of the Hessian matrix. Moreover, using a simple modification of the Marquardt algorithm, we ensure the monotonicity of the algorithm which is considered as a main advantage of the EM algorithm [8].

We compared our program HETMIXLIN with a SAS-macro developed by Spiessens et al. using an EM algorithm. This SAS-macro allows to estimate heterogeneous generalized linear mixed models, but when the model is linear, this macro has the drawback of computing numerically an integral across the random effects while it has a closed form. Our algorithm HETMIXLIN allows to estimate more complex linear models (models with a larger number of mixture components, a larger number of random effects and more covariate effects depending on the mixture components) and is suitable for much larger samples. Moreover, it converges faster.

This paper also illustrates the usefulness of heterogeneous linear mixed models on a study about cognitive ageing. These models allow to highlight various evolution profiles taking covariates into account. The cross-classification of the groups defined by the model and clinical events in the next years enables to evaluate whether the cognitive profiles are associated with different clinical evolutions.

As a conclusion, we hope this work will improve the availability and the use of heterogeneous linear mixed models.

Appendix A. Extract of the Schoolgirl data file (two first subjects)

```

1          ← identification number of the unit
           (subject)
5          ← number of measures
111 116.4 ← raw vector of the  $n_i$  responses
    121.7 126.3
    130.5
1 1 1 1 1 ← raw vector of the first covariate
6 7 8 9 10 ← raw vector of the second covariate
2          ← identification number of the next unit
           (subject)
5          ← number of measures
110 115.8 ← raw vector of the  $n_i$  responses
    121.5 126.6
    131.4
1 1 1 1 1 ← raw vector of the first covariate
6 7 8 9 10 ← raw vector of the second covariate
3          ← identification number of the next unit
           (subject)

```

Appendix B. Example of the parameter file

An example of HETMIXLIN.inf used in the application about the height of schoolgirls is given below. The user should notice that each asked piece of information is preceded by a line summing it up.

```

→ Filename for the data
schoolgirls.txt
→ Filename for the output
girls.out
→ Title of the procedure
(in inverted commas)
'G=2 : school girls'
→ Number of units (subjects)
20
→ Number of mixture components (G) and,
if and only if G>1, the initial
values for the G-1 first component
probabilities below and the filename
for the posterior probabilities
below again
2
0.5
p.out
→ Number of explanatory variables
(including the intercept) in the
data file
2
→ Indicator that the explanatory
variable is in the model
(1 if present 0 if not)
1 1

```

```

→ Indicator of random effect for each
variable in the model (variables
included in Z1 or Z2)
1 1
→ Indicator of mixture for each
variable in the model (variables
included in X2 or Z2)
1 1
→ Initial values for fixed effects.
First, initial values for common
fixed effects (without mixture) in
the same order as in the datafile,
then initial values for the
covariates with a mixture (G values
per covariate). ex : b1 b3 b21 b22
b23 b41 b42 b43 for a mixture on the
second and the fourth covariate and
G=3
86 80 5 7
→ Indicator of the random effect
covariance matrix structure (0 if
unstructured matrix/1 if diagonal
matrix)
0
→ Initial values for the
variance---covariance parameters of
the random effects (1/2 superior
matrix column by column)
3 1 1
→ Initial value for the variance of
the independent Gaussian errors
1

```

References

- [1] N.M. Laird, J.H. Ware, Random-effects models for longitudinal data, *Biometrics* 38 (1982) 963–974.
- [2] G. Verbeke, E. Lesaffre, A linear mixed-effects model with heterogeneity in the random-effects population, *JASA* 91 (1996) 217–221.
- [3] B. Spiessens, G. Verbeke, A. Komárek, A SAS-macro for the classification of longitudinal profiles using mixtures of normal distributions in nonlinear and generalized linear models. <http://www.med.kuleuven.ac.be/biostat/research/software.htm>, 2002.
- [4] B. Muthén, K. Shedden, Finite mixture modeling with mixture outcomes using a EM algorithm, *Biometrics* 55 (1999) 463–469.
- [5] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via EM algorithm (with discussion), *J. Roy. Statist. Soc. Ser. B* 39 (1977) 1–38.
- [6] A. Komárek, A SAS-macro for linear mixed models with a finite normal mixture as random-effects distribution. <http://www.med.kuleuven.ac.be/biostat/research/software.htm>, 2001.
- [7] M.J. Linstrom, D.M. Bates, Newton–Raphson and EM algorithms for linear mixed models for repeated-measures data, *JASA* 83 (1988) 1014–1022.

- [8] R.A. Redner, H.F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Rev.* 26 (1984) 195–239.
- [9] D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *SIAM J. Appl. Math.* 11 (1963) 431–441.
- [10] R. Fletcher, *Practical Methods of Optimization*, second ed., John Wiley & Sons, 2000 (Chapter 3).
- [11] K. Knight, *Mathematical Statistics*, Chapman & Hall/CRC, 2000 (Chapters 3 and 5).
- [12] D. Karlis, E. Xekalaki, Choosing initial values for the EM algorithm for finite mixtures, *Comput. Statist. Data Anal.* 41 (2003) 577–590.
- [13] P. Schlattmann, Estimating the number of components in a finite mixture model: the special case of homogeneity, *Comput. Statist. Data Anal.* 41 (2003) 441–451.
- [14] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automat. Control* 19 (1974) 716–723.
- [15] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* 6 (1978) 461–464.
- [16] H. Goldstein, *The Design and Analysis of Longitudinal Studies*, Academic Press, London, 1979.
- [17] G.J. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, 1997.
- [18] T.A. Louis, Finding the Observed Information Matrix when Using the EM algorithm, *J. Roy. Statist. Soc. Ser. B* 44 (1982) 226–233.
- [19] L. Letenneur, D. Commenges, J.F. Dartigues, P. Barberger-Gateau, Incidence of dementia and Alzheimer’s disease in elderly community residents of south-western France, *Int. J. Epidemiol.* 23 (1994) 1256–1261.
- [20] H. Jacqmin-Gadda, C. Fabrigoule, D. Commenges, J.F. Dartigues, A 5-year longitudinal study of the Mini Mental State Examination in normal aging, *Am. J. Epidemiol.* 145 (1997) 498–506.
- [21] L. Letenneur, V. Gilleron, D. Commenges, C. Helmer, J.M. Orgogozo, J.F. Dartigues, Are sex and educational level independent predictors of dementia and Alzheimer’s disease? Incidence data from the PAQUID project, *J. Neurol. Neurosurg. Psychiatr.* 66 (1999) 177–183.



Practice of Epidemiology

Sensitivity of Four Psychometric Tests to Measure Cognitive Changes in Brain Aging-Population-based Studies

Cécile Proust-Lima^{1,2}, Hélène Amieva^{2,3}, Jean-François Dartigues^{2,3}, and Hélène Jacqmin-Gadda^{1,2}

¹ INSERM E0338, Bordeaux, France.

² Université Victor Segalen Bordeaux 2, Bordeaux, France.

³ INSERM U593, Bordeaux, France.

Received for publication April 11, 2006; accepted for publication July 5, 2006.

Choosing the measure of cognition in an epidemiologic study investigating cognitive changes over time is a challenging question. A powerful measure must be able to detect small cognitive changes in all the range of cognition observed in the target population. This work aims at comparing the sensitivity to detect cognitive changes in the observed range of cognition of four widely used psychometric tests in an aging-population-based study through a nonlinear latent process model, assuming that the psychometric tests are nonlinear noisy transformations of their common factor. With data from the French prospective cohort study PAQUID (1989–2001), the authors found that the Mini-Mental State Examination and the Benton Visual Retention Test exhibited a better sensitivity to cognitive changes in low levels of cognition, while the Digit Symbol Substitution Test was more sensitive to changes in high levels of cognition. In contrast, the Isaacs Set Test shortened at 15 seconds appeared to be sensitive to small changes in all the range of cognition and, thus, represents an appropriate measure of cognition in population-based studies including both highly normal and severely impaired subjects.

aging; cognition; dementia; epidemiologic methods; neuropsychological tests

Abbreviations: BVRT, Benton Visual Retention Test; DSST, Digit Symbol Substitution Test; IQR, interquartile range; IST15, Isaacs Set Test shortened at 15 seconds; MMSE, Mini-Mental State Examination.

With the increasing number of people suffering from neurodegenerative diseases, especially Alzheimer's disease, investigating cognitive changes over time has received growing attention in population-based cohort studies for understanding the natural history of the neurodegenerative diseases and in intervention trials designed to assess the effects of drugs on neuropsychological functioning (1, 2). In these studies, cognition is generally assessed through a battery of psychometric tests repeatedly administered to the subjects. Collecting several cognitive tests may be useful, because this allows exploration of the various cognitive domains (memory functioning, attention, or executive functions) and because the tests often have different metrologic properties. In particular, some tests are very sensitive to small changes at

high levels of cognition, while others are more sensitive to changes at the lower levels. As a consequence, when studying the effects of drugs or the aging process on cognition, results may differ considerably according to the properties of the test being used (3). A way to deal with this problem is to study the change over time of the common latent cognitive level underlying the battery of cognitive tests used (4).

However, administering an extensive battery of neuropsychological tests can turn out to be difficult, because of the duration (and thus cost) of the evaluation and also because subjects presenting cognitive impairment are more prone to refuse long testing sessions than cognitively intact elderly subjects (5). In this way, it would be of substantial interest to compare the properties of commonly used cognitive tests

to highlight arguments for selecting a restricted number of tests and even only one according to the aim of the study.

Brevity and ease of use are obviously relevant criteria by which to select a test for population-based studies. Nonetheless, it is more crucial to select a test able to detect small changes in cognition in all the range of cognitive levels observed in the target population (1, 2), particularly in cases of prolonged follow-up or long-term prevention studies. Indeed, the range of general cognitive level targeted will be different according to whether the study sample consists of subjects from population-based cohorts, patients consulting memory clinics for memory troubles, or patients with Alzheimer's disease enrolled in pharmacologic trials. Until now, when a single psychometric test is collected, the Mini-Mental State Examination (MMSE) (6) is usually preferred, because it gives a brief measure of global cognitive functioning (7). However, as the MMSE suffers from a strong ceiling effect, it is not suitable to identify slight declines in high levels of cognition (7, 8) and, thus, is not appropriate to study normal cognitive aging in prospective studies, particularly among highly educated people.

The aim of this work is to compare the sensitivity to cognitive change of four tests widely used in clinical practice: the MMSE, the Isaacs Set Test, the Benton Visual Retention Test, and the Digit Symbol Substitution Test. More specifically, we would like to identify the most appropriate test to measure cognitive changes over time in heterogeneous populations, including both highly normal and severely impaired subjects, as encountered in population-based studies. To answer this purpose, we use a nonlinear latent variable model for longitudinal multivariate data in which psychometric tests are assumed to be nonlinear parameterized transformations of a common factor (4). The common factor is a latent process representing the latent cognitive process underlying the psychometric tests. It is related to age through a linear mixed model for describing the latent cognitive evolution according to age. By estimating flexible transformations between the psychometric tests and the common factor, we are able to compare the metrologic properties of the psychometric tests.

This approach may be viewed as an extension of Item Response Theory (9) to handle repeated measurements of quantitative outcomes (the summary scores of each test) instead of binary outcomes (the individual binary items of one test). The link functions estimate the mean score of each test given the values of the latent process similarly to the Item Characteristic Curve in Item Response Theory models that estimates the probability of correct response given the latent ability. In addition, we are able to estimate evolution with time of the latent process and the test scores.

MATERIALS AND METHODS

Population

PAQUID is a prospective cohort study initiated in 1988 in southwestern France (Dordogne and Gironde) to explore functional and cerebral aging. In brief, 3,777 subjects who were 65 years or older and were living at home at enrollment were included in the cohort and were followed up for six times with a visit at 1 year, 3 years, 5 years, 8 years,

10 years, and 13 years after the initial visit, except at 1 year in Dordogne. At each visit, a neuropsychological evaluation and a diagnosis of dementia were carried out at home. Letenneur et al. (10) offer a detailed description of the PAQUID program.

Neuropsychological evaluation

In PAQUID, a battery of psychometric tests was used to evaluate cognition. In this paper, we focus on the following four tests.

- The Mini-Mental State Examination (6) is a summed score evaluating various dimensions of cognition (memory, calculation, orientation in space and time, language, and word recognition). It is used as an index of global cognitive performance and ranges from 0 to 30.
- The Isaacs Set Test shortened at 15 seconds (11) evaluates verbal fluency abilities and speed of verbal production. Subjects have to give a list of words (with a maximum of 10) belonging to a specific semantic category in 15 seconds. Four semantic categories were successively used (cities, fruits, animals, and colors). The score ranges from 0 to 40.
- The Benton Visual Retention Test (12) evaluates immediate visual memory. After the presentation for 10 seconds of a stimulus card displaying a geometric figure, subjects are asked to choose the initial figure among four possibilities; 15 figures are successively presented. The score ranges from 0 to 15.
- The Digit Symbol Substitution Test (13) explores attention and psychomotor speed. Given a code table displaying the correspondence between pairs of digits (from 1 to 9) and symbols, the subjects have to fill in blank squares with the symbol that is paired to the digit displayed above the square. The subjects have to fill in as many squares as possible in 90 seconds. In PAQUID, the score ranges from 0 to 76 even if the theoretical maximum is 90.

Statistical model

The statistical model assumes that the correlation between the tests is induced by a latent common cognitive factor. Thus, each test is a noisy measure of a test-specific nonlinear transformation of the common factor. The evolution of the common factor was modeled using a linear mixed model (14) that aims at evaluating changes over time of a repeated outcome (here, the latent common factor), accounting for correlation between the repeated measures on each subject. The linear mixed model included random intercept, age, and age squared in accordance with other longitudinal aging studies (15, 16) that showed quadratic cognitive evolutions. We added a Brownian motion to account for individual deviations from this quadratic evolution and thus relaxed the parametric form of the model. Test-specific random intercepts were added to evaluate if, for a same level of the latent common factor, two subjects scored differently at the tests.

We used flexible nonlinear transformations to link each psychometric test with the latent common factor. The chosen flexible transformations were beta cumulative distribution functions that offer a large variety of shapes (concave, convex, or sigmoid) using only two estimated parameters per

TABLE 1. Demographic and health characteristics in the sample ($n = 2,252$), the PAQUID Study, France, 1989–2001

	No.	%
Male	974	43.3
No diploma	608	27.0
Age (years) at visit at 1 year		
≤ 70	1,100	48.9
$>70-75$	368	16.3
$>75-80$	412	18.3
>80	372	16.6
Prevalent dementia at visit at 1 year	14	0.6
Incident dementia between visits at 3 and 13 years	368	16.3

test. The estimated beta cumulative distribution functions represent the shapes of the transformations between the latent common factor and each of the four tests and provide interesting metrologic properties for each test. The complete methodology was previously detailed in the report by Proust et al. (4) and is described in the Appendix.

Sample

Measurements at the initial visit were excluded from the analysis because of a first passing effect previously described (17). Indeed a test-retest improvement was observed between the initial visit and the visit at 1 year, possibly due to the stress induced by the first evaluation at the initial visit or a practice effect observed after the first evaluation. Of the 3,777 subjects in PAQUID, we retained 3,043 subjects who were followed up after the initial visit, and we studied the evolution of the common factor between the visits at 1 and 13 years. The Digit Symbol Substitution Test (DSST) was not completed at the visit at 3 years. Moreover, a subsample of PAQUID completed a nutritional questionnaire at the visit at 3 years that could impact on the Isaacs Set Test shortened at 15 seconds (IST15) score for the fruit and animal categories. So, we excluded the measurements of IST15 at the visit at 3 years. From the sample of 3,043 subjects, we retained a sample of 2,252 subjects with at least one measure at each neuropsychological test between visits at 1 and 13 years to ensure that every subject brought information about each neuropsychological test. The median number of measures was four for the MMSE (interquartile range (IQR): 2–5) and the Benton Visual Retention Test (BVRT) (IQR: 2–5) and three and two for the IST15 (IQR: 1–4) and the DSST (IQR: 1–4), respectively.

RESULTS

Description of the sample

Characteristics of the subjects in the sample are described in table 1: 56.7 percent of the subjects were women, and 73 percent had at least graduated from primary school. The mean age at the beginning of the follow-up (visit at 1 year) was 75.16 (standard error: 6.24) years. Among the 2,252

subjects, 14 had a positive diagnosis of dementia at the visit at 1 year, and 368 subjects were diagnosed as demented during the follow-up. At the 1-year visit, 1,623 subjects completed the MMSE with a median score of 28 (IQR: 26–29), 1,622 subjects completed the IST15 with a median score of 28 (IQR: 24–33), 1,619 subjects completed the BVRT with a median score of 11 (IQR: 9–13), and 1,614 completed the DSST with a median score of 28 (IQR: 20–37) (table 2).

Cognitive evolution with age

The latent common cognitive factor ranges from 0 to 1, with 0 corresponding to a minimal cognitive performance and 1 corresponding to a maximal cognitive performance. The predicted mean evolution of the latent common cognitive factor and its 95 percent confidence interval are displayed in figure 1, part A. It shows a decline of cognitive performances in the population according to age, with an acceleration of the decline at older ages. This evolution can be interpreted as the mean cognitive decline with age in a population of subjects aged 65 years or older, including both demented and nondemented subjects.

The predicted mean evolutions of the four tests and their 95 percent confidence intervals are displayed in figure 1, part B. The decline of the DSST is quite linear, whereas the decline observed in the other tests is nonlinear, with an acceleration at older ages.

Comparison of the sensitivity of the four tests

The estimated transformation shapes between each test and the common factor are displayed in figure 2. They are very different for the four tests: The transformation is concave for the MMSE and the BVRT, convex for the DSST, and close to linear for the IST15. Moreover, the MMSE and the BVRT cover, respectively, only 78 percent and 85 percent of the common factor range, while the DSST and the IST15 cover almost all of the entire range with, respectively, 94 percent and 95 percent of the common factor range.

Ceiling effect. For the MMSE and the BVRT, the maximum value of the test is reached for a value of the common factor lower than its maximum (respectively, 0.78 and 0.87), which shows that these tests are not appropriate to assess cognitive changes in subjects with a very high general cognitive level of functioning. This underlines the ceiling effect of these two tests. In contrast, the maximum values of the IST15 and the DSST are reached for a value of the common factor close to 1, underlining that these tests do not suffer from a ceiling effect and can identify changes at the highest levels of cognition.

The MMSE transformation. The nonlinear shape of the MMSE transformation shows that a decline of the MMSE does not correspond to the same intensity of decline of the common factor in all the range of the test. Indeed, a loss of 0.1 in the common factor scale between 0.7 and 0.6 represents a loss of 1 point for the MMSE, whereas the same loss between 0.3 and 0.2 represents a loss of more than 4 points for the MMSE. This shows that a decline in the MMSE cannot be interpreted without accounting for the initial value and that the MMSE is more appropriate to detect a decline at low cognitive levels than at high cognitive levels.

TABLE 2. Description of the cognitive performances for the four psychometric tests at the visit at 1 year, with number of measures, range, median with interquartile range, and mean with standard error, the PAQUID Study, France, 1989–2001

	No. of measures	Range (score)	Median score (interquartile range)	Mean score (standard error)
Mini-Mental State Examination	1,623	12–30	28 (26–29)	27.09 (2.51)
Isaacs Set Test shortened at 15 seconds	1,622	9–40	28 (24–33)	28.46 (6.05)
Benton Visual Retention Test	1,619	0–15	11 (9–11)	10.97 (2.53)
Digit Symbol Substitution Test	1,614	0–76	28 (20–37)	28.60 (12.15)

The BVRT transformation. The shape of the BVRT transformation is quite linear at the low cognitive levels (corresponding to a value under 10 at the BVRT or under 0.5 on the

common factor scale). It reveals that, under a value of 0.5 for the common factor, a decline in the BVRT represents the same intensity of decline for the common factor whatever

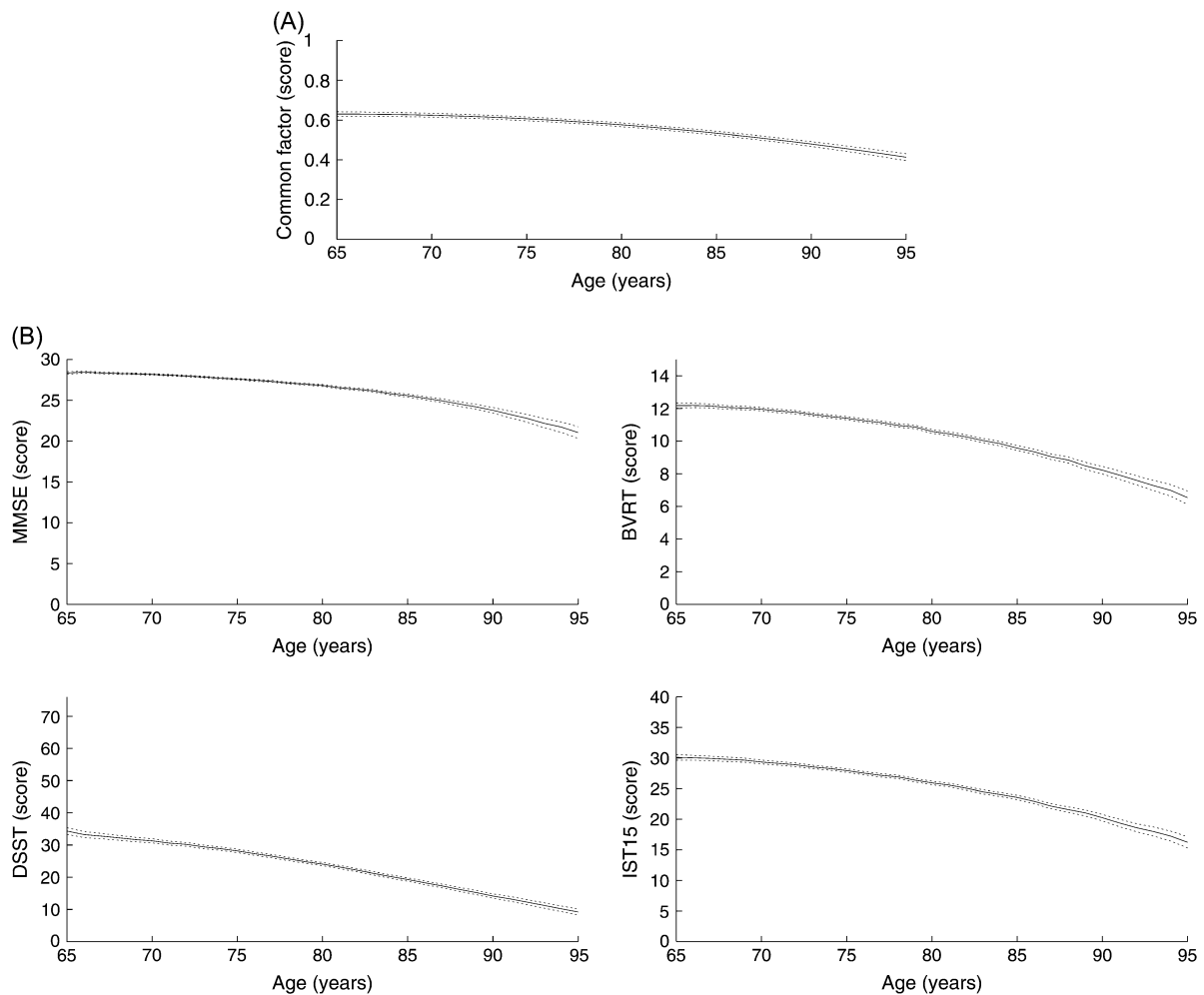


FIGURE 1. Predicted mean evolution (solid line) with 95% confidence interval (dashed lines) for the common factor (A) and for the four tests (Mini-Mental State Examination (MMSE), Benton Visual Retention Test (BVRT), Digital Symbol Substitution Test (DSST), and Isaacs Set Test shortened at 15 seconds (IST15)) (B), the PAQUID Study, France, 1989–2001.

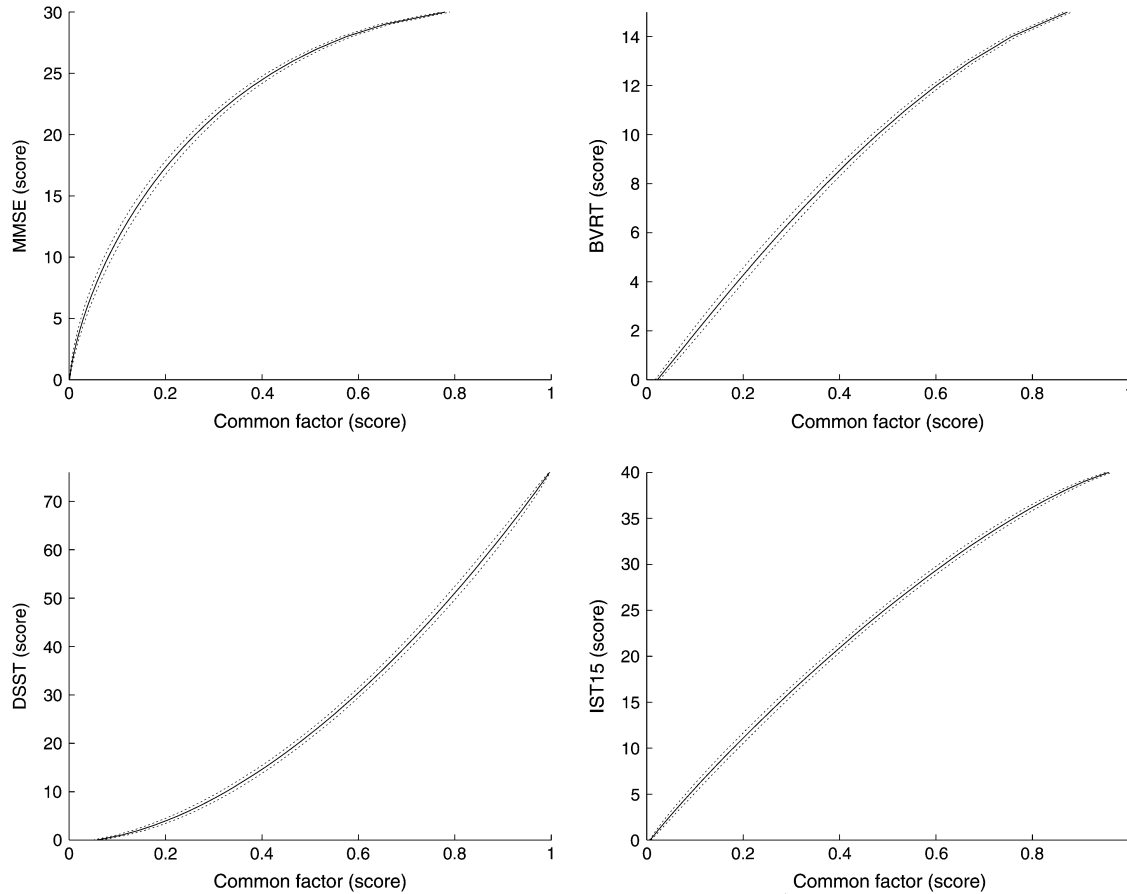


FIGURE 2. Estimated link function (solid line) between the common factor and each of the four tests (Mini-Mental State Examination (MMSE), Benton Visual Retention Test (BVRT), Digital Symbol Substitution Test (DSST), and Isaacs Set Test shortened at 15 seconds (IST15)) with the 95% pointwise confidence interval (dashed lines), the PAQUID Study, France, 1989–2001.

the initial cognitive level. However, at high levels of cognition, the slope of the curve decreases. Thus, at the higher levels of cognition, the BVRT is less sensitive to small cognitive changes.

The DSST transformation. The shape of the DSST transformation is quite linear at the high levels of cognition, corresponding to a score higher than 25 or a common factor level higher than 0.54. This means that, above this value, whatever the initial cognitive level, a decline of the DSST corresponds to the same decline in the latent cognitive level. In contrast, at the low levels of cognition, the shape is nonlinear, revealing that 1 point lost at the DSST corresponds to a greater decline of the common factor level under a score of 25. In this way, subjects with a very low level of general cognitive functioning (lower than 0.1) tend to score 0 on the DSST. This could be explained by the fact that subjects with a low level of cognition can probably not even understand the instructions of the DSST. Thus, DSST is appropriate to identify small

changes in cognition among subjects with a high to medium general cognitive level but not among impaired subjects.

The IST15 transformation. In contrast with the other tests, the IST15 transformation is close to linear for the entire range of cognition. This shows that the IST15 proves to be sensitive to change in all the ranges of cognition; 1 point lost at the IST15 score in higher values of the score has the same interpretation in terms of cognitive decline as 1 point lost in the lower values of the score. Thus, the IST15 can be useful for evaluating cognition in a heterogeneous population, including both subjects with a high general cognitive functioning and subjects with a severe cognitive impairment.

DISCUSSION

By modeling jointly four psychometric tests and their latent common factor, we have been able to compare their

distributions in all the range of cognition. In this way, we showed that MMSE and BVRT were not sensitive to cognitive changes at high levels of cognition and thus were not appropriate to study cognitive aging in prospective studies including highly educated people. On the contrary, we showed that the DSST was very sensitive to cognitive changes at high levels of cognition. However, as it was less sensitive to cognitive changes at low levels of cognition, it was not suitable for measuring cognitive changes in heterogeneous populations consisting of both highly normal and severely impaired subjects. In contrast, the IST15 appeared to be a satisfactory cognitive measure in all the range of cognition, which is of substantial interest when studying cognitive aging in population-based cohort studies.

The IST15 has several assets when compared with the three other tests. First, it does not suffer from a floor effect or a ceiling effect. Indeed, using cognitive measures with border effects can lead to misleading results (especially underestimated declines) when investigating cognitive changes, since initial scores are often differentially distributed among exposure groups and the sensitivity of the tests to identify cognitive changes is thus different among these groups (1). Second, the Isaacs Set Test when shortened at 15 seconds, as well as the DSST, includes a speed component that may explain its high sensitivity to changes at upper levels of cognition. Indeed, the speed component plays a key role in cognitive aging, and it has been shown, for example, that most age-related differences in cognition were due to the decrease in the processing speed (18). Finally, the IST15 is a very brief test, and its instructions are easily understandable. It can therefore be performed in large population-based studies even with severely impaired subjects.

The methodology we proposed in this paper has several advantages that should be discussed. First, the estimated link functions between the test scores and the latent process make it possible to compare properties of the tests and, especially, their sensitivity to detect cognitive changes within the entire range of cognition. This is done by modeling jointly various psychometric tests for which the hypothesis of a common factor is sensible. By the way, it is worth noting that the latent common factor in this model is actually defined according to the pool of psychometric tests used in the analysis. Computing the model with other tests involving different cognitive components could have an impact on the common factor evolution. In this analysis, we used tests that both are frequently used and explore different domains of cognition, because we wanted to select one test for exploring general cognitive decline in heterogeneous populations. The methodology could also be used for selecting sensitive measures in a specific domain of cognition. In this case, based on his/her knowledge or on other analyses such as principal component analyses, the researcher must choose the tests that are assumed to measure the same latent cognitive ability in this specific domain and then apply the methodology to the selected tests.

A second asset of the methodology is that, thanks to the estimated transformations of tests, the tests are no longer constrained to follow a Gaussian distribution as in a standard linear mixed model. In this way, even if longitudinal evolutions of the four tests, as presented in figure 1, part B, could

have actually been estimated using linear mixed models, they would have been obtained under the wrong Gaussian assumption.

Finally, as parameters are estimated using the maximum likelihood estimators, results are robust to data missing at random (i.e., when the probability that data are missing does not depend on unobserved values given the past observed values). Simpler analyses that aim at comparing empirical means of the tests for different age groups are often biased by the missing data process, especially when the cognitive level and the dropout are linked, as was previously shown in the PAQUID cohort (17). In this previous work, it was also shown that the missing at random assumption was probably not strictly true, but the impact on the estimated evolution was slight (17, 19). Moreover, even if missing data may blur the comparison of evolution of the tests' scores, it is very unlikely that they biased the comparison of test sensitivity, which is the main objective of this paper. This was checked by comparing transformations estimated on four subsamples defined by the time of dropout (dropout after the visits at 3, 5, 8, and 10 years or complete follow-up) in the spirit of pattern mixture analysis (20). Whatever the pattern of dropout, the estimated transformations were very similar (results not shown).

Some methodological issues of this analysis should, however, be discussed. First, as the results rely on a parametric model, adequate fit of the model to the data has been carefully checked using postfit methods based on the residuals and the predictions developed in Proust et al. (4) (results not shown). An essential part of the model is the link function between the tests and the common factor. The beta cumulative distribution function was chosen because this transformation was flexible enough to exhibit very different shapes and depended on only two parameters per test. However, complementary analyses have been performed estimating the link functions on a basis of splines instead of the beta cumulative distribution functions; they have led to very similar results while raising more numerical problems due to the large number of parameters.

Second, in the PAQUID Study, MMSE was the first test fulfilled during each testing session. Consequently, it was more frequently completed than the three other tests, particularly among impaired subjects. To ensure that test-specific parameters were estimated on the same sample and to maintain comparability between the tests, we required that every subject had at least one measure at each test. The 791 subjects excluded from the sample were older (median age: 78.6 vs. 73.1 years) and less educated (51.5 percent did not graduate from primary school vs. 27 percent in the sample) than the subjects included in the selected sample, but the range of the observed scores was the same. Note also that using longitudinal data and keeping incident cases of dementia in the sample increased the observed range of cognition and allowed us to compare evolution of each test over time.

In conclusion, our results show that the Isaacs Set Test shortened at 15 seconds could be a good candidate to measure cognitive changes in a general population. More generally, the methodology used in this study provides some clues to thoughtfully select the appropriate measures of cognition collected in a study according to the nature of the target population and the objective of the study.

ACKNOWLEDGMENTS

This study was funded by Novartis AG, SCOR insurance Agrica, Conseil Général de la Gironde, and Conseil général de la Dordogne.

Conflict of interest: none declared.

REFERENCES

1. Morris MC, Evans DA, Hebert LE, et al. Methodological issues in the study of cognitive decline. *Am J Epidemiol* 1999; 149:789–93.
2. Yesavage JA, Brooks JO 3rd. On the importance of longitudinal research in Alzheimer's disease. *J Am Geriatr Soc* 1991; 39:942–4.
3. Galasko DR, Gould RL, Abramson IS, et al. Measuring cognitive change in a cohort of patients with Alzheimer's disease. *Stat Med* 2000;19:1421–32.
4. Proust C, Jacqmin-Gadda H, Taylor JM, et al. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*. (Advance Access: doi: 10.1111/j.1541-0420.2006.00573.x.).
5. Dartigues JF, Commenges D, Letenneur D, et al. Cognitive predictors of dementia in elderly community residents. *Neuroepidemiology* 1997;16:29–39.
6. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–98.
7. Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: a comprehensive review. *J Am Geriatr Soc* 1992;40: 922–35.
8. Commenges D, Gagnon M, Letenneur L, et al. Statistical description of the Mini-Mental State Examination for French elderly community residents. Paquid Study Group. *J Nerv Ment Dis* 1992;180:28–32.
9. Baker F. The basics of Item Response Theory. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, 2001.
10. Letenneur L, Commenges D, Dartigues JF, et al. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *Int J Epidemiol* 1994;23: 1256–61.
11. Isaacs B, Kennie AT. The Set test as an aid to the detection of dementia in old people. *Br J Psychiatry* 1973;123:467–70.
12. Benton A. Manuel pour l'application du Test de Rétention Visuelle. Applications cliniques et expérimentales. (In French). Paris, France: Centre de Psychologie appliquée, 1965.
13. Wechsler D. WAIS-R manual. New York, NY: Psychological Corporation, 1981.
14. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–74.
15. Hall CB, Lipton RB, Sliwinski M, et al. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Stat Med* 2000;19:1555–66.
16. Amieva H, Jacqmin-Gadda H, Orgogozo JM, et al. The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain* 2005;128: 1093–101.
17. Jacqmin-Gadda H, Fabrigoule C, Commenges D, et al. A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *Am J Epidemiol* 1997;145:498–506.
18. Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychol Rev* 1996;103:403–28.
19. Jacqmin-Gadda H, Commenges D, Dartigues J. Analysis of longitudinal Gaussian data with missing data on the response variable. (In French). *Rev Epidemiol Sante Publique* 1999; 47:525–34.
20. Little R. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc* 1993;88:125–34.

APPENDIX

Model specification

We consider K neuropsychological tests. For each test k , $k = 1, \dots, K$, each subject i , $i = 1, \dots, N$, and each occasion j , $j = 1, \dots, n_{ik}$, the measure of the neuropsychological test y_{ijk} is collected at time t_{ijk} , t_{ijk} being different for each test and each subject. The latent process that represents the common factor of the K neuropsychological tests is modeled by use of the following linear mixed model, including a quadratic function of time and a Brownian motion $(w_i(t))_{t \geq 0}$ with variance term $\sigma_w^2 \times t$:

$$\Lambda_i(t) = (\mu_0 + u_{0i}) + (\mu_1 + u_{1i}) \times t + (\mu_2 + u_{2i}) \times t^2 + w_i(t).$$

The vector of random effects $\mathbf{u}_i = (u_{0i}, u_{1i}, u_{2i})^T$ follows a multivariate normal distribution with mean vector $\mathbf{0}$ and variance covariance matrix D . The mean evolution of the common factor is represented by the fixed effects μ_0 , μ_1 , and μ_2 .

The observed score value y_{ijk} is linked to the value of the common factor at the time of measurement $\Lambda_i(t_{ijk})$ through a nonlinear link function h_k that is a beta cumulative distribution function depending on two test-specific parameters $\boldsymbol{\eta}_k = (\eta_{1k}, \eta_{2k})$. This leads to the following measurement model:

$$h_k(y_{ijk}; \boldsymbol{\eta}_k) = \Lambda_i(t_{ijk}) + \alpha_{ik} + \varepsilon_{ijk}$$

where the test-specific random intercept α_{ik} follows a Gaussian distribution with mean 0 and variance $\sigma_{\alpha k}^2$. It takes into account the residual individual variability between tests after adjustment on the latent common factor, that is, the fact that two subjects with the same latent cognition can score differently at the psychometric tests. At last, ε_{ijk} are independent Gaussian errors with mean 0 and variance $\sigma_{\varepsilon k}^2$.

**Impact of gender and education on cognitive decline with age:
a longitudinal approach**

Cécile Proust-Lima^{a,b}, Hélène Amieva^{b,c}, Hélène Jacqmin-Gadda^{a,b,*}, Luc Letenneur^{b,c}, Jean-Marc Orgogozo^{b,c} and Jean-François Dartigues^{b,c}

^a INSERM, EMI 0338 (Biostatistique), Bordeaux , F-33076, FRANCE

^b Université Victor Segalen Bordeaux 2, Bordeaux , F-33076, FRANCE

^c INSERM, U 593, Bordeaux , F-33076, FRANCE

* correspondence to : Hélène Jacqmin-Gadda, INSERM E0338, ISPED, Université de Bordeaux2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France. Tel : (33) 5 57 57 45 18 ;Fax : (33) 5 56 24 00 81 ; e-mail : helene.jacqmin-gadda@bordeaux.inserm.fr

Abstract :

When investigating effects of gender and education on degradation of cognitive performances with age, one cannot distinguish the association of these variables with cognitive decline from their association with the response of the tools used for measuring cognition.

Using a longitudinal model for describing the cognitive factor underlying four psychometric tests, we quantify both the effects of gender and education on a comprehensive indicator of cognition and their differential effect on the four psychometric tests. The approach is applied to a large sample of elderly subjects from the French cohort PAQUID whatever the type and the course of their cognitive aging process.

Adjusting for vascular risk factors, the results suggest that women have a slightly worst cognitive decline in oldest age than men and that women better perform in tests involving verbal component while men better perform in tests involving visuospatial skills. The results confirm a large difference in cognitive level according to education but the difference in scores tends to decrease with age for tests involving a speed component.

170 words

Keywords : aging, Benton visual retention test, cognition, Digit Symbol Substitution Test, Isaacs set test, Mini mental state examination, mixed model, neuropsychological evaluation.

Introduction

In spite of a vast literature on the subject, the effect of gender and education on the evolution of cognitive performances with age remains controversial. Longitudinal studies of randomly selected population-based samples of subjects is so far the best way to study these relationships by the collection of repeated measures of psychometric tests over time [21, 30]. This method allows to analyse both the baseline level and the evolution with age of the cognitive performances. However, even in these studies, multiple methodological problems interfere with the statistical analysis and disturb the associations. The use of psychometric tests as surrogate markers of performances in several cognitive functions implies two difficulties. First whatever the battery of psychometric tests used, each test only measures some domains of cognition even when considered as a global test like the Mini Mental State Examination (MMSE) [5], a composite score derived from a limited number of cognitive operations. Second, several studies have reported that the ability to perform the tests was strongly related to gender and education. Females tended to have better performances than males on tests involving an important verbal component such as verbal memory tests while males were generally better than females on tests requiring visuospatial information processing [3, 15, 23, 29]. High educated subjects tended to perform better on a wide range of tests including the MMSE [14, 17, 27] and community-based studies strengthened this result by showing that the effects of education on numerous tests scores remained significant after adjusting for confounding factors such as age [4, 7, 18]. However, when considering test scores it is important to keep in mind their artificial nature and arbitrary meaning. For instance, education may influence test-taking behaviour, lower educated subjects having more difficulty to deal with general as well as specific testing conditions. So, one could argue that the degradation of cognitive performances with age related to a low level of education is mediated by the differential ability in developing test-taking skills between high and low

educated subjects. This may imply that one cannot distinguish the association between education and the cognitive level from the association between education and tools used for measuring cognition. A way to deal with these problems is to study the change over time of a common latent cognitive factor underlying the various cognitive tests that are measured repeatedly. This common latent cognitive level, a potentially better measure of global cognitive degradation with age, can be obtained by a nonlinear latent variable model for longitudinal multivariate data in which psychometric tests are assumed to be nonlinear parameterized transformations of their common factor [22]. The common factor is a latent process representing the common cognitive process underlying the psychometric tests. It is related to age through a linear mixed model [16, 22]. By introducing covariates like gender or education both in association with the latent process and in association with each psychometric test, this model allows to compare the intensities of association of covariates with the different psychometric tests adjusting on the latent cognitive level, in addition to the estimation of the global association with cognitive level. Thus the effect of the covariate on the ability to perform the test can be separated from its effect on the latent cognitive level itself. The present analysis consisted in applying such statistical modelling to the Paquid study, a French population-based cohort of elderly subjects who have been followed for 13 years with repeated measures of cognition. The aim was to estimate the true impact of gender and educational level on cognitive degradation with age.

Methods

Population

PAQUID is a prospective cohort study initiated in 1988 in South Western France (Dordogne and Gironde) to explore functional and cerebral aging. In brief, 3777 subjects who were 65 years and older and were living at home at enrolment were included in the cohort if they agreed, after being randomly selected on electoral rolls. They were extensively assessed at baseline (V0) and then were followed up 6 times with a visit at 1 year (V1), 3 years (V3), 5 years (V5), 8 years (V8), 10 years (V10) and 13 years (V13) (except at 1 year in Dordogne). At each visit a neuropsychological evaluation and a screening for dementia were carried out at home. A more detailed description of Paquid is given by Letenneur et al. [19] .

Neuropsychological evaluation

In PAQUID, a battery of psychometric tests was used to quantitatively assess cognition. In this paper we focus the four tests which were systematically proposed at each visit:

(i) the Mini Mental State Examination (MMSE) [5] is a sum-score evaluating various dimensions of cognition (memory, calculation, orientation in space and time, language and word recognition). It is often used as an index of global cognitive performance and ranges from 0 to 30.

(ii) the Isaacs Set Test (IST15) [10] shortened at 15 seconds evaluates semantic verbal fluency abilities and speed of verbal production. Subjects have to give a list of words (with a maximum of 10) belonging to a specific semantic category in 15 seconds. Four semantic categories were successively used (cities, fruits, animals and colours). The score ranges from 0 to 40 [1].

(iii) the recognition form of the Benton Visual Retention Test (BVRT) [2] evaluates immediate visual memory. After the presentation for 10 seconds of a stimulus card displaying geometric figures, subjects are asked to choose the initial figure among four possibilities. Fifteen figures are successively presented. The score ranges from 0 to 15.

(iv) the Digit Symbol Substitution Test (DSST) [28] explores attention and psychomotor speed. Given a code table displaying the correspondence between pairs of digits (from 1 to 9) and symbols, the subjects have to fill in blank squares with the symbol which is paired to the digit displayed above the square. The subjects have to fill in as many squares as possible in 90 seconds. The score, representing the number of correct symbols, ranges from 0 to 76 in the PAQUID sample.

Statistical model

The response dependent variables are the scores at repeated measures of the four psychometric tests. The statistical model assumes that the correlation between the four tests is induced by a latent common cognitive process. The model is thus divided into two parts: (i) the longitudinal model for the latent cognitive process which describes the evolution of the latent cognitive factor and evaluates the effects of covariates on the latent cognitive evolution, and (ii) the test-specific models which link the measures of the psychometric tests with the latent cognitive process, taking into account test-specific associations with covariates.

The evolution of the common factor was modelled using a linear mixed model [16] which aims at evaluating changes over time of a repeated outcome (here the latent cognitive level) accounting for correlation between the repeated measures on each subject. The linear mixed model included random intercept, age and age squared in accordance with other longitudinal studies [1, 9] which showed quadratic cognitive evolutions with age. This quadratic trend

allows an acceleration of the cognitive decline among oldest subjects. The intercept, linear and quadratic coefficients for age were subject-specific random coefficients to take into account intra-subject correlation. We added a Brownian motion to account for individual deviations from this quadratic evolution and thus relaxed the parametric form of the model.

The test-specific models defined the flexible links between the psychometric tests and the latent cognitive factor. The hypothesis is that a test-specific nonlinear transformation of each test is a noisy measure of the latent cognitive factor. More specifically, the value of the nonlinear transformation of each test at age t equalled:

- the latent cognitive level at age t ,
- plus test-specific covariates effects, which account for the effect of the covariates on the ability to pass this test after adjustment on the cognitive level,
- plus a test-and-subject-specific random effect, which accounts for inter-individual variations of ability to pass this test after adjustment on the latent cognitive level and effects of covariates,
- plus an independent test-specific measurement error.

The nonlinear test-specific transformations which link each psychometric test with the latent common cognitive factor were parametric functions depending on parameters to be estimated simultaneously with the other parameters in the model. The chosen flexible transformations were Beta Cumulative Distribution Functions which offer a large variety of shapes (concave, convex or sigmoid) using only two estimated parameters per test. The complete methodology was previously detailed in Proust et al. [22] and is described in Appendix.

Explanatory variables

In addition to age and age squared, the model included educational level (graduated from primary school *versus* did not graduated) and gender which were included as covariates both in the linear mixed model for the latent cognitive level and in the models for each psychometric test. Thus, we were able to estimate association of covariates with the underlying global cognitive level, on one hand, and on each test score adjusting on the latent level, on the other hand. Moreover, summing these two estimates, we may also compute the global association of the covariate with each psychometric test without adjustment on the latent cognitive level. We examined both the impact of covariates on the mean latent cognitive level and on its evolution (by including interaction between covariates and age or age squared). We also adjusted the latent cognitive level model on vascular factors collected at the initial visit V0, that is history of stroke, high blood pressure (HBP) (self-reported usual systolic and diastolic blood pressures respectively higher than 140 and 90 mm Hg or use of drug for HBP), diabetes (self-reported history of diabetes, diabetes diet or diabetes drug), hypercholesterolemia (diet or drug against hypercholesterolemia) and smoking status at V0 (current or past smoker *versus* no-smoker), the frequency of some of these factor being markedly different between men and women. Only significant factors (simple effects or interactions) were kept in the final model except that non significant effects of gender or education were maintained in the latent process model when the corresponding differential associations in the test-specific models were significant.

Sample selection

Cognitive measurements at the initial visit (V0) were excluded from the analysis because of a first passing effect previously described [12]. Indeed a test-retest improvement was observed

between V0 and V1 possibly due to a stress induced by the first evaluation at V0 or a practice effect observed after the first evaluation. From the 3,777 subjects in PAQUID, we retained 3,043 subjects who were followed up after V0 and we studied the evolution of the common factor between V1 and V13. The DSST was not completed at V3, and as a sub-sample of PAQUID completed a nutritional questionnaire at V3 which could impact on the IST15 score for the fruit and animal categories, we excluded the measurements of IST15 at V3. From the sample of 3,043 subjects, we retained a sample of 2,252 subjects with at least one measure at each neuropsychological test between V1 and V13 to ensure that every subject brought information about each neuropsychological test. At last, 2,228 subjects had no missing values for the main covariates. The median number of measures was 4 for the MMSE (Inter Quartile Range (IQR)=2-5) and BVRT (IQR=2-5) and respectively 3 for the IST15 (IQR=1-4) and 2 for the DSST (IQR=1-4).

Results

Description of the sample

Characteristics of the sample are described in table 1: 56.7% of the 2228 subjects included in the sample were women and 73% had at least graduated from primary school. The mean age at V1 was 75.20 (standard deviation (sd)=6.25). Among the 2228 subjects, 7 (0.4%) subjects were demented at the beginning of the follow-up (V1) and 364 subjects (16.3%) had a positive diagnosis of dementia during the follow-up. At inclusion (V0), 82.9% had a high blood pressure, 7.8% of the subjects were suffering from diabetes, 17.5% from hypercholesterolemia and 38.1% were either current smokers or past-smokers. At V1, 7.1% had history of stroke. At visit V1, including only subjects living in Gironde, 1,604 subjects completed the MMSE with a median score of 28 (IQR=26-29), 1,603 subjects completed the

IST15 (median=28, IQR=24-33), 1,600 subjects completed the BVRT (median=11, IQR=10-13) and 1,596 completed the DSST (median=28, IQR=20-37).

Longitudinal Multivariate model

Impact of gender on cognition:

Estimates of the multivariate model are displayed in table 2. Firstly, while gender was not globally associated with the latent common cognition when not taking into account the vascular factors (neither associated with the mean latent cognitive level ($\beta=0.0005$, $p=0.40$) nor with the cognitive evolution ($\beta=0.0019$, $p=0.18$)), it was associated with the cognitive evolution when adjusting on the vascular factors and especially the smoking status. Indeed, it was still not associated with the basic latent cognitive level ($\beta=-0.0056$, $p=0.23$) but was associated with the cognitive evolution ($\beta=0.0061$, $p=0.002$), women having a slightly worst cognitive decline with age than men in oldest age as seen in figure 1(a). In the following, the commented results come from the model adjusted for vascular factors.

After adjusting on the latent cognitive level, gender effect was significantly different over the four tests both on the mean level of the tests ($p<0.0001$ for contrasts on gender) and on the evolution of the tests with age ($p<0.0001$ for contrasts on gender \times time squared). Parameters for the contrasts on gender measure the association of gender with each psychometric test conditional on the latent cognitive level. Careful examination of these parameters shows that men have lower MMSE ($\beta=-0.139$, $p=0.0002$) and IST ($\beta=-0.0092$, $p=0.033$) scores at 65 years on average than women with the same latent cognitive level. On the contrary, at 65 years, the mean BVRT score is higher for men ($\beta=0.0225$, $p<0.0001$) compared to women with the same latent cognitive level. Contrasts on the interaction gender by time squared show that the sharper decline of women found on latent cognition was more marked on the IST ($\beta=-$

0.0029, $p=0.024$) and less marked on the MMS ($\beta=0.0058$, $p<0.0001$). No difference was found on the other tests.

Using estimates shown in table 2, we can also compute the global measure of association of gender with each psychometric test without adjusting on the latent cognitive level. These unadjusted measures are computed by summing the parameters for gender in the model for the common factor and the test-specific parameter (contrast) for gender. Thus, without adjustment on the global cognitive level, women succeed better than men to the IST at 65 years old ($\beta=-0.0056-0.0092=-0.0148$, $p=0.043$), their cognitive decline with age being not significantly different from the men's one ($\beta=0.0032$, $p=0.150$). On the contrary, men succeed better than women to the BVRT at 65 years old ($\beta=0.0169$, $p<0.0001$) but have a similar decline with age ($\beta=0.0039$, $p=0.096$). For the DSST, despite no difference of score between men and women at 65 years old ($\beta=-0.0050$, $p=0.293$), women tended to have a deeper decline with age than men ($\beta=0.0055$, $p=0.0145$). At last, although women succeed slightly better at 65 years old to the MMSE than men ($\beta=-0.0195$, $p=0.003$), they had a deeper decline in oldest ages ($\beta=0.0119$, $p<0.0001$). The figure 1(b) represents these predicted mean evolutions with age for the four tests.

Impact of educational level on cognition:

As the results for EL were not modified when adjusting for vascular factors, we only detail them in the adjusted model. Subjects with higher EL had on average a better cognitive level at 65 years than subjects with lower EL ($\beta=0.112$, $p<0.0001$) but had a sharper decline in older ages ($\beta=-0.0038$, $p=0.033$) as shown in figure 2(a).

The association between cognition and EL was significantly different over the four psychometric tests both on the mean initial level of each test ($p<0.0001$ for contrasts on EL)

and on the mean evolution with age ($p=0.0003$ for contrasts on $EL \times \text{time squared}$) which means that education did not only have an impact on the latent cognition but also on the tools used to measure cognition. Contrasts on EL shows that the effect of education at 65 years old was amplified for the DSST ($\beta=0.0416$, $p<0.0001$) and reduced for the MMS ($\beta=-0.0144$, $p=0.0019$) and the BVRT ($\beta=-0.0175$, $p=0.0004$): the difference between DSST scores for subjects with high EL compared to subjects with low EL is higher than expected given the impact of EL on the latent cognition. On the contrary, this difference is lower than expected for MMSE and BVRT. Contrasts on the interaction EL by time squared show that the difference in the rate of decline between high EL and low EL is more pronounced for the DSST ($\beta=-0.0039$, $p=0.005$) (and IST even if not strictly significant, $\beta=-0.0025$, $p=0.077$) than for the latent cognitive level and less pronounced for MMSE ($\beta=0.0037$, $p=0.005$) and BVRT ($\beta=0.027$, $p=0.05$).

Using parameters shown in table 2, we can also compute the global measures of association between EL and each test, that is without adjusting on the latent cognitive level. Thus, we found that the mean score of subjects with higher EL was significantly better for the four tests at 65 years old but the impact of EL was stronger for the DSST ($\beta=0.154$, $p<0.0001$) than for the IST ($\beta=0.102$, $p<0.0001$), the BVRT ($\beta=0.095$, $p<0.0001$) and the MMSE ($\beta=0.098$, $p<0.0001$). For the IST and the DSST, the mean decline was significantly faster for subjects with high EL compared to subjects with low EL ($\beta=-0.006$, $p=0.008$ and $\beta=-0.008$, $p=0.0005$ respectively). There was no significant difference in the decline between the educational levels for the MMSE ($\beta=-0.0002$, $p=0.397$) and the BVRT ($\beta=-0.001$, $p=0.354$). These results are illustrated on figure 2(b).

Discussion

By applying a nonlinear latent variable model [22] to the psychometric data of the PAQUID cohort, we have been able to estimate the evolution with age of a comprehensive indicator of cognition and to distinguish the impact of education and gender on cognition from their impact on four psychometric tests used for measuring cognition. Thus, after adjusting for vascular factors, gender was associated with the latent cognitive factor, women having the same baseline cognitive level but a deeper decline in older ages than men. This association suffered from confusion since without accounting for smoking status, the association was non-significant. Indeed, 82% of the smokers are men and smokers tend to have a deeper cognitive decline with age. Gender had also a differential effect on the four tests. Especially, we found that men performed better than women to the BVRT which is consistent with previous studies reporting that men use to be better performers than women in tests involving visuospatial skills [23, 29]. In contrast, we found that women performed better than men to the IST15 which is also in agreement with previous findings reporting that women tend to better perform than men on tasks with verbal modality [23, 29].

In the same vein, we distinguished the effect of education on the latent cognitive evolution from its specific effect on the psychometric tests. We found that subjects who graduated from primary school had in mean a better cognitive level than subjects who did not graduate which corroborates previous results about the influence of education on neuropsychological performances [4, 7, 18]. Nonetheless, we also showed that this effect differed significantly according to the psychometric test considered. Thus, the difference between a subject who graduated from primary school and a subject who did not graduate decreased with age when considering the IST15 and DSST while it remained the same whatever the age when considering the MMSE or the BVRT. In contrast with the MMSE or the BVRT, the IST15 and DSST include a speed component and it was found previously that the decrease in processing speed explained most of age-related differences in cognition [24].

These findings should be interpreted by keeping in mind the fact that our sample is constituted of a mixture of demented subjects, subjects in a pre-diagnosis phase of dementia, and subjects with a normal cognitive aging since we included in this study all the subjects with at least one neuropsychological evaluation during the follow-up. Thus, estimated cognitive declines and differential declines according to gender and the level of education represent declines in a population of elderly people whatever the type and the course of their cognitive aging process. Brain aging in a sample of subjects from the general population cumulates the proper effect of age on the central nervous system, particularly on neurones and synapses, the effect of Alzheimer lesions or other neurodegenerative lesions like Lewy bodies, and the effect of cerebro-vascular lesions. All these morbid processes increase with age. From a practical clinical point of view and a public health point of view, it would be relevant to separate these processes to evaluate for instance the impact of gender and education separately on normal cognitive aging and on cognitive decline toward dementia. However, it appears also relevant to study the evolution of the whole sample because there is a great overlap in brain lesions between demented and non demented subjects [13], there is a long progressing pre-dementia phase and the stage of dementia remains more or less arbitrary defined. Moreover, it is very difficult to study separately normal cognitive aging and especially to distinguish determinants of normal aging from those of pathological aging because we cannot be sure that a sample of normal elderly does not include subjects in a pre-diagnosis phase of dementia [26]. As a consequence, the association we observed between education and cognitive decline both on the latent cognition and on the psychometric tests could be the result of the effect of education on the risk of dementia [20], on the shape of the pre-diagnosis cognitive decline [1, 11, 25] and on the normal aging process [1, 8, 26]. On the other hand, the deeper cognitive decline of women in older ages could be explained by a higher risk of dementia for the women compared to men among the oldest subjects [6].

A limit of the study is that we considered missing data as ignorable and we did not account for dropout in our model whereas dropout is very frequent in population-based aging studies. Moreover, it is known that missing data and dropout occur more often among impaired subjects [12] and can be informative. Not accounting for informative missing data can bias the estimations of cognitive declines even if, by using jointly various psychometric tests and adjusting on vascular factors which may be associated with dropout, their impact could be diminished.

Another important point is that the latent common factor in this model is defined according to the pool of psychometric tests used in the analysis. Computing the model with other tests involving different cognitive components could have an impact on the common factor evolution. This is why, for this analysis, we selected tests which both are frequently used and explore different domains of cognition.

In spite of these limitations, this methodology allowed us firstly to model jointly four widely used psychometric tests, including non Gaussian scores or scores with floor and ceiling effects like the MMSE, and secondly to separate and quantify the differential effects of gender and educational level on the psychometric tests and their global effect on the latent cognition represented by the common factor of the four tests.

Acknowledgements

This study was funded by NOVARTIS pharma, SCOR insurance, Agrica, Conseil régional d'Aquitaine, Conseil Général de la Gironde and Conseil Général de la Dordogne.

Disclosure statement

There is no actual or potential conflicts of interest concerning this study.

Appendix : model specification

We consider K neuropsychological tests. For each test k , $k=1, \dots, K$, each subject i , $i=1, \dots, N$ and each occasion j , $j=1, \dots, n_{ik}$, the measure of the neuropsychological test y_{ijk} is collected at time t_{ijk} , t_{ijk} being different for each test and each subject. The latent process which represents the common factor of the K neuropsychological tests is modelled using the following linear mixed model including a quadratic function of time and a Brownian motion $(w_i(t))_{t \geq 0}$ with variance term $\mathcal{S}_w^2 t$:

$$\Lambda_i(t) = (X_i^T \mathbf{b}_0 + \mathbf{m}_0 + u_{0i}) + (X_i^T \mathbf{b}_1 + \mathbf{m}_1 + u_{1i}) \times t + (X_i^T \mathbf{b}_2 + \mathbf{m}_2 + u_{2i}) \times t^2 + w_i(t)$$

The vector of random effects $\mathbf{u}_i = (u_{0i}, u_{1i}, u_{2i})^T$ follows a multivariate normal distribution with mean vector $\mathbf{0}$ and variance covariance matrix D . The mean evolution of the common factor is represented by the fixed effects \mathbf{m}_0 , \mathbf{m}_1 and \mathbf{m}_2 . The vector of covariates X_i can either be associated with the mean latent common factor level through the vector \mathbf{b}_0 or the mean common factor evolution through the vectors \mathbf{b}_1 and \mathbf{b}_2 .

The observed score value y_{ijk} is linked to the value of the common factor at the time of measurement $L_i(t_{ijk})$ through a nonlinear link function h_k which is a Beta Cumulative Distribution Function depending on two test-specific parameters $\mathbf{h}_k = (\mathbf{h}_{1k}, \mathbf{h}_{2k})$. This leads to the following measurement model :

$$h_k(y_{ijk}; \mathbf{h}_k) = \Lambda_i(t_{ijk}) + X_i(t_{ijk})^T \mathbf{g}_k + \mathbf{a}_{ik} + \mathbf{e}_{ijk}$$

where $X_i(t_{ijk})$ is a vector of covariates including the quadratic function of time and associated with the neuropsychological tests through the vector of parameters \mathbf{g}_k . The vector \mathbf{g}_k represents the differential association (or contrast) of $X_i(t_{ijk})$ with the neuropsychological test k after adjusting on the common factor value $L_i(t_{ijk})$ ($\sum_{k=1}^K \mathbf{g}_{mk} = 0, \forall m$). The test-specific random intercept \mathbf{a}_{ik} follows a Gaussian distribution with mean 0 and variance \mathcal{S}_{ak}^2 . It takes into account the residual individual variability between tests after adjustment on the latent

common factor and the covariates. At last, \mathbf{e}_{ijk} are independent Gaussian errors with mean 0 and variance $\mathbf{S}_{\mathbf{e}_k}^2$.

References

- [1] Amieva H, Jacqmin-Gadda H, Orgogozo JM, Le Carret N, Helmer C, Letenneur L, Barberger-Gateau P, Fabrigoule C, Dartigues JF. The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain* 2005;128(Pt 5):1093-101
- [2] Benton A. Manuel pour l'application du Test de Rétention Visuelle. Applications cliniques et expérimentales. Paris:Centre de Psychologie appliquée; 1965
- [3] Collaer ML, Hines M. Human behavioral sex differences: a role for gonadal hormones during early development? *Psychol Bull* 1995;118(1):55-107
- [4] 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. *Exp Aging Res* 1997;23(3):201-35
- [5] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12(3):189-98
- [6] Fratiglioni L, Launer LJ, Andersen K, Breteler MM, Copeland JR, Dartigues JF, Lobo A, Martinez-Lage J, Soininen H, Hofman A. Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology* 2000;54(11 Suppl 5):S10-5
- [7] Ganguli M, Ratcliff G, Huff FJ, Belle S, Kancel MJ, Fischer L, Seaberg EC, Kuller LH. Effects of age, gender, and education on cognitive tests in a rural elderly community sample: norms from the Monongahela Valley Independent Elders Survey. *Neuroepidemiology* 1991;10(1):42-52
- [8] Ganguli M, Seaberg EC, Ratcliff GG, Belle SH, DeKosky ST. Cognitive stability over 2 years in a rural elderly population: the MoVIES project. *Neuroepidemiology* 1996;15(1):42-50

- [9] Hall CB, Lipton RB, Sliwinski M, Stewart WF. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Stat Med* 2000;19(11-12):1555-66
- [10] Isaacs B, Kennie AT. The Set test as an aid to the detection of dementia in old people. *Br J Psychiatry* 1973;123(575):467-70
- [11] Jacqmin-Gadda H, Commenges D, Dartigues JF. Random change point model for joint modeling of cognitive decline and dementia. *Biometrics* 2006;62(1):254-60
- [12] Jacqmin-Gadda H, Fabrigoule C, Commenges D, Dartigues JF. A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *Am J Epidemiol* 1997;145(6):498-506
- [13] Jicha GA, Parisi JE, Dickson DW, Johnson K, Cha R, Ivnik RJ, Tangalos EG, Boeve BF, Knopman DS, Braak H, Petersen RC. Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Arch Neurol* 2006;63(5):674-81
- [14] Jorm AF, Scott R, Henderson AS, Kay DW. Educational level differences on the Mini-Mental State: the role of test bias. *Psychol Med* 1988;18(3):727-31
- [15] Kramer J, Delis D, Daniel M. Sex differences in verbal learning. *J Clin Psychol* 1988;44(9):907-915
- [16] Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38(4):963-74
- [17] Launer LJ, Dinkgreve MA, Jonker C, Hooijer C, Lindeboom J. Are age and education independent correlates of the Mini-Mental State Exam performance of community-dwelling elderly? *J Gerontol* 1993;48(6):P271-7
- [18] 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. *Dev Neuropsychol* 2003;23(3):317-37

- [19] Letenneur L, Commenges D, Dartigues JF,Barberger-Gateau P. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *Int J Epidemiol* 1994;23(6):1256-61
- [20] Letenneur L, Gilleron V, Commenges D, Helmer C, Orgogozo JM,Dartigues JF. Are sex and educational level independent predictors of dementia and Alzheimer's disease? Incidence data from the PAQUID project. *J Neurol Neurosurg Psychiatry* 1999;66(2):177-83
- [21] Morris MC, Evans DA, Hebert LE,Bienias JL. Methodological issues in the study of cognitive decline. *Am J Epidemiol* 1999;149(9):789-93
- [22] Proust C, Jacqmin-Gadda H, Taylor JM, Ganiayre J,Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* 2006;Online publication date: 26-Apr-2006 doi: 10.1111/j.1541-0420.2006.00573.x(
- [23] Reite M, Cullum CM, Stocker J, Teale P,Kozora E. Neuropsychological test performance and MEG-based brain lateralization: sex differences. *Brain Res Bull* 1993;32(3):325-8
- [24] Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychol Rev* 1996;103(3):403-28
- [25] Scarmeas N, Albert SM, Manly JJ,Stern Y. Education and rates of cognitive decline in incident Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 2006;77(3):308-16
- [26] Sliwinski M, Lipton RB, Buschke H,Stewart W. The effects of preclinical dementia on estimates of normal cognitive functioning in aging. *J Gerontol B Psychol Sci Soc Sci* 1996;51(4):P217-25
- [27] Tombaugh TN,McIntyre NJ. The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc* 1992;40(9):922-35
- [28] Wechsler D.WAIS-R manual. New York:Psychological Corporation; 1981

[29] 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. *J Am Geriatr Soc* 1993;41(6):639-47

[30] Yesavage JA, Brooks JO, 3rd. On the importance of longitudinal research in Alzheimer's disease. *J Am Geriatr Soc* 1991;39(9):942-4

Figure 1 : (A) Predicted mean evolutions of the latent cognitive level with age according to gender for a non smoking subject with a low educational level and any antecedent of stroke or diabetes (with the 95% confidence bands (CB)) (B) predicted mean evolutions of the MMSE, IST15, BVRT and DSST according to gender for a non smoking subject with a low educational level and any antecedent of stroke or diabetes.

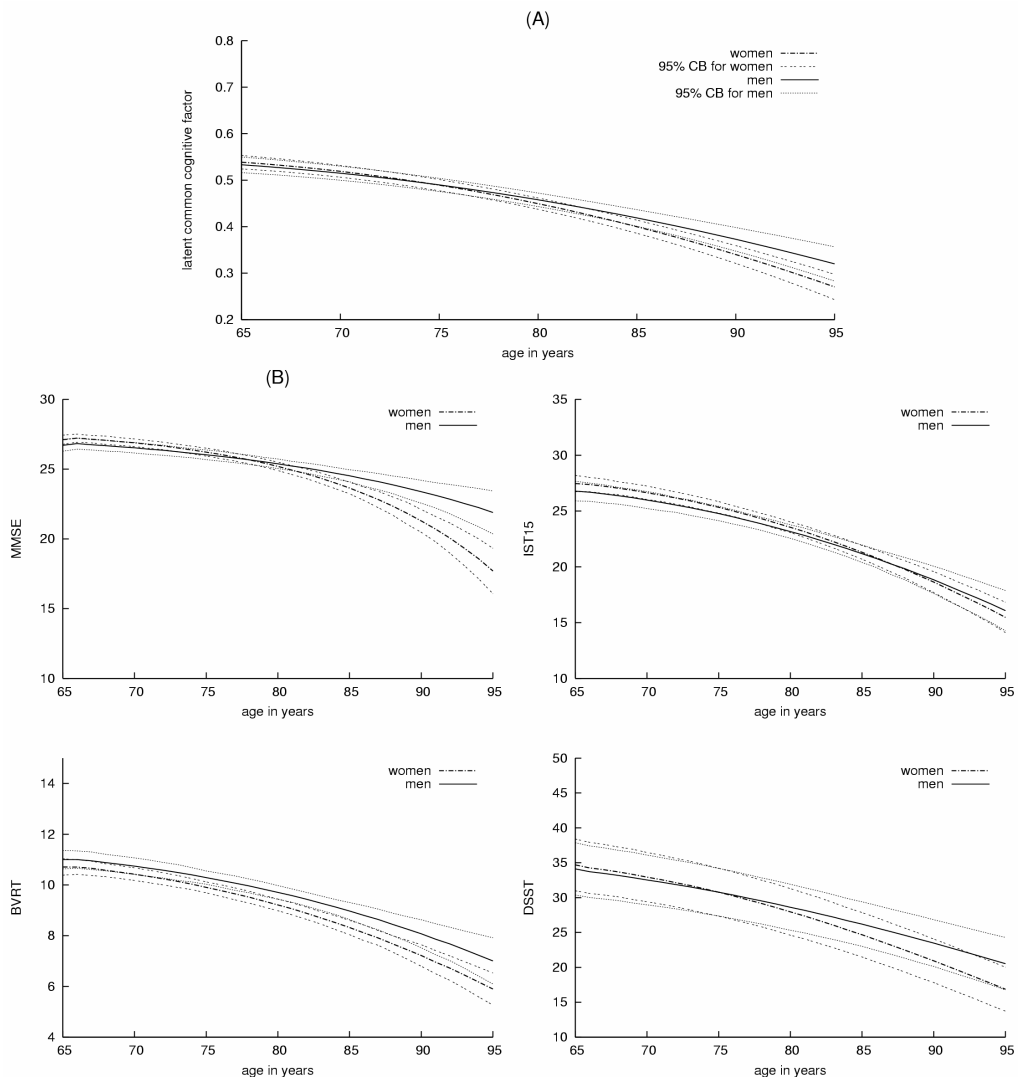


Figure 2 : (A) Predicted mean evolutions of the latent cognitive level with age according to the educational level (EL) for a non smoking women without any antecedent of stroke or diabetes (with the 95% confidence bands (CB)) (B) predicted mean evolutions of the MMSE, IST15, BVRT and DSST according to the educational level for a non smoking women without no antecedent of stroke or diabetes.

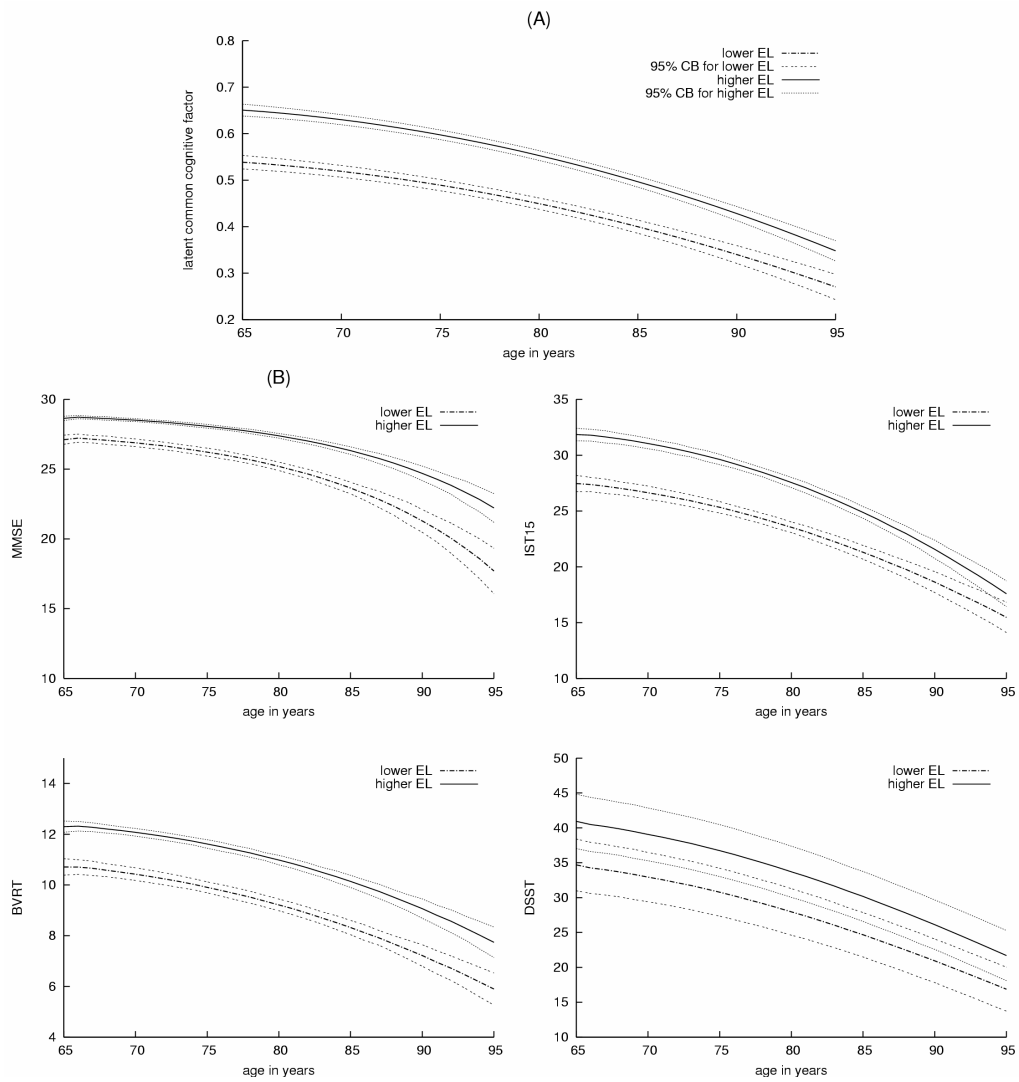


Table 1: demographic and health characteristics in the sample (N=2228)

	N	%
Gender		
Male	964	43.3
Female	1264	56.7
Educational level		
No diploma	602	27
Primary school with diploma or +	1626	73
Age at V1		
=<70 year old	1086	48.7
>70 and =<75 year old	365	16.4
>75 and =<80 year old	407	18.3
>80 year old	370	16.6
Prevalent dementia at V1	7	0.4
Incident dementia between V3 and V13	364	16.3

Table 2: Estimates and 95% confidence interval (95%CI) of the parameters in the nonlinear latent variable model using MMSE, IST15, BVRT and DSST without adjustment on vascular factors (VF) (on the left, Log-likelihood=-69993.86 and 42 parameters) and with adjustment on vascular factors (on the right, Log Likelihood=-69969.18 and 47 parameters).

Variable	Without VF adjustment		With VF adjustment	
	Estimate ^a	95% CI	Estimate ^a	95% CI
<u>Common factor evolution :</u>				
Intercept (at 65 years old)	0.538 ^{***}	0.523,0.552	0.539 ^{***}	0.524,0.553
Time ^b	-0.0305 ^{***}	-0.0425,-0.0185	-0.0294 ^{***}	-0.0414,-0.0174
Time squared ^b	-0.0209 ^{***}	-0.0261,-0.0158	-0.0200 ^{***}	-0.0251,-0.0149
Gender ^c	0.0005	-0.0081,0.0091	-0.0056	-0.0161,0.0050
Gender ^c × time squared	0.0019	-0.0011,0.0050	0.0061 [*]	0.0024,0.0099
Educational level (EL) ^d	0.114 ^{***}	0.103,0.124	0.112 ^{***}	0.102,0.123
EL ^d × time squared	-0.0038 [*]	-0.0071,-0.0004	-0.0038 [*]	-0.0072,-0.0005
<u>Contrasts on gender ^e :</u>				
	$p<0.0001$ ^f		$p<0.0001$ ^f	
MMSE	-0.0139 ^{**}	-0.0208,-0.0070	-0.0139 ^{**}	-0.0209,-0.0070
IST15	-0.0092 [*]	-0.0172,-0.0011	-0.0092 [*]	-0.0172,-0.0011
BVRT	0.0225 ^{***}	0.0152,0.0299	0.0225 ^{***}	0.0152,0.0299
DSST	0.0006	-0.0077,0.0079	0.0005	-0.0068,0.0079
<u>Contrasts on gender×time squared ^e :</u>				
	$p<0.0001$ ^f		$p<0.0001$ ^f	
MMSE	0.0058 ^{***}	0.0036,0.0079	0.0058 ^{***}	0.0036,0.0080
IST15	-0.0029 [*]	-0.0054,-0.0005	-0.0029 [*]	-0.0054,-0.0005
BVRT	-0.0022	-0.0046,0.0002	-0.0022	-0.0046,0.0002
DSST	-0.0006	-0.0028,0.0016	-0.0006	-0.0028,0.0016
<u>Contrasts on EL ^e :</u>				
	$p<0.0001$ ^f		$p<0.0001$ ^f	
MMSE	-0.0145 [*]	-0.0232,-0.0059	-0.0144 [*]	-0.0231,-0.0058
IST15	-0.00966	-0.0197,0.0004	-0.0097	-0.0198,0.0004
BVRT	-0.0175 ^{**}	-0.0268,-0.0083	-0.0175 ^{**}	-0.0267,-0.0082
DSST	0.0417 ^{***}	0.0316,0.0519	0.0416 ^{***}	0.0315,0.0517
<u>Contrasts on EL×time squared ^e :</u>				
	$p=0.0003$ ^f		$p=0.0003$ ^f	

MMSE	0.0037 [*]	0.0013,0.0060	0.0037 [*]	0.0013,0.0060
IST15	-0.0025	-0.0052,0.0002	-0.0025	-0.0052,0.0002
BVRT	0.0027 [*]	0,0.0055	0.0027 [*]	0,0.0054
DSST	-0.0039 [*]	-0.0065,-0.0013	-0.0039 [*]	-0.0064,-0.0013

^a p-value of the Wald test indicated by * if $p < 0.05$, ** if $p < 0.001$ or *** if $p < 0.0001$

^b time variable is age minus 65 by 10 years ($\text{time} = (\text{age} - 65) / 10$)

^c men vs. women

^d Graduated from primary school vs. not graduated

^e For a given covariate, the sum over the four contrasts is null

^f p-value of the likelihood ratio test (χ^2 with 3 degrees of freedom)