Study of dementia and cognitive decline accounting for selection by death

Anais Rouanet

université de BORDEAUX

Numéro national de thèse :

Thèse présentée pour obtenir le grade de

DOCTEURE
DE L'UNIVERSITÉ DE BORDEAUX

École Doctorale Sociétés, Politique, Santé Publique

Spécialité Santé Publique, option Biostatistique

Par Anaïs ROUANET

Prise en compte de la sélection par le décès dans l'étude de la démence et du déclin cognitif

Study of dementia and cognitive decline accounting for selection by death

Sous la direction d'Hélène JACQMIN-GADDA

Soutenue le 14 Décembre 2016

Membres du jury

COMMENGES Daniel	Directeur de Recherche Emerite, INSERM, Bordeaux	Président
RIZOPOULOS Dimitris	Associate Professor, Erasmus UMC, Rotterdam	Rapporteur
LECLERCQ SAMSON Adeline	Professeure, Université Grenoble Alpes, Grenoble	Rapporteure
BERR Claudine	Directrice de Recherche, INSERM, Montpellier	Examinatrice
DUFOUIL Carole	Directrice de Recherche, INSERM, Bordeaux	Examinatrice
JACQMIN-GADDA Hélène	Directrice de Recherche, INSERM, Bordeaux	Directrice de thèse

Contents

Remerciements

En premier lieu, je remercie tout naturellement ma directrice de thèse, Hélène Jacqmin-Gadda. J'ai énormément appris au cours de ces trois années de thèse avec toi, et je tiens à te remercier tout particulièrement pour ta pédagogie, ta bienveillance et ta disponibilité. Merci pour ces longues réunions toujours plus stimulantes, pour ton regard encourageant sur mon travail et pour ce bagage statistique que tu m'as aidé à consolider. Ton esprit scientifique est un modèle pour moi. J'espère sincèrement que notre collaboration continuera.

Monsieur Daniel Commenges, je suis honorée que vous présidiez mon jury de thèse. Votre rigueur et votre connaissance théorique approfondie des modèles statistiques sont un exemple pour chacun de nous, doctorants. Merci d'avoir bâti les fondations de cette équipe qui nous offre ce cadre d'étude et de travail si privilégié.

Je remercie également les rapporteurs et examinatrices de mon jury : Pr Dimitris Rizopoulos, I am very grateful that you accepted to review my Ph.D work, and honoured that you take part in my jury. Madame Adeline Leclercq Samson, je vous remercie chaleureusement d'avoir accepté de juger mon travail, et d'en être rapporteure. Madame Claudine Berr et Madame Carole Dufouil, je vous remercie d'avoir accepté de prendre part à ce jury, et vous suis reconnaissante pour votre contribution à la cohorte Trois Cités, qui nous offre des données de qualité pour l'étude de la démence.

Je souhaite à présent exprimer ma gratitude envers les personnes qui ont marqué mon cursus étudiant : tout d'abord mon professeur de maths Sup, M. Roussel, qui a su passionner ses étudiants à coup de gnous, de 12, et de grande fête. Sophie Desset, vous m'avez prise sous votre aile pour mon stage de M1, alors que je n'avais presqu'aucune connaissance en biologie, et que vous étiez en transition professionnelle. Merci pour cette riche expérience qui m'a donné un aperçu attrayant de la recherche. Céline Helbert, vous m'avez aiguillée en 3^{e} année à Centrale Lyon et je vous remercie de votre soutien tout au long de mon stage de M2. Merci à David Rousseau et Carole Frindel, avec qui j'ai découvert

le monde de l'imagerie cérébrale lors de mon stage de M2. Je garde un souvenir ému de mon premier cours de R, de mon premier séminaire et de nos rencontres enrichissantes avec les médecins. Merci pour cette belle opportunité.

Je remercie ensuite toute l'(ancienne) équipe Biostat, au sein de laquelle il est si agréable de travailler : Cécile, pour ta gentillesse, ton soutien (personnel et professionnel) et pour les nouveaux jeux que je connais désormais, Pierre, pour vos visites régulières et égayantes, Rodolphe (aka Arthur) pour ton énergie débordante, ainsi que Karen, Virginie, Marta et Alioum. Je tiens à remercier Jean-François Dartigues, qui partage volontiers son expertise dans le domaine de la démence. Un merci également à Luc, pour nos échanges sur mes travaux mais aussi sur les copies des élèves, et plus généralement à toute l'équipe pédagogique. Sandrine, merci pour ces virées-séminaires et pour ta bonne humeur ! Majid, Fleur et Cécilia, c'est toujours avec grand plaisir que je discutais avec vous au détour d'un couloir. Enfin, merci à Diena et à mes étudiants qui m'ont beaucoup apporté et m'ont conforté dans mon goût pour l'enseignement.

Un Special Big Up pour toute la clique : Matmat, ou Dr Wanneveich, tu m'as accueillie dès mon premier jour à l'Isped, avec ton big smile (et ta baguette de pain). Surtout, garde ce grain de folie décalée que j'adore. Milie, depuis un an tu me supportes en face à face H12. Rien que pour ça, je te décerne la médaille de la co-bureau la plus tenace, et la plus attentionnée. Trompette de 7 (Loïc et al., 1991), mon co-thésard, on aurait pu terminer en même temps et écrire une thèse pour deux. Mais comme d'hab, t'es jamais à l'heure. Papi, tu nous auras tous contaminés avec ton zzzzit. Pourtant, ça ne vient ni de Papi Touya, ni de Michel-à-peu-près ! Pépette, après nos six mois intenses lors de ton stage, tu as préféré t'envoler vers d'autres horizons (aka 3^e étage). Depuis, pour me consoler, j'écoute le connemara en boucle avant de m'endormir. Chloé, tes cours de danse du 8 resteront gravés dans ma mémoire. Vive la Kizumba ! Robin, malgré ton accent parfois un peu douteux, je suis fière de te compter parmi les Bordeaux lovers. Merci pour la session de 'back-surf', et surtout, pour tes TD de SAS. Viviane, merci pour tous tes bons gâteaux ! Agnieszka, bravo pour ton courage et ta ténacité pendant ces 3 ans, je te souhaite beaucoup de bonheur avec ta petite famille.

Un merci aux "anciens" : DDD, Asie ✓, USA ✓, where next ?, Loulou surtout garde ta joie de vivre, Paulo, merci pour tes remises en question scientifiques permanentes et tes conseils, Mel (Prague) pour ton enthousiasme à toute épreuve, Hind pour la relecture !,

Julie pour l'agencement du bureau 45 !, sans oublier Lingling (Guomin, Léo) et Jérémie.

Un merci à tous les autres doc' et ingé qui ont partagé cette expérience avec moi : Bachirou le chef pâtissier, Aaron aaron petit patapon, Laura et ses dizaines de colocs, Alex-grenouillette, Astou et Mehdi, Myriam et sa cafetière, TomTom, Marcel et son orchestre, Souf' et son big smile, Juanito et Caro, Alex et Frailtypack, Bernard, Rémy, Mélanie (Plazy), Bruno et ses chats, Hadrien, Mai, Arlette et le p'tit dernier Corentin. Yun-Hee, hope to see you again, in Canada, France (or Korea !). Cheers to the ones I met in conferences !

Une pensée nostalgique pour mes différents colocs : Béguéééé et Gabi du Château Lespiault, Birte et Johan de Gaston l'espion, et Bruno et Alice, les petits derniers ! Merci pour ces moments conviviaux et ces retrouvailles à chaque fois chaleureuses. Merci à ceux qui ont suivi de près ou de loin cette aventure : Cheers to Will and Peter my very first roomates, see you in Cambridge for a Never-have-I-ever session ! A ceux de l'ECM : Gabi mi amor de Paquita, Steph, Gianni, Rebecca et Clem, Teddy, Thib et particulièrement Martouze, prêt à tout moment à me replonger dans l'ambiance marseillaise. A Florent, François, TBT, Micheline de la prépa, Rudy, Rémi du lycée, Sarah, Ariane du collège, Chachou de l'enfance. A vous tous, je ne vous vois pas souvent, mais même loin des yeux, vous restez près du coeur.

Merci à ma famille : Agnès et Bébert, toujours présents dans les moments importants, Guillaume et Clément, mes deux modèles, Caroline, Lisa et mes petites cousines, tonton Jean-Paul, mi primo querido André, mon p'ti frère Coco, ma princesse Yun Qiu, ma soeur Audrey, ma nièce Ella, et mamie Claudine. Merci à Marc et Christine pour votre accueil chaleureux et anisé à Marseille, à ma Marraine chérie, à mon bébé Lolo, Agnès (et Bob !), Jorgelina et Kiki, la famille de coeur. Un grand merci à Gilles, Annie et Lucas, pour votre accueil empreint de gentillesse, d'attentions et de générosité (et merci pour la relecture !). Merci à petit Georges. Même si tu aimes parfois jouer le capitaine solitaire, cette fois-ci, on prend tous le bateau avec toi.

Un merci tout spécial à ma Lulu (et à Fab !), qui m'accompagne depuis 25 ans. Je suis tellement fière de toi, de ce que tu accomplis, et surtout de ton amitié. Merci à mon papa, mon papounet chéri qui est là et sera toujours là. Merci à ma maman, mon pilier. Si j'en suis là, c'est en grande partie grâce à toi. Merci pour ton amour et ton soutien inconditionnels, ce sont mes sources d'énergie. Je t'admire pour ta force, ta droiture, ton

espoir en la vie et l'humain, que j'espère avoir appris de toi. Je t'aime. Enfin, merci à Boris, mon coeur. Merci pour ton amour, ton soutien, ta confiance et ton humour, toujours. La vie est si douce avec toi. J'ai hâte qu'on commence à écrire ensemble cette nouvelle page d'aventures.

Je dédie ce manuscrit à mes grand-parents, Odette, Henri et Claudine.

Scientific production

Articles

Thesis publications

► A. Rouanet, P. Joly, JF. Dartigues, C. Proust-Lima, H. Jacqmin-Gadda, Joint latent class model for longitudinal data and interval-censored semi-competing events : Application to dementia, *Biometrics*, 2016.
DOI: 10.1111/biom.12530

► A. Rouanet, JF. Dartigues, C. Helmer, H. Jacqmin-Gadda, Interpretation of mixed models and marginal models with cohort attrition due to death and drop-out, submitted.

Communications

Oral communications at conferences

► A. Rouanet, H. Jacqmin-Gadda, Modèle conjoint pour données longitudinales et données multi-états avec censure par intervalle : application à l'étude de la démence, *Journées de la Statistique*, Rennes, France, 2014.

► A. Rouanet, H. Jacqmin-Gadda, Joint latent class model for longitudinal data and competing interval-censored events : Application to the study of Alzheimer's disease, *International Society for Clinical Biostatistics Conference*, Vienna, Austria, 2014.

► A. Rouanet, C. Proust-Lima, P. Joly, JF. Dartigues, H. Jacqmin-Gadda, Joint latent class model for longitudinal data and competing interval-censored events : Application to the study of Alzheimer's Disease, *IBS Channel Conference*, Nijmejen, The Netherlands, 2015.

► A. Rouanet, C. Proust-Lima, H. Jacqmin-Gadda, Assessment of the goodness-of-fit for joint latent class models for longitudinal data and multiple events, *Groupe de Recherche Statistiques et Santé*, Paris, France, 2015.

► A. Rouanet, C. Proust-Lima, P. Joly, JF. Dartigues, H. Jacqmin-Gadda, Modélisation conjointe du déclin cognitif, du risque de démence et du risque de décès, *Rencontre des Jeunes Statisticiens*, Arcachon, France, 2015.

► A. Rouanet, H. Jacqmin-Gadda, Analyses of longitudinal data with follow-up truncated by death, *International Workshop on Statistical Modelling*, Rennes, France, 2016.

► A. Rouanet, H. Jacqmin-Gadda, Comparison of methods for longitudinal data with follow-up truncated by death, *Population-based Time-to-event Analyses International Conference*, London, United Kingdom, 2016.

Written communications (posters) at international conferences

► A. Rouanet, H. Jacqmin-Gadda, Illness-death joint model for longitudinal data and competing interval-censored events: Application to the study of natural history of Alzheimer's disease, *Statistical Analysis of Multi-outcome data workshop*, Cambridge, United Kingdom, 2014.

► A. Rouanet, D. Sylla, H. Jacqmin-Gadda, Comparing GEE, weighted GEE and mixed models for longitudinal data with follow-up truncated by death, *Workshop on flexible models for longitudinal and survival data with applications in biostatistics*, Warwick, United Kingdom, 2015.

Scientific award

► Student award in IBS Channel Conference, rank 2nd, Nijmegen, the Netherlands, 2015

Notations and abbreviations

Notations

- $\mathcal{L}$: Log-likelihood
- G: number of latent classes
- B: Variance matrix of random effects
- $\Lambda(\cdot)$: Latent process
- u_i: Random effects
- Y: Longitudinal marker
- $\alpha(\cdot)$: instantaneous hazard function in survival models or transition intensity function in multi-state models
- $A(\cdot)$: Cumulative risk function
- $F(\cdot)$: Cumulative incidence function
- $S(\cdot)$: Survival function
- T: Time-to-event variable

Abbreviations

BVRT: Benton Visual Retention Test

CEP: Certificat d'Etudes Primaires (educational level)

DAR: Death at random

DCAR: Death Completely At Random

DNAR: Death Not At Random

EM: Expectation-Maximization

GEE: Generalized Estimating Equations

IEE: Generalized Estimating Equations with independent working correlation matrix

IST: Isaacs Score Test

JLCM: Joint latent class model

MCAR: Missing Completely At Random

MAR: Missing At Random

MLE: Maximum Likelihood Estimation

MNAR: Missing Not At Random

MMSE: Mini-Mental Score Examination

PA: Population-averaged

SS: Subject-specific

WIEE: Weighted Generalized Estimating Equations with independent working correlation matrix

WIEE1: Generalized Estimating Equations with independent working correlation matrix, weighted by the inverse probability to be observed

WIEE2: Generalized Estimating Equations with independent working correlation matrix, weighted by the inverse probability to be observed given the subject is alive

Résumé substantiel

1 Introduction

En 2012, l'Organisation Mondiale de la Santé a reconnu la démence comme une priorité de santé publique, sensibilisant sur ses coûts sanitaires, sociaux et économiques. La démence est une maladie chronique caractérisée par un long déclin cognitif, plus marqué que le déclin cognitif normal, avec des conséquences dans la vie quotidienne. En 2015, Prince et al. [2015] ont estimé le nombre de cas de démences à 46.8 millions dans le monde, avec 9.9 millions de nouveaux cas chaque année. La prévalence augmente avec l'âge, atteignant 4.6% en Europe Centrale et 8.7% en Afrique du nord chez les plus de 60 ans. Sous l'hypothèse que l'incidence de la démence reste constante et qu'aucune intervention de santé publique n'est mise en place, Wanneveich et al. [2016] ont prédit une augmentation de 47.2% des cas de démences entre 2015 et 2030 en France.

Plusieurs facteurs de risque de la démence ont été mis en évidence dans la littérature tels que l'âge [Letenneur et al., 1994], un faible niveau d'étude [Fabrigoule et al., 1995], le sexe féminin [Fratiglioni et al., 2004], ou l'allèle E4 du gène Apolipoprotéine E [Farrer et al., 1997]. L'identification de tels facteurs de risque est un enjeu majeur de la recherche actuelle, afin de mieux comprendre les mécanismes d'action de la démence.

Le diagnostic complet de démence comprend une évaluation des fonctions cognitives, de l'autonomie, un entretien clinique, l'entretien d'un proche, un examen d'imagerie cérébrale et un test sanguin. L'évaluation cognitive constitue une phase clé de cette procédure et est effectuée via des tests psychométriques, tels que le test global *Mini Mental State Examination* (MMSE) [Folstein et al., 1975] ou le test Isaacs qui cible la fluence verbale [Isaacs and Kennie, 1973]. Ces tests sont rapides et peu coûteux, avantageux pour la détection de la démence. Cependant, il n'y a aucun consensus concernant le test le plus discriminant pour le diagnostic de démence.

Plusieurs traitements ont été proposés mais aucun ne s'est montré efficace dans l'amélioration de l'état cognitif et clinique des patients [Birks, 2006; Courtney et al., 2004], probablement parce qu'ils sont administrés trop tard dans le processus de dégradation cognitive. Cibler la population à risque permettrait alors de mieux prendre en charge les personnes atteintes de démence en leur administrant un traitement ou en leur proposant des interventions non pharmaceutiques plus tôt. Il est également intéressant de savoir si seulement certains sous-groupes de la population sont réceptifs à un traitement [Vellas et al., 2012].

Afin d'étudier le déclin cognitif pré-démence, des outils statistiques sont requis pour l'analyse de données longitudinales, tels que des tests psychométriques ou marqueurs biologiques répétés dans le temps, et des temps d'événement, tel que le temps de survenue de la maladie. Le déclin cognitif est un facteur de risque de la démence d'une part, et la démence interrompt le suivi longitudinal d'autre part, induisant des données manquantes informatives du marqueur. Les modèles conjoints permettent de prendre en compte la corrélation entre le temps d'événement et les mesures répétées, de manière à décrire le déclin cognitif ou à prédire le risque de démence sans biais [Tsiatis and Davidian, 2004; Proust-Lima et al., 2014].

Appliquée à l'étude de la démence, cette approche soulève plusieurs challenges méthodologiques : premièrement, la prévalence de la démence est plus importante chez les personnes âgées, qui sont aussi à plus fort risque de décès. Le décès et la démence sont donc des événements compétitifs. Plus exactement, on parle de risques semi-compétitifs car les individus déments peuvent mourir mais la démence ne peut pas survenir après le décès. De plus, ces deux événements ont des facteurs de risque communs, tels que l'âge ou le sexe. Il est donc primordial de traiter le risque compétitif de décès dans nos analyses afin d'obtenir des résultats non biaisés. Deuxièmement, dans les études de cohortes, le temps de survenue de la démence est censuré par intervalle : les individus ne peuvent être diagnostiqués déments qu'à l'occasion des visites, parfois espacées de plusieurs années. L'âge exact en début de démence n'est donc pas connu et surtout, des individus déments peuvent décéder avant la visite de diagnostic. Enfin, l'hypothèse d'une population homogène est peu probable, au vu de la forte hétérogénéité des déclins cognitifs observée dans la population globale, chez les déments et chez les non déments [Schaie, 1988; Christensen

et al., 1999; Dartigues et al., 1996; Wilson et al., 2002]. Les modèles conjoints développés jusqu'ici ne prennent pas en compte à la fois l'hétérogénéité des déclins cognitifs, le risque compétitif de décès et la censure par intervalle [Rizopoulos, 2011; Dantan et al., 2011; Proust-Lima et al., 2016].

Les modèles conjoints offrent de nombreuses possibilités pour mieux capturer la complexité relative à l'analyse de données longitudinales corrélées à des temps d'événement. Cependant, dans le contexte de données tronquées par le décès, l'interprétation des modèles conjoints (ainsi que des modèles mixtes) estimés par maximum de vraisemblance, est débattue. Lorsque le décès est la seule cause d'attrition, cette méthode d'estimation est équivalente à l'imputation des données manquantes dues au décès. Ainsi, Kurland et al. [2009] considèrent que les estimateurs du maximum de vraisemblance ciblent l'évolution du marqueur dans la population immortelle, irréaliste, qui aurait été observée avec un risque de décès nul et sans sortie d'étude. L'approche partiellement conditionnelle, qui regroupe les modèles marginaux estimés par Equations d'Estimation Généralisées (GEE), cible quant à elle l'évolution du marqueur dans la population dynamique des survivants.

Par ailleurs, l'interprétation des estimateurs de ces deux approches est également différente sur des données complètes. Les estimateurs du maximum de vraisemblance des modèles mixtes sont dit 'spécifiques au sujet', car ils représentent l'évolution du marqueur conditionnellement aux effets aléatoires, qui peuvent être interprétés comme des covariables individuelles non mesurées. Les estimateurs GEE des modèles marginaux sont moyennés sur la population et représentent l'évolution du marqueur, moyennée sur tous les individus. L'utilisation de l'une ou l'autre de ces deux méthodes dépend de l'objectif de l'étude. De notre point de vue, les estimateurs spécifiques au sujet sont plus pertinents dans l'étude de l'histoire naturelle de la démence, où l'on cherche à caractériser le changement cognitif individuel plutôt que le changement cognitif moyen dans la population.

L'objectif de ce travail est de développer des outils statistiques pour étudier le déclin cognitif général ou pré-démence, en prenant en compte le décès. Dans la première partie, nous proposons un modèle conjoint à classes latentes pour données longitudinales corrélées à un événement censuré par intervalle, en compétition avec le décès. Appliqué à la cohorte Paquid, ce modèle permet d'identifier des profils de déclin cognitif associés à des risques

différents de démence et de décès. En utilisant cette méthodologie, nous comparons dans une deuxième partie des modèles pronostiques dynamiques pour la démence, traitant la censure par intervalle, et basés sur des mesures répétées de marqueurs cognitifs. Dans la troisième partie, nous conduisons une étude comparative afin de clarifier l'interprétation des estimateurs du maximum de vraisemblance des modèles mixtes et conjoints et des estimateurs GEE des modèles marginaux, dans le contexte de données longitudinales incomplètes et tronquées par le décès.

2 Modèles conjoints à classes latentes pour données longitudinales et événement censuré par intervalle en compétition avec le décès

Les modèles conjoints pour données longitudinales et temps d'événement se répartissent en deux catégories : les modèles conjoints à effets aléatoires partagés [Wulfsohn and Tsiatis, 1997], qui modélisent la relation entre les deux variables via des effets aléatoires communs, continus, et les modèles conjoints à classes latentes [Lin et al., 2002], où la relation est capturée par des classes latentes, représentées par une variable d'appartenance catégorielle. Dans le cadre de l'étude de la démence, le dernier modèle semble plus approprié car il permet de prendre en compte l'hétérogénéité des déclins cognitifs. L'objectif de ce travail est de développer un modèle conjoint à classes latentes pour données longitudinales corrélées à un temps d'événement censuré par intervalle, en compétition avec le décès.

Modèle

Le modèle que nous proposons repose sur l'hypothèse d'une population hétérogène, composée de G sous-groupes homogènes non connus *a priori*, appelés classes latentes. La probabilité que l'individu i, $i = 1, ..., N$ appartienne à la classe g, $g = 1, ..., G$, est notée $\pi_{ig} = P(c_i = g)$ et est définie par un modèle multinomial logistique. La variable d'appartenance c_i est égale à g si le sujet i appartient à la classe g.

Au temps t_{ij}, $j = 1, ..., n_i$, la valeur observée du marqueur Y_{ij} est la somme du processus latent $\Lambda_i(t_{ij})$, qui représente le vrai niveau cognitif, et d'une erreur de mesure ϵ_{ij}.

Conditionnellement à la classe g, la trajectoire du marqueur est modélisée par un modèle mixte :

$$Y_{ij} = \Lambda_i(t_{ij}) + \epsilon_{ij} = f_1(X_{ij};\ \beta_g) + f_2(Z_{ij};\ \beta_g)\ u_{ig} + \epsilon_{ij},\ \text{avec}\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2) \qquad (1)$$

où X_{ij} est un vecteur de covariables pour le sujet i au temps t_{ij}, β_g est le vecteur de paramètres de régression spécifiques aux classes et Z_{ij} est un sous-vecteur de X_{ij}. Sachant la classe g, le vecteur des effets aléatoires u_{ig} suit la loi $\mathcal{N}(0, \sigma_g^2 B)$ avec $\sigma_G^2 = 1$ et B une matrice définie positive. Les effets aléatoires sont indépendants des erreurs de mesure et ces dernières sont indépendantes entre elles. Les fonctions f_1 et f_2 peuvent inclure des fonctions non linéaires du temps, des covariables et des paramètres de régression. Une transformation non linéaire peut être utilisée pour lier les observations du marqueur, pas forcément gaussiennes, au processus latent gaussien. En effet, certains tests psychométriques ont des propriétés métrologiques particulières comme des effets plancher ou plafond.

Simultanément, les intensités de transition vers la démence et le décès sont modélisées par un modèle Sain-Dément-Décédé spécifique à la classe g. Sachant la classe g, l'intensité de transition de l'état k à l'état l dépend de l'âge t et est modélisé par un modèle à risques proportionnels :

$$\alpha_{klig}(t) = \alpha^0_{klg}(t)\ \exp(W_{kli}^\top\ \gamma_{klg}), \qquad (2)$$

où α^0_{klg} est l'intensité de transition de base, W_{kli} est un vecteur de covariables indépendantes du temps et γ_{klg} sont des paramètres spécifiques aux classes. Dans une version semi-markovienne, l'intensité de transition des déments vers le décès dépend du temps passé en démence et non pas de l'âge. L'hypothèse centrale de ce modèle porte sur l'indépendance des mesures répétées du marqueur et des temps d'événement, conditionnellement à la classe latente.

La censure par intervalle du temps de démence est ensuite prise en compte dans le calcul de la vraisemblance. Pour chaque individu vu sans démence à t_{ij} et mort à $T_i > t_{ij}$, on considère deux chemins possibles : soit il est resté sans démence jusqu'au décès, soit il est devenu dément entre t_{ij} et T_i. La probabilité de développer une démence augmente avec la longueur de l'intervalle de censure $]t_{ij}, T_i]$.

La log-vraisemblance, qui a une forme analytique grâce à l'hypothèse d'indépendance,

est maximisée par l'algorithme de Marquardt [Marquardt, 1963] et les variances des estimateurs sont obtenues par inversion de la matrice Hessienne. Afin de choisir le nombre optimal de classes latentes, G, le modèle doit être estimé avec différentes valeurs de G. Un des critères de sélection le plus utilisé est le BIC [Schwarz, 1978]. Enfin, il est conseillé d'estimer le modèle à partir de valeurs initiales différentes pour éviter les maxima locaux [Hipp and Bauer, 2006], caractéristiques des modèles de mélange [Redner and Walker, 1984].

Une fois le modèle estimé, les probabilités d'appartenance aux classes *a posteriori* $\hat{\pi}_{ig}$, sachant les observations du marqueur et les temps d'événement, permettent d'attribuer une classe à chaque individu puis de quantifier la discrimination du modèle. Le modèle sera d'autant plus discriminant que la probabilité moyenne $\overline{\hat{\pi}}_{ig}$ des individus alloués à la classe g est proche de 1.

Nous proposons une méthode d'évaluation des prédictions longitudinales conditionnellement ou marginalement aux classes et conditionnellement ou marginalement aux effets aléatoires, à partir des prédictions du marqueur conditionnelles aux classes latentes et de la moyenne des observations pondérée par les probabilités d'appartenance $\hat{\pi}_{ig}$. L'évaluation des prédictions du modèle Sain-Dément-Décédé peut se faire conditionnellement aux classes, en comparant les incidences cumulées spécifiques aux classes prédites aux incidences cumulées estimées par le modèle Sain-Dément-Décédé proposé par Joly et al. [2002], qui traite la censure par intervalle et où les contributions individuelles à la vraisemblance sont pondérées par les probabilités *a posteriori* d'appartenance aux classes $\hat{\pi}_{ig}$.

Application

Le modèle conjoint à classes latentes présenté ci-dessus est appliqué à la cohorte Paquid [Letenneur et al., 1994], mise en place pour étudier les déclins cognitifs normal et pathologique à partir de tests psychométriques, chez les individus de Gironde et Dordogne âgés de plus de 65 ans. L'objectif est de distinguer des profils de déclins cognitifs associés à des risques de démence et de décès différents.

L'échantillon sélectionné, comprenant 3 525 sujets, contient les individus sans démence à leur entrée dans la cohorte, qui ont effectué au moins une fois le test d'Isaacs de fluence verbale avant leur visite de diagnostic. Le suivi total est de 25 ans, au cours duquel les

sujets ont été vus tous les 2 ou 3 ans.

Nous avons comparé plusieurs modèles conjoints avec différentes spécifications : tout d'abord, le modèle markovien avec changement de pente spécifique à la classe a un BIC plus faible que le modèle markovien avec tendance quadratique, quel que soit le nombre de classes latentes (G=1,...,5). De plus, la comparaison avec un modèle semi-markovien avec changement de pente favorise l'hypothèse markovienne. L'intensité de transition des déments vers le décès semble donc plus dépendre de l'âge que du temps passé en démence. Nous avons également estimé un modèle semi-markovien, où l'intensité de transition des déments vers le décès dépend du temps passé en démence et de l'âge courant, avec changement de pente. Le BIC de ce dernier modèle est plus élevé que le modèle markovien, quelque soit le nombre de classes latentes.

Les estimations du modèle conjoint markovien à changement de pente, à 4 classes latentes, permettent de représenter les profils d'évolution cognitive dans chacune des classes ainsi que leurs intensités de transition vers la démence et vers le décès associées. Les courbes obtenues attestent une fois de plus la forte hétérogénéité dans la population. Nous avons également cherché à représenter des trajectoires cognitives typiques, pour des individus ayant développé la maladie ou étant décédé à un âge donné. Il est alors possible de différencier le déclin cognitif pré-démence du déclin cognitif pré-décés (appelé déclin terminal) sans démence. Ce déclin terminal a été décrit dans la littérature [Kleemeier, 1962; Siegler, 1975; Wilson et al., 2003], mais sans le distinguer du déclin pré-démence. Au vu des résultats, le déclin avant la démence est plus marqué que le déclin avant le décès sans démence.

3 Prédictions dynamiques pour la démence

Le modèle conjoint présenté a ensuite été utilisé pour prédire de manière dynamique le risque de démence depuis l'entrée dans la cohorte, à partir de mesures répétées d'un ou plusieurs tests cognitifs.

L'échantillon d'apprentissage, issu de la cohorte Paquid, inclut 2 490 sujets sans démence à leur entrée dans la cohorte, ayant effectué au moins une fois le test global MMSE, le test de fluence verbale d'Isaacs et le test de rétention visuelle de Benton, avant leur

visite de diagnostic. Trois modèles conjoints markoviens ont été estimés sur cet échantillon : un modèle basé sur le MMSE, un modèle basé sur l'Isaacs et un dernier basé sur les deux tests. Pour chacun des modèles, on fait l'hypothèse que l'évolution cognitive a une tendance quadratique, avec des effets communs des covariables sur les classes, mais des effets spécifiques aux classes sur les intensités de transition.

Une fois estimés, ces modèles ont été validés sur un échantillon issu de la cohorte Trois-Cités [3C Study Group, 2003], incluant 3 880 sujets de Bordeaux et Montpellier, sans démence à leur entrée dans la cohorte, ayant effectué au moins une fois le MMSE et l'Isaacs avant leur visite de diagnostic. A différents temps *landmark*, nous calculons chez les individus sans démence la probabilité *a posteriori* de développer une démence dans les 5 prochaines années, sachant les observations du marqueur. La qualité prédictive des trois modèles est évaluée via l'extension de l'aire sous la courbe (Area Under the Curve, AUC) et du Score de Brier proposés par Blanche et al. [2015], pour un marqueur longitudinal et un événement dépendant du temps, censuré à droite avec risque compétitif.

Les résultats obtenus confirment ceux de Blanche et al. [2015] : l'Isaacs a une meilleure capacité prédictive que le MMSE, en termes de discrimination (AUC) et de calibration (score de Brier). Cependant, combiner l'Isaacs au MMSE ne semble pas améliorer les performances prédictives. Différentes pistes d'amélioration de ces résultats préliminaires sont discutés dans le document principal.

4 Interprétation de modèles mixtes et modèles marginaux avec attrition due au décès et à la sortie d'étude

Dans les études épidémiologiques chez les personnes âgées, le suivi peut être interrompu par le décès ou la sortie d'étude. Les deux méthodes les plus utilisées dans ce contexte sont les modèles mixtes et les modèles marginaux. Dans cette partie, nous comparons les modèles mixtes (et modèles conjoints) estimés par maximum de vraisemblance aux modèles marginaux estimés par GEE avec matrice de corrélation de travail indépendante (IEE), pondérées et non pondérées, en termes d'interprétation, d'efficacité et de robustesse via une étude de simulations.

Nous nous intéressons à différentes structures d'association entre le marqueur et le risque de sortie d'étude, à partir des hypothèses de données manquantes complètement aléatoires (MCAR) ou aléatoires (MAR) définies par Kurland et al. [2009]. Pour caractériser le lien entre le marqueur et le décès, nous définissons les hypothèses de décès complètement aléatoire (DCAR), aléatoire (DAR) ou non aléatoire (DNAR) selon lesquelles le risque de décès dépend respectivement des covariables, des covariables et des observations du marqueur ou des covariables et des caractéristiques non observées du marqueur (telle que la vraie valeur courante ou pente courante du marqueur).

Estimateurs et interprétation

Sans attrition due au décès ou à la sortie d'étude, il y a consensus sur le fait que les paramètres du modèle mixte représentent le changement individuel ('*subject-specific*') tandis que les paramètres du modèle marginal décrivent la moyenne sur la population ('*population-averaged*'). Dans le contexte linéaire, les paramètres du modèle mixte ont les deux interprétations.

Lorsque le suivi est interrompu par le décès et par la sortie d'étude, l'interprétation des estimants est sujet à débat. La procédure d'estimation IEE des modèles marginaux fournit les estimateurs *population-averaged* dans la population dynamique des individus vivants et observés (non sortis de l'étude). Le décès et la sortie d'étude étant souvent associés au marqueur, plusieurs méthodes de pondération ont été proposées pour corriger le biais de sélection. Nous considérons deux méthodes de pondération : par l'inverse de la probabilité d'être vivant et observé (WIEE1) et par l'inverse de la probabilité d'être observé sachant que l'individu est vivant (WIEE2). Les estimateurs WIEE1 ciblent la trajectoire moyennée sur la population immortelle alors que les estimateurs WIEE2 ciblent la trajectoire moyennée sur la population des survivants à chaque âge.

Dans ce même contexte, la procédure d'estimation du maximum de vraisemblance des modèles mixtes ne fournit pas les estimateurs *population-averaged* parmi la population des survivants. Ainsi, certains auteurs [Kurland et al., 2009; Dufouil et al., 2004] considèrent que les estimateurs obtenus ne sont interprétables que dans la population immortelle, sans décès ni sorties d'étude. Nous démontrons qu'ils ciblent également le changement

individuel chez les individus vivants. Les estimateurs du maximum de vraisemblance des modèles mixtes sont robustes lorsque les processus de décès et de sortie d'étude sont aléatoires ou complètement aléatoires, tandis que les estimateurs du maximum de vraisemblance des modèles conjoints bien spécifiés le sont lorsque le processus de décès est informatif.

Application

Ces modèles sont appliqués à la cohorte Paquid pour étudier l'effet du sexe sur le déclin cognitif, quantifié par le test d'Isaacs. Les estimations du modèle conjoint montrent que l'effet du sexe sur le changement individuel parmi les sujets vivants n'est pas significatif : un homme et d'une femme, vivants, de même âge et même niveau d'études, ont des évolutions cognitives similaires. Cependant, les estimations de WIEE2 montrent que le sexe a un effet significatif sur l'évolution moyennée sur la population des survivants. Cette différence est due à une sélection par le décès plus forte chez les hommes à bas niveau cognitif. Les hommes vivants ont ainsi un niveau cognitif plus élevé que les femmes toujours en vie (les femmes ayant un risque de décès plus faible à niveau cognitif égal et âge égal).

5 Discussion

Dans les études de cohortes, la censure par intervalle est souvent négligée, même lorsque les visites sont espacées de plusieurs années. Le projet initial de ce travail était d'étendre les modèles conjoints pour prendre en compte la censure par intervalle du temps de démence. Cependant, le décès est un enjeu central dans ce type d'analyses, car il entraîne une sélection de la population et ne peut être considéré comme n'importe quelle autre cause de non-réponse, les données post-mortem n'existant pas. Des discussions au sein du groupe Melodem (Methods in longitudinal dementia research) ont soulevé des questions et des débats sur l'interprétation des modèles conjoints lorsque les mesures répétées sont tronquées par le décès. Ces discussions nous ont amenées à nous concentrer sur un marqueur longitudinal corrélé au temps de décès pour clarifier l'interprétation des modèles partiellement conditionnels (modèles marginaux) et des modèles non conditionnels (modèles mixtes et conjoints) et justifier l'utilisation de ces derniers.

Les analyses du déclin cognitif et de la démence nécessitent des outils sophistiqués pour décrire l'histoire naturelle de la démence sans biais, et pour comprendre ses mécanismes. Ces outils permettent de clarifier le rôle de facteurs de risque, en distinguant leur impact sur le déclin cognitif, sur le risque de démence et le risque de décès. Les modèles conjoints sont aussi utiles pour diagnostiquer de manière plus précoce, en identifiant les sujets à haut risque de démence. La recherche actuelle en pharmacologie s'oriente d'ailleurs vers le développement de nouveaux traitements, donnés à des stades plus précoces de la maladie. Enfin, ces travaux méthodologiques sont applicables à toute maladie chronique où le décès est un événement compétitif.

1 Introduction

In 2012, the World Health Organization recognized dementia as a public health priority, rising awareness on its health, social and economic costs. The burden of dementia is increasing as the world's population is ageing. This work aims at studying and proposing statistical methods to better understand the natural history of dementia and cognitive ageing and to enhance earlier diagnosis based on repeated measures of cognitive markers.

1.1 Epidemiology of dementia

1.1.1 Definition and risk factors

Dementia is a chronic disease characterized by a significant cognitive decline, steeper than the normal cognitive decline that comes with ageing. This insidious disease induces a deterioration of the global cognitive sphere, with alteration of memory, thinking, attention, orientation, language and reasoning, as well as behavioral changes and consequences in daily life, and notably undermines health. Among the degenerative forms of dementia, we find Alzheimer's disease (which represents 60 to 70% of dementia cases), fronto-temporal dementia, Lewy Body dementia and Parkinson's dementia. Nowadays, dementia is one of the main causes of dependency and disability among the elderly [Dartigues et al., 2012] and represents a strong burden for the diseased individuals, their family members and caregivers.

As of 2015, the number of individuals with dementia worldwide was estimated to 46.8 million, with 9.9 million new cases each year [Prince et al., 2015]. Some studies suggested a steady or decreasing time trend of dementia incidence [Matthews et al., 2013; Wu et al., 2016] in the recent years but low and middle-income countries are still dramatically affected, as their population growth is more important. The prevalence of dementia increases with age and reaches up to 4.6% in Central Europe and 8.7% in North Africa, among the population aged 60 and over. Assuming no health intervention and no change

in dementia incidence over time, Wanneveich et al. [2016] predicted an increase of 47.18% between 2015 and 2030 in France, based on the expected rise in life expectancy.

Global costs of dementia were estimated to $818 billion in 2015 [Prince et al., 2015], that is 1.09% of the global gross domestic product, based on the World Bank country classification by income. These costs include direct medical costs (20%), direct social and care costs provided by professional caregivers (40%) and costs of informal care, provided by the family for example (40%). The latter are valued based on the mean wage of each country. When comparing with the same estimates made in 2010, the greatest cost increases were observed in Africa and East Asia, where dementia prevalence increased considerably. Dementia is actually a global and major public health concern, and scientific research is currently focusing on exploring the biological processes of the disease, as well as the action mechanism of its risk factors in order to develop new treatments and prevention strategies.

Several risk factors of dementia have been identified, such as age [Letenneur et al., 1994], a low occupational attainment [Stern et al., 1994], a poor social environment [Fabrigoule et al., 1995], E4 allel of genetic factor Apolipoprotein E [Farrer et al., 1997], hypertension [Forette et al., 1998], a low educational level [Letenneur et al., 1999], female gender [Fratiglioni et al., 2004], depression [Ownby et al., 2006], a poor physical activity and consumption of antioxydant [Berr et al., 2009], and smoking [Peters et al., 2008]. Their impact may be accumulated throughout lifetime, possibly from midlife periods onwards [Fratiglioni et al., 2004]. Identifying such risk factors is essential first to understand the natural history of dementia and then to target the susceptible population and implement both health interventions on modifiable risk factors (such as hypertension) and individual care.

1.1.2 Diagnosis and treatment

A complete dementia diagnosis should include cognitive testing, disability assessment, clinical interview and informant interview, brain imaging and blood testing. In the Paquid study presented later in this manuscript, the procedure for cognitive assessment included the following criteria, based on the revised third version of the DSM-IIIR [American Psychiatric Association, 1987]: impairment of short-term and long-term memory and at least one other cognitive function (disturbance of executive functioning such as abstract

thinking, aphasia - language impairment, apraxia - motor disorder due to brain damage, agnosia - inability to associate an object with its meaning, to recognize objects or people) with consequences on social or professional life. Changes in behavior, such as irritability, anxiety, emotional lability (excessive emotional reactions and frequent mood changes) or apathy are cited as complementary criteria for dementia, when accompanied by memory impairment. The fourth version DSM-IV [American Psychiatric Association, 2000], used in the Three-city Study which is also exploited in this manuscript, does not distinguish short-term from long-term memory and does not include behavioral changes in its criteria.

At last, the different categories of dementia are distinguished according to a differential diagnosis, possibly based on biological and imaging biomarkers of the cerebrospinal fluid or on cognitive tests such as the Free and Cued Selective Reminding Test [Grober and Buschke, 1987] which distinguishes Alzheimer's disease from fronto-temporal dementia. Recently, dementia was renamed "major neurocognitive disorder" in the DSM-V [Association, 2013].

The neuropsychological examination for cognitive assessment is a central part of the diagnosis process. Based on several psychometric tests, it assesses the different cognitive functions to identify the altered ones. The most common test is the Mini Mental State Examination (MMSE) [Folstein et al., 1975], which assesses cognition globally through different items (registration, attention, calculation, recall, language, ability to follow simple commands and orientation). Other tests are more specific, such as the Isaacs Set Test [Isaacs and Kennie, 1973], which evaluates verbal fluency, or the Benton Visual Retention Test for visual perception and memory. So far, there is no consensus on the most discriminatory psychometric test for dementia diagnosis. Psychometric tests are also valuable tools for early detection of subjects at high risk of dementia, as they are cost-effective, not invasive and they enable a quantitative assessment of cognition over time. Besides, combining information from multiple longitudinal cognitive tests is likely to enhance predictive abilities of such tools. Ideally, a prediction could be obtained as soon as at least one measure of any cognitive tests is available and could be up-dated after each cognitive measurement.

To date, several treatments have been proposed against dementia, most of which are

based on the enhancement of the transmission of a chimical compound, acetylcholine, between neurons. No drug treatment has proved able to stop the progression of dementia and only a modest efficiency, sometimes associated with relatively important side effects, was shown on cognitive and clinical states [Birks, 2006; Courtney et al., 2004]. An hypothesis is that treatment administration comes too late in the disease process. Targeting the population at risk at earlier stages may improve the efficiency of such treatments, as well as non-pharmacological interventions, including memory training, stimulation of social environment, physical or psychological support to reduce the risk of daily life accidents.

1.1.3 Heterogeneity in natural history of dementia

Cognitive decline is heterogeneous among the general elderly population [Schaie, 1988; Colsher and Wallace, 1991; Christensen et al., 1999], but also at all ages of adulthood [Wilson et al., 2002]. Some studies highlighted different patterns between normal ageing and ageing with dementia from neuropathological and neuroradiological data [West et al., 1994; Jobst et al., 1994] and psychometric data [Dartigues et al., 1996; Backman et al., 2001].

Dementia is conceptualized as a continuous pathological process, falling along a continuum with normal ageing, which also implies a cognitive decline with age [Brayne and Calloway, 1988]. This concept is based, amongst others, on the difficulty to distinguish normal elderly from mild dementia regarding cognitive performances. Indeed, there is a lag period between physiological alterations, appearance of clinical symptoms and cognitive impairment. In Alzheimer's Disease for example, neurodegeneration can occur relatively late after the aggregation of a peptide called Aβ in the brain, which is the start of the pathological cascade defined in Jack et al. [2010]. This concept led to the hypothesis of a transitory period, called Mild Cognitive Impairment (MCI), during which the cognitive functions are altered with no significant consequences on daily life, so that dementia criteria are not met.

The length of this pre-clinical phase is variable among individuals living with dementia. This heterogeneity may be explained by the variability between diagnosis procedures but it is mainly due to inter-individual variability. Stern [2009] formulated the hypothesis of a 'brain reserve', linked to the structure and composition of the brain, which may

stimulate a 'cognitive reserve'. The latter is represented by all the cerebral processes which compensate the cognitive loss and help individuals to cope with the brain pathology during a certain period, delaying the appearance of the first clinical symptoms and lengthening this 'silent' phase.

Several studies focused on this pre-diagnosis phase to evaluate whether cognitive impairment could be detected from earlier stages. Thus, an acceleration of cognitive decline was observed long before diagnosis [Bäckman et al., 2005], from 5 [Hall et al., 2000] to 12 years earlier [Amieva et al., 2008], according to the cognitive function under consideration. As an example, a decline of semantic memory was observed 12 years ahead of diagnosis on average. It is also of interest to evaluate the impact of treatments given at earlier phases on the risk of dementia [Vellas et al., 2012] or to evaluate if sub-groups are receptive to a given treatment, based on data from prevention trials. This can lead to personalized medicine.

1.2 Cohorts on dementia

Several cohorts in the general population were set up to provide needed information for understanding the epidemiology of dementia, to build diagnosis and prediction tools or to evaluate interventions and guide public health policies. Among the studies which collected repeated measures of cognitive tests and time-to-dementia diagnosis, we can list: the Framingham study [Farmer et al., 1987], the Paquid cohort [Dartigues et al., 1992], the Three-city study [3C Study Group, 2003], the Rush Memory and Aging Project [Bennett et al., 2005], the Mayo Clinic Study of Ageing [Roberts et al., 2008] or the AMI cohort [Pérès et al., 2012]. This section details the Paquid and Three-city cohorts which took place in France and which were used in analyses presented in this manuscript.

1.2.1 The Paquid cohort

The epidemiological Paquid Cohort ("Personnes Âgées Quid") was initiated in 1988 in two French departments, Dordogne and Gironde, to study normal and pathological brain ageing from repeated psychometric tests. The main objectives included the identification of risk factors and pre-clinical symptoms of Alzheimer's Disease.

A total of 3,777 subjects, aged 65 and older, were followed during more than 25 years, as the study is still ongoing. They were chosen randomly from communal electoral lists to ensure representativeness of the population over 65, who lived independently in their home, in Gironde and Dordogne. The population, composed of 58.2% of women, entered the study on average at 75.5 years old (sd=6.9 years) and included a majority of individuals who obtained their primary school diploma or more (35.5% with no primary school diploma, 43.4% with primary school diploma as the highest degree and 21.1% with higher degrees).

When recruited, subjects were administered a questionnaire by a psychologist about, inter alia, demographic, socio-professional and health information. Then, cognitive functioning was assessed through a battery of psychometric tests: the MMSE on global mental status, the Isaacs Set Test on verbal fluency, the Benton's Visual Retention Test on visual memory, Wechsler's Paired-Associates on verbal memory, the Zazzo test on visual attention and the Digit Symbol Test on simple logical reasoning. Then, the subjects answered a last questionnaire based on the DSM-IIIR criteria for dementia. They were suspected for dementia if they had experienced an alteration of memory and of another cognitive function with repercussions in daily life and if their MMSE score, at follow-up visits, decreased by more than three points from previous visits. If so, they were assigned to a neurologist who possibly confirmed the diagnosis and established the etiology of dementia, based on the NINCDS-ADRDA criteria [McKhann et al., 1984] and the Hachinski score, to determine if the cognitive impairment had a vascular origin [Hachinski et al., 1975].

Interviews were then organized at home approximately at years 1, 3, 5, 8, 10, 13, 15, 17, 20, 22 and 25. Cognitive functioning (and possible clinical dementia diagnosis) was assessed with the same procedure at each follow-up visit. The cause of canceled visits was recorded: no response from the subject, refusal, relocation out of Gironde or Dordogne or death. Age at death was informed by families, general practitioners or retrieved from death registries.

1.2.2 The Three-City cohort

The Three-City study is another French prospective cohort study, designed to quantify the impact of vascular factors on the risk of dementia and cognitive impairment, and to distinguish different groups of subjects at high risk of dementia to guide future potential

prevention strategies.

This cohort recruited 9,285 individuals from electoral rolls in three French cities (Bordeaux, Dijon, Montpellier), who were aged 65 and older, and were not institutionalized between March 1999 and March 2001. On average, the participants, a majority of women (60.7%), entered the study at 74.3 years old (sd=5.6 years) and had studied longer than the Paquid's population (8.7% with no primary school diploma, 17.5% with primary school diploma as the highest degree and 73.9% with higher degrees).

Cognitive functioning was assessed through several tests: the MMSE, the Benton Visual Retention Test, the Isaacs Set Test, the Trail Making Test on visual attention and task switching, the Delayed Recall test on delayed verbal recall, and the Digit Span Test on short-term verbal memory. Other data were collected, such as cerebral Magnetic Resonance Imaging or biological markers. Screening for dementia diagnosis was made by a neuropsychologist based on DSM-IV criteria, then a neurologist performed a clinical examination and an independent panel of expert neurologists possibly confirmed the diagnosis and the etiology of dementia.

According to the center or to the wishes of the participants, visits were done at home or in an exam center at years 2, 4, 7, 10 and 12 approximately. Age at death was informed by practitioners or retrieved from the INSERM epidemiology center on medical causes of death, CépiDC.

1.3 Methodological challenges in studies on cognitive decline and dementia

To study cognitive decline before diagnosis, information must be derived from both repeated measures of markers over time, such as biological or psychometric markers, and from the time-to-diagnosis. Analyzing longitudinal data and time to dementia onset raises several methodological challenges. First, the follow-up of longitudinal data may be truncated by death, with the probability to die potentially linked to observed and unobserved characteristics of the marker trajectory. Second, time to dementia onset may not be known exactly since the data are collected intermittently. Moreover, death is a competing risk of dementia, which may complicate the interval censoring problem. This section details all these issues, which can have an important impact on analysis results.

1.3.1 Attrition due to drop-out in longitudinal studies

In longitudinal observational studies, subjects may drop out for different reasons, inducing missing data. In this manuscript, we will distinguish death and drop-out. In 1987, Little, Roderick JA and Rubin [1987] proposed a classification for the drop-out mechanism, defining missing data as:

- Completely at random (MCAR data) when the probability to drop out is independent from the marker trajectory, conditionally on covariates. This would be the case of an individual who missed a visit because a friend had an accident, for instance.

- At random (MAR data) when the probability to drop out depends on covariates as well as on past observed values of the marker. A particular case states that the probability depends on the very last observed value of the marker. As an example, consider an individual who realizes that his/her observed cognitive scores decline and decides on his own to drop out from the study.

- Not at random (MNAR data or informative missing data) when the probability to drop out depends also on unobserved characteristics of the cognitive marker trajectory, such as the current cognitive value or the cognitive slope at time of drop-out. The mechanism is MNAR if subjects who deteriorated since their last visit tend to drop out.

There is no statistical test to distinguish MAR from MNAR data. When the cohort is prone to attrition due to drop-out only, the target estimand is most often the expectation among the hypothetical population with no drop-out, that is the initial population, which would have been observed until the planned end of the study if no one had dropped out.

1.3.2 Attrition due to death in longitudinal studies

When the collection of longitudinal data may also be interrupted by death, the target expectation is controversial as the population with no death would be immortal and not realistic. However, selection by death is likely heterogeneous over the whole population

so the population of survivors may not be representative of the initial population. As an example, women have a lower risk to die, so at the end of the follow-up, the proportion of women survivors may be different from the proportion of women in the initial population. This selection issue is more critical as death is linked to the cognitive decline [Wilson et al., 2003]: individuals with a high cognitive level have a higher risk to survive so subjects who stay in the study tend to be healthier. Since the risk of death is higher among men than women regardless of cognitive level, a higher mean cognitive level among men survivors than women survivors can either reflect an actual cognitive difference according to gender, or a higher attrition due to death among cognitively impaired men.

Within this framework, the target estimand is most often the mean among the population currently alive [Kurland and Heagerty, 2005]. However, it is necessary to identify the dependence structure between the death process and the longitudinal marker for selecting the appropriate model and obtain unbiased estimates.

1.3.3 Competing risk of death and interval censoring

As dementia prevalence is higher among the elderly, the population at risk of dementia is also at a non-negligible risk of death. Moreover, death and dementia may have common risk factors, such as sex and age. If the correlation between death and dementia is not accounted for, the estimated effect of a common risk factor of interest on the risk of dementia may be biased. Hence, we consider death as a semi-competing event as dementia cannot occur after death while subjects living with dementia can die.

This competing risk issue gets more problematic when time-to-dementia is interval-censored. In cohort studies, the dementia status is evaluated intermittently as subjects can only be diagnosed at visit times. The exact time to dementia onset is then unknown and is interval-censored. As diseased individuals are more susceptible to die, they may die before the visit following dementia onset, without being diagnosed. Thus, interval censoring of time to dementia onset, in a framework of competing risks with death, leads to an under-estimation of dementia incidence.

One way to account for both the semi-competing risk of death and interval censoring is to use an Illness-Death model [Joly et al., 2002], which considers three states – Health, Dementia and Death – and estimates simultaneously the three transition intensities (Health-Dementia, Dementia-Death and Health-Death). It is then possible to account

for a possible unobserved transition to dementia for subjects free of dementia at their last visit who died later, with no information on dementia status at time of death. Moreover, this model differentiates the effect of covariates on death and dementia.

1.3.4 Joint modeling of cognition and dementia

In order to investigate the pre-dementia cognitive decline, the cognitive marker and time-to-dementia should not be analyzed separately as both are highly correlated. On the one hand, cognitive impairment is a predictor of the risk of dementia [Visser et al., 1999; Flicker et al., 1991]. On the other hand, the follow-up is truncated by dementia, inducing informative missing data of the repeated marker [Wu and Carroll, 1988]. Consequently, joint models of cognitive decline and dementia are necessary to estimate the cognitive trajectory in the pre-diagnostic phase or the risk of dementia within the next years given repeated measures of cognitive tests.

It is also imperative to account for the competing risk of death in the joint modeling framework to estimate without bias the effect of covariates on the cognitive decline or on the risk of dementia. Besides, joint modeling the cognitive decline, the risk of death and the risk of dementia would enable to distinguish the impact of covariates on these three variables. Moreover, it is of interest to differentiate the cognitive decline before dementia from the cognitive decline before death without dementia. No joint models were proposed so far to handle longitudinal data correlated to an interval-censored event, competing with death. However, this type of data is frequent in cohort studies and new statistical tools based on the development of such models would be applicable to other types of chronic diseases where death is a competing risk.

1.3.5 Other methodological challenges

As stated before, there is abundant literature on heterogeneity in cognitive declines [Schaie, 1988; Dartigues et al., 1996; Backman et al., 2001; Wilson et al., 2002], among the general population but also among the subjects living with dementia. Mixture models [Muthén, 2008] accommodate this heterogeneity through mixture components. As an example, latent class mixed models consider different unobserved groups, called latent classes, with specific mean trajectories. The distribution of the outcome and the random

effects then follow a mixture of multivariate gaussian distributions. The latent classes are not known *a priori*, but it is possible to label the posterior classes ('normal ageing', 'moderate decline','severe decline', regarding the estimated class-specific mean cognitive trajectories for example) and to compute the individual posterior probabilities to belong to each class.

Besides, standard models assume that the markers are gaussian but cognitive tests may follow other distributions, possibly asymmetric, and may have specific metrological properties such as floor or ceiling effects. For instance, the MMSE fails at discriminating between subjects with high cognitive scores: it has a ceiling effect. Conversely, a test has a floor effect when it cannot discriminate between subjects with low scores. More generally, markers can be curvilinear, that is to say they have a varying sensitivity to change: a one-point difference in low values will not represent the same cognitive loss as a one-point-difference at high values. Proust-Lima et al. [2006] proposed a mixed model with a parametric non-linear transformation, based on splines or Beta cumulative density function, for longitudinal modeling of curvilinear markers. Proust-Lima et al. [2012] also proposed other transformations to deal with discrete, asymmetric or bounded outcomes such as quality of life or autonomy scale.

Finally, cognition is measured by multiple cognitive tests and it may be useful to describe their change over time simultaneously instead of studying one arbitrary selected test. The approaches proposed by Proust-Lima et al. [2006] modeled the unobserved process underlying the different markers by a latent process. The observed cognitive tests are considered as correlated measures of this latent process, to which they may be linked through a non-linear transformation.

1.4 Objective and outline

The purpose of this work is to develop statistical tools to study the general cognitive ageing or the pre-diagnosis cognitive decline in dementia, accounting for selection by death.

In order to describe the pre-dementia phase and to propose prediction tools for the risk of dementia, joint models for time-to-dementia and longitudinal cognitive measures were previously developed but none of them accounted for both the competing risk of death and interval censoring. As a first step, this works aims at developing a joint latent class model for modeling the cognitive decline, the risk of dementia and the risk of death, accommodating heterogeneity of the data, selection by death and interval censoring. In a second part, we performed a comparative study to address the general debate on the best method to use in analyses for longitudinal data when the follow-up is truncated by death, clarifying the interpretation of the different methods proposed in this framework.

Section 2 first presents the state of the art relative to joint models, which capture the correlation between a longitudinal marker and a time-to-event through a latent structure. This structure can be represented by random effects or latent classes, which are unobserved homogeneous sub-groups of the population. Latent class models are more appropriate when data are heterogeneous. Then, we present the models proposed to accommodate selection by death and interval censoring.

Section 3 describes the joint latent class model that we proposed to handle longitudinal data and interval-censored time-to-event with semi-competing risk of death. This model captures heterogeneity in cognitive decline through latent classes. It combines a mixed model to describe the cognitive evolution and an Illness-Death model to estimate the risk of dementia, considering death as a semi-competing event, and interval censoring is accounted for in the computation of the likelihood. Simulations are performed under the markovian and semi-markovian assumptions. This model is then applied to the Paquid cohort in order to distinguish profiles of cognitive declines associated with risks of dementia and death.

Section 4 describes the application of the above model to build a dynamic prediction tool for dementia, using the Paquid cohort data. The prediction tool is then evaluated using the Three-city cohort as validation sample.

Section 5 focuses on methods used for longitudinal data only, when the population is prone to attrition due to drop-out and death. We clarify the interpretations of estimates from mixed models estimated by likelihood maximization and marginal models estimated by unweighted and weighted Generalized Estimating Equations, which are the main methods used to handle longitudinal data in epidemiological studies, when follow-up may be

truncated by death and drop-out. Simulations are carried out to quantify the differences in estimates and an application on Paquid cohort is presented to illustrate the impact of selection by death.

At last, a general discussion summarizes the advantages and the limits of the methods presented, as well as the possible perspectives.

2 State of the art

This chapter exposes models proposed in the literature for handling longitudinal data and possibly time-to-event data correlated to the longitudinal markers. We first describe the methodology of mixed models, marginal models and joint models and then review the different approaches for addressing the selection issue due to death in longitudinal studies on one hand and the competing risk of death and interval censoring in time-to-event analyses on the other hand.

2.1 Methods for longitudinal data

2.1.1 Mixed models estimated by likelihood maximisation

Linear mixed models were introduced by Laird and Ware [1982] to analyze longitudinal gaussian outcomes, accounting for the within-subject correlation.

Specification of linear mixed models

Let Y_{ij} be the observation of a gaussian variable for subject i, $i = 1, ..., N$ at time t_{ij}, $j = 1, ..., n_i$. The standard linear mixed model is written:

$$Y_{ij} = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top u_i + \epsilon_{ij} \tag{3}$$

where $u_i \sim \mathcal{N}(0, B)$ are the subject-specific random effects, with B the variance matrix. The observed value Y_{ij} at time t_{ij} is the sum of the true value of the marker $\tilde{Y}_i(t_{ij})$ and a gaussian measurement error $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, assumed independent across different times and independent from the random effects. The vectors X_{ij} and Z_{ij} include the covariates with fixed and random effects respectively, and Z_{ij} is a sub-vector of X_{ij}. Then, the vector of the marker observations $Y_i = (Y_{i1}, ..., Y_{in_i})^\top$ follows the normal distribution $\mathcal{N}(X_i\beta, V_i = Z_i B Z_i^\top + \Sigma_i)$ where Σ_i is the $n_i \times n_i$ variance matrix of the measurement error vector $\epsilon_i = (\epsilon_{1i}, ..., \epsilon_{in_i})^\top$ and X_i, Z_i are the matrices with row vectors $X_{ij}^\top$ and $Z_{ij}^\top$

respectively.

Estimation

The log-likelihood of the linear mixed model is written:

$$\mathcal{L}(\beta,\phi) = -\frac{1}{2}\sum_{i=1}^{N}\{n_i\ log(2\pi) + log|V_i(\phi)| + (Y_i - X_i\beta)^\top V_i(\phi)^{-1}(Y_i - X_i\beta)\} \quad (4)$$

with ϕ the vector of all the variance parameters (variance matrices of the random effects and measurement errors) and $|V_i(\phi)|$ the determinant of $V_i(\phi)$. The log-likelihood is maximised in (β,ϕ) satisfying this score equation:

$$\frac{\partial\mathcal{L}(\beta,\phi)}{\partial\beta} = \sum_{i=1}^{N} X_i^\top V_i(\phi)^{-1}(Y_i - X_i\beta) = 0 \quad (5)$$

In practice, estimators of regression parameters are obtained by

$$\hat{\beta} = (\textstyle\sum_{i=1}^{N} X_i^\top V_i(\phi)^{-1}X_i)^{-1}(\sum_{i=1}^{N} X_i^\top V_i(\phi)^{-1}Y_i)$$

and then ϕ is estimated by maximisation of $\mathcal{L}(\hat{\beta},\phi)$ via an iteration procedure (such as the Newton-Raphson algorithm). Otherwise, $\mathcal{L}(\beta,\phi)$ can be maximised with respect to both β and ϕ simultaneously, also in an iterative procedure. Note that if the correlation of the random effects is well-specified, the estimators from mixed models are consistent.

Specification of generalized linear mixed models

The theory of mixed models was applied to generalized linear models, as detailed in McCullagh, Peter and Nelder [1989], to describe the evolution of non-gaussian continuous outcomes over time. In generalized linear mixed models, a non-linear function links the expectation of the outcome to the linear predictor. The distribution of the outcome belongs to the exponential family:

$$f_{Y_{ij}}(y;\theta_{ij},\Psi) = \exp\{\Psi^{-1}[\theta_{ij}y - a(\theta_{ij})] + C(y,\Psi)\} \quad (6)$$

where $\theta_{ij} = X_{ij}^\top\beta + Z_{ij}^\top u_i$ is the linear predictor, $a(\cdot)$ and $C(\cdot)$ are some functions and Ψ is the vector of variance parameters. Then, conditionally to the random effects:

$$E(Y_{ij}|u_i) = a'(\theta_{ij}) \text{ and } var(Y_{ij}|u_i) = a''(\theta_{ij})\Psi.$$

The link function is then $[a']^{-1}$, the inverse of the derivative of a. The likelihood is written:

$$L(\beta, \phi) = \prod_{i=1}^{N} log\Big(\prod_{j=1}^{n_i} \int_{u_i} f_{Y_{ij}|u_i}(Y_{ij}|u_i) f_{u_i}(u_i) du_i\Big) \tag{7}$$

with $f_{Y_{ij}|u_i}$ the gaussian density with mean $a'(\theta_{ij})$ and variance $a''(\theta_{ij})\Psi$.

Interpretation

Equation (3) describes the mean trajectory of the outcome conditionally to the subject-specific random effects, which can be considered as individual unmeasured covariates (independent from the observed ones X_{ij}). Thus, the parameters β quantify the change in the outcome for a unit change in X, the other covariates and the random effect being constant: $\beta = E(Y_{ij}|X_{ij} = 1, u_i) - E(Y_{ij}|X_{ij} = 0, u_i)$. They have a 'subject-specific' interpretation as they represent the individual change of the marker for a change in X.

From Equation (3), we also obtain the expectation of the marker marginally to the random effects: $E(Y_{ij}) = X_{ij}^{\top}\beta$, which represents the expectation of the marker averaged over the random effects, i.e the expectation among the whole population. Thus, the parameters β also quantify the mean effect of X on Y averaged over the random effects: $\beta = E(Y_{ij}|X_{ij} = 1) - E(Y_{ij}|X_{ij} = 0)$. They also have a 'population-averaged' interpretation as they represent the mean effect of X on Y over the whole population. In the linear framework, parameters from mixed models have both interpretations.

Parameters from generalized linear mixed models have only a subject-specific interpretation as they are interpretable conditionally to the random effects exclusively: $E(Y_{ij}|X_{ij}, u_i) = a'(X_{ij}^{\top}\beta + Z_{ij}^{\top}u_i)$. The population-averaged interpretation does not hold as $E(Y_{ij}|X_{ij}) = E_{u_i}(E(Y_{ij}|X_{ij}, u_i)) = E_{u_i}(a'(X_{ij}^{\top}\beta + Z_{ij}^{\top}u_i)) \neq a'(X_{ij}^{\top}\beta)$.

Mixed models and missing data

Maximum likelihood estimates of mixed models are robust to MCAR and MAR data: indeed, by modeling the within-subject correlation, the likelihood maximisation procedure implicitly imputes missing data, under the assumption that the conditional distributions of future missing data and future observed data, given past observations, are identical. Thus, the maximum likelihood estimates obtained from observed data are unbiased. Let Y^o be the vector of observed responses, Y^m the vector of missing responses and R the observation process (with $R_{ij} = 1$ if the response at time t_{ij} is observed, and 0 otherwise),

depending on parameter ψ_R. When missing data are MAR, $f(R|Y^o, Y^m, X) = f(R|Y^o, X)$ so the joint density of the observed data and R is:

$$\begin{aligned} f(Y^o, R|\theta, \psi_R) &= \int f(Y^o, Y^m|\theta) f(R|Y^o, Y^m, \psi_R) dY^m \\ &= f(R|Y^o, \psi_R) \int f(Y^o, Y^m|\theta) dY^m \\ &= f(R|Y^o, \psi_R) f(Y^o|\theta) \end{aligned}$$

Thus, the log-likelihood is written:

$$\mathcal{L}(\theta, \psi_R|Y^o, R) = \mathcal{L}(\theta|Y^o) + \mathcal{L}(\psi_R|Y^o, R)$$

so that θ can be estimated without bias by maximisation of $\mathcal{L}(\theta|Y^o)$ on available data. When missing data are not at random, mixed models are not robust anymore, but joint models, which combine a mixed sub-model and a survival sub-model for the time-to-drop-out linked by a latent structure, provide unbiased estimates when the dependence between the two processes is well-specified.

2.1.2 Marginal models estimated by GEE

Liang and Zeger [1986] focused on the marginal distribution of a non-gaussian repeated outcome. They proposed an alternative estimation procedure to likelihood maximisation in order to obtain estimates and variance estimates robust to the misspecification of the within-subject correlation, considered as a nuisance.

Specification of marginal models

The marginal distribution of the outcome Y_{ij} is modeled by a generalized linear model defined by:

$$f_{Y_{ij}}(y; \theta^*_{ij}, \psi^*) = \exp\{\psi^{*-1}[\theta^*_{ij} y - a(\theta^*_{ij})] + C(y, \psi^*)\} \tag{8}$$

where $\theta^*_{ij} = X_{ij}^\top \beta^*$ is the marginal linear predictor with β^* the marginal regression parameters associated with covariates X, and ψ^* is the variance parameters. Then, Y_{ij} has mean $E(Y_{ij}) = a'(\theta^*_{ij})$ and variance $var(Y_{ij}) = a''(\theta^*_{ij})\psi^*$.

Estimation

The generalized estimating equation is written:

$$U(\beta^*) = \sum_{i=1}^{N} D_i^\top V_i(\psi^*)^{-1}(Y_i - \mu_i) = 0 \tag{9}$$

where $Y_i = (Y_{i1}, ..., Y_{in_i})^\top$, $\mu_i = (\mu_{i1}, ..., \mu_{in_i})^\top$ with $\mu_{ij} = a'(X_{ij}^\top \beta^*)$, $D_i = \frac{\mathrm{d}\mu_i}{\mathrm{d}\beta^*} = A_i X_i$ and $V_i = \psi^* A_i^{1/2} R(\alpha) A_i^{1/2}$ the working covariance matrix, with $A_i = diag(a''(X_{ij}^\top \beta^*))$ and α the vector which characterizes the working correlation matrix $R(\alpha)$. If $R(\alpha)$ is the true correlation matrix of Y_i, then $V_i = \mathrm{cov}(Y_i)$. When the repeated measurements are assumed independent, $R(\alpha)$ is the identity matrix and the equation reduces to the likelihood score equation (5).

Once ψ^* is replaced by a consistent estimate $\hat{\psi}^*$, the solution of Equation (9) provides an asymptotically multivariate gaussian estimator with mean β^* and robust variance:

$$V_R(\hat{\beta}^*) = \Big\{ \sum_{i=1}^{N} D_i^\top V_i(\hat{\psi}^*)^{-1} D_i \Big\}^{-1} \Big\{ \sum_{i=1}^{N} D_i^\top V_i(\hat{\psi}^*)^{-1} var(Y_i) V_i(\hat{\psi}^*)^{-1} D_i \Big\} \Big\{ \sum_{i=1}^{N} D_i^\top V_i(\hat{\psi}*)^{-1} D_i \Big\}^{-1}$$

The computation of robust variances is based on the assumed independence between subjects. Then, the estimators $\hat{\beta}^*$ and $\hat{V}_R$ are consistent if the model for $E(Y_{ij})$ is correct, whatever the specification of the working correlation matrix $R(\alpha)$, when covariates are time-independent.

Interpretation

The GEE estimators of marginal models can only be interpreted as population-averaged, since the within-subject correlation is not modeled explicitly: $E(Y_{ij}) = a'(X_{ij}^\top \beta^*)$. For linear models (identity link function $[a']^{-1}$), the parameters β^* represent the mean effect of X on the marker over the whole population.

Subject-specific or population-averaged parameters may be preferred according to the aim of the study. If the purpose of the analysis is to explore the individual change in cognition over time, the individual change for a variation in a given risk factor, or to predict individual decline, subject-specific estimands may be more appropriate. In contrast, if the objective is to quantify the impact of a public health intervention on the population mean, then population-averaged estimands may be more suited.

Missing data

GEE estimates are robust to MCAR data only. Robins et al. [1995] introduced a weighted GEE estimation procedure, with an independent working correlation matrix, to handle MAR data. The individual contributions to the GEE score equation (9) are weighted to correct for selection induced by the missingness mechanism:

$$U_D(\beta^*) = \sum_{i=1}^{N} D_i^\top V_i(\psi^*)^{-1} \mathcal{W}_i (Y_i - \mu_i) = 0 \tag{10}$$

with $\mathcal{W}_i$ the diagonal $n_i \times n_i$-matrix with the j^{th} diagonal element equal to $R_{ij} w_{ij}$ where $R_{ij} = 1$ if subject i is observed at occasion j and 0 otherwise, and w_{ij} is the occasion-specific weight defined by the inverse probability for subject i to be observed at occasion j. Fitzmaurice et al. [1995] defined individual-specific weights as the inverse probability to be observed at the observed time of drop-out, but Preisser et al. [2002] showed that these weights led to biased estimates. At last, no method provides directly marginal estimates in the MNAR framework.

2.2 Joint models for longitudinal and time-to-event data

Joint models were developed in order to account for the relationship between the time to an event and a longitudinal marker [Tsiatis and Davidian, 2004]. They are useful to describe the change over time in the repeated marker in the presence of informative drop-out, to investigate the association between the longitudinal and survival processes, or to predict the time-to-event given repeated values of the marker. The idea of joint models is to combine a mixed sub-model for the trajectory of the longitudinal marker and a failure time sub-model for the risk of the event, both linked by a latent structure. Two approaches were developed, considering either random effects or latent classes as latent structures to capture the dependence between the marker and the time to an event.

2.2.1 Joint shared-random-effect model

Wulfsohn and Tsiatis [1997] proposed a joint model for a gaussian longitudinal marker and the time to an event, in which their relationship is modeled by random effects, assumed to be gaussian. Conditionally to the random effects, the two variables are independent.

Specification of joint shared-random-effect models

The shared-random-effect model incorporates two sub-models. On one hand, a linear mixed model describes the trajectory of the marker, conditional on the random effects:

$$Y_{ij} = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top u_i + \epsilon_{ij} \tag{11}$$

where $u_i \sim \mathcal{N}(0, B)$ are the subject-specific random effects, with B the variance matrix. The observed value Y_{ij} at time t_{ij} is the sum of the true value of the marker $\tilde{Y}_i(t_{ij})$ and a gaussian measurement error $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, assumed independent from measurement errors at different times. The individual random effects and measurement errors are also independent.

On the other hand, the instantaneous hazard rate of the event is modeled by a proportional hazards model, conditional on the random effects:

$$\alpha_i(t|u_i; \beta, \gamma, \gamma^{(a)}) = \alpha_0(t) exp(W_i^\top \gamma + h(u_i, \beta, t_{ij}, Z_{ij}, X_{ij})^\top \gamma^{(a)}) \tag{12}$$

with $\alpha_0(\cdot)$ the baseline hazard function, γ the vector of regression parameters associated with covariates W_i. The function $h(\cdot)$, specified *a priori*, depends on the random effects and possibly on time, on covariates and parameters from the mixed model. For instance, $h(\cdot)$ can give the true current value $\tilde{Y}(t)$ or true current slope of the marker $\frac{\partial \tilde{Y}(t)}{\partial t}$ [Yu et al., 2004]. At last, $\gamma^{(a)}$ is the vector of association parameters between the two processes.

This model was extended to handle multiple events as it will be discussed in section 2.3.3.

Estimation

In the following, we will denote by T_i the time-to-event and by δ_i the occurrence indicator of the event, equal to 1 if the event was observed before the end of the follow-up and 0 otherwise. Based on the independence assumption of Y_i and (T_i, δ_i) conditionally to the random effects u_i, the individual contribution to the joint distribution can be written as follows:

$$f_{Y_i,T_i,\delta_i}(Y_i, T_i, \delta_i; \theta) = \int_{\mathbf{R}^{n_u}} f_{Y_i|u_i}(Y_i|u_i; \theta) f_{T_i,\delta_i|u_i}(T_i, \delta_i|u_i; \theta) f_{u_i}(u_i; \theta) du_i$$

with n_u the dimension of the random effects, θ including the regression and variance parameters from model (11) and the regression, association parameters from model (12).

The multivariate gaussian density $f_{Y_i|u_i}$ has mean $X_i\beta + Z_i u_i$ and variance $\sigma_\epsilon^2 I_{n_i}$ and f_{u_i} is a multivariate gaussian density with mean 0 and variance matrix B. The likelihood is then:

$$L(\theta) = \prod_{i=1}^{N} \int_{\mathbf{R}^{n_u}} f_{Y_i|u_i}(Y_i|u_i;\theta) S_i(T_i|u_i;\theta) \alpha_i(T_i|u_i;\theta)^{\delta_i} f_{u_i}(u_i;\theta) du_i.$$

The Expectation-Maximisation algorithm [Dempster et al., 1977] was used for likelihood maximisation in this longitudinal framework by De Gruttola and Tu [1994]. The idea is to iteratively compute the expectation of the complete data log-likelihood $Q(\theta|\theta^{it}) = \sum_i \int log(P(T_i, \delta_i, Y_i, u_i; \theta)) p(u_i|T_i, \delta_i, Y_i; \theta^{it}) du_i$ from current estimates θ^{it}, and then to find the estimates θ which maximise it. However, the Expectation step requires the computation of an integral over the random effects since the complete data log-likelihood does not have a closed form. Rizopoulos et al. [2009] used laplace approximations to bypass intractable integrals when the number of random effects increases, but Gaussian quadratures or Monte-Carlo Markov Chain methods can also be used [Wulfsohn and Tsiatis, 1997; Henderson et al., 2000]. The likelihood can also be maximised numerically through Newton-like algorithms [Jacqmin-Gadda et al., 2006], which also involves the computation of an integral over the random effects. However, the convergence rate of the latter algorithm is higher than the rate of the EM algorithm. Thus, Rizopoulos et al. [2009] proposed an optimization procedure combining both EM and quasi-Newton algorithms, to speed up the convergence rate.

Advantages and limits

The specification of the dependence through the random effects allows to better understand the link between the longitudinal marker and the event. However, the link has to be specified *a priori*. It is then possible to test different assumptions regarding the relationship between the marker and the time-to-event, either via likelihood ratio tests when the corresponding models are embedded or via the AIC criterion.

The computation of multiple integrals over the random effects remains the main drawback of joint shared-random-effect models. This issue may be accentuated when considering several markers simultaneously if each marker-specific mixed sub-model involves different vectors of random effects. A solution would be to assume that all the markers

are linked by a same underlying process, modeled by one single mixed sub-model, with possible marker-specific intercepts or effects of covariates.

Besides, in joint shared-random effect models, the random effects account for both the correlation between repeated measures of the marker and the correlation between the two processes. Rizopoulos et al. [2008] proposed an alternative parameterization where the survival model depends on an extra random effect $u_i^{(T)} \sim \mathcal{N}(0, B^{(T)})$, linked to the random effects from the longitudinal model via a copula function.

Finally, the trajectory of the marker is assumed homogeneous, linked continuously to the risk of event. However, the literature has reported that cognitive decline was heterogeneous in the whole population, and also within the population with dementia and among subjects free of dementia. The following approach for joint modeling may be more appropriate to accommodate heterogeneity in cognitive decline.

2.2.2 Joint latent class models

Lin et al. [2002] proposed a joint model capturing the association between the marker and the risk of the event through latent classes. This model assumes an heterogeneous population, composed of G unobserved homogeneous sub-groups represented by G latent classes, with specific marker trajectory and risk of the event. A conditional assumption states that given the latent classes, the longitudinal and survival processes are independent, such that latent classes capture completely their correlation.

Specification of joint latent class models

We define the probability for subject i, $i = 1, ..., N$, to belong to class g, for $g = 1, ..., G$, by a multinomial logistic model:

$$\pi_{ig} = P(c_i = g | X_i^{(P)}) = \frac{\exp(X_i^{(P)\top} \zeta_g)}{\sum_{m=1}^{G} \exp(X_i^{(P)\top} \zeta_m)} \tag{13}$$

with c_i the membership variable, equal to g if subject i belongs to class g. The vector $X_i^{(P)}$ includes time-independent variables and 1 for the intercept. For identifiability, we fix ζ_G to 0.

Besides, the trajectory of the marker is modeled by a linear mixed sub-model, conditionally to the latent class g:

$$Y_{ij} = X_{ij}^\top \beta_g + Z_{ij}^\top u_{ig} + \epsilon_{ij} \tag{14}$$

where β_g are the class-specific regression parameters, $u_{ig} \sim \mathcal{N}(0, B_g)$ are the random effects, with $B_g = \sigma_g^2 B$ the class-specific variance matrix. Here again, $\sigma_G^2 = 1$ for identifiability. The vector X_{ij} contains the covariates of subject i, at time t_{ij}, $j = 1, ..., n_i$, and Z_{ij} is a sub-vector of X_{ij}. The measurement errors $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ are identically independently distributed and independent from the individual random effects.

Besides, the risk of the event given the latent class g is modeled by a proportional hazards model:

$$\alpha_i(t|c_i = g; \gamma_g) = \alpha_{0g}(t) exp(W_i^\top \gamma_g) \tag{15}$$

with a class-specific baseline function $\alpha_{0g}(\cdot)$ and γ_g the vector of class-specific regression parameters associated with covariates W_i.

Estimation

Based on the conditional independence assumption, the joint distribution can be written as follows:

$$f_{Y_i,T_i,\delta_i}(Y_i, T_i, \delta_i; \theta_G) = \sum_{g=1}^{G} f_{Y_i|c_i}(Y_i|c_i = g; \theta_G) P(T_i, \delta_i|c_i = g; \theta_G) \pi_{ig}$$

where $Y_i = (Y_{i1}, ..., Y_{in_i})^\top$, $f_{Y_i|c_i}$ is a multivariate gaussian density function with mean $X_i\beta_g$ and variance matrices $V_{ig} = \sigma_g Z_i B Z_i^\top + \sigma_\epsilon^2 I_{n_i}$, with X_i and Z_i the matrices with row vectors $X_{ij}^\top$ and $Z_{ij}^\top$ respectively. The vector θ_G includes the parameters from the membership probability model (13), parameters from the conditional mixed sub-model (14) and survival sub-model (15). Thus the likelihood has a closed form:

$$L(\theta_G) = \prod_{i=1}^{N} \sum_{g=1}^{G} f_{Y_i|c_i}(Y_i|c_i = g; \theta_G) \alpha_i(T_i|c_i = g; \theta_G)^{\delta_i} S_i(T_i|c_i = g; \theta_G) \pi_{ig}$$

For a given number of classes, the likelihood can be maximised by the EM algorithm. Some authors prefer algorithms based on Newton-Raphson optimization, especially for

their speed of convergence, and for the possibility to define a convergence criterion based on the Hessian matrix [Proust-Lima et al., 2014] in addition to criteria on log-likelihood stability and parameter stability. Besides, the likelihood function has local maxima, since the model is an extension of mixture models [Redner and Walker, 1984]. To counter this problem, it is recommended to run the model from different sets of initial values for a same value of G [Hipp and Bauer, 2006].

The model should also be run with different values of G. Several criteria were proposed to select the optimal number of classes: BIC, AIC, the size of the classes, the quality of the discrimination between classes or the result to the conditional independence test when available. There is no consensus so far on the best criterion, even if the BIC [Schwarz, 1978] is recommended for mixture models [Hawkins et al., 2001; Zhang and Cheng, 2004].

Discrimination of the model

Once the parameters are estimated, the subjects can be allocated to the latent class corresponding to the highest posterior probability $\arg\max_g(\hat{\pi}_{ig}, g = 1, ..., G)$ with:

$$\hat{\pi}_{ig}^{Y,T} = P(c_i = g|Y_i, T_i, \delta_i; \hat{\theta}_G) = \frac{\hat{\pi}_{ig} f(Y_i|c_i = g; \hat{\theta}_G) P(T_i, \delta_i|c_i = g; \hat{\theta}_G)}{\sum_{m=1}^{G} \hat{\pi}_{im} f(Y_i|c_i = m; \hat{\theta}_G) P(T_i, \delta_i|c_i = m; \hat{\theta}_G)}.$$

Based on these posterior probabilities, we can compute the classification matrix, with elements P_{lm} representing the mean probability, among all the subjects allocated to class l, to belong to class m. The ideal discriminatory model would have its classification matrix equal to the identity matrix. The discriminatory ability of the model can also be evaluated based on the number of individuals with their maximal posterior probability higher than a threshold, such as 0.8.

Goodness-of-fit

The goodness-of-fit assessment concerns both the longitudinal and survival predictions. For the longitudinal part, it can be done in different ways, by comparing the marker observations to the longitudinal predictions, either conditional or marginal on the random effects, and either conditional or marginal on the latent classes.

The goodness-of-fit assessment is made from the following comparisons:

	Class-specific	Marginal on class
Conditional on u_{ig}	$\overline{\hat{Y}}^{(SS)}_{.tg} = \frac{\sum_{(i,j)\in N_t} \hat{\pi}_{ig} \hat{Y}^{(SS)}_{ijg}}{\sum_{(i,j)\in N_t} \hat{\pi}_{ig}}$	$\overline{\hat{Y}}^{(SS)}_{.t} = \frac{\sum_{(i,j)\in N_t} \sum_g^G \hat{\pi}_{ig} \hat{Y}^{(SS)}_{ijg}}{N_t}$
(compared to)	$\overline{Y}_{.tg} = \frac{\sum_{(i,j)\in N_t} Y_{ij} \hat{\pi}_{ig}}{\sum_{(i,j)\in N_t} \hat{\pi}_{ig}}$	$\overline{Y}_{.t} = \sum_{(i,j)\in N_t} \frac{Y_{ij}}{N_t}$
Marginal on u_{ig}	$\overline{\hat{Y}}^{(M)}_{.tg} = \frac{\sum_{(i,j)\in N_t} \hat{\pi}_{ig} \hat{Y}^{(M)}_{ijg}}{\sum_{(i,j)\in N_t} \hat{\pi}_{ig}}$	$\overline{\hat{Y}}^{(M)}_{.t} = \frac{\sum_{(i,j)\in N_t} \sum_g^G \hat{\pi}_{ig} \hat{Y}^{(M)}_{ijg}}{N_t}$
(compared to)	$\overline{Y}_{.tg} = \frac{\sum_{(i,j)\in N_t} Y_{ij} \hat{\pi}_{ig}}{\sum_{(i,j)\in N_t} \hat{\pi}_{ig}}$	$\overline{Y}_{.t} = \sum_{(i,j)\in N_t} \frac{Y_{ij}}{N_t}$

where $\hat{Y}^{(SS)}_{ijg} = X_{ij}^\top \hat{\beta}_g + Z_{ij}^\top \hat{u}_{ig}$ are the class-specific predictions conditional on the random effects, computed from $\hat{u}_{ig} = E(u_{ig}|Y_i, c_i = g; \hat{\theta}_G)$ the bayesian estimates of random effects, and $\hat{Y}^{(M)}_{ijg} = X_{ij}^\top \hat{\beta}_g$ are the class-specific predictions, marginal on the random effects. The set N_t includes all pairs (i, j) such that the value Y_{ij} is observed at time $t_{ij} = t$ or during the t^{th} interval if time is discretised.

As an example, when the assessment is made conditionally to the random effects and conditionally to the latent class g, the weighted predicted mean $\overline{\hat{Y}}^{(SS)}_{.tg}$ at time t is compared to the weighted observed mean $\overline{Y}_{.tg}$.

Besides, the goodness-of-fit of survival predictions can be made by comparing the weighted class-specific estimated survival functions $\sum_{i=1}^N \hat{\pi}_{ig} S_i(t|c_i = g; \hat{\theta}_G)$ with the weighted class-specific Kaplan Meier estimates.

Assessment of the conditional independence assumption

The joint latent class model relies on the assumption that the longitudinal and the survival processes are independent conditionally to the latent classes. Lin et al. [2002] assessed this assumption by using a weighted survival analysis adjusting for a function of the marker, stratified on the posterior latent classes. To prescind from the posterior classification, Proust-Lima et al. [2009] proposed a test comparing the means of the standardized conditional residuals of the marker given the event, $Y_i - E(Y_i|T_i, \delta_i)$, between censored and

uncensored subjects. However, the power of this test depends on the dependence structure between the marker and the time-to-event. Then, Jacqmin-Gadda et al. [2010] developed a score test to detect any residual correlation between the marker and the time-to-event. Under the alternative hypothesis $\mathcal{H}_1$, the survival sub-model depends on the n_u-vector of random effects u_{ig} from the longitudinal sub-model, conditionally to the latent class:

$$\alpha_i(t|c_i = g; \gamma_g, \gamma^{(a)}) = \alpha_{0g}(t)exp(W_i^\top \gamma_g + u_{ig}^\top \gamma^{(a)})$$

Defining the score statistic by $U = \sum_{i=1}^{N} \sum_{g=1}^{G} \hat{\pi}_{ig}(\delta_i - A_{ig}(T_i))\hat{u}_{ig}$, with $A_{ig}(t) = \int_0^t \alpha_{ig}(s)ds$ the class-specific cumulative hazard, under the null hypothesis $\mathcal{H}_0 : \gamma^{(a)} = 0$, the test statistic $U^\top \mathrm{Var}(U)^{-1}U$ follows a chi-squared distribution, with n_u degrees of freedom. A simulation study comparing the two last tests showed that the latter was more powerful.

Advantages and limits

The main advantage of the joint latent class approach is its flexibility to accommodate heterogeneity and describe the profiles of the outcome in the population. For instance, in dementia applications, joint latent class models allow to distinguish different homogeneous sub-groups and to describe the corresponding specific profiles of cognitive decline associated with specific profiles of risk of dementia. Another asset is the possibility to test class-specific effects of covariates. Nonetheless, when the model has a low discriminatory ability, interpreting the latent classes becomes difficult.

Secondly, through the use of a categorical latent structure, the model can approximate any kind of association structures between the marker and the risk of the event with no *a priori*, which makes it appropriate for predictions. Proust-Lima et al. [2014] highlighted a substantial gain in predictive ability with a joint latent class model compared to a joint shared-random-effect model, in the first years of prediction, in an application on prostate specific antigene trajectory in prostate cancer. In return, the relationship between the marker and the time-to-event cannot be described explicitly, and this might be a shortcoming in an etiologic study.

The joint latent class approach has an additional strength: the correlation between repeated measures of cognitive decline is differentiated from the correlation between the longitudinal and the survival processes. Indeed, their relationship is entirely captured by

latent classes.

A computational advantage is the closed form of the likelihood, since both mixed models and survival models have analytic likelihoods and the latent shared variable is discrete. Thus, the likelihood can be maximised without approximations, bypassing the computation of integrals over the distribution of the random effects. However, the likelihood is multimodal so the model should be run from different sets of initial values. The model should also be run several times to select the optimal number of latent classes G. As the sub-models are class-specific, the number of parameters may rise dramatically as the number of latent classes increases. Assumptions of proportionality between classes can be made in order to limit the number of parameters.

Finally, the conditional independence assumption is quite strong, but a similar assumption is made for joint shared-random-effect models. The difference is that random effects are continuous whereas latent classes are discrete. The conditional independence assumption in joint latent class models can be assessed by the score test developed by Jacqmin-Gadda et al. [2010].

Extensions of joint latent class models

In the joint latent class model, the longitudinal marker is assumed to be gaussian. When this is not the case, the outcome can be pre-transformed to meet the normality assumption but the standard transformations may not be suitable to any kind of continuous data. Proust-Lima et al. [2009] extended the joint latent class model to handle non-gaussian continuous outcomes, by simultaneously estimating a parametric non-linear transformation ψ. The transformed values of the marker are obtained by

$$\psi(Y_{ij};\eta^{(1)},\eta^{(2)},\eta^{(3)},\eta^{(4)}) = \frac{h(Y_{ij}^r;\eta^{(1)},\eta^{(2)}) - \eta^{(3)}}{\eta^{(4)}} \text{ with } h(Y_{ij}^r;\eta^{(1)},\eta^{(2)}) = \frac{\int_0^{Y_{ij}^r} x^{\eta_1-1}(1-x)^{\eta_2-1}dx}{\int_0^1 u^{\eta_1-1}(1-u)^{\eta_2-1}du}$$

where Y_{ij}^r is the rescaled version of Y_{ij} in [0,1]. This transformation is advantageous as it offers a high flexibility of shapes with only 4 parameters [Proust-Lima et al., 2006]. Proust-Lima et al. [2014] presented other transformations to analyze discrete, asymmetric or bounded data with joint latent class models.

Besides, the joint latent class model was also extended to handle K markers, considered as different correlated measures of an underlying latent process Λ_i:

$$\psi_k(Y_{ijk};\eta_k) = \Lambda_i(t_{ijk}) + X_{ij}^{(k)\top}\beta^{(k)} + \alpha_{ki} + \epsilon_{ijk} \tag{16}$$

for marker $k = 1, ..., K$, with $\alpha_{ki} \sim \mathcal{N}(0, \sigma^2_{\alpha_k})$ and $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2_{\epsilon_k})$, independent from $\epsilon_{ij'k}$ with $j' \neq j$ and from α_{ki}. The model accounts for covariates $X_{ij}^{(k)}$ with possible marker-specific effects $\beta^{(k)}$ called contrasts, and marker-specific inter-subject variability via random effects α_{ki}. This model is detailed in section 3.2.4 in the framework of the extension we propose, to handle competing risks and interval censoring. At last, joint latent class models were extended to competing risks, as detailed in section 2.3.3.

In studies on the pre-diagnosis or general cognitive decline, joint latent class models may be more appropriate than joint shared-random-effect models as they accommodate heterogeneity in the population, but both would account for the correlation between the longitudinal marker and the time-to-event. The next section presents different approaches to consider death when handling longitudinal data, possibly jointly to time-to-event data, correlated with death.

2.3 Dealing with death

In cohort studies among the elderly, the follow-up is often interrupted by death. When analyzing the change over time of a longitudinal marker, death must be differentiated from any other causes of non-response as it is unrealistic to consider data beyond death. However, when dealing with drop-out, the aim is to recover longitudinal response values beyond the time-to-drop-out to describe the trajectory which would have been observed with complete data.

Besides, when studying the risk of dementia, death may induce a dependent censoring as it has common risk factors with dementia. Moreover, death should be considered as a semi-competing risk as subjects living with dementia can die later while dementia cannot occur after death. Finally, when the data are collected intermittently, death may accentuate the interval censoring issue of time-to-dementia. As death have common risk

factors with both the cognitive decline and dementia and as the risk of death increases among cognitively impaired subjects and subjects with dementia, selection by death must be carefully addressed in both longitudinal and time-to-event analyses as well as in joint analyses of cognitive decline and dementia.

2.3.1 Death in longitudinal studies

Within the framework of longitudinal data truncated by death, different models have been suggested for regression analysis. They can be classified into three categories: unconditional, partly conditional and fully conditional approaches. Kurland et al. [2009] linked each of these with the corresponding research aims.

Unconditional models target the evolution of the marker among the immortal cohort, i.e. as if subjects were not susceptible to die. The unconditional expectation can be obtained by imputing missing values beyond death or by using methods that implicitly do the same, such as likelihood-based methods (mixed models and joint models) through the modeling of the intra-subject correlation, or weighted GEE, through the modeling of the death process. When death is assumed to be the only cause of attrition, Kurland et al. [2009] suggested that these models are appropriate only if death is independent from the longitudinal marker or if death does not shorten the follow-up. Otherwise, unconditional models estimate the averaged conditional expectation over the distribution of the time-to-event: $E(Y_{ij}) = E(Y_{ij}|T_i > t_{ij})P(T_i > t_{ij}) + E(Y_{ij}|T_i \leqslant t_{ij})P(T_i \leqslant t_{ij})$.

Partly conditional models condition on being alive, targeting the expectation among the dynamical cohort of survivors at each time-point [Dufouil et al., 2004]. In order not to implicitly impute data after death, data collected for a same individual should be considered as independent. When death is the only cause of non-response, Kurland and Heagerty [2005] showed that estimates provided by GEE, with an independent working covariance matrix, were appropriate for the partly conditional expectation whereas estimates from likelihood-based methods or GEE, with a non-independent correlation matrix, suited better the unconditional expectation. Several authors [Brayne et al., 1999; Dufouil et al., 2004] concur that partly conditional estimands are the most relevant ones in analyses of longitudinal data when the follow-up is interrupted by death.

At last, fully conditional models condition on time-to-death. This category includes pattern-mixture models [Little, 1995] and principal stratification models [Frangakis and Rubin, 2002]. Pattern-mixture models estimate profiles of evolution for different strata, defined by survival time. Such models are useful for example to demonstrate the terminal decline before death, presented in the next sub-section. Principal stratification models also define strata, based on observed and counterfactual survival status.

In studies of longitudinal data, death must be considered as a methodological challenge as it leads to a selection in the population, but it is obviously also of clinical importance. This may be interesting to characterize the cognitive decline preceding death to better highlight the cognitive decline preceding dementia.

2.3.2 Terminal decline

The terminal decline refers to an acceleration of cognitive decline before death, for which evidence is well documented in the literature [Kleemeier, 1962; Siegler, 1975; Wilson et al., 2003]. The normal ageing process involves a cognitive alteration, but sharper declines were observed in late life among deceased subjects as compared to those observed among long-term survivors [Colsher and Wallace, 1991]. This pre-death decline may not be completely related to age but also to an alteration of physical and health states. Kleemeier [1962] formulated the hypothesis that risk factors of death could be the cause of this pre-death cognitive decline and that the onset of this phase may be predictable several years before.

Several studies then aimed at testing the hypothesis of an acceleration of the rate of cognitive decline in the last years of life, as well as identifying the time of onset of this acceleration. Wilson et al. [2003] used change-point random effect models and observed a terminal decline starting 3 to 4 years prior to death, in the general population, with a high variability in the rate of cognitive change. In contrast, survivors experienced a small rate of cognitive decline, and an increase in some cognitive abilities was even reported. Thorvaldsson et al. [2008] also used change-point mixed models and noticed a substantial acceleration in cognitive decline from 6.6 to 14.8 years prior to death among individuals

without dementia, depending on the cognitive ability involved, also with a high variability in the rate of cognitive decline. Kurland et al. [2009] presented a terminal decline model, with the time scale going backward from death, thus falling within the fully conditional approach. However, this leads to a selection issue as the analysis is restricted to subjects who died before the end of the study. At last, the joint modeling of the longitudinal marker and the time-to-death bypasses this selection bias, but provides an indirect estimation of the terminal decline through $E(Y|T = t; \hat{\theta})$, computed *a posteriori*.

It is of interest to differentiate the decline before death from the cognitive decline before dementia. So far, the studies either considered the population without dementia [Thorvaldsson et al., 2008] or quantified the terminal decline among the general population, without distinguishing the cognitive decline attributable to dementia and to death without dementia [Wilson et al., 2003, 2007].

The sections 2.3.1 and 2.3.2 explicited the impact of death on longitudinal data analyses and presented approaches for tackling this issue. However, death also censors time to dementia onset in a likely informative way. Time-to-event analyses should thus also account for death.

2.3.3 Death as a competing risk of dementia

Putter et al. [2007] reviewed the methods proposed for handling several correlated causes of failure. The standard competing risk model, applied to dementia and death without dementia, includes three states:

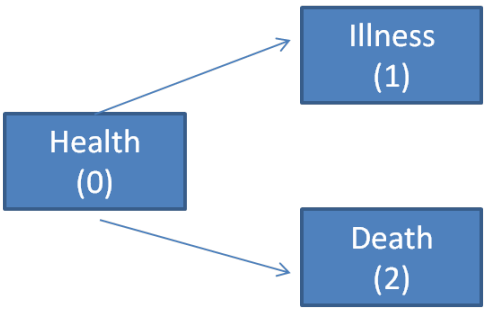


Figure 2.1: Competing risk Model.

The two transition intensities from the Health state to Death or Illness are defined by

cause-specific proportional hazards models:

$$\alpha_{0li}(t) = \lim_{\Delta t \to 0} \frac{P[t < T_i \leqslant t + \Delta t, \delta_i^{(l)} = 1 | T_i \geqslant t]}{\Delta t} = \alpha_{0l}^{(0)}(t) exp(X_{0li}^{\top} \gamma_{0l}) \text{ with } l \in \{1, 2\}$$

with $\alpha_{0l}^{(0)}$ the cause-specific baseline hazard for the transition from the Health state to state l (1 for Illness, 2 for Death), γ_{0l} the cause-specific regression parameters associated with covariates X_{0li} and $\delta_i^{(1)}$, $\delta_i^{(2)}$ the dementia and death indicators respectively with $\delta_i^{(1)} = \delta_i^{(2)} = 0$ if the subject is still alive without dementia at the end of the follow-up, $\delta_i^{(1)} = 1$ if he/she was diagnosed with dementia and $\delta_i^{(2)} = 1$ if he/she died. The likelihood is written:

$$L(\theta) = \prod_{i=1}^{N} \prod_{l=1}^{L} \alpha_{0li}(T_i; \theta)^{\delta_i^{(l)}} S_i(T_i; \theta)$$

with θ the vector including the regression parameters. The overall survival function is defined by $S_i(t) = exp(-\int_0^t (\alpha_{01i}(s) + \alpha_{02i}(s) ds)$ and stands for the probability to be free of both events, i.e. to be alive without dementia at t. The estimators can be obtained by maximisation of the partial likelihood when the baseline hazard function is not specified.

The cumulative incidence of cause l is defined by:

$$F_{0li}(t) = P[T_i \leqslant t, \delta_i^{(l)} = 1] = \int_0^t S_i(s) \alpha_{0li}(s) ds$$

Then, $F_{0li}(t) + F_{02i}(t) + S_i(t) = 1$ at any time t. However, $F_{0li}(\cdot)$ is not a proper distribution function as $\lim_{t \to \infty} F_{0li}(t) = P(\delta_i^{(l)} = 1) < 1$, explaining the denomination of F_{0li} as 'sub-distribution' function. Cumulative incidences are easier to interpret than hazard functions as they represent a probability. However, as the one-to-one relationship between the cause-specific hazard $\alpha_{0li}(\cdot)$ and the cumulative incidence $F_{0li}(\cdot)$ is lost with the cause-specific multi-state model defined above, $F_{0li}(\cdot)$ depends on all the cause-specific hazards. Fine and Gray [1999] then proposed to model directly the effects of covariates on cumulative incidences. However, within the framework of competing risk of death, the use of such models may be questionable as they are based on the idea that subjects who experienced the competing event (who died) are still at risk of the event [Andersen and Keiding, 2012].

Competing risks in joint models

In the joint shared-random-effect model, Elashoff et al. [2008] extended the joint shared-random-effect model to L_e competing causes of event, with cause-specific transition intensities depending on the characteristics of the longitudinal trajectory:

$$\alpha_{0li}(t) = \alpha_{0l}^{(0)}(t) exp(X_{0li}^{\top}\gamma_{0l} + h(u_i, t)^{\top}\eta_{0l}) \text{ with } l \in \{1, ..., \mathrm{L}_e\} \quad (17)$$

with $h(\cdot)$ a function of the random effects from the mixed model and $\alpha_{0li}(\cdot)$ the baseline transition intensity of the l^{th} event. Conditionally to the random effects u_i, the marker and the L_e causes of failure are assumed to be independent. The likelihood is maximised via the EM algorithm. Proust-Lima et al. [2016] also extended joint latent class models to multiple competing risks:

$$\alpha_{0li,g}(t) = \alpha_{0l,g}^{(0)}(t) exp(X_{0li,g}^{\top}\gamma_{0l,g}) \text{ with } l \in \{1, ..., \mathrm{L}_e\} \text{ and } g \in \{1, ..., G\}$$

In this model, the marker and the events are assumed to be independent, conditionally to the classes.

Dantan et al. [2011] proposed a joint illness-death shared-random-effect model with a random change-point in the mixed sub-model to account for the non-linear cognitive decline before dementia diagnosis. The change-point defines the beginning of a phase of accelerated cognitive decline preceding dementia. In this model, the risk of death is assumed independent from the previous state (Health, Pre-diagnosis or Illness) given the current marker value. At last, Ferrer et al. [2016] generalized the joint shared-random-effect model to multiple (i.e. more than 2) transitory or absorbing states, with all the transition intensities modeled by proportional hazards models, as functions of the random effects from the longitudinal sub-model.

These joint models handle the competing risk of death and distinguish the impact of risk factors on dementia and death. However, they should not be used when the time-to-event is interval-censored as the status is not known at any time until the time-to-death or to right censoring.

2.4 Methods for interval-censored time-to-event data

In cohort studies on dementia, the time to dementia onset is interval-censored as subjects can be diagnosed with dementia at visit times only. Thus, the time of onset of the disease is comprised in an interval $]J_i; R_i]$, with J_i the last visit before the event

occurs and R_i the first visit after the event [Finkelstein, 1986; Kalbfleisch and Prentice, 2002]. The methods detailed below rely on the assumption that the visit process is non-informative [Self and Grossman, 1986; Zhang et al., 2005].

2.4.1 Single interval-censored event

The standard method consists in imputing the time to illness onset by the middle of the censoring interval [Law and Brookmeyer, 1992; Helmer et al., 2001] or by the time to diagnosis [Al Hazzouri et al., 2011]. However, these approximations may lead to biased estimations [Freitag et al., 2006; Odell et al., 1992] and Kim [2003] showed that standard errors were likely under-estimated, depending on the length (and variability of lengths among individuals) of censoring intervals.

Sun [2007] reviewed the different approaches proposed to handle interval-censored data. Finkelstein [1986] extended the Cox proportional hazards model, considering time as dicrete. The likelihood, written as:

$$L = \prod_{i=1}^{N} [P(T_i > J_i|X_i) - P(T_i > R_i|X_i)]$$

is maximised by the EM algorithm. Huang and Wellner [1997] showed that the maximum likelihood estimators were asymptotically normal and efficient. This model was extended to left-truncated data in continuous time by Alioum and Commenges [1996], accounting for the fact that subjects are not included in the study from birth but along life, provided they are still free of dementia. The likelihood is computed conditionally to the entry time, under the assumption that this entry time is independent from the time-to-disease. Joly et al. [1998] proposed to model the hazard function either parametrically or semi-parametrically via M-splines, for avoiding any *a priori* specification of its form. In the latter case, the likelihood was penalized, based on the norm of its second derivative, in order to provide smooth transition intensities.

Other parametric models have been proposed for handling interval-censored data, such as the proportional odds model [Huang and Rossini, 1997], the logistic model [Sun, 1997], the accelerated failure model [Betensky, 2001] or the additive hazards model [Zeng et al., 2006].

2.4.2 Semi-competing risks with interval censoring

When death is a competing risk, the interval censoring issue gets more critical as death may occur during the interval $]J_i; R_i]$. Thus, the disease status of the subject at time-to-death is unknown.

In the standard cause-specific approach for competing risks, the time to disease onset is censored at time-to-death. However, this approach may over-estimate the risk of dementia, considering subjects at risk of dementia until their death. On the other hand, censoring the time-to-disease at the last visit may under-estimate the risk of dementia since the subjects with dementia have a higher risk to die.

Joly et al. [2002] considered a multi-state model with three states (Health, Illness, Death on Figure 2.2), where the transition intensity from state k to state l is modeled by a proportional hazards model:

$$\alpha_{kli}(t) = \alpha_{kl}^{(0)}(t) exp(X_{kli}^{\top} \gamma_{kl}) \text{ with } (k,l) \in \{(0,1),(0,2),(1,2)\}$$

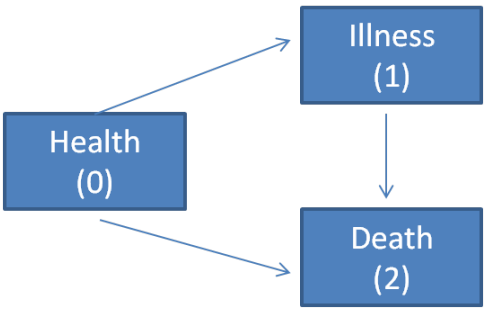


Figure 2.2: Illness-death Model.

The likelihood accounts for interval censoring, by considering a possible unobserved transition to dementia for any subject seen free of dementia at their last visit J_i, with information on vital status collected later, at $T_i > J_i$. This is the case of all subjects who died, as the dementia status is not known at their time-to-death. This would also be the case of a subject who answered at the phone but could not be visited such that his/her vital status is known whereas his/her dementia status is not.

The individual contribution to the log-likelihood can be written as follows:

$$\begin{aligned}\mathcal{L}_i =& log\Big[\delta_i^A \int_{J_i}^{R_i} p_{00i}(0,u)\alpha_{01i}(u)p_{11i}(u,T_i)du\ \alpha_{12i}(T_i)^{\delta_i^D} \\ &+ (1-\delta_i^A)\Big(p_{00i}(0,T_i)\alpha_{02i}(T_i)^{\delta_i^D} + \int_{J_i}^{T_i} p_{00i}(0,u)\alpha_{01i}(u)p_{11i}(u,T_i)du\ \alpha_{12i}(T_i)^{\delta_i^D}\Big)\Big]\end{aligned} \tag{18}$$

with $\delta_i^A = 1$ if the subject is diagnosed with dementia and 0 otherwise, and $\delta_i^D = 1$ if the subject died before the end of the follow-up and 0 otherwise. The probability $p_{lli}(s,t)$ represents the probability that subject i stays in state l between times s and t. Thus, $p_{00i}(0,t)$ is the probability that he/she is alive without dementia at t.

The transition intensities can be modeled by Weibull or M-splines functions. The likelihood, possibly penalized, is maximised via the Levenberg-Marquardt algorithm [Levenberg, 1944; Marquardt, 1963] which combines the Newton-Raphson algorithm and the Gradient descent algorithm. The R SmoothHazard package [Touraine et al., 2016] was made available to fit such model.

Leffondré et al. [2013] compared the illness-death model to the Cox model on simulated interval-censored time-to-event data,within the framework of semi-competing risk of death. They showed that the illness-death model performed better than the Cox model when the exposure was linked to both the event of interest and death. Moreover, the illness-death model does not require to redefine arbitrarily the event of interest (imputing the time at the last visit or the middle of the censoring interval) nor the censoring event (time at the last visit or time-to-death), contrary to the survival model [Boucquemont et al., 2014].

Interval censoring in joint models

In the joint shared-random-effect framework, Gueorguieva et al. [2012] handled competing risks where all the time-to-events were interval-censored. However, the type of failure was known for each subject who experienced an event as they considered strict competing events and not semi-competing ones. In Dantan's joint model, the interval-censored time to dementia onset was imputed by the middle of the censoring interval. With this approximation, interval censoring was no more an issue because under the assumption that the transition intensities to death do not depend on the previous state, conditionally to

the current true value of the marker, the naïve likelihood and the likelihood accounting for interval censoring are equivalent.

In the joint latent class framework, no developments permitted to handle interval censoring so far. In the next section, we present a joint latent class model to handle a longitudinal marker correlated to an event, accounting for interval censoring of the time-to-event when death is a semi-competing event.

3 Joint latent class model for semi-competing interval-censored events

Anaïs Rouanet[1,2,*], Pierre Joly[1,2], Jean-François Dartigues[1,2], Cécile Proust-Lima[1,2] and Hélène Jacqmin-Gadda[1,2]

[1]INSERM, Centre INSERM U1219 - Epidemiologie - Biostatistiques, F-33076 Bordeaux, France

[2]Université de Bordeaux, ISPED, 146 rue Léo Saignat, F-33076 Bordeaux, France

Accepted in *Biometrics*. DOI: 10.1111/biom.12530

SUMMARY. Joint models are used in ageing studies to investigate the association between longitudinal markers and a time-to-event, and have been extended to multiple markers and/or competing risks. The competing risk of death must be considered in the elderly because death and dementia have common risk factors. Moreover, in cohort studies, time-to-dementia is interval-censored because dementia is assessed intermittently. So subjects can develop dementia and die between two visits without being diagnosed. To study pre-dementia cognitive decline, we propose a joint latent class model combining a (possibly multivariate) mixed model and an illness-death model handling both interval censoring (by accounting for a possible unobserved transition to dementia) and semi-competing risks. Parameters are estimated by maximum likelihood handling interval censoring. The correlation between the marker and the times-to-event is captured by latent classes, homogeneous sub-groups with specific risks of death, dementia and profiles of cognitive decline. We propose markovian and semi-markovian versions. Both approaches are compared to a joint latent-class model for competing risks through a simulation study, and applied in a prospective cohort study of cerebral and functional ageing to distinguish different profiles of cognitive decline associated with risks of dementia and death. The comparison highlights that among subjects with dementia, mortality depends more on age than on duration of dementia. This model distinguishes the so-called terminal pre-death decline (among healthy subjects) from the pre-dementia decline.

KEYWORDS. Illness–death; Interval censoring; Joint model; Mixed model; Semi-competing risks.

3.1 Introduction

Joint models are becoming increasingly popular as they allow an analysis of the association between the risk of an event and the change over time of a longitudinal marker [Tsiatis and Davidian, 2004; Rizopoulos, 2012]. In a cognitive ageing study, the link between cognitive decline and dementia needs to be understood to better describe the course of the disease in the pre-diagnostic stage, and to develop prediction tools for the risk of dementia. Moreover, modeling the evolution of cognitive markers without modeling jointly the risk of dementia may lead to biased estimations of the change over time of the marker as the collection of cognitive measures is often stopped after dementia onset (or subjects leave the cohort), inducing non-random missing data. Joint models correct for this bias by accounting for the association between the marker and the time-to-event.

Wulfsohn and Tsiatis [1997] proposed shared-random-effect models where a function of the random effects from the longitudinal model is included in the survival model, thus capturing the correlation between the time-to-event and the marker. The risk of the event is then partly explained by the individual dynamics of the marker trajectory. An alternative is the joint latent class mixed model, developed by Lin et al. [2002] which considers a heterogeneous population, divisible into several homogeneous latent sub-groups, with a specific risk of the event and a specific evolution of the marker.

When studying the risk factors of dementia, it is important to account for the competing risk of death as dementia and death have common risk factors. Elashoff, Li and Li [2008] and Williamson et al. [2008] proposed shared-random-effect models accounting for competing risks and Proust-Lima, Dartigues and Jacqmin-Gadda [2016] developed a joint latent class model for competing risks to study multiple longitudinal markers of different natures. However, studies of the risk of dementia are made more difficult by the interval censoring of time-to-dementia. Indeed, in cohort studies, patients are observed intermittently and the age at dementia onset is not precisely known as dementia can only be diagnosed at clinical follow-up visits. This induces an interval of uncertainty between the last visit where the patient has been seen to be healthy and the visit where a diagnosis was made. More importantly, a patient can develop dementia and die between two visits

without being diagnosed with dementia. Consequently, the risk of dementia may be underestimated when interval censoring is not accounted for. This is the case, for example, when considering only the first observed event in the standard competing risks model. Joly et al. [2002] proposed an illness-death model to fix this issue but it has not yet been implemented in a joint model. The illness-death model is well adapted for modeling semi-competing events such as death and dementia. These events are semi-competing because the risk of dementia is null after death but death can occur after dementia.

To our knowledge, only one joint model combining a multi-state model and a mixed model for a longitudinal marker has previously been proposed. Within the framework of the shared-random-effect approach, Dantan et al. [2011] described a joint model combining a two-phase mixed model with a random change-point and a multi-state model. The underlying clinical idea was an acceleration of the cognitive decline before dementia onset, which was modeled by a second phase with a different slope in the mixed model. In this model, interval censoring was not a critical issue as death depended on the current value of the marker and not on the current state.

In this work, we propose a joint latent class illness-death model for semi-competing interval-censored events and a longitudinal marker. We developed two versions of the model, a markovian and a semi-markovian versions, as well as an extension for the joint analysis of multiple longitudinal markers. In the following section, we detail the model and the estimation procedure that are then evaluated in a simulation study in section 3. In section 4, the model is applied to study cognitive decline before dementia and death using data from the French Paquid cohort, including 3,777 subjects followed over 20 years with regular cognitive evaluation.

3.2 Methods

Notations

Let Y_{ij} denote the score of the psychometric test for subject i $(i = 1, ..., N)$ at time t_{ij} $(j = 1, ..., n_i)$. We denote by T_i^A the unobserved age at dementia onset and T_i^D the age at death. We assume that age at dementia onset is interval-censored while age

at death is only right-censored as exact ages of death are generally collected in cohort studies. Thus, D_i denotes the vector of collected variables for the times-to-events: $D_i = (T_{0i}, L_i, R_i, \delta_i^A, T_i, \delta_i^D)^\top$ where T_{0i} is the age at inclusion, L_i is the age at the last visit where the subject has been seen to be healthy, R_i is the age at the visit of diagnosis if the subject was diagnosed with dementia ($R_i = +\infty$ if not diagnosed), T_i is the age at death or at the end of the follow-up, δ_i^A is the indicator of dementia diagnosis ($\delta_i^A = 1$ if $R_i \leq T_i$ and 0 otherwise) and δ_i^D is the indicator of death ($\delta_i^D = 1$ if $T_i^D = T_i$ and 0 otherwise).

3.2.1 Joint latent class illness-death model

The model relies on the hypothesis that the population is heterogeneous and can be divided into G homogeneous latent classes. Each class has specific transition intensities to dementia and death and a specific marker trajectory, as displayed in Figure 3.1. A central assumption states that the marker and the times-to-events are independent conditionally on the classes and covariates.

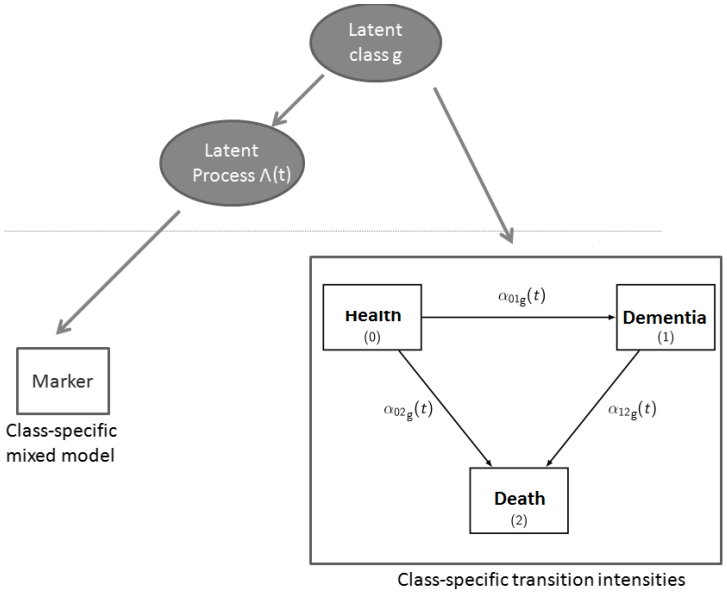


Figure 3.1: Joint latent class illness-death model: the latent classes correspond to homogeneous sub-groups of subjects with a specific marker trajectory and specific transition intensities to dementia and death. Ovales indicate latent quantities and squares indicate observed quantities.

We first describe the probability π_{ig}, for subject i, of belonging to class g ($g = 1, ..., G$)

with a multinomial logistic model:

$$\pi_{ig} = P(c_i = g) = \frac{\exp(X_i^{(\mathrm{P})\top}\ \zeta_g)}{\sum_{m=1}^{G} \exp(X_i^{(\mathrm{P})\top}\ \zeta_m)} \tag{19}$$

The latent class membership variable c_i is $c_i = g$ if subject i belongs to class g and $X_i^{(\mathrm{P})}$, where P stands for probability, is a vector including covariates at baseline and the intercept. We choose class G as the reference class so that $\zeta_G = 0$ to ensure identifiability. The vector ζ_g contains the log odd ratios of the explanatory variables for belonging to class g versus class G.

We denote by $\Lambda(\cdot)$ the latent process which stands for the true cognitive level. The conditional distribution of $\Lambda(t)$ given the latent class is defined by a mixed model, without residual error, with class-specific parameters:

$$\Lambda_i(t_{ij}) = f_1(X_{ij};\ \beta_g) + f_2(Z_{ij};\ \beta_g)\ u_{ig}, \tag{20}$$

where X_{ij} is a vector of covariates for subject i at time t_{ij}, β_g is the vector of class-specific regression parameters and Z_{ij} is a sub-vector of X_{ij}. Given class g, the vector of random effects u_{ig} is $\mathcal{N}(0, \sigma_g^2 B)$ with $\sigma_G^2 = 1$ and B a positive definite matrix. We denote by U the Cholesky transformation of matrix B, which is a lower triangular matrix satisfying $UU^\top = B$. The functions f_1 and f_2 can include nonlinear functions of time, covariates and regression parameters. This formulation especially encompasses both the linear mixed model with polynomial time-trend and models with a class-specific change-point, both linear in the random effects. Note that $f_1(X_{ij};\ \beta_g)$ is a scalar while $f_2(Z_{ij};\ \beta_g)$ is a row n_u-vector with n_u the number of random effects. The marker Y_{ij} is considered as a measure with error of the latent process $\Lambda_i(t_{ij})$ at time t_{ij}:

$$Y_{ij} = \Lambda_i(t_{ij}) + \epsilon_{ij} \text{ with } \epsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2). \tag{21}$$

Simultaneously, the transition intensities to dementia and death are modeled using an illness-death model with three class-specific transition intensities (Figure 3.1). We propose both a markovian and a semi-markovian models.

3.2.2 Markovian model

Given the latent class g, the transition intensity from state k to state l depends on age t and it is modeled by a proportional hazards model with class-specific parameters:

$$\alpha_{klig}(t) = \alpha^0_{klg}(t)\ \exp(W^\top_{kli}\ \gamma_{klg}), \tag{22}$$

where α^0_{klg} is the baseline transition intensity (modeled by Weibull or M-splines functions with equidistant nodes or nodes at quantiles in our program), W_{kli} is a vector of time-independent covariates and γ_{klg} are class-specific regression parameters. The cumulative transition intensities are denoted by $A_{klig}(t) = \int_0^t \alpha_{klig}(s)ds$.

3.2.3 Semi-markovian model

Alternatively, the transition intensity to death among subjects with dementia may depend on the time spent in the Dementia state instead of age, leading to a semi-markovian illness-death model:

$$\alpha_{12ig}(t, T^A_i) = \alpha_{12ig}(t - T^A_i) = \alpha^0_{12g}(t - T^A_i)\ \exp(W^\top_{12i}\ \gamma_{12g}), \tag{23}$$

where T^A_i is the age at dementia onset so $(t - T^A_i)$ is the time spent in the Dementia state. The other two transition intensities depend on age only.

3.2.4 Log-likelihood of the markovian model

Let θ_G denote the vector including the regression, variance and baseline transition intensities parameters. The contribution $\mathcal{L}_i$ of any subject i to the global log-likelihood $\mathcal{L}(\theta_G)$ is the weighted sum over the G classes of his/her class-specific contributions. According to the conditional independence assumption, the individual contribution to the log-likelihood given the class is the product of the conditional contributions to the mixed model and to the multi-state model. The model also accounts for delayed entry, the second part of Eq.24 representing the probability of being alive and healthy at entry, which is the condition for inclusion in the sample.

$$\mathcal{L}(\theta_G) = \sum_{i=1}^{N} \mathcal{L}_i = \sum_{i=1}^{N} \log \Big[\sum_{g=1}^{G} \pi_{ig} f(Y_i|c_i = g; \theta_G) P(D_i|c_i = g; \theta_G) \Big] - \sum_{i=1}^{N} \log \Big[\sum_{g=1}^{G} \pi_{ig} \, \mathrm{e}^{-A_{01ig}(T_{0i};\theta_G) - A_{02ig}(T_{0i};\theta_G)} \Big] \quad (24)$$

where $Y_i = (Y_{i1}, ..., Y_{in_i})^\top$ with n_i the number of repeated measures of subject i, $f(Y_i|c_i = g; \theta_G)$ is a multivariate gaussian density with mean $E_{ig} = [f_1(X_{i1}; \beta_g), ..., f_1(X_{in_i}; \beta_g)]^\top$ and variance matrix $V_{ig} = \sigma_g^2 \, F_{2ig} B F_{2ig}^\top + \sigma_e^2 I_{n_i}$ with F_{2ig} the $(n_i \times n_u)$-matrix with lines given by $f_2(Z_{ij}; \beta_g)$. Then, $P(D_i|c_i = g; \theta_G)$ is detailed below for each possible observation pattern for death and dementia illustrated in Figure 3.2.

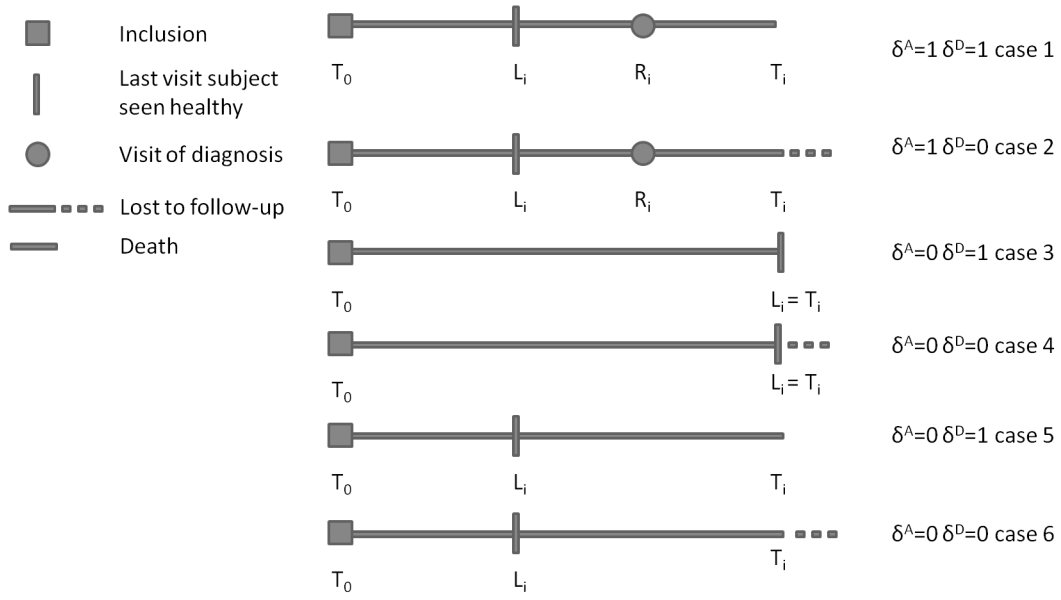


Figure 3.2: Possible observation patterns for dementia and death. To obtain a more flexible program, we also implemented the computation of the likelihood for subjects with an exact date of dementia onset ($L_i = R_i$) although this date was never known in our dataset.

– Subject diagnosed with dementia (cases 1 and 2, Figure 3.2):

$$P_{ig}^{(d)}(T_{0i}, L_i, R_i, 1, T_i, \delta_i^D; \theta_G) = \int_{L_i}^{R_i} \mathrm{e}^{-A_{01ig}(u)} \, \mathrm{e}^{-A_{02ig}(u)} \, \alpha_{01ig}(u) \, \mathrm{e}^{-(A_{12ig}(T_i) - A_{12ig}(u))} \, \alpha_{12ig}(T_i)^{\delta_i^D} \, \mathrm{d}u$$

The subject remained healthy and alive until age u between L_i and R_i, developed dementia at u, remained alive until T_i and possibly died at T_i (if $\delta_i^D = 1$).

– Subject observed healthy at the end of the follow-up (cases 3 and 4, Figure 3.2):

$$P_{ig}^{(h)}(T_{0i}, L_i, R_i, 0, T_i, \delta_i^D; \theta_G) = \mathrm{e}^{-A_{01ig}(T_i)-A_{02ig}(T_i)}\, \alpha_{02ig}(T_i)^{\delta_i^D}$$

The subject remained healthy and alive until T_i and possibly died at T_i. Case 4 corresponds to subjects healthy at the last visit and with no information on vital status after this visit. Case 3 is not observed in the Paquid study because subjects never die the very day of the visit. Consequently, we can never be totally sure that a subject who died was free of dementia. Nevertheless, this case may be observed for other diseases.

– Subject with unknown dementia status at the end of the follow-up (cases 5 and 6, Fig.3.2):

$$\begin{aligned} P_{ig}^{(u)}(T_{0i}, L_i, R_i, 0, T_i, \delta_i^D; \theta_G) &= \mathrm{e}^{-A_{01ig}(T_i)-A_{02ig}(T_i)}\, \alpha_{02ig}(T_i)^{\delta_i^D} \\ &+ \int_{L_i}^{T_i} \mathrm{e}^{-A_{01ig}(u)-A_{02ig}(u)}\, \alpha_{01ig}(u)\, \mathrm{e}^{-(A_{12ig}(T_i)-A_{12ig}(u))}\, \alpha_{12ig}(T_i)^{\delta_i^D}\, \mathrm{d}u \end{aligned}$$

The term $P_{ig}^{(u)}$ accounts for the two possible trajectories: either the subject remained healthy until the end of the follow-up T_i (and possibly died), or he/she developed dementia between the last visit L_i, where he/she was observed healthy, and T_i. If so, the subject remained healthy and alive until age u between L_i and T_i, developed dementia at u, remained alive until T_i and was lost to follow-up or died at T_i.

The likelihood of the semi-markovian model is detailed in Web Appendix A.

Optimisation algorithm

The maximum likelihood estimators are obtained for a fixed number of classes G by a Newton-Raphson-like algorithm [Marquardt, 1963]. If necessary, at each iteration p, the Hessian matrix $H^{(p)}$ is diagonal-inflated to obtain a positive definite matrix $H^{*(p)}$. The vector of parameters is then updated by $\theta_G^{(p+1)} = \theta_G^{(p)} - \kappa\ \mathrm{U}(\theta_G^{(p)})^\top [H^{*(p)}]^{-1} \mathrm{U}(\theta_G^{(p)})$ with $\mathrm{U}(\theta_G^{(p)})$ the gradient at iteration p and κ the improvement control parameter, optimized using a line search strategy. The convergence criteria are reached when the squared euclidean distance between the estimates of two consecutive iterations, the absolute change in the log-likelihood as well as the most stringent criterion $\frac{\mathrm{U}(\theta_G^{(p)})^\top [H^{(p)}]^{-1} \mathrm{U}(\theta_G^{(p)})}{n_{param}}$ are less than 10^{-3}, 10^{-3} and 10^{-2} respectively, with n_{param} the total number of parameters. The variances of the estimates are obtained with the inverse of $H^{(p)}$. For each value of G, the estimation process is repeated with different initial values to ensure convergence. Finally,

the number of classes G is chosen by minimising the Bayesian Information Criterion (BIC) [Schwartz, 1978].

Extension to multiple non-gaussian markers

The model can be extended to the analysis of K non-gaussian markers as in Proust-Lima et al. [2009]. In psychometrics, cognitive tests are frequently considered as quantitative variables, with asymmetric distributions and ceiling or floor effects. Moreover, as all these tests measure cognition, they may be considered as measures with error of a common latent process that stands for the true latent cognitive level underlying the various tests. A parametric monotonic function $\Psi_k(\cdot, \eta_k)$, such as a Beta cumulative distribution function or a linear combination of spline functions, with η_k a vector of marker-specific parameters, can then be used to model the link between the observed markers Y_{ijk} and the latent process $\Lambda(t_{ijk})$:

$$\Psi_k(Y_{ijk};\eta_k) = \Lambda_i(t_{ijk}) + X_{ij}^{(k)\top}\,\beta^{(k)} + \alpha_{ki} + \epsilon_{ijk} \text{ with } \alpha_{ki} \sim \mathcal{N}(0,\sigma^2_{\alpha_k}), \epsilon_{ijk} \sim \mathcal{N}(0,\sigma^2_{e_k}) \quad (25)$$

The random intercept α_{ki} captures the marker-specific inter-individual variability. Note that for more flexibility, the model (25) may also include marker-specific effects $\beta^{(k)}$ of some covariates $X_{ij}^{(k)}$. Consequently, we can define transformed scores $\tilde{Y}_{ijk}$ for test k and subject i at time t_{ijk}, for $i = 1, ..., N, j = 1, ..., n_{ik}$ and $k = 1, ..., K$, on the scale of the latent process:

$$\tilde{Y}_{ijk} = \Psi_k(Y_{ijk};\eta_k) \quad (26)$$

The log-likelihood is then defined by:

$$\begin{aligned}
\mathcal{L}(\theta_G) &= \sum_{i=1}^{N} \log\Big(\sum_{g=1}^{G} \pi_{ig}\Phi_g(\widetilde{Y}_i|c_i=g;\theta_G)\ \Big[\prod_{k=1}^{K}\prod_{j=1}^{n_{ik}} J(\Psi_k(Y_{ijk};\eta_k))\Big]\ P(D_i|c_i=g;\theta_G)\Big) \\
&\quad - \sum_{i=1}^{N} \log\Big[\sum_{g=1}^{G} \pi_{ig}\,\mathrm{e}^{-A_{01ig}(T_{0i};\theta_G)-A_{02ig}(T_{0i};\theta_G)}\Big] \\
&= \sum_{i=1}^{N} \log\Big[\sum_{g=1}^{G} \pi_{ig}\Phi_g(\widetilde{Y}_i|c_i=g;\theta_G)\ P(D_i|c_i=g;\theta_G)\Big] + \sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{j=1}^{n_{ik}} \log\Big[J(\Psi_k(Y_{ijk};\eta_k))\Big] \\
&\quad - \sum_{i=1}^{N} \log\Big[\sum_{g=1}^{G} \pi_{ig}\,\mathrm{e}^{-A_{01ig}(T_{0i};\theta_G)-A_{02ig}(T_{0i};\theta_G)}\Big],
\end{aligned}$$

where $J(\cdot)$ is the Jacobian, $\Phi_g(\widetilde{Y}_i|c_i = g;\theta_G)$ is the multivariate gaussian density for the vector of transformed scores for subject i with mean $E_{ig} = (E_{i1g}^\top, ..., E_{iKg}^\top)^\top$ and the

elements of E_{ikg} are $E_{ijkg} = f_1(X_{ijk}; \beta_g) + X_{ij}^{(k)\top}\ \beta^{(k)}$, with X_{ijk} the vector of covariate values at time t_{ijk}, and variance matrix:

$$V_{ig} = \begin{pmatrix} F_{2ig}^{(1)} \\ \vdots \\ F_{2ig}^{(K)} \end{pmatrix} \sigma_g^2 B \ \left(F_{2ig}^{(1)\top} \ ... \ F_{2ig}^{(K)\top} \right) + \begin{pmatrix} \sigma_{\alpha_1}^2 J_{n_{i1}} + \sigma_{e_1}^2 I_{n_{i1}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\alpha_K}^2 J_{n_{iK}} + \sigma_{e_K}^2 I_{n_{iK}} \end{pmatrix},$$

where $F_{2ig}^{(k)}$ is the matrix with row vectors $f_2(Z_{ijk}; \beta_g)$, with Z_{ijk} the vector of covariates of Eq.20 at time t_{ijk} and J_n is the square matrix of size n with all elements equal to 1.

3.2.5 Goodness-of-fit

Once the parameters are estimated, we can compute the individual posterior probability of belonging to class g,

$$P(c_i = g|Y_i, D_i; \hat{\theta}_G) = \frac{\hat{\pi}_{ig}\ f(Y_i|c_i = g; \hat{\theta}_G)\ P(D_i|c_i = g; \hat{\theta}_G)}{\sum_{m=1}^{G} \hat{\pi}_{im} f(Y_i|c_i = m; \hat{\theta}_G)\ P(D_i|c_i = m; \hat{\theta}_G)}, \tag{27}$$

and subjects are assigned to the class with the highest probability. First, we propose to assess the goodness-of-fit of the longitudinal predictions, conditional on the classes. To do so, we split the timescale into five-year age groups $[\tau_q, \tau_{q+1}]$. For each class g, we then compare the class-specific predicted mean evolution of the marker, weighted by the posterior class membership probability

$$\hat{\mu}_{gq} = \frac{\sum_{(i,j)|\tau_q < t_{ij} < \tau_{q+1}} E(Y_{ij}|c_i = g; \hat{\theta}_G)\ P(c_i = g|Y_i, D_i; \hat{\theta}_G)}{\sum_{(i,j)|\tau_q < t_{ij} < \tau_{q+1}}\ P(c_i = g|Y_i, D_i; \hat{\theta}_G)},$$

to the observed mean evolution weighted by the same posterior probability:

$$\hat{\mu}_{gq}^{(o)} = \frac{\sum_{(i,j)|\tau_q < t_{ij} < \tau_{q+1}} Y_{ij}\ P(c_i = g|Y_i, D_i; \hat{\theta}_G)}{\sum_{(i,j)|\tau_q < t_{ij} < \tau_{q+1}}\ P(c_i = g|Y_i, D_i; \hat{\theta}_G)}$$

The assessment can also be done conditionally on the random effects, comparing $\hat{\mu}_{gq}^{(o)}$ to

$$\hat{\mu}_{gq}^{(u)} = \frac{\sum_{(i,j)|\tau_q < t_{ij} < \tau_{q+1}} E(Y_{ij}|c_i = g, \hat{u}_{ig}; \hat{\theta}_G)\ P(c_i = g|Y_i, D_i; \hat{\theta}_G)}{\sum_{(i,j)|\tau_q < t_{ij} < \tau_{q+1}}\ P(c_i = g|Y_i, D_i; \hat{\theta}_G)}$$

with $\hat{u}_{ig} = E(u_i|Y_i, c_i = g; \hat{\theta}_G)$ the Bayesian estimates of the random effects given class g.

Secondly, to evaluate the goodness-of-fit of the parametric illness-death model, we compare the predicted class-specific cumulative incidences of the three transitions to the class-specific predictions obtained by a semi-parametric illness-death model [Touraine, Helmer and Joly, 2013], both marginal on the covariates. Each transition intensity is modeled by a proportional hazards model with baseline transition intensities modeled by M-splines and estimated by penalized likelihood. The contribution to the likelihood of any subject i is weighted by the individual posterior probability $P(c_i = g|Y_i, D_i; \hat{\theta}_G)$ obtained by the joint latent class illness-death model. Note that the compared cumulative incidences for transitions 0-1 and 0-2 are computed given that the subject is alive and healthy at age 65, and the cumulative incidence for transition 1-2 is estimated given that the subject developed dementia at 65 and is alive at age 65, as follows:

$$F_{0lg}(t) = \frac{\int_{65}^{t} \mathrm{e}^{-A_{01g}(u)-A_{02g}(u)}\, \alpha_{0lg}(u)du}{\mathrm{e}^{-A_{01g}(65)-A_{02g}(65)}}, F_{12g}(t) = \frac{\int_{65}^{t} \mathrm{e}^{-A_{12g}(u)}\, \alpha_{12g}(u)du}{\mathrm{e}^{-A_{12g}(65)}} \quad l = 1, 2$$

3.3 Simulations

Design

We carried out simulations in order to evaluate the estimation procedure and compare the estimations with those obtained using a joint latent class model for competing events that does not account for interval censoring.

Data were generated with a model with 2 latent classes and membership probability $\pi_1 = 0.5$ ($\zeta_1 = 0$) for the first class. Of note, π_1 represents the probability of belonging to the first class in the general population. As we only include in the cohort subjects who were healthy and alive at the first visit, we introduce a selection bias. The proportion of each class in the selected sample may then be different. For each subject, age at entry is generated from a uniform distribution on [65,85]. Age at dementia onset and age at death (for healthy subjects and subjects with dementia) are generated from Weibull distributions with class-specific parameters (shape parameter $\lambda_{klg}^{(1)}$ and scale parameter $\lambda_{klg}^{(2)}$ for transition intensity from state k to state l in class g). The transition intensities account for a common effect of a binary covariate X, also generated from a binomial

distribution with parameter 0.5:

$$\alpha_{klig}(t) = \lambda_{klg}^{(1)} \; \lambda_{klg}^{(2)} {}^{\lambda_{klg}^{(1)}} \; t^{(\lambda_{klg}^{(1)}-1)} \; \mathrm{e}^{\gamma_{kl} X_i} \text{ for } k = 0, 1; \; l = 1, 2; \; g = 1, 2 \tag{28}$$

The scores of the psychometric test are generated by a linear mixed model including fixed and random effects on the intercept and the slope, with an adjustment for the covariate X, common over the classes:

$$Y_{ij} = \beta_{0g} + \beta_{1g} \; \tilde{t}_{ij} + u_{ig}^{(0)} + u_{ig}^{(1)} \; \tilde{t}_{ij} + \beta_X \; X_i + \epsilon_{ij}, \quad u_{ig} \sim \mathcal{N}(0, \sigma_g^2 B), \epsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2) \tag{29}$$

where $\tilde{t}$ is a scaled transformation of age t : $\tilde{t} = \frac{t-65}{10}$. Two designs of follow-up were generated: the follow-up visits were scheduled either every 2 or 4 years from inclusion to the minimum between the visit following dementia onset, death or the administrative right-censoring which is 20 years after inclusion. For each design, we generated 500 samples of 500 subjects. The simulated parameters were similar to the ones obtained by a joint markovian illness-death model with two latent classes on the Paquid dataset, without any linear transformation Ψ. We define the age at diagnosis as the age at the first visit following dementia onset if the generated age at death is older. Subjects who die before this next visit are considered as censored for dementia at the last visit before dementia onset.

On average, over the 500 samples simulated within the two-year visit interval framework, 23% of the subjects were observed with dementia, 17.7% died after dementia diagnosis and 62% died without dementia diagnosis. An average of 33.8% were allocated to the first class, of which 33.8% were observed with dementia, 65.3% died with no dementia diagnosis and 30.9% died after dementia diagnosis. In the second class, 17.9% were seen with dementia, 60.5% died with no dementia diagnosis and 10.9% died after dementia diagnosis.

We also estimated a joint latent class model for competing risks, which does not account for interval censoring, on the same simulated data. In the standard competing risks framework, the outcome is a couple (T, δ) with T the time to the first event or the censoring time and δ the indicator of the cause of the first event ($\delta = 1$ if observed with dementia, $\delta = 2$ if dead and $\delta = 0$ if censored). The estimation of this model on interval-censored

data requires to impute this couple (T, δ) for some subjects. If the subject is observed with dementia before death, the recorded transition is the Health-Dementia transition at age of diagnosis. If the subject dies before the dementia diagnosis, we consider that the Health-Death transition occurs at age of death and time-to-dementia is censored at age of death.

Results

Table 3.1 displays the results of the simulation study for (a) visits every 2 years, (b) visits every 4 years, for the joint latent class model for interval-censored semi-competing events on the left and for the joint latent class model for standard competing risks on the right.

For the model accounting for interval censoring and visits every two years, the top left part of Table 3.1 shows small biases and good coverage rates of the 95% confidence interval for the 25 parameters except for the two scale parameters of the Dementia-Death transition, $\lambda_{121}^{(2)}$ and $\lambda_{122}^{(2)}$, which have lower coverage rates because their standard errors are underestimated. This may be due to the small number of observed transitions from dementia to death, in these 500-subject samples. Indeed, the simulations made on 1000 subjects and presented in Web Table 3.3 show better coverage rates.

With visits every four years, the model accounting for interval censoring provides unbiased estimators which have higher variances, especially for the illness-death parameters, as the number of unobserved transitions increases when the censoring interval gets bigger. Indeed, 12.8% of the subjects with dementia were not observed within the two-year visit interval while 24.3% were unobserved with a censoring interval of four years.

When comparing these results with the estimators of the model with competing risks, we observe higher biases for the shape parameters for the transition toward dementia in both classes ($\lambda_{011}^{(1)}$ and $\lambda_{012}^{(1)}$), and toward death ($\lambda_{021}^{(1)}$) as well as an underestimation of the standard errors of the four parameters for the transition toward dementia. These trends are more pronounced for the 4-year-visit-interval data (see part (b) in Table 3.1), leading to poor coverage rates that worsen further when the sample size increases (see part (b) Web Table 3.3 for N=1000).

(a) Visits every 2 years

		Joint illness-death model*					Joint competing risks model*					
		β	$\hat{\beta}$	ASE	ESE	Cover Rate		β	$\hat{\beta}$	ASE	ESE	Cover Rate
Class Membership	ζ_1	0.00	0.04	0.2949	0.3208	0.95	ζ_1	0.00	-0.19	0.4384	0.5344	0.91
Baseline transition intensities of events	$\lambda_{011}^{(1)}$	3.20	3.23	0.4392	0.4908	0.93	$\lambda_{011}^{(1)}$	3.20	3.04	0.5571	0.6202	0.94
	$\lambda_{012}^{(1)}$	3.50	3.58	0.3863	0.4333	0.95	$\lambda_{012}^{(1)}$	3.50	3.09	0.3999	0.4444	0.80
	$\lambda_{011}^{(2)}$	0.11	0.11	0.0017	0.0027	0.96	$\lambda_{011}^{(2)}$	0.11	0.11	0.0099	0.0397	0.94
	$\lambda_{012}^{(2)}$	0.10	0.10	0.0009	0.0009	0.92	$\lambda_{012}^{(2)}$	0.10	0.10	0.0354	0.0572	0.97
	$\lambda_{021}^{(1)}$	3.50	3.54	0.3476	0.3680	0.94	$\lambda_{021}^{(1)}$	3.50	3.69	0.3421	0.3477	0.92
	$\lambda_{022}^{(1)}$	3.40	3.44	0.2324	0.2493	0.93	$\lambda_{022}^{(1)}$	3.40	3.42	0.2162	0.2354	0.93
	$\lambda_{021}^{(2)}$	0.11	0.11	0.0008	0.0010	0.93	$\lambda_{021}^{(2)}$	0.11	0.11	0.0008	0.0008	0.85
	$\lambda_{022}^{(2)}$	0.10	0.10	0.0006	0.0006	0.94	$\lambda_{022}^{(2)}$	0.10	0.10	0.0006	0.0006	0.92
	$\lambda_{121}^{(1)}$	2.78	2.85	0.5938	0.6210	0.92						
	$\lambda_{122}^{(1)}$	3.14	3.30	0.6908	0.7144	0.92						
	$\lambda_{121}^{(2)}$	0.12	0.12	0.0138	0.0224	0.84						
	$\lambda_{122}^{(2)}$	0.11	0.11	0.0064	0.0091	0.85						
Event covariates	γ_{01}	0.02	0.03	0.2308	0.2323	0.95	γ_{01}	0.02	-0.08	0.2145	0.2157	0.93
	γ_{02}	0.67	0.69	0.1514	0.1525	0.97	γ_{02}	0.67	0.66	0.1456	0.1471	0.95
	γ_{12}	0.47	0.49	0.2737	0.3104	0.92						
Latent process	β_{01}	30.22	30.27	0.7851	0.8325	0.93	β_{01}	30.22	30.05	0.9237	0.9604	0.94
	β_{02}	32.96	32.98	0.5162	0.5276	0.95	β_{02}	32.96	32.82	0.5106	0.5133	0.94
	β_{11}	-5.76	-5.76	0.5678	0.5921	0.93	β_{11}	-5.76	-6.04	0.6569	0.6606	0.94
	β_{12}	-3.53	-3.51	0.2029	0.2038	0.94	β_{12}	-3.53	-3.53	0.1988	0.1988	0.95
	β_X	0.08	0.03	0.4495	0.4660	0.94	β_X	0.08	0.06	0.4479	0.4698	0.94
Cholesky transformation	U(1,1)	4.93	4.88	0.2924	0.2820	0.95	U(1,1)	4.93	4.85	0.2994	0.2914	0.94
of the B matrix	U(1,2)	-1.15	-1.11	0.2069	0.1992	0.96	U(1,2)	-1.15	-1.10	0.2081	0.1991	0.95
$UU^\top = B$	U(2,2)	1.46	1.42	0.1392	0.1385	0.95	U(2,2)	1.46	1.43	0.1391	0.1389	0.94
Measurement error	σ_e	3.47	3.47	0.0515	0.0534	0.91	σ_e	3.47	3.47	0.0515	0.0535	0.94

*based on 492 samples with convergence criteria fulfilled

*based on 497 samples with convergence criteria fulfilled

(b) Visits every 4 years

		Joint illness-death model*					Joint competing risks model*					
		β	$\hat{\beta}$	ASE	ESE	Cover Rate		β	$\hat{\beta}$	ASE	ESE	Cover Rate
Class Membership	ζ_1	0.00	0.05	0.3563	0.3961	0.95	ζ_1	0.00	-0.33	0.5909	0.6620	0.82
Baseline transition intensities of events	$\lambda_{011}^{(1)}$	3.20	3.24	0.5170	0.5556	0.93	$\lambda_{011}^{(1)}$	3.20	2.85	0.8301	1.1372	0.88
	$\lambda_{012}^{(1)}$	3.50	3.57	0.4232	0.4842	0.93	$\lambda_{012}^{(1)}$	3.50	2.69	0.4083	0.4822	0.48
	$\lambda_{011}^{(2)}$	0.11	0.11	0.0033	0.0053	0.96	$\lambda_{011}^{(2)}$	0.11	0.12	0.0410	0.0839	0.83
	$\lambda_{012}^{(2)}$	0.10	0.10	0.0030	0.0053	0.94	$\lambda_{012}^{(2)}$	0.10	0.10	0.0111	0.0391	0.96
	$\lambda_{021}^{(1)}$	3.50	3.52	0.3799	0.3944	0.94	$\lambda_{021}^{(1)}$	3.50	3.74	0.3412	0.3618	0.89
	$\lambda_{022}^{(1)}$	3.40	3.44	0.2438	0.2649	0.93	$\lambda_{022}^{(1)}$	3.40	3.41	0.2106	0.2300	0.91
	$\lambda_{021}^{(2)}$	0.11	0.11	0.0010	0.0013	0.94	$\lambda_{021}^{(2)}$	0.11	0.11	0.0008	0.0008	0.63
	$\lambda_{022}^{(2)}$	0.10	0.10	0.0007	0.0007	0.94	$\lambda_{022}^{(2)}$	0.10	0.10	0.0006	0.0007	0.82
	$\lambda_{121}^{(1)}$	2.78	2.90	0.6621	0.7284	0.89						
	$\lambda_{122}^{(1)}$	3.14	3.31	0.7530	0.7818	0.92						
	$\lambda_{121}^{(2)}$	0.12	0.12	0.0233	0.0357	0.81						
	$\lambda_{122}^{(2)}$	0.11	0.11	0.0102	0.0256	0.83						
Event covariates	γ_{01}	0.02	0.04	0.2556	0.2680	0.94	γ_{01}	0.02	-0.18	0.2190	0.2282	0.85
	γ_{02}	0.67	0.69	0.1595	0.1639	0.95	γ_{02}	0.67	0.65	0.1445	0.1486	0.94
	γ_{12}	0.47	0.51	0.3073	0.3658	0.91						
Latent process	β_{01}	30.22	30.28	0.7995	0.8250	0.94	β_{01}	30.22	29.44	1.0022	1.0924	0.94
	β_{02}	32.96	32.99	0.5210	0.5334	0.95	β_{02}	32.96	32.68	0.5026	0.5193	0.92

	β_{11}	-5.76	-5.77	0.5758	0.6007	0.94	β_{11}	-5.76	-6.20	0.7162	0.7353	0.93
	β_{12}	-3.53	-3.51	0.2046	0.2077	0.94	β_{12}	-3.53	-3.53	0.1957	0.1976	0.94
	β_X	0.08	0.02	0.4528	0.4766	0.95	β_X	0.08	0.10	0.4454	0.4799	0.92
Cholesky transformation	U(1,1)	4.93	4.88	0.2937	0.2844	0.96	U(1,1)	4.93	4.85	0.3010	0.2963	0.94
of the B matrix	U(1,2)	-1.15	1.11	0.2073	0.1994	0.96	U(1,2)	-1.15	-1.09	0.2082	0.1991	0.95
$UU^\top = B$	U(2,2)	1.46	1.42	0.1408	0.1418	0.94	U(2,2)	1.46	1.44	0.1388	0.1393	0.95
Measurement error	σ_e	3.47	3.47	0.0515	0.0534	0.94	σ_e	3.47	3.47	0.0515	0.0534	0.94
*based on 490 samples with convergence criteria fulfilled							*based on 490 samples with convergence criteria fulfilled					

Table 3.1: Results of the simulation study comparing estimates of the joint latent class markovian illness-death model for interval-censored events and the joint latent class competing risks model. A total of 500 samples of 500 subjects were generated with a joint markovian illness-death model with visits every 2 or 4 years. ASE is the asymptotic standard error, ESE is the empirical standard error and the coverage rate is calculated from the 95% confidence interval.

In addition, we ran a series of simulations to assess the semi-markovian model, with visit intervals of two and four years. The estimates have small biases and good coverage rates in the longitudinal and the illness-death parts (see Web Table 3.4). Additional simulations were carried out with three and six latent classes (see Web Tables 3.5 and 3.7 respectively), also showing good results. To assess the robustness of the model to the violation of the conditional independence assumption, simulations were then performed generating the data with a three-class joint model where the transition intensity to dementia depended on both the latent classes and the individual random (a) intercept or (b) slope. Results, displayed in Web Table 3.6, highlight the robustness of the estimators. At last, we assessed the adequacy of the BIC to choose the optimal number of classes. We generated data from a four-class joint model and the model was estimated successively with 1 to 6 classes. Over 100 replicates, the models with 1, 2, 3, 4, 5 and 6 classes were selected 0, 0, 21, 48, 27 and 4 times respectively. When generating data from a two-class joint model, the models with 1, 2, 3 and 4 classes were selected 0, 92, 4 and 4 times respectively, over 100 replicates. These results are in agreement with Hawkins, Allen and Stromberg [2001] who performed extensive simulations and highlighted that the BIC criterion was better when the number of classes was small.

3.4 Application to the Paquid cohort

The joint latent class illness-death model was applied to a French prospective cohort, the Paquid cohort, to distinguish different profiles of cognitive decline in the elderly associated with transition intensities to dementia and death. We compared markovian and semi-markovian models, in order to determine whether the transition intensity to death among subjects with dementia depended on age or on duration of dementia.

Data

The Paquid cohort [Letenneur et al., 1994] involves 3,777 subjects from two French administrative departments, Dordogne and Gironde. The subjects were older than 65 at entry and they were visited at years 1, 3, 5, 8, 10, 13, 15, 17, 20, 22 and 25 at home to undergo a battery of psychometric tests. The diagnosis of dementia was based on a screening phase according to DSM III R criteria for dementia [American Psychiatric Association, 1987] and a final clinical diagnosis assessed by a neurologist. In this work, we focused on the Isaacs Set Test, scored from 0 to 40, assessing verbal fluency. Subjects had to produce up to 10 words from four different semantic categories within 15 seconds for each category.

We selected subjects who were healthy at inclusion and who completed at least one Isaacs Set Test before their diagnosis for subjects diagnosed with dementia. The main analysis was performed on 3,525 subjects, excluding 102 subjects who were prevalent cases at the first visit and 150 who had completed no tests during the follow-up. The sample under study included 42.2% of men and 34.1% of subjects with a low level of education at the initial visit. A total of 23.8% of subjects were diagnosed with dementia, including 19.8% who died during the follow-up, and 65.1% died before the dementia diagnosis. The mean age at entry into the study was 75 years (sd=6.7 years), the mean age at death was 86.7 years (sd=6.9 years), the mean age at dementia diagnosis was 86.4 years (sd=5.8 years) and the mean number of measurements before dementia diagnosis or end of follow-up was 3.99 (sd=2.9).

3.4.1 Comparison of models

First, we compared models assuming a class-specific quadratic time-trend (Eq.30) and a linear-linear time trend with class specific change-point (Eq.31) estimated on the Paquid cohort. In both models, the sub-model for the longitudinal marker assumed common effects for gender ($Sex = 0$ for men, 1 for women) and educational level ($Educ = 1$ for subjects who obtained their primary school diploma, 0 for others). The quadratic model accounted for three class-specific fixed and random effects while the change-point model assumed one more class-specific fixed effect relative to the change-point time. Given class g, the first model was defined by:

$$\begin{aligned}\Lambda_i(\tilde{t}) =& \beta_{0g} + u_{ig}^{(0)} + [\beta_{1g} + u_{ig}^{(1)}]\ \tilde{t} + [\beta_{2g} + u_{ig}^{(2)}]\ \frac{\tilde{t}^2}{10} + \beta_3\ Educ_i \\ &+ \beta_4\ Educ_i \times \tilde{t} + \beta_5\ Educ_i \times \frac{\tilde{t}^2}{10} + \beta_6\ Sex_i, \qquad (u_{ig}^{(0)}, u_{ig}^{(1)}, u_{ig}^{(2)})^\top \sim \mathcal{N}(0, \sigma_g^2 B)\end{aligned} \tag{30}$$

where $\tilde{t}$ is a scaled transformation of age t: $\tilde{t} = \frac{t-65}{10}$. The interaction between time and gender was not accounted for because it was not significant in previous analyses. The second model was defined by:

$$\Lambda_i(t) = \beta_{0g} + u_{ig}^{(0)} + [\beta_{1g} + \frac{u_{ig}^{(1)}}{2} + \frac{u_{ig}^{(2)}}{2}]\ (\tilde{t} - \tilde{\tau}_g) + [\beta_{2g} - \frac{u_{ig}^{(1)}}{2} + \frac{u_{ig}^{(2)}}{2}](\tilde{t} - \tilde{\tau}_g)\ trn\big(\tilde{t} - \tilde{\tau}_g; \upsilon\big)$$

$$+ \beta_3\ Educ_i + \beta_4\ Educ_i\ (\tilde{t} - \tilde{\tau}_g) + \beta_5\ Educ_i\ (\tilde{t} - \tilde{\tau}_g) trn\big(\tilde{t} - \tilde{\tau}_g; \upsilon\big) + \beta_6\ Sex_i \tag{31}$$

with $\tilde{\tau}_g = \frac{\tau_g - 65}{10}$, $u_{ig} = (u_{ig}^{(0)}, u_{ig}^{(1)}, u_{ig}^{(2)})^\top \sim \mathcal{N}(0, \sigma_g^2 B)$, $u_{ig}^{(1)}$ standing for the first random slope in class g, before the change-point τ_g, and $u_{ig}^{(2)}$ for the second random slope, after τ_g. The function $trn(t, \upsilon) = \frac{1}{t}\sqrt{t^2 + \upsilon}$ (Seber and Wild, 2003) ensures a smooth transition as $\lim_{\upsilon \to 0} trn(t, \upsilon) = sign(t)$. In this application, $\upsilon = 0.01$. The parameter β_{0g} represents the marker value at $t = \tau_g$ in class g, β_{1g} and β_{2g} stand for half the sum of the first and second slopes and half their difference, respectively, as detailed in Dantan et al. [2011].

For both models, a Beta cumulative distribution function was used to link the observed scores to the latent process:

$$\tilde{Y}_{ij} = \Psi(Y_{ij}; \eta^{(1)}, \eta^{(2)}, \eta^{(3)}, \eta^{(4)}) = \frac{Beta\big(Y_{ij}; \eta^{(1)}, \eta^{(2)}\big) - \eta^{(3)}}{\eta^{(4)}} = \Lambda_i(t_{ij}) + \epsilon_{ij}, \tag{32}$$

	BIC			
	Quadratic Markovian	Change-point Markovian	Change-point Semi-markovian	Change-point adjusted Semi-markovian
G=1	106,928	106,901	107,055	106,954
G=2	106,315	106,270	106,356	106,313
G=3	106,120	106,081	106,177	106,144
G=4	106,058	106,005	106,091	106,066
G=5	106,091	106,027	106,107	106,090

Table 3.2: Comparison of BIC of joint quadratic markovian illness-death mixed model and joint change-point markovian/semi-markovian illness-death mixed models, with a total number of classes varying from 1 to 5 (Paquid, N=3,525).

where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$. The models for the transition intensities were proportional hazards models with common effects of the two covariates and class-specific Weibull baseline transition intensities, as functions of age t:

$$\alpha_{klig}(t) = \alpha^0_{klg}(t) \exp\big(\gamma^s_{kl}\ Sex_i + \gamma^e_{kl}\ Educ_i\big),\ k \in [0,1],\ l \in [1,2], \tag{33}$$

The two first columns of Table 3.2 show that the joint change-point model, which has G more parameters (G class-specific change-point times) fits the data better, with smaller BIC values irrespective of the number of classes. Note that the four-class models have the smallest values in both cases. In the following, we thus focus on the joint change-point model.

Secondly, we compared the joint change-point markovian model defined by Eq.31 and Eq.33 and the joint change-point semi-markovian model where the Dementia-Death transition intensity depends on the time since T^A, age at dementia onset:

$$\alpha_{12ig}(t, T^A_i) = \alpha^0_{12g}(t - T^A_i) \exp\big(\gamma^s_{12}\ Sex_i + \gamma^e_{12}\ Educ_i\big), \tag{34}$$

According to Table 3.2, the markovian model fits the data better, irrespective of the number of classes. Thus, the transition intensity to death among subjects with dementia depends more on age than on dementia duration. When adding T^A as an explanatory variable in Eq.34, we obtain a more flexible model where mortality of subjects with

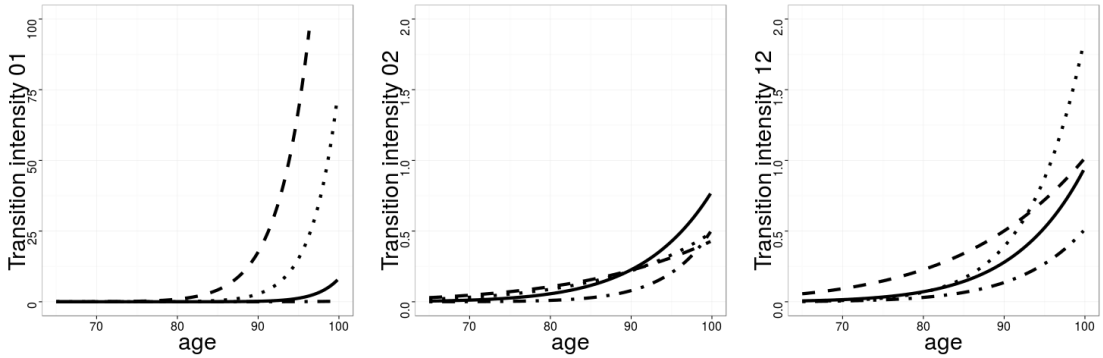
dementia depends on both duration of dementia and age at dementia. Table 3.2 shows that this adjusted model is better than the standard semi-markovian model but does not fit the data as well as the markovian model. We finally selected the joint change-point markovian model with 4 classes that had the best BIC.

3.4.2 Results

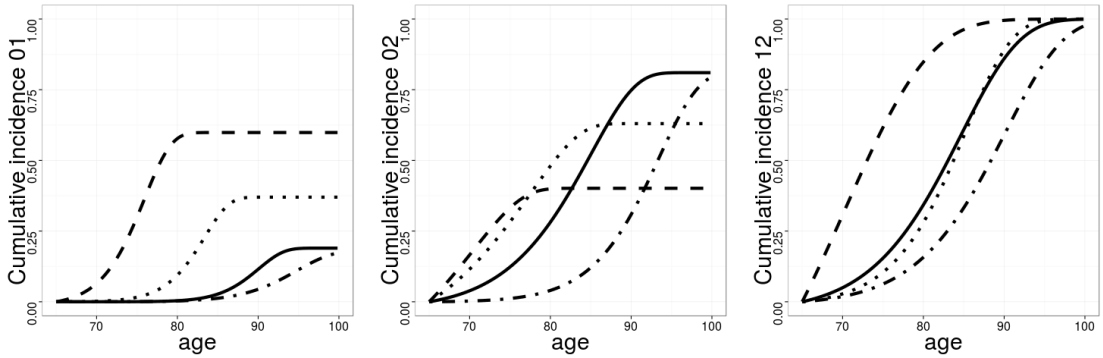
Figure 3.3 displays the three transition intensities (Part A), the three cumulative incidences (Part B) and the estimated mean Isaacs trajectories (Part C) for men with a low level of education (with no primary school diploma), in each class.

The first class, including 11.12% of the population, has the highest transition intensities to dementia and death for subjects with dementia until advanced ages and the fastest and deepest decline of Isaacs scores. As shown in Figure 3.3B, most of these subjects developed dementia or died before age 80. The second class, corresponding to 32.91% of the population, also has high transition intensities to dementia and death among subjects with dementia and a steep cognitive decline, but these occur at later ages than in the first class. The cumulative incidences show that 63% of this group died without dementia before age 85 and 37% developed dementia before age 90. The third class includes 9.73% of the sample and has the lowest transition intensities to dementia and death until advanced ages, as well as the slightest mean decline of the Isaacs Set Test scores (Figure 3.3C). The cumulative incidences show that most dementia and death occurrences arose after age 85 in this class (Figure 3.3B). Finally, the fourth class includes 46.24% of the population and is quite similar to the third one, with a more pronounced cognitive decline and higher transition intensities to death among both healthy subjects and subjects with dementia. About 80% died without dementia and 20% developed dementia before age 95 in this class whereas these figures are only reached 5 to 10 years later in the third class.

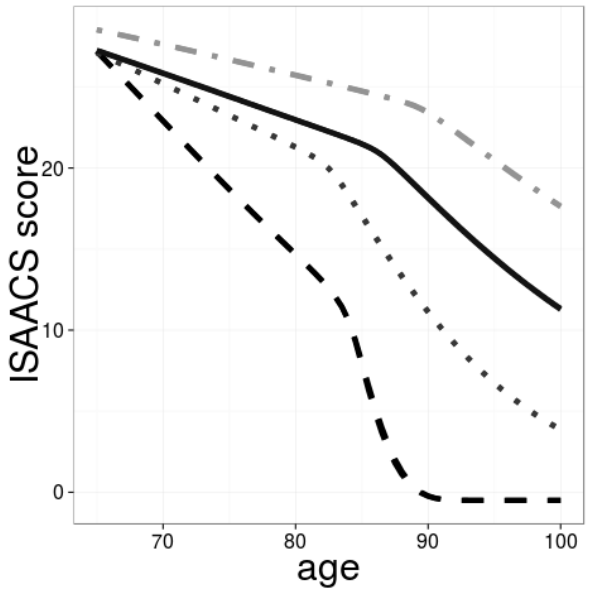
Estimates of the main parameters of the selected model are presented in Web Table 3.8. High education is associated with lower transition intensities to dementia and death among healthy subjects but not with the transition intensity to death among subjects with dementia, after accounting for the heterogeneity of the transition intensities due to the classes. Gender is not associated with the transition intensity to dementia but women,



A) Class-specific transition intensities of the illness-death model for men with a low level of education.



B) Class-specific cumulative incidences of the illness-death model for men with a low level of education.



C) Class-specific Isaacs trajectories for men with a low level of education.

Figure 3.3: Class-specific estimated transition intensities, cumulative incidences and mean longitudinal trajectories of the joint change-point latent class illness-death mixed model for each class (class 1: dashed line, class 2: dotted line, class 3: dotdashed line, class 4: solid line).

with or without dementia, have a lower transition intensity to death. A secondary analysis accounting for Apolipoprotein E4 (ApoE4), which is a known risk factor of dementia, is presented in Web Appendix B on a sub-sample of 619 subjects with ApoE4 measurement. We selected the two latent class model, based on the BIC criterion. ApoE4 appears associated with a higher transition intensity to dementia but not with a steeper decline of Isaacs score within the classes.

3.4.3 Goodness-of-fit

As loss of follow-up may be linked to a change in the cognitive test [Jacqmin-Gadda et al., 1997], missing data are not missing completely at random; so we assessed the goodness-of-fit of the main model estimated on 3525 subjects, conditionally on the random effects. Web Figure 3.5 displays the predicted weighted mean of Isaacs scores given the random effects and the classes, and the weighted mean of the observed scores for each class. The predicted mean is close to the observed mean and is within its 95% confidence interval which increases over time as there is less and less data. The estimated class-specific cumulative incidences, marginal on the covariates, on part B of Web Figure 3.5, are compared to the estimations, also marginal on the covariates, obtained by a semi-parametric illness-death model with baseline transition intensities estimated by penalized likelihood on a basis of M-splines. Here again, the estimations are close and the graphs show that the model fits the data well.

3.4.4 Posterior classification

Considering each posterior class, we can compute the mean probability of belonging to each of the four classes in order to quantify the discriminatory ability of the model (see Web Table 3.9). For each class, the mean probability of belonging to the allocated class is between 61% and 79%, which means that the discrimination between classes is correct but not very good. The description of the posterior classes according to the covariates is given in Web Table 3.10.

3.4.5 Estimated trajectories given the age at dementia and/or death

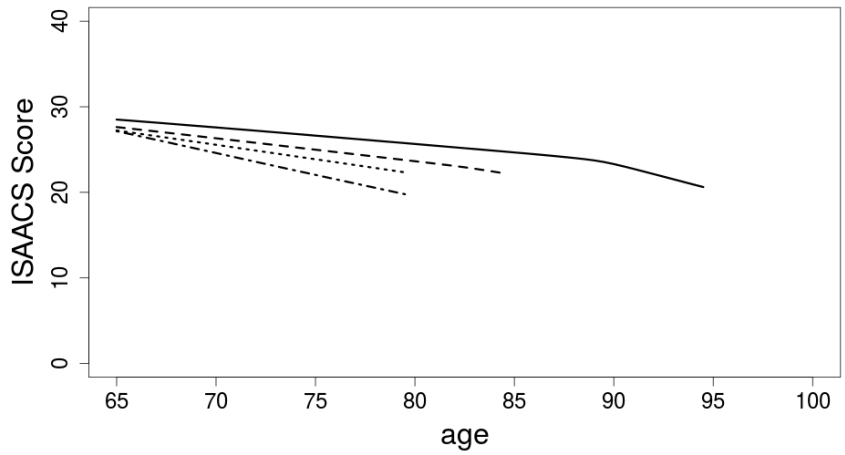
It is of interest to estimate the typical cognitive decline of subjects who developed dementia or were deceased at a given age, as well as the evolution of subjects who were alive and healthy at an advanced age. Thus, we computed the mean trajectories of the Isaacs scores for a man with a low or high level of education for 4 different cases: alive and healthy at age 95 or at 85, dead without dementia at age 80 and a man with dementia onset at age 80. The expectation for the first case is given by

$$E(Y(t)|T_i^A > 95, T_i^D > 95; \hat{\theta}_G) = \sum_{g=1}^{G} E(Y(t)|c_i = g; \hat{\theta}_G) P(c_i = g|T_i^A > 95, T_i^D > 95; \hat{\theta}_G),$$

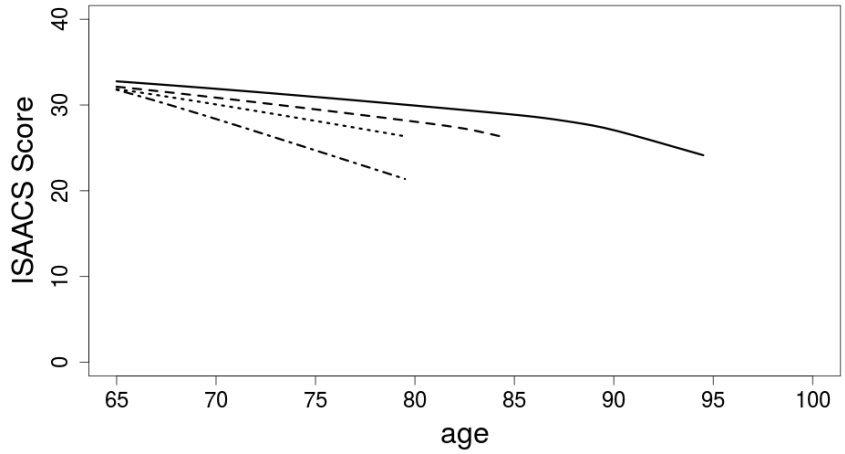
and the other cases are computed in the same way. As expected, we observe in Figure 3.4 that a man alive and healthy at 95 has the smallest cognitive decline (solid lines of parts A and B). Nevertheless, these estimates show a slight decline of the Isaacs scores in older ages among healthy subjects, probably due to a slowing of information processing with age. When comparing the decline of subjects alive and free of dementia at 85 and the decline of subjects who die at age 80 without dementia, we can highlight the so-called terminal decline before death. In the same way, the trajectory of a man who develops dementia at 80 (dotdashed curves) highlights the decline before dementia. The decline before death is more marked among subjects with low education and rather close to the decline before dementia. Moreover, pre-dementia decline is steeper for highly educated subjects compared to subjects with lower education. This is in agreement with the cognitive reserve hypothesis [Stern et. al, 1999]. The 'terminal decline' has been described by other authors [Wilson et al., 2003]. However, it had not been distinguished from the decline toward dementia. As expected, the pre-dementia decline is steeper.

3.5 Discussion

We proposed a joint latent class illness-death model for semi-competing interval-censored events and longitudinal data. Joint models have previously been developed to capture the correlation between a longitudinal marker and competing risks but no previous model has accounted for interval censoring [Elashoff, Li and Li, 2008; Williamson



A) Predicted Isaacs trajectories for a man with a low level of education.



B) Predicted Isaacs trajectories for a man with a high level of education.

Figure 3.4: Predicted Isaacs trajectories for a man with low (A) or high (B) educational level, alive and healthy at age 95 (solid line) or 85 (dashed line), a man who dies at age 80 without dementia (dotted line) and a man who develops dementia at age 80 (dotdashed line). The trajectories are plotted from 65 years old until the age at dementia diagnosis, loss of follow-up or death.

et al., 2008; Proust-Lima et al., 2016]. Our simulations highlighted biased estimates of the joint model for standard competing risks. By dealing with interval censoring, the proposed method corrects for this bias and highlights different profiles of cognitive decline associated with different transition intensities to death and dementia. Subsequently, the mean trajectories of cognitive decline for subjects who developed dementia or were deceased at a given age can be estimated. This makes it possible to distinguish the cognitive decline of the healthy elderly from the cognitive decline before death without dementia and the cognitive decline in the pre-dementia phase that appears to be the steepest.

By jointly modeling the transition intensities to death and dementia, this model accounts for informative drop-outs due to dementia and death. Thus, estimates are robust under the weaker assumption that missing data for other causes (especially intermittent missing data) are missing at random, i.e. they are not linked to the missing values of the marker, conditional on the observations.

In this model, we assumed that any subject was at risk to develop dementia. By setting the parameters for the transition intensities to dementia and from dementia to death to fixed values ($\lambda^{(1)}_{01g} = 1$, $\lambda^{(2)}_{01g} = 0$, $\lambda^{(1)}_{12g} = 1$, $\lambda^{(2)}_{12g} = 0$) for one latent class, we obtained a "cure" joint model assuming that a part of the population was not susceptible for dementia. In the Paquid application, this constraint introduced in the four-class model lead to a worse fit.

Contrary to Dantan et al. [2011], we chose a latent class approach to account for the heterogeneity of cognitive aging. Compared to joint models with shared-random effects, joint latent class models allow a more flexible modeling of the link between dementia, death and cognition and of the distributions of each outcome as they are mixtures of distributions (see Proust-Lima et al. [2014] for a review of the joint latent class models). Then, the model presented distinguishes death intensity for healthy subjects and subjects with dementia, with either a markovian or semi-markovian assumption. This makes it possible to disentangle cognitive decline before dementia from cognitive decline before death without dementia but requires careful handling of interval censoring. In Dantan et al. [2011], the death intensity was assumed to be independent of the dementia status given the current cognitive level. With this assumption, it was shown that the likelihood for interval-censored data was identical to the one without interval censoring, as long as we imputed the middle of the censoring interval for subjects diagnosed with dementia. Nevertheless, alternative joint models for longitudinal markers and illness-death data could be proposed in the shared-random-effect framework that would require to adapt the present methodology to account for interval censoring. Recently, Gueorguieva, Rosenheck and Lin [2012] developed a shared-random-effect model for longitudinal data and competing risks for an application where the time-to-event can be interval-censored. However, the

type of event (different types of drop-out) is assumed to be known, which is not the case in our context.

For selecting the number of classes in mixture models, several criteria, such as BIC [Bauer and Curran, 2003] and DIC [Celeux et al., 2006], were proposed and previously compared. None of them is perfect. In practice, we recommend to combine clinical and statistical criteria such as BIC, size of the classes, quality of the discrimination between classes or test for conditional independence when available. We are currently working on an extension of such a score test [Jacqmin-Gadda et al., 2010] to the proposed model but interval censoring makes the computation cumbersome.

3.6 References

American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders.* (DSM-III-R. 3rd edition revised). 82

Bauer, D. J. and Curran, P. J. (2003). Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. *Psychological Methods* **8(3)**, 338–363. 91

Celeux, G., Forbes, F., Robert, C. and Titterington, D. (2006). Deviance information criteria for missing data models. *Bayesian Analysis* **1**, 651–706. 91

Dantan, E., Joly, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2011). Joint model with latent state for longitudinal and multistate data. *Biostatistics* **12(4)**, 723–736. 69, 83, 90

Elashoff, R. M., Li, G., and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64(3)**, 762–771. 68, 88

Gueorguieva, R., Rosenheck, R., and Lin, H. (2012). Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *Journal of the Royal Statistical Society A* **175(2)**, 417–433. 90

Hawkins, D. S., Allen, D. M., and Stromberg, A. J. (2001). Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis* **38(1)**, 15–48. 81

Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., and Dartigues, J.-F. (1997). A five year longitudinal study of mini mental state examination in normal aging. *American Journal of Epidemiology* **145(6)**, 498–506. 87

Jacqmin-Gadda, H., Proust-Lima, C., Taylor, J. M. G., and Commenges, D. (2010). Score Test for Conditional Independence Between Longitudinal Outcome and Time to Event Given the Classes in the Joint Latent Class Model. *Biometrics* **66(1)**, 11–19. 91

Joly, P., Commenges, D., Helmer, C., and Letenneur, L. (2002). A penalized likelihood approach for an illness–death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* **3(3)**, 433–443. 69

Leffondre, K., Touraine, C., Helmer, C., and Joly, P. (2013). Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model? *International Journal of Epidemiology* **42(4)**, 1177–1186.

Letenneur, L., Commenges, D., Dartigues, J.-F., and Barberger-Gateau, P. (1994). Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *International Journal of Epidemiology* **23(6)**, 1256–1261. 22, 82

Lin, H., Turnbull, B. W., McCulloch, C. E., and 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(457)**, 53–65. 68

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* **11(2)**, 431–441. 22, 65, 74

Proust-Lima, C., Joly, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2009). Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics & Data Analysis* **53(4)**, 1142–1154. 75

Proust-Lima, C., Séne, M., Taylor, J. M. G., and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review *Statistical Methods in Medicine Research* **23**(1), 74–90. 90

Proust-Lima, C., Dartigues, J.-F. and Jacqmin-Gadda, H. (2016). Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach *Statistics in Medicine* **35**(3), 382–398. 89, 95

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-event Data: With Applications in R*, Vol. 6. Boca Raton: Chapman & Hall. 68

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6(2)**, 461–464. 75

Stern, Y., Albert, S., Tang, M.-X., and Tsai, W.-Y. (1999). Rate of memory decline in AD is related to education and occupation: Cognitive reserve? *Neurology* **53(9)**, 1942–1947. 88

Touraine, C., Helmer, C., and Joly, P. (2013). Predictions in an illness-death model. *Statistical Methods in Medical Research* DOI: 10.1177/0962280213489234. 77

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data : An overview. *Statistica Sinica* **14**, 809–834. 68

Williamson, P. R., Kolamunnage-Dona, R., Philipson, P., and Marson, A. G. (2008). Joint modelling of longitudinal and competing risks data. *Statistics in Medicine* **27(30)**, 6426–6438. 68, 88

Wilson, R. S., Beckett, L. A., Bienias, J. L., Evans, D. A., and Bennett, D. A. (2003). Terminal decline in cognitive function. *Neurology* **60(10)**, 1781–1787. 88

Winnock, M., Letenneur, L., Jacqmin-Gadda, H., Dallongeville, J., Amouyel, P. and Dartigues, J. F (2002). Longitudinal analysis of the effect of apolipoprotein E epsilon4 and education on cognitive performance in elderly subjects: the PAQUID study. *Journal of Neurology, Neurosurgery & Psychiatry* **72(6)**:794–7. 95

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53(1)**, 330–339. 68

3.7 Web Supplementary Materials: "Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia"

Anaïs Rouanet[1,2], Pierre Joly[1,2], Jean-François Dartigues[1,2],
Cécile Proust-Lima[1,2] and Hélène Jacqmin-Gadda[1,2]

[1]INSERM, Centre INSERM U1219 - Epidemiologie - Biostatistiques, F-33076 Bordeaux, France
[2]Université de Bordeaux, ISPED, 146 rue Léo Saignat, F-33076 Bordeaux, France

Web Appendix A: Conditional log-likelihood of a semi-markovian model

In the case of a semi-markovian model, the individual conditional contribution to the log-likelihood of the multi-state model is changed for the observation patterns 1, 2, 5 and 6 as follows:

- Subject diagnosed with dementia (cases 1 and 2, Figure 3.2):

$$P_{ig}^{(d)}(T_{0i}, L_i, R_i, 1, T_i, \delta_i^D; \theta_G) = \int_{L_i}^{R_i} \mathrm{e}^{-A_{01ig}(u)-A_{02ig}(u)}\, \alpha_{01ig}(u)\, \mathrm{e}^{-A_{12ig}(T_i-u)}\, \alpha_{12ig}(T_i-u)^{\delta_i^D}\, \mathrm{d}u$$

- Subject with unknown dementia status at the end of follow-up T_i (cases 5 and 6, Figure 3.2):

$$\begin{aligned} P_{ig}^{(u)}(T_{0i}, L_i, R_i, 0, T_i, \delta_i^D; \theta_G) = {} & \mathrm{e}^{-A_{01ig}(T_i)-A_{02ig}(T_i)}\, \alpha_{02ig}(T_i)^{\delta_i^D} \\ & + \int_{L_i}^{T_i} \mathrm{e}^{-A_{01ig}(u)-A_{02ig}(u)}\, \alpha_{01ig}(u)\, \mathrm{e}^{-A_{12ig}(T_i-u)}\, \alpha_{12ig}(T_i-u)^{\delta_i^D}\, \mathrm{d}u \end{aligned}$$

Web Appendix B: Application of the joint latent class illness-death model on Paquid, adjusting on ApoE4

The model was applied on the sub-sample of subjects with ApoE4 measurement available. The sample under study involved only 619 subjects, who accepted blood testing for ApoE4 measurement, including 42.7% of men and 22.8% of ApoE4 carriers. These subjects were younger (mean age at entry: 73.9 vs 75.3, p-value < 0.001) and more educated (72.5% vs 64.4% with a high educational level, p-value < 0.001) than the sample with no

ApoE4 measurement. A total of 31.3% of subjects were diagnosed with dementia, including 27.3% who died during the follow-up, and 58.8% died before the dementia diagnosis.

The three transition intensities of the illness-death sub-model were adjusted on ApoE4, as well as the change-point mixed sub-model (intercept and slope), with all these regression parameters being common to the classes. The BIC of the models from 1 to 5 classes were respectively 25252, 25161, 25164, 25168 and 25216. Web Table 3.11 presents the estimated parameters for the model with 2 latent classes that had the lowest BIC. Education is associated with a lower transition intensity to dementia but not to death among healthy subjects. Gender is associated with lower transition intensities to death and ApoE4 carriers have a higher intensity transition to dementia.

These results are in line with previous analyses on Paquid data which found that ApoE4 was not associated with cognitive level after adjustment on educational level and found no association with cognitive decline [Winnock et al., 2002] or a non significant trend to faster decline Proust-Lima et al. [2016] depending on the cognitive tests under study.

Since the BIC of the models with two, three and four latent classes estimated on this ApoE sub-sample are very close, we also present the classification coming from the four latent class model for comparison with the analysis of the whole Paquid dataset. The proportions of subjects classified in the four posterior latent classes are similar in the two samples (see captions of Web Tables 3.9 and 3.12), as well as the estimated class-specific mean Isaacs trajectories (part C of Web Figure 3.6). The main differences are a delayed intensity transition to Death in class 2 that leads to a much higher proportion of subjects that develop dementia before death and a lower transition intensity to Dementia in class 3 (the healthiest class). Besides, this model has a higher discriminatory ability, as shown in Web Table 3.12.

(a) Visits every 2 years

		Joint illness-death model*					Joint competing risks model*					
		β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate		β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate
Class Membership	ζ_1	0.00	0.44	0.1983	0.2044	0.95	ζ_1	0.00	-0.22	0.2442	0.2785	0.85
Baseline Risks of events	$\lambda^{(1)}_{011}$	3.20	3.19	0.3068	0.3305	0.94	$\lambda^{(1)}_{011}$	3.20	3.00	0.3739	0.4144	0.92
	$\lambda^{(1)}_{012}$	3.50	3.52	0.2701	0.2720	0.94	$\lambda^{(1)}_{012}$	3.50	3.08	0.2756	0.2735	0.68
	$\lambda^{(2)}_{011}$	0.11	0.11	0.0010	0.0010	0.96	$\lambda^{(2)}_{011}$	0.11	0.11	0.0018	0.0035	0.90
	$\lambda^{(2)}_{012}$	0.10	0.10	0.0006	0.0006	0.94	$\lambda^{(2)}_{012}$	0.10	0.10	0.0007	0.0007	0.94
	$\lambda^{(1)}_{021}$	3.50	3.54	0.2365	0.2700	0.92	$\lambda^{(1)}_{021}$	3.50	3.67	0.2312	0.2636	0.88
	$\lambda^{(1)}_{022}$	3.40	3.42	0.1594	0.1582	0.95	$\lambda^{(1)}_{022}$	3.40	3.40	0.1487	0.1479	0.96
	$\lambda^{(2)}_{021}$	0.11	0.11	0.0005	0.0006	0.94	$\lambda^{(2)}_{021}$	0.11	0.11	0.0005	0.0006	0.68
	$\lambda^{(2)}_{022}$	0.10	0.10	0.0004	0.0004	0.94	$\lambda^{(2)}_{022}$	0.10	0.10	0.0004	0.0004	0.88
	$\lambda^{(1)}_{121}$	2.78	2.79	0.4198	0.4514	0.94						
	$\lambda^{(1)}_{122}$	3.14	3.19	0.4782	0.4945	0.95						
	$\lambda^{(2)}_{121}$	0.12	0.12	0.0067	0.0102	0.88						
	$\lambda^{(2)}_{122}$	0.11	0.11	0.0035	0.0049	0.89						
Event covariates	γ_{01}	0.02	0.03	0.1589	0.1609	0.95	γ_{01}	0.02	-0.08	0.1500	0.1539	0.89
	γ_{02}	0.67	0.68	0.1056	0.1061	0.95	γ_{02}	0.67	0.66	0.1011	0.1024	0.94
	γ_{12}	0.47	0.48	0.1893	0.1980	0.93						
Latent process	β_{01}	30.22	30.19	0.5534	0.5884	0.93	β_{01}	30.22	29.98	0.6363	0.6844	0.93
	β_{02}	32.96	33.00	0.3617	0.3678	0.95	β_{02}	32.96	32.85	0.3573	0.3693	0.92
	β_{11}	-5.76	-5.77	0.4028	0.4134	0.94	β_{11}	-5.76	-6.04	0.4539	0.4560	0.90
	β_{12}	-3.53	-3.54	0.1420	0.1443	0.95	β_{12}	-3.53	-3.55	0.1395	0.1395	0.95
	β_X	0.08	0.08	0.3108	0.3242	0.93	β_X	0.08	0.11	0.3103	0.3279	0.93
Cholesky transformation	U(1,1)	4.93	4.88	0.2070	0.2072	0.95	U(1,1)	4.93	4.85	0.2112	0.2119	0.94
of the B matrix	U(1,2)	-1.15	-1.13	0.1454	0.1454	0.94	U(1,2)	-1.15	-1.12	0.1458	0.1465	0.94
	U(2,2)	1.46	1.42	0.0964	0.1002	0.93	U(2,2)	1.46	1.43	0.0965	0.1000	0.94
Measurement error	σ_e	3.47	3.45	0.0364	0.0365	0.96	σ_e	3.47	3.47	0.0364	0.0364	0.95

*Models converged based on 500 samples with convergence criteria fulfilled

* 500 estimations had convergence criteria fulfilled

(b) Visits every 4 years

		Joint illness-death model*					Joint competing risks model*					
		β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate		β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate
Class Membership	ζ_1	0.00	0.01	0.2035	0.2133	0.96	ζ_1	0.00	-0.37	0.3988	0.3910	0.70
Baseline Risks of events	$\lambda^{(1)}_{011}$	3.20	3.20	0.3524	0.3770	0.95	$\lambda^{(1)}_{011}$	3.20	2.76	0.5509	0.6043	0.88
	$\lambda^{(1)}_{012}$	3.50	3.52	0.2940	0.2953	0.94	$\lambda^{(1)}_{012}$	3.50	2.71	0.2863	0.2953	0.17
	$\lambda^{(2)}_{011}$	0.11	0.11	0.0011	0.0010	0.95	$\lambda^{(2)}_{011}$	0.11	0.11	0.0161	0.0116	0.72
	$\lambda^{(2)}_{012}$	0.10	0.10	0.0007	0.0007	0.94	$\lambda^{(2)}_{012}$	0.10	0.10	0.0010	0.0009	0.84
	$\lambda^{(1)}_{021}$	3.50	3.54	0.2520	0.2848	0.93	$\lambda^{(1)}_{021}$	3.50	3.72	0.2345	0.2574	0.84
	$\lambda^{(1)}_{022}$	3.40	3.42	0.1666	0.1656	0.96	$\lambda^{(1)}_{022}$	3.40	3.39	0.1453	0.1445	0.96
	$\lambda^{(2)}_{021}$	0.11	0.11	0.0006	0.0006	0.95	$\lambda^{(2)}_{021}$	0.11	0.11	0.0005	0.0006	0.34
	$\lambda^{(2)}_{022}$	0.10	0.10	0.0005	0.0005	0.95	$\lambda^{(2)}_{022}$	0.10	0.10	0.0004	0.0004	0.72
	$\lambda^{(1)}_{121}$	2.78	2.80	0.4750	0.4935	0.94						
	$\lambda^{(1)}_{122}$	3.14	3.20	0.5242	0.5592	0.94						
	$\lambda^{(2)}_{121}$	0.12	0.12	0.0082	0.0112	0.88						
	$\lambda^{(2)}_{122}$	0.11	0.11	0.0047	0.0104	0.89						
Event covariates	γ_{01}	0.02	0.02	0.1729	0.1708	0.94	γ_{01}	0.02	-0.17	0.1525	0.1563	0.77
	γ_{02}	0.67	0.68	0.1099	0.1117	0.94	γ_{02}	0.67	0.64	0.1002	0.1013	0.93
	γ_{12}	0.47	0.48	0.2111	0.2301	0.92						
Latent process	β_{01}	30.22	30.18	0.5603	0.5967	0.92	β_{01}	30.22	29.85	0.6935	0.7516	0.91
	β_{02}	32.96	33.00	0.3635	0.3710	0.94	β_{02}	32.96	32.71	0.3561	0.3747	0.87
	β_{11}	-5.76	-5.77	0.4063	0.4170	0.94	β_{11}	-5.76	-6.19	0.4963	0.4955	0.88
	β_{12}	-3.53	-3.54	0.1426	0.1436	0.95	β_{12}	-3.53	-3.55	0.1379	0.1379	0.95
	β_X	0.08	0.08	0.3121	0.3248	0.95	β_X	0.08	0.14	0.3123	0.3289	0.92
Cholesky transformation	U(1,1)	4.93	4.88	0.2080	0.2077	0.95	U(1,1)	4.93	4.85	0.2130	0.2165	0.94
of the B matrix	U(1,2)	-1.15	-1.13	0.1455	0.1457	0.94	U(1,2)	-1.15	-1.11	0.1462	0.1464	0.93
	U(2,2)	1.46	1.42	0.0968	0.0999	0.93	U(2,2)	1.46	1.45	0.0970	0.1010	0.94
Measurement error	σ_e	3.47	3.47	0.0364	0.0365	0.96	σ_e	3.47	3.47	0.0364	0.0365	0.95

*Models converged based on 500 samples with convergence criteria fulfilled

* 495 estimations had convergence criteria fulfilled

Web Table 3.3: Results of the simulation study comparing estimates of the two-latent-class joint linear markovian illness-death model for interval-censored data and the joint competing risks model based on 500 samples of 1000 subjects generated with a joint markovian illness-death model with visits every 2 years or every 4 years. ASE is the asymptotic standard error, ESE is the empirical standard error and the coverage rate is calculated from the 95% confidence interval.

(a) Visits every 2 years

		β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate
Class Membership	ζ_1	0.00	0.03	0.3061	0.3389	0.93
Baseline Risks of events	$\lambda^{(1)}_{011}$	3.20	3.20	0.4250	0.4838	0.93
	$\lambda^{(1)}_{012}$	3.50	3.55	0.4335	0.4743	0.92
	$\lambda^{(2)}_{011}$	0.11	0.11	0.0020	0.0026	0.93
	$\lambda^{(2)}_{012}$	0.10	0.10	0.0010	0.0010	0.95
	$\lambda^{(1)}_{021}$	3.50	3.52	0.3312	0.3532	0.93
	$\lambda^{(1)}_{022}$	3.40	3.43	0.2293	0.2433	0.94
	$\lambda^{(2)}_{021}$	0.11	0.11	0.0008	0.0009	0.94
	$\lambda^{(2)}_{022}$	0.10	0.10	0.0006	0.0007	0.94
	$\lambda^{(1)}_{121}$	1.32	1.35	0.1122	0.1317	0.93
	$\lambda^{(1)}_{122}$	1.13	1.18	0.1466	0.1717	0.93
	$\lambda^{(2)}_{121}$	0.37	0.37	0.0211	0.0209	0.95
	$\lambda^{(2)}_{122}$	0.48	0.49	0.0479	0.0520	0.91
Event covariates	γ_{01}	0.02	0.02	0.2310	0.2562	0.94
	γ_{02}	0.67	0.69	0.1499	0.1550	0.95
	γ_{12}	0.47	0.50	0.2248	0.2214	0.96
Latent process	β_{01}	30.22	30.20	0.7926	0.8398	0.94
	β_{02}	32.96	32.96	0.5208	0.4931	0.97
	β_{11}	-5.76	-5.80	0.5676	0.6036	0.93
	β_{12}	-3.53	-3.51	0.2029	0.2048	0.95
	β_X	0.08	0.08	0.4549	0.4251	0.96
Cholesky transformation	U(1,1)	4.93	4.88	0.2935	0.2874	0.95
of the B matrix	U(1,2)	-1.15	-1.12	0.2065	0.2008	0.95
$UU^\top = B$	U(2,2)	1.46	1.41	0.1391	0.1489	0.94
Measurement error	σ_e	3.47	3.47	0.0514	0.0538	0.94

*Models converged based on 500 samples with convergence criteria fulfilled

(b) Visits every 4 years

		β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate
Class Membership	ζ_1	0.00	0.02	0.3307	0.4023	0.94
Baseline Risks of events	$\lambda^{(1)}_{011}$	3.20	3.19	0.4815	0.5332	0.94
	$\lambda^{(1)}_{012}$	3.50	3.52	0.5103	0.5920	0.92
	$\lambda^{(2)}_{011}$	0.11	0.11	0.0056	0.0296	0.95
	$\lambda^{(2)}_{012}$	0.10	0.10	0.0014	0.0014	0.93
	$\lambda^{(1)}_{021}$	3.50	3.53	0.3499	0.3844	0.93
	$\lambda^{(1)}_{022}$	3.40	3.43	0.2491	0.2548	0.95
	$\lambda^{(2)}_{021}$	0.11	0.11	0.0009	0.0011	0.93
	$\lambda^{(2)}_{022}$	0.10	0.10	0.0008	0.0008	0.95
	$\lambda^{(1)}_{121}$	1.32	1.37	0.1540	0.1736	0.93
	$\lambda^{(1)}_{122}$	1.13	1.30	0.2479	0.6555	0.93
	$\lambda^{(2)}_{121}$	0.37	0.37	0.0260	0.0282	0.94
	$\lambda^{(2)}_{122}$	0.48	0.49	0.0637	0.0762	0.88
Event covariates	γ_{01}	0.02	0.01	0.2616	0.2976	0.93
	γ_{02}	0.67	0.69	0.1602	0.1700	0.95
	γ_{12}	0.47	0.52	0.2635	0.2577	0.96
Latent process	β_{01}	30.22	30.15	0.8042	0.8603	0.94
	β_{02}	32.96	32.95	0.5227	0.4987	0.97
	β_{11}	-5.76	-5.78	0.5739	0.6086	0.93
	β_{12}	-3.53	-3.51	0.2038	0.2065	0.95
	β_X	0.08	0.09	0.4571	0.4277	0.96
Cholesky transformation	U(1,1)	4.93	4.87	0.2949	0.2873	0.95
of the B matrix	U(1,2)	-1.15	-1.12	0.2066	0.2006	0.94
$UU^\top = B$	U(2,2)	1.46	1.41	0.1403	0.1500	0.95
Measurement error	σ_e	3.47	3.47	0.0514	0.0535	0.94

*Models converged based on 498 samples with convergence criteria fulfilled

Web Table 3.4: Results of the simulation study for the two-latent-class joint linear semi-markovian illness-death model for interval-censored data based on 500 samples of 500 subjects generated with visits every 2 years or every 4 years. ASE is the asymptotic standard error, ESE is the empirical standard error and the coverage rate is calculated from the 95% confidence interval.

		(a) Visits every 2 years					(b) Visits every 4 years				
		β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate	β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate
Class Membership	ζ_1	0.00	-0.00	0.1409	0.1353	0.96	0.00	-0.00	0.1439	0.1379	0.97
	ζ_2	0.00	0.00	0.1607	0.1637	0.95	0.00	0.00	0.1683	0.1709	0.95
Baseline Risks of events	$\lambda_{011}^{(1)}$	4.95	4.98	0.3064	0.3325	0.93	4.95	4.98	0.3308	0.3537	0.94
	$\lambda_{012}^{(1)}$	4.70	4.74	0.2684	0.2737	0.95	4.70	4.74	0.2909	0.3054	0.94
	$\lambda_{013}^{(1)}$	4.84	4.88	0.2514	0.2439	0.96	4.84	4.89	0.2693	0.2611	0.95
	$\lambda_{011}^{(2)}$	0.10	0.10	0.0004	0.0004	0.96	0.10	0.10	0.0004	0.0004	0.96
	$\lambda_{012}^{(2)}$	0.11	0.11	0.0004	0.0003	0.96	0.11	0.11	0.0004	0.0004	0.96
	$\lambda_{013}^{(2)}$	0.10	0.10	0.0003	0.0003	0.94	0.10	0.10	0.0003	0.0003	0.95
	$\lambda_{021}^{(1)}$	3.31	3.30	0.2814	0.2813	0.95	3.31	3.30	0.2929	0.2936	0.95
	$\lambda_{022}^{(1)}$	3.28	3.26	0.3490	0.3738	0.94	3.28	3.26	0.3684	0.3925	0.95
	$\lambda_{023}^{(1)}$	3.10	3.09	0.3528	0.3540	0.95	3.10	3.09	0.3719	0.3701	0.96
	$\lambda_{021}^{(2)}$	0.10	0.10	0.0007	0.0007	0.94	0.10	0.10	0.0007	0.0007	0.95
	$\lambda_{022}^{(2)}$	0.10	0.11	0.0009	0.0009	0.98	0.10	0.10	0.0010	0.0010	0.97
	$\lambda_{023}^{(2)}$	0.10	0.10	0.0009	0.0009	0.96	0.10	0.10	0.0010	0.0010	0.95
	$\lambda_{121}^{(1)}$	4.80	4.95	0.5603	0.5594	0.95	4.80	4.96	0.6132	0.6171	0.95
	$\lambda_{122}^{(1)}$	4.55	4.63	0.4481	0.4698	0.94	4.55	4.64	0.4855	0.5283	0.94
	$\lambda_{123}^{(1)}$	4.70	4.83	0.4506	0.4577	0.95	4.70	4.83	0.4868	0.4928	0.94
	$\lambda_{121}^{(2)}$	0.10	0.10	0.0007	0.0007	0.94	0.10	0.10	0.0008	0.0008	0.93
	$\lambda_{122}^{(2)}$	0.10	0.10	0.0007	0.0007	0.94	0.10	0.10	0.0007	0.0009	0.92
	$\lambda_{123}^{(2)}$	0.10	0.10	0.0006	0.0006	0.95	0.10	0.10	0.0006	0.0006	0.94
Event covariates	γ_{01}	-1.07	-1.08	0.1647	0.1630	0.95	-1.07	-1.09	0.1758	0.1782	0.94
	γ_{02}	-0.12	-0.12	0.1433	0.1475	0.94	-0.12	-0.12	0.1451	0.1495	0.94
	γ_{12}	-0.03	-0.04	0.2019	0.2116	0.93	-0.03	-0.05	0.2108	0.2212	0.94
Latent process	β_{01}	32.05	32.04	0.2216	0.2283	0.94	32.05	32.04	0.2216	0.2289	0.94
	β_{02}	30.99	31.00	0.2345	0.2221	0.97	30.99	31.00	0.2345	0.2213	0.97
	β_{03}	28.97	28.97	0.2208	0.2192	0.97	28.97	28.97	0.2208	0.2190	0.96
	β_{11}	-3.08	-3.07	0.1051	0.1111	0.93	-3.08	-3.07	0.1051	0.1112	0.93
	β_{12}	-5.60	-5.61	0.1136	0.1113	0.96	-5.60	-5.61	0.1136	0.1111	0.96
	β_{13}	-7.69	-7.69	0.1043	0.1026	0.96	-7.69	-7.69	0.1043	0.1025	0.96
	β_X	4.83	4.83	0.1109	0.1150	0.94	4.83	4.83	0.1109	0.1151	0.94
Cholesky transformation	U(1,1)	2.24	2.23	0.1027	0.1042	0.95	U(1,1)	2.24	2.23	0.1027	0.1042	0.95
of the B matrix	U(1,2)	-0.89	-0.89	0.0580	0.0599	0.94	U(1,2)	-0.89	-0.89	0.0580	0.0599	0.94
$UU^\top = B$	U(2,2)	0.45	0.44	0.0230	0.0228	0.95	U(2,2)	0.45	0.44	0.0230	0.0228	0.95
Measurement error	σ_e	1.00	1.00	0.0137	0.0133	0.96	σ_e	1.00	1.00	0.0137	0.0133	0.96
*Models converged based on 500 samples with convergence criteria fulfilled							*Models converged based on 499 samples				

Web Table 3.5: Results of the simulation study of the three-latent-class joint linear markovian illness-death model for interval-censored data based on 500 samples of 500 subjects generated with visits every 2 years or every 4 years. ASE is the asymptotic standard error, ESE is the empirical standard error and the coverage rate is calculated from the 95% confidence interval.

(a) $\lambda_{01}(t) = \lambda_{01}^{0}(t)exp(\beta X_i + u_{i0})$

		β	$\hat{\beta}$	ASE	ESE	Cover Rate
Class Membership	ζ_1	0.00	-0.01	0.1370	0.1429	0.95
	ζ_2	0.00	0.00	0.1565	0.1497	0.97
Baseline Risks of events	$\lambda_{011}^{(1)}$	4.95	4.80	0.2798	0.2833	0.91
	$\lambda_{012}^{(1)}$	4.70	4.54	0.2473	0.2494	0.89
	$\lambda_{013}^{(1)}$	4.84	4.70	0.2326	0.2228	0.90
	$\lambda_{011}^{(2)}$	0.10	0.10	0.0004	0.0004	0.95
	$\lambda_{012}^{(2)}$	0.11	0.11	0.0004	0.0004	0.93
	$\lambda_{013}^{(2)}$	0.11	0.10	0.0003	0.0003	0.94
	$\lambda_{021}^{(1)}$	3.31	3.33	0.2634	0.2584	0.94
	$\lambda_{022}^{(1)}$	3.28	3.31	0.3203	0.3334	0.94
	$\lambda_{023}^{(1)}$	3.10	3.10	0.3185	0.3225	0.96
	$\lambda_{021}^{(2)}$	0.10	0.10	0.0007	0.0007	0.94
	$\lambda_{022}^{(2)}$	0.11	0.11	0.0008	0.0009	0.97
	$\lambda_{023}^{(2)}$	0.10	0.10	0.0009	0.0008	0.95
	$\lambda_{121}^{(1)}$	4.80	4.93	0.5037	0.5083	0.94
	$\lambda_{122}^{(1)}$	4.55	4.63	0.4059	0.4078	0.94
	$\lambda_{123}^{(1)}$	4.70	4.79	0.4122	0.4089	0.95
	$\lambda_{121}^{(2)}$	0.10	0.10	0.0006	0.0007	0.95
	$\lambda_{122}^{(2)}$	0.10	0.10	0.0006	0.0006	0.94
	$\lambda_{123}^{(2)}$	0.10	0.10	0.0005	0.0005	0.95
Event covariates	γ_{01}	-1.07	-1.00	0.1545	0.1591	0.92
	γ_{02}	-0.12	-0.13	0.1383	0.1493	0.93
	γ_{12}	-0.03	-0.05	0.1886	0.2032	0.93
Latent process	β_{01}	32.05	32.00	0.2205	0.2179	0.96
	β_{02}	30.99	30.91	0.2331	0.2230	0.95
	β_{03}	28.97	28.91	0.2197	0.2253	0.94
	β_{11}	-3.08	-3.06	0.1047	0.1044	0.95
	β_{12}	-5.60	-5.56	0.1133	0.1080	0.96
	β_{13}	-7.69	-7.66	0.1038	0.1071	0.94
	β_X	4.83	4.83	0.1114	0.1072	0.96
Cholesky transformation	U(1,1)	2.24	2.21	0.1023	0.1048	0.93
of the B matrix	U(1,2)	-0.89	-0.88	0.0581	0.0591	0.93
$UU^\top = B$	U(2,2)	0.45	0.44	0.0233	0.0243	0.93
Measurement error	σ_e	1.00	1.00	0.0137	0.0124	0.97

*Models converged based on 500 samples with convergence criteria fulfilled

(b) $\lambda_{01}(t) = \lambda_{01}^{0}(t)exp(\beta X_i + 0.2\ u_{i1})$

	β	$\hat{\beta}$	ASE	ESE	Cover Rate
ζ_1	0.00	-0.01	0.1357	0.1374	0.97
ζ_2	0.00	-0.00	0.1545	0.1565	0.96
$\lambda_{011}^{(1)}$	4.95	4.94	0.2855	0.2758	0.95
$\lambda_{012}^{(1)}$	4.70	4.67	0.2490	0.2486	0.94
$\lambda_{013}^{(1)}$	4.84	4.82	0.2356	0.2352	0.95
$\lambda_{011}^{(2)}$	0.10	0.10	0.0003	0.0003	0.93
$\lambda_{012}^{(2)}$	0.11	0.11	0.0003	0.0004	0.93
$\lambda_{013}^{(2)}$	0.11	0.11	0.0003	0.0003	0.94
$\lambda_{021}^{(1)}$	3.31	3.32	0.2617	0.2600	0.95
$\lambda_{022}^{(1)}$	3.28	3.28	0.3222	0.3112	0.96
$\lambda_{023}^{(1)}$	3.10	3.11	0.3212	0.3254	0.95
$\lambda_{021}^{(2)}$	0.10	0.10	0.0007	0.0007	0.96
$\lambda_{022}^{(2)}$	0.11	0.11	0.0008	0.0009	0.96
$\lambda_{023}^{(2)}$	0.10	0.10	0.0009	0.0010	0.96
$\lambda_{121}^{(1)}$	4.80	4.93	0.5162	0.5405	0.94
$\lambda_{122}^{(1)}$	4.55	4.63	0.4130	0.3841	0.97
$\lambda_{123}^{(1)}$	4.70	4.78	0.4137	0.3941	0.96
$\lambda_{121}^{(2)}$	0.10	0.10	0.0006	0.0007	0.93
$\lambda_{122}^{(2)}$	0.10	0.10	0.0006	0.0005	0.94
$\lambda_{123}^{(2)}$	0.10	0.10	0.0005	0.0005	0.96
γ_{01}	-1.07	-1.06	0.1558	0.1617	0.92
γ_{02}	-0.12	-0.12	0.1386	0.1423	0.94
γ_{12}	-0.03	-0.04	0.1903	0.2007	0.94
β_{01}	32.05	32.08	0.2221	0.2149	0.95
β_{02}	30.99	31.03	0.2344	0.2320	0.96
β_{03}	28.97	29.02	0.2197	0.2268	0.94
β_{11}	-3.08	-3.10	0.1054	0.1019	0.95
β_{12}	-5.60	-5.63	0.1136	0.1163	0.93
β_{13}	-7.69	-7.72	0.1038	0.1067	0.93
β_X	4.83	4.83	0.1115	0.1191	0.94
U(1,1)	2.24	2.22	0.1027	0.1027	0.93
U(1,2)	-0.89	-0.89	0.0581	0.0576	0.94
U(2,2)	0.45	0.45	0.0232	0.0222	0.97
σ_e	1.00	1.00	0.0137	0.0139	0.94

*Models converged based on 500 samples

Web Table 3.6: Results of the simulation study of the misspecified three-latent-class joint linear markovian illness-death model for interval-censored data based on 500 samples of 500 subjects generated with visits every 2 years. The simulated transition intensity to dementia depends on (a) the individual random intercept or (b) the individual random slope.

	β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate
ζ_1	0.00	-0.00	0.1158	0.1171	0.94
ζ_2	0.00	0.00	0.1385	0.1375	0.95
ζ_3	0.00	0.00	0.1488	0.1545	0.92
ζ_4	0.00	0.00	0.1962	0.1955	0.96
ζ_5	0.00	0.01	0.2021	0.2244	0.93
$\lambda_{011}^{(1)}$	4.95	5.00	0.2285	0.2347	0.95
$\lambda_{012}^{(1)}$	4.70	4.73	0.2172	0.2225	0.94
$\lambda_{013}^{(1)}$	4.84	4.88	0.2244	0.2441	0.94
$\lambda_{014}^{(1)}$	4.30	4.34	0.2404	0.2670	0.94
$\lambda_{015}^{(1)}$	4.81	4.95	0.5355	0.6618	0.92
$\lambda_{016}^{(1)}$	4.50	4.53	0.1814	0.1935	0.93
$\lambda_{011}^{(2)}$	0.10	0.10	0.0002	0.0003	0.93
$\lambda_{012}^{(2)}$	0.11	0.11	0.0003	0.0003	0.96
$\lambda_{013}^{(2)}$	0.11	0.11	0.0003	0.0003	0.94
$\lambda_{014}^{(2)}$	0.11	0.11	0.0004	0.0004	0.93
$\lambda_{015}^{(2)}$	0.10	0.10	0.0006	0.0006	0.91
$\lambda_{016}^{(2)}$	0.10	0.10	0.0003	0.0003	0.94
$\lambda_{021}^{(1)}$	3.31	3.33	0.2160	0.2099	0.95
$\lambda_{022}^{(1)}$	3.28	3.32	0.3123	0.3078	0.94
$\lambda_{023}^{(1)}$	3.10	3.15	0.3814	0.3776	0.94
$\lambda_{024}^{(1)}$	3.12	3.16	0.3986	0.4080	0.94
$\lambda_{025}^{(1)}$	3.15	3.18	0.2953	0.3105	0.93
$\lambda_{026}^{(1)}$	3.17	3.19	0.2981	0.2962	0.95
$\lambda_{021}^{(2)}$	0.10	0.10	0.0005	0.0005	0.96
$\lambda_{022}^{(2)}$	0.11	0.11	0.0007	0.0007	0.95
$\lambda_{023}^{(2)}$	0.10	0.10	0.0008	0.0008	0.96
$\lambda_{024}^{(2)}$	0.10	0.10	0.0010	0.0011	0.94
$\lambda_{025}^{(2)}$	0.11	0.11	0.0009	0.0010	0.92
$\lambda_{026}^{(2)}$	0.10	0.10	0.0007	0.0006	0.95
$\lambda_{121}^{(1)}$	4.90	4.97	0.4082	0.4019	0.95
$\lambda_{122}^{(1)}$	4.55	4.58	0.3525	0.3590	0.95
$\lambda_{123}^{(1)}$	4.70	4.78	0.3822	0.3866	0.96
$\lambda_{124}^{(1)}$	4.50	4.58	0.4211	0.4378	0.96
$\lambda_{125}^{(1)}$	4.80	5.04	1.0264	1.3517	0.94
$\lambda_{126}^{(1)}$	4.10	4.15	0.3777	0.3629	0.95
$\lambda_{121}^{(2)}$	0.10	0.10	0.0005	0.0005	0.96
$\lambda_{122}^{(2)}$	0.10	0.10	0.0005	0.0006	0.93
$\lambda_{123}^{(2)}$	0.10	0.10	0.0005	0.0005	0.94
$\lambda_{124}^{(2)}$	0.10	0.10	0.0005	0.0005	0.96
$\lambda_{125}^{(2)}$	0.10	0.10	0.0036	0.0087	0.90
$\lambda_{126}^{(2)}$	0.10	0.10	0.0005	0.0006	0.95

	β	$\hat{\beta}$	*ASE*	*ESE*	Cover Rate
γ_{01}	-1.07	-1.08	0.0868	0.0878	0.93
γ_{02}	-0.12	-0.13	0.0705	0.0731	0.93
γ_{12}	-0.03	-0.03	0.1106	0.1127	0.94
β_{01}	32.05	32.06	0.1643	0.1591	0.95
β_{02}	30.99	31.00	0.1901	0.1878	0.96
β_{03}	29.97	29.97	0.1904	0.2016	0.94
β_{04}	28.05	28.08	0.2321	0.2495	0.93
β_{05}	27.20	27.21	0.2290	0.2387	0.95
β_{06}	26.00	25.99	0.1721	0.1772	0.95
β_{11}	-3.08	-3.08	0.0774	0.0754	0.95
β_{12}	-4.60	-4.60	0.0948	0.0957	0.95
β_{13}	-5.69	-5.68	0.0882	0.0914	0.94
β_{14}	-6.45	-6.45	0.1139	0.1154	0.94
β_{15}	-7.10	-7.10	0.1127	0.1182	0.93
β_{16}	-8.00	-8.00	0.0805	0.0860	0.93
β_X	4.83	4.83	0.0861	0.0870	0.95
U(1,1)	2.24	2.22	0.0556	0.0568	0.93
U(1,2)	-0.89	-0.89	0.0320	0.0320	0.95
U(2,2)	0.45	0.44	0.0187	0.0198	0.91
σ_e	1.00	1.00	0.0068	0.0068	0.95

*Models converged based on 456 samples

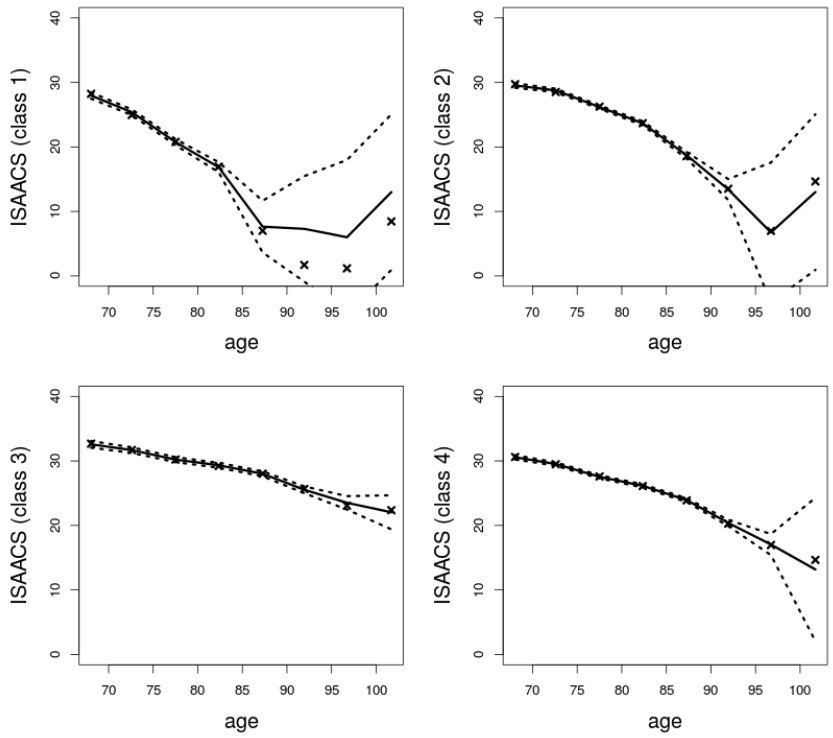
Web Table 3.7: Results of the simulation study of the six-latent-class joint linear markovian illness-death model for interval-censored data based on 500 samples of 2000 subjects generated with visits every 2 years. ASE is the asymptotic standard error, ESE is the empirical standard error and the coverage rate is calculated from the 95% confidence interval.

		$\hat{\beta}$	SE	CI inf	CI sup
Class Membership					
	π_1	-0.40	0.23	-0.86	0.05
	π_2	0.05	0.23	-0.40	0.50
	π_3	-1.32	0.23	-1.76	-0.87
Illness-death model					
Education	γ_{01}^e	-1.18	0.12	-1.43	-0.94
Gender	γ_{01}^s	-0.07	0.11	-0.29	0.15
Education	γ_{02}^e	-0.29	0.07	-0.43	-0.16
Gender	γ_{02}^s	-0.70	0.06	-0.83	-0.58
Education	γ_{12}^e	-0.00	0.05	-0.09	0.08
Gender	γ_{12}^s	-0.54	0.08	-0.70	-0.37
Latent process					
Class-specific	τ_1	84.22	0.86	82.52	85.91
Change-point times	τ_2	82.62	0.44	81.75	83.48
	τ_3	89.42	1.12	87.23	91.61
	τ_4	86.30	0.62	85.08	87.52
Education	β_3	1.280	0.09	1.11	1.45
Education $\times\, t$ if $t < \tau_g$	β_4	-0.002	0.03	-0.07	0.07
Education $\times\, t$ if $t > \tau_g$	β_5	-0.004	0.04	-0.09	0.08
Gender	β_6	0.090	0.05	0.00	0.18

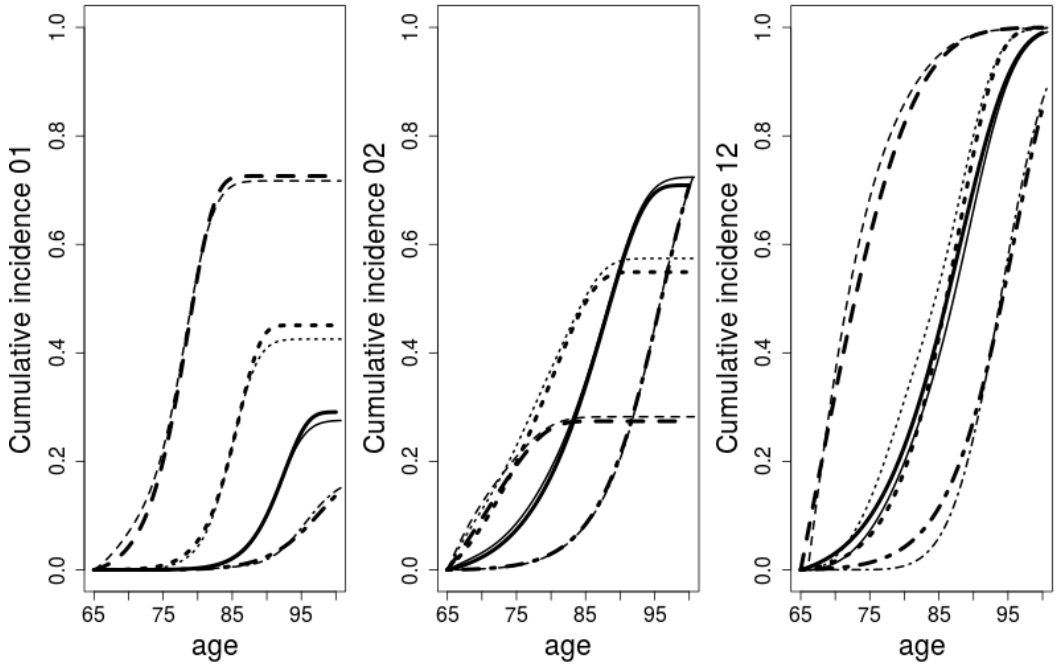
Web Table 3.8: Estimation of the parameters of the four-latent-class joint change-point markovian illness-death model for semi-competing interval-censored events and longitudinal data (Paquid data, N=3,525).

Class	1	2	3	4
1	71.36	21.92	0.18	6.54
2	12.83	61.54	1.24	24.39
3	0.01	0.47	79.11	20.40
4	0.96	19.11	12.96	66.97

Web Table 3.9: Mean probabilities (in percentages) to belong to each class according to the posterior classification, allocating 392 (11.12%) subjects in class 1, 1160 (32.91%) to class 2, 343 (9.73%) to class 3 and 1630 (46.24%) to class 4 (Paquid data, N=3,525).

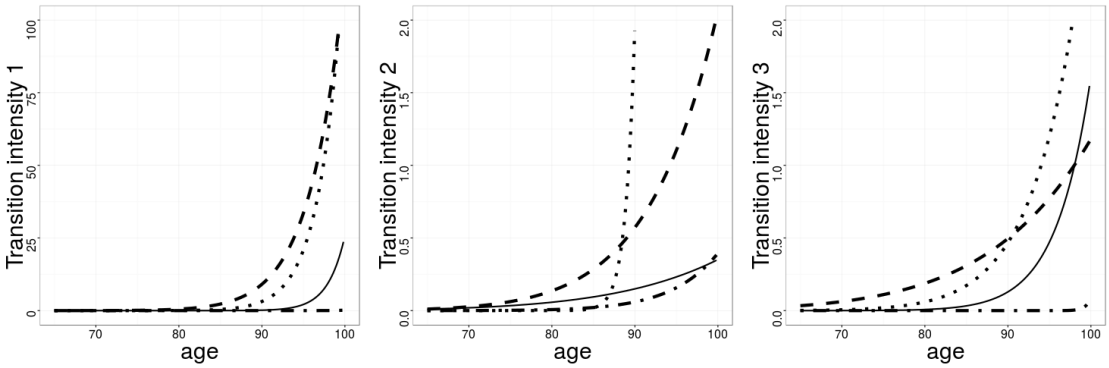


A) Predicted means of Isaacs scores given the random effects and the classes by age range (x) and weighted means of the observed scores (solid lines) with their 95% confidence intervals (dashed lines) for each class.

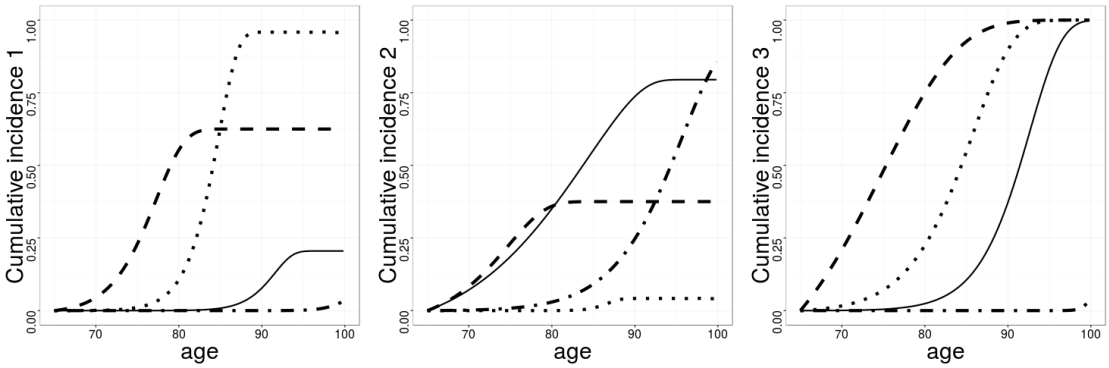


B) Class-specific predicted cumulative incidences estimated by the joint latent class illness-death model (thick lines) and by the weighted semi-parametric illness-death model (thin lines), for each class (class 1: dashed line, class 2: dotted line, class 3: dotdashed line, class 4: solid line).

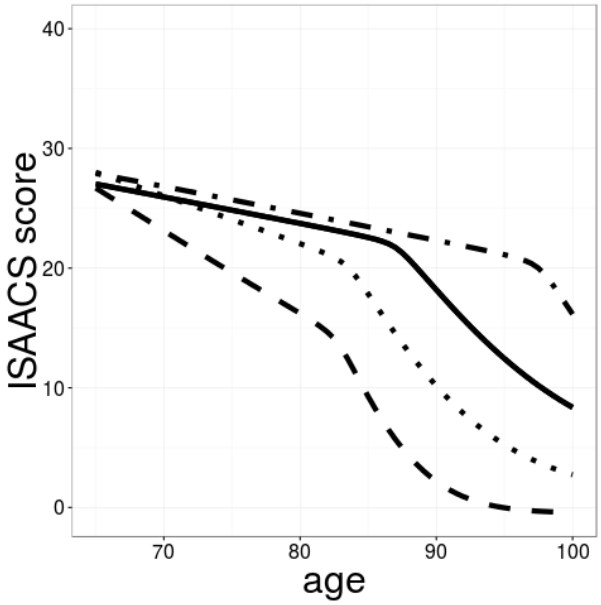
Web Figure 3.5: Goodness-of-fit assessment of the four-latent-class joint change-point illness-death model estimated on Paquid data (N= 3,525).



A) Class-specific transition intensities of the illness-death model for men, ApoE4 non-carrier, with a low level of education.



B) Class-specific cumulative incidences of the illness-death model for men, ApoE4 non-carrier, with a low level of education.



C) Class-specific Isaacs trajectories for men, ApoE4 non-carrier, with a low level of education.

Web Figure 3.6: Class-specific estimated transition intensities, cumulative incidences and mean longitudinal trajectories of the joint change-point latent class illness-death mixed model for each class (class 1: dashed line, class 2: dotted line, class 3: dotdashed line, class 4: solid line, sub-sample from Paquid cohort with ApoE4 measurement, N= 619).

Class	CEP=0	CEP=1	men	women
1	23.2	76.8	40.1	59.9
2	29.7	70.3	40.7	59.3
3	42.9	57.1	42.0	58.0
4	38.0	62.0	43.8	56.2

Web Table 3.10: Class-specific description of the posterior classes of the four-latent-class joint change-point markovian illness-death model, for gender and educational level, in percentages (Paquid data, N=3,525).

		$\hat{\beta}$	SE	CI inf	CI sup
Class Membership					
	π_1	-0.11	0.2607	-0.62	0.40
Illness-death model					
Education	γ^e_{01}	-0.41	0.1906	-0.79	-0.04
Gender	γ^s_{01}	0.24	0.1877	-0.12	0.61
ApoE4	γ^a_{01}	0.43	0.2176	0.00	0.85
Education	γ^e_{02}	0.01	0.1578	-0.30	0.31
Gender	γ^s_{02}	-0.72	0.1535	-1.02	-0.42
ApoE4	γ^a_{02}	0.20	0.1845	-0.16	0.56
Education	γ^e_{12}	0.16	0.1665	-0.17	0.48
Gender	γ^s_{12}	-0.41	0.1697	-0.74	-0.07
ApoE4	γ^a_{12}	-0.24	0.1724	-0.58	0.09
Latent process					
Class-specific	τ_1	86.07	0.7235	84.65	87.49
Change-point times	τ_2	83.12	0.6076	81.93	84.31
Education	β_3	0.91	0.1387	0.64	1.18
Education $\times\, t$ if $t < \tau_g$	β_4	-0.11	0.1288	-0.37	0.14
Education $\times\, t$ if $t > \tau_g$	β_5	-0.04	0.1651	-0.36	0.28
Gender	β_6	0.12	0.0796	-0.03	0.28
ApoE4	β_7	-0.14	0.1331	-0.41	0.11
ApoE4 $\times\, t$ if $t < \tau_g$	β_8	-0.23	0.1652	-0.55	0.09
ApoE4 $\times\, t$ if $t > \tau_g$	β_9	-0.03	0.1969	-0.42	0.36

Web Table 3.11: Estimation of the parameters of the two-latent-class joint change-point markovian illness-death model, adjusted on gender, educational level and ApoE4 (subsample of Paquid cohort with ApoE4 measurement, N=619).

Class	1	2	3	4
1	76.21	9.30	0.35	14.14
2	4.63	77.36	1.99	16.02
3	0.00	0.00	87.70	12.29
4	6.42	8.88	10.37	74.33

Web Table 3.12: Mean probabilities (in percentages) to belong to each class according to the posterior classification, allocating 103 (16.64%) subjects in class 1, 145 (23.42%) to class 2, 64 (10.34%) to class 3 and 307 (49.60%) to class 4 (sub-sample of PAQUID cohort with ApoE4 measurement, N=619).

4 Dynamic predictions for dementia

Joint models are useful tools to predict the risk to have dementia within the next few years, based on repeated cognitive markers. In this part, the model developed in the previous chapter is applied to propose dynamic prediction tools for dementia. Using the Paquid cohort as learning data set and the Three-city cohort as validation set, we compare the predictive abilities of joint models based on a single or multiple cognitive tests accounting for interval censoring of dementia and the competing risk of death.

4.1 Introduction

The usual tools to assess predictive accuracy are the time-dependent receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC) as well as the Brier Score. The AUC is a measure of discrimination that quantifies the probability that the marker value of a case is higher than the marker value of a control, provided that the higher the marker value, the higher the risk of having the disease. Thus, the AUC can be used to assess the ability of a model to predict the time-to-onset of the disease. On the other hand, the Brier Score is a mean squared error which quantifies both the discrimination and the calibration of the model. The second component depends on the distribution of the risk predictions and quantifies how close these predictions are from the true underlying risk of disease.

Within the framework of competing risks, Blanche et al. [2013] proposed fully non-parametric estimators of the AUC based on the inverse probability of censoring weighting (IPCW) method, as well as a test to compare the AUCs of two markers, measured at baseline. Handling competing risks requires to redefine the set of controls, especially by precising if the set includes subjects free of the disease who experienced the competing event. This can lead to two different definitions of the sets of controls and two different corresponding AUC estimators. Jacqmin-Gadda et al. [2014] proposed an approximation

to apply the IPCW estimators within the setting of semi-competing risks with interval censoring. A comparison between the IPCW estimators with this approximation and parametric illness-death estimators showed that the misspecification of the illness-death models could lead to substantial biases while IPCW estimators remained relatively efficient.

In the joint modeling context, predictions of the risk of dementia can be computed from repeated measures of the marker, and updated at each new measure. Blanche et al. [2015] extended the IPCW estimators of the AUC to assess and compare the dynamic predictive accuracy of such prognostic models.

This chapter presents a dynamic predictive model based on repeated measures of cognitive markers, accounting for interval censoring of dementia and the competing risk of death. The methodology for this model was described in Chapter 3 and the inference for AUC relied on the method proposed by Blanche et al. [2015].

4.2 Methods

4.2.1 Data

The prognostic models were constructed from the two French prospective cohorts presented in the section 1.2, fitted on the Paquid cohort and validated on the Three-city Study.

The Paquid training sample

To build the training sample, we selected subjects from the Paquid cohort, located in Gironde, who were not prevalent at entry and who completed at least once the Isaacs Set Test (IST), the MMSE and the Benton Visual Retention Test (BVRT) before the visit of dementia diagnosis, death or right censoring. This sample included 2,490 subjects, who entered the study on average at 74.5 years old (sd=6.5 years), with a majority of women (58.47%) and a majority of subjects with a high educational level (30.36%). During 25 years of follow-up, the subjects were seen every 2 or 3 years and 600 subjects (24.10%) were diagnosed with dementia, of which 519 died later (20.84% of the total sample), while 2212 (88.84%) died with no dementia diagnosis. The mean number of collected responses

was 4.58 (sd=3.02) for the MMSE and 4.31 (sd=3.04) for the IST.

The Three-city validation sample

The validation set was built from the Bordeaux and Montpellier samples of the Three-city Study, including initially 4,363 subjects. We excluded subjects who were deaf, blind or confined to bed at entry and we selected subjects who were free of dementia and who performed at least once the IST or the MMSE before their dementia diagnosis. Finally, the validation sample was composed of 3,880 subjects, including 59.9% of women and 91.4% of subjects with a high level of education, who entered the study at 73.7 years old on average (sd= 5.3 years). Subjects were seen every 2 or 3 years during a 10-year follow-up. To apply the IPCW estimators of AUC, we used the imputation rule proposed by Jacqmin-Gadda et al. [2014] to define the cases and controls despite interval censoring of the time-to-dementia: the time-to-dementia was imputed by the middle of the censoring interval for subjects diagnosed with dementia and subjects who died with no dementia diagnosis more than 2 years after their last visit were considered as right-censored at their last visit, otherwise they were considered as non-demented at their time-to-death. Thus, a total of 397 (10.2%) subjects were diagnosed with dementia, 403 (10.4%) individuals died before the dementia diagnosis and 3,080 (79.4%) were right-censored.

4.2.2 Cognitive tests

We focused on the Isaacs Set Test and the MMSE. The first one evaluates verbal fluency and consists in asking subjects to give as many names of cities, fruits, animals and colors as possible, within 1 minute for each category. The maximum number of names was truncated at 10. The test scores range from 0 to 40 points. Due to the ceiling effect of the 1-minute score, we used the score at 15 seconds which does not have any floor or ceiling effects. This test was shown to be sensitive to small changes in high levels of cognition, due to its speed component, but also in other ranges of cognition [Proust-Lima et al., 2007]. At last, it is a short test with simple instructions which can be easily proposed to large samples, including cognitively impaired individuals.

The MMSE assesses the global cognitive functioning through different items on memory, calculation, time and space orientation, language, and word recognition. The scores

range from 0 to 30 points. The MMSE is widely used in clinical practice as a screening test for dementia but also in population-based studies to quantify the cognitive change. However, it has a strong ceiling effect: it is not sensitive to changes in high levels of cognition and may not be appropriate to study the cognitive decline among subjects with high educational level. This test also has a poor sensitivity in low levels of cognition, which makes the discrimination of severely impaired subjects difficult. Nevertheless, the MMSE has a good discriminatory ability in the medium range of cognition and remains an interesting tool for dementia diagnosis. Thus, the MMSE is curvilinear: it shows a varying sensitivity to cognitive changes, that is to say that 1-point decrease in MMSE does not represent the same clinical deterioration according to the initial score. In order to avoid this curvilinearity issue, Philipps et al. [2014] proposed a normalizing transformation of the MMSE.

4.2.3 Prognostic models for dementia

We estimated three joint latent class models accounting for interval censoring of the time-to-dementia and the competing risk of death, on the Paquid cohort. The first model was based on IST data, the second on MMSE data and the third combined both IST and MMSE data.

The mixed sub-models of the three joint latent class models were specified in the same way. Given the latent class g, the latent cognitive process was modeled as follows:

$$\begin{aligned}\Lambda_i(t_{ij}) =& \beta_{0g} + u_{ig}^{(0)} + \beta_{0,\text{age}}\ \text{Age}_{i0} + \beta_{0,\text{CEP}}\ \text{CEP}_i + \beta_{0,\text{learn}}\ \mathbb{1}_{(t_{ij}=0)} \\ &+ (\beta_{1g} + u_{ig}^{(1)} + \beta_{1,\text{age}}\ \text{Age}_{i0}) \times t_{ij} \\ &+ (\beta_{2g} + u_{ig}^{(2)} + \beta_{2,\text{age}}\ \text{Age}_{i0}) \times t_{ij}^2\end{aligned}$$

with $u_{ig} = (u_{ig}^{(0)}, u_{ig}^{(1)}, u_{ig}^{(2)})^\top \sim \mathcal{N}(0, \sigma_g B)$ the class-specific random effects. The time scale was the delay from the entry in the cohort, in decades: $t_{ij} = \frac{\text{Age}_{ij} - \text{Age}_{\text{entry},i}}{10}$. The trajectory of the marker was adjusted on the scaled age at entry $\text{Age}_{i0} = \frac{\text{Age}_{\text{entry},i} - 65}{10}$, as well as on educational level ($\text{CEP}_i = 1$ if the subject obtained the primary school diploma and 0 otherwise). The time-dependent covariate $\mathbb{1}_{(t_{ij}=0)}$ allowed to account for the improvement in cognitive scores observed after the first interview, possibly due to the stress of the testing procedure at the first visit or the learning effect at the second

visit [Jacqmin-Gadda et al., 1997]. We assumed that each of the three above mentioned covariates has a common effect on the latent classes.

Marker-specific beta cumulative distribution functions $\psi_k(\cdot;\eta_k)$ depending on parameters η_k, were used to link the observed score Y_{ijk} of subject i at time t_{ij} of the repeated marker k to the latent cognitive process: $\psi_k(Y_{ijk};\eta_k) = \Lambda_i(t_{ij}) + \epsilon_{ijk}$ with $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2_{\epsilon_k})$ and σ_{ϵ_k} accounting for the marker-specific within-subject variability. Thus, the model based on the IST involved one transformation ψ^{M1}_{IST}, the model based on the MMSE involved ψ^{M2}_{MMSE} and the combined model involved both ψ^{M3}_{IST} and ψ^{M3}_{MMSE}.

Conditionally to the latent classes, the transition intensities of the illness-death submodels were modeled by proportional hazards models:

$$\alpha_{01ig}(t) = \alpha^0_{01g}(t)\ \exp(\gamma_{01g,\text{age}}\ \text{Age}_{i0} + \gamma_{01g,\text{CEP}}\ \text{CEP}_i)$$
$$\alpha_{\ell 2ig}(t) = \alpha^0_{\ell 2g}(t)\ \exp(\gamma_{\ell 2g,\text{age}}\ \text{Age}_{i0} + \gamma_{\ell 2g,\text{Sex}}\ \text{Sex}_i + \gamma_{\ell 2g,\text{CEP}}\ \text{CEP}_i),\ \text{with}\ \ell = 0, 1$$

with class-specific effects of the scaled age at entry and educational level for the transition toward dementia, and an additional class-specific effect of gender ($\text{Sex}_i = 1$ for women and 0 for men) for the two transitions toward Death. The class-specific baseline transition intensities were parameterized by Weibull functions.

4.2.4 Assessment criteria

In the joint modeling setting with competing risks and interval censoring, the dynamic area under the curve is defined at landmark time s for a prediction horizon t, as follows:

$$AUC(s,t) = P(\pi_i(s,t) > \pi_j(s,t) | D_i(s,t) = 1, D_j(s,t) = 0, T^A_i > s, T^D_i > s, T^A_j > s, T^D_j > s)$$

with T^A_i and T^D_i the time-to-dementia and the time-to-death for subject i respectively and

$$\pi_i(s,t) = P(s < T^A_i \leqslant s+t, T^D_i > T^A_i | T^A_i > s, T^D_i > s, \mathcal{H}^Y_i(s), X_i) \tag{35}$$

the probability for subject i to become demented before dying during the interval $]s, s+t]$, given the subject is alive and not suffering from dementia at time s and given his/her responses to the marker(s) under consideration collected up to s, denoted by $\mathcal{H}^Y_i(s)$. The dementia indicator $D_i(s,t) = \mathbb{1}_{(s<T^A_i\leqslant s+t, T^D_i > T^A_i)}$ is equal to 1 for cases and to 0 for controls, at each landmark s. Thus, the set of cases includes subjects alive and free of dementia at

s, who became demented before $s+t$ and before dying while the set of controls includes subjects alive and free of dementia at s, who were still alive and free of dementia at $s+t$ or who died without dementia within the interval $]s, s+t]$.

Besides, the definition of the expected Brier Score can also be extended to dynamic predictions:

$$\begin{aligned} BS(s,t) &= E\big[\big(D(s,t) - \pi(s,t)\big)^2 | T^A > s, T^D > s\big] \\ &= E\big[\big(E(D(s,t)|\mathcal{H}^Y(s)] - \pi(s,t)\big)^2 | T^A > s, T^D > s\big] \\ &\quad + E\big[\big(D(s,t) - E[D(s,t)|\mathcal{H}^Y(s)]\big)^2 | T^A > s, T^D > s\big] \end{aligned}$$

The first part of the sum above quantifies the calibration of the model, and the second part assesses the inherent discrimination ability of the information of the marker collected up to s.

Blanche et al. [2015] proposed IPCW estimators of the AUC and the Brier Score to handle right-censored time-to-event data. The dementia indicator is not known for subjects who were lost to follow-up during the interval $]s, s+t]$. The weights are then equal to the inverse probability to be alive and free of dementia at s for cases and equal to the inverse probability to be alive and free of dementia at $s+t$ for controls, computed by the non-parametric Kaplan-Meier estimator. The main advantage of these IPCW estimators of AUC and Brier Score is that they are model-free and do not assume that the prognostic joint models are well-specified.

At last, Blanche et al. [2015] also proposed point-wise tests to compare the AUC and the Brier Score of two different prognostic models at each landmark time.

4.3 Results

We estimated the three joint models with G=1, 2, 3 latent classes and retained 3 latent classes for each model. We set up the landmark times to s=0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4 years with a prediction horizon of t =5 years. The corresponding probabilities of Eq.(35) were computed for the subjects in the Three-city validation sample, from the estimations obtained on the Paquid cohort, for each model.

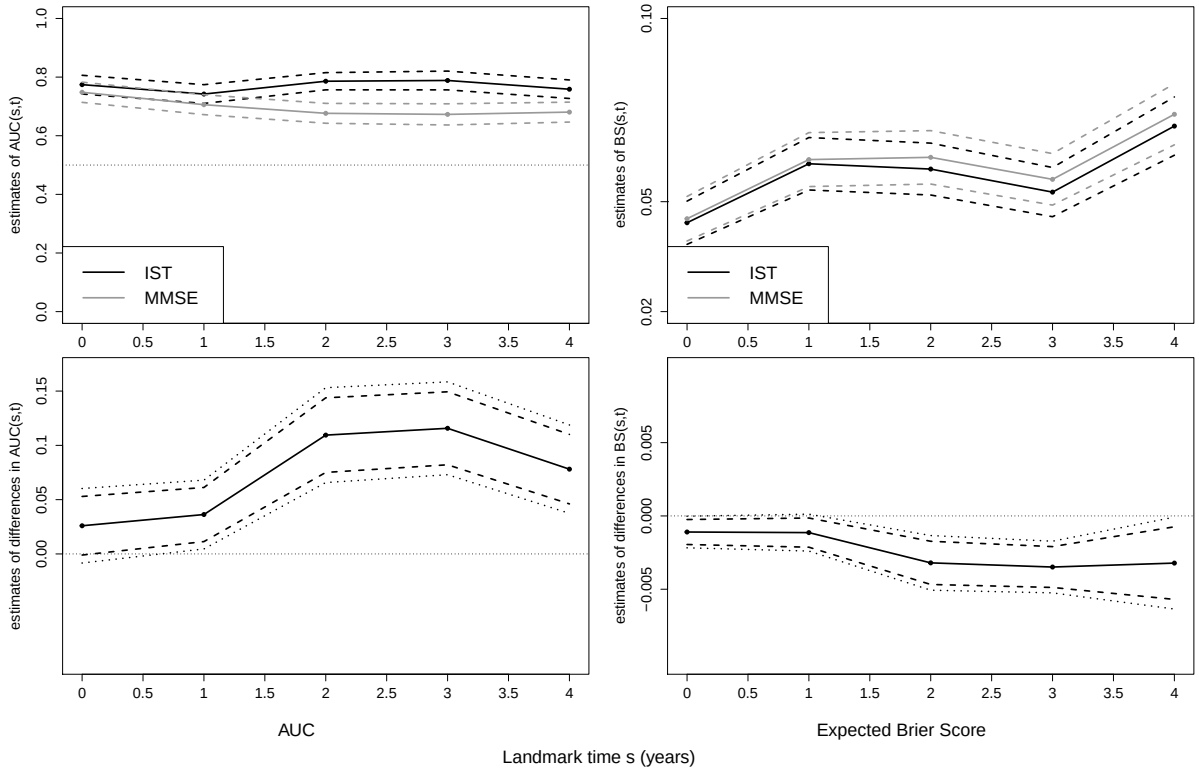


Figure 4.1: Comparison of predictive accuracy of the predicted risks of dementia, based on the IST and on the MMSE, within time window (s, s+t) when s =0, 0.5, 1, 1.5,..., 4 and t = 5 years. 95% point-wise confidence intervals are displayed as dashed lines, 95% simultaneous confidence bands as dotted lines. Three-city data, n = 3,880 subjects.

Figure 4.1 presents a graphical comparison of predictive abilities between the model based on the IST and the model based on the MMSE. The estimated AUCs represented in the top left part show that the IST marker has a better discrimination ability than the MMSE, particularly after a delay of one year. Note that the AUCs do not increase with time, despite the accumulation of collected data. This may be due to the stronger selection of the population with time: the subjects who remain in the study may be more difficult to discriminate, as subjects with high probabilities to become demented already became demented or died. The bottom left part displays the estimated difference between the two AUCs, at each landmark time. The point-wise confidence intervals confirm that the model based on the IST has a significantly better discriminatory ability after a delay of one year. The right part of the figure presents the expected Brier Score (top right part) and the estimated difference of expected Brier Score between the two models (bot-

tom right part). The Brier Score curves show that the model based on the IST is better calibrated from a delay of one year. The increase observed at $s = 4$ years may be due to the selection of the population or to the fact that all the information of the prediction window (from 4 to 9 years) is not collected yet, as the follow-up is considered up to 10 years.

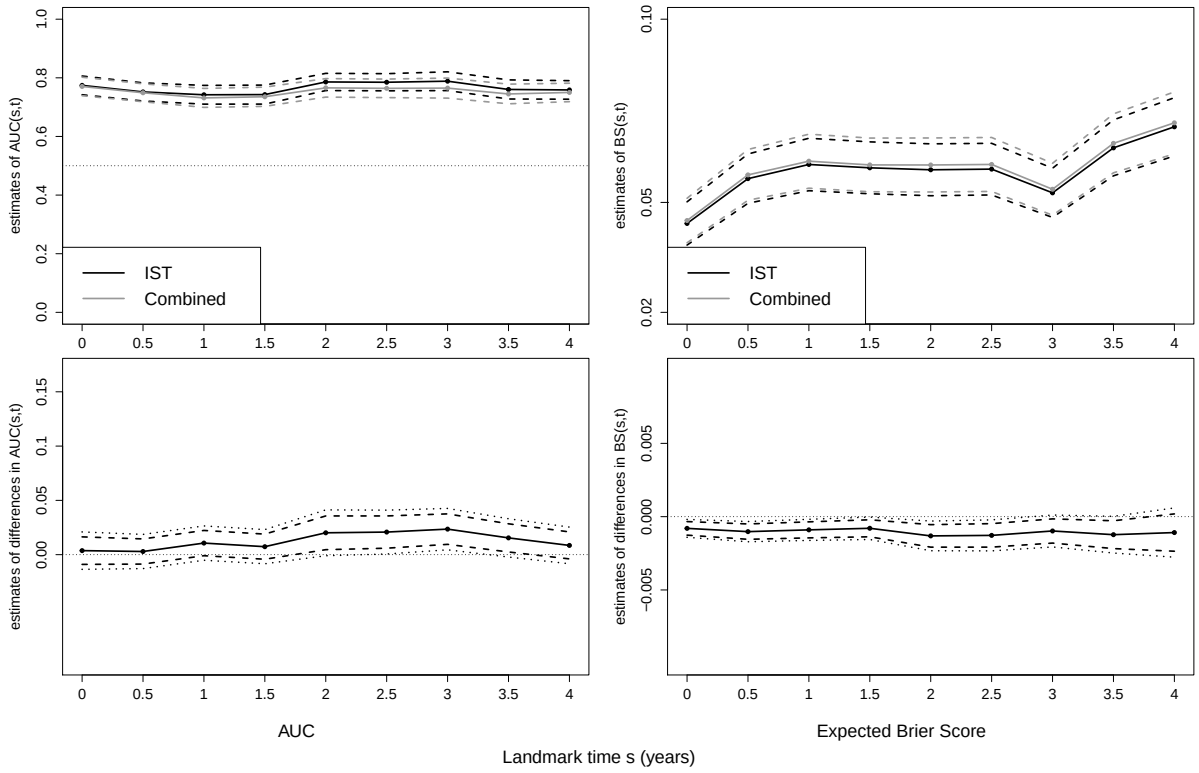


Figure 4.2: Comparison of predictive accuracy of the predicted risks of dementia, based on the IST and on both the IST and the MMSE, within time window (s, s+t) when s =0, 0.5, 1, 1.5,..., 4 and t = 5 years. 95% point-wise confidence intervals are displayed as dashed lines, 95% simultaneous confidence bands as dotted lines. Three-city data, n = 3,880 subjects.

Figure 4.2 presents the comparison between the model based on the IST and the combined model. The AUCs and Brier Scores of the two models are very similar, with a non significant advantage for the model based on the IST. Combining the IST with the MMSE does not improve the predictive ability in terms of discrimination and calibration, compared to the model based on the IST alone.

As expected, the combined model has a better discrimination and calibration ability

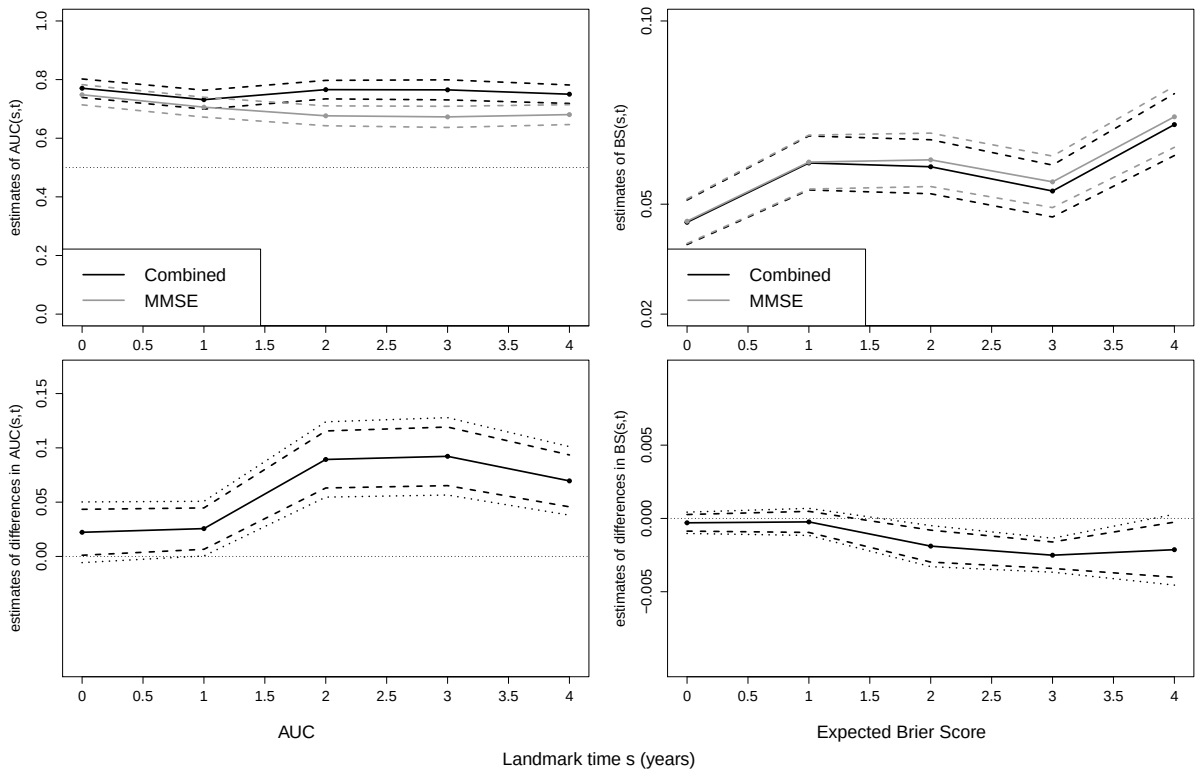


Figure 4.3: Comparison of predictive accuracy of the predicted risks of dementia, based on the MMSE and on both the IST and the MMSE, within time window (s, s+t) when s =0, 0.5, 1, 1.5,..., 4 and t = 5 years. 95% point-wise confidence intervals are displayed as dashed lines, 95% simultaneous confidence bands as dotted lines. Three-city data, n = 3,880 subjects.

than the model based on the MMSE, as shown in Figure 4.3.

4.4 Discussion

We compared the predictive abilities of joint latent class models, accounting for interval censoring and semi-competing risks, based on repeated measures of different cognitive tests. The initial objective was to compare the models based on the IST, the MMSE and the BVRT with the model combining the three tests, that is why we selected subjects who completed the BVRT at least once. We first focused on the two tests retained by Blanche et al. [2015] in order to compare our results. This work is in progress but the current results confirm that the IST is more suitable than the MMSE to predict dementia and also suggest that combining the MMSE and the IST is not better in terms of predictive ability than the IST alone.

However, we made some assumptions in the combined model which could be made more flexible. First, the effects of the covariates on the latent process are assumed common to the two markers. This assumption is likely too restrictive, given the differences in estimations between the three models. Also, we assumed that the variability of the random intercept was common to the two markers. Nevertheless, in a supplementary analysis, we accounted for a marker-specific inter-subject variability and this yielded similar results. At last, the comparisons of the three models were made with three latent classes for each model as in Blanche et al. [2015], but analyses will be continued to choose the optimal number of classes for each marker. We also intend to combine other tests such as the BVRT.

We considered the delay from entry as time scale in order to gather a sufficient number of cases in a 5-year prediction window at each landmark time. When dealing with age as time scale, the predictive assessment can get difficult as the range of ages at entry is relatively wide. Thus, we may not have enough subjects who became demented within the same 5-year age interval and we may obtain large confidence intervals of IPCW estimators, also resulting in a lower power of the point-wise test.

We obtained worse AUC and Brier Scores than the ones obtained by Blanche et al. [2015] with a joint three-latent-class model for competing risks, not accounting for interval censoring. We did not expect to have better predictive abilities but such a difference is not

negligible (mean AUC=0.8 in our analysis for IST versus 0.9 in Blanche et al. [2015]). The first difference between the two analyses is the training sample: in our analysis, we did not exclude deaf subjects, blind or confined to bed as in Blanche et al., and we included only subjects who completed at least the BVRT once before the dementia diagnosis, death or right-censoring. In total, our training set contains 400 fewer subjects and the follow-up was 5-year longer. Note also that the selection procedure is different in our training and in our validation samples.

Second, Blanche et al. analyzed directly the repeated measurements values of IST and the normalized MMSE while we estimated non-linear transformations. These transformations may be close to the identity function as the IST and the normalized version of the MMSE are gaussian. However, if the distributions of these tests are different in the Paquid and in the Three-city cohorts, the estimated transformations on the Paquid cohort may not be transferable to the Three-city cohort and this may alter the preditive abilities of the model.

Third, the number of classes is not optimized in our study and it would be useful to run the three models with more latent classes to check that the heterogeneity in cognitive decline and risk of dementia is correctly accounted for.

Fourth, interval censoring is not treated in the same way in the training sample and in the validation sample. In the latter set, we imputed the middle of the censoring interval for subjects diagnosed with dementia and we used an imputation rule for subjects who died before the dementia diagnosis. Jacqmin-Gadda et al. [2014] compared the properties of the IPCW estimators of AUC based on this imputation rule with an estimator based on an illness-death model accounting for interval censoring and showed that there was little difference, provided the illness-death model was well-specified. However, imputation may artificially increase information in both the training and the validation sets in Blanche et al.'s analysis.

As the objective of this work was to compare the predictive abilities of the different markers, we estimated the three joint latent class models on the same training sample from the Paquid cohort, which included only subjects who had at least performed one IST, one MMSE and one BVRT before the dementia diagnosis, death or right-censoring. However, when the purpose is to build the best predictive model, it is more relevant to

estimate the joint latent class model on the biggest training set available. Thus, the combined model could be estimated on a larger sample, including subjects who performed at least one of the different tests before the dementia diagnosis.

As a conclusion, accounting for interval censoring in predictive ability assessment is difficult: either we use arbitrary imputations, or we estimate the AUC with the same joint multi-state model than the one used in the estimation phase, which accounts for interval censoring. The comparison between the two methods, within the framework of a fixed marker, showed little difference [Jacqmin-Gadda et al., 2014]. However, as the model may be misspecified within the joint modeling framework, we preferred the non-parametric approach.

5 Interpretation of mixed models and marginal models with cohort attrition due to death and drop-out

Anaïs Rouanet[1,2,*], Jean-François Dartigues[1,2], Catherine Helmer[1,2], Hélène Jacqmin-Gadda[1,2]

[1]INSERM, Centre INSERM U1219 - Epidemiologie - Biostatistiques, F-33076 Bordeaux, France

[2]Université de Bordeaux, ISPED, 146 rue Léo Saignat, F-33076 Bordeaux, France

Submitted

Background: The most frequently used models for the analysis of longitudinal data with attrition due to death and drop-out are mixed models, joint models (for repeated measures of the marker and time-to-death) and marginal models estimated either by weighted or unweighted Generalized Estimating Equations (GEE). Our objective is to explain the interpretation and compare estimates from these methods under different assumptions regarding the death process and the drop-out process.

Methods: We compared through simulations maximum likelihood estimates (MLE) of mixed models and joint models and GEE with independent working correlation structure (IEE) and weighted IEE for marginal models. We considered two weighting methods: by the inverse probability to be observed (WIEE1) and by the inverse probability to be observed given the subject is alive (WIEE2). These models were then applied to the Paquid cohort to estimate the cognitive decline in the elderly.

Results: Estimates from mixed models and joint models may be interpreted as subject-specific estimates among the population alive while their population-averaged interpretation is possible only in an immortal cohort. IEE provides population-averaged estimates in the currently observed population, while WIEE2 provides population-averaged estimates in the population currently alive and WIEE1 provides population-averaged estimates in the immortal population.

Conclusions: When the follow-up may be truncated by death and the interest is in the impact of covariates on the individual change, we recommend mixed model or joint model analyses. When the interest is in the impact of covariates on the change in the population mean of survivors, we recommend WIEE2.

Keywords: Death, Immortal cohort, GEE, Mixed models, Partly conditional, Population-averaged, Subject-specific.

5.1 Introduction

Mixed models estimated by Maximum Likelihood [1] and marginal models estimated by Generalized Estimating Equations (GEE) [2] are the two main methods for the analysis of longitudinal data in epidemiology. Their differences regarding parameter interpretation led to an abundant literature [3; 4; 5]. Subject-specific parameters from mixed models represent effects of covariates on the individual change while population-averaged parameters from marginal models represent effects of covariates on the change in the population mean. The former is useful to measure the etiologic effect of the exposure while the latter measures the change in the population mean if the exposure disappeared from the population. In this paper, we focus on the linear case as this is the unique case where parameters of mixed models have both interpretations.

To analyze cohort studies prone to attrition due to drop-out, mixed models estimated on available data are the recommended approach because these estimators are robust to missing at random data (MAR) while GEE estimators from marginal models are unbiased only if the missing data are completely at random (MCAR). The MCAR and MAR assumptions state that the probability to drop out depends on covariates only, or on both covariates and observed past value of the outcome, respectively. However, weighted GEE approaches have been proposed [6; 7; 8] to provide unbiased population-averaged estimates under the MAR assumption. When missing data are at random, mixed models and weighted GEE provide estimates that would have been obtained without drop-out. Missing data are not at random (MNAR or informative) when the drop-out probability may depend on unobserved characteristics of the marker trajectory (for instance current true value or current slope) given past observed marker values and covariates. In this case, joint models of the longitudinal marker and time-to-drop-out, which combines a mixed model and a time-to-event model, may lead to unbiased estimates if the dependence structure between the two outcomes is well-specified.

When attrition is due to death, as in cohort of elderly subjects, the use of mixed models is debated because they estimate the unconditional expectation of the marker, which would have been observed in an immortal cohort, if nobody had died (i.e. initial cohort)

[9; 10]. Indeed, by modeling intra-subject correlation, maximum likelihood estimation (MLE) implicitly imputes data after death [9]. The same argument may be applied for the estimates from the mixed sub-model in a joint model suggesting that they provide an unbiased estimation of the trend in an immortal cohort, when the risk of death depends on unobserved characteristics of the longitudinal trajectory of the marker. Most authors consider that the target estimand should be the expectation of the marker given the subject is currently alive, also called partly conditional expectation (or mortal cohort inference), which can be estimated by GEE with an independence working correlation, denoted IEE [9; 10; 11; 12]. When attrition is due to death and drop-out, Dufouil et al. [11] target the same estimand by weighting IEE by the inverse probability to be observed given the subject is alive. Nevertheless, in the literature, other weights defined as the inverse probability to be observed (i.e. to be alive and observed) are used [13].

The objective of this paper is to clarify the interpretation of parameters from linear mixed models and joint models estimated by MLE, in longitudinal studies prone to attrition due to death and possibly drop-out. Our main point is that MLE parameters may be interpreted as subject-specific effects in the population alive whereas their population-averaged interpretation is possible in an immortal cohort only. We compare these estimates with unweighted and weighted IEE estimates. We show in a simulation study and by applying the different methods to the Paquid cohort of elderly that unweighted IEE provides the population-averaged estimates among the currently observed population, IEE weighted by the inverse probability to be observed given that the subject is alive provides the population-averaged estimates in the currently alive population whereas IEE weighted by the inverse probability to be observed (and alive) provides the population-averaged estimates in an immortal population.

5.2 Notations

For each subject i, $i = 1, \ldots, N$ we denote by $\tilde{Y}_i(t)$ the true marker value defined at any time t and by Y_{ij} the measure of the marker at visit times t_{ij}, $j = 1, \ldots, n_i$. Then X_{ij} is the set of covariates at visit j, T_{1i} the time-to-death and T_{2i} the time-to-drop-out.

We describe the marker trajectory by a linear mixed model:

$$Y_{ij} = \tilde{Y}_i(t_{ij}) + \varepsilon_{ij} = X_{ij}^T\beta + Z_{ij}^T\boldsymbol{u}_i + \varepsilon_{ij} \tag{1}$$

with $\boldsymbol{u}_i \sim N(0, B)$ random effects and ε_{ij} the measurement error at occasion j with $\varepsilon_{ij} \sim N(0, \sigma^2)$.

We assume that the longitudinal follow-up may be stopped by death or drop-out only, excluding intermittent missing data, and that covariate values X_{ij} are known at any visit times among subjects alive (irrespective of the observation of Y_{ij}). Following Kurland and Heagerty [9], the drop-out mechanism is defined as MCAR when the probability to be observed given the subject is alive is independent from the future time of death s and the marker values:

$$\begin{aligned} S_{2i}(t) &= P\left(T_{2i} > t | T_{1i} = s, H_i^X(s), H_i^Y(s), H_i^{\tilde{Y}}(s)\right) \\ &= P\left(T_{2i} > t | T_{1i} > t, H_i^X(t)\right) \qquad \text{with } t < s \end{aligned}$$

with $H_i^X(s) = \{X_{ij}, t_{ij} < s\}$ and $H_i^Y(s) = \{Y_{ij}, t_{ij} < s\}$ the sets of covariates and marker measures, respectively, collected for subject i at all visits before time t, and $H_i^{\tilde{Y}}(s) = \left\{\tilde{Y}_i(u), u \leq s\right\}$ the history of the true marker values up to and including time t. The drop-out process is MAR when the probability to be observed given the subject is alive is independent from the future time of death and the unobserved values of the marker:

$$\begin{aligned} S_{2i}(t) &= P\left(T_{2i} > t | T_{1i} = s, H_i^X(s), H_i^Y(s), H_i^{\tilde{Y}}(s)\right) \\ &= P\left(T_{2i} > t | T_{1i} > t, H_i^X(t), H_i^Y(t)\right) \end{aligned}$$

Similarly, we define the death process as completely at random (DCAR) or at random (DAR) if the survival probability $S_{1i}(t) = P(T_{1i} > t)$ is a function of, respectively¸ the history of covariate values until t, or a function of the histories of both the covariate and marker values until t. The death process is not at random (DNAR) when S_{1i} is a function of the histories of the covariate values and observed marker values until t and the history of true marker values up to and including t.

5.3 Target estimands with death and drop-out

One of the estimands of interest is $E(Y_{ij}|X_{ij})$, the population-averaged mean of the marker in the immortal cohort, which would have been observed if subjects were not prone to death. Parameters in model (1) have this population-averaged interpretation since:

$$\beta = E(Y_{ij}|X_{ij} = 1) - E(Y_{ij}|X_{ij} = 0)$$

Thus, β represents the change in the population mean of Y associated with a unit change in X.

When considering the individual change, one can be interested in $E(Y_{ij}|X_{ij}, u_i)$ which represents the subject-specific mean in the immortal population. Regression parameters in model (1) also have the subject-specific interpretation as:

$$\beta = E(Y_{ij}|X_{ij} = 1, u_i) - E(Y_{ij}|X_{ij} = 0, u_i)$$

Thus, β also represents the individual change in the immortal population associated with a unit change in X, adjusting for u_i that stands for the subject-specific unmeasured covariates considered independent from the observed covariates [14].

As the immortal population, with no risk of dying, is not realistic, it is of interest to focus on the same expectations among the population currently alive, said 'partly conditional' expectations [9]. The subject-specific expectation given the subject is currently alive is:

$$E(Y_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}) = X_{ij}^T\beta + Z_{ij}^T\boldsymbol{u}_i + E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij})$$

We demonstrate in Web Appendix 1 that $E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}) = 0$ whatever the death mechanism (DCAR, DAR and DNAR when the risk of death depends on the histories of covariate and true marker values only). Hence the subject-specific expectation of Y in the immortal cohort equals the partly conditional subject-specific expectation and β can also be interpreted as the individual change in Y associated with a unit change in X among subjects currently alive.

Besides, the population-averaged expectation given the subject is currently alive can be derived from Eq. (1):

$$E\left(Y_{ij}|X_{ij}, T_{1i} > t_{ij}\right) = X_{ij}^T\beta + Z_{ij}^T E\left(u_i|X_{ij}, T_{1i} > t_{ij}\right) + E\left(\varepsilon_{ij}|X_{ij}, T_{1i} > t_{ij}\right)$$

In both DAR and DNAR frameworks, the risk of death depends on u_i, either through the past values of Y (DAR), through the expected true current value $\tilde{Y}$ or directly on u_i (DNAR). Thus, $E(u_i|X_{ij}, T_{1i} > t_{ij}) \neq 0$ and $E(Y_{ij}|X_{ij}, T_{1i} > t_{ij}) \neq E(Y_{ij}|X_{ij}) = X_{ij}^T\beta$. Thus, the unconditional and the partly conditional population-averaged expectations are different and the parameter β from a linear mixed model has a population-averaged interpretation only in an immortal cohort. It represents the impact of a change in X on the population mean of Y whether subjects were not permitted to die.

Based on McCulloch et al. [15], we propose in Web Appendix 2 an analytical approximation of the population-averaged partly conditional expectation of the marker $E(Y_{ij}|X_{ij}, T_{1i} > t_{ij})$ assuming a proportional hazards model for death depending on X and the random effects u_i (DNAR assumption). When the unconditional population-averaged trajectory is linear on time and does not depend on X, the population-averaged expectation among subjects currently alive turns to have a quadratic time-trend and to be affected by a change in X, due to selection by death.

We concur with most authors that the objective of longitudinal data analyses is to estimate the mean trajectory, or the impact of covariates, among subjects alive. Thus, when interested in the population-averaged effect of predictors, mixed models or joint models are not appropriate. However, for most real data analyses, the target estimand is $E(Y_{ij}|X_{ij}, u_i, T_{1i} > t_{ij})$ rather than $E(Y_{ij}|X_{ij}, T_{1i} > t_{ij})$ as the analyses aim at estimating the association between a change in X and the individual change in Y among subjects alive rather than the change in the population mean.

Two other possible estimands $E(Y_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}, T_{2i} > t_{ij})$ and $E(Y_{ij}|X_{ij}, T_{1i} > t_{ij}, T_{2i} > t_{ij})$, represent respectively the subject-specific and population-averaged expectations among the subjects currently alive and observed, but usually they are not the target of analyses.

5.4 Standard estimators with death and drop-out

When the follow-up may be terminated by death and drop-out in the DAR/MAR framework, the MLE of linear mixed models are unbiased and represent the population-averaged and subject-specific effects of covariates in the immortal cohort, as likelihood maximisation procedure is equivalent to imputing data after drop-out and death [9]. Since unconditional and partly conditional subject-specific expectations are equal as demonstrated above, these estimates can also be interpreted as subject-specific effects among subjects alive. When the risk of death may depend on unobserved characteristics of the marker trajectory (DNAR), these estimates may be biased as mixed models are not robust in this framework. However, the MLE of joint models are unbiased when the correlation between the marker and the time-to-event is modeled correctly. Thus, they estimate both unconditional population-averaged and subject-specific effects, as well as partly conditional subject-specific effects.

When attrition is due to death and drop-out, marginal models estimated by unweighted IEE estimate the expectation $E(Y_{ij}|X_{ij}, T_{1i} > t_{ij}, T_{2i} > t_{ij})$ among subjects currently alive and observed [9; 10]. In the MAR/DAR case, Weuve et al. [13] used a weighted GEE approach, with time-dependent weights equal to the inverse probability to be currently alive and observed:

$$w_{ij}^{(1)} = \frac{1}{P\left(T_{2i} > t_{ij}, T_{1i} > t_{ij} | H_i^X(t_{ij}), H_i^Y(t_{ij})\right)}$$

With the independence working correlation structure, this weighted IEE (WIEE1) estimates the population mean in an immortal cohort without drop-out [9], $E(Y_{ij}|X_{ij})$ which is rarely the target estimand.

To correct for drop-out only, Dufouil et al. [11] proposed another weighted IEE with time-dependent weights equal to the inverse probability to be currently observed given the subject is alive:

$$w_{ij}^{(2)} = \frac{1}{P\left(T_{2i} > t_{ij} | T_{1i} > t_{ij}, H_i^X(t_{ij}), H_i^Y(t_{ij})\right)}$$

This weighted IEE (WIEE2) provides estimates of the population mean among subjects alive that would have been observed without drop-out: $E\left(Y_{ij}|X_{ij}, T_{1i} > t_{ij}\right)$ [9; 11].

5.5 Simulations

In order to check empirically the above interpretations and to quantify the differences between the unconditional and partly conditional estimates, we carried out several sets of simulations according to different mechanisms for drop-out and death.

For each scenario, 500 datasets were generated. Longitudinal data were generated by a linear mixed model: $Y_{ij} = 20 - 0.3t_{ij} + u_{i0} + u_{i1}t_{ij} + \varepsilon_{ij}$ with $u_i \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.3 & -0.1\\-0.1 & 0.1\end{bmatrix}\right)$, $\varepsilon_{ij} \sim N\left(0, 0.9^2\right)$ for $i = 1, \ldots, 500$ subjects and $t_{ij} = 0, 4, 12, 16, 20$ year.

Within non-informative frameworks, times to death and drop-out were successively generated by logistic models:

$$\text{logit}\left(P(T_{1i} > t_{ij} | T_{1i} > t_{ij-1})\right) = \eta_0 + \eta_1 X_i + \eta_2 Y_{ij-1}$$
$$\text{logit}\left(P(T_{2i} > t_{ij} | T_{1i} > t_{ij}, T_{2i} > t_{ij-1})\right) = \gamma_0 + \gamma_1 X_i + \gamma_2 Y_{ij-1}$$

The binary covariate X had a Bernoulli distribution with probability 50%. When death and drop-out were DCAR/MCAR or DAR/MAR, parameter values were equal to (η_0, η_1, η_2, γ_0, γ_1, γ_2) = (2.5, 0.5, 0, 2.5, 0.5, 0) or (η_0, η_1, η_2, γ_0, γ_1, γ_2) = (-0.5, 0.5, 0.15, 0.5, 0.5, 0.1) respectively. Within the DNAR framework, time-to-death was generated by a proportional hazards model with a constant baseline hazard depending on X and on the random slope from the mixed model:

$$\lambda(t) = 0.05 * \exp\left(-X_i - u_{1i}\right)$$

At last, the marker values were truncated at the first time-to-event met by each subject.

We compared maximum likelihood estimates of the mixed model, with marginal models estimated by unweighted IEE, WIEE1 or WIEE2 as well as maximum likelihood estimates of joint models in the DNAR case. While marker values were generated with a linear time-trend without exposure effect, all estimated models accounted for a possible quadratic time-trend and an effect of X, with interactions with t and t^2. Weights were estimated by pooled logistic regressions using the product from $k = 2$ to j of $P\left(T_{2i} > t_{ik} | T_{1i} > t_{ik}, T_{2i} > t_{ik-1}, X_i, Y_{ik-1}\right) P\left(T_{1i} > t_{ik} | T_{1i} > t_{ik-1}, T_{2i} > t_{ik-1}, X_i, Y_{ik-1}\right)$ for

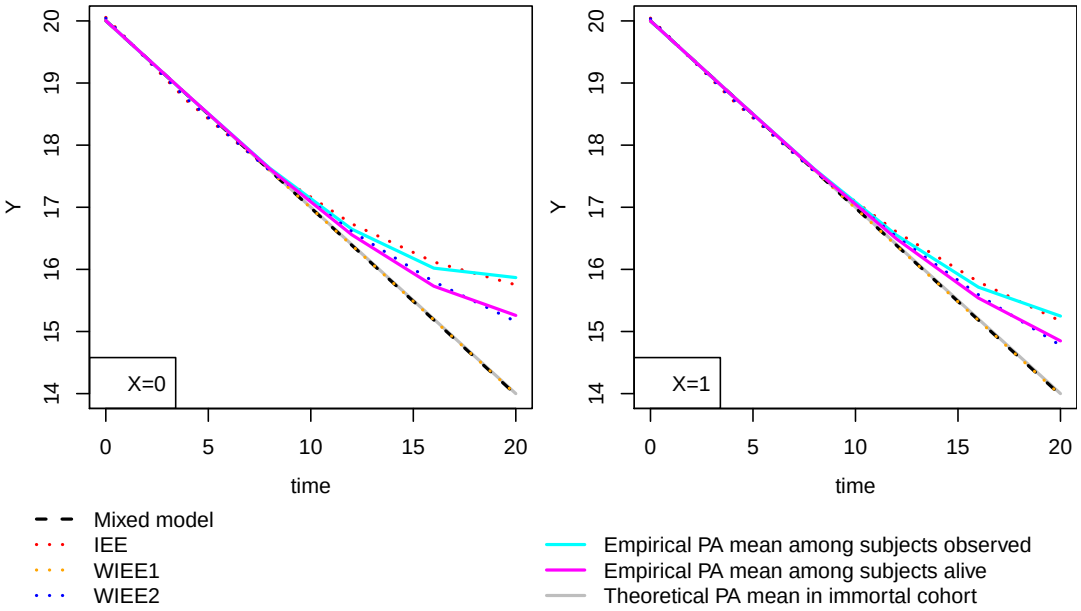
$w_{ij}^{(1)}$ and of $P\left(T_{2i} > t_{ik} | T_{1i} > t_{ik}, T_{2i} > t_{ik-1}, \ X_i, Y_{ik-1}\right)$ for $w_{ij}^{(2)}$ [9; 11]. Mixed models, IEE and joint models were estimated with R packages NLME, geeM and JM respectively. The main results are presented below and complementary simulations are displayed in Web Appendix 3.

When the risks of death and drop-out depend only on covariates (DCAR/MCAR), as expected, estimated parameters and estimated mean trajectories of all the methods are almost equal confirming that subject-specific and population-averaged estimates are identical in this case (Web Figure 5.5 and Web Table 5.4).

When the risks of death and drop-out depend on past marker values (DAR/MAR), Figure 5.1 displays the empirical means at each time among subjects currently observed, subjects currently alive and in the immortal cohort. These curves are different, the first two exhibiting a non-linear time-trend as death and drop-out led to a selection of the sample, subjects with lower marker values leaving the cohort and dying earlier. These three empirical means are well fitted by the marginal models estimated respectively by unweighted IEE, WIEE2 and WIEE1. The mixed model also fits well the empirical population mean among the immortal population.

Table 5.1 displays the estimates in the DAR/MAR framework. As expected, the mixed model and WIEE1 both lead to unbiased estimates of the parameters of the generated model but estimates from WIEE1 have larger variances due to the use of the independence working correlation. On the other hand, unweighted IEE, that estimates the population mean among subjects observed, finds a quadratic time-trend (significant in 96% of the samples) and a significant interaction between X and t^2 in 14% of the samples. WIEE2 that estimates the population mean among subjects alive also highlights a quadratic time-trend (power=63%) but the interaction between X and t^2 is generally not significant. Indeed, WIEE2 estimates are impacted by selection by death only while unweighted IEE highlights the selection effect of both death and drop-out. When the risk of death depends on an interaction between the exposure X and the previous marker value Y_{ij-1}, the estimated interaction terms Xt and Xt^2 are larger for both unweighted IEE and WIEE2 (Web Figure 5.6 and Web Table 5.5).

Figure 5.1: Empirical means and estimated expectations, for $X = 0$ and $X = 1$, of the mixed model and marginal models using IEE, WIEE1 WIEE2, when death and drop-out mechanisms are at random (DAR/MAR): logit $(P(T_{1i} > t_{ij}|T_{1i} > t_{ij-1})) = -0.5 + 0.5X_i + 0.15Y_{ij-1}$ and logit $(P(T_{2i} > t_{ij}|T_{1i} > t_{ij}, T_{2i} > t_{ij-1})) = 0.5 + 0.5X_i + 0.1Y_{ij-1}$ $(N = 500)$.



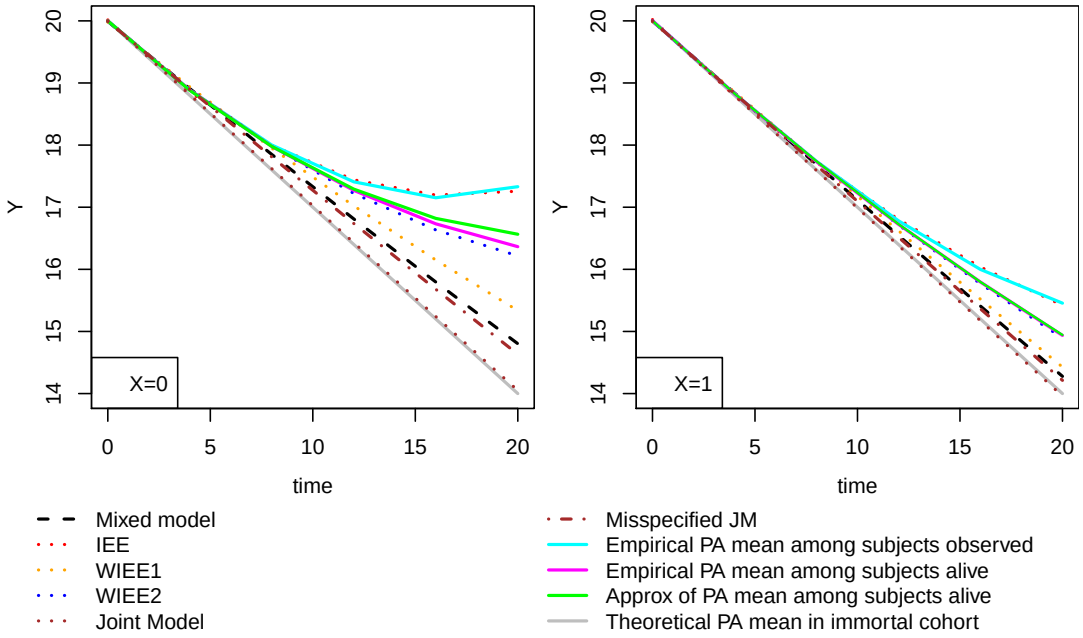
When the risk of death depends on the unobserved random slope of the marker (DNAR) while the risk of drop-out depends only on the last observed marker value (MAR), unweighted IEE and WIEE2 find a significant interaction term Xt^2 in 32% and 12% of the samples respectively. Besides, the mixed model and WIEE1 are biased (Table 5.2 and Figure 5.2). However, these biases are relatively small because the time-trend is linear and the number of observed values is large enough to estimate individual random slopes of subjects who died, without large bias. These biases are corrected by the well-specified joint model. At last, mixed model and WIEE1 estimates from the marginal model differ as the weights are misspecified.

Figure 5.2 shows that the approximation, presented in Web Appendix 2, fits well the population mean among subjects currently alive. Two joint models were estimated, accounting for a dependence of the death process on either the true current value or the

Table 5.1: Estimates of the mixed model and marginal models using IEE, WIEE1 WIEE2, when death and drop-out mechanisms are at random (DAR/MAR): $\text{logit}\left(P(T_{1i} > t_{ij} | T_{1i} > t_{ij-1})\right) = -0.5 + 0.5X_i + 0.15Y_{ij-1}$ and $\text{logit}\left(P(T_{2i} => t_{ij} | T_{1i} > t_{ij}, T_{2i} > t_{ij-1})\right) = 0.5 + 0.5X_i + 0.1Y_{ij-1}$. The last column corresponds to the probability of rejection of the null hypothesis: $\beta = \beta_G$, with β_G the parameter values for data generation in the mixed model. ASE and ESE are respectively the asymptotical and empirical standard errors.

		β_G	$\overline{\hat{\beta}}$	ASE	ESE	Proba rejection H0: $\beta = \beta_G$
Mixed Model	intercept	20	20.005	0.0638	0.0598	0.04
	X	0	-0.005	0.0898	0.0878	0.04
	t	-0.3	-0.302	0.0263	0.0251	0.05
	t^2	0	0.000	0.0009	0.0008	0.04
	Xt	0	0.001	0.0364	0.0369	0.06
	Xt^2	0	0.000	0.0012	0.0011	0.05
IEE	intercept	20	20.056	0.0690	0.0658	0.12
	X	0	-0.013	0.0961	0.0953	0.06
	t	-0.3	-0.369	0.0357	0.0355	0.49
	t^2	0	0.008	0.0022	0.0022	0.96
	Xt	0	0.018	0.0478	0.0488	0.06
	Xt^2	0	-0.002	0.0027	0.0027	0.14
WIEE1	intercept	20	20.007	0.0744	0.0696	0.04
	X	0	-0.004	0.1021	0.0986	0.05
	t	-0.3	-0.303	0.0430	0.0389	0.02
	t^2	0	0.000	0.0030	0.0022	0.01
	Xt	0	0.001	0.0559	0.0519	0.03
	Xt^2	0	0.000	0.0037	0.0028	0.00
WIEE2	intercept	20	20.042	0.0698	0.0665	0.09
	X	0	-0.010	0.0970	0.0958	0.06
	t	-0.3	-0.349	0.0368	0.0355	0.25
	t^2	0	0.005	0.0023	0.0020	0.63
	Xt	0	0.012	0.0490	0.0488	0.06
	Xt^2	0	-0.002	0.0029	0.0026	0.06

Figure 5.2: Empirical means, approximated population-averaged expectation among subjects alive and estimated expectations, for $X = 0$ and $X = 1$, of the mixed model and marginal models using IEE, WIEE1 WIEE2, when the drop-out mechanism is at random (MAR) and death risk depends on the random slope (DNAR): $\text{logit}\left(P(T_{2i} > t_{ij} | T_{1i} > t_{ij}, T_{2i} > t_{ij-1})\right) = -1 + X_i + 0.2Y_{ij-1}$ and $\lambda(T_{1i}) = 0.05 * \exp(-X_i - u_{1i})$ $(N = 500)$.



slope of the marker. The first joint model, which is misspecified, under-estimates the decline like the mixed model, but this is corrected by the well-specified joint model. When the risk of death depends on an interaction between the exposure X and the random slope u_{1i}, the estimated interaction term Xt^2 is larger for both unweighted IEE and WIEE2 (Web Figure 5.7 and Web Table 5.7). Web Figure 5.8 and Web Table 5.7 display the results when the risk of death depends on the true current value of the marker (DNAR), and drop-out leads to MAR data. The differences between β_G and estimates from GEE models are higher, mainly due to a stronger selection by death.

5.6 Application to the Paquid cohort

To illustrate the impact of selection by death, we applied the models to the Paquid cohort [16] set up to characterize the pathological brain ageing. Paquid is a population-based

Table 5.2: Estimates of the mixed model, joint model and marginal models using IEE, WIEE1 WIEE2, when the drop-out mechanism is at random (MAR) and death risk depends on the random slope (DNAR): logit $(P(T_{2i} > t_{ij}|T_{1i} > t_{ij}, T_{2i} > t_{ij-1})) = -1 + X_i + 0.2Y_{ij-1}$ and $\lambda(T_{1i}) = 0.05 * \exp(-X_i - u_{1i})$. The last column corresponds to the probability of rejection of the null hypothesis: $\beta = \beta_G$, with β_G the parameter values for data generation in the mixed model. ASE and ESE are respectively the asymptotical and empirical standard errors.

		β_G	$\overline{\hat{\beta}}$	ASE	ESE	Proba rejection H0: $\beta = \beta_G$
Mixed Model	intercept	20	19.991	0.0646	0.0645	0.05
	X	0	0.005	0.0903	0.0880	0.04
	t	-0.3	-0.273	0.0286	0.0295	0.16
	t^2	0	0.001	0.0011	0.0011	0.09
	Xt	0	-0.017	0.0381	0.0380	0.08
	Xt^2	0	0.000	0.0014	0.0014	0.07
IEE	intercept	20	20.015	0.0693	0.0715	0.06
	X	0	0.006	0.0962	0.0966	0.04
	t	-0.3	-0.329	0.0412	0.0430	0.11
	t^2	0	0.010	0.0029	0.0030	0.91
	Xt	0	0.004	0.0516	0.0526	0.06
	Xt^2	0	-0.005	0.0033	0.0034	0.32
WIEE1	intercept	20	19.980	0.0747	0.0744	0.06
	X	0	0.006	0.1015	0.0999	0.04
	t	-0.3	-0.268	0.0501	0.0470	0.08
	t^2	0	0.002	0.0039	0.0029	0.06
	Xt	0	-0.016	0.0606	0.0561	0.03
	Xt^2	0	-0.001	0.0044	0.0033	0.02
WIEE2	intercept	20	19.994	0.0718	0.0719	0.06
	X	0	0.005	0.0986	0.0977	0.05
	t	-0.3	-0.294	0.0454	0.0443	0.04
	t^2	0	0.005	0.0034	0.0030	0.36
	Xt	0	-0.007	0.0557	0.0533	0.05
	Xt^2	0	-0.003	0.0039	0.0034	0.12
Joint Model	intercept	20	20.001	0.0646	0.0645	0.05
	X	0	0.000	0.0903	0.0883	0.04
	t	-0.3	-0.298	0.0301	0.0308	0.06
	t^2	0	0.000	0.0012	0.0011	0.05
	Xt	0	-0.003	0.0395	0.0390	0.05
	Xt^2	0	0.000	0.0014	0.0014	0.05

cohort study including 3777 subjects aged 65 years and older at baseline and representative of two departments of South-Western France. Subjects were followed every 2 or 3 years with repeated cognitive assessments over 25 years.

We focused on the effect of gender on the change over time of verbal fluency, assessed by the Isaacs Set Test (IST). This test consists in giving up to 10 names of cities, fruits, colors and animals, within 15 seconds for each category, for a total of 40 points. As weighted IEE approaches require only monotone missing data, we reduced the sample to 1,379 subjects (55.1% of women) without intermittent missing data. Along the 25 years of follow-up, 615 subjects dropped out (including 564 who died later), 677 subjects died before dropping out, and 87 were followed until 25 years. The average number of visits was 4.18.

The estimated models included a quadratic time-trend and were adjusted on gender (1 for women and 0 for men), age at baseline Age_0 (in decades and centered on 65 years), educational level (CEP for 'Certificat d'Etudes Primaires' =1 if primary school diploma obtained, 0 otherwise), with interactions of gender with time and time square. Time was defined as the delay since the first visit, in decades. The logistic models for death and drop-out used to compute the weights and the survival sub-model of joint models were adjusted on educational level and age at baseline. The mixed model and joint model included random intercept and slope. Using likelihood ratio tests, we compared several joint models with different dependence structures between the risk of death and the marker trajectory and we retained a dependence on the current true value without interaction with gender.

Figure 5.3 displays the estimated mean curves. The large differences between curves estimated by the mixed model, the joint model and WIEE1 suggest that the death process is informative. The mean curves estimated by IEE and WIEE2 clearly highlight a growing selection in the cohort over time, which is more pronounced among men. Indeed, the risks of death and drop-out are higher among subjects with low cognitive score and among men. The trajectory estimated by the joint model has the steepest decline. Estimates from the joint model show that the hazard ratio for death associated with a one-point-increase of

IST is 0.92 ($p < 0.01\%$).

Figure 5.3: Isaacs estimated expectations of the mixed model and marginal models using IEE, WIEE1 WIEE2, and the joint model, for men ($X = 0$) and women ($X = 1$) with a low level of education (CEP=0) and entered in the study at 65 years old ($N = 1,379$).

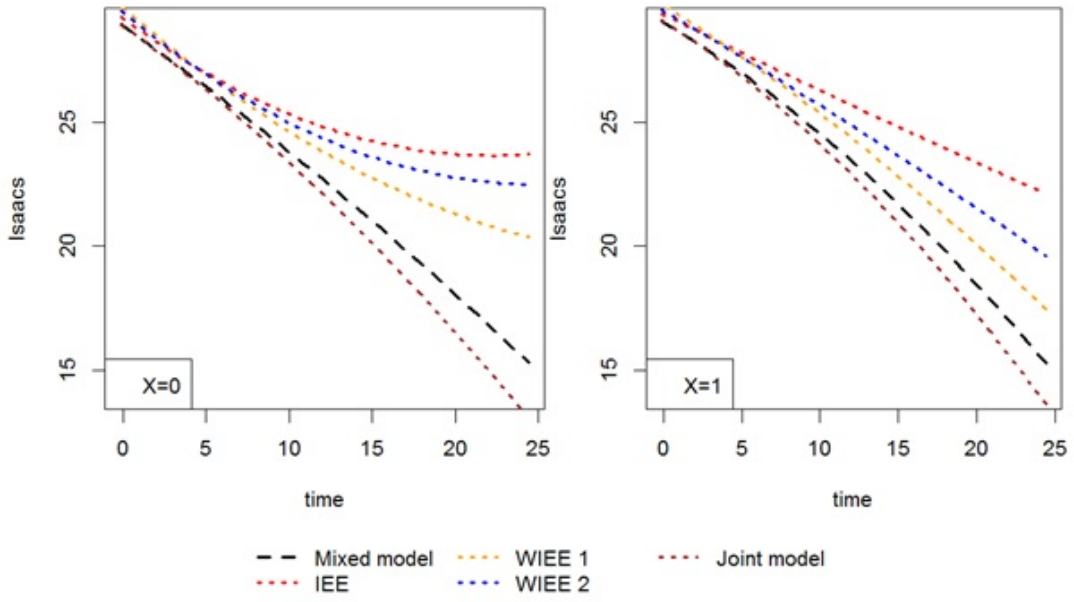


Table 5.3 displays the MLE and IEE estimates. The mixed model, joint model and WIEE1 model show no significant effect of gender on IST evolution. By contrast, the IEE and WIEE2 methods find a significant association of gender with IST change over time. Considering a man and a woman alive, with the same characteristics (same random effects), the joint model results show that both have similar cognitive evolutions. However, WIEE2 suggests that the difference between population means of IST among men and women alive increases over time. Indeed, the risk of death is higher among men, so men survivors tend to be healthier than women survivors even if the rate of cognitive decline is the same among men and women. Note however that WIEE2 estimates are probably biased as the death process appears to be informative.

5.7 Discussion

To summarize these results, we propose the following recommendations. For etiologic studies where the aim is to quantify the impact of an exposure on the individual change,

Table 5.3: Estimates of the mixed model and marginal models using IEE, WIEE1 WIEE2, and the joint model, Paquid ($N = 1,379$).

	MLE					
	Mixed model			Joint Model		
	$\hat{\beta}$	SE($\hat{\beta}$)	p-value	$\hat{\beta}$	SE($\hat{\beta}$)	p-value
intercept	28.88	0.40	0.0000	28.87	0.41	0.0000
Sex	0.15	0.29	0.6208	0.19	0.29	0.5165
Delay	-4.74	0.42	0.0000	-4.81	0.42	0.0000
Delay^2	-0.34	0.23	0.1444	-0.68	0.24	0.0040
Age$_0$	-3.86	0.21	0.0000	-3.91	0.21	0.0000
CEP	4.38	0.31	0.0000	4.42	0.31	0.0000
Sex.delay	1.04	0.55	0.0600	0.85	0.55	0.1230
Sex.delay^2	-0.45	0.29	0.1147	-0.29	0.28	0.3093

	GEE								
	IEE			WIEE1			WIEE2		
	$\hat{\beta}$	SE($\hat{\beta}$)	p-value	$\hat{\beta}$	SE($\hat{\beta}$)	p-value	$\hat{\beta}$	SE($\hat{\beta}$)	p-value
intercept	29.19	0.45	0.0000	29.63	0.51	0.0000	29.43	0.47	0.0000
Sex	0.15	0.30	0.6266	0.08	0.32	0.7958	0.04	0.31	0.8841
Delay	-5.00	0.62	0.0000	-5.86	0.74	0.0000	-5.58	0.67	0.0000
Delay^2	1.13	0.33	0.0006	0.85	0.46	0.0671	1.12	0.38	0.0031
Age$_0$	-3.79	0.24	0.0000	-4.21	0.28	0.0000	-3.99	0.25	0.0000
CEP	4.05	0.35	0.0000	3.97	0.39	0.0000	3.99	0.36	0.0000
Sex.delay	1.88	0.81	0.0197	1.92	1.10	0.0792	1.95	0.92	0.0339
Sex.delay^2	-1.06	0.43	0.0126	-1.29	0.72	0.0741	-1.29	0.55	0.0186

subject-specific effects can be estimated by mixed models when the risk of attrition, whatever the cause (death or drop-out) may be considered as independent from unobserved characteristics of the marker after adjusting on the past observed values (DAR/MAR assumption). If the DAR assumption is unlikely, a joint model can be used, with careful specification of the dependence structure between death and the marker. For pragmatic studies, where the aim is to evaluate the impact of a health intervention on the population mean, the marginal model estimated by WIEE2 weighted by the inverse probability to be observed given that the subject is alive is more appropriate. However, this method is robust under the DAR assumption only.

When the follow-up may be terminated by death or drop-out, we assumed that the drop-out mechanism was MAR. If the drop-out probability depends on unobserved characteristics, for instance the current true value of the marker (MNAR case), to our knowledge, it is not possible to estimate without bias the population-averaged parameters in the population alive. The subject-specific parameters could be estimated by modeling jointly the risks of death and drop-out [17].

Joint models and mixed models may be used with intermittent missing data if they can be considered as MAR. However, computing the weights for WIEE2 requires monotone missing data and should be extended for intermittent missing data. Another limit of weighted IEE is that the package geeM used in this work slightly over-estimates the variance of the estimations as weights are considered as known [18]. However, Dufouil et al. [11] showed that this loss in efficiency was relatively small.

In this work, we focused on linear models because parameters of linear mixed models have both the subject-specific and population-averaged interpretations with complete data. Parameters from non-linear mixed models can only be interpreted as subject-specific. It is easy to demonstrate that the unconditional and partly conditional subject-specific expectations are still identical in the non-linear framework, contrary to the population-averaged expectations. Thus the same recommendations apply. However, WIEE1 estimates will differ from mixed models estimates in the DAR context.

Key Messages

- To quantify the impact of an exposure on the individual change, mixed models estimated by MLE provide unbiased subject-specific effects among subjects currently alive, when the risk of attrition whatever the cause (death or drop-out) may be considered as independent from unobserved characteristics of the marker after adjusting on the past observed values (DAR/MAR assumption).
- When the risk of death may depend on unobserved characteristics of the marker trajectory (DNAR), subject-specific effects may be estimated by joint models.
- To evaluate the impact of an exposure on the population mean among subjects currently alive, marginal models estimated by IEE weighted by the inverse probability to be observed given the subject is alive WIEE2, is appropriate under the DAR/MAR assumption.
- The same recommendations apply with non-linear models.

Acknowledgements

This work was part of the French SMALA project funded by French National Agency for Research and Anaïs Rouanet was funded by an INSERM/Région Aquitaine PhD allocation. The Paquid study is funded by IPSEN and Novartis laboratories and the Caisse Nationale de Solidarité et d'Autonomie. Computer time for this study was provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l'Adour. We thank the members of the MELODEM initiative (Methods in longitudinal dementia research) for helpful discussion.

5.8 References

[1] Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. Biometrics. 1982; 963–974. 120

[2] Liang KY, Zeger SL. Longitudinal Data Analysis using Generalized Linear Models. Biometrika. 1986; 73(1): 13–22. 120

[3] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics. 1988; 44: 1049–1060. 120

[4] Kenward M, Lesaffre E, Molenberghs G. An Application of Maximum Likelihood and Generalized Estimating Equations to the Analysis of Ordinal Data from a Longitudinal Study with Cases Missing at Random. Biometrics. 1994; 945–953. 120

[5] McCulloch CE, Neuhaus JM. Generalized Linear Mixed Models. In: Encyclopedia of Biostatistics; 2005. 2085–2089. 120

[6] Little RJ, Rubin DB. Statistical Analysis with Missing Data. New York: John Wiley & Sons; 1987. 120

[7] Robins JM, Rotnitzky A, Zhao LP. Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. Journal of the American Statistical Association. 1995; 90(429): 106–121. 120

[8] Fitzmaurice GM, Molenberghs G, Lipsitz SR. Regression Models for Longitudinal Binary Responses with Informative Drop-Outs. Journal of the Royal Statistical Society, Series B. 1995; 691–704. 120

[9] Kurland BF, Heagerty PJ. Directly Parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. Biostatistics. 2005; 6(2): 241–258. 121, 122, 123, 125, 127

[10] Kurland BF, Johnson LL, Egleston BL, Diehr PH. Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. Statistical science: a review journal of the Institute of Mathematical Statistics. 2009; 24(2): 211–222. 121, 125

[11] Dufouil C, Brayne C, Clayton D. Analysis of longitudinal studies with death and drop-out: a case study. Statistics in medicine. 2004; 23(14): 2215–2226. 121, 125, 127, 135

[12] Seaman SR, Farewell D, White IR. Linear Increments with Non-monotone Missing Data and Measurement Error. Scandinavian Journal of Statistics. 2016; DOI: 10.1111/sjos.12225. 121

[13] Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. Epidemiology. 2012; 23(1): 119. 121, 125

[14] Gilks W, Wang C, Yvonnet B, Coursaget P. Random-Effects Models for Longitudinal Data Using Gibbs Sampling. Biometrics. 1993; 441–453. 123

[15] McCulloch CE, Neuhaus JM, Olin RL. Biased and unbiased estimation in longitudinal studies with informative visit processes. Biometrics. 2016; PMID: 26990830. 124, 139, 140

[16] Letenneur L, Commenges D, Dartigues JF, Barberger-Gateau P. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. International Journal of Epidemiology. 1994; 23(6): 1256–1261. 130

[17] Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data: with Applications in R. Boca Raton: Chapman and Hall/CRC; 2012. 135

[18] Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Hoboken: John Wiley & Sons; 2011. 135

5.9 Web Supplementary Materials: "Interpretation of mixed models and marginal models with cohort attrition due to death and drop-out"

Anaïs Rouanet[1,2,*], Jean-François Dartigues[1,2], Catherine Helmer[1,2], Hélène Jacqmin-Gadda[1,2]

[1]INSERM, Centre INSERM U1219 - Epidemiologie - Biostatistiques, F-33076 Bordeaux, France
[2]Université de Bordeaux, ISPED, 146 rue Léo Saignat, F-33076 Bordeaux, France

Web Appendix 1: Subject-specific interpretation among subjects alive

In order to demonstrate that unconditional and partly conditional subject-specific expectations are equal, we show that, whatever the death mechanism, $E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}) = 0$ so that:

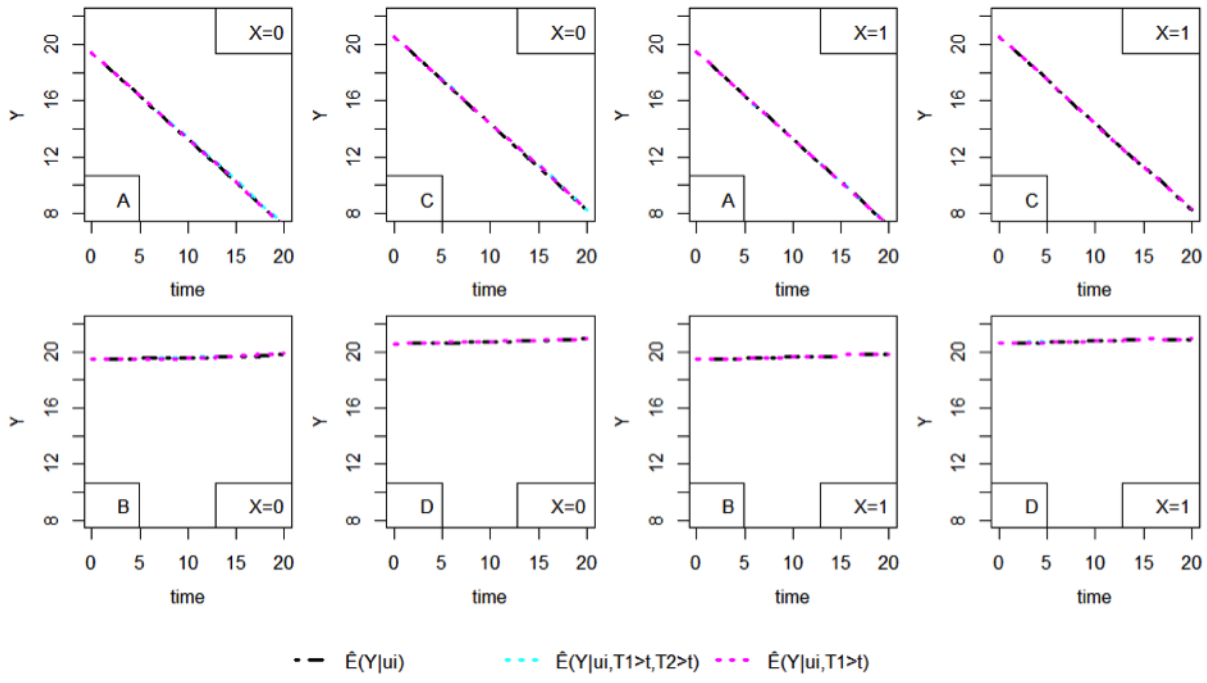
$$E(Y_{ij}|X_{ij}, u_i, T_i > t_{ij}) = X_{ij}^T\beta + Z_{ij}^T u_i + E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}) = E(Y_{ij}|X_{(ij)}, u_i) \quad (1.1)$$

By definition of the mixed model, $E(\varepsilon_{ij}|X_{ij}, u_i) = E(\varepsilon_{ij}) = 0$ as $\varepsilon_{ij} \perp X_{ij}$ and $\varepsilon_{ij} \perp u_i$. When the risk of death depends on covariates only (DCAR case), whatever i and j, $T_{1i} \perp Y_{ij}$ given X_{ij} thus $T_{1i} \perp \varepsilon_{ij}$ given X_{ij} and $E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}) = 0$. When the survival probability $S_1(t_{ij}) = P(T_{1i} > t_{ij})$ depends on past covariate values $H_i^X(t_{ij})$ and past values of the outcome $H_i^Y(t_{ij})$ (DAR assumption), it depends on the past residual errors $\{\varepsilon_{ik}, t_{ik} < t_{ij}\}$, but not on ε_{ij} at the current time if the residual errors at different times are independent. Thus, $E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}) = 0$.

Often, ε_{ij} represents only the measurement error which is typically independent from time to time. In some cases, ε_{ij} is the sum of an auto-correlated process and the measurement error so that ε_{ij} and ε_{ij-1} are not independent. In this case, Equation (1.1) does not strictly hold but $E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij})$ is expected to be close to 0 as shown in the simulation study presented below.

At last, when $S_{1i}(t_{ij})$ depends on past covariate values and unobserved characteristics of the marker trajectory $H_i^{\tilde{Y}}(t_{ij})$ but not on observed values of the marker $H_i^{Y}(t_{ij})$ (restricted DNAR assumption), it depends on the random effects u_i, either directly or through the dependence on the unobserved true value $\tilde{Y}_i(t_{ij})$ and possibly on covariates but it does not depend on ε_{ij}. Consequently, $E(\varepsilon_{ij}|X_{ij}, u_i, T_{1i} > t_{ij}) = E(\varepsilon_{ij}) = 0$ and the equality (1.1) holds.

Web Figure 5.4: Empirical means of the marker Y, on the simulated sample ($N = 500$), among the population currently alive (dashed pink line), currently observed (dashed light blue line), and the immortal population (dash-dotted black line), given X_i and u_i: A) $u_i = (-\sqrt{0.3}; -\sqrt{0.1})$ B) $u_i = (-\sqrt{0.3}; \sqrt{0.1})$ C) $u_i = (\sqrt{0.3}; -\sqrt{0.1})$ D) $u_i = (\sqrt{0.3}; \sqrt{0.1})$.



Web Appendix 2 : Partly conditional population-averaged expectation approximation

We can approximate $E(Y_{ij}|X_{ij}, T_{1i} > t_{ij})$ by applying the results of McCulloch et al. (2016) [15] for informative visiting process. Assuming a log-linear model for the probability of observation, these authors derive the distribution of the random effects given the subject is observed. In the framework of truncation by death, the probability to be

observed at t_{ij} is the survival probability at t_{ij}. We consider the following DNAR case:

$$P(T_{1i} > t_{ij}) = \exp\left(\mu_{ij} + \gamma_{ij}^T u_i\right) \tag{2.1}$$

with μ_{ij} and γ_{ij} time-dependent functions possibly dependent on covariates. Then McCulloch et al. (2016) [15] show that $E(u_i|X_{ij}, T_{1i} > t_{ij}) = B\gamma_{ij}$. Therefore, with independent residual errors,

$$E\left(Y_{ij}|X_{ij}, T_{1i} > t_{ij}\right) = X_{ij}^T\beta + Z_{ij}^T B\gamma_{ij} \tag{2.2}$$

Note also that the conditional distribution of Y_{ij} given $T_{1i} > t_{ij}$ is still gaussian with unchanged variance.

Assuming an additive risk model for death,

$$\lambda(t) = \eta_0(t) + \eta_1^T(t)X_i(t) + \eta_2^T(t)u_i$$

the survival function takes the log-linear form of Eq.(2.1) with $\mu_{ij} = \int_0^{t_{ij}} \eta_0(t) + \eta_1^T(t)X_i(t)\mathrm{d}t$ and $\gamma_{ij} = \int_0^{t_{ij}} \eta_2^T(t)\mathrm{d}t$ and the partly conditional expectation has the form of Eq.(2.2). Therefore, the effect of X on the population mean of the hypothetical immortal cohort will be different from its effect on the partly conditional population mean when X is included in Z_{ij} (random effect on X) or in γ_{ij} (interaction between X and the random effects u_i in the death model).

Using a more standard proportional hazards model for the death risk with time-fixed covariates,

$$\lambda(t) = \lambda_0(t)\exp\left(\eta_1^T X_i + \eta_2^T u_i\right)$$

the cumulative hazard is

$$\Lambda(t) = \Lambda_0(t)\exp\left(\eta_1^T X_i + \eta_2^T u_i\right) = \mu_{it}\exp\left(\eta_2^T u_i\right)$$

with $\Lambda_0(t)$ the cumulative baseline hazard. By first-order Taylor development around $u_i = 0$, we have: $\Lambda(t) \approx \mu_{it} + \mu_{it}\eta_2^T u_i$. Then the survival function has approximately the log-linear form of Eq.(2.1) and the partly conditional expectation may be approximated by formula Eq.(2.2) with $\gamma_{ij} = \Lambda_0(t_{ij})\exp\left(\eta_1^T X_i\right)\eta_2^T$. This suggests that the effect of X on the population mean of the hypothetical immortal cohort will be different from its effect on the partly conditional population mean as soon as X and u_i are predictors of death.

As an example, let consider a standard mixed model with linear time-trend, and random intercept and slope and assuming that the risk of death depends on the current slope and one fixed covariate which is not associated with the marker Y:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + u_i 0 + u_{i1} t_{ij} + \varepsilon_{ij}$$

$$\lambda(t) = \lambda_0 \exp\left(\eta_1 X_i + \eta_2 u_{i1}\right)$$

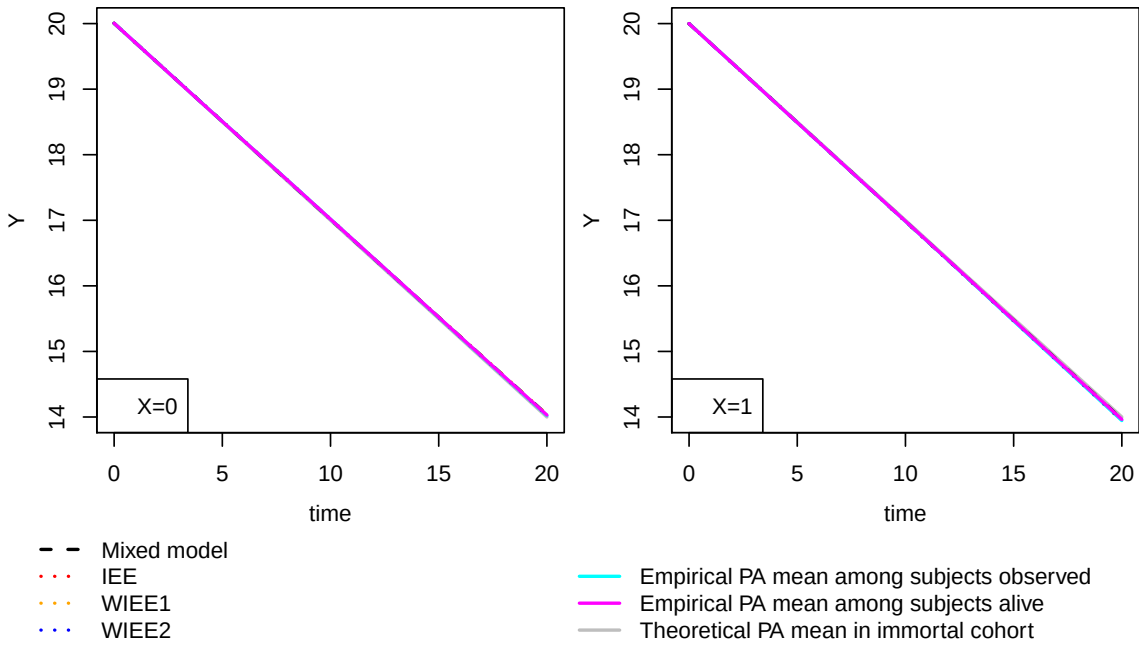
Applying the results above, we found that:

$$E\left(Y_{ij} | X_i, T_{1i} > s\right) \approx \beta_0 + \beta_1 t_{ij} + \eta_2 \Lambda_0 \exp\left(\eta_1 X_i\right)\left(B_{01} t_{ij} + B_{11} t_{ij}^2\right)$$

The population mean among subjects currently alive has thus a quadratic time-trend and depends on X while the time-trend for the population mean in the immortal cohort is linear and independent on X.

Web Appendix 3 : Additional simulations

Web Figure 5.5: Empirical means and estimated expectations of the mixed model and marginal models using IEE, WIEE1 WIEE2, when death and drop-out mechanisms are completely at random[1] (DCAR/MCAR), for $X = 0$ and $X = 1$ ($N = 500$).



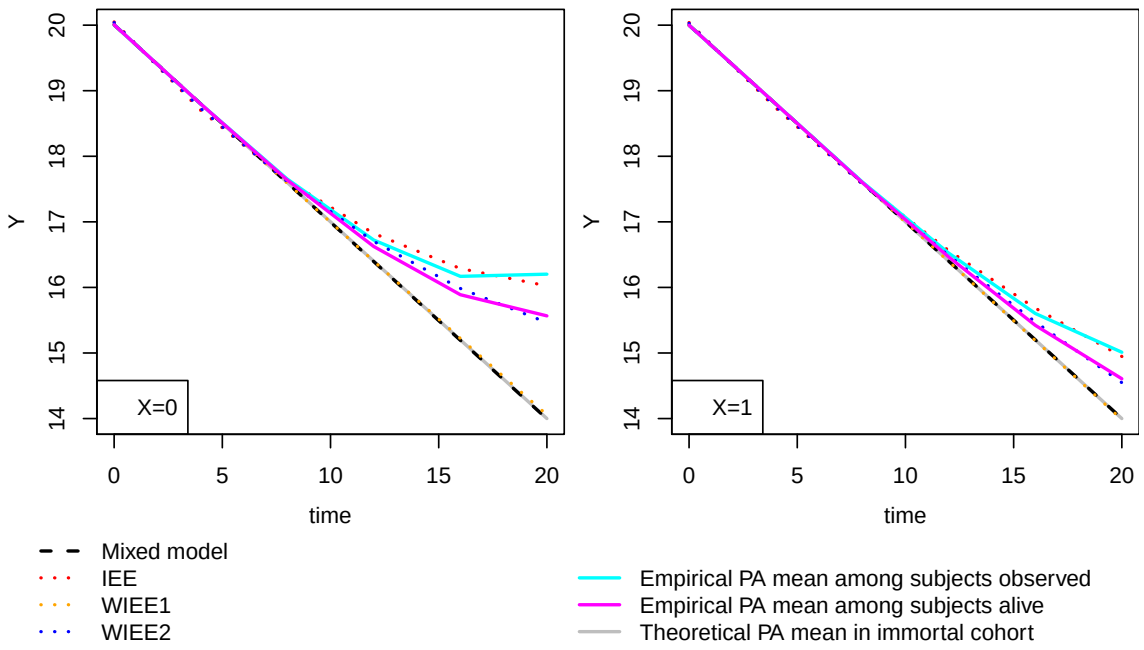
1. Generation models: $\text{logit}\left(P(T_{1i} > t_{ij} | T_{1i} > t_{ij-1})\right) = 2.5 + 0.5 X_i$

$\text{logit}\left(P(T_{2i} > t_{ij} | T_{1i} > t_{ij}, T_{2i} > t_{ij-1})\right) = 2.5 + 0.5 X_i$

Web Table 5.4: Estimates of the mixed model and marginal models using IEE, WIEE1 WIEE2 when death and drop-out mechanisms are completely at random (DCAR/M-CAR)[1]. The last column corresponds to the probability of rejection of the null hypothesis: $\beta = \beta_G$, with β_G the parameter values for data generation in the mixed model. ASE and ESE are respectively the asymptotical and empirical standard errors.

		β_G	$\overline{\hat{\beta}}$	ASE	ESE	Proba rejection H0: $\beta = \beta_G$
Mixed Model	intercept	20	20.004	0.0636	0.0591	0.04
	X	0	-0.007	0.0896	0.0866	0.04
	t	-0.3	-0.300	0.0259	0.0250	0.04
	t^2	0	0.000	0.0008	0.0008	0.05
	Xt	0	-0.002	0.0360	0.0342	0.03
	Xt^2	0	0.000	0.0011	0.0011	0.05
IEE	intercept	20	20.004	0.0661	0.0630	0.04
	X	0	-0.006	0.0927	0.0897	0.03
	t	-0.3	-0.299	0.0308	0.0308	0.07
	t^2	0	0.000	0.0017	0.0017	0.05
	Xt	0	-0.002	0.0416	0.0399	0.03
	Xt^2	0	0.000	0.0021	0.0020	0.05
WIEE1	intercept	20	20.004	0.0662	0.0621	0.04
	X	0	-0.007	0.0927	0.0888	0.03
	t	-0.3	-0.299	0.0308	0.0283	0.04
	t^2	0	0.000	0.0017	0.0010	0.00
	Xt	0	-0.002	0.0416	0.0376	0.03
	Xt^2	0	0.000	0.0021	0.0015	0.00
WIEE2	intercept	20	20.004	0.0661	0.0623	0.04
	X	0	-0.006	0.0927	0.0894	0.03
	t	-0.3	-0.299	0.0308	0.0294	0.04
	t^2	0	0.000	0.0017	0.0014	0.02
	Xt	0	-0.002	0.0416	0.0389	0.02
	Xt^2	0	0.000	0.0021	0.0018	0.03

Web Figure 5.6: Empirical means and estimated expectations of the mixed model and marginal models using IEE, WIEE1 WIEE2, when death and drop-out mechanisms are at random (DAR/MAR) and the death risk depends on an interaction between the exposure and the last observed marker value[2] for $X = 0$ and $X = 1$ ($N = 500$).



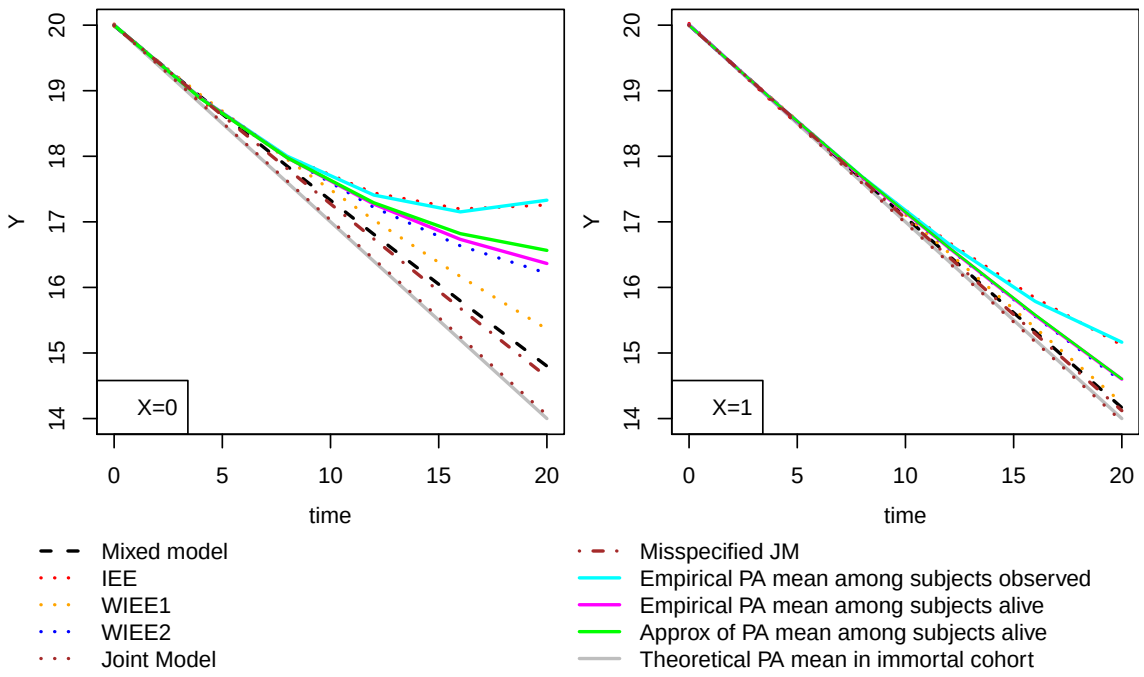
2. Generation models: $\text{logit}\left(P(T_{1i} > t_{ij} | T_{1i} > t_{ij-1})\right) = -0.5 + 0.5X_i + 0.1Y_{ij-1} + 0.1X_iY_{ij-1}$

$\text{logit}\left(P(T_{2i} > t_{ij} | T_{1i} > t_{ij}, T_{2i} > t_{ij-1})\right) = 0.5 + 0.5X_i + 0.1Y_{ij-1}$

Web Table 5.5: Estimates of the mixed model and marginal models using IEE, WIEE1 WIEE2, when death and drop-out mechanisms are at random (DAR/MAR) and the death risk depends on an interaction between the exposure and the last observed marker value^2. The last column corresponds to the probability of rejection of the null hypothesis: $\beta = \beta_G$, with β_G the parameter values for data generation in the mixed model. ASE and ESE are respectively the asymptotical and empirical standard errors.

		β_G	$\overline{\hat{\beta}}$	ASE	ESE	Proba rejection H0: $\beta = \beta_G$
Mixed Model	intercept	20	20.006	0.0646	0.0665	0.06
	X	0	-0.007	0.0902	0.0883	0.03
	t	-0.3	-0.302	0.0285	0.0268	0.03
	t^2	0	0.000	0.0011	0.0011	0.05
	Xt	0	0.002	0.0376	0.0373	0.05
	Xt^2	0	0.000	0.0013	0.0014	0.06
IEE	intercept	20	20.048	0.0706	0.0752	0.12
	X	0	-0.009	0.0967	0.0980	0.05
	t	-0.3	-0.370	0.0423	0.0420	0.39
	t^2	0	0.008	0.0031	0.0032	0.77
	Xt	0	0.027	0.0519	0.0530	0.09
	Xt^2	0	-0.004	0.0034	0.0035	0.21
WIEE1	intercept	20	20.010	0.0754	0.0771	0.06
	X	0	-0.010	0.1016	0.0989	0.04
	t	-0.3	-0.306	0.0500	0.0435	0.02
	t^2	0	0.000	0.0039	0.0027	0.01
	Xt	0	0.006	0.0597	0.0536	0.02
	Xt^2	0	0.000	0.0043	0.0030	0.01
WIEE2	intercept	20	20.037	0.0714	0.0748	0.10
	X	0	-0.012	0.0975	0.0981	0.04
	t	-0.3	-0.352	0.0436	0.0414	0.21
	t^2	0	0.006	0.0032	0.0030	0.47
	Xt	0	0.025	0.0532	0.0524	0.08
	Xt^2	0	-0.004	0.0036	0.0032	0.15

Web Figure 5.7: Empirical means and estimated expectations of the mixed model and marginal models using IEE, WIEE1 WIEE2 when the drop-out mechanism is at random (MAR) and the death risk depends on an interaction between X and the random slope (DNAR)[3] for $X = 0$ and $X = 1$ ($N = 500$).



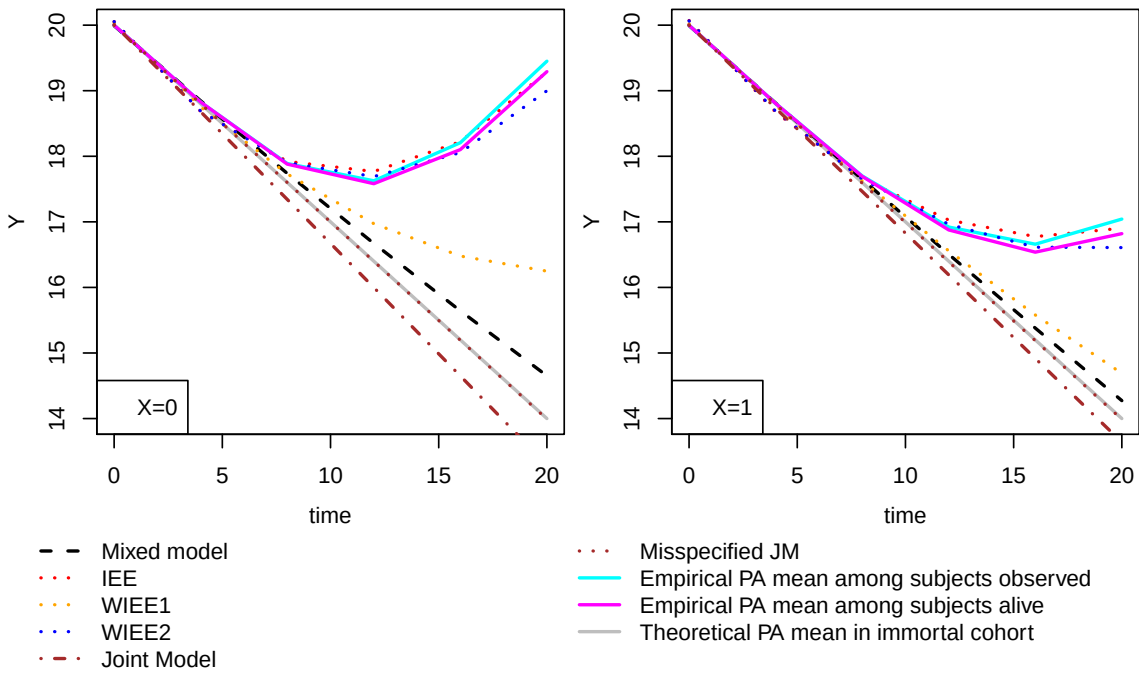
3. Generation models: $\lambda(t) = 0.05 * \exp(-X_i - u_{1i} + 0.3X_i u_{1i})$

$$\text{logit}\left(P(T_{2i} > t_{ij} | T_{1i} > t_{ij}, T_{2i} > t_{ij-1})\right) = -1 + X_i + 0.2Y_{ij-1}$$

Web Table 5.6: Estimates of the mixed model and marginal models using IEE, WIEE1 WIEE2 when the drop-out mechanism is at random (MAR) and the death risk depends on an interaction between X and the random slope (DNAR)[3]. The last column corresponds to the probability of rejection of the null hypothesis: $\beta = \beta_G$, with β_G the parameter values for data generation in the mixed model. ASE and ESE are respectively the asymptotical and empirical standard errors.

		β_G	$\overline{\hat{\beta}}$	ASE	ESE	Proba rejection H0: $\beta = \beta_G$
WIEE2	intercept	20	19.991	0.0646	0.0645	0.05
	X	0	0.007	0.0903	0.0878	0.04
	t	-0.3	-0.273	0.0287	0.0295	0.16
	t^2	0	0.001	0.0011	0.0011	0.09
	Xt	0	-0.021	0.0380	0.0379	0.09
	Xt^2	0	-0.001	0.0013	0.0014	0.07
IEE	intercept	20	20.015	0.0693	0.0715	0.06
	X	0	0.009	0.0961	0.0962	0.05
	t	-0.3	-0.329	0.0412	0.0430	0.11
	t^2	0	0.010	0.0029	0.0030	0.91
	Xt	0	0.002	0.0514	0.0525	0.06
	Xt^2	0	-0.005	0.0033	0.0034	0.40
WIEE1	intercept	20	19.980	0.0746	0.0743	0.05
	X	0	0.011	0.1011	0.0987	0.05
	t	-0.3	-0.269	0.0500	0.0468	0.07
	t^2	0	0.002	0.0039	0.0029	0.06
	Xt	0	-0.021	0.0600	0.0552	0.03
	Xt^2	0	-0.002	0.0043	0.0032	0.03
WIEE2	intercept	20	19.994	0.0718	0.0719	0.06
	X	0	0.007	0.0985	0.0968	0.04
	t	-0.3	-0.294	0.0454	0.0443	0.04
	t^2	0	0.005	0.0034	0.0030	0.36
	Xt	0	-0.008	0.0555	0.0529	0.04
	Xt^2	0	-0.004	0.0039	0.0033	0.17
Joint Model	intercept	20	20.000	0.0646	0.0646	0.05
	X	0	0.001	0.0903	0.0878	0.04
	t	-0.3	-0.297	0.0301	0.0308	0.06
	t^2	0	0.000	0.0012	0.0011	0.05
	Xt	0	-0.004	0.0394	0.0388	0.05
	Xt^2	0	0.000	0.0014	0.0014	0.05

Web Figure 5.8: Empirical means and estimated expectations of the mixed model and marginal models using IEE, WIEE1 WIEE2 when the drop-out mechanism is at random (MAR) and the death risk depends on the current unobserved true value (DNAR)[4] for $X = 0$ and $X = 1$ ($N = 500$).



4. Generation models: $\lambda(t) = 3 * \exp\left(-X_i - 0.2\tilde{Y}_i(t)\right)$
$\text{logit}\left(P(T_{2i} > t_{ij} | T_{1i} > t_{ij}, T_{2i} > t_{ij-1})\right) = 0.3X_i + 0.2Y_{ij-1}$

Web Table 5.7: Estimates of the mixed model and marginal models using IEE, WIEE1 WIEE 2when the drop-out mechanism is at random (MAR) and the death risk depends on the current unobserved true value (DNAR)[4]. The last column corresponds to the probability of rejection of the null hypothesis: $\beta = \beta_G$, with β_G the parameter values for data generation in the mixed model. ASE and ESE are respectively the asymptotical and empirical standard errors.

		β_G	$\overline{\hat{\beta}}$	ASE	ESE	Proba rejection H0: $\beta = \beta_G$
WIEE2	intercept	20	19.991	0.0642	0.0620	0.04
	X	0	0.003	0.0902	0.0918	0.06
	t	-0.3	-0.291	0.0284	0.0282	0.06
	t^2	0	0.001	0.0012	0.0012	0.16
	Xt	0	-0.007	0.0380	0.0377	0.05
	Xt^2	0	-0.001	0.0015	0.0014	0.06
IEE	intercept	20	20.054	0.0702	0.0700	0.11
	X	0	0.017	0.0984	0.1019	0.06
	t	-0.3	-0.418	0.0433	0.0447	0.78
	t^2	0	0.019	0.0030	0.0032	1.00
	Xt	0	0.022	0.0560	0.0559	0.07
	Xt^2	0	-0.007	0.0036	0.0038	0.51
WIEE1	intercept	20	20.037	0.0787	0.0835	0.09
	X	0	-0.019	0.1135	0.1236	0.06
	t	-0.3	-0.354	0.0550	0.0608	0.22
	t^2	0	0.008	0.0041	0.0044	0.62
	Xt	0	0.033	0.0746	0.0786	0.09
	Xt^2	0	-0.006	0.0054	0.0055	0.20
WIEE2	intercept	20	20.053	0.0707	0.0706	0.10
	X	0	0.011	0.0992	0.1027	0.07
	t	-0.3	-0.413	0.0441	0.0450	0.73
	t^2	0	0.018	0.0031	0.0032	1.00
	Xt	0	0.026	0.0570	0.0567	0.07
	Xt^2	0	-0.007	0.0037	0.0038	0.51
Joint Model	intercept	20	19.999	0.0643	0.0622	0.04
	X	0	-0.001	0.0902	0.0921	0.06
	t	-0.3	-0.300	0.0286	0.0281	0.03
	t^2	0	0.000	0.0012	0.0012	0.04
	Xt	0	-0.001	0.0383	0.0375	0.05
	Xt^2	0	0.000	0.0015	0.0014	0.03

6 Discussion

The work presented in this thesis addressed several challenges raised by current research on cognitive decline and dementia. First, we extended joint models in order to account for interval censoring of the time-to-dementia with the competing risk of death. However, when longitudinal data are truncated by death, the interpretation of maximum likelihood estimators is debated as they target the mean trajectory among the immortal cohort. In a second part, we compared likelihood-based and GEE estimators in terms of interpretation to justify the use of mixed models and joint models within this framework.

6.1 Joint latent class model for semi-competing risks and interval censoring

In the first part of this work, we proposed a joint latent class model for handling longitudinal data correlated to an interval-censored time-to-event, while accounting for the semi-competing risk of death. This model was proposed under the markovian and semi-markovian assumptions. The estimating procedure was assessed on simulated data, validating the efficiency and consistency of the estimators. This model was applied to the Paquid cohort to distinguish different profiles of cognitive decline associated with specific risks of dementia and death. We were also able to differentiate the terminal cognitive decline before death among subjects free of dementia from the cognitive decline before dementia, which is steeper.

The joint latent class model can handle multiple longitudinal markers, which are considered as correlated measures of an underlying latent process. In a second application, we built a predictive model, based on the Isaacs Set Test and the MMSE, to predict the risk of dementia within the next five years. This model was estimated on the Paquid cohort and then validated on the Three-city cohort. We obtained dynamic predictions which can be updated at each new measurement of any of these two tests. The results

suggest that combining MMSE and IST data is not better in terms of predictability (both AUC and Brier Score) than the IST.

The joint latent class model relies on the assumption that given the classes, the longitudinal marker and the times-to-events are independent. We extended the score test proposed by Proust-Lima et al. [2014] to handle interval censoring and semi-competing risks. Under the alternative hypothesis $\mathcal{H}_1$, the transition intensities depend on characteristics of the longitudinal process:

$$Y_{ij} = f_1(X_{ij};\ \beta_g) + f_2(Z_{ij};\ \beta_g)\ u_{ig} + \epsilon_{ij}$$
$$\alpha_{klig}(t) = \alpha^0_{klg}(t)\ \exp(W^\top_{kli}\ \gamma_{klg} + \gamma^{(a)}_{kl} u_{ig})$$

Under the null hypothesis, $\gamma^{(a)}_{kl} = 0$. The score is written:

$$U(\gamma^{(a)}_{kl}, \theta) = \sum_{i=1}^{N} U_i(\gamma^{(a)}_{kl}, \theta) = \sum_{i=1}^{N} \frac{\partial L_i^{H_1}}{\partial \gamma^{(a)}_{kl}}$$

Under $\mathcal{H}_0$, the test statistic $U(0,\hat{\theta})^\top Var(U)^{-1} U(0,\hat{\theta})$ follows a Chi square distribution with $3n_u$ degrees of freedom, with n_u the dimension of u_{ig}. The variance under the null hypothesis can be estimated by the empirical variance [Freedman, 2007] defined as $V_{em} = \sum_{i=1}^{N} U_i(0,\hat{\theta}) U_i(0,\hat{\theta})^\top - \frac{U_i(0,\hat{\theta}) U_i(0,\hat{\theta})^\top}{N}$. It would be useful to perform a simulation study to evaluate the type-I and type-II errors of this test. This score test can also be applied to test the conditional independence between a specific transition intensity and the longitudinal marker.

Besides, it is also possible to test if there is any residual correlation between the times-to-events given the classes, using a score test. In the alternative hypothesis, all the transition intensities depend on a shared frailty $\nu_i \sim \mathcal{N}(0, D^2)$ with $D > 0$, which is not included in the mixed sub-model. Under $\mathcal{H}_0$, $D = 0$ and the test statistic $\mathcal{U}(0,\hat{\theta})^\top Var(\mathcal{U})^{-1} \mathcal{U}(0,\hat{\theta})$ would then follow a $\chi^2_{0:1}$ distribution, with $\mathcal{U}(D,\theta) = \sum_{i=1}^{N} \mathcal{U}_i(D,\theta) = \sum_{i=1}^{N} \frac{\partial L_i^{H_1}}{\partial D}$. This score test was more deeply investigated by Loïc Ferrer and Cécile Proust-Lima, in the joint shared-random-effect setting.

There is a growing evidence of heterogeneity in cognitive decline. To take into account this heterogeneity, we can optimize the number of latent classes using either the BIC criterion or a combination of such criteria. However, there is no real scientific consensus on

what would be the best criterion to use, or on the optimal number of latent classes as the biological reasons for heterogeneity remain unclear. Another approach would be to estimate the number of latent classes directly in the model, for instance using non-parametric bayesian methods such as Dirichlet process mixtures [Antoniak, 1974], in order to bypass this model selection issue.

It would be interesting to extend the joint shared-random-effect model to competing risks with interval censoring. The survival sub-model would be replaced by an Illness-Death model as in the model presented in Dantan et al. [2011]. However, as stated in the discussion of the third chapter, the interval censoring issue is less sensitive in joint shared-random-effect models as the assumption that the risk of death is independent from the dementia or health status, given the true current value of the marker, is reasonable and simplifies the likelihood, provided that the middle of the censoring interval is imputed for subjects with dementia. Nevertheless, extending joint shared-random effect models to handle interval censoring will be necessary for a rigorous comparison of the joint shared-random-effect and joint latent class approaches.

In the proposed joint latent class model, every subject is considered at risk to get demented. If it is plausible to consider that a sub-population is not at risk of dementia, it is possible to fix the risk of dementia of a specific latent class to null. Besides, a referee suggested that error in dementia diagnosis was frequent and should be handled in the model by adding a reverse transition from Dementia to Health. First of all, we think that error in dementia diagnosis in the Paquid cohort is unlikely because the diagnosis, carried out by experts, included different steps: three criteria of the DSM IIIR were first clinically evaluated by a psychologist and then, the subjects who had met the three criteria were seen by a neurologist who confirmed and completed the diagnosis. Biological examinations were performed if accepted by the practitioner and the subject. Finally, each case was classified after a consensus meeting, based on the whole follow-up (up to 25 years) of the subjects, and no false positive was recorded. Moreover, we do not think that a reverse transition from Dementia to Health would be a good way to account for misclassification of dementia as this is clinically impossible. In the end, interval censoring is untractable with reversibility between Health and Dementia, unless constraining the

number of possible transitions. It would be more realistic for handling a classification error.

When applied to multivariate markers, the joint latent class model handles the different markers as noisy measures of a common latent process. However, the cognitive tests quantify different cognitive functions, with specific components such as speed (Isaacs Score Test) or culture (MMSE) for instance, which could be represented by correlated latent processes. This is even more valid when considering different types of longitudinal markers, such as biological markers or imaging markers. To better understand the mechanisms involved in the natural history of dementia, it could be interesting to consider several correlated latent processes, as an integrative analysis, linked to the time-to-dementia and time-to-death. Bachirou Tadde and Cécile Proust-Lima are currently working on a multivariate latent process measured by different types of markers. This model could be extended as a joint model in the shared-random-effect or latent class approach. However, in the latter case, the number of parameters may be untractable.

At last, we implemented the Fortran program to estimate the model developed in this thesis and made it available. We plan to include those routines in the R package 'jointlcmm'.

In our applications, the longitudinal follow-up can be stopped by either drop-out or death. As maximum likelihood estimators are equivalent to imputing data beyond the end of the follow-up, the interpretation of their estimand is questioned. In this case, it would mean imputing beyond the time-to-death, targeting the immortal mean trajectory. In a more general way, joint models correspond to a factorisation of the joint density function of the longitudinal marker and the time-to-event $f(Y, T)$ involving latent variables, but other factorisations exist such as pattern mixture models or partly conditional models, which also account for the correlation between the survival and the longitudinal processes. In the second part of the thesis, we investigated and compared the most frequently used methods to handle longitudinal data when the follow-up is truncated by death.

6.2 Methods for longitudinal data truncated by death

In the fifth chapter, we compared likelihood-based estimators of joint and mixed models to unweighted and weighted IEE estimators, when longitudinal data are truncated by death and possibly right-censored by another cause of non-response. In the linear framework, the individual change in the marker over time among the population currently alive is consistently estimated by mixed models under the DCAR/MCAR and DAR/MAR assumptions. If the probability of dying may depend on the true current value of the marker (DNAR), joint models should be used. On the other hand, the population-averaged mean among the population currently alive is consistently estimated by unweighted IEE under the DCAR/MCAR assumption and by IEE weighted by the inverse probability to be observed given that the subject is alive, under the DAR/MAR assumption. We carried out simulations to check these interpretations and to assess the robustness of the different methods. At last, we applied them on the Paquid cohort to quantify the effect of gender on the cognitive decline, assessed by the Isaacs Set Test.

Kurland and Heagerty [2005] defined the missing at random (MAR) assumption as follows: $P(T_{2i} > t|T_{1i} = s, \mathcal{H}_i^{\tilde{Y}}(s), \mathcal{H}_i^{Y}(s), \mathcal{H}_i^{X}(s)) = P(T_{2i} > t|T_{1i} > t, \mathcal{H}_i^{Y}(t), \mathcal{H}_i^{X}(t))$, with $s > t$. The MAR assumption implies that the probability to be observed at t is independent from the future time-to-death s and depends only on past values of the marker observed up to t. This assumption is required to compute the weights as product integrals:

$$\frac{1}{w_{ij}^{(2)}} = P(T_{2i} > t_{ij}|T_{1i} > t_{ij}) = \prod_{k=2}^{j} P(T_{2i} > t_{ik}|T_{1i} > t_{ik}, T_{2i} > t_{ik-1}, \mathcal{H}_i^{Y}(t_{ik-1}), \mathcal{H}_i^{X}(t_{ik-1}))$$

Kurland and Heagerty [2005] proposed a more flexible assumption, called MAR-S, where $P(T_{2i} > t|T_{1i} = s, \mathcal{H}_i^{\tilde{Y}}(s), \mathcal{H}_i^{Y}(s), \mathcal{H}_i^{X}(s)) = P(T_{2i} > t|T_{1i} = s, \mathcal{H}_i^{Y}(t), \mathcal{H}_i^{X}(t))$, with $s > t$. The probability to be observed at t depends on the future time-to-death and on the past observed values of the marker. In this case, the weights need to be adapted by stratifying on the time-to-death (or the corresponding visit). Then, the weights $w_{ij}^{(2)}$, $i = 1, ..., N$ at occasion j would be estimated by an occasion-specific logistic regression on the sub-sample including subjects alive at t_j, sub-sample which may be smaller and smaller as j increases. A future work aims at comparing the two computation methods on the Paquid

cohort.

The MAR assumption implies that the risk of death is the same for subjects still in the cohort and censored subjects. To be compatible with this assumption, the DAR assumption must be precisely defined as $P(T_{1i} = s|T_{2i} = t < s, \mathcal{H}_i^{\tilde{Y}}(s), \mathcal{H}_i^{Y}(s), \mathcal{H}_i^{X}(s)) = P(T_{1i} = s|\mathcal{H}_i^{Y}(t), \mathcal{H}_i^{X}(t))$. This means that the risk of death does not depend on the possible time-to-drop-out. Therefore, in the simulation study presented in section 5.5 under the DAR/MAR assumption, the probability of dying depends on the marker value at the previous visit. If the subject is already censored, the risk of dying then depends on an unobserved value of the marker, and this is not in agreement with the usual missing at random concept. To avoid the dependence on unobserved values between the time-to-drop-out and the time-to-death, we could define another assumption denoted DAR-DO (DO for drop-out) by: $P(T_{1i} = s|T_{2i} = t < s, \mathcal{H}_i^{\tilde{Y}}(s), \mathcal{H}_i^{Y}(s), \mathcal{H}_i^{X}(s)) = P(T_{1i} = s|T_{2i} = t < s, \mathcal{H}_i^{Y}(t), \mathcal{H}_i^{X}(t))$. Here, the risk of death depends on the time-to-drop-out given the observed marker values, which is incompatible with Kurland et al.'s MAR assumption but is compatible with the more flexible MAR-S assumption.

Regarding the simulation results presented in section 5.5, none of the methods (except the joint model) uses the information of the time-to-death for dropped out subjects, so the way the time-to-death is generated for censored subjects does not matter. In practice, the DNAR/MAR assumption is more realistic and we showed that joint models were robust in this case, while mixed models were slightly biased.

The robustness of mixed models and joint models under the MAR-S assumption is questionable. Under the DAR/MAR-S assumption, the probability to be observed at t depends on the future time-to-death s (MAR-S), which may depend on unobserved values of the marker between t and s (DAR). Thus, without conditioning on the time-to-death, the drop-out probability may depend on unobserved marker values between t and s, which defines the MNAR hypothesis. Joint models and mixed models may be biased in this case. However, it would be of interest to investigate their properties under the DAR-DO/MAR-S assumption, when the probability to die in s, given the subject dropped out in $t < s$, only depends on values of the marker observed before the time-to-drop-out t (DAR-DO).

In the presented application on the Paquid cohort, we excluded subjects with intermittent missing data because the methods and hypotheses used to compute the weights did not handle intermittent missing data. When the sample includes intermittent MAR data, both mixed models and joint models are unbiased. On the other hand, the computation of the weights for WIEE2 should be adapted, for example by adjusting the logistic regressions on the last observed value (instead of the observed value at the previous visit) and on the observation indicator at the previous visit. This extension could also be compared on the Paquid cohort.

A type of models where the weighted IEE approach could be very useful is quantile regressions. Indeed, quantile regression aims at estimating quantiles of the marker distribution in the population and not individual quantiles that have an unclear meaning. Thus, the GEE approach is more adapted to quantile regression for longitudinal data than the mixed model approach. A future work consists in developing the estimation algorithm for WIEE2, handling intermittent missing data, in order to target the quantiles of the trajectory of the marker among the dynamic population of survivors.

6.3 Conclusion

In cohort studies on dementia, interval censoring is most often neglected. This may not be an issue if visit intervals are short but visits are often spaced by more than two years. The initial purpose of this work was to investigate joint models to deal with longitudinal markers and the time-to-dementia, accounting for interval censoring. However, death is a central issue in this type of analyses, as it implies a selection of the population, and it can not be considered as any other cause of non-response as data beyond death do not exist. Discussions within the Melodem group (Methods in longitudinal dementia research) raised questions and debates about the interpretations of joint models, when longitudinal data are truncated by death. Those discussions prompted us to focus on a longitudinal marker correlated to death to clarify the interpretations of partly conditional models (marginal models) and unconditional models (mixed models and joint models).

Analyses on cognitive decline and dementia require sophisticated tools to describe the

natural history of dementia, without bias, and gain insights into its mechanisms. These tools help clarify the role of risk factors, distinguishing their impact on the cognitive decline, on the risk of dementia and on the risk of death. As an example, men seem to have a slower mean cognitive decline than women over time but this work actually shows that this is due to a stronger selection of men by death. Joint models are also useful tools for earlier diagnosis, as they can help target subjects at high risk of dementia. The current research in pharmacology focuses on the development of new treatments given at earlier stages of dementia, as the existing ones have showed limited efficiency on the improvement of health and cognitive states. At last, these works can be applied to any chronic disease with the competing risk of death.

Bibliography

3C Study Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology*, 22(6): 316–325, 2003. 24, 33

Al Hazzouri A. Z., Haan M. N., Kalbfleisch J. D., Galea S., Lisabeth L. D., and Aiello A. E. Life-course socioeconomic position and incidence of dementia and cognitive impairment without dementia in older Mexican Americans: Results from the Sacramento Area Latino Study on Aging. *American Journal of Epidemiology*, 173(10): 1148–1158, 2011. 63

Alioum A. and Commenges D. A Proportional Hazards Model for Arbitrarily Censored and Truncated Data. *Biometrics*, 52(2): 512–524, 1996. 63

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. (DSM-III-R. 3rd edition revised). 1987. 30

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). 2000. 31

Amieva H., Le Goff M., Millet X., Orgogozo J. M., Pérès K., Barberger-Gateau P., Jacqmin-Gadda H., and Dartigues J.-F. Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Annals of Neurology*, 64(5): 492–498, 2008. 33

Andersen P. K. and Keiding N. Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine*, 31(11-12): 1074–1088, 2012. 61

Antoniak C. E. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics*, 2(6): 1152–1174, 1974. 151

Association A. P. Diagnostic and statistical manual of mental disorders (5th ed.). 2013. 31

Backman L., Small B. J., and Fratiglioni L. Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain*, 124: 96–102, 2001. 32, 38

Bäckman L., Jones S., Berger A.-K., Laukka E. J., and Small B. J. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology*, 19(4): 520–531, 2005. 33

Bennett D. A., Schneider J. A., Buchman A. S., De Leon C. M., Bienias J. L., and Wilson R. S. The rush memory and aging project: Study design and baseline characteristics of the study cohort. *Neuroepidemiology*, 25(4): 163–175, 2005. 33

Berr C., Vercambre M.-N., and Akbaraly T. N. Epidemiology of Alzheimer's disease: methodological approaches and new perspectives. *Psychologie & neuropsychiatrie du vieillissement*, 7(1): 7–14, 2009. 30

Betensky R. a. Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, 88(3): 703–711, 2001. 63

Birks J. Cholinesterase inhibitors for Alzheimer's disease. *The Cochrane database of systematic reviews*, (1): CD005593, 2006. 18, 32

Blanche P., Dartigues J.-F., and Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine*, 32(30): 5381–5397, 2013. 107

Blanche P., Proust-Lima C., Loubère L., Berr C., Dartigues J.-F., and Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, 71(1): 102–113, 2015. 24, 108, 112, 116, 117

Boucquemont J., Metzger M., Combe C., Stengel B., Leffondré K., and NephroTest Study Group. Should we use standard survival models or the illness-death model for interval-censored data to investigate risk factors of chronic kidney disease progression? *PloS One*, 9(12): e114839, 2014. 65

Brayne C. and Calloway P. Normal ageing, impaired cognitive function, and senile dementia of the Alzheimer's type: a continuum? *The Lancet*, 331(8597): 1265–1267, 1988. 32

Brayne C., Spiegelhalter D., Dufouil C., Chi L.-y., Dening T. R., Paykel E. S., Connor D. W. O., Ahmed A., Mcgee M. A., and Huppert F. A. Estimating the True Extent of Cognitive Decline in the Old Old. *JAGS*, 47: 1283–1288, 1999. 58

Christensen H., Mackinnon A. J., Korten A. E., Jorm A. F., Henderson A. S., Jacomb P., and Rodgers B. An analysis of diversity in the cognitive performance of elderly community dwellers: individual differences in change scores as a function of age. *Psychology and aging*, 14(3): 365–379, 1999. 18, 32

Colsher P. L. and Wallace R. B. Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology*, 1(3): 215–230, 1991. 32, 59

Courtney C., Farrell D., Gray R., Hills R., Lynch L., Sellwood E., Edwards S., Hardyman W., Raftery J., Crome P., Lendon C., Shaw H., Bentham P., and AD2000 Collaborative Group. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *The Lancet*, 363(9427): 2105–2115, 2004. 18, 32

Dantan E., Joly P., Dartigues J.-F., and Jacqmin-Gadda H. Joint model with latent state for longitudinal and multistate data. *Biostatistics*, 12(4): 723–736, 2011. 19, 62, 151

Dartigues J.-F., Gagnon M., Barberger-Gateau P., Letenneur L., Commenges D., Sauvel C., Michel P., and Salamon R. The Paquid Epidemiological Program on Brain Ageing. *Neuroepodemiology*, 11(1): 14–18, 1992. 33

Dartigues J.-F., Commenges D., Letenneur L., Barberger-Gateau P., Gilleron V., Fabrigoule C., Mazaux J. M., Orgogozo J. M., and Salamon R. Cognitive Predictors of Dementia in Elderly Community Residents. *Neuroepidemiology*, 16(1): 29–39, 1996. 19, 32, 38

Dartigues J.-F., Helmer C., Letenneur L., Péres K., Amieva H., Auriacombe S., Orgogozo J. M., Commenges D., Jacqmin-Gadda H., Richard-Harston S., Delva F., Foubert-Samier A., and Barberger-Gateau P. Paquid 2012: illustration et bilan. *Geriatrie et Psychologie Neuropsychiatrie du Vieillissement*, 10(3): 325–331, 2012. 29

De Gruttola V. and Tu X. M. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, 50(4): 1003–1014, 1994. 50

Dempster A., Laird N., and Rubin D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, 39(1): 1–38, 1977. 50

Dufouil C., Brayne C., and Clayton D. Analysis of longitudinal studies with death and drop-out: a case study. *Statistics in Medicine*, 23(14): 2215–2226, 2004. 25, 58

Elashoff R. M., Li G., and Li N. A Joint Model for Longitudinal Measurements and Survival Data in the Presence of Multiple Failure Types. *Biometrics*, 64(3): 762–771, 2008. 61

Fabrigoule C., Letenneur L., Dartigues J.-F., Zarrouk M., Commenges D., and Barberger-Gateau P. Social and leisure activities and risk of dementia: a prospective longitudinal study. *Journal of the American Geriatrics Society*, 43: 485–490, 1995. 17, 30

Farmer M. E., White L. R., Kittner S. J., Kaplan E., Moes E., McNamara P., Wolz M. M., Wolf P. A., and Feinleib M. Neuropsychological test performance in Framingham: a descriptive study. *Psychological reports*, 60: 1023–1040, 1987. 33

Farrer L. A., Cupples L. A., Haines J. L., Hyman B., Kukull W. A., Mayeux R., Myers R. H., Pericak-Vance M. A., Risch N., and van Duijn C. M. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: A meta-analysis. *JAMA*, 278(16): 1349–1356, 1997. 17, 30

Ferrer L., Rondeau V., Dignam J., Pickles T., Jacqmin-Gadda H., and Proust-Lima C. Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in Medicine*, 35(22): 3933–3948, 2016. 62

Fine J. P. and Gray R. J. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 94(446): 496–509, 1999. 61

Finkelstein D. M. A proportional Hazards Model for Interval-Censored Failure Time Data. *Biometrics*, 42(4): 845–854, 1986. 63

Fitzmaurice G. M., Molenberghs G., and Lipsitz S. R. Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(4): 691–704, 1995. 48

Flicker C., Ferris S. H., and Reisberg B. Mild cognitive impairment in the elderly: Predictors of dementia. *Neurology*, 41(7): 1006–1009, 1991. 38

Folstein M. F., Folstein S. E., and McHugh P. R. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3): 189–198, 1975. 17, 31

Forette F., Seux M. L., Staessen J. A., Thijs L., Birkenhäger W. H., Babarskiene M. R., Babeanu S., Bossini A., Gil-Extremera B., Girerd X., Laks T., Lilov E., Moisseyev V., Tuomilehto J., Vanhanen H., Webster J., Yodfat Y., and Fagard R. Prevention of dementia in randomised double-blind placebo-controlled Systolic Hypertension in Europe (Syst-Eur) trial. *The Lancet*, 352(9137): 1347–1351, 1998. 30

Frangakis C. E. and Rubin D. B. Principal Stratification in Causal Inference. *Biometrics*, 58: 21–29, 2002. 59

Fratiglioni L., Paillard-Borg S., and Winblad B. An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet Neurology*, 3(6): 343–353, 2004. 17, 30

Freedman D. a. How Can the Score Test Be Inconsistent? *The American Statistician*, 61(4): 291–295, 2007. 150

Freitag M. H., Peila R., Masaki K., Petrovitch H., Ross G. W., White L. R., and Launer L. J. Midlife pulse pressure and incidence of dementia: The Honolulu-Asia Aging Study. *Stroke*, 37 (1): 33–37, 2006. 63

Grober E. and Buschke H. Genuine memory deficits in dementia. *Developmental Neuropsychology*, 3(1): 13–36, 1987. 31

Gueorguieva R., Rosenheck R., and Lin H. Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(2): 417–433, 2012. 65

Hachinski V. C., Iliff L. D., Zilhka E., Du Boulay G. H., McAllister V. L., Marshall J., Russell R. W., and Symon L. Cerebral blood flow in dementia. *Archives of neurology*, 32(9): 632–637, 1975. 34

Hall C. B., Lipton R. B., Sliwinski M., and Stewart W. F. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in medicine*, 19 (11-12): 1555–1566, 2000. 33

Hawkins D. S., Allen D. M., and Stromberg A. J. Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis*, 38(1): 15–48, 2001. 53

Helmer C., Joly P., Letenneur L., Commenges D., and Dartigues J.-F. Mortality with dementia: Results from a French prospective community-based cohort. *American Journal of Epidemiology*, 154(7): 642–648, 2001. 63

Henderson R., Diggle P., and Dobson a. Joint modelling of longitudinal measurements and event time data. *Biostatistics (Oxford, England)*, 1(4): 465–480, 2000. 50

Hipp J. R. and Bauer D. J. Local solutions in the estimation of growth mixture models. *Psychological methods*, 11(1): 36–53, 2006. 22, 53

Huang J. and Rossini A. J. Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, 92(439): 960–967, 1997. 63

Huang J. and Wellner J. A. Interval censored survival data: A review of recent progress. *Proceedings of the First Seattle Symposium in Biostatistics*, 1997. 63

Isaacs B. and Kennie A. T. The Set Test as an Aid to the Detection of Dementia in Old People. *The British Journal of Psychiatry*, 123(575): 467–470, 1973. 17, 31

Jack C. R., Knopman D. S., Jagust W. J., Shaw L. M., Aisen P. S., Weiner M. W., Petersen R. C., and Trojanowski J. Q. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1): 119–128, 2010. 32

Jacqmin-Gadda H., Fabrigoule C., Commenges D., and Dartigues J.-F. A 5-Year Longitudinal Study of the Mini-Mental State Examination in Normal Aging. *American Journal of Epidemiology*, 145(6): 498–506, 1997. 111

Jacqmin-Gadda H., Commenges D., and Dartigues J.-F. Random change point model for joint modeling of cognitive decline and dementia. *Biometrics*, 62(1): 254–260, 2006. 50

Jacqmin-Gadda H., Proust-Lima C., Taylor J. M. G., and Commenges D. Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics*, 66(1): 11–19, 2010. 55, 56

Jacqmin-Gadda H., Blanche P., Chary E., Touraine C., and Dartigues J.-F. Receiver operating characteristic curve estimation for time to event with semicompeting risks and interval censoring. *Statistical methods in medical research*, in press, 2014. 107, 109, 117, 118

Jobst K. A., Smith A. D., Szatmari M., Esiri M. M., Jaskowski A., Hindley N., McDonald B., and Molyneux A. J. Rapidly progressing atrophy of medial temporal lobe in Alzheimer's disease. *The Lancet*, 343(8901): 829–830, 1994. 32

Joly P., Commenges D., Letenneur L., Joly P., and Commenges D. A Penalized Likelihood Approach for Arbitrarily Censored and Truncated Data : Application to Age-Specific Incidence of Dementia. *Biometrics*, 54(1): 185–194, 1998. 63

Joly P., Commenges D., Helmer C., and Letenneur L. A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics*, 3(3): 433–443, 2002. 22, 37, 64

Kalbfleisch J. D. and Prentice R. L. Relative risk (Cox) Regression Models. In *The Statistical Analysis of Failure Time Data*, chapter 4, pages 95–147. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2nd edition, 2002. 63

Kim J. S. Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65(2): 489–502, 2003. 63

Kleemeier R. W. Intellectual change in the senium. *Proceedings of the Social Statistics Section of the American Statistical Association, Washington, DC*, pages 290–295, 1962. 23, 59

Kurland B. F. and Heagerty P. J. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics*, 6(2): 241–258, 2005. 37, 58, 153

Kurland B. F., Johnson L. L., Egleston B. L., and Diehr P. H. Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. *Statistical Science*, 24 (2): 211–222, 2009. 19, 25, 58, 60

Laird N. M. and Ware J. H. Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4): 963–974, 1982. 43

Law C. G. and Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine*, 11(12): 1569–1578, 1992. 63

Leffondré K., Touraine C., Helmer C., and Joly P. Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model? *International journal of epidemiology*, 42(4): 1177–1186, 2013. 65

Letenneur L., Gilleron V., Commenges D., Helmer C., Orgogozo J. M., and Dartigues J.-F. 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(2): 177–183, 1999. 30

Letenneur L., Commenges D., Dartigues J.-F., and Barberger-Gateau P. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *International journal of epidemiology*, 23(6): 1256–1261, 1994. 17, 30

Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2): 164–168, 1944. 65

Liang K.-Y. and Zeger S. L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73(1): 13–22, 1986. 46

Lin H., Turnbull B. W., McCulloch C. E., and Slate E. H. Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data. *Journal of the American Statistical Association*, 97(457): 53–65, 2002. 20, 51, 54

Little R. J. A. Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association*, 90(431): 1112–1121, 1995. 59

Little, Roderick JA and Rubin D. *Statistical analysis with missing data*. Wiley, New York, 1987. 36

Marquardt D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2): 431–441, 1963. 22, 65, 74

Matthews F. E., Arthur A., Barnes L. E., Bond J., Jagger C., Robinson L., and Brayne C. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: Results of the cognitive function and ageing study I and II. *The Lancet*, 382(9902): 1405–1412, 2013. 29

McCullagh, Peter and Nelder J. A. *Generalized linear models*. CRC press, 1989. 44

McKhann G. M., Drachman D., Folstein M. F., Katzman R., Price D., and Stadlan E. M. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7): 939–944, 1984. 34

Muthén B. Latent variable hybrids: Overview of old and new models. In Hancock G. R. and Samuelsen K. M., editors, *Advances in latent variable mixture models*, pages 1–24. Information Age Publishing, 2008. 38

Odell P. M., Anderson K. M., and D'Agostino R. B. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, 48(3): 951–959, 1992. 63

Ownby R. L., Crocco E., Acevedo a., John V., and Loewenstein D. Depression and risk for Alzheimer disease. *Arch Gen Psychiatry*, 63(May): 530–538, 2006. 30

Pérès K., Matharan F., Allard M., Amieva H., Baldi I., Barberger-Gateau P., Bergua V., Bourdel-Marchasson I., Delcourt C., Foubert-Samier A., Fourrier-Réglat A., Gaimard M., Laberon S., Maubaret C., Postal V., Chantal C., Rainfray M., Rascle N., and Dartigues J.-F. Health and aging in elderly farmers: the AMI cohort. *BMC public health*, 12: 558, 2012. 33

Peters R., Poulter R., Warner J., Beckett N., Burch L., and Bulpitt C. Smoking, dementia and cognitive decline in the elderly, a systematic review. *BMC Geriatrics*, 8: 36, 2008. 30

Philipps V., Amieva H., Andrieu S., Dufouil C., Berr C., Dartigues J.-F., Jacqmin-Gadda H., and Proust-Lima C. Normalized mini-mental state examination for assessing cognitive change in population-based brain aging studies. *Neuroepidemiology*, 43(1): 15–25, 2014. 110

Preisser J. S., Lohman K. K., and Rathouz P. J. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine*, 21(20): 3035–3054, 2002. 48

Prince M., Wimo A., Guerchet M., Gemma-Claire A., Wu Y.-T., and Prina M. World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends, 2015. 17, 29, 30

Proust-Lima C., Jacqmin-Gadda H., Taylor J. M. G., Ganiayre J., and Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*, 62(4): 1014–1024, 2006. 39, 56

Proust-Lima C., Amieva H., Dartigues J.-F., and Jacqmin-Gadda H. Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. *American Journal of Epidemiology*, 165(3): 344–350, 2007. 109

Proust-Lima C., Joly P., Dartigues J.-F., and Jacqmin-Gadda H. Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics & Data Analysis*, 53(4): 1142–1154, 2009. 54, 56

Proust-Lima C., Amieva H., and Jacqmin-Gadda H. Analysis of multivariate mixed longitudinal data: a flexible latent process approach. *The British journal of mathematical and statistical psychology*, 66(3): 470–487, 2012. 39

Proust-Lima C., Séne M., Taylor J. M. G., and Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical methods in medical research*, 23(1): 74–90, 2014. 18, 53, 55, 56, 150

Proust-Lima C., Dartigues J.-F., and Jacqmin-Gadda H. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Statistics in Medicine*, 35(3): 382–398, 2016. 19, 62

Putter H., Fiocco M., and Geskus R. B. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11): 2389–2430, 2007. 60

Redner R. A. and Walker H. F. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2): 195–239, 1984. 22, 53

Rizopoulos D. Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data. *Biometrics*, 67(3): 819–829, 2011. 19

Rizopoulos D., Verbeke G., and Molenberghs G. Shared parameter models under random effects misspecification. *Biometrika*, 95(1): 63–74, 2008. 51

Rizopoulos D., Verbeke G., and Lesaffre E. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3): 637–654, 2009. 50

Roberts R. O., Geda Y. E., Knopman D. S., Cha R. H., Pankratz V. S., Boeve B. F., Ivnik R. J., Tangalos E. G., Petersen R. C., and Rocca W. A. The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30(1): 58–69, 2008. 33

Robins J. M., Rotnitzky A., and Zhao L. P. Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*, 90(429): 106–121, 1995. 48

Schaie K. W. Variability in cognitive function in the elderly: implications for societal participation. *Basic life sciences*, 43: 191–211, 1988. 18, 32, 38

Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978. 22, 53

Self S. G. and Grossman E. a. Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics*, 42(3): 521–530, 1986. 63

Siegler I. C. The terminal drop hypothesis: fact or artifact? *Experimental Aging Research*, 1(1): 169–185, 1975. 23, 59

Stern Y., Gurland B., Tatemichi T. K., Tang M. X., Wilder D., and Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA : the journal of the American Medical Association*, 271(13): 1004–1010, 1994. 30

Stern Y. Cognitive reserve. *Neuropsychologia*, 47(10): 2015–2028, 2009. 32

Sun J. Regression analysis of interval-censored failure time data. *Statistics in medicine*, 16(5): 497–504, 1997. 63

Sun J. *The Statistical Analysis of Interval-censored Failure Time Data*. Springer Science & Business Media, Columbia, 2007. 63

Thorvaldsson V., Hofer S. M., Berg S., Skoog I., Sacuiu S., and Johansson B. Onset of terminal decline in cognitive abilities in individuals without dementia. *Neurology*, 71(12): 882–887, 2008. 59, 60

Touraine C., Helmer C., and Joly P. Predictions in an illness-death model. *Statistical Methods in Medical Research*, 25(4): 1452–1470, 2016. 65

Tsiatis A. A. and Davidian M. Joint Modeling of Longitudinal and Time-to-Event Data : An Overview. *Statistica Sinica*, 14: 809–834, 2004. 18, 48

Vellas B., Coley N., Ousset P. J., Berrut G., Dartigues J.-F., Dubois B., Grandjean H., Pasquier F., Piette F., Robert P., Touchon J., Garnier P., Mathiex-Fortunet H., and Andrieu S. Long-term use of standardised ginkgo biloba extract for the prevention of Alzheimer's disease (GuidAge): A randomised placebo-controlled trial. *The Lancet Neurology*, 11(10): 851–859, 2012. 18, 33

Visser P. J., Scheltens P., Verhey F. R., Schmand B., Launer L. J., Jolles J., and Jonker C. Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. *J Neurol*, 246(6): 477–485, 1999. 38

Wanneveich M., Jacqmin-Gadda H., Dartigues J.-F., and Joly P. Impact of intervention targeting risk factors on chronic disease burden. *Statistical methods in medical research*, in press, 2016. 17, 30

West M. J., Coleman P. D., Flood D. G., and Troncoso J. C. Differences in the Pattern of Hippocampal Neuronal Loss in Normal Aging and Alzheimers-Disease. *The Lancet*, 344(8925): 769–772, 1994. 32

Wilson R. S., Beckett L. A., Barnes L. L., Schneider J. A., Bach J., Evans D. A., and Bennett D. A. Individual differences in rates of change in cognitive abilities of older persons. *Psychol Aging*, 17(2): 179–193, 2002. 19, 32, 38

Wilson R. S., Beckett L. A., Bienias J. L., Evans D. A., and Bennett D. A. Terminal decline in cognitive function. *Neurology*, 60(11): 1782–1787, 2003. 23, 37, 59, 60

Wilson R. S., Beck T. L., Bienias J. L., and Bennett D. a. Terminal cognitive decline: accelerated loss of cognition in the last years of life. *Psychosomatic medicine*, 69(2): 131–137, 2007. 60

Wu M. C. and Carroll R. J. Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process. *Biometrics*, 44(1): 175–188, 1988. 38

Wu Y.-T., Fratiglioni L., Matthews F. E., Lobo A., Breteler M. M. B., Skoog I., and Brayne C. Dementia in western Europe: epidemiological evidence and implications for policy making. *The Lancet Neurology*, 15(1): 116–124, 2016. 29

Wulfsohn M. S. and Tsiatis A. A. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1): 330–339, 1997. 20, 48, 50

Yu M., Law N. J., Taylor J. M., and Sandler H. M. Joint Longitudinal-Survival-Cure Models and Their Application to Prostate Cancer. *Statistica Sinica*, 14: 835–862, 2004. 49

Zeng D., Cai J., and Shen Y. Semiparametric additive risks model for interval-censored data. *Statistica Sinica*, 16: 287–302, 2006. 63

Zhang M. H. and Cheng Q. S. Determine the number of components in a mixture model by the extended KS test. *Pattern Recognition Letters*, 25(2): 211–216, 2004. 53

Zhang Z., Sun L., Zhao X., and Sun J. Regression analysis of interval-censored failure time data with linear transformation models. *Canadian Journal of Statistics*, 33(1): 61–70, 2005. 63

Prise en compte de la sélection par le décès dans l'étude de la démence et du déclin cognitif

Ce travail a pour but de développer des outils statistiques pour l'étude du déclin cognitif général ou précédant le diagnostic de démence, à partir de données de cohorte en tenant compte du risque compétitif de décès et de la censure par intervalle. Le temps de démence est censuré par intervalle dans les études de cohortes car le diagnostic de démence ne peut être établi qu'à l'occasion des visites qui peuvent être espacées de plusieurs années. Ceci induit une sous-estimation du risque de démence à cause du risque compétitif de décès : les sujets déments sont à fort risque de mourir, et peuvent donc décéder avant la visite de diagnostic. Dans la première partie, nous proposons un modèle conjoint à classes latentes pour données longitudinales corrélées à un événement censuré par intervalle, en compétition avec le décès. Appliqué à la cohorte Paquid, ce modèle permet d'identifier des profils de déclin cognitif associés à des risques différents de démence et de décès. En utilisant cette méthodologie, nous comparons ensuite des modèles pronostiques dynamiques pour la démence, traitant la censure par intervalle, basés sur des mesures répétées de marqueurs cognitifs. Dans la seconde partie, nous conduisons une étude comparative afin de clarifier l'interprétation des estimateurs du maximum de vraisemblance des modèles mixtes et conjoints et estimateurs par équations d'estimation généralisées (GEE), couramment utilisés dans le contexte de données longitudinales incomplètes et tronquées par le décès. Les estimateurs de maximum de vraisemblance ciblent le changement individuel chez les individus vivants. Les estimateurs GEE avec matrice de corrélation de travail indépendante, pondérés par l'inverse de la probabilité d'être observé sachant que le sujet est vivant, ciblent la trajectoire moyennée sur la population des survivants à chaque âge. Ces résultats justifient l'utilisation des modèles conjoints dans l'étude de la démence, qui sont des outils prometteurs pour mieux comprendre l'histoire naturelle de la maladie.

Mots clés : Censure par intervalle, Décès, Estimateur moyen sur la population, Estimateur spécifique au sujet, GEE, Modèles conjoints, Modèles mixtes, Prédictions dynamiques, Risques semi-competitifs.

Study of dementia and cognitive decline accounting for selection by death

The purpose of this work is to develop statistical tools to study the general or the pre-diagnosis cognitive decline, while accounting for the selection by death and interval censoring. In cohort studies, the time-to-dementia-onset is interval-censored as the dementia status is assessed intermittently. This issue can lead to an under-estimation of the risk of dementia, due to the competing risk of death: subjects with dementia are at high risk to die and can thus die prior to the diagnosis visit. First, we propose a joint latent class illness-death model for longitudinal data correlated to an interval-censored time-to-event, competing with the time-to-death. This model is applied on the Paquid cohort to identify profiles of pre-dementia cognitive declines associated with different risks of dementia and death. Using this methodology, we compare dynamic prognostic models for dementia based on repeated measures of cognitive markers, accounting for interval censoring. Secondly, we conduct a simulation study to clarify the interpretation of maximum likelihood estimators of joint and mixed models as well as GEE estimators, frequently used to handle incomplete longitudinal data truncated by death. Maximum likelihood estimators target the individual change among the subjects currently alive. GEE estimators with independent working correlation matrix, weighted by the inverse probability to be observed given that the subject is alive, target the population-averaged change among the dynamic population of survivors. These results justify the use of joint models in dementia studies, which are promising statistical tools to better understand the natural history of dementia.

Key words: Death, Dynamic predictions, GEE, Interval censoring, Joint models, Mixed models, Population-averaged, Semi-competing risks, Subject-specific.

Discipline : Santé publique – option : Biostatistiques
Laboratoire : Unité INSERM U1219 - Université de Bordeaux - ISPED
146 rue Léo Saignat 33076 Bordeaux, FRANCE