



HAL
open science

Modélisation de l'histoire naturelle des démences et implications des facteurs de risques vasculaires dans la cohorte MEMENTO

Jérémie Lespinasse

► **To cite this version:**

Jérémie Lespinasse. Modélisation de l'histoire naturelle des démences et implications des facteurs de risques vasculaires dans la cohorte MEMENTO. Médecine humaine et pathologie. Université de Bordeaux, 2023. Français. NNT : 2023BORD0019 . tel-04048640

HAL Id: tel-04048640

<https://theses.hal.science/tel-04048640>

Submitted on 28 Mar 2023

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

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

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE

**DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE SOCIÉTÉS, POLITIQUE, SANTÉ PUBLIQUE
SPÉCIALITÉ SANTÉ PUBLIQUE - OPTION ÉPIDÉMIOLOGIE

Par Jérémie Lespinasse

**Modélisation de l'histoire naturelle des démences et
implications des facteurs de risques vasculaires dans la
cohorte MEMENTO**

Sous la codirection de Carole Dufouil et Cécile Proust-Lima

Soutenue le mardi 31 janvier 2023

Membres du jury :

Mme. DEBETTE, Stéphanie	PU-PH. Université de Bordeaux	Présidente
M. DANTAN, Etienne	MCU. Université de Nantes	Rapporteur
M. MURA, Thibault	PU-PH. CHU de Nîmes	Rapporteur
M. VIDAL, Jean-Sébastien	PH. Assistance Publique Hôpitaux de Paris	Examinateur
M. TOURDIAS, Thomas	PU-PH. Université de Bordeaux	Examinateur
Mme. SAMIERI, Cécilia	DR. Inserm, Université de Bordeaux	Examinatrice
Mme DUFOUIL Carole	DR. Inserm, Université de Bordeaux	Invitée
Mme PROUST-LIMA Cécile	DR. Inserm, Université de Bordeaux	Invitée

Titre : Modélisation de l'histoire naturelle des démences et implications des facteurs de risques vasculaires dans la cohorte MEMENTO

Résumé : La démence est un syndrome qui se manifeste par une altération progressive des fonctions cognitives jusqu'à la perte de l'autonomie. Il existe plusieurs étiologies de la démence dont la plus fréquente est la maladie d'Alzheimer qui est la conséquence de lésions cérébrales spécifiques. La cascade amyloïde en est la principale hypothèse physiopathologique selon laquelle le déclin cognitif est causé par la mort des neurones consécutive à une neuropathologie initiée par l'agrégation anormale de molécules, les peptides tau et amyloïde. Cette hypothèse est cependant souvent remise en question et l'histoire naturelle de la maladie reste insuffisamment comprise, notamment les interrelations et la séquence temporelle des marqueurs de la maladie d'Alzheimer. De plus, la contribution des conséquences cérébrales des facteurs de risque vasculaires pourrait être sous-estimée dans les mécanismes physiopathologiques de la démence. En particulier l'hypertension artérielle entraîne des anomalies des petits vaisseaux sanguins du cerveau qui pourraient accélérer l'entrée dans la démence.

Mes travaux de recherche s'appuient sur les données multimodales de la cohorte MEMENTO (2011-2019) qui inclut 2323 participants issus de 28 centres mémoire français et suivis de façon régulière avec des données répétées cliniques, psychométriques, biologiques et d'imageries cérébrales. Les objectifs sont (1) de décrire la séquence temporelle d'évolution de marqueurs multimodaux avec l'application d'un modèle mixte à recalage temporel qui permet de modéliser la progression des marqueurs selon un temps latent de la maladie et (2) d'étudier les voies d'actions de l'hypertension et de son contrôle dans l'association avec la cognition à l'aide de modèles d'équations structurelles qui permettent d'analyser la médiation. Pour ces travaux de thèse, j'ai étudié particulièrement (i) des tests neuropsychologiques qui évaluent la mémoire, le langage, la fonction exécutive et la fonction visuo-spatiale ; (ii) des marqueurs de neurodégénérescence : volume de l'hippocampe, épaisseur du cortex cérébral, scanner du métabolisme glucidique (TEP FDG) et fraction du volume crânien occupée par le cerveau ; (iii) des marqueurs de neuropathologie : mesures des concentrations dans le liquide céphalo-rachidien des peptides amyloïdes et tau et scanner de l'agrégation des peptides amyloïdes (TEP amyloïde) ; (iv) un marqueur d'anomalies des petits vaisseaux sanguins du cerveau : le volume des hypersignaux de la substance blanche.

Les résultats ont permis d'une part de mettre en évidence une incertitude sur le rôle initiateur du peptide amyloïde et l'influence substantielle des caractéristiques individuelles (facteur de risque génétique et niveau d'études) sur la séquence temporelle des marqueurs de progression de la maladie. D'autre part les résultats mettent en évidence l'importance de la prise en charge et du contrôle de l'hypertension. Nous avons en effet montré que l'hypertension non contrôlée est associée à une dégradation de la structure cérébrale et des performances cognitives inférieures.

Ces travaux de thèse ont montré que la modélisation de la progression selon le temps latent de maladie pourrait permettre de mieux caractériser les individus avant l'apparition des symptômes, d'améliorer leurs prises en charge, et d'identifier des axes de prévention potentiels. Les stratégies destinées à renforcer le contrôle de la pression artérielle pourraient contribuer à ralentir la dégénérescence du cerveau et à prévenir la démence. Les perspectives de ces travaux incluent l'exploitation d'un nombre plus important de marqueurs et de données répétées pour décrire plus précisément l'influence des facteurs de risque de démence sur la dynamique de progression des marqueurs.

Mots clés : démence, maladie d'Alzheimer, biomarqueurs, hypertension, modélisation

Unité de recherche

Unité INSERM U1219 – Bordeaux Population Health
Université de Bordeaux – ISPED
146 rue Léo Saignat 33076 Bordeaux, France

Title : Modeling the natural history of dementia and implications of vascular risk factors in the MEMENTO cohort

Abstract : Dementia is a syndrome characterized by a progressive alteration of cognitive functions until the loss of autonomy. There are several etiologies of dementia, the most common of which is Alzheimer's disease resulting from specific brain lesions. The amyloid cascade is the main pathophysiological hypothesis according to which cognitive decline is caused by the death of neurons following a neuropathology initiated by the abnormal aggregation of molecules, the tau and amyloid peptides. However, this hypothesis is often questioned, and the natural history of the disease is still poorly understood, especially the interrelationships and temporal sequence of the markers of Alzheimer's disease. In addition, the contribution of cerebral consequences of the vascular risk factors may be underestimated in the pathophysiological mechanisms of dementia. High blood pressure leads to abnormalities in the small blood vessels of the brain that may accelerate the onset of dementia.

My thesis work is based on multimodal data from the MEMENTO cohort (2011-2019) which includes 2323 participants from 28 French memory centers and followed-up regularly with repeated clinical, psychometric, biological, and brain imaging data. The objectives are (1) to describe the temporal sequence of evolution of multimodal markers using time-shift mixed models that allow to model the progression of markers according to a latent time of the disease and (2) to assess the pathways of action of high blood pressure and its control in the association with cognition using structural equation models that allow mediation analysis. For this thesis work I have studied in particular (i) neuropsychological test scores that assess memory, language, executive function and visual-spatial function; (ii) Markers of neurodegeneration: hippocampal volume, cerebral cortex thickness, glucose metabolism scan (PET FDG) and fraction of cranial volume occupied by the brain; (iii) Markers of neuropathology: measurements of cerebrospinal fluid concentrations of amyloid and tau peptides and amyloid peptide aggregation scan (amyloid PET); (iv) A marker of abnormalities of small blood vessels in the brain: the volume of white matter hyperintensities.

On the one hand, the results revealed uncertainty about the initiating role of amyloid peptide and the substantial influence of individual characteristics (genetic risk factor and level of education) on the temporal sequence of markers of disease progression. On the other hand, the results highlight the importance of the management and control of hypertension. Indeed, we have shown that uncontrolled hypertension is associated with a degradation of the brain structure and inferior cognitive performance.

This thesis work showed that modeling the progression according to the latent disease time could allow to better characterize individuals before the onset of symptoms, to improve their management, and to identify potential prevention axes. The strategies aiming at reinforcing blood pressure control could contribute to slowing brain degeneration and preventing dementia. The perspectives of this work include the exploitation of a larger number of markers and repeated data to describe more precisely the influence of dementia risk factors on the dynamics of the progression of the markers.

Keywords : dementia, Alzheimer's disease, biomarkers, hypertension, modelling

Unité de recherche

Unité INSERM U1219 – Bordeaux Population Health
Université de Bordeaux – ISPED
146 rue Léo Saignat 33076 Bordeaux, France

Remerciements

A madame Stéphanie Debette,

Je vous remercie d'avoir accepté de présider ce jury. Soyez assurée de toute ma gratitude.

A monsieur Thibault Mura,

Je suis honoré que vous ayez accepté de juger ce travail de thèse en qualité de rapporteur. Votre grande expérience dans l'analyse de données cliniques longitudinales m'apporta beaucoup. Je vous remercie.

A monsieur Etienne Dantan,

Je suis honoré que vous ayez accepté de juger ce travail de thèse. Votre expertise dans l'analyse de données longitudinales pour l'étude de l'histoire naturelle de maladies apportera un regard précieux à ce travail. Je vous remercie d'avoir accepté d'être rapporteur.

A monsieur Thomas Tourdias,

Je vous remercie d'avoir accepté de participer à ce jury. Votre expertise dans les techniques d'imagerie cérébrale apportera un éclairage nouveau à ce travail.

A monsieur Jean-Sébastien Vidal,

Je vous remercie d'avoir accepté de participer au jury de ma thèse. Soyez assuré de mon plus grand respect.

A madame Cécilia Samieri,

Je suis ravi que tu sois présente pour juger ce travail de thèse, après avoir participé au comité de suivi du projet et m'avoir encouragé. Je te remercie d'avoir accepté de participer à ce jury.

A madame Cécile Proust-Lima,

Merci beaucoup d'avoir dirigé ce travail de thèse. Je te remercie pour ton soutien, ta disponibilité, ta pédagogie et ta bienveillance tout au long de ce projet de thèse. Sois assurée de toute ma reconnaissance.

A madame Carole Dufouil,

Je te remercie sincèrement de m'avoir fait confiance en me proposant ce projet de thèse passionnant. Merci d'avoir dirigé ces travaux, ta rigueur scientifique d'épidémiologiste m'a beaucoup appris pendant ces 4 dernières années. Sois assurée de toute ma reconnaissance.

A l'équipe MEMENTO

Je remercie toutes les personnes de l'équipe MEMENTO pour ces années passées ensemble. Je garderai de très bons souvenirs de notre petite équipe stat (Vincent, Leslie, Jonviea, Federica, Ariane). Merci Vincent d'avoir toujours été disponible pour m'aider à comprendre les données de la cohorte. Je tiens à remercier aussi tous les participants de l'étude MEMENTO.

A l'équipe Biostatistique

Je remercie toute l'équipe Biostatistique pour votre accueil et l'organisation des séminaires et SOS qui ont souvent éveillé ma curiosité. J'ai beaucoup apprécié l'entraide dans l'équipe. Je remercie aussi Pierre-Louis du CREDIM pour sa réactivité et son travail sur la gestion du serveur de calcul TURING.

A mes amis,

Je tiens à remercier tous mes amis qui ont toujours été présents pour me soutenir et profiter de la vie, Camille, Loïc, Christophe, Adrien, Clément, Anaïs, Charlotte, Marie, Jade, Noémie, et ceux m'ont soutenu à distance Paul, Auguste, Clémence, Simon et Flavie.

A ma famille,

Mes chers parents je vous remercie infiniment. Vous m'avez toujours soutenu dans mes choix et vous m'avez offert les moyens d'être heureux. Je n'aurais pas réalisé mon parcours universitaire sans vous. Vous êtes formidables. Ma chère sœur, Salomé, je te remercie d'être là et de partager avec moi cette passion pour la montagne. Ton énergie est toute autant épuisante pour ton entourage qu'elle en est une source d'inspiration. Je suis fier d'être ton frère et je te souhaite le meilleur dans ta carrière d'urgentiste.

A Justine,

Merci pour ton soutien dans les moments les plus difficiles. J'ai de la chance de t'avoir dans ma vie, tu as su être patiente et me rassurer quand il fallait. J'ai hâte de pouvoir repartir à l'aventure avec toi.

A mon vélo, pour nos évasions salutaires en altitude.

Valorisations scientifiques

Articles dans le cadre du projet de thèse

- Lespinasse, J., Chêne, G., Mangin, J.-F., Dubois, B., Blanc, F., Paquet, C., Hanon, O., Planche, V., Gabelle, A., Ceccaldi, M., Annweiler, C., Krolak-Salmon, P., Godefroy, O., Wallon, D., Sauvée, M., Bergeret, S., Chupin, M., Proust-Lima, C., Dufouil, C., & group for the M. study. (2022). Associations among hypertension, dementia biomarkers, and cognition: The MEMENTO cohort. *Alzheimer's & Dementia*. <https://doi.org/https://doi.org/10.1002/alz.12866>
- Lespinasse, J., Dufouil, C., Proust-Lima, C., & group for the M. study. Disease progression model anchored around clinical diagnosis in longitudinal cohorts: example of Alzheimer's disease and related dementia. **(En préparation)**

Autres contributions dans le domaine du projet de thèse

- Rouanet, A., Avila-Rieger, J., Dugravot, A., Lespinasse, J., Stuckwisch, R., Merrick, R., Anderson, E., Long, L., Helmer, C., Jacqmin-Gadda, H., Dufouil, C., Judd, S., Manly, J., Sabia, S., Gross, A., & Proust-Lima, C. (2022). How Selection Over Time Contributes to the Inconsistency of the Association Between Sex/Gender and Cognitive Decline Across Cognitive Aging Cohorts. *American Journal of Epidemiology*, *191*(3), 441–452. <https://doi.org/10.1093/aje/kwab227>
- Kühnel, L., Bouteloup, V., Lespinasse, J., Chêne, G., Dufouil, C., Molinuevo, J. L., & Raket, L. L. (2021). Personalized prediction of progression in pre-dementia patients based on individual biomarker profile: A development and validation study. *Alzheimer's and Dementia*, *17*(12), 1938–1949. <https://doi.org/10.1002/alz.12363>

Présentations orales

- **AAIC 2021** (Alzheimer's Association International Conference), *Virtual event*
“Modeling the pathological cascade of Alzheimer's disease and related dementia markers: the MEMENTO cohort”.
- **ISCB 2020** (International Society for Clinical Biostatistics), *Virtual event*
“Modeling multivariate longitudinal markers in a latent disease time scale: Application in the MEMENTO cohort to Alzheimer's disease and related dementia”
- **MELODEM 2020** (MEthods for LOngitudinal studies in DEMentia), *Virtual event*
“Modeling of multivariate longitudinal markers in a latent disease time scale”
- **AAIC 2019** (Alzheimer's Association International Conference), *Los Angeles*
“Association between hypertension and Alzheimer's disease and related causes of dementia: a pathway analysis in the MEMENTO cohort”.

Liste des abréviations

$A\beta_{42}$ = protéine Amyloïde β 42
 $A\beta_{40}$ = protéine Amyloïde β 40
APOE = gène de l'apolipoprotéine E
APOE4 = génotype de l'allèle $\epsilon 4$ de l'*APOE*
ATN = Amyloïde ;Tau ;Neurodégénérescence (classification)
CDR = « Clinical Dementia Rating » pour évaluation clinique de la démence
CDR-SB = « CDR Sum of Boxes »
CSF = « cerebrospinal fluid » pour liquide céphalo rachidien
DAG = « Directed Acyclic Graph » pour graphe acyclique orienté
FDG = 18F-fluorodésoxyglucose
FIML = « Full Information Maximum Likelihood » pour maximum de vraisemblance
HMC. « Hamiltonian Monte Carlo » pour algorithme de Monte Carlo Hamiltonien
HR = « Hazard Ratio » pour rapport de risques instantanés
HSB = Hypersignaux de la Substance Blanche
HTA = hypertension artérielle
IC95% = Intervalle de Confiance à 95%
IRM = Imagerie par Résonance Magnétique
LCR = Liquide Céphalo Rachidien
MAMA = Maladie d'Alzheimer et Maladies Apparentées
MAR = « Missing At Random » pour manquant au hasard
MCAR = « Missing Completely At Random » pour manquant complètement au hasard
MCI = « Mild Cognitive Impairment » pour déficit cognitif léger
MCMC = « Markov Chain Monte Carlo »
PAD = Pression Artérielle Diastolique
PAS = Pression Artérielle Systolique
p-tau = protéine tau phosphorylée
RC = Rapport de Côtes
RLRI-16 = rappels libres et indicés à 16 items
RMSE = « Root Mean Squared Error » pour racine de l'erreur quadratique moyenne
sd = « standard deviation » pour écart-type
SEM = « Structural Equation Model » pour modèle à équations structurelles
SUVr = « Standard Uptake Value ratio »
TEP = Tomographie par Emission de Positrons
TDP-43 = protéine 43 liée à l'ADN TAR
TMT-A et TMT-B = Trail Making Test version A et version B
t-tau = protéine tau totale

Liste des figures

Figure 1. Proportion des étiologies du trouble neurocognitif en fonction de l'âge. Adapté de Knopman et al. 2021.....	23
Figure 2. Représentation schématique des principaux tissus du cerveau. Adapté de Mastorakos et McGavern (2019) et Blausen Medical, BruceBlas.....	24
Figure 3. Stades de Braak qui caractérisent la sévérité et la propagation des plaques de protéines amyloïdes (à gauche) et des enchevêtrements neurofibrillaires de protéines tau (à droite). Adapté de Braak et Braak 1991.....	25
Figure 4. Infographie résumant la nature et la physiopathologie ciblée des principaux biomarqueurs de la maladie d'Alzheimer. Adapté de Hampel et al. 2021.....	33
Figure 5. Représentation schématique des marqueurs d'imagerie de la maladie des petits vaisseaux du cerveau. Adapté de Ter Telgte et al. 2018.	34
Figure 6. Évolution de la prévalence en fonction de l'âge et du sexe. Estimation mondiale 2019. Adapté de Nichols et al. 2022	35
Figure 7. Modèle théorique de la dynamique des biomarqueurs de la maladie d'Alzheimer. Adapté de Jack et al. 2013	45
Figure 8. Estimation de la durée des stades de la maladie d'Alzheimer (N = 3268). Adapté de Vermunt et al. 2019	47
Figure 9. Séquence temporelle de biomarqueur dans l'étude ADNI. Exemple de résultats obtenus par un modèle basé sur les évènements (EBM). Adapté de Archetti et al. 2019.....	49
Figure 10. Modèle hypothétique de la dynamique de progression des biomarqueurs de la maladie d'Alzheimer ses principales difficultés statistiques. Adapté de Jack et al. 2013....	50
Figure 11. Progression et temporalité des biomarqueurs estimés avec un modèle à score de progression de la maladie d'Alzheimer. Adapté de Jedynak et al. 2012.....	53
Figure 12. Prédiction de la progression de la sévérité de 16 marqueurs (protéines amyloïde, protéines tau, dégénérescence et cognition) dans l'étude ADNI (N=1554) obtenues avec le modèle LTJMM. Adapté de Li et al. 2018.....	54
Figure 13. Localisations des 26 centres de mémoire, de ressources et de recherche qui participent à l'étude MEMENTO et nombres de participants recrutés par centre. Cohorte MEMENTO (2011-2014).	58

Figure 14. Nombre de participants et de mesures répétées pour les différentes évaluations et examens cliniques. Cohorte MEMENTO 2011-2021	64
Figure 15. Incidences cumulées de démence et de décès sans démence par l'estimateur non paramétrique de Aalen Johansen. Cohorte MEMENTO 2011-2021.....	64
Figure 16. Histogramme des temps de mesure depuis l'inclusion selon le type d'examen. Cohorte MEMENTO (2011-2021).	66
Figure 17. Représentation graphique d'un système d'équation structurelle.....	68
Figure 18. Représentation graphique d'un système d'équation structurelle qui inclus une variable explicative observée.....	69
Figure 19. Représentation schématique d'une variable latente. Adapté de Hair et al. 2014.....	70
Figure 20. Représentation graphique d'une équation de mesure.....	70
Figure 21. Illustration d'un facteur de confusion.....	72
Figure 22. Illustration d'un médiateur d'effet.....	72
Figure 23. Illustration d'un modificateur d'effet.....	73
Figure 24. Exemple de prise en compte de l'hétérogénéité inter-individuelle par recalage temporel pour estimer les trajectoires à long terme des marqueurs au cours de la progression de la maladie.	82
Figure 25. Illustration de la construction de l'échelle de sévérité des marqueurs par la pondération ou la non-pondération des percentiles.....	87
Figure 26. Exemple de chaînes MCMC qui ne convergent pas vers une distribution stationnaire commune.....	97
Figure 27. Exemple de chaînes MCMC ayant convergées sur la même distribution stationnaire.....	97
Figure 28. Distribution postérieure des temps de recalage individuels estimés à partir d'un modèle de progression non-contraint. Cohorte MEMENTO 2011-2019 (N=2186).....	143
Figure 29. Residual mean square error per marker for the main disease progression anchored model (DPAM) with constraint of 1.5 years around the observed clinical diagnosis, the DPAM with a weaker constraint of 3 years, and a non-anchored disease progression..	144
Figure 30. Facteur d'accélération de l'échantillonnage HMC-NUTS (Stan) avec calcul parallèle en fonction du nombre de processeurs à partir d'un jeu de données simulé.....	148

Liste des tableaux

Tableau 1. Calendrier prévisionnel du recueil de données dans l'étude MEMENTO. Adapté de Dufouil et al. 2017.	59
Tableau 2. Caractéristiques des participants de la cohorte MEMENTO à l'inclusion dans l'étude. Cohorte MEMENTO 2011-2014.	65
Tableau 3. Caractéristiques des principaux travaux sur les modèles de progression de la MAMA dans la littérature ; k marqueur ($k=1, \dots, K$) ; i sujet ($i=1, \dots, N$) ; j occasion ($j=1, \dots, n_{ik}$)	84
Tableau 4. e-Table 1. Distribution of missing data for each variable of interest by hypertension status – N=2048, The MEMENTO cohort, 2011-2014	113
Tableau 5. e-Table 2. Estimates of direct and total effects of covariates on dementia biomarkers and cognition computed from structural equation model - The MEMENTO cohort, 2011-2014	113
Tableau 6. e-Table 3. Estimates of the factor loadings and residual variances of the observed variables for each latent variable – N=2048, The MEMENTO cohort, 2011-2014	114
Tableau 7. e-Table 4. Estimates of associations between hypertension status on dementia biomarkers and cognition according to sex. The MEMENTO cohort, 2011-2014.	114
Tableau 8. e-Table 5. Estimates of associations between hypertension status on dementia biomarkers and cognition according to APOE4 status. The MEMENTO cohort, 2011-2014.....	115
Tableau 9. e-Table 6. Estimates of associations between hypertension status on dementia biomarkers and cognition according to age. The MEMENTO cohort, 2011-2014	116

Tables des matières

1	<i>Préambule</i>	15
2	<i>Introduction</i>	18
2.1	La démence	18
2.1.1	Histoire	18
2.1.2	Critères diagnostiques et sévérité	19
2.1.3	Étiologies des démences	22
2.1.4	Neuropathologie	24
2.2	Biomarqueurs	29
2.2.1	Biomarqueurs de neuropathologie de type Alzheimer	29
2.2.2	Biomarqueurs de mort des neurones	31
2.2.3	Biomarqueurs de neuropathologies cérébrales vasculaires.....	34
2.3	Épidémiologie de la démence	35
2.3.1	Prévalence et incidence	35
2.3.2	Facteurs de risque	36
2.3.3	Prise en charge et voies d'actions de l'hypertension artérielle	40
2.3.4	Incidence de la démence	42
2.4	Temporalité des atteintes dans la MAMA	44
2.4.1	Modèle théorique de la séquence des atteintes	44
2.4.2	Temporalité de la MAMA	46
2.4.3	Séquence des atteintes : une approche évènementielle	48
2.4.4	Séquence des atteintes : une approche par le temps de maladie latent	49
2.5	Problématiques et objectifs du travail de thèse	56
3	<i>Matériel et méthodes</i>	57
3.1	Applications aux données de la cohorte clinique MEMENTO	57
3.1.1	La cohorte.....	57
3.1.2	Critères d'inclusions	57
3.1.3	Calendrier du recueil des données	58
3.1.4	Facteurs de risque, marqueurs de MAMA et évaluations cognitives	60
3.1.5	Description de la population	62
3.2	Études des chemins d'association par les modèles à équations structurelles	67
3.2.1	Représentation graphique d'un SEM	67
3.2.2	Système d'équations structurelles.....	68
3.2.3	Équation d'observation	69

3.2.4	Étude des mécanismes d'association.....	71
3.2.5	Estimation d'un SEM	73
3.2.6	Évaluation des performances.....	75
3.3	Études des marqueurs de progression par des modèles à temps de maladie	
latent	78	
3.3.1	Modèles mixtes	78
3.3.2	Modèles de progression.....	80
3.3.3	Estimation des modèles	90
3.3.4	Prédictions.....	92
3.4	Inférence bayésienne avec HMC-NUTS et Stan	92
3.4.1	Algorithmes pour constituer une chaîne MCMC	92
3.4.2	Le logiciel Stan	95
3.4.3	Indicateurs diagnostiques	96
3.4.4	Ajustement des données et comparaison de modèles.....	99
4	<i>Association entre l'hypertension artérielle, la maladie des petits vaisseaux, les lésions de type Alzheimer, l'atrophie du cerveau et la cognition.....</i>	100
4.1	Résumé.....	100
4.2	Article 1: « Associations among hypertension, dementia biomarkers, and cognition: The MEMENTO cohort »	101
4.3	Compléments.....	113
4.3.1	Résultats supplémentaires de l'article	113
4.3.2	Spécification du modèle avec <i>R</i> et le package <i>lavaan</i>	116
4.3.3	Estimation du modèle avec <i>R</i> et le package <i>lavaan</i>	118
4.4	Conclusions.....	119
5	<i>Modèle de progression de la MAMA : Applications à la cohorte MEMENTO. 121</i>	
5.1	Résumé.....	121
5.2	Article 2 : « Disease progression model anchored around clinical diagnosis in longitudinal cohorts: example of Alzheimer's disease and related dementia »	122
5.3	Compléments.....	143
5.3.1	Résultats supplémentaires de l'article.....	143
5.3.2	Développement par étape du modèle de progression autour du diagnostic	144
5.3.3	Parallélisation des calculs de la distribution postérieure $P(\theta X)$	146
5.3.4	Spécification du modèle avec Stan	149

5.4	Conclusions.....	153
6	<i>Discussion générale</i>	154
6.1	Résumé des travaux et analyses complémentaires	154
6.2	Association entre l'hypertension artérielle, la maladie des petits vaisseaux, les lésions de type Alzheimer, l'atrophie du cerveau et la cognition	154
6.3	Modèle de progression de la MAMA : applications à la cohorte MEMENTO	155
6.4	Extrapolation des résultats à la population générale	156
6.5	Implications des facteurs de risques vasculaires	158
6.6	Conclusion générale	159
7	<i>Références</i>	161

1 PREAMBULE

La démence se caractérise par une réduction progressive des capacités dans un ou plusieurs domaines cognitifs et se traduit par une dégradation de l'autonomie et de l'aptitude à réaliser les activités quotidiennes. Les conséquences sur les plans physiques, psychologiques et sociaux sont considérables pour les patients mais aussi pour les proches et le personnel soignant. La prévalence de la démence augmente fortement avec l'âge qui est le principal facteur de risque. Or l'espérance de vie à la naissance a fortement augmenté au niveau mondial elle était de 64 ans en 1990 et elle devrait atteindre 77,2 ans en 2050 avec la plus forte augmentation enregistrée dans les pays de l'hémisphère sud. Les transitions démographiques successives des pays s'accompagnent d'un vieillissement global de la population mondiale, avec une proportion d'individus âgés de plus de 65 ans qui devrait passer de 9,7% en 2022 à 16,4% en 2050, ce qui se traduit par une augmentation importante du nombre attendu de cas de démence (United Nations Department of Economic and Social Affairs 2022). La prévalence dans le monde est évaluée à 50 millions de personnes et elle pourrait tripler à l'horizon de 2050 (Nichols et al. 2022). Malgré les nombreux efforts de recherche réalisés au cours des 3 dernières décennies, la démence reste un enjeu majeur de santé publique et à ce jour il n'existe pas de traitement efficace pour ralentir ou stopper la progression de la maladie.

La démence est un syndrome causé par une variété d'étiologies parmi lesquelles la maladie d'Alzheimer est considérée comme la plus fréquente. Cependant de nombreuses études documentent la coexistence de plusieurs étiologies à l'origine de démences dites mixtes et plus particulièrement la contribution des anomalies vasculaires cérébrales aux troubles cognitifs et à la démence pourrait être sous-estimée (Kapasi, DeCarli, and Schneider 2017). Les études épidémiologiques ont permis également de mettre en évidence l'origine multifactorielle de la démence avec l'identification de prédispositions génétiques ainsi que des facteurs de risques modifiables, notamment cardiovasculaires (Livingston et al. 2017).

L'essor des biomarqueurs au cours des 20 dernières années a permis de nombreux progrès techniques dans le domaine de la recherche biomédicale qui ont permis d'identifier des biomarqueurs *in vivo* associés au vieillissement et à la démence, d'une part avec les nouvelles techniques d'imagerie cérébrale et d'autre part avec l'évolution des méthodes de

dosage protéique. Il est aujourd'hui possible de quantifier les lésions spécifiques de type Alzheimer, les anomalies vasculaires cérébrales, et l'atrophie des structures du cerveau.

L'absence de symptômes dans les premiers stades pré-clinique de la progression de la maladie d'Alzheimer et des maladies apparentées (MAMA) complique l'étude de leur histoire naturelle. Les biomarqueurs permettent de compléter l'évaluation clinique en mesurant les atteintes infracliniques que l'on observe dans le cerveau des patients en amont du diagnostic.

Les études de cohortes permettent d'évaluer l'intérêt des biomarqueurs depuis les signes les plus précoces jusqu'aux stades sévères de détérioration cognitive et de caractériser les phénotypes multidimensionnels des patients. Les mesures répétées dans le temps contribuent à la compréhension des facteurs de risques et des mécanismes physiopathologiques des démences. Cependant il existe une grande hétérogénéité intra-individuelle et inter-individuelle dans les données recueillies qui pose de nouveaux défis en termes d'analyses et nécessite de nouvelles approches statistiques afin d'exploiter au mieux les aspects multidimensionnels et temporels des mesures longitudinales des marqueurs de MAMA.

Les travaux de la thèse visent à développer et utiliser des approches multidimensionnelles en exploitant un ensemble de données cliniques, psychométriques, biologiques et de neuroimagerie recueillies dans le cadre de l'étude MEMENTO : une cohorte française multicentrique avec un suivi clinique intensif de 2323 participants à risque de développer une MAMA. La première partie de la thèse introduit les connaissances nécessaires à la compréhension de l'épidémiologie de la démence et des problèmes statistiques posés par l'étude de la MAMA afin de définir le contexte et les objectifs dans lesquels s'inscrivent les recherches réalisées. La deuxième partie présente l'étude MEMENTO et les outils statistiques utilisés pour répondre aux objectifs. Par la suite les travaux de recherche de la thèse s'articulent autour de 2 axes. Le premier axe a pour objectif d'investiguer les associations et voies d'actions entre les marqueurs de MAMA dans le cadre du lien entre hypertension et cognition en utilisant des modèles à équations structurelles. Le deuxième axe porte sur la modélisation du temps de maladie pour décrire à partir d'observations individuelles à court terme la progression à long terme des biomarqueurs étudiés. Les résultats permettent d'une part de suggérer une description de la séquence temporelle des événements de l'histoire naturelle de la MAMA et d'autre part d'étudier leurs interrelations. La dernière partie aborde

les forces et limites des résultats à l'issue de ces travaux ainsi que de leurs perspectives dans le contexte de la littérature scientifique actuelle.

2 INTRODUCTION

2.1 La démence

La démence est un état confusionnel suivi d'un syndrome défini par un ensemble de signes cliniques et de symptômes qui se manifestent progressivement et pour lequel il existe un ensemble de sous-types étiologiques (American Psychiatric Association 2013).

2.1.1 Histoire

L'étymologie du mot démence trouve son origine dans les termes latin « de » (sans) et « ment » (esprit). Au XVIIIe et au début du XIXe siècles la démence est souvent considérée comme une conséquence normale du vieillissement, les termes « démence » et « sénilité » sont utilisés pour indiquer l'incapacité à s'occuper de soi-même physiquement et mentalement (Borenstein and Mortimer 2016a).

La principale cause de la démence est la maladie d'Alzheimer (Gauthier S, Rosa-Neto P, Morais JA 2021) dont on attribue la découverte au Dr. Alois Alzheimer qui est le premier médecin à décrire en 1906 le cas d'une patiente âgée de 51 ans avec des symptômes psychiatriques et cognitifs d'évolution progressive (Möller and Graeber 1998). C'est l'examen anatomopathologique de son cerveau qui a permis d'identifier la présence de deux types de lésions particulières : d'une part des plaques séniles, petites, très nombreuses et diffuses dans le cortex cérébral ; d'autre part des enchevêtrements neurofibrillaires à l'intérieur des neurones. La maladie d'Alzheimer connue alors sous le nom de démence présénile est une maladie qui survient à un âge relativement jeune et se distingue de la démence sénile souvent considérée comme la conséquence naturelle du vieillissement (Borenstein and Mortimer 2016b).

Cette différenciation va persister jusqu'aux années 60 quand des études d'autopsie pour étudier la neuropathologie des cerveaux de patients déments vont montrer la présence des mêmes lésions caractéristiques et conclure que la démence présénile et la démence sénile à des âges avancés sont causées par la même maladie (Tomlinson, Blessed, and Roth 1968; 1970). La vision clinique de la démence évolue puisqu'elle n'est plus considérée comme un trouble psychiatrique mais une pathologie avec une cause biologique bien identifiée.

A partir de 1940 l'intérêt pour les maladies psychiatriques évolue à la suite des nombreux traumatismes de la seconde guerre mondiale. Aux Etats-Unis les pensions militaires pour les troubles psychiatriques concernaient un demi-million de vétérans (Scull 2010). Les efforts d'harmonisation de la classification des troubles mentaux motivent la publication par l'association américaine de psychiatrie du manuel diagnostique et statistique des troubles mentaux (DSM), qui intègre les critères diagnostiques de la démence. Ces critères ont évolué dans les différentes versions de l'ouvrage et les avancées des connaissances scientifiques.

2.1.2 Critères diagnostiques et sévérité

DSM-IV-TR (2000)

Les critères diagnostiques de la 4^{ème} édition du DSM (American Psychiatric Association 2000) sont encore utilisés en clinique notamment dans le cadre de protocoles d'études (Alpérovitch et al. 2003; Roberts et al. 2008; Ellis et al. 2009; Dufouil et al. 2017) pour définir un cas de démence qui se caractérise par : (1) une atteinte de la mémoire dans les capacités d'apprentissage (mémoire à court terme) et de restitution de l'information (mémoire des faits récents et à long terme), (2) une atteinte d'un autre domaine cognitif incluant les troubles du langage (aphasie), de la reconnaissance des objets (agnosie), des fonctions motrices (apraxie) et des fonctions exécutives, (3) un trouble du comportement social et de l'attention consécutif aux déficits cognitifs identifiés, et (4) ces troubles présentent une évolution progressive et ne dépendent pas d'un épisode délirant.

DSM-V (2013)

La dernière version du DSM propose une définition moins restrictive que le DSM-IV-TR sur les atteintes cognitives et préfère le terme de trouble cognitif majeur (plutôt que démence) qui est défini comme une réduction acquise, significative et évolutive des capacités cognitives suffisamment importante pour ne plus être capable d'effectuer seul les activités de la vie quotidienne selon les critères suivant (American Psychiatric Association 2013) :

- A. Preuves d'un déclin cognitif significatif par rapport à un niveau antérieur de fonctionnement dans un ou plusieurs domaines cognitifs (attention complexe, fonctions exécutives, apprentissage et mémorisation, langage, activités perceptivo motrices ou cognition sociale) reposant sur (1) une préoccupation du sujet, d'un informant fiable, ou du clinicien concernant un déclin significatif du

fonctionnement cognitif ; et (2) une altération importante des performances cognitives, idéalement documentée par un bilan neuropsychologique standardisé ou, à défaut, par une évaluation clinique quantifiée.

- B. Les déficits cognitifs interfèrent avec l'autonomie dans les actes du quotidien (c'est à dire tout au moins une aide nécessaire dans les activités instrumentales complexes de la vie quotidienne comme payer ses factures ou gérer la prise de ses médicaments).
- C. Les déficits cognitifs ne surviennent pas exclusivement dans le contexte d'un état confusionnel (delirium).
- D. Les altérations cognitives ne sont pas mieux expliquées par un autre trouble mental (par exemple un trouble dépressif caractérisé, une schizophrénie).

Le DSM-V reconnaît un degré de sévérité moindre du trouble neurocognitif avec le trouble cognitif léger (ou MCI : « mild cognitive impairment ») qui était défini comme un trouble cognitif non-spécifié dans la version IV. Le trouble cognitif léger est à distinguer de la démence, il se caractérise par un léger déclin du fonctionnement cognitif et une altération modeste des performances cognitives (critère A) plus importante que celle attendue dans le cadre du vieillissement naturel sans pour autant impacter l'autonomie du patient (critères B) (American Psychiatric Association 2013).

NIA-AA (2018)

Le NIA-AA (« National Institute of Ageing – Alzheimer's Association ») (Jack et al. 2018) propose une classification des critères diagnostiques appelée « continuum » qui souligne le caractère progressif du trouble cognitif chez le patient avec trois stades syndromiques « sans déficience cognitive », « trouble cognitif léger » et « démence » dans le cas spécifique où le trouble est dû à la maladie d'Alzheimer. Les stades syndromiques sont complétés par une classification clinique en 6 stades qui peut être résumée ainsi: (1) l'absence de déficience cognitive, (2) l'absence de déficience cognitive mais la présence d'un déclin par rapport à une évaluation antérieure, (3) déficience cognitive avec déclin par rapport à une évaluation antérieure, (4) démence légère avec trouble cognitif progressif substantiel, éventuellement des changements neurocomportementaux et nécessité d'une aide occasionnelle pour les activités de la vie quotidienne, (5) démence modérée avec trouble cognitif ou changements neurocomportementaux, dépendance à une aide fréquente pour les activités de la vie

quotidienne, (6) démence sévère avec dépendance complète pour effectuer les activités de la vie quotidienne y compris les soins personnels.

Évaluation clinique de la démence

En pratique clinique, la CDR (« Clinical Dementia Rating » ou évaluation clinique de la démence) est très utilisée pour réaliser une évaluation globale de la cognition et des troubles fonctionnels de la vie quotidienne (Morris 1993). Les informations nécessaires à l'évaluation des domaines cognitifs sont obtenues à l'aide d'un entretien semi-structuré du patient avec l'aide d'un informant si possible. L'évaluation comprend 6 domaines spécifiques qui sont la mémoire, l'orientation, le jugement /résolution de problèmes, les activités sociales, la vie à domicile / loisirs, ainsi que les soins personnels. Un algorithme permet de résumer les informations cliniques pour quantifier les symptômes selon une échelle divisée en 5 stades, notés de 0 à 3 : absence de symptômes (0), symptômes légers (0,5), démence légère (1), démence modérée (2) et démence grave (3).

La CDR « sum of the boxes » (CDR-SB) est une variante de la CDR généralement utilisée dans un cadre de recherche clinique. Elle permet une évaluation quantitative plus fine par la somme des 6 domaines qui sont chacun évalué de 0 à 3 sur l'échelle en 5 stades, le score total est compris entre 0 et 18 (O'Bryant et al. 2008).

Évaluation cognitive

Plusieurs tests neuropsychologiques permettent d'évaluer les atteintes cognitives et leur caractère progressif dans le cadre d'un trouble cognitif léger et d'une démence. Les tests permettent de cibler plus spécifiquement un domaine cognitif particulier. Le DSM-V définit 6 domaines (American Psychiatric Association 2013) :

- L'apprentissage et la mémoire : il s'agit principalement de la mémoire épisodique (immédiate et pour les faits récents).
- L'attention : par exemple maintenir l'attention au cours du temps, en présence d'un élément de distraction, ou réaliser deux tâches simultanément.
- Les fonctions exécutives : elles incluent la planification, la prise de décision, la mémoire de travail et la flexibilité mentale.
- Le langage : l'identification d'objets, la fluence verbale sémantique ou littérale et la compréhension.

- Les activités perceptivo motrices : elles incluent la perception visuelle, les gnosies (reconnaissance visuelle), la coordination du mouvement, et les praxies (mémorisation et apprentissage des gestes).
- La cognition sociale : par exemple la reconnaissance des émotions

Les 4 principaux domaines généralement affectés dans le cas de la démence sont la mémoire épisodique, les fonctions exécutives, le langage, et les activités perceptivo motrices (Weintraub, Wicklund, and Salmon 2012).

Principales présentations cliniques

Dans la pratique clinique on distingue la forme typique de la démence des autres formes dites démences atypiques.

La démence amnésique typique est le cas le plus fréquent avec une apparition des symptômes habituellement après 65 ans. Les symptômes sont principalement liés à un trouble de la mémoire et les autres fonctions cognitives peuvent être affectées ultérieurement avec la progression de la pathologie (McKhann et al. 1984). Dans la majorité des cas on estime que la démence typique est due à une maladie d'Alzheimer (Gauthier S, Rosa-Neto P, Morais JA 2021).

Les démences atypiques sont un groupe hétérogène en termes d'étiologies, d'atteintes cognitives et de l'âge au début des symptômes, habituellement avant 65 ans (McKhann et al. 1984). Les principaux symptômes cliniques ne se manifestent pas par une atteinte de la mémoire mais par une autre fonction cognitive (langage, attention, fonctions exécutives, trouble de la cognition sociale par exemple). La progression des symptômes est généralement plus rapide que les cas de démences typiques (Dubois et al. 2007).

2.1.3 Étiologies des démences

Il existe plusieurs types de démences, un groupe de travail international estimait en 2009 que la maladie d'Alzheimer est le plus fréquent avec une proportion de cas comprise entre 50 et 75%, suivi par la démence vasculaire qui concernerait 20 à 30% des cas. Moins fréquentes, les démences fronto-temporales seraient responsables de 5 à 10% des cas et les démences à corps de Lewy moins de 5% (Prince et al. 2009).

Des investigations cliniques supplémentaires permettent de rechercher la cause du syndrome démentiel. Généralement l'étiologie de la démence est identifiée à l'aide d'un ou

plusieurs examens cliniques par imagerie cérébrale ou prélèvement biologique qui suggèrent la mort des neurones ou la présence d'une neuropathologie de type Alzheimer (Dubois et al. 2007). Le diagnostic d'une maladie d'Alzheimer comme étiologie de la démence est établi en l'absence d'éléments supplémentaires suggérant la présence d'une étiologie mixte associée à une neuropathologie cérébrale vasculaire ou d'une autre maladie neurodégénérative (McKhann et al. 2011). Parmi les autres étiologies possibles il existe la sclérose hippocampique, l'accumulation de TDP-43 (protéine 43 liée à l'ADN TAR) particulièrement impliquée dans les cas de démence fronto-temporale et l'accumulation de l' α -synucléine impliquée dans les maladies à corps de Lewy, ou encore la maladie de Parkinson (Kapasi, DeCarli, and Schneider 2017).

Seule l'autopsie permet d'identifier avec certitude les lésions causées par les différentes maladies neurodégénératives et lésions cérébrales vasculaires. La figure 1 est issue d'une revue de la littérature d'études d'autopsie de cerveau (Nelson et al. 2011) et représente l'évolution de la prévalence des différentes neuropathologies responsable du trouble cognitif en fonction de l'âge. On observe qu'avec l'âge la présence des neuropathologies vasculaires cérébrales est nettement plus importante.

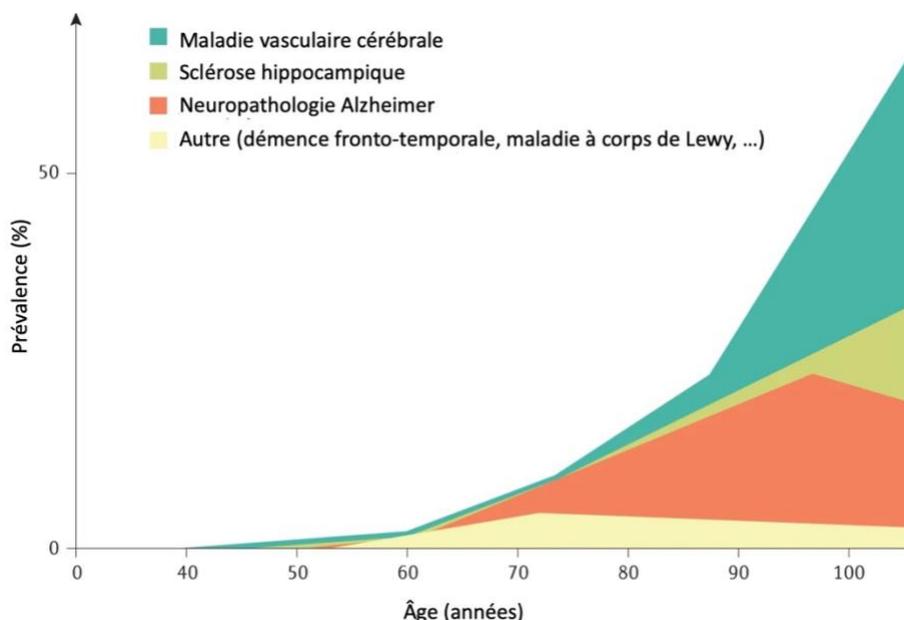


Figure 1. Proportion des étiologies du trouble neurocognitif en fonction de l'âge. Adapté de Knopman et al. 2021

Une étude à montrer qu'entre 2005 et 2010 la sensibilité et la spécificité du diagnostic de maladie d'Alzheimer « probable » pour 526 participants dont le cerveau a été autopsié étaient respectivement comprises entre 70,9%-76,6% et 59,5%-70,8% en fonction des critères retenus pour le nombre et la sévérité des lésions observées (Beach et al. 2012).

2.1.4 Neuropathologie

Le cerveau (figure 2) est constitué de 2 tissus principaux qui sont la substance grise et la substance blanche. La substance grise est localisée sur la partie externe du cerveau et contient les corps cellulaires des neurones, elle forme l'ensemble du cortex cérébral. La substance blanche localisée sur la partie interne contient les axones des neurones qui sont les fibres nerveuses conductrices de l'information. L'ensemble du cerveau est entouré par le liquide céphalo-rachidien (LCR) qui joue un rôle principal de protection en amortissant les mouvements de la boîte crânienne. L'oxygène, le glucose et les autres éléments nutritifs nécessaires au bon fonctionnement des cellules du cerveau sont apportés par le sang. La vascularisation du cerveau se fait par plusieurs artères cérébrales qui pénètrent la substance blanche et le cortex sous forme de petits vaisseaux (Brewer 2011).

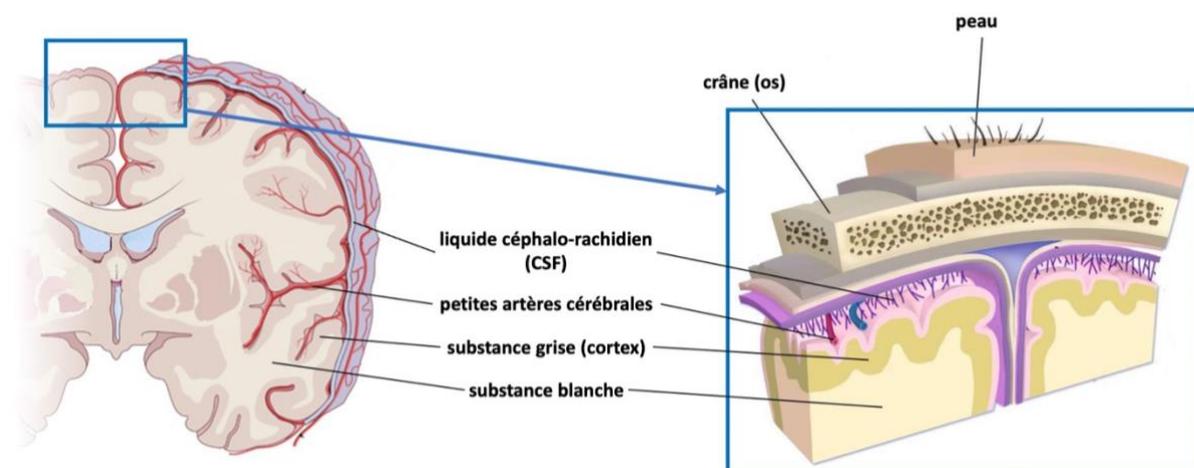


Figure 2. Représentation schématique des principaux tissus du cerveau. Adapté de Mastorakos et McGavern (2019) et Blausen Medical, BruceBlas

La neuropathologie de type Alzheimer se caractérise par la présence de deux protéinopathies qui sont des lésions consécutives aux dysfonctions des protéines amyloïdes et des protéines tau.

Protéines amyloïdes

Les plaques amyloïdes aussi connues sous le nom de plaques séniles sont formées principalement de l'agrégation de peptides amyloïde β_{42} ($A\beta_{42}$). Les plaques se situent dans l'espace extracellulaire des neurones et en particulier avec une localisation sous-corticale (Duyckaerts, Delatour, and Potier 2009). La localisation et l'étendue des lésions ont été

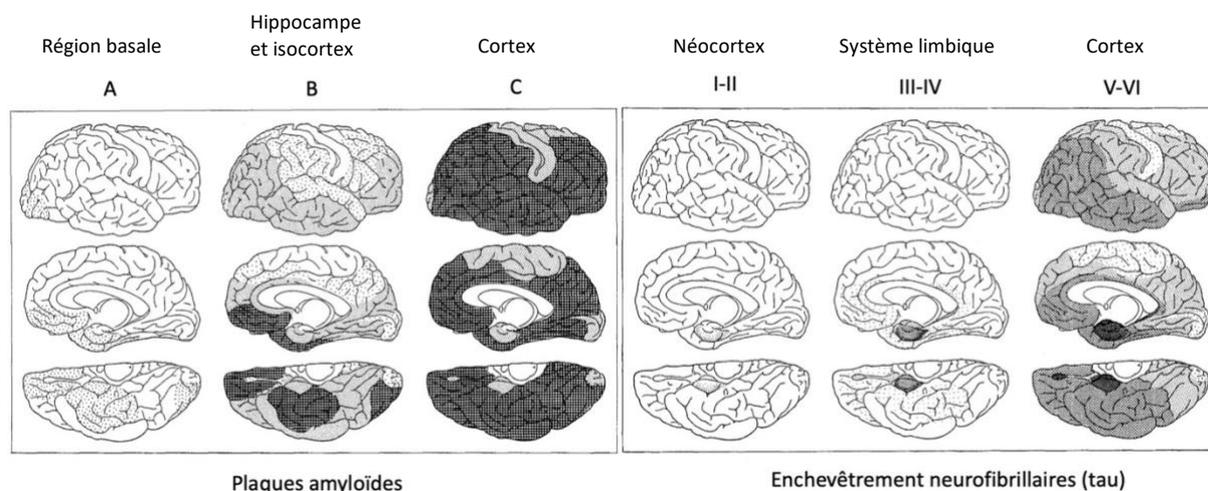


Figure 3. Stades de Braak qui caractérisent la sévérité et la propagation des plaques de protéines amyloïdes (à gauche) et des enchevêtrements neurofibrillaires de protéines tau (à droite). Adapté de Braak et Braak 1991.

étudiées par l'autopsie de 83 cerveaux de personnes démentes et non-démentes (Braak and Braak 1991) pour définir une classification en 3 stades d'évolution progressive de la déposition des protéines amyloïdes (figure 3) avec les premières lésions situées dans le néocortex : la partie basale des lobes frontal, temporal et occipital (stade A), puis elles atteignent l'hippocampe et une partie de l'isocortex : partie externe des lobes du cerveau (stade B) et s'étendent à l'ensemble du cortex (stade C).

Protéines tau

Les protéines tau sont des protéines très présentes dans l'environnement intracellulaire des neurones, elles interagissent avec les microtubules qui sont des structures permettant la caractérisation architecturale et mécanique des neurones : elles sont essentielles pour maintenir l'intégrité physique des neurones. Le changement de conformation de protéines tau serait dû à un défaut de leur processus d'activation par une phosphorylation incontrôlée (Fischer et al. 2009). Les protéines tau phosphorylées (p-tau) s'agrègent de manière à former des enchevêtrements neurofibrillaires et provoquer une dysfonction du neurone (Selkoe and Hardy 2016). Les études d'autopsie ont aussi permis de caractériser la localisation et la propagation des lésions neurofibrillaires avec les stades I/II (néocortex avec les cortex transentorhinal, entorhinal et l'hippocampe), les stades III/IV (système limbique) puis les stades V/VI (progression à l'ensemble du cortex) qui sont représentés dans la figure 3 (Braak and Braak 1991).

Le diagnostic étiologique différentiel des démences est d'autant plus complexe que les résultats d'analyses neuropathologiques démontrent l'existence d'une étiologie multiple pour

une majorité des cas de démences avec la coexistence d'une neuropathologie de type Alzheimer avec d'autres pathologies neurodégénératives et vasculaires (Kapasi, DeCarli, and Schneider 2017).

Coexistence de pathologies neurodégénératives

Une analyse de la neuropathologie des cerveaux de 766 patients atteints d'une pathologie neurodégénérative décrit que 65% des cas qui présentaient des lésions spécifiques de la maladie d'Alzheimer présentaient aussi d'autres types de lésions neurodégénératives (protéinopathies liées à TDP-43 et à l' α -synucléine) (Robinson et al. 2018).

Une autre analyse de 1024 cerveaux de participants à l'étude Religious Orders Study/Memory and Aging Project (ROSMAP), âgés de 89 ans en moyenne au moment du décès et dont 30,5% ne présentaient pas de troubles cognitifs, reporte que 94% des participants ont au moins une pathologie neuronale. Parmi eux seulement 9% avaient uniquement des lésions de type Alzheimer alors que 77,8% et 52,8% avaient respectivement plus de 2 ou 3 pathologies neurodégénératives (Boyle et al. 2018).

Similairement, au centre de la maladie d'Alzheimer de l'université de Kentucky, l'analyse des cerveaux de 375 participants avec un syndrome démentiel ou un déficit cognitif léger montre que 87% présentaient les deux types de lésions caractéristiques de la maladie d'Alzheimer (liées aux protéines amyloïdes et tau) dont 58% présentaient également une protéinopathie liée à TDP-43, à l' α -synucléine ou les deux (Karanth et al. 2020).

Neuropathologies cérébrales vasculaires

La leucoaraiose ou « leukoaraiosis » qui signifie raréfaction de la substance blanche désigne les lésions de la substance blanche qui provoquent un changement de son apparence (Hachinski, Potter, and Merskey 1987). On considère que la principale cause de la leucoaraiose est la maladie des petits vaisseaux. Le terme de maladie des petits vaisseaux du cerveau englobe tous les processus pathologiques qui affectent les petits vaisseaux du cerveau y compris les petites artères, les artérioles, les capillaires et les petites veines (Pantoni 2010).

Les causes et conséquences de la maladie des petits vaisseaux ne sont pas encore bien comprises, cependant le principal mécanisme sous-jacent suspecté est une atteinte de l'endothélium vasculaire (paroi interne des vaisseaux sanguins) provoquant des altérations de la régulation du transport du sang et des nutriments, des fuites sanguines et des dysfonctions de la barrière hémato-encéphalique (Wardlaw, Smith, and Dichgans 2019). Elle se caractérise

notamment par des changements structuraux au niveau de la substance blanche et de la substance grise sous-corticale. Il s'agit d'anormalités majoritairement microscopiques qui performent les petits vaisseaux (Pantoni 2010). Les petites lésions peuvent éventuellement régresser dans certains cas mais l'évolution de la neuropathologie montre fréquemment une augmentation du nombre de lésions et de leurs dimensions formant des petits infarctus ou lacunes (assimilable à une cavité dans la substance blanche) (Ter Telgte et al. 2018).

Plusieurs études ont montré que la maladie des petits vaisseaux pourrait contribuer à la survenue de la démence. Les données de 3913 participants provenant de 4 cohortes en population générale et âgés en moyenne de 69 à 80 ans ont été regroupées dans une méta-analyse. Les résultats montraient que la maladie des petits vaisseaux était associée à une augmentation du risque de survenue de démence de 39% (hazard ratio (HR)=1,39 ; intervalle de confiance à 95% (IC95%)=[1,00-1,94]), 47% (HR=1,47 ; IC95%=[0,97-2,22]) ou 25% (HR=1,25 ; IC95%=[0,66-2,38]) selon le type de lésions vasculaires étudié (D. Bos et al. 2018).

Les lésions associées à la maladie des petits vaisseaux sont suspectées de provoquer des altérations de la connectivité cérébrale (l'information nerveuse qui transite dans la substance blanche). Ces altérations seraient responsables de la dysfonction des neurones à distance dans le cortex cérébrale (substance grise) (Ter Telgte et al. 2018). Les résultats d'une analyse des images cérébrales de 32 participants ayant eu un accident vasculaire cérébral soutiennent l'existence d'une atrophie du cortex consécutive aux lésions identifiées dans la substance blanche (Duering et al. 2015). Les chercheurs ont utilisé l'imagerie de diffusion qui est une méthode d'imagerie par résonance magnétique (IRM) spécifique de tractographie pour mettre en évidence les faisceaux de neurones et associer les régions de la substance blanche et du cortex qui sont connectées entre elles. Les images réalisées 6 mois après l'évènement montraient que le cortex des régions constituées par les neurones connectés était significativement diminué comparativement à l'hémisphère cérébral opposé (utilisé comme référence). De plus les résultats d'une analyse de médiation qui portait sur les données de 519 participants montraient que l'association entre les lésions de la substance et la cognition était en partie expliquée par une plus petite épaisseur du cortex cérébral (Rizvi et al. 2018).

Enfin, la présence de lésions évoquant une neuropathologie cérébrale vasculaire est fréquemment concomitante avec la présence d'autres lésions neurodégénératives. Dans l'étude ROSMAP, 40% des participants qui présentaient des lésions de type Alzheimer avaient

également des lésions vasculaires (Boyle et al. 2018), et 20% à 60% des participants de l'étude menée à l'université de Kentucky présentaient une ou plusieurs pathologies neurodégénératives et des lésions cérébrales vasculaires (Karanth et al. 2020).

La maladie d'Alzheimer et les lésions vasculaires cérébrales sont très associées à l'âge et peuvent présenter des manifestations cliniques similaires (Pantoni, Garcia, and Brown 1996). Une revue de la littérature a estimé que la prévalence de la démence mixte définie par la coexistence d'une neuropathologie de type Alzheimer et de neuropathologies vasculaires cérébrales est fréquente puisqu'elle était comprise entre 22% et 45% dans plusieurs cohortes de participants principalement caucasiens (Chui and Ramirez-Gomez 2015).

2.2 Biomarqueurs

Un biomarqueur est un indicateur biologique mesurable d'un processus physiologique, pathologique ou de la réaction à une exposition ou une intervention. L'Organisation Mondiale de la Santé définit les biomarqueurs comme toute « substance, structure, ou processus qui peut être mesuré dans le corps ou ses produits et influence ou prédit une réponse » (World Health Organization 2001). Dans le contexte de la maladie d'Alzheimer et plus généralement des démences les biomarqueurs sont des indicateurs de la présence et de la sévérité des processus pathologiques qui sont associés à la maladie.

2.2.1 Biomarqueurs de neuropathologie de type Alzheimer

Plusieurs méthodes permettent aujourd'hui d'étudier et de quantifier *in vivo* les protéines amyloïdes et les protéines tau. L'amyloïdopathie et la tauopathie désignent respectivement les accumulations pathologiques au-delà de valeurs seuils des protéines amyloïdes et tau dans le cerveau.

Protéines amyloïde et tau dans le liquide céphalo rachidien

Les formes circulantes des protéines tau et amyloïde (en opposition aux formes agrégées) sont présentes dans le LCR qui peut être prélevé par ponction lombaire (Hansson 2021). Une revue de la littérature portant sur 231 publications comparant les biomarqueurs du LCR entre cas (maladie d'Alzheimer, N=15699) et témoins (volontaires sains ou patients ayant réalisé une ponction lombaire dans le cadre d'une hospitalisation, N=13018) met en évidence les associations entre les concentrations de plusieurs protéines d'intérêts et la maladie d'Alzheimer. Les auteurs concluent que la concentration en protéine A β ₄₂ est plus faible en moyenne de 44% (ratio=0,56 ; IC95%=[0,55-0,58]) chez les cas (Olsson et al. 2016). Le ratio des concentrations en protéine A β ₄₂/A β ₄₀ est également utilisé et pourrait être un meilleur biomarqueur de la neuropathologie de type Alzheimer que la concentration en protéine A β ₄₂ seule (Hansson et al. 2019). Inversement, les concentrations en protéines tau totale (t-tau) et en protéines tau phosphorylée (p-tau) sont respectivement supérieures en moyenne de 154% (ratio=2,54 ; IC95%=[2,44-2,64]) et 88% (ratio=1,88 ; IC95%=[1,79-1,97]) chez les cas (Olsson et al. 2016).

Imagerie moléculaire des protéines amyloïdes et tau

Les techniques d'imagerie moléculaire permettent de visualiser les agrégations protéiques présentes au niveau du tissu cérébral grâce au développement de différents radioligands. Ils sont habituellement administrés par voie veineuse et ont d'une part une capacité plus ou moins spécifique à se fixer à un élément d'intérêt et d'autre part la propriété d'être visible au scanner TEP (tomographie à émissions de positons) à l'aide d'un élément faiblement radioactif (généralement le fluor 18). Il existe aujourd'hui le 18F-flutemetamol, le 18F-florbetapir, le 18F-florbetaben et le composé B de Pittsburgh (11C-PIB) qui permettent de visualiser les agrégations de protéines amyloïdes, et le 18F-flortaucipir qui permet de visualiser les agrégations de protéines tau (Hansson 2021).

Des chercheurs ont réalisé l'autopsie du cerveau de participants décédés et ayant réalisé l'examen d'imagerie dans l'année passée (N = 29), en moyenne 99 jours séparaient l'imagerie du décès. L'étude a montré une bonne corrélation de 0,75 entre la quantification du signal obtenue par l'imagerie TEP au 18F-florbetapir et la sévérité des plaques amyloïdes au niveau du cortex (Clark et al. 2011). De manière similaire, 4 études analysées conjointement ont montré un niveau de corrélation à 0,74 entre le signal obtenu par l'imagerie TEP au 18F-flutemetamol et la présence de plaques amyloïdes sur des biopsies corticales de patients souffrant d'hydrocéphalie chronique (N = 52) (Rinne et al. 2012).

Des chercheurs ont aussi étudié la concordance des marqueurs de protéines amyloïdes en imagerie moléculaire avec la concentration en protéines $A\beta_{42}$ dans le LCR. Une analyse transversale sur 374 participants d'une cohorte clinique ayant eu une ponction lombaire et un examen d'imagerie TEP au 18F-florbetapir dans un délai moyen de 2 semaines a mis en évidence une corrélation négative de -0,74 entre les deux marqueurs (puisque la concentration en protéine $A\beta_{42}$ dans le LCR diminue avec la progression de la pathologie) (Landau et al. 2013). Dans une autre étude de cohorte, une analyse similaire a été réalisée sur un échantillon de 118 participants ayant eu une ponction lombaire et un examen d'imagerie moléculaire au 18F-flutemetamol, la corrélation était de -0,69 (Palmqvist et al. 2014).

Il existe un troisième groupe récent de biomarqueurs de l'accumulation des protéines amyloïde et tau dans le cerveau par le dosage plasmatique. La possibilité d'obtenir une information substantielle sur la maladie par une simple prise de sang est très attendue en pratique clinique. Cependant de plus amples recherches sont encore nécessaires pour

comprendre la performance de ces biomarqueurs et pouvoir les utiliser comme marqueurs de la maladie d'Alzheimer sous-jacente (Teunissen et al. 2022).

2.2.2 Biomarqueurs de mort des neurones

La dysfonction et la mort des neurones ont pour conséquence des processus d'atrophies également appelés dégénérescence. Les techniques d'imageries permettent d'étudier le phénomène par le biais d'une variété de biomarqueurs. En particulier les progrès de l'IRM permettent aujourd'hui d'obtenir un aperçu tridimensionnel du cerveau pour différencier les tissus et quantifier les volumes, surfaces et épaisseurs des différentes structures cérébrales. Il est possible d'évaluer la mort des neurones en mesurant l'atrophie de la substance grise en différentes localisation.

L'hippocampe

L'hippocampe est une structure du lobe temporal médian avec un rôle fondamental dans la fonction de mémorisation (J. C. S. Johnson et al. 2021). Dans l'étude MEMENTO, une précédente analyse sur 2157 participants ayant réalisé une IRM à l'inclusion montre qu'un volume plus important de l'hippocampe (+1 écart-type) est associé à une probabilité plus importante de 30% et 453% d'obtenir respectivement le score maximal à l'évaluation de la mémoire épisodique immédiate (rapport de cotes (RC)=1,30 ; IC95%=[1,15-1,46]) et différée (RC=4,53 ; IC95%=[4,72-6,35]) (Epelbaum et al. 2018). La diminution du volume global de l'hippocampe est particulièrement associée à la maladie d'Alzheimer, une analyse discriminante linéaire utilisant les mesures d'imagerie ajustées sur l'âge et le sexe permettait notamment de classer avec une précision de 81% (IC95%=[57%-100%]) un échantillon de 231 participants cognitivement sains de 198 participants avec un diagnostic de maladie d'Alzheimer (Wolz et al. 2011). Il s'agit aussi d'une localisation précoce pour les stades des lésions neurofibrillaires de Braak (Gosche et al. 2002).

L'épaisseur corticale

L'épaisseur du cortex permet d'étudier l'atrophie du cerveau de manière globale ou en fonction de différentes localisations anatomiques qui peuvent par exemple être définies avec l'atlas de Desikan-Killiany (Desikan et al. 2006). Des chercheurs ont notamment comparé des patients qui présentaient une démence légère de type Alzheimer (N=29 ; CDR=1) et des patients sans trouble cognitifs (N=115 ; CDR = 0) dans l'étude ADRC (« Washington University

AD Research Center ») pour identifier un groupe de 9 régions du cortex préférentiellement affectées par la maladie d'Alzheimer. En combinant les images précédentes avec celles de 3 autres groupes de patients du « Massachusetts General Hospital » (N=194 ; CDR=0 et N=76 ;CDR=1) ils ont confirmé la pertinence des régions étudiées en observant une diminution significative de l'épaisseur corticale pour ces 9 régions de 4,39% à 9,65% (Dickerson et al. 2009). Il existe une relation phénotypique entre les régions cérébrales atrophiées et les atteintes fonctionnelles des différents domaines cognitifs, notamment les lobes frontaux sont associés aux fonctions exécutives, les lobes temporaux au langage et les lobes occipitaux aux fonctions visuo-spatiales (J. C. S. Johnson et al. 2021).

L'atrophie globale

L'atrophie globale du cerveau peut être appréciée par le volume du cerveau qui est la proportion du volume intracrânien total occupée par les tissus cérébraux (par exemple le parenchyme formé des substances grise et blanche, cette proportion est appelée fraction du parenchyme). En effet le volume du crâne n'est pas susceptible de grandes variations à l'âge adulte comparativement au volume du cerveau (DeCarli et al. 2005), la diminution de cette proportion peut donc être utilisée comme le reflet d'une dégénérescence avec la mort des neurones. Des chercheurs ont réalisé une analyse multivariable dans une étude longitudinale incluant 156 participants avec un trouble cognitif léger (CDR=0,5) et ayant réalisé un examen d'imagerie. Le suivi clinique de 6,4 ans en moyenne a permis d'identifier 54 participants avec un diagnostic de démence. Les résultats indiquaient que chaque diminution de 1% de la fraction du parenchyme mesurée à l'entrée dans l'étude était associée à une augmentation du risque de conversion vers la démence de 10% (HR=1,10 ; IC95%=[1,02-1,19]) (Smith et al. 2008).

Le métabolisme du glucose

L'imagerie TEP au 18-fluorodésoxyglucose (FDG) permet d'évaluer l'activité cellulaire qui dépend du métabolisme glucidique. On utilise dans ce but un radio marqueur fluoré analogue du glucose, qui permet de visualiser l'intensité et la localisation de ce métabolisme. La diminution de l'intensité du marqueur dans la maladie d'Alzheimer est alors interprétée comme marqueurs des dysfonctions synaptiques et de la mort des neurones (K. A. Johnson et al. 2012). Les auteurs d'une revue de la littérature ont analysé 24 études ayant évalué la précision de l'imagerie TEP FDG à identifier les participants qui vont évoluer vers le diagnostic

de maladie d'Alzheimer parmi ceux avec un trouble cognitif léger et ont conclu qu'il existe une variabilité importante des performances de ce biomarqueur (Smailagic et al. 2018).

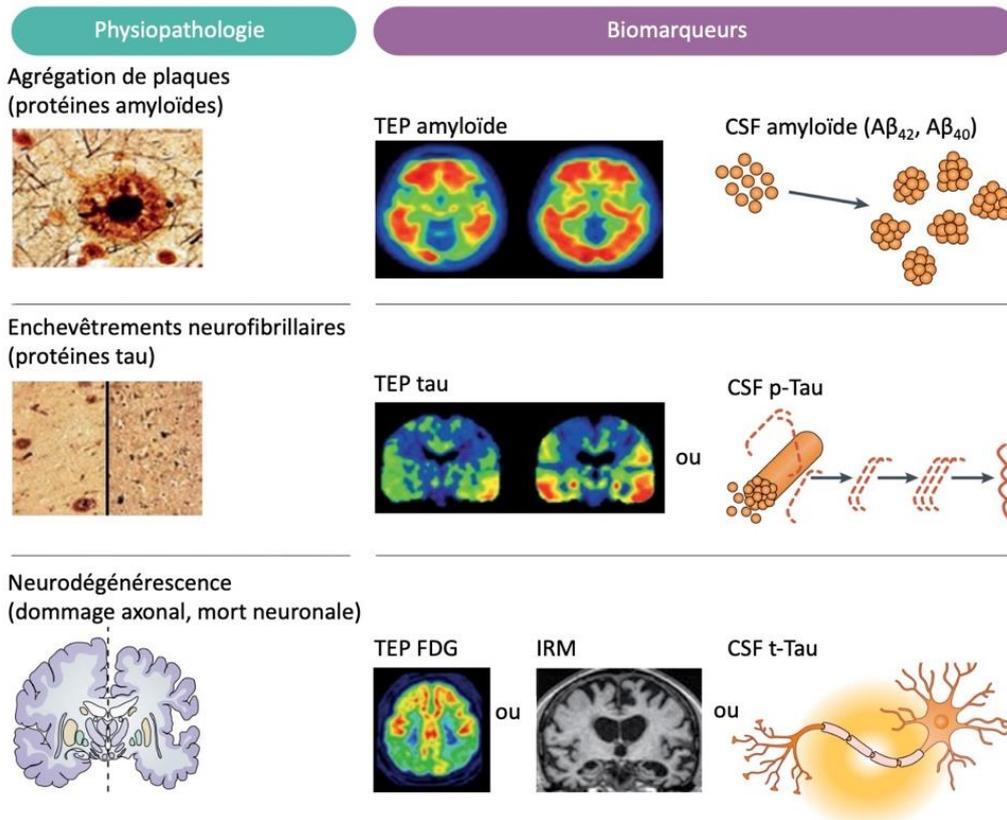


Figure 4. Infographie résumant la nature et la physiopathologie ciblée des principaux biomarqueurs de la maladie d'Alzheimer. Adapté de Hampel et al. 2021.

La figure 4 résume les principaux marqueurs de la maladie d'Alzheimer qui sont à ce jour bien établis dans la littérature scientifique (Hansson 2021) pour évaluer la sévérité de la pathologie et comprendre son évolution jusqu'à l'apparition d'un trouble neurocognitif majeur. On peut distinguer 3 sous-groupes qui permettent de quantifier l'accumulation pathologique de protéines amyloïdes, l'accumulation pathologique de protéines tau et la neurodégénérescence. A noter qu'il n'est pas certain que le dosage des protéines circulantes (non-agrégées) dans le LCR reflète le même phénomène que les lésions visualisées par les scanners TEP (protéines agrégées) (Mattsson et al. 2017). De plus il existe une distinction entre le dosage des protéines tau phosphorylée (p-tau) considéré comme le marqueur d'une dysfonction pathologique des protéines tau et l'ensemble des protéines tau (t-tau) considéré comme un marqueur de mort des neurones (neurodégénérescence) (Jack et al. 2018).

2.2.3 Biomarqueurs de neuropathologies cérébrales vasculaires

Plusieurs catégories de lésions sont identifiables avec une IRM (figure 5) et sont considérées comme des biomarqueurs de la maladie des petits vaisseaux : les hypersignaux de la substance blanche (HSB), les lacunes, les petits infarctus sous-corticaux, les micro-saignements, les espaces périvasculaires et l'atrophie du cortex (Ter Telgte et al. 2018).

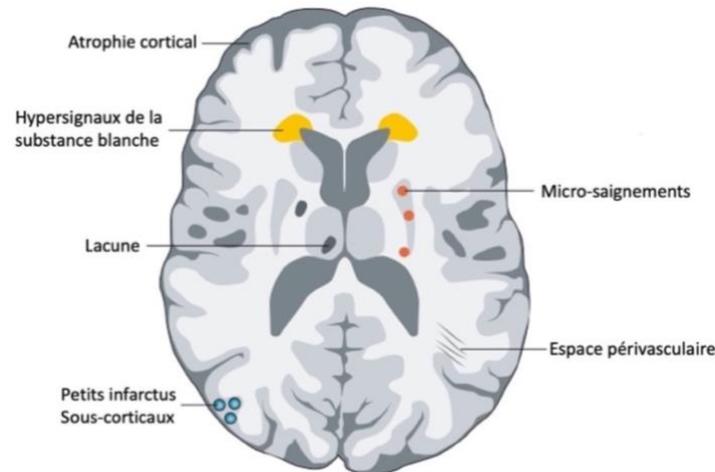


Figure 5. Représentation schématique des marqueurs d'imagerie de la maladie des petits vaisseaux du cerveau. Adapté de Ter Telgte et al. 2018.

Les auteurs d'une méta-analyse se sont intéressés à la signification clinique de ces lésions vasculaires cérébrales (Debette et al. 2019). Les données de 17 études incluant 2402 cas de démence et 9338 témoins ont permis de conclure que la présence d'HSB, d'infarctus cérébraux et de micro-saignements sont respectivement associées à une augmentation de 84% (HR=1,84 ; IC95%=[1,40-2,43]), de 29% (HR=1,29 ; IC95%=[1,02-1,65]) et de 41% (HR=1,41 ; IC95%=[0,90-2,21]) du risque de survenue de démence.

Les HSB représentent le principal biomarqueur de la maladie des petits vaisseaux (Gouw et al. 2011), ils sont un facteur de risque important en clinique pour la survenue d'accident vasculaire cérébral, de démence et de décès (Debette and Markus 2010). De plus leur prévalence dans la population est importante. Dans une précédente étude, les analyses de 1077 IRM en population générale montraient que chez les participants âgés de 60 à 70 ans, 68% avaient des HSB avec une localisation périventriculaire et 87% avec une localisation sous-corticale. Chez les participants âgés de 80 à 90 ans, 95% avait des HSB avec une localisation périventriculaire et 100% avec une localisation sous-corticale (Alber et al. 2019).

2.3 Épidémiologie de la démence

2.3.1 Prévalence et incidence

La prévalence de la démence a fortement augmenté avec la croissance de la population mondiale. Un groupe de travail de l'Organisation des Nations Unies explique cette croissance par l'augmentation de l'espérance de vie à la naissance (liée à une diminution de la mortalité), et par un taux de natalité mondiale qui reste élevé (United Nations Department of Economic and Social Affairs 2022). Des estimations récentes de la prévalence de la démence pour l'année 2019 et les prédictions pour l'année 2050 ont été obtenues à partir de l'étude GBD (« Global Burden of Disease ») qui s'appuie sur les données de plus de 204 pays et territoires. Les auteurs estiment que la prévalence en 2019 était de 54,7 millions (IC95%=[50,4-65,1]) et elle devrait atteindre 83,2 millions (IC95%=[73,0-94,6]) en 2030 et 152,8 millions (IC95%=[130,8-175,9]) en 2050 (Nichols et al. 2022). L'analyse des données socio-démographiques et des facteurs de risques de démence a permis aux auteurs de montrer que la prévalence augmente exponentiellement avec l'âge et elle est légèrement plus importante chez les femmes (figure 6). Les prédictions pour 2050 des prévalences globales rapportées à l'âge et au sexe montrent qu'elles seraient respectivement pour les hommes et les femmes de 0,5% (IC95%=[0,4-0,7]) et 0,6% (IC95%=[0,5-0,8]) chez les personnes âgées de 40 à 69 ans, de 6,5% (IC95%=[5,3-8,0]) et 8,5% (IC95%=[7,0-10,4]) chez les personnes âgées de 70 à 84 ans

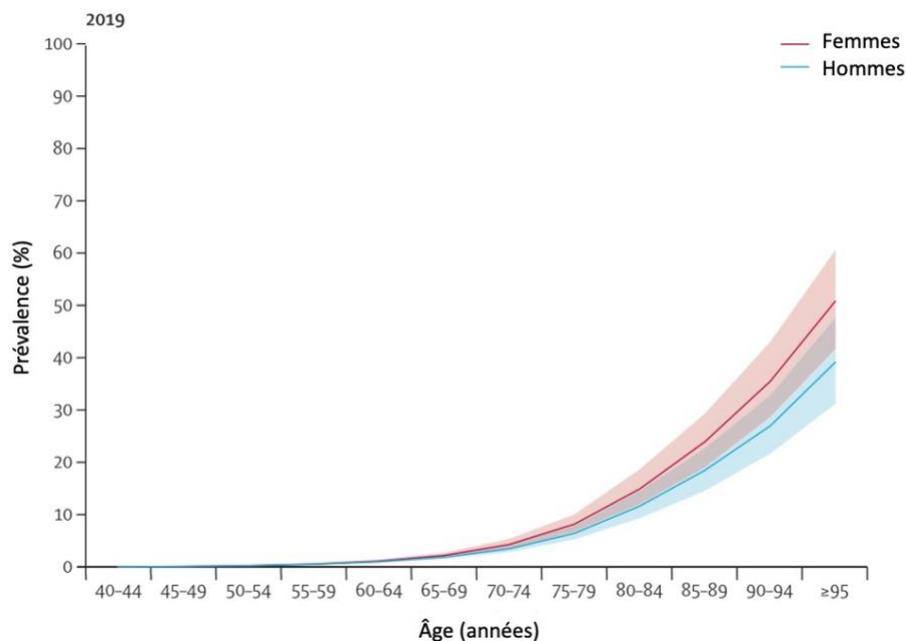


Figure 6. Évolution de la prévalence en fonction de l'âge et du sexe. Estimation mondiale 2019. Adapté de Nichols et al. 2022

et de 23,5% (IC95%=[19,1-29,2]) et 30,5% (IC95%=[25,0-36,8]) chez les individus âgés de plus de 85 ans (Nichols et al. 2022).

2.3.2 Facteurs de risque

La maladie d'Alzheimer et les autres démences apparentées ont une étiologie multifactorielle clairement établie. Des chercheurs ont étudié 11 884 paires de jumeaux monozygotes et dizygotes parmi lesquels 392 paires présentaient 1 (paire discordante) ou 2 (paire concordante) cas de démence. Les auteurs ont pu estimer la part d'héritabilité entre 58% et 79% en comparant les paires monozygotes (100% des gènes identiques) et les paires dizygotes (environ 50% des gènes identiques). Ils ont aussi estimé entre 21% et 42% la part attribuable à des facteurs environnementaux avec les données des paires monozygotes discordantes (Gatz et al. 2006).

Une revue de la littérature sur la prévention, les interventions et les soins concernant la démence a aussi établi en 2017 (Livingston et al. 2017) que 35% des cas de démences sont attribuables aux seuls facteurs de risques modifiables. Ces estimations renforcent l'importance de la recherche sur les rôles des facteurs de risque et la pertinence de la mise en place d'éventuelles stratégies de prévention de la démence. Parmi les principaux facteurs de risque modifiables on trouve l'obésité, le diabète, l'isolement social, l'absence d'activité physique, l'hypertension artérielle, la dépression, le tabagisme, un faible niveau d'étude et la perte d'audition.

Il existe une quarantaine de variants génétiques qui sont associés au risque de démence. Les paragraphes suivants développent plusieurs résultats concernant (1) l'allèle $\epsilon 4$ du gène de l'apolipoprotéine E (*APOE4*), qui est un facteur de risque génétique majeur de la maladie d'Alzheimer (Jansen et al. 2019), et le niveau d'études qui seront les principaux facteurs de confusion considérés dans le cadre des travaux de la thèse ; ainsi que (2) l'hypertension artérielle qui sera d'un intérêt spécifique pour l'étude de l'implication des facteurs de risque vasculaires dans l'histoire naturelle des démences.

Génotypes du gène APOE

Parmi les 65% de cas de démence attribuables à des facteurs non-modifiables 7% sont attribuables à l'allèle $\epsilon 4$ (Livingston et al. 2017). Il existe 3 formes principales du gène *APOE* dont les fréquences dans la population mondiale sont estimées à 78,3% pour $\epsilon 3$, 14,5% pour

$\epsilon 4$ et 6,4% pour $\epsilon 2$ (Eisenberg, Kuzawa, and Hayes 2010). Cette proportion varie en fonction de la population étudiée et du groupe ethnique (Corbo and Scacchp 1999), par exemple la fréquence de l'allèle $\epsilon 4$ augmente à mesure qu'on s'éloigne de l'équateur vers les populations qui ont évolué dans un environnement plus froid et diminue avec l'âge ce qui s'explique probablement par un biais de sélection lié aux comorbidités qui sont associées au gène (Eisenberg, Kuzawa, and Hayes 2010). L'analyse génétique de 337 484 britanniques caucasiens, âgés de 39 à 73 ans et sans lien de parenté qui participent à l'étude UKbiobank donne une estimation de la fréquence des 6 génotypes possibles en population générale : $\epsilon 3/\epsilon 3$ (58,3%), $\epsilon 3/\epsilon 4$ (23,9%), $\epsilon 3/\epsilon 2$ (12,3%), $\epsilon 4/\epsilon 2$ (2,6%), $\epsilon 4/\epsilon 4$ (2,4%), $\epsilon 2/\epsilon 2$ (0,6%) (Lumsden et al. 2020).

Les données génétiques de 7531 cas de maladie d'Alzheimer et 10132 témoins essentiellement caucasiens et européens ont montré qu'en fonction de l'âge la probabilité d'un diagnostic de maladie d'Alzheimer étaient 2,09 (IC95%=[1,61-2,71]) à 4,54 (IC95%=[4,09-5,05]) fois plus grande chez les porteurs hétérozygotes $\epsilon 4/\epsilon 3$ et 5,60 (IC95%=[3,17-9,89]) à 35,07 (IC95%=[23,8-51,68]) fois plus grande chez les porteurs homozygotes $\epsilon 4/\epsilon 4$ comparativement aux porteurs homozygotes $\epsilon 3/\epsilon 3$ (Genin et al. 2011).

Une équipe de recherche s'est plus récemment intéressée à la quantification de ce risque génétique d'un échantillon de 12 255 personnes non-démences à l'inclusion avec un suivi médian de 11 ans (van der Lee et al. 2018). Ils ont mis en évidence que la probabilité de survenue de la démence avant 85 ans était de 48,3% (IC95%=[40,1-57,3]) chez les porteurs homozygotes $\epsilon 4/\epsilon 4$, de 18,4% (IC95%=[16,5-20,4]) chez les porteurs hétérozygotes $\epsilon 3/\epsilon 4$ et $\epsilon 2/\epsilon 4$, de 8,6% (IC95%=[7,7-9,6]) chez les porteurs homozygotes $\epsilon 3/\epsilon 3$ et de 5,5% (IC95%=[4,1-7,4]) chez les porteurs homozygotes $\epsilon 2/\epsilon 2$ et hétérozygotes $\epsilon 3/\epsilon 2$. Les analyses incluaient également un ensemble d'autres variants génétiques qui étaient regroupés sous la forme d'un score de risque combiné aux génotypes de l'*APOE* pour étudier l'évolution de la probabilité de survenue de la maladie d'Alzheimer en fonction des caractéristiques génétiques et de l'âge. Les auteurs estiment que l'âge moyen au diagnostic de la maladie d'Alzheimer chez le groupe le plus à risque survient 18 à 29 ans plus tôt comparativement au groupe le moins à risque.

Le niveau d'études

Le niveau d'études est un facteur protecteur de déclin cognitif et de démence, parmi les 35% de cas attribuables à des facteurs de risque modifiables 8% sont attribuables à l'absence d'études secondaires (Livingston et al. 2017) ce qui correspond approximativement en France à un arrêt des études avant l'obtention du brevet des collèges. Les personnes concernées ont un risque de démence est 1,59 (IC95%=[1,26-2,01]) fois plus important relativement aux personnes avec un niveau d'études plus important.

Le niveau d'éducation est lié aux concepts de résiliences cognitives et cérébrales qui traduisent les différences de susceptibilité aux changements cérébraux liées à l'âge ou à la pathologie (Stern 2012). Ces différences de susceptibilité individuelles pourraient exercer une influence sur l'évolution de la maladie et son histoire naturelle. La résilience cérébrale représente les processus biologiques sous-jacents qui permettent à une personne de résister à la progression de la neuropathologie. La résilience cognitive se réfère aux processus fonctionnels qui permettent à une personne de maintenir un niveau cognitif supérieur à celui attendu comparativement à son degré d'atteinte de neuropathologie (Arenaza-Urquijo and Vemuri 2018).

L'hypertension artérielle

L'hypertension artérielle est définie par la mesure de la pression artérielle pour laquelle on distingue la pression artérielle systolique (phase de contraction du cœur) de la pression artérielle diastolique (phase de décontraction du cœur). En accord avec les recommandations internationales un individu est considéré hypertendu si la pression artérielle systolique est supérieure ou égale à 140mmHg ou bien si la pression artérielle diastolique est supérieure ou égale à 90mmHg (Williams et al. 2018).

L'hypertension artérielle est considérée parmi les facteurs de risque vasculaires comme celui ayant le plus fort impact sur le risque de démence (Livingston et al. 2017). Les conclusions d'une autre revue de la littérature présentent les preuves d'une association forte entre l'hypertension artérielle à laquelle peuvent être exposés les adultes au cours de leur vie et l'impact sur les fonctions cognitives à un âge plus avancé (Iadecola et al. 2016).

C'est notamment le cas dans l'étude « Honolulu-Asia Aging Study » où une analyse des données de 3735 participants montrait que l'augmentation de 10mmHg de pression artérielle

systolique était associée à une augmentation de 9% (CI95%=[3%-16%]) du risque de présenter des performances cognitives plus faibles 23 à 28 ans plus tard (Launer et al. 1995).

Compte tenu de l'association qui existe entre l'hypertension artérielle et les fonctions cognitives, plusieurs travaux de recherche ont été menés afin d'explorer les associations entre les traitements antihypertenseurs et le risque de démence. A ce jour aucun essai clinique n'a été mené spécifiquement pour étudier l'efficacité des traitements antihypertenseurs mais l'analyse secondaire SPRINT MIND dans le cadre de l'essai SPRINT (« Systolic Blood Pressure Intervention Trial ») avait pour objectif d'étudier la cognition comme critère de jugement secondaire (Williamson et al. 2019). L'objectif principal de l'essai clinique était d'évaluer l'intérêt d'une stratégie de prise en charge intensive (objectif de pression artérielle inférieure à 120mmHg) comparativement à une prise en charge classique (objectif de pression artérielle inférieur à 140mmHg). La comparaison des groupes d'intervention montre un bénéfice de la prise en charge intensive avec une réduction du risque de survenue d'un déficit cognitif léger de 19% (RR = 0,81, CI95% = [0,69-0,95]) et une réduction du risque de survenue d'une démence de 17% bien que statistiquement non significative (RR = 0,83, CI95% = [0,67-1,04]).

Plusieurs autres études ont été combinées dans une méta-analyse de données observationnelles en population générale. Les auteurs se sont intéressés à l'utilisation des différentes classes thérapeutiques d'antihypertenseurs et leurs associations avec le risque de démence (Ding et al. 2019). L'analyse s'appuyait sur les données de 15553 participants avec une pression artérielle normale et 15537 participants avec une pression artérielle élevée. Ils ont montré une réduction du risque de survenue de démence de 12% (HR=0,88 ; CI95%=[0,79-0,98]) chez les personnes qui bénéficient d'un traitement antihypertenseur comparativement à ceux qui ne prennent pas de traitement dans le groupe des personnes avec une pression artérielle élevée. Dans le groupe avec une pression artérielle normale ils n'ont pas retrouvé de différences significatives sur le risque de démence associé à la prise d'un traitement antihypertenseur. Les auteurs concluent que l'absence d'association avec le risque de démence dans le groupe avec une pression artérielle normale et l'absence de différences entre les classes thérapeutiques pour l'association retrouvée dans le groupe avec une pression artérielle élevée ne sont pas en faveur d'un effet protecteur intrinsèque des médicaments hypertenseurs mais en faveur d'un effet délétère d'une pression artérielle non contrôlée.

2.3.3 Prise en charge et voies d'actions de l'hypertension artérielle

Plusieurs aspects de l'hypertension artérielle motivent l'étude de son rôle et de ses potentielles voies d'action dans l'histoire naturelle de la MAMA.

Prise en charge de l'hypertension artérielle

L'hypertension artérielle est responsable de 10 millions de décès dans le monde chaque année. Il s'agit de la pathologie chronique la plus fréquente, elle concerne 1 personne sur 3, et elle est un facteur de risque majeur de pathologies cardiovasculaires et neurovasculaires (Gakidou et al. 2017). En France on estime que la prévalence de l'hypertension artérielle est de 25,1% chez les femmes et de 36,5% chez les hommes. Comme la démence la prévalence de l'hypertension artérielle augmente fortement avec l'âge puisqu'elle est de 28,6% chez les individus de 45 à 54 ans, de 53,2% chez les individus âgés de 55 à 65 ans et de 67,8% chez les individus âgés de 65 à 74 ans (Perrine et al. 2018).

La prise en charge de l'hypertension artérielle doit s'appuyer prioritairement sur l'amélioration de l'hygiène de vie. Les recommandations hygiéno-diététiques consistent en une réduction de l'indice de masse corporelle en-dessous de 25, une diminution de la consommation d'alcool, de sel et de lipides saturés ainsi que la pratique d'une activité physique régulière. En cas d'échec, les recommandations hygiéno-diététiques sont renforcées par une prise en charge médicamenteuse. Il existe 5 classes thérapeutiques majeures qui peuvent être utilisées en monothérapie ou dans le cadre d'une association thérapeutique : les inhibiteurs calciques, les bêtabloquants, les diurétiques thiazidiques, les inhibiteurs de l'enzyme de conversion et les antagonistes des récepteurs de l'angiotensine II (HAS - Service des bonnes pratiques professionnelles 2016).

Cependant l'hypertension artérielle est une maladie dont la prise en charge est perfectible en France et plus généralement dans le monde (Perrine et al. 2018). Les causes principales sont l'absence de dépistage, de suivi médical et d'observance au traitement. Les résultats d'une méta-analyse sur la prise en charge de l'hypertension à l'échelle mondiale estiment que parmi les individus hypertendus, 53,5% ne connaissaient pas leur pathologie et 63,1% ne bénéficiaient pas d'une prise en charge médicamenteuse. De plus seulement 37,1% des hypertendus sous traitement avaient une pression artérielle effectivement contrôlée (Mills et al. 2016).

L'hypertension artérielle est un facteur de risque de démence. Elle concerne une grande majorité de la population vieillissante et est souvent mal prise en charge alors qu'il existe déjà des outils cliniques et thérapeutiques efficaces.

Voies d'actions de l'hypertension artérielle dans la démence

L'hypertension artérielle pourrait donc constituer un levier d'action pour réduire le nombre et la sévérité des démences. Cependant son implication dans les mécanismes physiopathologiques reste encore méconnue. Les voies d'actions pour expliquer l'association entre l'hypertension et la démence sont insuffisamment comprises et plusieurs résultats justifient l'utilisation de données multimodales pour étudier les interrelations de l'hypertension avec les différents biomarqueurs de la MAMA. Une piste possible est la maladie des petits vaisseaux du cerveau. Les HSB sont associés au déclin cognitif et au risque de survenue de démence. Or l'hypertension est très associée à la présence des HSB.

Dans l'étude EVA (« Epidemiology of Vascular Aging ») l'hypertension (PA systolique >160mmHg ou diastolique >95mmHg ou prise d'un traitement anti-hypertenseur) a été évaluée à l'inclusion et un examen IRM réalisé 4 ans plus tard pour 830 participants. Les participants étaient âgés en moyenne de 69 ans. La sévérité des HSB était classifiée en légère, modérée et sévère selon leur taille et nombre. Les participants avec une hypertension avait une probabilité 2,19 (IC95%=[1,43-3,36]) fois plus importante de présenter des HSB sévères (N = 138) comparativement à la présence de petits HSB (N = 318) (Dufouil et al. 2001).

La prise en compte de mesures répétées d'imagerie a permis d'établir un lien entre l'hypertension artérielle au cours de la vie et la vitesse de progression des HSB. Dans l'étude de Framingham une analyse sur un sous-échantillon de 1352 participants portait sur l'association entre l'hypertension évaluée à l'âge moyen de 54 ans (PAS > 140mmHg ou PAD > 90mmHg ou prise d'un traitement antihypertenseur) et la progression des HSB évaluée avec 2 IRM aux âges moyens de 61 ans et 67 ans, soit 7 et 13 ans après l'évaluation de l'hypertension. Les résultats montrent que les participants avec une hypertension à l'inclusion ont une vitesse de progression annuelle significativement plus importante ($\beta=0,23$ écart-type; IC95%=[0,17-0,29]; p-value<0,001) pour une mesure centrée-réduite des HSB comparativement aux participants sans hypertension (Debette et al. 2011). L'étude des 3 cités (« 3C Study ») a également permis de quantifier cette vitesse de progression en population générale. Dans un échantillon de 1319 participants ayant réalisé 2 IRM à 4 ans d'intervalle, la

progression du volume d'HSB était plus importante de $0,33\text{cm}^3$ (IC95%=[0,02-0,64]) chez les participants hypertendus à l'inclusion (PAS>140mmHg ou PAD>90mmHg ou prise d'un antihypertenseur) comparativement aux participants non hypertendus (ajusté sur l'âge, le sexe et le volume intracrânien), ce qui correspondait approximativement à une augmentation de 17% chez les non hypertendus contre 20% chez les hypertendus (Godin et al. 2011).

De plus il a été montré qu'une prise en charge renforcée de l'hypertension artérielle pourrait permettre de retarder ou stopper la progression des lésions. Un sous-échantillon de 192 participants de l'essai clinique PROGRESS âgés en moyenne de 61 ans ont bénéficié de 2 examens d'IRM à un intervalle moyen de 36 mois. Dans le groupe d'intervention (traitement anti-hypertenseur, N=89) la réduction des pressions artérielles systolique et diastolique étaient en moyenne respectivement de 12,5mmHg et 8,6mmHg ; dans le groupe contrôle (placebo, N=103) elles étaient respectivement de 1,3mmHg et 3,9mmHg. L'analyse des images montrait une réduction non statistiquement significative de 43% (IC95%=[7%-89%]) du risque d'apparition de nouveaux HSB dans le groupe d'intervention comparé au groupe contrôle. Plus spécifiquement le volume des nouveaux HSB était significativement plus faible ($6,5\text{mm}^3$) dans le groupe d'intervention par rapport au groupe contrôle ($7,6\text{mm}^3$) (Dufouil et al. 2005).

Un groupe de travail de l'« American Heart Society » conclut qu'il existe un nombre suffisant de preuves de l'impact délétère de l'hypertension au cours de la vie sur les capacités cognitives des années plus tard, à des âges plus avancés (Iadecola et al. 2016). En effet, l'hypertension perturbe la structure des vaisseaux sanguins du cerveau et peut causer des lésions ischémiques et hémorragiques dans la substance blanche qui est essentielle à la fonction cognitive. Les auteurs concluent que ces lésions pourraient donc favoriser la survenue d'un syndrome démentiel.

D'autre part, les auteurs d'une revue de la littérature suggèrent que la pression artérielle élevée est le plus important facteur de risque de sévérité et de progression des HSB (Alber et al. 2019).

2.3.4 Incidence de la démence

Un nombre croissant de données provenant d'Amérique du Nord et d'Europe suggère une tendance à la baisse de l'incidence de la démence. Deux échantillons d'analyses des données des cohortes françaises PAQUID (participants recrutés en 1988/1989, N=1469) et 3C (participants recrutés en 1999/2000, N=2104) ont été comparées pour étudier l'évolution de

l'incidence de la démence sur une période de 10 ans de suivi (Grasset et al. 2016). Les résultats montrent une diminution de l'incidence de la démence chez les femmes de 38% (HR=0,62 ; IC95%=[0,48-0,80]). Cette diminution n'a pas été observée chez les hommes.

L'évolution de l'incidence a aussi été étudiée en s'appuyant sur les données de l'étude de Framingham (« Framingham Heart Study ») aux Etats-Unis (Satizabal et al. 2016). Il s'agit d'une cohorte emblématique du fait de son ancienneté et ses contributions scientifiques. Elle a débuté en 1948 avec 5209 participants recrutés aléatoirement parmi les 10 000 résidents de la ville de Framingham (État du Massachusetts, États-Unis) de manière à représenter au mieux les disparités sociales (Mahmood et al. 2014). L'analyse de l'évolution de l'incidence porte sur la comparaison de 4 périodes de 5 ans de suivi pour les participants âgés de plus de 60 ans à chaque époque distincte : 1977-1983 (N=2457), 1986-1991 (N=2135), 1992-1998 (N=2333) et 2004-2008 (N=2090). La comparaison de l'incidence de la démence à 5 ans ajustée sur l'âge et le sexe de la période 1977-1983 avec les 3 périodes suivantes montre des diminutions respectives de l'incidence de 22% (HR=0,78 ; IC95%=[0,59-1,04]), 38% (HR=0,62 ; IC95%=[0,47-0,83]) et 44% (HR=0,56 ; IC95%=[0,41-0,77]) (Satizabal et al. 2016).

La diminution de l'incidence au cours du temps pourrait s'expliquer par une diminution de l'exposition à certains facteurs de risque de démence, notamment par l'augmentation du niveau d'études et l'amélioration de la prise en charge des maladies cardiovasculaires et de leurs facteurs de risque (Wu et al. 2017). Les études qui ont analysé la tendance à la baisse de l'incidence au cours du temps suggèrent un rôle significatif du niveau d'études qui a considérablement augmenté dans la population. Par exemple en France dans l'analyse comparative des cohortes PAQUID et 3C (Grasset et al. 2016), 10,1% des femmes avaient le baccalauréat dans les années 1990 contre 28,9% dans les années 2000. Les auteurs ont montré que niveau d'étude plus grand et un meilleur contrôle des facteurs de risque vasculaires expliquaient partiellement la diminution d'incidence observée chez les femmes. Dans l'analyse de l'évolution de l'incidence de la démence dans l'étude de Framingham 36% des participants avaient un niveau d'étude inférieur au baccalauréat (« high school diploma ») pour la période 1977-1983 contre seulement 5% pour la période 2004-2008. Les résultats supportaient que la diminution d'incidence globale observée entre les 2 périodes soit partiellement expliquée par l'augmentation du niveau d'éducation et une meilleure prise en charge des facteurs de risque vasculaires (Satizabal et al. 2016).

2.4 Temporalité des atteintes dans la MAMA

2.4.1 Modèle théorique de la séquence des atteintes

Séquence hypothétique des atteintes

La communauté scientifique s'accorde sur l'existence d'une séquence d'évènements qui définissent l'histoire naturelle de la MAMA. Le modèle théorique de la dynamique des biomarqueurs publié et discuté par (Jack et al. 2010) est de loin le modèle le plus cité dans la littérature. La figure 7 est une représentation de ce modèle, l'abscisse représente la progression de la maladie dans le temps et l'ordonnée la sévérité pathologique du marqueur considéré. Ce modèle est cohérent avec l'hypothèse de la cascade amyloïde (Selkoe and Hardy 2016) qui suppose que l'agrégation des protéines amyloïdes ($A\beta_{42}$) sous forme de plaques est à l'origine d'un mécanisme intracellulaire pathologique par phosphorylation des protéines tau. Les protéines tau phosphorylées (p-tau) s'agrègent pour former de manière incontrôlée des enchevêtrements neurofibrillaires intracellulaires provoquant la dysfonction neuronale et la mort de la cellule. La perte des neurones se traduit par l'observation des phénomènes d'atrophies des structures cérébrales (dégénérescence). En dernier survient le déclin cognitif progressif qui présente un tableau clinique multiple selon les régions fonctionnelles atteintes.

Il faut bien comprendre que ces évènements se chevauchent dans le temps. La sévérité des accumulations de protéines amyloïdes et de protéines tau continue de progresser en même temps que l'apparition de phénomènes d'atrophie du cerveau et les premiers symptômes cliniques (Jack et al. 2013).

Définition de la maladie d'Alzheimer centrée sur la protéine amyloïde

Les critères diagnostiques spécifiques de la démence due à la maladie d'Alzheimer ont progressivement évolué avec le développement de l'imagerie cérébrale et l'accumulation de preuves sur l'intérêt clinique des biomarqueurs de neuropathologie et d'atrophie du cerveau (McKhann et al. 1984; Dubois et al. 2007; McKhann et al. 2011). Ils complètent les critères diagnostiques de la démence toutes causes confondues (mentionnés dans la partie 2.1.2).

Le déploiement des technologies de neuroimagerie et d'analyse biologique sont le centre d'intérêt d'une redéfinition biologique de la maladie comme le propose la classification

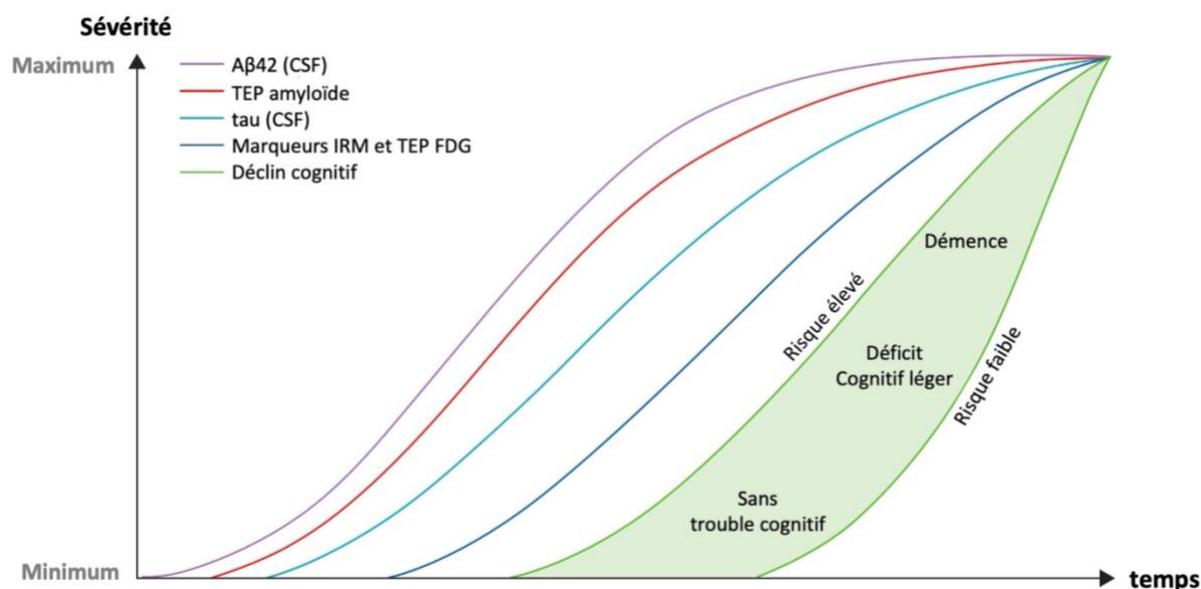


Figure 7. Modèle théorique de la dynamique des biomarqueurs de la maladie d'Alzheimer. Adapté de Jack et al. 2013

ATN (Amyloïde ; Tau phosphorylé ; Neurodégénérescence) : un ensemble de définitions et de concepts qui propose un cadre de travail pour la recherche (Jack et al. 2018). La classification recentre la maladie d'Alzheimer sur l'existence d'un *continuum* et une définition basée sur la présence de lésions associées aux protéines amyloïde et tau phosphorylée. Cette définition purement biologique suppose que la maladie peut être diagnostiquée à partir des biomarqueurs indépendamment des critères cliniques avec comme perspective de pouvoir distinguer la maladie d'Alzheimer des autres étiologies de la démence.

Le modèle théorique de la figure 7 et la classification ATN sont débattus dans la communauté scientifique et restent hypothétiques. La protéine amyloïde est définie comme la cause initiale de la maladie d'Alzheimer mais ce modèle ne propose pas de notions de durées ni de relations de causalité dans la séquence des atteintes.

De plus la contribution vasculaire à la démence et au déficit cognitif, pourtant largement documentée, n'est pas prise en compte (Sweeney et al. 2019). Nous avons vu que l'étiologie de la démence est majoritairement multiple et il est actuellement difficile d'imputer la survenue du trouble cognitif à la seule présence de la neuropathologie de type Alzheimer. Les connaissances scientifiques actuelles ne permettent pas d'affirmer l'absence de relation entre les lésions caractéristiques d'une neuropathologie de type Alzheimer et le rôle des anomalies vasculaires cérébrales. Le cadre de recherche sur la classification ATN est donc susceptible d'évoluer en intégrant d'autres marqueurs selon les avancées de la science (Silverberg et al. 2018).

Éléments empiriques

La construction du modèle théorique (figure 7) repose sur un ensemble d'observations sur les associations entre les biomarqueurs, obtenues à partir de données de participants à des stades cliniques différents. Des résultats sont cohérents avec l'hypothèse que l'amyloïdopathie précède la dégénérescence et le déclin cognitif. Une analyse concomitante des IRM et d'imageries moléculaires amyloïde (N=61) montraient l'absence de plaques amyloïdes et d'atrophie cérébrale chez des participants sans symptômes cognitifs, la présence de plaques amyloïdes sans atrophie cérébrale chez des participants avec un trouble cognitif léger et la présence de plaques amyloïdes et d'une atrophie cérébrale chez des participants avec un diagnostic de démence (Jack et al. 2009). La dégénérescence précéderait aussi le déclin cognitif, des études ont trouvé une valeur prédictive de la protéine tau totale (LCR) (Ewers et al. 2007), de l'atrophie cérébrale (Sluimer et al. 2010) et de l'imagerie moléculaire FDG (Jaagust et al. 2006) sur le déclin cognitif ou la conversion vers la démence.

Le modèle théorique propose une seule vision globale de la maladie qui s'appuie sur une revue d'un ensemble d'études indépendantes (Jack et al. 2010; 2013) mais elle ne s'appuie pas sur des données longitudinales de cohortes incluant les mesures répétées de l'ensemble des biomarqueurs étudiés depuis le premier stade de la maladie jusqu'au stade de démence sévère. Ces résultats peuvent donc être nuancés car ils posent la question de la durée et de la séquence temporelle des atteintes.

2.4.2 Temporalité de la MAMA

Cas particulier de la forme autosomique dominante

L'étude DIAN (« Dominantly Inherited Alzheimer Network ») (Bateman et al. 2012) a permis de contribuer à la compréhension de la dynamique des biomarqueurs en disposant de données chez des patients jeunes atteints d'une forme autosomique dominante de la maladie d'Alzheimer (forme familiale). Les patients développent une démence avec la présence d'une neuropathologie de type Alzheimer (amyloïde et tau) et surtout ils développent la maladie à un âge relativement identique d'une génération à l'autre. L'étude DIAN est donc une opportunité de comprendre l'histoire naturelle de la maladie.

Parmi 128 descendants de cas de maladie d'Alzheimer autosomique dominante, 88 étaient porteurs de la mutation génétique et 40 non-porteurs. Les données recueillies comprenaient des évaluations de la cognition, des imageries cérébrales (IRM, TEP amyloïde,

TEP FDG) et des analyses du LCR qui ont été comparées entre les porteurs et non-porteurs de la mutation génétique. Les chercheurs ont utilisé l'âge d'apparition des symptômes chez l'ascendant pour associer à chaque observation une estimation du temps avant l'apparition des symptômes. Ils ont ensuite estimé un modèle linéaire mixte en fonction du temps avant l'apparition des symptômes pour identifier des changements notables 20 ans avant pour l'amyloïde dans le LCR, 15 avant pour l'imagerie TEP amyloïde, la protéine tau dans le LCR et le volume de l'hippocampe, 10 avant pour l'imagerie TEP FDG et l'évaluation de la mémoire et 5 avant pour la CDR-SB.

Cependant il s'agit de données transversales qui ne permettent pas d'étudier des trajectoires individuelles de progression des biomarqueurs. De plus les caractéristiques phénotypiques et cliniques sont similaires mais la forme génétique pourrait ne pas représenter la réalité complexe et multifactorielle de l'histoire naturelle de la forme sporadique de la maladie d'Alzheimer.

Temporalité des stades cliniques

Les auteurs d'une méta-analyse incluant 3268 participants de 6 cohortes ont proposé une méthode d'analyse originale pour estimer la durée des différents stades cliniques à partir d'un modèle multi-états (Vermunt et al. 2019). Ils ont décomposé l'histoire naturelle de la maladie d'Alzheimer avec 4 phases successives : préclinique (présence d'une amyloïdopathie sans trouble cognitif), prodromale (présence d'une amyloïdopathie et trouble cognitif léger), démence légère (CDR < 2 ou CDR-SB ≤ 9) et démence modérée à sévère (CDR ≥ 2 ou CDR-SB > 9). Les résultats ont permis d'estimer une période totale de la maladie d'Alzheimer comprise entre 15 (15,0 ; IC95%=[11,0-16,9]) et 25 ans (24,1 ; IC95%=[21,8-25,4]) selon les

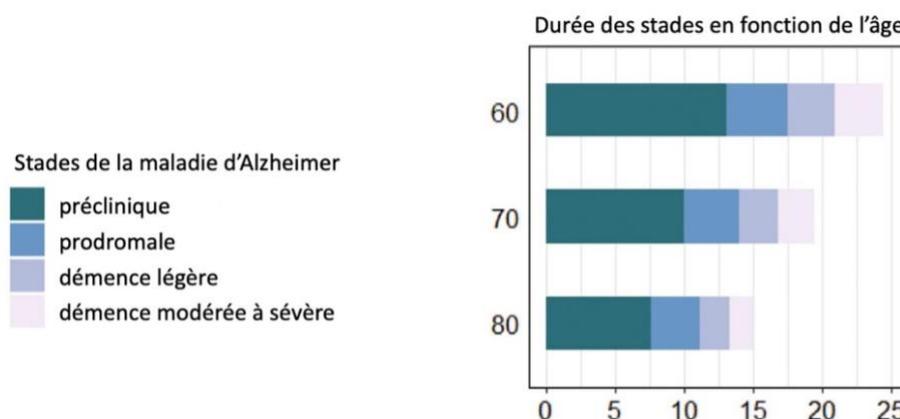


Figure 8. Estimation de la durée des stades de la maladie d'Alzheimer (N = 3268). Adapté de Vermunt et al. 2019

caractéristiques individuelles (figure 8). Par exemple pour une personne de 70 ans la phase

préclinique durerait 9,9 ans (IC95%=[8,4-11,5]), la phase prodromale durerait 4,0 ans (IC95%=[3,3-4,7]), la démence légère et la démence modérée à sévère durerait respectivement 2,9 ans (IC95%=[2,4-3,3]) et 2,6 ans (IC95%=[2,1-3,3]).

2.4.3 Séquence des atteintes : une approche évènementielle

Plusieurs travaux ont proposé des modèles dit évènementiels (« event-based model ») avec un temps discret de la maladie. L'intérêt de cette méthode est qu'elle peut être appliquée à des jeux de données transversales qui sont plus petits et plus largement disponibles comparativement aux données longitudinales des études de cohortes. L'estimation de la séquence d'évènements liés à la progression de la maladie repose sur l'hypothèse que la maladie est irréversible, les premiers travaux supposaient que la séquence est identique chez tous les individus (Fontein et al. 2011; Young et al. 2014), puis des extensions ont été proposées pour identifier des sous-types de la maladie (Young et al. 2018).

Par exemple supposons que l'on souhaite déterminer l'ordre dans lequel l'imagerie moléculaire de protéines amyloïdes, le volume de l'hippocampe et un test de mémoire sont affectés au cours de la progression de la maladie. Si d'une part les individus qui ont des plaques amyloïdes visibles en imagerie et un faible volume d'hippocampe n'ont pas nécessairement de trouble mnésique identifié au test de mémoire et d'autre part les individus avec un trouble mnésique identifié ont pour la plupart des plaques amyloïdes visibles en imagerie et un faible volume d'hippocampe, alors on peut estimer que l'atteinte de la mémoire est postérieure aux deux autres biomarqueurs. Ces modèles utilisent des mélanges gaussiens pour déterminer les probabilités des différents stades d'atteinte en distinguant pour chaque marqueur l'état normal et l'état anormal. Il est ensuite possible d'en déduire la séquence la plus probable des atteintes.

L'étude ADNI est une cohorte clinique multicentrique (principalement aux Etats-Unis et au Canada) de participants sélectionnés à différents stades cliniques (sans trouble cognitif, avec un trouble cognitif léger ou avec un diagnostic de démence). Le recrutement des participants a débuté en 2004 avec ADNI1 pour un suivi clinique initial de 5 ans et s'est poursuivi en plusieurs phases avec ADNIGO, ADNI2 puis ADNI3 débuté en 2016. Les principaux objectifs comprennent d'une part l'étude de l'histoire naturelle de la maladie et les relations entre la clinique, l'imagerie, la génétique et les biomarqueurs biologiques ; et d'autre part le

développement de l'usage des biomarqueurs comme critère de jugement dans le cadre d'essais cliniques.

La figure 9 représente les résultats obtenus par l'application d'un modèle

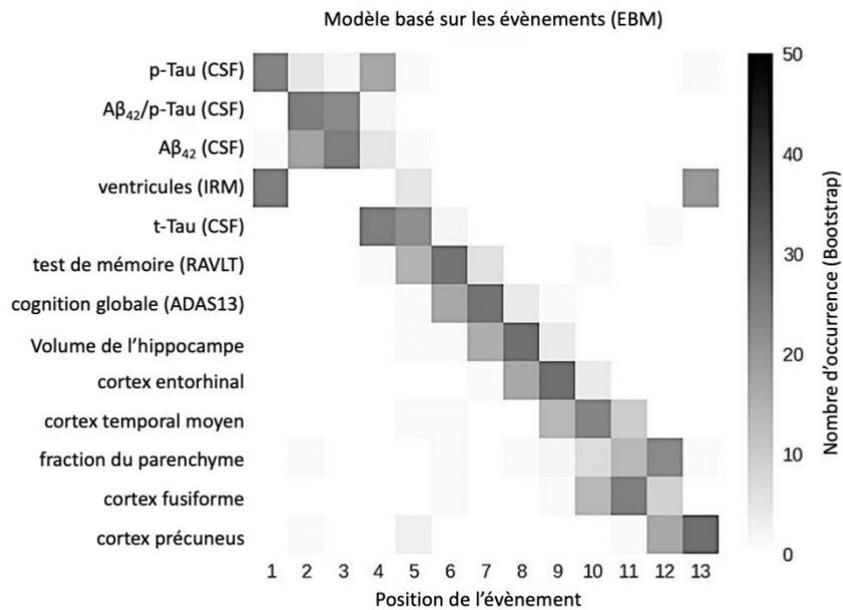


Figure 9. Séquence temporelle de biomarqueur dans l'étude ADNI. Exemple de résultats obtenus par un modèle basé sur les événements (EBM). Adapté de Archetti et al. 2019

évènementiel aux données de l'étude ADNI : les marqueurs de neuropathologie de type Alzheimer (p-tau, Aβ₄₂) deviennent anormaux en premier, cependant la cognition (RAVLT, ADAS13) semble atteinte avant la dégénérescence avec les différentes épaisseurs corticale (Archetti et al. 2019).

Cette approche permet de définir l'ordre de la séquence d'atteinte des biomarqueurs et l'incertitude associée mais elle est limitée à une approche transversale et ne permet pas d'exploiter les données longitudinales des cohortes pour étudier les notions de durée de progression dans le temps.

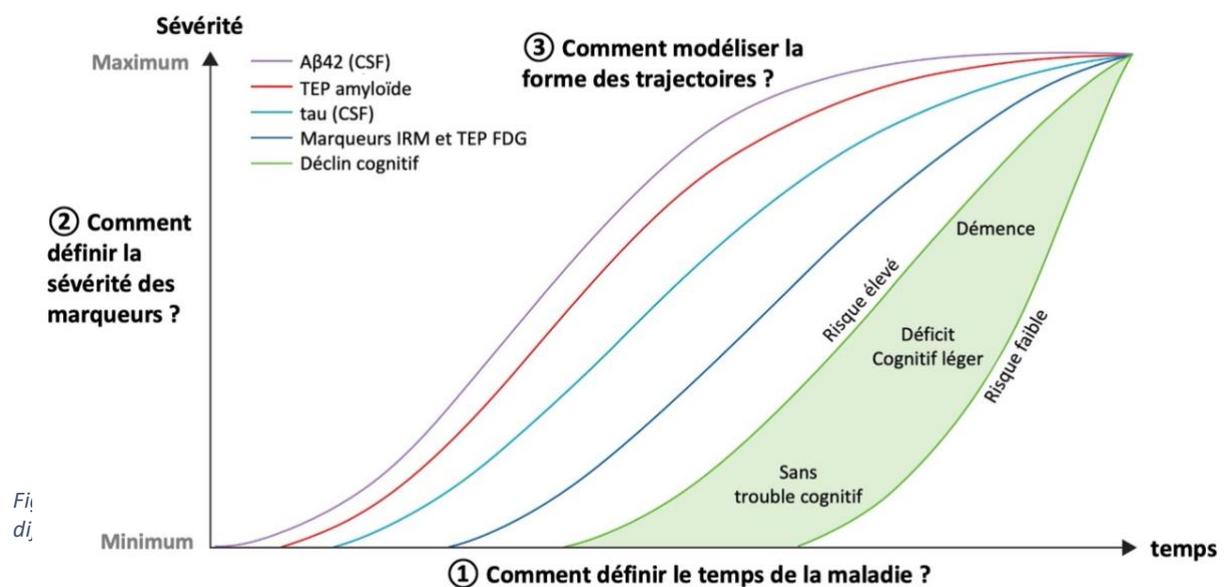
2.4.4 Séquence des atteintes : une approche par le temps de maladie latent

D'autres auteurs ont cherché à traduire directement le modèle théorique de Jack et al. (figure 7), plutôt que d'étudier la succession des événements. Une approche est d'utiliser des modèles avec un temps continu pour exploiter les données répétées des études de cohortes. Cependant cela induit plusieurs questions et choix statistiques résumés en figure 10 : (1) définir l'échelle du temps de la maladie, (2) définir l'échelle d'anormalité/sévérité des marqueurs, et (3) définir la forme des trajectoires.

La confrontation du modèle théorique aux données longitudinales de cohortes a donné lieu à de nombreux travaux en statistiques et en épidémiologie qui ont proposé des modèles longitudinaux d'évolution des marqueurs selon des modèles de progression dans lesquels le temps est le temps de maladie (Jedynak et al. 2012; Donohue et al. 2014; Bilgel et al. 2016; Guerrero et al. 2016; Li et al. 2017; Lorenzi et al. 2017; Marinescu et al. 2017; Schiratti et al. 2017; Bilgel and Jedynak 2019; Raket 2020; Garbarino and Lorenzi 2021; Koval et al. 2021; Kühnel et al. 2021). Ces travaux avaient différents objectifs concernant la nature et le nombre de marqueurs modélisés : atteintes des domaines cognitifs, progression spatio-temporelle des lésions de type Alzheimer ou de l'atrophie du cortex en imagerie, ou bien des approches multimodales qui modélisent à la fois les biomarqueurs de maladie d'Alzheimer, de dégénérescence et l'évaluation cognitive. Cependant l'idée des modèles de progression reste la même et ces travaux ont proposé des approches différentes pour traiter les 3 complexités statistiques mentionnées.

Définir le temps de maladie (1)

Nous avons vu que les biomarqueurs de neuropathologie de la maladie d'Alzheimer



pourraient commencer à changer entre 15 et 20 ans avant le début des symptômes. Les mesures répétées qui sont recueillies dans les études de cohortes ne concernent généralement que des observations à court terme (par exemple le temps de suivi moyen est inférieur à 5 ans dans les cohortes cliniques comme MEMENTO et ADNI), il est donc impossible d'observer les trajectoires entières des biomarqueurs pour tous les participants au cours de l'histoire naturelle de la maladie jusqu'au diagnostic de la démence. Le délai depuis l'entrée

dans l'étude ou le délai jusqu'au diagnostic ne permettent pas de décrire l'histoire naturelle de la maladie. On pourrait utiliser l'âge pour retrouver une dynamique globale des différents biomarqueurs (Luo et al. 2022) puisque l'âge est de loin le principal facteur de risque de la MAMA mais les données observées restent trop hétérogènes vis-à-vis de l'histoire naturelle de la maladie puisqu'elle ne débute pas au même âge chez tous les patients. De plus tous les participants ne développeront pas la démence, le temps de la maladie est donc inconnu.

Pour décrire les trajectoires entières il est donc nécessaire de travailler sur une autre échelle de temps que le délai jusqu'au diagnostic, le délai depuis l'entrée dans l'étude et l'âge. Une solution est la mise en œuvre d'une méthode spécifique pour combiner les informations intra-individuelles avec les informations inter-individuelles. On peut réaligner les trajectoires à court terme des observations individuelles et estimer la dynamique globale à long terme des différents marqueurs selon le temps de la maladie.

Le temps de maladie ne peut pas être observé directement, une façon de surmonter le problème est de l'inclure comme un paramètre individuel du modèle statistique en supposant que les différentes observations individuelles sont relatives à une trajectoire sous-jacente du marqueur qui se définit dans une échelle de temps latent de la maladie. Dans les travaux publiés, le temps latent peut être défini par un décalage temporel individuel, une combinaison d'un décalage temporel et d'une vitesse de progression individuels, ou une transformation exponentielle du temps observé. Dans un modèle de progression, les mesures peuvent alors être utilisées comme le reflet de la progression de la maladie.

Définir l'anormalité/sévérité des mesures des marqueurs (2)

Les valeurs recueillies pour chaque marqueur appartiennent à des échelles de mesures différentes et ne peuvent donc pas être comparées directement. Interpréter la dynamique de progression des différents biomarqueurs d'une façon similaire au modèle théorique nécessite une méthode pour définir la sévérité des marqueurs et les comparer. Une solution peut être d'utiliser les percentiles des mesures pour construire une échelle commune de valeurs normales à anormales avec par exemple les valeurs minimales et maximales qui ont été mesurées pour chaque biomarqueur. Une autre solution est d'estimer dans le modèle les valeurs minimales et maximales de la trajectoire de chaque marqueur avec par exemple une fonction logistique bornée.

Définir la forme de la trajectoire des biomarqueurs (3)

Dans le modèle théorique la forme sigmoïdale des trajectoires se compose d'une asymptote au départ, d'une accélération et d'une asymptote finale. Cette hypothèse s'appuie sur plusieurs analyses pour lesquelles une trajectoire non-linéaire sigmoïdale semblait la plus adaptée aux observations avec une phase d'accélération puis de décélération pour les biomarqueurs du LCR (Caroli and Frisoni 2010) et les biomarqueurs en imagerie (Sabuncu et al. 2011; Schuff et al. 2012) voire l'atteinte d'un plateau au moment de la progression des symptômes cliniques (Jack et al. 2012). L'essentiel est de choisir une méthode suffisamment flexible pour s'ajuster aux données, tout en étant bornée entre une valeur minimale et une valeur maximale ; une trajectoire sigmoïde semble donc adaptée. Dans la littérature plusieurs approches ont été proposées pour modéliser une trajectoire sigmoïdale avec par exemple une fonction logistique ou une fonction de distribution des percentiles.

Modèles de progression de la MAMA

Les travaux précédemment cités sur l'utilisation de modèles de progression de la maladie ont très majoritairement été appliqués aux données de l'étude ADNI. Les résultats ont notamment mis en évidence que les modèles de progression sont effectivement applicables à la maladie d'Alzheimer et peuvent contribuer à la compréhension de son histoire naturelle. Pour illustrer ces modèles de progression, 3 contributions sont détaillées ci-après.

Par exemple (Jedynak et al. 2012) ont défini un modèle statistique pour estimer le score de progression de la maladie de 687 participants de l'étude ADNI avec au moins 1 répétition de mesures pour au moins 1 biomarqueur. Ils ont estimé une séquence temporelle de 7 biomarqueurs et évaluations cognitives : $A\beta_{42}$ et t-tau dans le LCR, le volume de l'hippocampe, les scores aux tests RAVLT (mémoire), ADAS (cognition globale), MMSE (cognition globale) et la CDR-SB (figure 11). Les résultats étaient cohérents avec les stades d'évaluation clinique puisque les participants avec un diagnostic de démence avaient un score de progression (ADPS) plus élevé (+6,49 ans) que les participants avec un trouble cognitif léger (+2,85 ans) et les participants sans trouble cognitif (la référence). La vitesse de progression différait aussi, elle était 1,46 fois plus rapide chez les participants au stade de maladie d'Alzheimer par rapport aux participants sans trouble cognitif. Les résultats de la temporalité des marqueurs étaient surprenants puisque l'atteinte initiale était la mémoire, suivie du volume de l'hippocampe, de l' $A\beta_{42}$ puis t-tau dans le LCR.

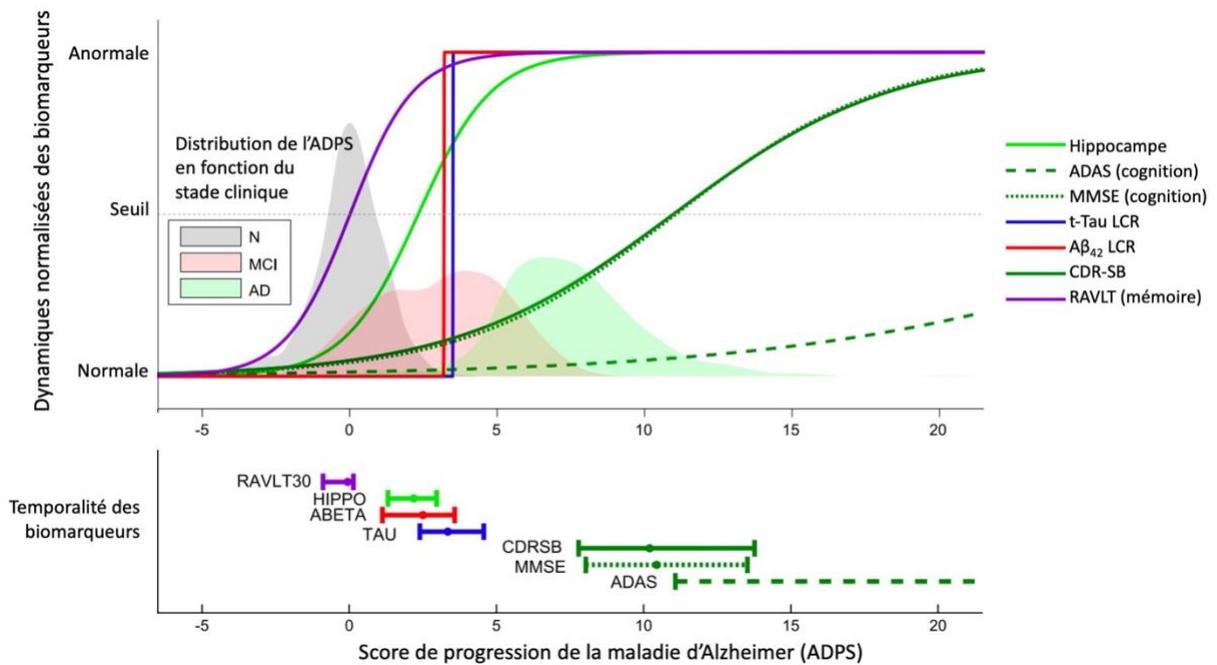


Figure 11. Progression et temporalité des biomarqueurs estimés avec un modèle à score de progression de la maladie d'Alzheimer. Adapté de Jedynak et al. 2012

(Li et al. 2018) ont décrit 16 biomarqueurs et évaluations cognitives par 1 modèle de progression qui supposait que tous les participants partagent la même dynamique de progression de la maladie. Le modèle appliqué aux données d'ADNI a mis en évidence un temps latent de maladie individuel (ajusté sur l'âge, le sexe, le génotype de l'*APOE* et le niveau d'études) qui illustre bien la variabilité inter-individuelle qui s'étendait sur plus de 40 ans (écart-type=10,7 ; IC95%=[9,6-11,6]). Les résultats étaient partiellement cohérents avec l'hypothèse de la cascade amyloïde. L'échantillon d'analyse incluait 1554 participants dont 26% sans troubles cognitifs, 6,8% avec une plainte mnésique subjective, 49,7% avec un trouble cognitif léger et 17,6% avec un diagnostic de maladie d'Alzheimer. Les prédictions obtenues pour une femme avec le génotype *APOE4* supposaient que la déposition amyloïde (imagerie TEP) surviendrait en premier lieu, suivie des marqueurs amyloïde et tau du LCR (figure 12). Plusieurs régions corticales seraient ensuite affectées ainsi que la mémoire et la cognition globale sur une période approximative de 10 ans avec en dernier les atteintes de l'hippocampe et l'atrophie globale du cerveau suivi de la CDR-SB. Les symptômes cliniques sévères avec une atteinte de l'autonomie dans les activités quotidiennes surviendraient une vingtaine d'années après la déposition des protéines amyloïdes. De façon inattendue la temporalité des protéines amyloïdes dans le LCR survenait plus de 15 ans après la déposition de protéine amyloïde visible en imagerie TEP.

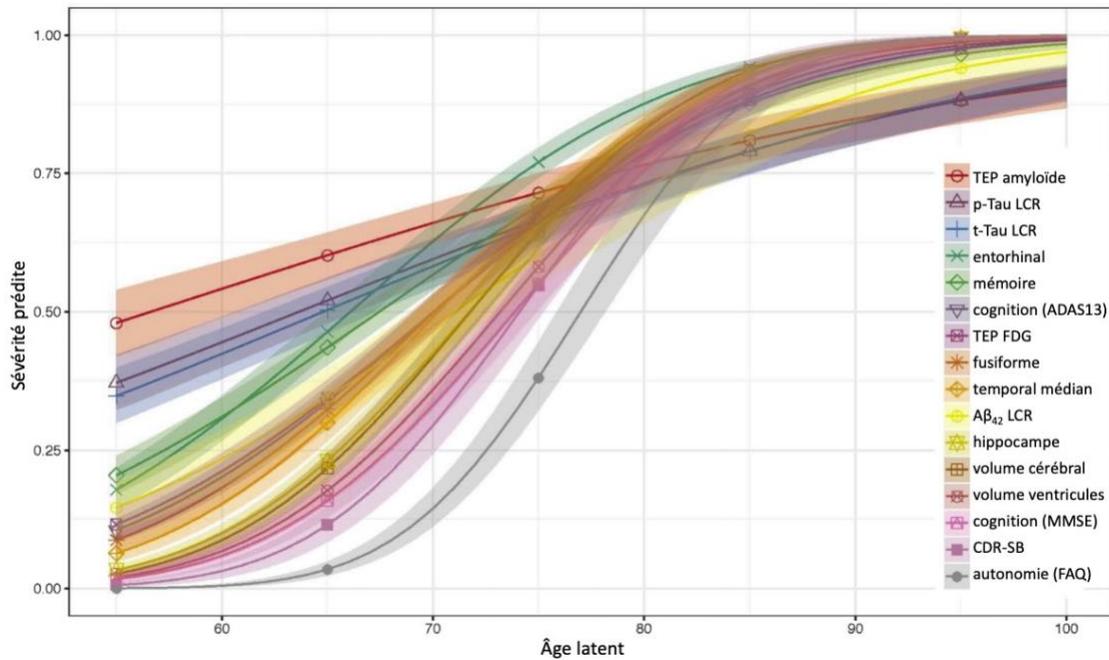


Figure 12. Prédications de la progression de la sévérité de 16 marqueurs (protéines amyloïde, protéines tau, dégénérescence et cognition) dans l'étude ADNI (N=1554) obtenues avec le modèle LTJMM. Adapté de Li et al. 2018

A nouveau sur les données de l'étude ADNI, Lorenzi et al. ont évalué la capacité prédictive d'un modèle de progression (Lorenzi et al. 2017). Ils ont sélectionné 782 participants avec une amyloïdopathie (mesurée par $A\beta_{42}$ dans le LCR) et ont entraîné le modèle de progression sur un échantillon aléatoire de 200 participants avant de le tester sur les 582 participants restants. Les résultats montraient une échelle de temps de la maladie qui s'étendait sur une vingtaine d'années et étaient cohérents avec l'hypothèse de la cascade amyloïde. Le premier marqueur atteint était l'imagerie TEP amyloïde suivi de l'imagerie TEP FDG puis du volume global du cerveau. La cognition (ADAS13) et l'autonomie (FAQ) étaient les derniers marqueurs atteints

A ma connaissance seulement 2 articles présentent les résultats d'un modèle de progression de la maladie d'Alzheimer appliqué sur les données d'une autre étude. (Bilgel et al. 2016) ont appliqué un modèle de progression sur les données d'imagerie TEP amyloïde de l'étude BLSA (« Baltimore Longitudinal Study of Aging » ; N=104). Ils ont montré que la progression du dépôt de protéines amyloïde dans plusieurs régions du cerveau n'était que partiellement cohérente avec les stades de déposition amyloïde de Braak (Braak and Braak 1991). Dans une deuxième application d'un modèle de progression à un ensemble d'évaluations cognitives mesurées dans les études BLSA (N=1661) et WRAP (« Wisconsin Registry for Alzheimer's Prevention » ; N=1510), (Bilgel et al. 2017) ont mis en évidence des

résultats cohérents entre les 2 études avec une atteinte de la fonction exécutive-attentionnelle, suivie de la mémoire, puis de la fonction exécutive-motrice et enfin du langage.

Dans la littérature, les modèles de progression font l'hypothèse que tous les participants suivent la même trajectoire de progression de la maladie et que le temps individuel de maladie n'est pas connu. Ainsi le temps latent de maladie permet de capturer les différences entre les individus mais celui est toujours traité comme une caractéristique totalement inconnue et défini uniquement par les mesures répétées des biomarqueurs.

Or dans les cohortes cliniques l'information du temps de diagnostic de démence est connue pour une partie des participants. Il est alors envisageable d'exploiter cette information partielle dans un modèle de progression de la maladie dans le but d'améliorer la pertinence clinique des trajectoires estimées.

2.5 Problématiques et objectifs du travail de thèse

Les aspects cliniques et physiopathologiques évoqués précédemment introduisent brièvement la complexité de la MAMA. L'atrophie des structures cérébrales est associée aux lésions de type Alzheimer et aux anomalies vasculaires cérébrales. Ces associations sont en faveur du rôle central de la dégénérescence dans l'évolution clinique des troubles neurocognitifs vers la démence et motive l'étude d'une part de l'histoire naturelle de la MAMA et d'autre part de l'implication de la maladie des petits vaisseaux cérébraux et du rôle de l'hypertension dans les mécanismes physiopathologiques de la MAMA. La mise en place de grandes cohortes cliniques et les progrès techniques ont permis d'accroître le nombre de marqueurs et la quantité de données à disposition de la communauté scientifique. C'est le cas de l'étude MEMENTO, une cohorte clinique multicentrique française, présentée dans la partie 3.1 du chapitre matériel et méthodes.

L'objectif général était de mieux comprendre l'implication de la composante vasculaire dans la MAMA en s'appuyant sur les données de la cohorte MEMENTO. Plus spécifiquement les objectifs étaient :

1. Comprendre les voies d'action et les interrelations entre l'hypertension artérielle, la maladie des petits vaisseaux, les lésions de type Alzheimer, l'atrophie du cerveau et la cognition. Pour cela, j'ai appliqué une analyse de médiation avec un modèle d'équation structurelle.
2. Décrire la séquence temporelle de la progression des biomarqueurs au cours de l'histoire naturelle de la maladie (neuropathologie de type Alzheimer, neuropathologie vasculaire cérébrale, atrophie du cerveau et déclin cognitif). Pour cela, j'ai développé un modèle de progression de la maladie avec recalage temporel pour prendre en compte l'hétérogénéité inter-individuelle des données longitudinales tout en conservant l'information clinique du temps de diagnostic.

3 MATERIEL ET METHODES

Mes travaux de thèse reposent principalement sur l'utilisation des données de l'étude MEMENTO et le développement de modèles mixtes et de modèles à équations structurelles pour les analyser. La cohorte permet de répondre aux questions soulevées grâce à un suivi clinique et biologique approfondi.

3.1 Applications aux données de la cohorte clinique MEMENTO

3.1.1 La cohorte

La cohorte MEMENTO (« deterMinants and Evolution of AlzheiMer's disEase aNd relaTed disOrders ») est une étude clinique française et multicentrique. L'objectif principal de MEMENTO est de mieux comprendre l'histoire naturelle de la MAMA, et plus spécifiquement d'étudier l'évolution clinique de personnes présentant des symptômes évocateurs, notamment en estimant la valeur pronostique et diagnostique de biomarqueurs innovants dans le passage à des états sévères de détérioration cognitive, en particulier de démence. Les biomarqueurs innovants concernent la neuropsychologie, la biologie, la génétique, et la neuroimagerie. Le recrutement des participants a été réalisé dans 26 centres de mémoires, de ressources et de recherche (CMRR) parmi les 28 centres existants sur le territoire métropolitain (figure 13). L'étude a reçu un avis favorable du comité de protection des personnes Sud-Ouest et Outre-Mer III ; elle suit les recommandations de la déclaration d'Helsinki et tous les participants ont signé un consentement éclairé.

3.1.2 Critères d'inclusions

Pour sélectionner une population de personnes permettant d'étudier l'évolution clinique en amont du diagnostic, les participants devaient (1) soit être âgés de 18 ans ou plus et présenter un déficit cognitif léger, (2) soit être âgés de 60 ans ou plus et exprimer une plainte cognitive subjective. Le déficit cognitif léger a été établi à partir des scores obtenus à une batterie de tests neuropsychologiques qui devait être réalisée dans les 6 mois précédant l'inclusion. Si un score pour au moins un des domaines cognitifs évalués était inférieur d'au

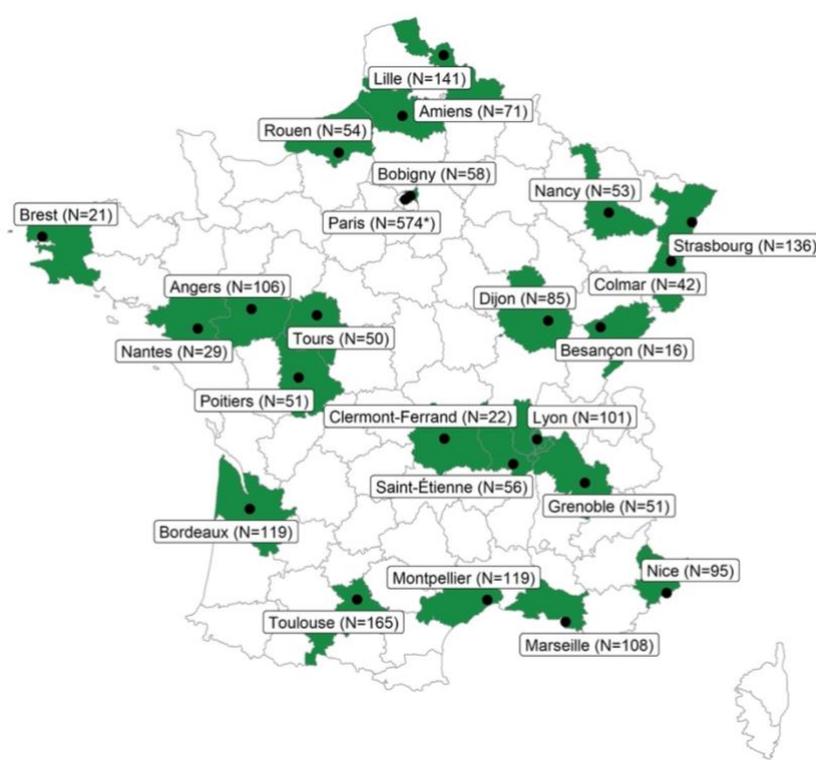


Figure 13. Localisations des 26 centres de mémoire, de ressources et de recherche qui participent à l'étude MEMENTO et nombres de participants recrutés par centre. Cohorte MEMENTO (2011-2014).

moins un écart-type par rapport à la norme moyenne (tenant compte de l'âge, du sexe et du niveau d'étude), le participant était considéré comme présentant un déficit cognitif léger. Les autres critères d'inclusion vérifiaient (1) l'absence de démence ($CDR < 1$), (2) des capacités visuelles et auditives suffisantes pour la passation des tests neuropsychologiques, (3) une couverture sociale. Les personnes exclues de l'étude comprenaient (1) les refus et contre-indications à l'IRM, (2) les tutelles et résidents en institution, (3) les femmes enceintes ou allaitantes, (4) les maladies neurologiques et maladies d'Alzheimer précoce autosomique dominante, et (5) les antécédents de chirurgie intracrânienne, de déficit neurologique secondaire à un trauma crânien, de schizophrénie et d'accident vasculaire cérébral survenu dans les 3 mois.

3.1.3 Calendrier du recueil des données

Le recrutement des participants a débuté en avril 2011 et s'est terminé en juin 2014 pour constituer une cohorte de 2323 participants. Le protocole de l'étude comprenait un suivi intensif des participants pendant une durée initiale de 5 ans (60 mois). Les participants ont été vus tous les 6 mois pour réaliser un ensemble d'examen cliniques et paracliniques, ce qui représente un total de 11 visites :

Tableau 1. Calendrier prévisionnel du recueil de données dans l'étude MEMENTO. Adapté de Dufouil et al. 2017.

Données recueillies	Inclusion et visites planifiées dans le suivi de l'étude (60 mois)										
	M0	M06	M12	M18	M24	M30	M36	M42	M48	M54	M60
Suivi clinique	■	□	■	□	■	□	■	□	■	□	■
Tests neuropsychologiques	■	▲	■	▲	■	▲	■	▲	■	▲	■
IRM	■				■				■		
TEP FDG	●				●				●		
Ponction lombaire	●				●				●		
TEP amyloïde	<i>Dans le cadre de l'étude MEMENTO-Amyging (A l'inclusion et 24 mois après)</i>										

■ : suivi clinique sur site □ : suivi clinique possiblement simplifié par téléphone ; ● : examen optionnel ; ▲ : examen réalisé si CDR ≥ 1 à la visite précédente ; IRM : imagerie par résonance magnétique ; TEP : tomographie par émission de positron ; FDG : 18F-fluorodésoxyglucose ; CDR : évaluation clinique de la démence

- Tous les 6 mois, un suivi clinique était réalisé incluant notamment le recueil des données socio-démographiques, des antécédents médicaux, des traitements médicamenteux, un ensemble de mesures physiologiques (y compris de la pression artérielle) et l'évaluation de la CDR. Le recueil de données concernant les antécédents médicaux et les traitements médicamenteux étaient complétés à chaque visite par les centres investigateurs à partir des informations disponibles ou des déclarations des participants. Il était conseillé de réaliser les examens de suivi intermédiaires (6 mois, 18 mois, 30 mois, 42 mois et 54 mois) sur site au CMRR, cependant l'investigateur clinique pouvait décider de réaliser un suivi simplifié par téléphone. Dans ce cas les examens cliniques (mesures physiologiques et CDR) n'étaient pas réalisés.
- Tous les ans, une batterie de tests neuropsychologiques était réalisée sur site ou tous les 6 mois si le participant avait une CDR ≥ 1 à la visite précédente.
- Tous les 2 ans, les examens comprenaient une imagerie cérébrale IRM, une imagerie moléculaire FDG ainsi qu'une biopsie de LCR par ponction lombaire. L'imagerie moléculaire et la ponction lombaire était des examens optionnels.
- Un examen d'imagerie moléculaire pour visualiser les accumulations de protéines amyloïdes (avec les traceurs 18F-flutémétamol ou 18F-florbétapir) était prévu dans le cadre de l'étude MEMENTO-Amyging. Le but était notamment d'estimer la relation entre la mesure de la charge amyloïde présente dans le cerveau et l'incidence de la démence. Une première mesure était réalisée à l'inclusion dans

l'étude ancillaire puis une deuxième était réalisée 2 ans après pour évaluer son évolution.

3.1.4 Facteurs de risque, marqueurs de MAMA et évaluations cognitives

Les travaux de la thèse s'intéressent aux données socio-démographiques qui incluent essentiellement l'âge, le sexe, le niveau d'études et les antécédents médicaux et médicamenteux. Les travaux ont aussi étudié le génotype des allèles du gène *APOE* qui était déterminé sur échantillon sanguin par KBiosciences (Hoddesdon, UK ; www.kbioscience.co.uk), ainsi qu'une grande partie des données multimodales recueillies dans le cadre de l'étude et mentionnées ci-dessous.

Mesure de la pression artérielle

Lorsque la mesure de la pression artérielle est réalisée dans un environnement médical il existe un effet « blouse blanche » qui se traduit par une augmentation transitoire de la pression artérielle. La prévalence serait comprise entre 15% et 25% chez les sujets âgés (Aronow et al. 2011). Afin de limiter le biais associé à cet effet, la mesure de la pression artérielle était réalisée 3 fois et après 2 minutes de repos. Les différentes mesures étaient reportées pour les pressions artérielles systoliques et diastoliques et les moyennes respectives ont été utilisées dans le cadre des travaux de la thèse.

Analyse du liquide céphalo-rachidien

Les concentrations en protéines tau totales (t-tau), tau phosphorylées (p-tau181) et amyloïdes ($A\beta_{42}$ et $A\beta_{40}$) dans le LCR ont été analysées par INNOTEST, une méthode immuno-enzymatique de type sandwich (ELISA) commercialisée par la société Fujirebio. La ponction lombaire a été acceptée par 17% des participants.

Marqueurs d'imageries cérébrales

Les examens IRM ont été réalisés avec des machines à aimants 3,0T permettant une acquisition haute résolution pour 86% des participants, les autres ayant bénéficié de machine 1,5T. Les examens d'imagerie moléculaire ont aussi été réalisés avec différentes machines en fonction des centres investigateurs. Cependant un ensemble de méthodes visant à harmoniser ces différences de machines a été prévu dans le protocole de l'étude (Dufouil et al. 2017). Toutes les acquisitions d'images cérébrales ont été traitées et analysées par le CATI

(Centre d'Acquisition et de Traitement d'Images pour la maladie d'Alzheimer, <http://www.cati-neuroimaging.com>) à l'aide de logiciels spécialisés :

- SACHA (Chupin et al. 2009) est un algorithme de segmentation automatique de l'image IRM-T1 pour fournir une estimation précise des volumes des lobes droit et gauche de l'hippocampe. Un contrôle qualité visuel a également été réalisé.
- FreeSurfer (Fischl et al. 2004) version 5.3 se base sur l'atlas de Desikan-Killiany (Desikan et al. 2006) pour segmenter les IRM-T1 et estimer les volumes, surfaces et épaisseurs des différentes régions du cortex.
- SPM8 (Ashburner and Friston 2005) permet de différencier à partir des IRM-T1 les tissus du cerveau (substance blanche, substance grise, LCR) et estimer par exemple l'atrophie globale du cerveau par la fraction du parenchyme.
- WHASA (Samaille et al. 2012) est spécialisé dans l'analyse des IRM-T2 (FLAIR) et permet notamment d'identifier le nombre et le volume des HSB.

Les intensités des signaux mesurés lors des examens d'imagerie moléculaire TEP FDG et TEP amyloïde sont analysées sous la forme d'un ratio SUVr (« standard uptake value ratio ») qui est notamment normalisé par rapport au pons (une structure cérébrale choisie en référence) à l'aide de l'IRM-T1 et de SPM8 (Buchert et al. 2005; Habert et al. 2016). Pour l'imagerie TEP amyloïde, 438 participants (19%) ont accepté l'examen : 49% des acquisitions ont été réalisées avec le radioligand flutemetamol et 51% avec le radioligand florbetapir. Pour l'imagerie TEP FDG, 1396 participants (60%) ont accepté l'examen.

Dans le cadre des travaux de la thèse toutes les données d'imagerie analysées dans les modèles statistiques sont des données numériques fournies par le CATI.

Évaluations cognitives

L'étude MEMENTO inclut une large évaluation cognitive avec une dizaine de tests neuropsychologiques qui s'ajoute à un ensemble de questionnaires (Dufouil et al. 2017). Une partie seulement a été analysée dans le cadre de la thèse. Ces données incluent l'évaluation clinique de la démence (CDR) et de la CDR-SB pour la cognition globale ainsi qu'une sélection de tests neuropsychologiques pour l'évaluation des différents domaines cognitifs :

(1) RLRI-16 : le test de rappels libres et indicés à 16 items est une adaptation française du test de Grober & Buschke (Grober et al. 1988). Il permet d'évaluer les capacités d'encodage et de récupération d'informations qui affectent la mémoire. La première phase d'encodage

sans limite de temps consistait à apprendre 4 listes de 4 mots associés à une catégorie sémantique (par exemple « poisson – hareng ») pour un total de 16 mots. Après une tâche distractive de 20 secondes (compter à l'envers par exemple), le participant avait 2 minutes pour se rappeler des 16 mots sans tenir compte de l'ordre, ce sont les rappels libres. Puis il avait 10 secondes avec un indice pour les mots qu'il n'avait pas réussi à retrouver (par exemple « quel est le poisson ? ») avant que la réponse ne lui soit donnée, ce sont les rappels indicés. La procédure était répétée 2 fois, les scores totaux des rappels libres et des rappels libres/indicés étant sur 48 points.

(2) Les tests de fluence verbale (Thurstone 1987) qui permettent d'évaluer le langage consistaient à énumérer un maximum de mots en 120 secondes. On différencie la fluence sémantique pour laquelle on demandait au participant de citer un maximum de noms d'animaux, et la fluence littérale pour laquelle on demandait au participant de citer un maximum de mots commençant par la lettre « p ».

(3) Le « Trail Making Test » (Tombaugh 2004) évalue les fonctions exécutives et l'attention. Dans sa version A (TMT-A), le participant disposait de 3 minutes pour relier dans l'ordre une série de 25 points numérotés de 1 à 25. Dans sa version B (TMT-B), qui fait aussi appel à la flexibilité mentale, le participant disposait de 5 minutes pour relier dans l'ordre une série de 25 points numérotés de 1 à 13 ou notés de A à L en alternant chiffres et lettres. La performance aux tests est généralement évaluée par le temps écoulé ou le temps moyen par bon déplacement.

(4) Le test de la figure complexe de Rey-Osterrieth fait intervenir la perception visuo-spatiale et la mémoire visuelle. On demandait dans un premier temps au participant de recopier l'image qui comporte 18 éléments sans limite de temps. Selon l'absence ou la présence et le bon placement de chaque élément, le score attribué était respectivement de 0, 1 ou 2 points pour un score maximal de 36 points. Toujours sans limite de temps, on demandait 3 minutes plus tard de reproduire la figure en faisant intervenir la mémoire du participant pour attribuer aussi un score total sur 36 points (Meyers and Meyers 1995).

3.1.5 Description de la population

Caractéristiques des participants à l'inclusion

Les 2323 participants étaient âgés en moyenne de 70,9 ans et 62% étaient des femmes. Le niveau d'études était supérieur au baccalauréat (environ 12 ans d'études) pour 39% des

participants, 30% étaient porteurs d'au moins 1 allèle $\epsilon 4$ du gène *APOE* et 65% étaient considérés hypertendus (PAS \geq 140mmHg ou PAD \geq 90mmHg ou prise d'un traitement antihypertenseur). Le tableau 2 présente les caractéristiques de la population de l'étude MEMENTO à l'inclusion dans la cohorte. Les variables continues sont résumées par la médiane et l'intervalle interquartile, les variables catégorielles par l'effectif et le pourcentage correspondant.

Nombre de mesures répétées au cours du suivi

Les données analysées sont celles qui étaient disponibles pendant les travaux de recherche de la thèse, excluant en particulier les analyses d'images à 4 ans de suivi (M48). Le suivi moyen des participants était de 4,1 ans (écart type = 1,8 ans) et le suivi médian de 5 ans (1^{er} quartile = 3,5ans ; 3^{ème} quartile = 5,1 ans). Parmi les 2323 participants inclus 1481 ont complété le suivi initialement prévu de 5 ans, soit une attrition globale de 36%. Les données recueillies comptaient 3791 IRM (N = 2199 participants), 2402 images moléculaires TEP FDG (N = 1524 participants) ainsi que 536 ponctions lombaires (N = 413 participants). La figure 14 décrit le nombre de mesures répétées disponibles par type d'examen et la figure 16 présente la temporalité de l'étude avec les délais entre les mesures et la date d'inclusion selon le type d'examen.

Diagnostics de démence et décès

Dans le cadre de l'étude MEMENTO, les cas incidents de démence font l'objet d'un diagnostic clinique basé sur les critères DSM-IV (American Psychiatric Association 2000) et NINCDS-ADRDA (Dubois et al. 2007). La sévérité de la démence est évaluée avec la CDR (\geq à 1). Tous les cas incidents sont analysés et validés par un comité d'experts indépendants, composé de neurologues et de gériatres, en charge de revoir le dossier clinique du participant pour confirmer le diagnostic et déterminer l'étiologie de la démence. Au cours du suivi, il y a eu 301 diagnostics cliniques de démence et 71 participants sont décédés sans diagnostic de démence clinique. Le biais d'analyse lié au risque compétitif du décès est a priori relativement faible compte tenu du faible nombre de décès sur le nombre total de participants. La figure 15 représente les incidences cumulées de cas de démence clinique et de décès avant la démence.

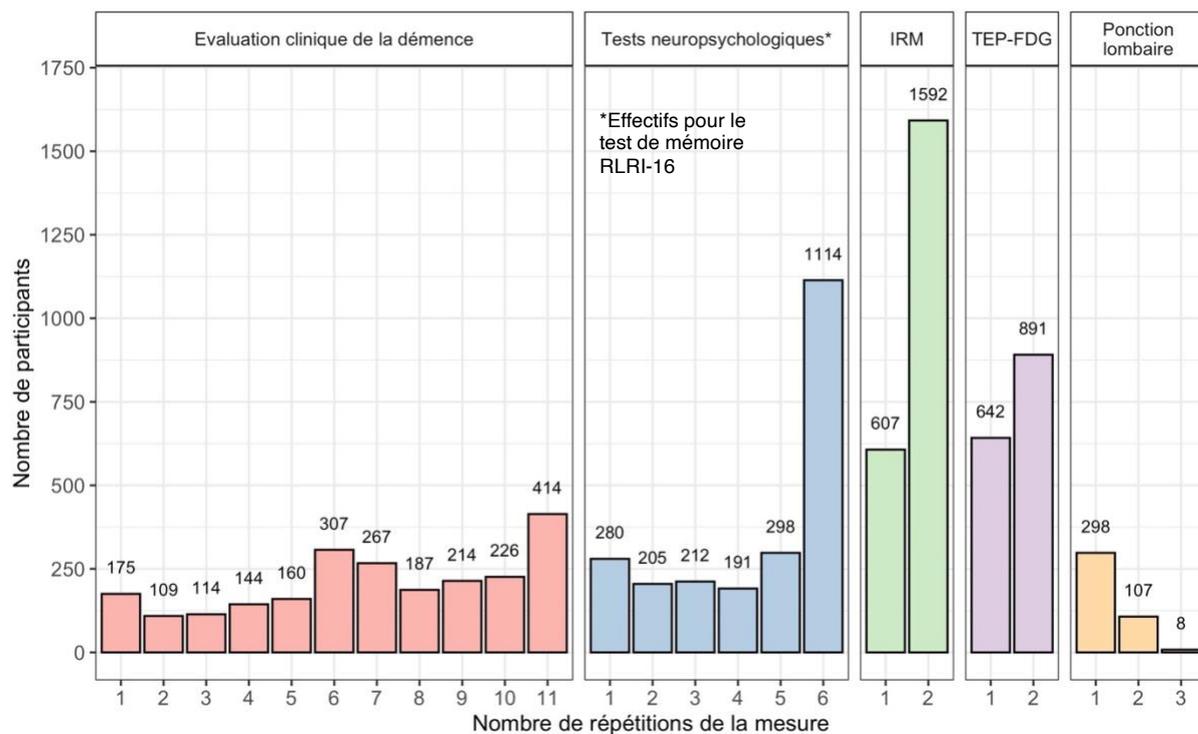


Figure 14. Nombre de participants et de mesures répétées pour les différentes évaluations et examens cliniques. Cohorte MEMENTO 2011-2021

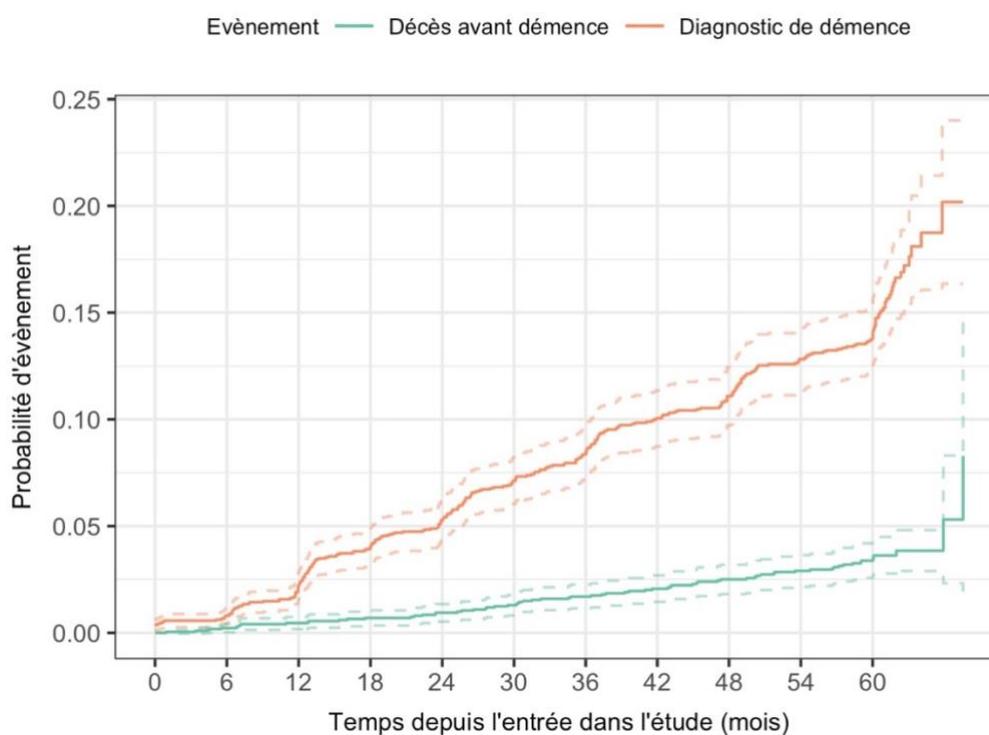


Figure 15. Incidences cumulées de démence et de décès sans démence par l'estimateur non paramétrique de Aalen Johansen. Cohorte MEMENTO 2011-2021.

Tableau 2. Caractéristiques des participants de la cohorte MEMENTO à l'inclusion dans l'étude. Cohorte MEMENTO 2011-2014.

Variable	N	N = 2,323 ¹
<i>Caractéristiques socio-démographiques</i>		
Âges (années)	2323	72 (65 ; 77)
Femmes	2323	1438 (62%)
Niveau d'études supérieur au baccalauréat	2318	908 (39%)
Trouble cognitif léger / MCI (CDR = 0.5)	2308	1375 (60%)
<i>Caractéristique génétique</i>		
Nombre d'allèle ε4 du gène de l'APOE	2196	
0		1538 (70%)
1		584 (27%)
2		74 (3,4%)
<i>Caractéristiques cardio-vasculaires</i>		
Prise en charge de l'hypertension artérielle	2227	
Non traité et <140/90mmHg		772 (35%)
Non traité et ≥140/90mmHg		452 (20%)
Traité et <140/90mmHg		530 (24%)
Traité et ≥140/90mmHg		473 (21%)
Antécédents cardio-vasculaires	2323	313 (13%)
<i>Ponction lombaire de liquide céphalo-rachidien</i>		
Protéine amyloïde Aβ ₄₂ (pg/ml)	356	1088 (723 ; 1,400)
Protéine amyloïde Aβ ₄₀ (pg/ml)	348	13788 (10922 ; 16588)
Protéine tau phosphorylée (pg/ml)	356	55 (44 ; 73)
Protéine tau totale (pg/ml)	355	294 (212 ; 439)
<i>Imagerie moléculaire des protéines amyloïdes</i>		
Radiotracer utilisé	438	
Florbétapir		224 (68%)
Flutémétamol		214 (32%)
TEP amyloïde (SUVr moyen)		0,70 (0,63 ; 0,94)
<i>Imagerie par résonance magnétique (IRM)</i>		
Fraction du parenchyme (%)	2183	71,9 (67,6 ; 75,5)
Volume de l'hippocampe (cm ³)	2130	5,50 (4,96 ; 5,96)
Épaisseur moyenne du cortex (mm)	2151	2,34 (2,27 ; 2,40)
Volume des hypersignaux de la substance blanche (cm ³)	2104	5,63 (2,66 ; 11,49)
<i>Imagerie moléculaire du métabolisme des glucides</i>		
TEP FDG (SUVr moyen)	1337	1,79 (1,68 ; 1,92)
<i>Évaluations cognitives</i>		
CDR-SB (score /18)	2308	0,5 (0,0 ; 1,0)
RLRI-16- Somme des rappels libres (score /48)	2302	27 (21 ; 32)
Reproduction de la figure de Rey à 3 minutes (score /36)	2155	16 (10 ; 20)
Trail Making Test A (secondes/bon déplacement)	2298	1,88 (1,46 ; 2,42)
Fluence sémantique - animaux (Nombre de noms)	2275	28 (22 ; 34)
Fluence littérale - lettre P (Nombre de mots)	2284	20 (15 ; 25)

¹Médiane (1^{er} quartile ; 3^{ème} quartile) ou n (%)

BAC = baccalauréat ; MCI = déficit cognitif léger ; CDR = évaluation clinique de la démence ; APOE = gène de l'apolipoprotéine E ; TEP = tomographie par émission de positrons ; SUVr = ratio de valeurs standardisées du signal ; FDG = fluorodésoxyglucose ; CDR-SB = sommes des domaines de la CDR ; RLRI-16 = rappels libres et indicés à 16 items

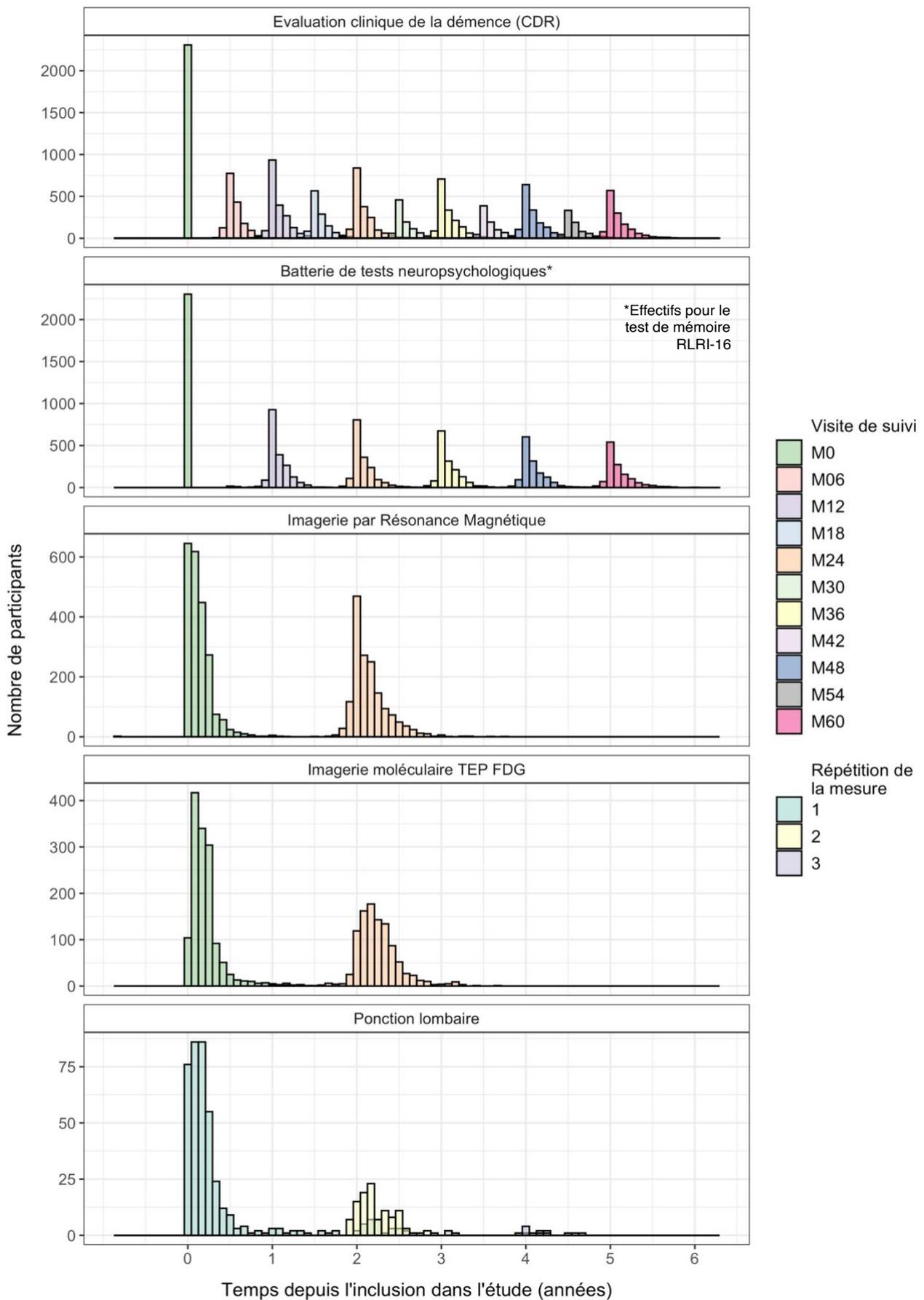


Figure 16. Histogramme des temps de mesure depuis l'inclusion selon le type d'examen. Cohorte MEMENTO (2011-2021).

3.2 Études des chemins d'association par les modèles à équations structurelles

Les modèles à équations structurelles (SEM) ont d'abord été développés pour l'analyse de données en sciences sociales et psychologiques (Cronbach and Meehl 1955). C'est un domaine dans lequel il est fréquent que la recherche soit centrée sur des concepts théoriques. Les SEM ont la particularité de pouvoir étudier ces concepts en les définissant comme des variables ou quantités d'intérêt qui ne sont pas mesurables directement (Jöreskog and Sörbom 1979). Les quantités conceptuelles sont appelées dimensions ou variables latentes. Par exemple dans l'étude des démences et du vieillissement cognitif, on peut définir le concept de cognition plutôt qu'utiliser directement les observations d'un score à un test neuropsychologique.

3.2.1 Représentation graphique d'un SEM

Un système d'équations structurelles permet d'étudier les associations entre plusieurs quantités conceptuelles (variables latentes). Il se définit à l'aide d'un graphe très similaire au graphique acyclique orienté (DAG ou « Directed Acyclic Graph ») utilisé en épidémiologie pour étudier les chemins causaux (Greenland, Pearl, and Robins 1999).

La nomenclature graphique dans un SEM permet de différencier les variables observées (carrés ou rectangles) et les variables latentes (ronds ou ovales) (Mcardle 1980). Les régressions sont représentées par des flèches unidirectionnelles. Cette relation implique une hypothèse de causalité pour expliquer l'association entre les 2 variables. Les flèches bidirectionnelles représentent la covariance entre 2 variables (une association sans hypothèse de causalité). On peut également utiliser une flèche bidirectionnelle dont les 2 extrémités pointent la même variable pour représenter la variance de cette variable. Les ronds ou ovales peuvent aussi être utilisés pour représenter les erreurs résiduelles qui sont des variables latentes particulières (McDonald et Ho 2002). La suivante est un exemple simple de graph que l'on peut étudier avec un SEM.

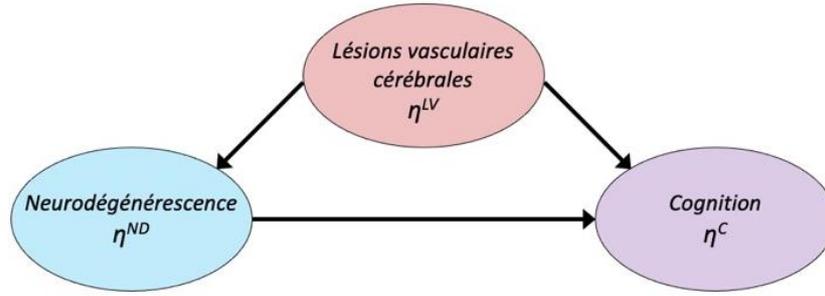


Figure 17. Représentation graphique d'un système d'équation structurelle.

Les lésions vasculaires cérébrales (LV), la neurodégénérescence (ND) et la cognition (C), sont 3 concepts définis par 3 variables latentes (η^{LV} , η^{ND} et η^C). On peut les structurer les unes par rapport aux autres avec des hypothèses sur le sens des relations.

3.2.2 Système d'équations structurelles

A la différence du DAG qui ne représente que des relations ; les SEM adossent une régression linéaire à chaque flèche du graphe (dans chacune des équations structurelles) pour spécifier la forme de la relation et permet donc d'estimer les associations entre variables (Spirtes et al. 1998). La spécification d'un SEM permet de confronter les données aux hypothèses sur les relations entre variables. Si on reprend l'exemple de la figure 17, le graphique se traduit par 2 équations structurelles :

$$\eta_i^C = \beta_0^C + \beta_{1LV}^C \times \eta_i^{LV} + \beta_{1ND}^C \times \eta_i^{ND} + \varepsilon^C \quad (1.1)$$

$$\eta_i^{ND} = \beta_0^{ND} + \beta_{1LV}^{ND} \times \eta_i^{LV} + \varepsilon^{ND}$$

$$\text{avec } \varepsilon^C \sim \mathcal{N}(0, \sigma_C^2), \varepsilon^{ND} \sim \mathcal{N}(0, \sigma_{ND}^2) \text{ et } \eta^{LV} \sim \mathcal{N}(\beta_0^{LV}, \sigma_{LV}^2)$$

où β_0^C , β_0^{ND} et β_0^{LV} sont les intercepts des 3 variables latentes, β_{1LV}^C , β_{1LV}^{ND} et β_{1ND}^C sont les coefficients de régressions entre les variables latentes et ε^C , ε^{ND} sont les erreurs de mesures.

On peut ensuite complexifier le SEM en ajoutant d'autres variables pour expliquer les différents concepts. Par exemple dans le cas des travaux de la thèse, notre objectif est de comprendre la relation entre l'hypertension artérielle et les différents concepts associés à la MAMA. Nous pouvons définir un SEM dans lequel l'hypertension artérielle (HTA) joue un rôle sur les lésions vasculaires cérébrales, la neurodégénérescence et la cognition.

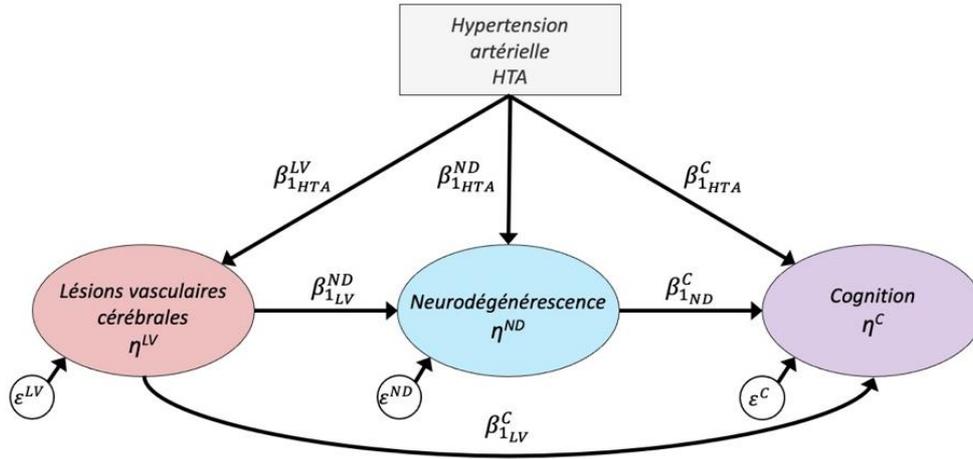


Figure 18. Représentation graphique d'un système d'équation structurelle qui inclut une variable explicative observée.

Le système d'équations structurelles est défini comme suit :

$$\begin{aligned} \eta_i^C &= \beta_0^C + \beta_{1LV}^C \times \eta_i^{LV} + \beta_{1ND}^C \times \eta_i^{ND} + \beta_{1HTA}^C \times HTA_i + \varepsilon^C & (1.2) \\ \eta_i^{ND} &= \beta_0^{ND} + \beta_{1LV}^{ND} \times \eta_i^{LV} + \beta_{1HTA}^{ND} \times HTA_i + \varepsilon^{ND} \\ \eta_i^{LV} &= \beta_0^{LV} + \beta_{1HTA}^{LV} \times HTA_i + \varepsilon^{LV} \end{aligned}$$

Où $\beta_{1HTA}^C, \beta_{1HTA}^{ND}, \beta_{1HTA}^{LV}$ sont les coefficients de régression de l'hypertension dans les 3 équations structurelles respectives. Et $\varepsilon^C, \varepsilon^{ND}$ sont les erreurs de mesures définies comme précédemment avec $\varepsilon^{LV} \sim \mathcal{N}(0, \sigma_{LV}^2)$. Par exemple β_{1HTA}^C représente l'effet de l'hypertension artérielle sur la cognition après ajustement sur les lésions vasculaires cérébrales et la neurodégénérescence.

3.2.3 Équation d'observation

Les équations structurelles définissent les relations entre les concepts du modèle par un système de régressions, mais chaque concept est mesuré par des variables observées appelées variables manifestes. Un SEM se compose donc d'un deuxième groupe de régressions appelées équations d'observations qui définissent le lien entre concepts et variables manifestes (Kline 2012).

Si on souhaite conceptualiser la cognition par une variable latente comme dans l'exemple précédent (eq 1.1), on peut utiliser les équations d'observations pour définir le lien entre le concept de cognition et un ensemble de tests neuropsychologiques mesurant les

fonctions cognitives. La variable latente peut être vue comme le facteur commun à un ensemble de variables manifestes, c'est-à-dire qu'elle capture la variance commune.

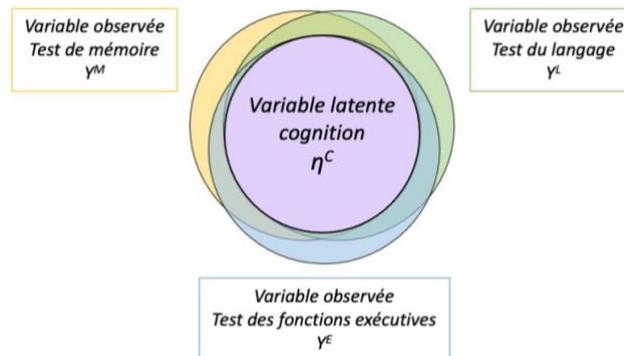


Figure 19. Représentation schématique d'une variable latente. Adapté de Hair et al. 2014

La figure ci-dessus est une représentation schématique d'une variable latente : dans cet exemple, la variable latente η^C qui conceptualise la cognition représente le chevauchement des variables observées Y^M , Y^L et Y^E qui sont respectivement les observations d'un test de mémoire (M), de langage (L) et de fonction exécutive (E).

Dans les équations d'observations on fait l'hypothèse que la variable latente est la cause sous-jacente des variables manifestes qui sont donc des observations bruitées de la variable latente. En reprenant la nomenclature graphique des SEM, cela se traduit par le graphique suivant :

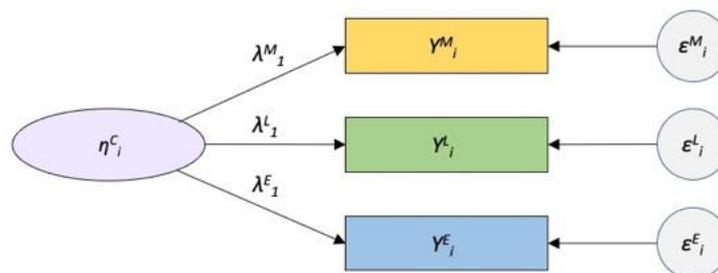


Figure 20. Représentation graphique d'une équation de mesure.

Soit les observations Y_i^M , Y_i^L et Y_i^E les scores des 3 tests neuropsychologiques, les 3 équations d'observations sont définies comme suit :

$$\begin{aligned}
 Y_i^M &= \lambda_0^M + \lambda_1^M \times \eta_i^C + \varepsilon_i^M \text{ avec } \varepsilon_i^M \sim \mathcal{N}(0, \sigma_M^2) \\
 Y_i^L &= \lambda_0^L + \lambda_1^L \times \eta_i^C + \varepsilon_i^L \text{ avec } \varepsilon_i^L \sim \mathcal{N}(0, \sigma_L^2) \\
 Y_i^E &= \lambda_0^E + \lambda_1^E \times \eta_i^C + \varepsilon_i^E \text{ avec } \varepsilon_i^E \sim \mathcal{N}(0, \sigma_E^2)
 \end{aligned}
 \tag{1.3}$$

avec λ_0^M, λ_0^L et λ_0^E respectivement les intercepts des tests de mémoire, de langage et de la fonction exécutive, λ_1^M, λ_1^L et λ_1^E les coefficients de régression associés à la variable latente cognition η^C , et $\varepsilon^M, \varepsilon^L, \varepsilon^E$ les erreurs de mesure. L'erreur de mesure est distribuée normalement et de façon indépendante. Elle correspond à la part de la variance de chaque variable manifeste qui n'est pas expliquée par la variable latente (le facteur commun).

Un SEM défini par les équations structurelles (en eq 1.1 par exemple) et les équations d'observations (en eq 1.3 pour la cognition) est identifiable seulement si on spécifie certaines contraintes. En effet la variable latente n'est pas mesurable, elle peut prendre n'importe quelle dimension et on ne peut donc pas estimer à la fois l'intercept et la variance de la variable latente ainsi que les intercepts et coefficients de proportionnalités associés aux variables manifestes. Dans l'exemple de la figure 17 et 20 pour la cognition, on ne peut à la fois estimer σ_c^2 et les 3 coefficients de régressions λ_1^M, λ_1^L et λ_1^E , un de ces paramètres d'échelle doit être fixé. Il en est de même pour β_0^C et λ_0^M, λ_0^L et λ_0^E . Pour fixer l'échelle de la variable latente, le plus souvent on définit la variable latente comme une variable aléatoire gaussienne de moyenne 0 ($\beta_0^C = 0$) et de variance 1 ($\sigma_c^2 = 1$). Dans ce cas on peut estimer librement les autres paramètres de l'équation d'observation. Une autre possibilité est de fixer l'intercept et le coefficient associé à η^C d'une seule variable manifeste à 0 ($\lambda_0^M = 0$) et 1 ($\lambda_1^M = 1$) respectivement (Hoyle 2012).

3.2.4 Étude des mécanismes d'association

Les SEM sont adaptés aux analyses multidimensionnelles pour étudier les chemins d'associations entre un ensemble de variables mesurées (observées) et non mesurées (latentes). Cette approche statistique est particulièrement pertinente pour l'étude de maladie complexe telle que la MAMA dans laquelle on souhaite vérifier un certain nombre hypothèses sur les mécanismes physiopathologiques et leurs voies d'actions.

Le premier type de relation pris en compte est le biais de confusion. Comme dans les modèles de régression classiques, si les variables X_1 et Y sont toutes les deux dépendantes de la variable X_2 alors on ajustera le modèle qui explique Y en fonction de X_1 pour estimer l'effet de X_1 sur Y indépendamment de X_2 .

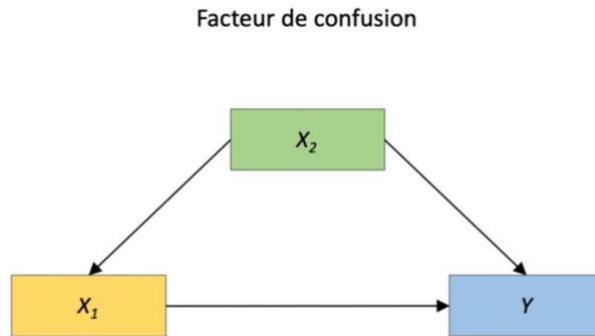


Figure 21. Illustration d'un facteur de confusion.

Les SEM facilitent aussi l'analyse de médiation lorsqu'une troisième variable se trouve sur le chemin causal dans la relation entre 2 autres variables. Dans l'exemple suivant la relation entre les variables X_1 et Y est définie par un effet direct (la flèche qui va de X_1 vers Y) et un effet indirect (via X_2 qui est un médiateur de la relation entre X_1 et Y).

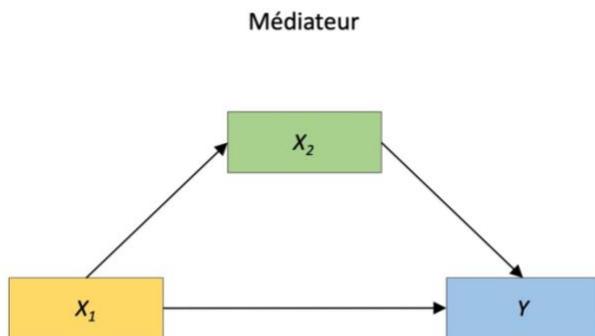


Figure 22. Illustration d'un médiateur d'effet.

Dans le cadre d'une recherche sur une maladie, les effets indirects sont particulièrement utiles pour clarifier ou expliquer pourquoi la relation entre X_1 et Y existe et conforter ou non l'hypothèse que X_2 est un mécanisme d'action physiopathologique dans cette relation.

Un troisième type de relation est représenté ci-dessous, un modificateur d'effet est une relation dans laquelle la variable X_2 affecte directement la force et/ou le sens de la relation entre X_1 et Y , en supposant que X_2 est indépendante de X_1 . L'existence d'un modificateur d'effet peut être mis en évidence soit par l'estimation d'un terme d'interaction associé au produit ($X_1 * X_2$) soit par une analyse stratifiée sur X_2 (Cheong and MacKinnon 2012).

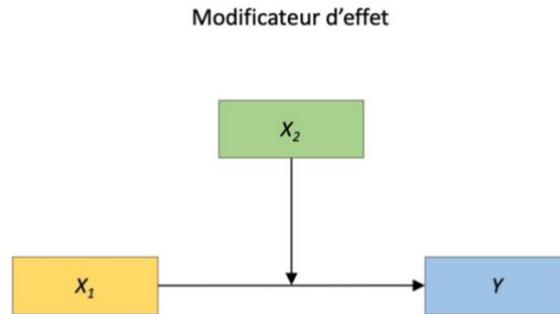


Figure 23. Illustration d'un modificateur d'effet.

3.2.5 Estimation d'un SEM

L'estimation d'un SEM consiste à estimer l'ensemble des paramètres des équations structurelles (coefficients de régression et variances) et des équations d'observations (coefficients de régression et variances). Ce vecteur est noté θ . Dans le cas d'une régression classique, les estimations $\hat{\theta}$ sont généralement obtenues par la méthode des moindres carrés où on calcule les valeurs de θ qui minimisent la somme des écarts au carré entre la valeur prédite \hat{Y}_i et l'observation Y_i . Dans le cas d'un SEM l'approche la plus classique, ou historique, n'utilise pas les observations individuelles mais travaille sur les données résumées au travers de la matrice de variance-covariances de l'échantillon de données notée S . Cette matrice est confrontée à la matrice de variance-covariance des observations déterminée via le modèle d'équation structurelle, notée $\Sigma(\theta)$. A l'aide d'une fonction d'optimisation F , on recherche les valeurs $\hat{\theta}$ qui minimisent l'écart entre la matrice de variance-covariance empirique S (celle des observations) et la matrice de variance-covariances définie par le modèle $\Sigma(\theta)$ (Lei and Wu 2012).

Par exemple, en supposant les données Gaussiennes, la distribution de S est une distribution de Wishart et on peut obtenir les estimations du maximum de vraisemblance pour la matrice observée S en minimisant la déviance F :

$$F(S, \Sigma(\theta)) = \log|\Sigma(\theta)| + tr(S\Sigma^{-1}(\theta)) - \log|S| - p \quad (1.4)$$

où p est le nombre de variables observées et tr est la fonction trace (somme des éléments de la diagonale de la matrice). L'estimation d'un SEM est réalisée par un algorithme itératif car généralement il n'existe pas de solution analytique pour déterminer $\hat{\theta}$.

Cette technique a été étendue pour améliorer les performances. On peut noter les estimateurs MLM (« mean adjusted ML ») et MLMV (« mean and variance adjusted ML ») qui sont ajustés sur la moyenne et la variance par la correction de Satorra-Bentler (Satorra and Bentler 1994) lorsque l'hypothèse de normalité n'est pas respectée. L'estimateur MLR (« robust ML ») propose aussi une approche robuste à la non-normalité des données par une correction de la variance et de la statistique de test (Yuan and Bentler 1998; T Asparouhov and Muthén 2005).

Estimation en présence de données manquantes

Le problème majeur de cette approche est qu'elle nécessite des données complètes pour calculer S. Or, la présence de données manquantes est courante dans les études de cohorte et elles peuvent avoir plusieurs causes. En fonction du mécanisme de survenue des données manquantes celles-ci peuvent induire un biais dans les analyses. Little et Rubin ont décrit 3 mécanismes du processus d'observations (Little and Rubin 2019) ; (1) les données sont manquantes complètement au hasard (MCAR) lorsque la probabilité d'observer la variable Y est indépendante des valeurs observées Y^o et non-observées Y^m mais peut dépendre de variables explicatives X inclus dans les modèles ; (2) les données sont manquantes au hasard (MAR) lorsque la probabilité d'observer la variable Y est indépendante des valeurs non-observées Y^m mais peut dépendre des valeurs observées Y^o et explicatives X ; et (3) les données sont manquantes non au hasard (MNAR) lorsque la probabilité d'observer la variable Y dépend des valeurs non-observées Y^m .

Plutôt que de définir un estimateur de maximum de vraisemblance (ML) sur la matrice de variances-covariances (issue des données complètes), on peut directement maximiser la vraisemblance des données individuelles. Cette technique est appelée FIML. Elle correspond à la technique ML dans les autres domaines d'études statistiques. Cette technique utilise toute l'information disponible et est robuste aux données manquantes MCAR et MAR (Finkelstein 1979).

Notons L_i , la contribution individuelle à la vraisemblance ($i = 1, \dots, N$) définie comme suit :

$$L_{i(\theta)} = K_i - \frac{1}{2} \log |\Sigma_{i(\theta)}| - \frac{1}{2} (x_i - \mu_{i(\theta)})' \Sigma_{i(\theta)}^{-1} (x_i - \mu_{i(\theta)}) \quad (1.5)$$

Où x_i est l'ensemble des observations du sujet i , $\mu_{i(\theta)}$ est la moyenne individuelle dérivée du modèle, $\Sigma_{i(\theta)}$ est la variance-covariance dérivée du modèle et K_i est une constante. La somme des contributions individuelles est égale à la log-vraisemblance L de l'échantillon. Supposons par exemple que pour l'individu i on dispose des observations pour q variables parmi les p variables du modèle : x_i est le vecteur des q observations, μ_i est un vecteur de taille q et Σ_i de dimensions $q \times q$ (Enders 2006).

Logiciels

Plusieurs logiciels permettent de spécifier, de visualiser et d'estimer un SEM. Les plus répandus sont LISREL (Jöreskog and Sörbom 2018), Mplus (Muthén and Muthen 2017) et Amos (Artbuckle 2019). Dans la catégorie des logiciels libres de droits, on retrouve dans le langage de programmation statistique R les packages lava (Holst and Budtz-Jørgensen 2013) et lavaan (Rosseel 2012) pour l'inférence fréquentiste et blavaan (Merkle and Rosseel 2018) pour l'inférence bayésienne qui permet une interface avec les langages JAGS et Stan. La plupart des méthodes d'estimation sont implémentées dans les logiciels de SEM pour s'adapter aux objectifs de l'analyse et aux caractéristiques des données. Les modèles développés dans le cadre des travaux de la thèse ont été estimés par FIML (sous l'hypothèse que les données soient MAR) avec le package lavaan.

3.2.6 Évaluation des performances

Du fait de la structure très paramétrique du modèle et des hypothèses structurelles réalisées, il est important d'évaluer a posteriori l'adéquation d'un SEM aux données. Il existe de nombreux indices ponctuels.

Test du Chi-2

Le test du Chi-2 est largement répandu pour évaluer l'adéquation globale d'un SEM aux données, l'hypothèse nulle est $\Sigma(\theta) = S$. Il peut être vu comme un cas particulier du rapport de vraisemblance où les modèles emboîtés seraient la matrice de variance-covariances des observations et la matrice de variance-covariances définie par le modèle (Jöreskog 1969). Soit \hat{F} la déviance minimisée (correspondant à la vraisemblance maximisée) et N la taille de l'échantillon, la statistique du test T est définie comme : $T = (N - 1) \times \hat{F}$.

Sous l'hypothèse nulle, la statistique de test T suit une distribution de χ^2 à df degrés de liberté avec $df = p^* - q$ où p^* est le nombre d'éléments non redondants de la matrice variance-covariances S du modèle et q est le nombre de paramètres à estimer que contient $\Sigma(\theta)$.

Si la statistique du test n'est pas dans la région critique, l'hypothèse nulle n'est pas rejetée et on peut conclure que le modèle est ajusté aux données. Les SEM qui portent sur l'analyse de grands échantillons sont rarement évalués avec le Chi-2 seul puisque la statistique de test devient très sensible à un faible écart et l'hypothèse nulle souvent rejetée.

D'autres indicateurs ont été développés avec l'objectif d'évaluer l'ajustement d'un SEM (indiqué par k) en comparaison du modèle dit nul qui suppose l'absence de toutes relations entre les variables (indiqué par 0). Ils sont généralement basés sur la même statistique T que le test du chi-2, notées T_k et T_0 respectivement. Cependant ces indices comparatifs ne reposent pas sur des lois de distributions sous l'hypothèse nulle et ne permettent donc pas d'évaluer la performance d'un modèle à l'aide d'un test statistique. Ils reposent donc sur des règles arbitraires de valeurs seuils qui sont parfois controversées (Hayduk 2018; Kline 2018).

Erreur quadratique moyenne standardisée

Le RMR est la moyenne des carrés des corrélations résiduelles absolues, c'est-à-dire la différence entre les paramètres de variances-covariances de $\Sigma(\theta)$ et de S . Comme la dimension du RMR dépend des données, on utilise sa version standardisée (SRMR) qui permet de donner des valeurs seuils. On considère que l'ajustement aux données est bon avec un $SRMR \leq 0,05$ et acceptable si $\leq 0,1$ (Bentler 1995).

Erreur quadratique moyenne d'approximation

C'est un indicateur, noté RMSEA, ≥ 0 qui mesure l'écart entre la matrice de variance-covariance supposée par le modèle et celle observée dans les données. Le RMSEA est estimé pour un modèle indiqué par k à partir de la déviance via la statistique T du test du Chi-2 et le nombre de degrés de liberté df comme suit (Steiger and Lind 1980) :

$$\widehat{RMSEA} = \sqrt{\frac{\max(T_k - df_k, 0)}{df_k(N - 1)}} \quad (1.6)$$

Un RMSEA $\leq 0,05$ indique un modèle très bien ajusté et $\leq 0,08$ un modèle bien ajusté (Browne and Cudeck 1992).

Indicateur de Tucker-Lewis

Le TLI (« Tucker-Lewis Index ») est un indicateur qui mesure l'augmentation de la qualité d'ajustement entre le modèle proposé (k) et le modèle nul (0) (Tucker and Lewis 1973). Il est estimé dans les échantillons à partir des statistiques T du test de Chi 2 et leur nombre de paramètres df comme suit :

$$\widehat{TLI} = \frac{\frac{T_0}{df_0} - \frac{T_k}{df_k}}{\frac{T_0}{df_0} - 1} \quad (1.7)$$

Le TLI est généralement compris entre 0 et 1. Une valeur supérieure ou égale à 0,90 (Bentler and Bonett 1980) ou 0,95 (Hu and Bentler 1999) indique un bon ajustement aux données.

Indicateur comparatif d'ajustement

Le CFI (« comparative fit index ») (Bentler 1990) est aussi un indicateur qui mesure l'amélioration de l'ajustement aux données obtenu dans le modèle à tester par rapport au modèle nul. Il est estimé dans les échantillons comme suit :

$$\widehat{CFI} = \frac{\max(T_0 - df_0, 0) - \max(T_k - df_k, 0)}{\max(T_0 - df_0, 0)} \quad (1.8)$$

Tout comme le TLI, le CFI est compris entre 0 et 1, une valeur supérieure ou égale à 0,90 (Bentler and Bonett 1980) ou 0,95 (Hu and Bentler 1999) indique un bon ajustement aux données.

3.3 Études des marqueurs de progression par des modèles à temps de maladie latent

Pour décrire l'histoire naturelle de la maladie, on souhaite analyser les trajectoires de marqueurs de l'évolution clinique. Les études de cohortes permettent de recueillir des mesures répétées des marqueurs chez le même individu et à des temps distincts, on appelle cela des données longitudinales. La répétition de la mesure chez un même individu au cours du temps induit une corrélation intra-individuelle qui ne permet pas d'appliquer les modèles de régression classiques car les observations ne sont plus indépendantes entre elles. L'analyse de données longitudinales nécessite donc, l'utilisation de méthodes spécifiques pour tenir compte de la corrélation intra-individuelle, principalement les modèles mixtes décrits ci-dessous.

3.3.1 Modèles mixtes

Les modèles mixtes sont des modèles de régression dans lesquels un ou plusieurs paramètres spécifiques à l'individu sont ajoutés pour tenir compte de la corrélation intra-individuelle entre les observations d'un même individu.

Modèles linéaires mixtes

Soit un échantillon d'analyse constitué de N individus notés i ($i = 1, \dots, N$) pour lesquels on aurait collecté les mesures du marqueur Y à j occasions ($j = 1, \dots, n_i$) notées Y_{ij} où n_i est le nombre de mesures pour l'individu i . La formulation du modèle linéaire mixte standard (eq 2.1) de (Laird and Ware 1982) s'écrit comme suit :

$$Y_{ij} = X_{ij}\beta + Z_{ij}u_i + \epsilon_{ij} \text{ avec } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \text{ et } u_i \sim \mathcal{N}(0, B) \quad (2.1)$$

On parle de modèle mixte car ce type de modèle se compose d'une part fixe et d'une part aléatoire. La part fixe correspond aux paramètres de régression au niveau populationnel (β) qui sont associés à un ensemble de p variables explicatives X_{ij} pour l'individu i à l'occasion j . La part aléatoire correspond aux paramètres de régression individuels (u_i) qui sont associés à un ensemble de q ($q \leq p$) variables explicatives Z_{ij} pour l'individu i à l'occasion j . Les paramètres individuels sont des effets aléatoires distribués selon une loi normale centrée en 0 de variance-covariance B . Le modèle inclut aussi des erreurs de mesure ϵ_{ij} pour

l'individu i à l'occasion j qui sont généralement indépendantes et identiquement distribuées selon une loi normale de moyenne 0 et de variance σ^2 . D'autres structures de covariances sont envisageables pour les erreurs de mesures comme les structures autorégressives mais nous ne les utiliserons pas dans le cadre des travaux de la thèse (Jones and Boadi-Boateng 1991). Puisque la part aléatoire $(Z_{ij}u_i)$ est d'espérance nulle, elle représente la déviation de l'individu i à l'occasion j par rapport à l'espérance de la trajectoire populationnelle $(X_{ij}\beta)$. Les effets aléatoires u_i prennent en compte la corrélation entre les données répétées d'un même sujet ou de façon similaire la variabilité interindividuelle.

La matrice variance-covariance B des effets aléatoires définit la structure de corrélation des effets aléatoires qui permet généralement la corrélation de tous les effets aléatoires. Cependant, il est possible de limiter la corrélation à une partie seulement des effets aléatoires selon des hypothèses spécifiques ou encore de ne permettre aucune corrélation pour des modèles qui sont difficiles à estimer à cause du nombre important de paramètres.

Un exemple classique de modèle linéaire mixte est celui qui suppose une évolution linéaire au cours du temps avec intercept et pente aléatoires. Si le temps d'observation t_{ij} de la mesure Y_{ij} est une variable explicative, on peut écrire le modèle de la façon suivante pour décrire la trajectoire de Y en fonction du temps

$$Y_{ij} = \gamma_0 + \gamma_1 t_{ij} + X_{ij}\beta + u_{0i} + u_{1i}t_{ij} + \epsilon_{ij}$$

$$\text{avec } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \text{ et } u_i = (u_{0i}, u_{1i})^\top \sim \mathcal{N}\left(0, B = \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0u_1} \\ \sigma_{u_0u_1} & \sigma_{u_1}^2 \end{bmatrix}\right)$$

où γ_0 est l'intercept populationnel et γ_1 représente la pente moyenne (populationnelle) de l'évolution en fonction du temps t_{ij} . X_{ij} est un vecteur regroupant un ensemble de p variables explicatives de l'individu i à l'occasion j . La déviation individuelle par rapport à la trajectoire moyenne au cours du temps est définie par un intercept aléatoire u_{0i} et une pente aléatoire u_{1i} . Les variances $\sigma_{u_0}^2$ et $\sigma_{u_1}^2$ de l'intercept et de la pente aléatoires ainsi que leur covariance $\sigma_{u_0u_1}$ définissent la matrice B de variance-covariance des effets aléatoires. La variance σ^2 capture la variabilité des erreurs de mesures.

Pour décrire des trajectoires non linéaires, X_{ij} et Z_{ij} peuvent inclure des fonctions du temps non linéaires, comme des Bases de Splines ou des polynômes fractionnaires.

Modèles non linéaires mixtes

Les modèles linéaires mixtes sont limités à l'estimation d'un prédicteur linéaire des observations Y_{ij} . Pour aller plus loin, on peut appliquer la théorie des effets aléatoires à un prédicteur non linéaire défini par une fonction f spécifique (Lindstrom and Bates 1990) :

$$Y_{ij} = f(X_{ij}, \phi_i) + \epsilon_{ij} \text{ avec } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \quad (2.2)$$

$$\phi_i = \beta + u_i \text{ avec } u_i \sim \mathcal{N}(0, B)$$

avec X_{ij} un ensemble de variables explicatives qui peut comprendre des fonctions du temps t_{ij} , et ϕ_i un ensemble de paramètres associés aux variables explicatives. Comme dans le modèle linéaire mixte, les paramètres ϕ_i de la fonction f se différencient avec une part d'effets fixes populationnels β et une part d'effets aléatoires individuels u_i qui sont définis selon une loi généralement gaussienne centrée en 0 de variance-covariance B . Les erreurs ϵ_{ij} sont définies comme dans le modèle linéaire mixte.

3.3.2 Modèles de progression

L'objectif des modèles de progression de la maladie en temps continu est de remplacer le temps classiquement utilisé dans les modèles mixtes classiques par une fonction g qui transforme l'échelle du temps comme évoqué en introduction, sous l'hypothèse d'une trajectoire commune de progression de la maladie. Le temps transformé par g est appelé temps latent et noté $s_i(t)$:

$$s_i(t) = g(t, \omega_i) \quad (2.3)$$

g est une fonction croissante monotone de transformation du temps observé t qui est définie par une combinaison d'effets aléatoires ω_i de distribution \mathcal{G} . Le temps latent est donc une caractéristique spécifique à l'individu de façon similaire aux effets aléatoires sur l'intercept et la pente des modèles mixtes classiques.

Les données répétées des marqueurs sont ensuite définies selon un modèle mixte qui s'exprime selon $s_i(t)$:

$$Y_{ij} = f(s_i(t_{ij}), X_{ij}, \phi_i) + \epsilon_{ij} \quad (2.4)$$

où ϕ_i et ϵ_{ij} sont définis comme en eq 2.2.

Plusieurs modèles de progression ont été proposés dans la littérature. Ils se différencient par la fonction g pour définir le temps latent, et la fonction f pour définir la forme des trajectoires. Les choix méthodologiques dépendaient des contraintes numériques pour l'estimation des modèles et des hypothèses physiopathologiques sur la progression de la maladie. Les principales caractéristiques de ces modèles sont résumées dans le tableau 3 avec la fonction utilisée pour définir le temps latent de maladie, la fonction utilisée pour modéliser la trajectoire des marqueurs, les effets aléatoires spécifiés et la méthode d'estimation.

Définition du temps latent

La transformation du temps observé t peut notamment se faire avec un paramètre individuel de décalage temporel (Donohue et al. 2014; Li et al. 2017; Lorenzi et al. 2017; Raket 2020; Garbarino and Lorenzi 2021) :

$$s_i(t) = a_i + t \quad (2.5)$$

Le paramètre individuel a_i est un effet aléatoire qui représente la déviation temporelle de l'individu i par rapport à la moyenne populationnelle, il est souvent défini selon une loi normale et centré en 0. Par exemple un individu i caractérisé par un décalage temporel $a_i = -3$ signifie que sa progression dans l'échelle de temps de la maladie se situe 3 ans plus tôt par rapport à la progression moyenne de tous les individus.

D'autres auteurs proposent de prendre aussi en compte une variabilité inter-individuelle sur la vitesse de progression de la maladie avec une combinaison de paramètres individuels de vitesse de progression et de décalage temporel (Jedynak et al. 2012; Bilgel et al. 2016; Marinescu et al. 2017; Schiratti et al. 2017; Koval et al. 2021) :

$$s_i(t) = a_i + b_i t \quad (2.6)$$

Le paramètre b_i est un deuxième effet aléatoire qui représente une déviation de l'individu i par rapport à la vitesse moyenne populationnelle, b_i est souvent défini selon une loi normale de moyenne 1. Par exemple un individu i caractérisé par une vitesse de progression $b_i = 1,5$ signifie que sa vitesse de progression dans l'échelle de temps de la maladie est 1,5 fois plus rapide que la moyenne des individus.

Enfin (Bilgel and Jedynak 2019) proposent d'utiliser une fonction du temps exponentielle avec un paramètre unique qui combine décalage temporel et vitesse

individuelle pour tenir compte d'une possible accélération de la progression de la maladie avec l'évolution des stades cliniques (sans trouble cognitif, trouble cognitif léger et dément) (Jedynak et al. 2012) :

$$s_i(t) = c_i e^{\gamma t} \quad (2.7)$$

Le paramètre c_i est un effet aléatoire individuel sur la fonction exponentielle du temps définie par un paramètre fixe populationnel γ qui décrit la progression de la maladie.

Exemple de décalage temporel

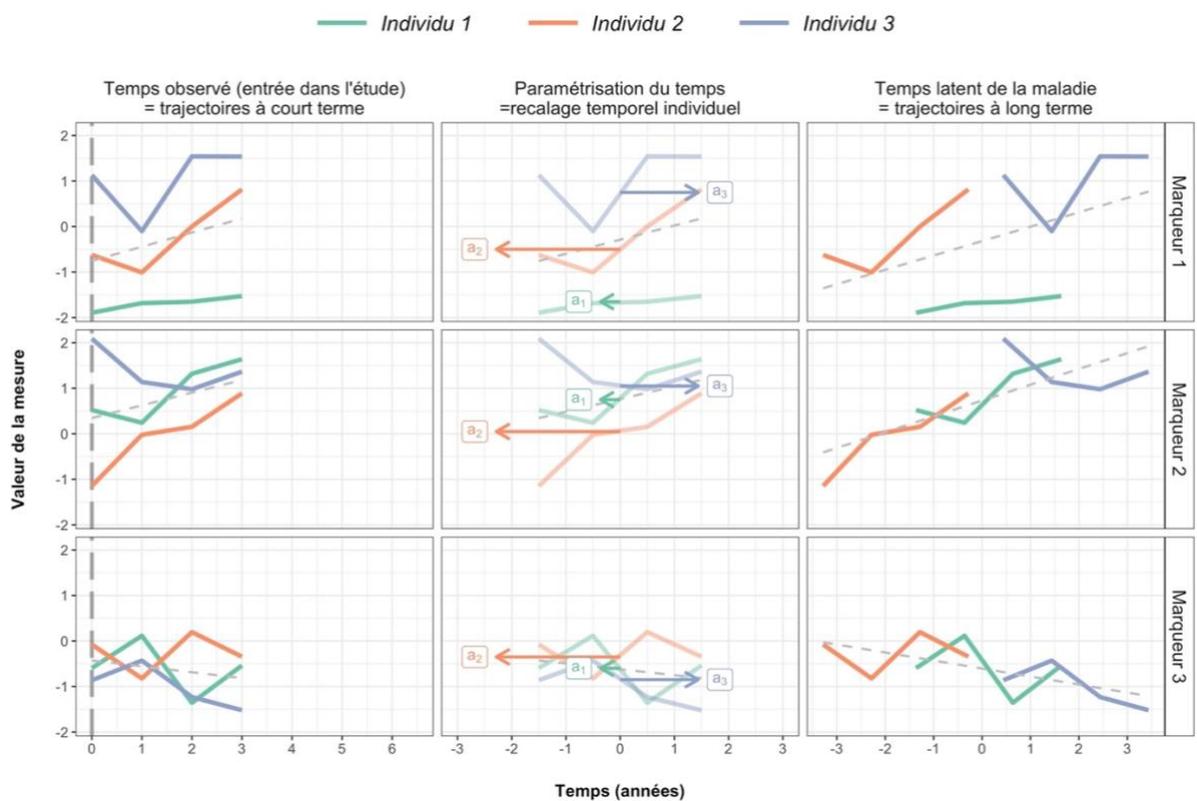


Figure 24. Exemple de prise en compte de l'hétérogénéité inter-individuelle par recalage temporel pour estimer les trajectoires à long terme des marqueurs au cours de la progression de la maladie.

La figure 24 permet d'illustrer un exemple où une partie de la variabilité inter-individuelle de 3 individus peut être prise en compte par une transformation du temps avec décalage temporel comme spécifié par l'équation (eq 2.5). La colonne de gauche représente les mesures à l'inclusion et répétées à 1 an, 2 ans et 3 ans selon le temps depuis l'inclusion pour 3 individus et 3 marqueurs différents de haut en bas. On suppose que toutes les mesures appartiennent à une progression commune de la maladie qui décrit la trajectoire des 3 marqueurs en fonction du temps de maladie. La colonne du milieu représente les paramètres

a_1 , a_2 et a_3 qui sont les effets aléatoires de décalage temporel pour réaligner au mieux les trajectoires des individus 1, 2 et 3 dans l'échelle du temps latent. La colonne de droite représente les mêmes trajectoires individuelles après prise en compte du décalage temporelle pour décrire les trajectoires globales des 3 marqueurs en fonction du temps latent de maladie.

Tableau 3. Caractéristiques des principaux travaux sur les modèles de progression de la MAMA dans la littérature ; k marqueur ($k=1, \dots, K$) ; i sujet ($i=1, \dots, N$) ; j occasion ($j=1, \dots, n_{ik}$)

Auteurs	Paramètre individuel du temps de maladie	Modélisation des trajectoires	Estimation
Jedynak et al. 2012	Recalage temporel et vitesse de progression individuels $s_{ijk} = b_i \times t_{ijk} + a_i$	Fonction logistique $Y_{ijk} = l_k \times (1 + e^{-m_k \times (s_{ijk} - r_k)})^{-1} + v_k + \epsilon_{ijk}$	Alternating Least Square (ALS) avec l'algorithme de Levenberg-Marquardt
Donohue et al. 2014	Recalage temporel individuel $s_{ijk} = t_{ijk} + a_i$	Fonction percentiles F_{Y_k} et modèle non linéaire mixte $\tilde{Y}_{ijk} = \Phi^{-1}(F_{Y_k}(Y_{ijk})) = g_k(s_{ijk}) + u_{0ik} + u_{1ik} \times t_{ijk} + \epsilon_{ijk}$ avec g_k une fonction monotone continue	Optimisation des moindres carrés résiduels (modèle additif généralisé) [packages R fda et lme4]
Marinescu et al. 2017	Recalage temporel et vitesse de progression individuels $s_{ijk} = b_i \times t_{ijk} + a_i$	Fonction logistique $Y_{ijk} = l_k \times (1 + e^{-m_k \times (s_{ijk} - r_k)})^{-1} + v_k + \epsilon_{ijk}$	Approximation stochastique EM (MCMC SAEM)
Schiratti et al. 2017	Recalage temporel et vitesse de progression individuels $s_{ijk} = b_i(t_{ijk} - t_0 - a_i)$	Fonction logistique $Y_{ijk} = \left(1 + \left(\frac{1}{p_0} - 1\right) e^{-\frac{v_0 s_{ijk}}{p_0(1-p_0)}}\right)^{-1} + \epsilon_{ijk}$	Approximation stochastique EM (MCMC SAEM) [Monolix]
Li et al. 2017	Recalage temporel individuel $s_{ijk} = t_{ijk} + a_i$	Fonction percentiles F_{Y_k} et modèle linéaire mixte $\tilde{Y}_{ijk} = \Phi^{-1}(F_{Y_k}(Y_{ijk})) = X_{ijk}\beta_k + \gamma_k s_{ijk} + u_{0ik} + u_{1ik} t_{ijk} + \epsilon_{ijk}$	MCMC avec HMC-NUTS [Stan]
Lorenzi et al. 2019	Recalage temporel individuel $s_{ijk} = t_{ijk} + a_i$	Fonction percentiles F_{Y_k} et processus gaussiens $\tilde{Y}_{ijk} = \Phi^{-1}(F_{Y_k}(Y_{ijk})) = f(s_{ijk}) + v^i(s_{ijk}) + \epsilon_{ijk}$ avec $f \sim \mathcal{GP}(0, \Sigma_G)$	Maximisation de la log-vraisemblance approchée
Bilgel et al. 2019	Transformation exponentielle individuelle $s_{ijk} = c_i e^{\gamma(t_{ijk} - t_0)}$	Fonction logistique $Y_{ijk} = \pi_k \times \sum_{z=1}^Z l_{kz} \times \left(1 + e^{-\frac{s_{ijk} - r_z}{\sigma_{basis}^2}}\right)^{-1} + v_k + \epsilon_{ijk}$	MCMC avec HMC-NUTS [Package Python PyMC3]
Koval et al. 2021	Recalage temporel individuel $s_{ijk} = b_i(t_{ijk} - t_0 - a_i) + t_0$	Modèle non linéaire mixte $Y_{ijk} = f_k(\theta_k, z_i, s_{ijk}) + \epsilon_{ijk}$ avec θ_k un vecteur d'effets fixes et z_i un recalage individuel dans l'espace	Approximation stochastique EM (MCMC SAEM)
Garbarino et al. 2021	Recalage temporel individuel $s_{ijk} = t_{ijk} + a_i$	Processus gaussiens $Y_{ijk} = f(s_{ijk}) + v^i(s_{ijk}) + \epsilon$ avec $f \sim \mathcal{GP}(0, \Sigma_G)$	Optimisation de la borne inférieure de la vraisemblance marginale (ELBO)
Künhel et al. 2021	Recalage temporel individuel dépendant de covariables $s_{ijk} = t_{ijk} + a_i + X_{ijk}\beta_k$	Fonction exponentielle $Y_{ijk} = l_k \times \exp\left(\frac{s_{ijk}}{e^h}\right) + m_k + u_{ik} + \epsilon_{ijk}$	Maximisation de la vraisemblance restreinte (REML) [Package R : Progmod basé sur NLME]

Définition des formes des trajectoires

Les modèles de progression de la maladie sont applicables à un marqueur unique Y_{ij} dans le cadre d'une analyse univariée (Raket 2020). Cependant l'intérêt des modèles de progression de la maladie est d'estimer le temps latent en exploitant tous les aspects physiopathologiques de la maladie qui sont mesurés par les données multimodales dans des études de cohortes. La formulation du modèle de progression peut être étendue aux cas multivariés où on modélise simultanément les trajectoires Y_{ijk} de K marqueurs ($k = 1, \dots, K$) dans l'échelle de temps de la maladie :

$$Y_{ijk} = f_k(s_{ijk}; X_{ijk}, \beta_k + u_{ik}) + \epsilon_{ijk} \quad (2.8)$$

La forme de la trajectoire des marqueurs peut être modélisée par différentes méthodes selon la fonction f_k choisie. Dans la littérature ont été proposé par exemple :

- Une fonction exponentielle (Kühnel et al. 2021)

$$Y_{ijk} = l_k \cdot \exp\left(\frac{s_i(t_{ijk}) + X_{ijk}\beta_k}{\exp(h)}\right) + m_k + u_{ik} + \epsilon_{ijk} \quad (2.9)$$

où l_k, m_k, β_k sont des effets fixes propres au marqueur k , u_{ik} un effet aléatoire spécifique au marqueur k , et h un paramètre d'échelle du temps.

- Une fonction logistique (Jedynak et al. 2012; Bilgel and Jedynak 2019)

$$Y_{ijk} = l_k \left(1 + e^{-m_k(s_i(t_{ijk}) - r_k)}\right)^{-1} + v_k + \epsilon_{ijk} \quad (2.10)$$

où l_k, m_k, r_k et v_k sont des effets fixes propres au marqueur k .

- Une fonction non paramétrique avec un processus gaussien (Lorenzi et al. 2017)

$$Y^i(s_{ijk}) = f(s_i(t_{ijk})) + v^i(s_i(t_{ijk})) + \epsilon_{ijk} \text{ avec } f \sim \mathcal{GP}(0, \Sigma_G) \text{ et } v^i \sim \mathcal{N}(0, \Sigma_s) \quad (2.11)$$

où Y^i, f, v^i, ϵ sont des structures de données de dimensions K marqueurs et K^j temps d'observation t .

- Une fonction quantile qui permet de restreindre les valeurs à l'espace $[0 ; 1]$ puis de ramener ces valeurs dans l'espace normal pour estimer une régression linéaire (Donohue et al. 2014; Li et al. 2017; Lorenzi et al. 2017). Par exemple le modèle LTJMM :

$$\tilde{Y}_{ijk} = \Phi^{-1}\left(F_{Y_k}(Y_{ijk})\right) = X_{ijk}\beta_k + \gamma_k S_{ijk} + u_{0ik} + u_{1ik}t_{ijk} + \epsilon_{ijk} \quad (2.12)$$

Où (Li et al. 2017) proposent d'utiliser la distribution des mesures d'un marqueur dans l'échelle des percentiles avec une fonction quantile F_{Y_k} , les observations Y_{ijk} deviennent P_{ijk} dans l'échelle des percentiles. Puis l'inverse de la fonction de distribution normale Φ^{-1} est appliquée pour estimer un modèle linéaire mixte sur les observations transformées \tilde{Y}_{ijk} qui ont une distribution gaussienne.

Définition de l'échelle de sévérité

En définissant la forme des trajectoires, la fonction f_k spécifie aussi comment l'échelle de sévérité des marqueurs est prise en compte. Les observations Y_{ijk} peuvent être modélisées dans l'échelle naturelle des marqueurs (Bilgel and Jedynak 2019; Kühnel et al. 2021) ce qui permet d'interpréter la mesure réelle en fonction du temps de progression de la maladie mais ne permet pas de superposer les trajectoires selon la même échelle dans un graphe pour comparer la temporalité des marqueurs comme le propose le modèle théorique de la figure 7 (Jack et al. 2010).

La fonction f_k de type logistique de (Jedynak et al. 2012) est estimée dans l'échelle naturelle des marqueurs tout en estimant l_k et v_k qui définissent les bornes l_k et $v_k + l_k$ de la sigmoïde pour chaque marqueur, les auteurs proposent ensuite de définir la temporalité des marqueurs par le point d'inflexion de la courbe sigmoïde.

Enfin, la fonction quantile F_{Y_k} (Donohue et al. 2014; Li et al. 2017; Lorenzi et al. 2017) est pratique puisqu'elle permet de modéliser la phase d'accélération de la progression des marqueurs de façon non linéaire dans l'échelle de sévérité des marqueurs comme dans le modèle théorique puisque toutes les mesures sont situées dans l'intervalle $[0 ; 1]$, elles sont donc comparables entre les marqueurs.

Avec la transformation quantile, l'échelle de sévérité du marqueur est cependant dépendante de l'échantillon d'étude. Si l'échantillon est composé d'une proportion plus importante de personnes sans troubles cognitifs que de personnes avec un trouble cognitif léger ou au stade de démence les quantiles, même haut, représentent des niveaux d'atteinte faible et l'échelle de sévérité ne sera pas représentative de tous les stades de progression de la maladie.

Une solution est alors d'utiliser $F_{Y_k}(Y_{ijk}; w_{ij})$ une fonction pondérée de percentiles où les mesures Y_{ijk} sont pondérées par une caractéristique individuelle w_{ij} , par exemple une donnée clinique (comme le statut clinique ou la CDR) pour homogénéiser la représentation des différents stades de progression de la maladie. De cette façon le 50^{ème} percentile correspond approximativement à la sévérité médiane du marqueur qui est atteinte au cours de l'histoire naturelle de la maladie.

La figure ci-dessous permet d'illustrer cette transformation en quantiles en représentant les échelles de percentiles du marqueur K obtenues sans et avec pondération pour une population fictive ($N = 1000$) qui comprendrait 40% de participants sans trouble cognitif léger, 30% avec un trouble cognitif léger, 25% avec un trouble cognitif modéré et 5% avec un trouble cognitif sévère.

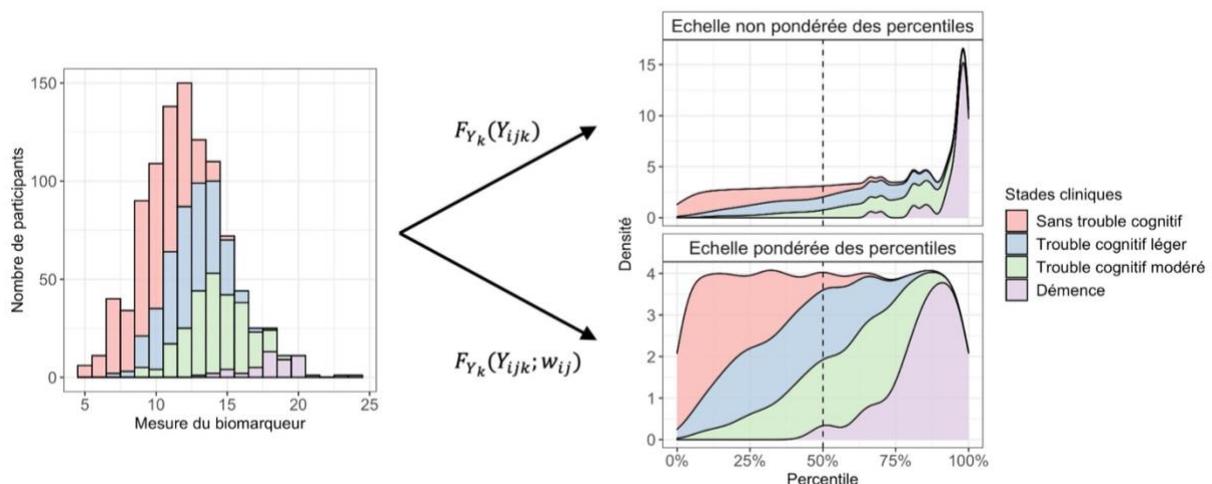


Figure 25. Illustration de la construction de l'échelle de sévérité des marqueurs par la pondération ou la non-pondération des percentiles.

L'histogramme de gauche représente la distribution des mesures du marqueur K selon le stade clinique. On voit que les mesures recueillies chez les participants avec un trouble cognitif modéré et avec un diagnostic de démence sont sous-représentées dans la distribution. Si on utilise l'échelle des percentiles de cette distribution sans pondération $F_{Y_k}(Y_{ijk})$, on voit que le 50^{ème} percentile est globalement aussi bien représenté par les mesures recueillies chez les participants avec un trouble modéré que ceux avec un trouble cognitif léger et ceux sans trouble cognitif. Or le 50^{ème} percentile est supposé correspondre à une sévérité de 50% du marqueur K au cours de l'histoire naturelle de la maladie. On peut aussi remarquer que les mesures recueillies chez les participants avec un diagnostic de démence sont très majoritairement représentées entre le 95^{ème} et le 100^{ème} percentile, ce qui

signifie que l'échelle de sévérité représente essentiellement l'étendue des valeurs possibles du marqueur K chez les participants avec un trouble cognitif léger ou sans trouble cognitif.

Dans le cas où on utilise l'échelle pondérée des percentiles $F_{Y_k}(Y_{ijk}; w_{ij})$ chaque valeur Y_{ijk} est associée à un poids $w_{ij} = 1/n$ où n est le nombre de participants au même stade clinique. On voit ici que le 50^{ème} percentile est majoritairement représenté par les participants avec un trouble cognitif modéré et léger. L'échelle de sévérité entre le 1^{er} et le 25^{ème} percentile est majoritairement représentées par les mesures recueillies chez les participants sans troubles cognitifs et entre le 75^{ème} et 100^{ème} percentile majoritairement représentées par les mesures recueillies chez les participants avec un diagnostic de démence.

Effets aléatoires

Les précédents travaux présentaient aussi des différences méthodologiques dans la spécification des effets aléatoires pour prendre en compte la variabilité inter-individuelle. Plusieurs modèles spécifiaient uniquement les effets aléatoires ω_i dans la fonction de transformation du temps, le reste de la variabilité étant expliquée par l'erreur de mesure ε (Jedynak et al. 2012; Bilgel and Jedynak 2019). Or dans l'analyse de données longitudinales multivariées, on peut alors distinguer 2 sources de variabilité inter-individuelle distinctes :

- Une source de variabilité commune aux trajectoires des marqueurs. Il s'agit d'une déviation systématique des observations individuelles par rapport à la trajectoire populationnelle et qui est partagée par l'ensemble des marqueurs. Sous l'hypothèse que la progression de la maladie est définie par une fonction monotone croissante, on peut considérer que cette variabilité représente l'état individuel dans la progression de la maladie. Cette variabilité est estimée par les effets aléatoires ω_i spécifiques à l'individu dans la fonction g (eq 2.3) qui définit le temps latent de la maladie.
- Une source de variabilité spécifique à chaque marqueur. Il s'agit d'une déviation systématique des observations individuelles par rapport à la trajectoire populationnelle d'un marqueur. On peut l'estimer par les effets aléatoires u_{ik} spécifiques à l'individu i et au marqueur k .

On remarque que pour bien différencier les 2 sources de variabilité il faut tenir compte de la variabilité partagée et de la variabilité non-partagée par les marqueurs, ce qui implique de considérer la distribution des effets aléatoires u_{ik} indépendante entre les marqueurs pour capturer le maximum de la variabilité partagée par les effets aléatoires partagés ω_i . Or

plusieurs modèles ont proposé des distributions corrélées de ces effets aléatoires ou de l'erreur de mesure (Jedynak et al. 2012; Lorenzi et al. 2017; Li et al. 2017; Bilgel and Jedynak 2019; Kühnel et al. 2021). Cependant supposer une distribution indépendante des u_{ik} est également une hypothèse forte, il est en réalité probable que les mesures des différents marqueurs soient corrélées entre elles pour d'autres raisons que la progression de la maladie. Une solution intermédiaire serait de définir la matrice de variance-covariance B par blocs de marqueurs. Par exemple on peut spécifier d'une part que les mesures obtenues par la même acquisition IRM soient corrélées entre elles puisqu'elles partagent des caractéristiques identiques (aimant, logiciel, manipulateur radio) et d'autre part que ces mesures sont indépendantes des scores aux tests neuropsychologiques.

Ajustement sur les variables explicatives

Nous avons vu en introduction qu'il existe des facteurs de risque de la MAMA avec notamment le génotype de l'*APOE* et le niveau d'études qui peuvent influencer la progression de la maladie. Les données modélisées sont issues de populations d'études qui ont leurs caractéristiques propres, il est donc essentiel de tenir compte de ces facteurs de risque dans les modèles de progression de maladie puisqu'ils peuvent expliquer une partie de la variabilité inter-individuelle qui n'est pas expliquée par le temps latent de maladie. Cependant une partie des travaux de la littérature ne tient pas compte des facteurs de risque pour modéliser le temps latent de la maladie. Les autres travaux proposaient des approches différentes de prise en compte des caractéristiques individuelles :

- Analyser à posteriori les associations des effets aléatoires ω_i avec les caractéristiques individuelles dans un modèle multivariable (Lorenzi et al. 2017; Bilgel and Jedynak 2019; Koval et al. 2021).
- Stratifier les modèles sur les facteurs de risques dans le cadre d'analyses de sensibilité (Donohue et al. 2014).
- Estimer des paramètres associés aux caractéristiques individuelles qui interviennent dans la fonction g de transformation du temps (Raket 2020; Kühnel et al. 2021). Cette approche suppose que les variables explicatives (facteurs de risque) ont un effet global sur la progression de la maladie, il s'agit d'un ajustement sur le temps latent.
- Estimer des paramètres associés aux caractéristiques individuelles qui interviennent dans la fonction f_k de modélisation des marqueurs (Li et al. 2017). Cette approche

suppose que les variables explicatives ont un effet sur la progression d'un ou plusieurs marqueurs indépendamment de la progression de la maladie, il s'agit d'un ajustement sur les mesures des marqueurs.

3.3.3 Estimation des modèles

Les modèles mixtes peuvent être estimés par l'approche fréquentiste qui recherche les estimations ponctuelles des paramètres et des intervalles de confiance correspondants ou par l'approche bayésienne qui estime les lois de probabilités de distribution des paramètres pour définir dans un second temps les moyennes et les intervalles de crédibilité respectifs.

Estimation fréquentiste

L'estimation des modèles linéaires mixtes se fait par maximisation de la vraisemblance (ML) à partir de la formulation marginale du modèle $Y_i \sim \mathcal{N}(X_i\beta, V_i)$ où β est le vecteur d'effets fixes et V_i la matrice de covariance qui inclut les paramètres de variances-covariances des effets aléatoires et des erreurs de mesures. Contrairement aux modèles linéaires mixtes, la distribution marginale des modèles non linéaires mixtes a rarement une solution analytique. Il est donc nécessaire de s'appuyer sur la distribution conditionnelle aux effets aléatoires et de calculer la vraisemblance avec une résolution numérique de l'intégrale sur les effets aléatoires. Cette étape de calcul supplémentaire est généralement faite par la méthode de quadrature gaussienne adaptative, la méthode de Monte-Carlo ou l'approximation de Laplace. La maximisation de la vraisemblance est réalisée par une procédure itérative le plus souvent avec les algorithmes de Newton-Raphson, Quasi-Newton ou EM (Commenges and Jacqmin-Gadda 2015).

Estimation bayésienne

L'inférence bayésienne repose sur le théorème de Bayes (Bayes and Price 1763) (eq 3.1) qui permet d'exprimer la distribution a posteriori des valeurs des paramètres θ sachant les données observées X . Elle dépend de la probabilité d'observer les données X sachant les valeurs des paramètres θ (vraisemblance), de la distribution des valeurs des paramètres θ (a priori) et de la distribution marginale des données observées X .

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)}{P(X)} \quad (3.1)$$

Alors que l'inférence fréquentiste recherche l'estimation des paramètres qui maximisent la vraisemblance, l'inférence bayésienne calcule la densité a posteriori des paramètres sachant les données. Puisque la distribution des données $P(X)$ est connue et indépendante de θ , on peut dire que la probabilité $P(\theta|X)$ est proportionnelle au produit de la distribution a priori des paramètres $P(\theta)$ et de la probabilité d'observer les données sous cette hypothèse $P(X|\theta)$ (vraisemblance). Pour le calcul on a donc besoin d'une hypothèse sur la distribution a priori des paramètres qui peut exercer une influence forte (a priori informatif ; par exemple une loi avec une petite variance) ou une faible influence (a priori non informatif).

La plupart du temps il n'existe pas de solution analytique au calcul du postérieur dans des modèles complexes si bien que l'inférence bayésienne s'appuie sur des méthodes itératives d'approximation de Monte Carlo pour construire une chaîne de Markov (MCMC) dont la distribution stationnaire π est la distribution a posteriori $P(\theta|X)$ recherchée (Metropolis et al. 1953). Une chaîne de Markov est une marche aléatoire ergodique c'est-à-dire qu'elle est irréductible (tout état est accessible à partir de n'importe quel autre état), et apériodique (la probabilité de se retrouver dans un état est indépendant des états passés, il n'est donc pas atteint selon une fréquence déterminable). Ces propriétés signifient que noter les états de la chaîne de Markov revient à constituer un échantillon à partir de sa distribution stationnaire π . L'ergodicité de la chaîne est vérifiée en implémentant une contrainte de réversibilité, c'est-à-dire que la probabilité de passer d'un état θ à un état θ^* est équivalente à la probabilité de de passer de l'état θ^* à l'état θ (Chib and Greenberg 1995). Soit K la fonction de transition et π la distribution stationnaire de la chaîne comme suit :

$$K(\theta^*|\theta) \times \pi(\theta) = K(\theta|\theta^*) \times \pi(\theta^*) \quad (3.2)$$

La chaîne de Markov est donc constituée des L vecteurs de valeurs des paramètres θ_l ($l = 1, \dots, L$) correspondants aux L itérations. A chaque itération, l'algorithme définit un ensemble de valeurs pour les paramètres θ qui appartiennent à l'espace des paramètres. En pratique on laisse de côté un certain nombre d'itérations de « chauffe » (ou « warm-up » ou « burn-in »), souvent ≥ 1000 , le temps que la chaîne de Markov converge vers sa distribution stationnaire. On garde généralement 1 proposition sur x (« éclaircissement » ou « thinning ») pour éviter les problèmes d'autocorrélation liés au fait que les propositions successives θ^* ne sont pas totalement indépendantes. Enfin il est important de s'assurer de la fiabilité des

estimations, pour cela on peut échantillonner plusieurs chaînes de Markov avec différentes valeurs initiales pour vérifier qu'elles convergent autour de la même chaîne stationnaire π .

3.3.4 Prédiction

On peut approximer la prédiction de la distribution de \bar{Y} (valeur prédite), à partir de sa distribution conditionnellement au prédicteur \bar{X} et aux L valeurs échantillonnées de θ qui constituent la chaîne MCMC comme suit :

$$P(\bar{Y}|\bar{X}, X, Y) = \int P(\bar{Y}|X, \theta) \cdot P(\theta|Y, X) d\theta \approx \frac{1}{L} \sum_{l=1}^L P(\bar{Y}|\bar{X}, \theta^l) \quad (3.3)$$

On peut alors résumer la distribution prédite par son espérance.

$$\mathbb{E}[\bar{Y}|\bar{X}, X, Y] \approx \frac{1}{L} \sum_{l=1}^L \bar{Y}^l \text{ avec } \bar{Y}^l \sim P(\bar{Y}|\bar{X}, X, Y) \quad (3.4)$$

Dans l'exemple du modèle mixte non linéaire défini à l'équation (eq 2.2) on peut calculer la moyenne \bar{Y}_{ij} et l'intervalle de crédibilité à 95% (IC95%) avec L valeurs prédites \bar{Y}_{ij}^l à partir d'un prédicteur X_{ij} et des L itérations pour les paramètres θ^l :

$$\mathbb{E}[\bar{Y}] \approx \frac{1}{L} \sum_{l=1}^L f(X, \theta^l) \text{ et } IC95 \approx [Q_{\sum_{l=1}^L \bar{Y}^l}(0,025); Q_{\sum_{l=1}^L \bar{Y}^l}(0,975)] \quad (3.5)$$

Avec Q est la fonction quantile qui permet définir la valeur du 2,5^{ème} et du 97,5^{ème} percentile.

3.4 Inférence bayésienne avec HMC-NUTS et Stan

3.4.1 Algorithmes pour constituer une chaîne MCMC

Algorithmes de type Metropolis-Hasting

L'inférence bayésienne a été en partie popularisée par le logiciel BUGS (« Bayesian inference Using Gibbs Sampling ») qui permet de construire une chaîne de Markov par échantillonnage aléatoire. Le fonctionnement est basé sur l'algorithme de Metropolis-Hasting

(Hastings 1970) qui propose à chaque itération de nouvelles valeurs de paramètres θ^* pour les p dimensions de θ , choisies aléatoirement à partir de la distribution postérieure :

$$\theta^* \sim P(\theta|X) \quad (3.6)$$

L'algorithme part d'une position initiale de distribution $\Phi(\theta)$, puis évalue la probabilité que θ^* appartienne à π qui est la distribution stationnaire de la chaîne de Markov (c'est-à-dire la distribution du postérieure $P(\theta|X)$ recherchée). Pour cela le critère d'acceptation est un calcul de probabilité qui vérifie la condition de réversibilité du changement d'état de θ à θ^* . Soit la valeur est acceptée et l'algorithme commence une nouvelle itération en partant de cette nouvelle valeur θ^* soit la valeur est rejetée et l'algorithme recommence avec une nouvelle proposition aléatoire. L'ensemble des valeurs acceptées constitue la chaîne de Markov, elles sont un échantillon tiré aléatoirement à partir de π et permettent une approximation de la distribution postérieure des paramètres $P(\theta|X)$.

L'échantillonnage de Gibbs (Geman and Geman 1984) est un cas particulier de l'algorithme de Metropolis-Hastings où les propositions des nouvelles valeurs pour les p dimensions de θ sont échantillonnées en p étapes à partir de la distribution de θ_p conditionnellement aux valeurs de θ_{p-1} . Les propositions sont donc toujours acceptées comme suit :

$$\theta_p^* \sim P(\theta_p^l | X, \theta_1^{l+1}, \dots, \theta_{p-1}^{l+1}, \theta_{p+1}^l, \dots, \theta_p^l) \quad (3.7)$$

Le problème des algorithmes de type Metropolis Hasting est qu'ils deviennent insuffisamment efficaces face à des inférences statistiques complexes qui impliquent un grand nombre de paramètres p (modèles à grandes dimensions, distributions multimodales, etc.). Dans le cas d'une inférence complexe, les valeurs échantillonnées se situent dans un espace restreint de l'espace des paramètres. Le fonctionnement très aléatoire de l'algorithme le rend alors très lent à cause d'un nombre très important de propositions de valeurs rejetées ou en explorant plusieurs fois la même partie de l'espace des paramètres. Le comportement de l'algorithme peut alors affecter la fiabilité de l'échantillonnage de la chaîne de Markov avec une exploration incomplète de π .

Une meilleure approche pour construire une chaîne de Markov est l'algorithme de Monte Carlo hamiltonien (HMC) (Duane et al. 1987). Le principe de cet algorithme est de rendre les propositions moins aléatoires pour échantillonner les valeurs de π et construire la chaîne de Markov. L'analogie de la physique est souvent utilisée pour comprendre le principe puisque ce type d'algorithme traite l'exploration de l'espace des paramètres comme un problème d'ordre physique. Les algorithmes hamiltonien permettent d'explorer l'espace des paramètres plus efficacement tout en limitant le risque que les propositions θ^* soient rejetées.

La trajectoire pour aller de l'état θ à θ^* est assimilée en mécanique hamiltonienne à une trajectoire qui dépend des lois de Newton. On peut voir le vecteur θ de p paramètres comme un objet physique (une bille par exemple) situé dans un espace à p dimensions. Dans cet espace le logarithme du postérieur recherché $P(\theta|X)$ représente la surface, avec des « creux » qui représentent une densité de probabilité plus importante que les « bosses ». On peut lancer θ dans une direction aléatoire et tirer des échantillons de son état le long de sa trajectoire pour déterminer la forme de la surface. Dans l'analogie de la bille, plus on la lance fort plus elle va s'éloigner d'un « creux ».

Le mouvement auxiliaire global ρ de la simulation est constitué de p vecteurs de mouvements pour se déplacer dans les p directions de l'espace. La transition d'état est déterminée par une fonction hamiltonienne \mathcal{H} qui définit la distribution conjointe ρ et θ . \mathcal{H} peut se décomposer en 2 termes : la densité du mouvement auxiliaire appelé énergie cinétique T , et la densité de la distribution postérieure $P(\theta|X)$ appelée énergie potentielle V (Betancourt 2017) comme suit :

$$\mathcal{H}(\rho, \theta) = T(\rho|\theta) + V(\theta) \text{ avec } \rho \sim \mathcal{N}(0, \Sigma) \quad (3.8)$$

L'énergie cinétique à une distribution gaussienne multivariée et on tire aléatoirement une nouvelle valeur pour chaque nouvelle simulation de trajectoire. Les deux types d'énergie s'équilibrent puisque \mathcal{H} est constante, ce qui permet d'explorer rapidement l'espace des paramètres tout en s'assurant de rester dans l'espace π de la distribution stationnaire de la chaîne de Markov où $P(\theta|X)$ est la plus dense. Une fois la simulation de la trajectoire effectuée la proposition θ^* est évaluée avec un critère d'acceptation de type Metropolis-Hasting.

Cependant l'approche HMC nécessite beaucoup plus de calculs que l'approche Metropolis-Hasting, le bénéfice d'explorer rapidement l'intégralité de $P(\theta|X)$ n'est perceptible que pour des cas d'inférence relativement complexes où une majorité des propositions θ^* seraient rejetées par l'approche très aléatoire de Metropolis-Hasting. Il existe aussi quelques limites à HMC. A chaque itération l'algorithme effectue une simulation de la trajectoire dans une direction aléatoire pour aller de l'état θ à θ^* . La longueur de la trajectoire est définie par le nombre et la taille des pas de la simulation qui sont 2 paramètres fixés par l'utilisateur. Si la longueur est trop petite, on peut rencontrer des difficultés à explorer l'intégralité de l'espace des paramètres et si elle est trop grande la trajectoire finit par opérer un demi-tour. Dans les deux cas, le risque est de se retrouver avec un échantillon de propositions auto corrélées qui n'explore que localement l'espace de π . L'utilisateur peut ajuster les paramètres qui définissent le nombre et la taille du pas mais cela demande d'être adapté à chaque application et peut se révéler complexe.

L'algorithme d'échantillonnage HMC-NUTS (« Hamiltonian Monte Carlo No-U-Turn Sampling ») a été développé pour faire face à une partie des limites de l'algorithme HMC en éliminant la nécessité de spécifier le nombre de pas pour les simulations de trajectoires, ce paramètre est alors déterminé de manière adaptative (Hoffman and Gelman 2014). Pour cela les simulations de trajectoires sont réalisées dans les 2 sens opposés afin d'estimer à quel moment la trajectoire commence à opérer un demi-tour, la simulation est alors arrêtée pour échantillonner θ^* .

3.4.2 Le logiciel Stan

Les inférences bayésiennes réalisées dans le cadre des travaux de la thèse ont été réalisées avec le logiciel Stan (Stan Development Team 2022c). Stan désigne à la fois un logiciel et un langage de programmation probabiliste. Il permet de faire de la modélisation statistique par inférence bayésienne avec l'avantage de pouvoir estimer des modèles complexes. La principale méthode d'estimation de Stan est l'échantillonnage MCMC par HMC-NUTS avec une amélioration supplémentaire qui permet, en plus du nombre, d'ajuster aussi de manière adaptative la taille des pas effectués lors des simulations de trajectoires (Betancourt 2017).

L'algorithme HMC-NUTS de Stan est complexe à comprendre mais les améliorations apportées pour son utilisation rendent l'échantillonnage hamiltonien accessible sans connaissances approfondies de l'inférence bayésienne. Stan a été pensé pour être utilisé par

un large public de statisticiens. Il s'agit d'un projet collaboratif open-source qui vit au travers d'un forum communautaire (<https://discourse.mc-stan.org>) et de sa plateforme GitHub pour ses différents répertoires de code (<https://github.com/stan-dev>). Stan est utilisable dans les interfaces des logiciels R, Python, MATLAB, Julia, Stata et il est également possible de l'utiliser en ligne de commande.

Nous avons fait le choix d'utiliser Stan car son usage répondait bien aux objectifs du projet de recherche puisque qu'il permet une grande liberté de spécifications d'effets fixes, d'effets aléatoires et de contraintes paramétriques. De plus le modèle de progression LTJMM (Li et al. 2017) qui a motivé une partie des recherches de la thèse était également estimé avec Stan.

Les principaux paramètres utilisateurs à renseigner pour estimer un modèle sont le nombre de chaînes MCMC à construire, le nombre d'itérations de chauffe (« warm-up »), le nombre d'itérations à échantillonner (« sampling ») et la proportion d'itérations à garder (« thinning »). Si nécessaire il est possible de spécifier des valeurs initiales θ_0 , modifier $\delta \in [0; 1]$ qui est le critère d'acceptation-rejet des propositions θ^* , ou les paramètres de comportement pour les simulations des trajectoires.

3.4.3 Indicateurs diagnostiques

Convergence des chaînes MCMC

La convergence des chaînes MCMC à l'équilibre sur la chaîne stationnaire π peut se vérifier graphiquement par l'examen des traces, c'est-à-dire que pour chaque paramètre on vérifie que la valeur du paramètre à chaque itération est stabilisée autour de la même position pour les différentes chaînes. Par exemple la figure suivante est représentative de chaînes qui ne se sont pas stabilisées :

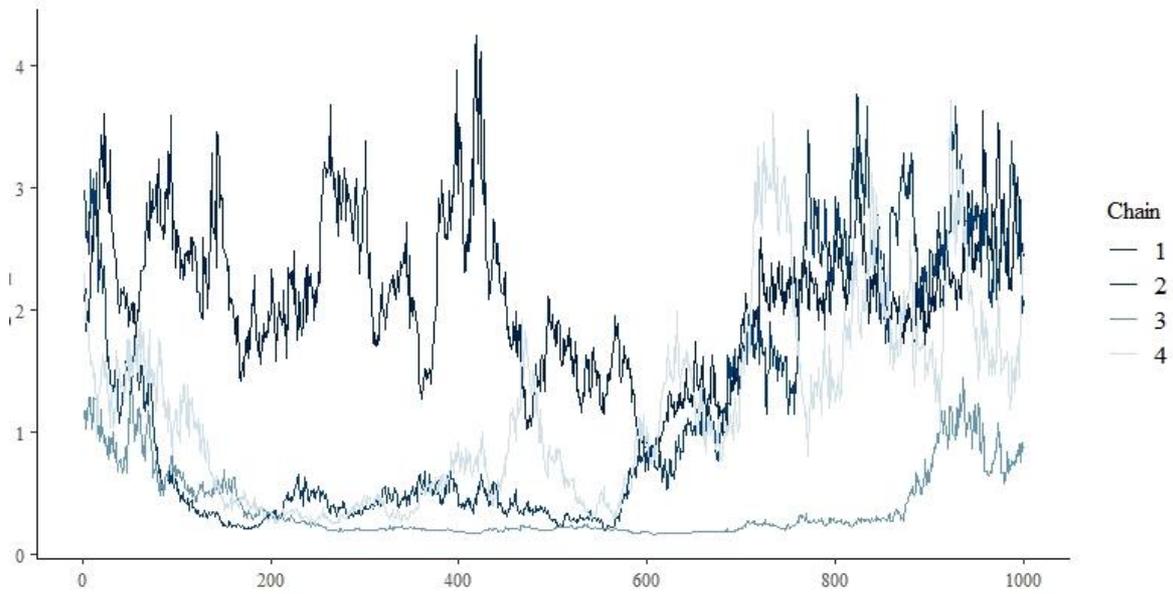


Figure 26. Exemple de chaînes MCMC qui ne convergent pas vers une distribution stationnaire commune.

Si les chaînes se stabilisent sur la même distribution stationnaire, elles sont bien mélangées et le graphique doit ressembler à cette image :

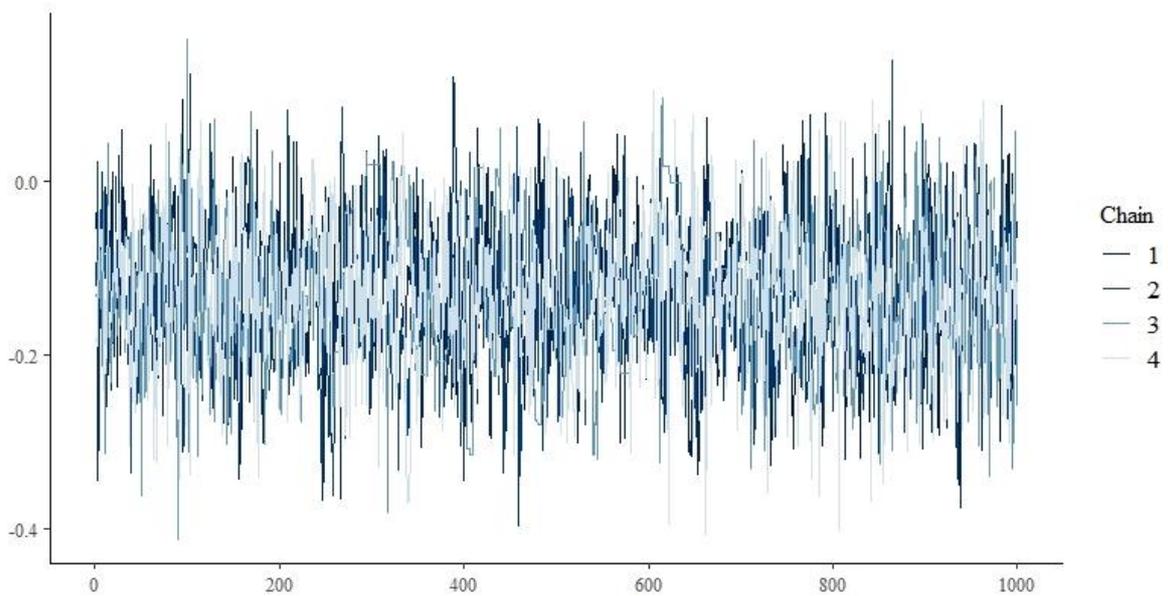


Figure 27. Exemple de chaînes MCMC ayant convergées sur la même distribution stationnaire.

Cette convergence peut aussi être vérifiée en comparant le comportement des chaînes par la statistique de réduction d'échelle \hat{R} (Gelman and Rubin 1992) qui est un rapport entre les variances intra-chaînes et inter-chaînes. Ce rapport est calculé pour chaque paramètre de θ comme suit :

$$\hat{R} = \sqrt{\frac{\frac{N-1}{N} \times W + \frac{1}{N} \times B}{W}} \quad (3.9)$$

Avec N le nombre d'échantillons constituant les chaînes MCMC, W la variance moyenne intra-chaîne et B la variance inter-chaîne. Dans le cas où toutes les chaînes sont à l'équilibre $\hat{R} = 1$. Si les chaînes n'ont pas convergées vers une distribution stationnaire π commune alors $\hat{R} > 1$. Les recommandations classiques sont d'obtenir un $\hat{R} \leq 1,1$ pour tous les paramètres estimés (Stan Development Team 2022b).

Rapport de la taille de l'échantillon

En notant N le nombre d'échantillons analysés par chaîne, l'ESS (« effective sample size ») estime le nombre d'échantillons indépendants qui seraient équivalents en terme d'incertitude au nombre N d'échantillons (potentiellement corrélés) analysés par chaîne. L'ESS, noté N_{eff} , est calculé pour chaque paramètre p de θ à partir de l'autocorrélation observée sur l'ensemble des lags t dans les chaînes.

Le rapport de la taille d'échantillon effective vaut N_{eff}/N . Lorsque le rapport est trop petit ($<0,1$ si on se réfère au manuel de Stan), la taille effective de l'échantillonnage est considérée trop faible pour réaliser une inférence avec une précision suffisante. La solution est souvent d'augmenter le nombre d'itérations, et plus particulièrement des itérations de chauffe (« warm-up ») pour s'assurer que la chaîne MCMC soit stabilisée sur sa chaîne stationnaire π .

Rapport de l'énergie cinétique

L'énergie de la fraction bayésienne de l'information manquante (BFMI « Bayesian Fraction of Missing Information ») est un indicateur de l'efficacité de l'algorithme pour explorer tout l'espace des paramètres. Les transitions de Markov peuvent adopter un comportement pathologique dans le cas d'une courbure importante de l'espace des paramètres. En effet, l'algorithme peut rester bloqué aux alentours de cette région de l'espace et ainsi difficilement explorer l'ensemble de l'espace (Betancourt 2017). Ce type de situation est une particularité des modèles hiérarchiques, ce qui inclut les modèles non linéaires mixtes (Betancourt and Girolami 2015).

Une solution peut être d'augmenter le nombre d'itérations de chauffe (« warm up ») ou de reparamétriser le modèle pour rechercher les valeurs θ^{**} qui appartiennent à une distribution dite non-centrée. Par exemple dans le cas d'un prior $p(\theta)$ gaussien on peut échantillonner une valeur de $\theta^{**} \sim \mathcal{N}(0,1)$ et la transformer par un paramètre de position μ et d'échelle σ sachant que $\theta^* = \mu + \theta^{**} \times \sigma$ plutôt qu'échantillonner les valeurs de $\theta^* \sim \mathcal{N}(\mu, \sigma)$.

3.4.4 Ajustement des données et comparaison de modèles

Le RMSE (« root mean squared error » ou racine de l'erreur quadratique moyenne) est un indicateur de l'ajustement du modèle aux données basé sur l'écart entre la valeur prédite \bar{Y}_i et l'observation Y_i . Le RMSE peut être indiqué pour différents niveaux : un indicateur global pour évaluer le modèle sur l'ensemble des observations ou bien par exemple un indicateur par marqueur dans le cas d'un modèle multivarié.

4 ASSOCIATION ENTRE L'HYPERTENSION ARTERIELLE, LA MALADIE DES PETITS VAISSEAUX, LES LESIONS DE TYPE ALZHEIMER, L'ATROPHIE DU CERVEAU ET LA COGNITION

4.1 Résumé

Nous avons vu en introduction qu'environ 35% des cas de démence pourraient être retardés ou évités en agissant sur les facteurs de risque modifiables, dont l'hypertension. Nous avons aussi vu que l'hypertension est associée au déclin des fonctions cognitives et à la maladie des petits vaisseaux du cerveau. Cependant, les mécanismes sous-jacents de l'association entre l'hypertension et la démence sont encore mal compris. Leur étude est importante pour identifier les cibles d'éventuelles stratégies de prévention de la démence.

Pour ces travaux nous avons réalisé une analyse transversale à partir des 2048 patients de la cohorte MEMENTO pour lesquels nous disposions de mesures de la pression artérielle. L'exposition à l'hypertension a été définie en 4 catégories par la combinaison de la pression artérielle élevée (PAS \geq 140mmHg et/ou PAD \geq 90mmHg) et de la prise d'un traitement antihypertenseur : 666 participants n'avaient pas d'hypertension, 379 avaient une pression artérielle élevée sans prise d'antihypertenseur, 530 avaient une pression artérielle contrôlée par une prise d'antihypertenseur et 473 avaient une pression artérielle élevée non contrôlée par la prise d'antihypertenseur.

Les chemins d'association entre l'hypertension, la maladie des petits vaisseaux, la neuropathologie de type Alzheimer, la mort des neurones et la cognition ont été examinés au moyen d'un modèle à équations structurelles intégrant un ensemble de biomarqueurs du LCR, de biomarqueurs de neuro-imagerie et de données cliniques à l'inclusion dans l'étude. Quatre dimensions ont été étudiées : (i) la fonction cognitive comme facteur commun des scores aux tests de mémoire, de fluence verbale et de la fonction exécutive ; (ii) la neurodégénérescence comme facteur commun de l'épaisseur corticale moyenne, du volume de l'hippocampe, de la fraction du parenchyme et du métabolisme du glucose à l'imagerie TEP-FDG ; (iii) la neuropathologie de type Alzheimer comme facteur commun des niveaux de protéines tau et amyloïdes dans le LCR et l'imagerie TEP amyloïde ; et (iv) la maladie des petits vaisseaux par le volume des HSB. Le modèle était ajusté sur l'âge, le sexe, le niveau d'étude, le génotype de

l'*APOE* et les principaux facteurs de risque cardio-métaboliques (antécédents cardiovasculaires, diabète, indice de masse corporelle). Enfin les données manquantes ont été prises en compte sous l'hypothèse d'un mécanisme MAR avec l'estimateur FIML.

Les résultats montraient que les participants avec une pression artérielle élevée malgré une prise en charge thérapeutique avaient une cognition significativement plus faible que les autres. Cette association était en partie expliquée par une neurodégénérescence plus importante et un volume plus important d'HSB (médiateurs de l'association) mais pas par les biomarqueurs de la neuropathologie de type d'Alzheimer. Les résultats ont aussi mis en évidence que les participants hypertendus avec une pression artérielle contrôlée par une prise en charge médicamenteuse ne présentaient pas de différence sur la fonction cognitive comparativement aux participants sans hypertension.

4.2 Article 1: « Associations among hypertension, dementia biomarkers, and cognition: The MEMENTO cohort »

FEATURED ARTICLE

Associations among hypertension, dementia biomarkers, and cognition: The MEMENTO cohort

Jérémy Lespinasse^{1,2} | Geneviève Chêne^{1,2} | Jean-Francois Mangin^{3,4} |
Bruno Dubois^{5,6} | Frederic Blanc⁷ | Claire Paquet⁸ | Olivier Hanon⁹ |
Vincent Planche¹⁰ | Audrey Gabelle¹¹ | Mathieu Ceccaldi¹² | Cedric Annweiler¹³ |
Pierre Krolak-Salmon¹⁴ | Olivier Godefroy¹⁵ | David Wallon¹⁶ | Mathilde Sauvée^{17,18} |
Sébastien Bergeret¹⁹ | Marie Chupin³ | Cécile Proust-Lima¹ | Carole Dufouil^{1,2} | for
the MEMENTO study group

¹Inserm Research Center « Bordeaux Population Health », Bordeaux School of Public Health, CIC 1401-EC, Bordeaux University, Bordeaux, France

²Pôle de santé publique, Centre Hospitalier Universitaire (CHU) de Bordeaux, Bordeaux, France

³CATI, US52-UAR2031, CEA, ICM, SU, CNRS, INSERM, APHP, Paris, France

⁴Université Paris-Saclay, CEA, CNRS, Neurospin, UMR9027 Baobab, Gif-sur-Yvette, France

⁵Sorbonne Université, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale, Paris, France

⁶Sorbonne-Université, Service des maladies cognitives et comportementales et Institut de la mémoire et de la maladie d'Alzheimer (IM2A), Hôpital de la Salpêtrière, Paris AP-PH, France

⁷Univ. Strasbourg, CNRS, ICube laboratory, UMR 7357, Fédération de Médecine Translationnelle de Strasbourg, Centre Mémoire de Ressources et de Recherches, Département de Gériatrie, Strasbourg, France

⁸Univ. Paris, Inserm U1144, GHU APHP Nord Lariboisière Fernand-Widal, Paris, France

⁹Univ. de Paris, EA 4468, Service de Gériatrie, AP-HP, Hôpital Broca, Paris, France

¹⁰Univ. Bordeaux, CNRS UMR 5293, Institut des Maladies Neurodégénératives, Centre Mémoire de Ressources et de Recherches, Pôle de Neurosciences Cliniques, CHU de Bordeaux, Bordeaux, France

¹¹Univ. Montpellier, i-site MUSE, Inserm U1061, Centre Mémoire de Ressources et de Recherches, Pôle de Neurosciences, Département de Neurologie, CHU de Montpellier, Montpellier, France

¹²Univ. Aix Marseille, Inserm UMR 1106, Institut de Neurosciences des Systèmes, Centre Mémoire de Ressources et de Recherches, Département de Neurologie et de Neuropsychologie, AP-HM, Marseille, France

¹³Univ. Angers, UPRES EA 4638, Centre Mémoire de Ressources et de Recherches, Département de Gériatrie, CHU d'Angers, Angers, France

¹⁴Univ. Lyon, Inserm U1028, CNRS UMR5292, Centre de Recherche en Neurosciences de Lyon, Centre Mémoire Ressource et Recherche de Lyon (CMRR), Hôpital des Charpennes, Hospices Civils de Lyon, Lyon, France

¹⁵Neurology Department and Functional Neurosciences Lab. (UR UPJV 4559), Amiens University Hospital, Amiens, France

¹⁶Normandie Univ, UNIROUEN, Inserm U1245, CHU Rouen, Department of Neurology and CNR-MAJ, Normandy Center for Genomic and Personalized Medicine, CIC-CRB1404, Rouen, France

¹⁷CMRR Grenoble Arc Alpin, CHU Grenoble, Grenoble, France

¹⁸Laboratoire de Psychologie et NeuroCognition: LPNC CNRS 5105, Université Grenoble Alpes, Grenoble, France

¹⁹Département de Médecine Nucléaire, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France

Cécile Proust-Lima and Carole Dufouil have contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

Correspondence

Carole Dufouil, Inserm research Center
"Bordeaux Population Health," Bordeaux
University, 146 rue Léo Saignat, 33076
Bordeaux, France.
E-mail: carole.dufouil@inserm.eu

Funding information

Fondation Plan Alzheimer, Grant/Award
Number: 2008-2012; French Ministry of
Research; French National Research Agency;
AVID; GE Healthcare; Fujirebio

Abstract

Introduction: Approximately 40% of dementia cases could be delayed or prevented acting on modifiable risk factors including hypertension. However, the mechanisms underlying the hypertension–dementia association are still poorly understood.

Methods: We conducted a cross-sectional analysis in 2048 patients from the MEMENTO cohort, a French multicenter clinic-based study of outpatients with either isolated cognitive complaints or mild cognitive impairment. Exposure to hypertension was defined as a combination of high blood pressure (BP) status and antihypertensive treatment intake. Pathway associations were examined through structural equation modeling integrating extensive collection of neuroimaging biomarkers and clinical data.

Results: Participants treated with high BP had significantly lower cognition compared to the others. This association was mediated by higher neurodegeneration and higher white matter hyperintensities load but not by Alzheimer's disease (AD) biomarkers.

Discussion: These results highlight the importance of controlling hypertension for prevention of cognitive decline and offer new insights on mechanisms underlying the hypertension–dementia association.

KEYWORDS

amyloid beta 42, cognition, cortical thickness, fluorodeoxyglucose positron emission tomography, hippocampal volume, hypertension, mediation, positron emission tomography amyloid, structural equation model, tau, white matter hyperintensities

Highlights

- Paths of hypertension–cognition association were assessed by structural equation models.
- The hypertension–cognition association is not mediated by Alzheimer's disease biomarkers.
- The hypertension–cognition association is mediated by neurodegeneration and leukoaraiosis.
- Lower cognition was limited to participants treated with uncontrolled blood pressure.
- Blood pressure control could contribute to promote healthier brain aging.

1 | INTRODUCTION

While the global population ages, the number of persons with dementia is increasing and could rise from 57.4 million globally in 2019 to 152.8 million in 2050.¹ Dementia is not inevitable as up to 40% of cases could be prevented or delayed by modifying 12 risk factors, including hypertension.^{2–4}

Hypertension is one of the most common conditions that degrade cerebral circulation, and prolonged high blood pressure (BP) is a cause of stroke and vascular dementia.^{5–8} Several studies suggest that high BP manifested in midlife may also contribute to an increased

risk of dementia due to Alzheimer's disease (AD) in late life.^{9–11} Two meta-analyses supported this result^{11,12} while two other studies concluded an inverse association between late-life hypertension and AD.^{13,14}

To better understand the association between hypertension and cognition and build successful intervention strategies, integrated pathways analyses are required to elucidate the mechanistic underlying processes. Imaging and cerebrospinal fluid (CSF) biomarkers have the potential to objectively measure normal biological or pathogenic processes in vivo, and allow taking into account simultaneously multifactorial cerebral processes that can mediate the association between

hypertension and cognition including white matter lesions (WMLs), neurodegeneration, or AD biomarkers.^{15–20}

We addressed this question in the MEMENTO cohort, a large clinic-based study in France, and sought to assess the mediating effect of dementia biomarkers, CSF and neuroimaging, on the association between hypertension and lower cognitive performance.

2 | METHODS

2.1 | Study population

The MEMENTO cohort is a clinic-based study of patients presenting with a large variety of cognitive symptoms or subjective cognitive complaints, who were enrolled between April 2011 and June 2014, within the French national network of memory clinics.²¹ Main inclusion criteria were a Clinical Dementia Rating (CDR) scale score ≤ 0.5 , mild cognitive impairment (< 1 standard deviation [SD] below the age, sex, and education-level thresholds in one or more cognitive test[s]), or isolated subjective cognitive complaint (for people older than 60). Exclusion criteria included history of head trauma with persistent neurological deficits, stroke in the last 3 months or with persistent neurological deficits, brain tumor, epilepsy, schizophrenia, known mutation in familial AD genes, and illiteracy. All examinations (including neuropsychological battery administration, clinical examinations, brain magnetic resonance imaging [MRI], CSF samples, and fluorodeoxyglucose [FDG] and amyloid positron emission tomography [PET]) followed standardized procedures. Among the 2323 participants included in the MEMENTO cohort, 2048 participants from 26 study centers were included in this analysis after exclusion of participants with all blood pressure measurements missing.

2.2 | Standard protocol approvals, registrations, and patient consents

This study was performed in accordance with the Declaration of Helsinki. All participants provided written informed consent. The MEMENTO cohort protocol has been approved by the local ethics committee ("Comité de Protection des Personnes Sud-Ouest et Outre Mer III"; approval number 2010-A01394-35) and was registered in ClinicalTrials.gov (Identifier: NCT01926249).

2.3 | Data collection

BP was measured three times in a seated position after 2 minutes of rest, and the mean of systolic (SBP) and diastolic (DPB) measures were calculated. Participants were classified as having high BP at baseline visit if the mean of SBP measures was ≥ 140 mm Hg or the mean of DBP measures was ≥ 90 mm Hg. Medications were recorded based on participants' reports, medical records, and prescriptions. Active substances were centrally coded using the Anatomical

RESEARCH IN CONTEXT

- 1. Systematic Review:** We reviewed literature using PubMed. Despite numerous studies and reviews on the link between hypertension and dementia risk, few studies have used integrative methods to study the association of hypertension simultaneously with brain imaging markers and cognition. To our knowledge, none has studied biomarkers of Alzheimer's disease (AD) pathology to investigate in a unique model the interrelationships among hypertension, biomarkers of AD pathology, biomarkers of neurodegeneration, and cognition in a large sample of patients recruited consecutively.
- 2. Interpretation:** Our results suggest that the impact of hypertension on cognition is mediated by neurodegeneration and white matter hyperintensities but not through AD pathology. Only participants with uncontrolled hypertension (treated by antihypertensives, high blood pressure) had lower cognition compared to the others. Our findings highlight the importance of controlling hypertension for prevention of cognitive decline.
- 3. Future Directions:** Future studies integrating longitudinal data will be necessary to confirm the causality of the associations observed and to consider the temporality of the different biomarkers.

Therapeutic Chemical (ATC) classification system. Participants were considered treated by antihypertensive drug if at least one of the five major drug classes intake was recorded: angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, beta-blockers, calcium channel blockers, and thiazide diuretics.

Four groups of exposure were defined as a combination of high BP status and antihypertensive drugs intake: (1) untreated, normal BP (reference category = no hypertension), (2) untreated, high BP, (3) treated, normal BP (controlled BP), (4) treated, high BP (uncontrolled BP).

2.4 | Neuropsychological evaluation

Memory was assessed by the total score at the three free recalls of the Free and Cued Selective Reminding Test (FCSRT),²² semantic verbal fluency via "animal" words (number cited in 120 seconds), and executive functions by Trail Making Test (TMT) Part B²³ (mean number of correct move per second).

2.5 | Dementia biomarkers assessment

Brain MRI was mandatory. Images were acquired after a standardization of the imaging processes and coordinated by the CATI ([15525279, 0, Downloaded from <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12866> by Cochrane France, Wiley Online Library on \[05/12/2022\]. See the Terms and Conditions \(<https://onlinelibrary.wiley.com/terms-and-conditions>\) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License](http://cati-</p>
</div>
<div data-bbox=)

neuroimaging.com), a neuroimaging platform dedicated to multicenter studies.²⁴ MRI scanners with 1.5 and 3 Tesla were used across centers. All MRI scans acquired were centralized, quality checked, and post-processed to obtain standardized measurements for each participant. Total intracranial volume, whole-brain, gray matter, and white matter volumes were assessed with Statistical Parametric Mapping 12,²⁵ hippocampal volumes with the SACHA software,²⁶ and mean cortical thickness of each hemisphere with FreeSurfer 5.3 averaged in the regions of interest (ROIs) of the Desikan–Killiany atlas.²⁷ WML volumes were estimated using WHASA software.²⁸ Brain parenchymal fraction was computed as the sum of gray matter and white matter volumes divided by total intracranial volume. Total hippocampal volume was computed as the sum of left and right volumes.

2.6 | [18F] FDG-PET

18F-FDG-PET was optional and available for 1187 individuals of the analytical sample. PET images were acquired after a standardization of the acquisition and reconstruction imaging parameters, coordinated by the CATI.²⁹ After a centralized quality check and postprocessing, the following measures were obtained: Mean FDG-PET uptake for the ROIs of the Automated Anatomical Labeling atlas relative to the pons reference region,³⁰ including partial volume correction, and mean FDG-PET uptake for a set of AD-specific ROIs inferred from the Alzheimer's Disease Neuroimaging Initiative database,³¹ expressed as standardized uptake value ratios (SUVRs).

2.7 | PET amyloid imaging

PET amyloid imaging was available for 577 participants of the analytical sample, using either 18F-florbetapir (Amyvid, Eli Lilly; $N = 396$) or 18F-flutemetamol (Vizamyl, GE Healthcare; $N = 181$) radioligands. Mean brain amyloid SUVR was standardized (z-score) by radioligand.

2.8 | CSF sampling

Lumbar puncture was proposed to all participants and available for 304 participants of the analytical sample. Centralized measurements of amyloid beta ($A\beta$) 42 peptide ($A\beta_{42}$), $A\beta$ 40 peptide ($A\beta_{40}$), and phosphorylated tau (p-tau) levels were performed using the standardized INNOTEST sandwich enzyme-linked immunosorbent assay (Fujirebio).

2.9 | Confounding factors

Sociodemographic information recorded at baseline included age, sex, and education (baccalaureate and above vs. less). Diabetes was defined as the presence of fasting blood glucose ≥ 7 mmol/L (≥ 126 mg/dl) or non-fasting blood glucose ≥ 11.1 mmol/L (≥ 200 mg/dl) or antidiabetic

drug intake or self-reported history of diabetes. Body mass index (BMI) was computed from measured height and weight. History of cardiovascular disease was defined as a self-reported history of myocardial infarction, angina pectoris, coronary artery, or peripheral artery disease. Apolipoprotein E (APOE) $\epsilon 2$, $\epsilon 3$, or $\epsilon 4$ alleles were determined by KBiosciences (www.kbioscience.co.uk; now Biosearch Technologies). APOE- $\epsilon 4$ status was defined as presence of at least one $\epsilon 4$ allele.

2.10 | Statistical analyses

Baseline characteristics were compared according to hypertension groups as previously defined using chi-square test and analysis of variance for categorical and continuous variables comparisons, respectively.

A structural equation model (SEM) was used to examine a potential mediating role of dementia CSF and neuroimaging biomarkers in the association between hypertension status and cognition (Figure 1). Four dimensions were considered: white matter hyperintensity (WMH) volume and three latent variables defined as the common factor of their manifest variables:

- AD pathology measured by CSF $A\beta_{42}/A\beta_{40}$ ratio and p-tau, and amyloid PET SUVR;
- Neurodegeneration measured by mean cortical thickness, hippocampal volume, brain parenchymal fraction, and FDG PET SUVR;
- Cognition measured by the verbal fluency, the total free reminding, and the TMT B scores;

For ease of interpretation, the four dimensions were standardized (mean 0, variance 1) so that one unit corresponds to the SD of a given dimension. The indirect effects of each hypertension category (vs. reference = "untreated, normal blood pressure") on cognition through the dimensions were estimated with their 95% confidence interval (95% CI), using a path analysis technique. All linear regressions of mediators and cognition were adjusted for the following potential confounders factors: age, sex, education, diabetes, BMI, cardiovascular history, and APOE $\epsilon 4$ status. Missing values for observed indicators of latent variables and for confounding factors (see description in Table S1 in supporting information) were handled using a full information maximum likelihood (FIML) approach, robust to missingness at random.³² The main model, assessed by fit indices robust to non-normal data³³ showed a satisfying fit (comparative fit index [CFI] = 0.92, Tucker–Lewis index [TLI] = 0.88, root mean square error of approximation [RMSEA] = 0.042 [90% confidence interval = 0.038–0.046], and standardized root mean square residual [SRMR] = 0.039). Sensitivity analyses explored potential differences in the findings by sex, age (dichotomized according to median age), and APOE status by adding interaction terms with the exposure of interest.

Analyses were conducted using R version 4.0.2 and the lavaan package version 0.6.9 for SEM analysis.³⁴

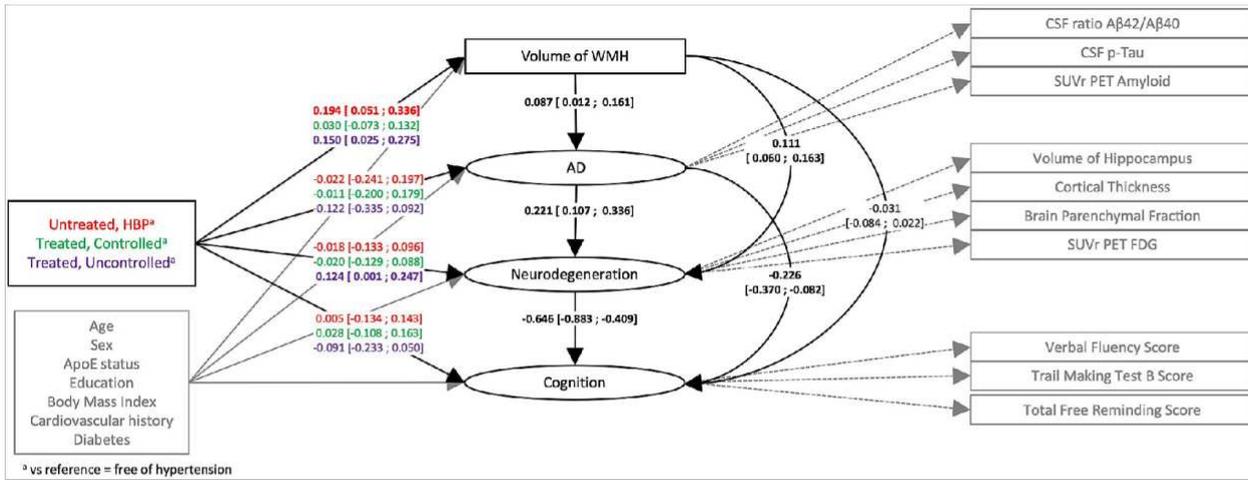


FIGURE 1 Hypothetical structural equation model and estimates of the associations described: The MEMENTO cohort, 2011–2014. Aβ42/Aβ40, ratio of amyloid beta 42 on amyloid beta 40 proteins; AD, Alzheimer’s disease; APOE, apolipoprotein E; CSF, cerebrospinal fluid; FDG, fluorodeoxyglucose; HBP, high blood pressure; PET, positron emission tomography; p-tau, phosphorylated tau; SUVR, standardized uptake value ratio; WMH, white matter hyperintensities

2.11 | Data availability statement

Anonymized data will be shared by request from any qualified investigator for the sole purpose of replicating procedures and results presented in the article, and as long as data transfer is in agreement with EU legislation on the general data protection regulation.

3 | RESULTS

Mean age (SD) of participants was 71.3 (8.5) years, 60% were female, and 40% had education level of baccalaureate and above (Table 1). Sixty percent of participants had mild cognitive impairment (CDR scale score = 0.5) versus isolated cognitive complaints (CDR score = 0). The mean CDR Sum of Boxes was 0.61 (SD = 0.72). Approximately two thirds of participants (67.5%) had hypertension and 23.1% were treated for hypertension and uncontrolled. There were statistically significant differences in participant characteristics according to hypertension status (Table 1). Compared to participants untreated with normal blood pressure, those “treated, uncontrolled” tended to have the more pejorative profile: they were 6.7 years older on average, they were less educated, and had more often cardiovascular history or diabetes. They also had lower cognitive performances on average and worse brain FDG-PET and MRI markers. Compared to participants “untreated, normal BP,” those “treated, uncontrolled” had worse brain structural and cognitive profile, a finding observed to a lower extent in “treated, controlled” and “untreated, high BP” groups. AD biomarkers (measured through CSF or amyloid PET) distribution did not vary by hypertension status.

The pathway analysis performed using SEM is summarized in Figure 1, and estimates of direct and total associations of hypertension status with AD pathology, WML volume, neurodegeneration, and

cognition are further presented in Table 2. See Table S2 in supporting information for the direct and total estimated associations of adjustment covariates on dementia biomarkers and cognition; Table 3 for those of WMH, AD pathology, and neurodegeneration; and Table S3 in supporting information for factor loadings and percentage of residual variance of the observed variables for each latent dimension. AD pathology captured 67.7%, 49.4%, and 74.1% of the total variance of, respectively, CSF Aβ42/Aβ40 ratio, CSF p-tau, and amyloid PET SUVR. Neurodegeneration captured 25.8%, 45.7%, 55.5%, and 37.4% of the total variance of, respectively, the mean cortical thickness, hippocampal volume, brain parenchymal fraction, and FDG PET SUVR. Cognition captured 51%, 40.4%, and 37.1% of the total variance of, respectively, the free Reminding Test, Verbal Fluency, and TMT B.

Adjusted for potential confounders and hypertension, higher levels of neurodegeneration and AD pathology were independently and directly associated with lower cognition (mean difference [MD] = -0.646 SD, 95% CI = [-0.883; -0.409], MD = -0.226 SD, 95% CI = [-0.370; -0.082], respectively). In contrast, WMH was associated with cognition through its effect on AD pathology (MD = 0.087 SD, 95% CI = [0.012; 0.161]) and neurodegeneration (MD = 0.111 SD, 95% CI = [0.060; 0.163]). Adjusted on WMH volume, AD pathology was also indirectly associated with cognition through its association with neurodegeneration (MD = 0.221 SD, 95% CI = [0.107; 0.336]).

The whole pathway of association did not differ in the “treated, controlled” group compared to the reference “untreated, normal BP” group (free of hypertension). Compared to the reference group, “untreated, high BP” or “treated, uncontrolled” groups were significantly associated with higher WML load (MD = 0.194 SD, 95% CI = [0.051; 0.336] and MD = 0.150 SD, 95% CI = [0.025; 0.275], respectively). The “treated, uncontrolled” group was also significantly directly associated with higher neurodegeneration than in the refer-

TABLE 1 Study sample characteristics globally and by hypertension management status: The MEMENTO cohort, 2011–2014

	Full sample	Sample stratified by antihypertensive treatment high blood pressure status				P-value
		Untreated, normal blood pressure	Untreated, high blood pressure	Treated, controlled	Treated, uncontrolled	
DEMOGRAPHIC AND CLINICAL CHARACTERISTICS	N = 2048	N = 666	N = 379	N = 530	N = 473	
Age, mean (SD), years	71.3 (8.5)	68.0 (9.1)	72.1 (7.7)	71.7 (7.7)	74.7 (7.4)	<0.001
Male sex, no. (%)	777 (40%)	213 (32.0%)	142 (37.5%)	224 (42.3%)	198 (41.9%)	<0.001
At least one APOE ε4 carried, no. (%)	571 (30%)	186 (29.4%)	113 (31.7%)	146 (29.1%)	126 (27.7%)	0.675
Baccalaureate or higher, no. (%)	803 (40%)	303 (45.6%)	154 (40.6%)	188 (35.5%)	158 (33.4%)	<0.001
BMI, mean (SD), kg/m ²	25.6 (4.3)	24.3 (4.0)	25.5 (4.0)	26.6 (4.5)	26.6 (4.3)	<0.001
Cardiovascular history, no. (%)	303 (10%)	28 (4.2%)	31 (8.2%)	125 (23.6%)	119 (25.2%)	<0.001
of which Myocardial infarction, no. (%)	68 (3.3%)	2 (0.3%)	6 (1.6%)	31 (5.8%)	29 (6.1%)	<0.001
of which Stroke, no. (%)	85 (4.2%)	11 (1.7%)	8 (2.1%)	33 (6.2%)	33 (7.0%)	<0.001
Diabetes, no. (%)	226 (10%)	38 (5.7%)	24 (6.3%)	85 (16.0%)	79 (16.7%)	<0.001
Dyslipidemia, no. (%)	1090 (53.2%)	317 (47.6%)	197 (52.0%)	304 (57.4%)	272 (57.5%)	<0.001
Glomerular Filtration Rate (mL/min)	78.8 (17.9)	85.5 (16.8)	78.8 (16.4)	77.1 (19.5)	75.5 (17.8)	<0.001
Averaged SBP, mean (SD), mmHg	136.9 (18.5)	123.0 (10.8)	152.1 (11.6)	126.3 (9.7)	156.1 (10.8)	<0.001
Averaged DBP, mean (SD), mmHg	76.8 (10.2)	72.8 (7.9)	83.6 (8.3)	72.6 (8.3)	81.9 (10.1)	<0.001
NEUROPSYCHOLOGICAL TESTING RESULTS	N = 2032	N = 665	N = 374	N = 525	N = 468	
Clinical Dementia Rating = 0.5, no. (%)	1209 (60%)	381 (57.5%)	209 (55.3%)	326 (61.7%)	293 (62.1%)	0.100
CDR Sum of Boxes, mean (SD), score	0.61 (0.72)	0.54 (0.60)	0.57 (0.72)	0.64 (0.70)	0.72 (0.86)	<0.001
Free Reminding Test, mean (SD), tot. score	25.9 (8.4)	27.3 (8.1)	26.2 (8.3)	25.8 (8.4)	23.8 (8.4)	<0.001
Trail Making Test B, mean (SD), good move/s	0.25 (0.12)	0.27 (0.12)	0.25 (0.12)	0.26 (0.13)	0.23 (0.11)	<0.001
Verbal Fluency, mean (SD), words	28.3 (8.7)	29.4 (8.8)	28.2 (8.3)	28.2 (8.9)	26.8 (8.4)	<0.001
Brain Imaging markers	N = 1938	N = 628	N = 354	N = 505	N = 451	
Cortical thickness, mean (SD), mm	2.60 (0.15)	2.62 (0.15)	2.60 (0.15)	2.60 (0.14)	2.58 (0.15)	<0.001
Volume of hippocampus, mean (SD), cm ²	5.40 (0.77)	5.52 (0.77)	5.38 (0.71)	5.44 (0.79)	5.19 (0.77)	<0.001
Brain parenchymal fraction, mean (SD), %	71.18 (6.65)	73.29 (6.40)	70.95 (5.52)	70.80 (6.60)	68.86 (6.98)	<0.001
WMH volume, mean (SD), log(mm ³)	1.73 (1.11)	1.47 (1.08)	1.82 (1.16)	1.74 (1.09)	2.02 (1.06)	<0.001
FDG PET MARKERS	N = 1187	N = 382	N = 226	N = 309	N = 270	
SUVr, mean (SD), ratio	1.73 (0.19)	1.76 (0.19)	1.73 (0.19)	1.72 (0.19)	1.68 (0.19)	<0.001
CEREBROSPINAL FLUID BIOMARKERS	N = 304	N = 105	N = 58	N = 74	N = 67	
CSF Aβ42/Aβ40, mean (SD), ratio	0.08 (0.04)	0.09 (0.04)	0.08 (0.04)	0.08 (0.04)	0.08 (0.03)	0.145
CSF p-tau, mean (SD), pg/ml	62.79 (29.49)	61.48 (32.54)	62.75 (28.41)	62.13 (28.34)	65.63 (30.86)	0.835
AMYLOID PET MARKERS	N = 577	N = 178	N = 124	N = 143	N = 132	
SUVr, mean (SD), ratio	0.020 (1.020)	0.007 (1.040)	0.128 (1.080)	0.005 (0.966)	-0.046 (1.000)	0.576

Note: Statistical tests used are chi-square and analysis of variance for categorical and continuous variables, respectively. Abbreviations: SD, standard deviation; APOE, apolipoprotein E; BMI, body mass index; WMH, white matter hyperintensities; PET, positron emission tomography; FDG, fluorodeoxyglucose; SUVr, standardized uptake value ratio; CSF, cerebrospinal fluid; Aβ, amyloid beta; p-tau, phosphorylated tau; SBP, systolic blood pressure; DBP, diastolic blood pressure.

TABLE 2 Estimates of direct and total effects of hypertension status on dementia biomarkers and cognition computed from structural equation model: The MEMENTO cohort, 2011–2014

Antihypertensive treatment High blood pressure status	Direct effects ^a				Total effects ^a			
	Pathway variable	β , SD	[95% CI]	P-value	β , SD	[95% CI]	P-value	
Untreated, high blood pressure	Volume of WMH	0.194	[0.051; 0.336]	0.008				
	AD Biomarkers	-0.022	[-0.241; 0.197]	0.842	-0.005	[-0.223; 0.212]	0.961	
	Neurodegeneration	-0.018	[-0.133; 0.096]	0.753	0.002	[-0.112; 0.116]	0.973	
Treated, controlled	Cognition	0.005	[-0.134; 0.143]	0.949	-0.001	[-0.137; 0.134]	0.983	
	Volume of WMH	0.030	[-0.073; 0.132]	0.572				
	AD biomarkers	-0.011	[-0.200; 0.179]	0.912	-0.008	[-0.198; 0.182]	0.934	
Treated, uncontrolled	Neurodegeneration	-0.020	[-0.129; 0.088]	0.711	-0.019	[-0.124; 0.086]	0.725	
	Cognition	0.028	[-0.108; 0.163]	0.689	0.041	[-0.091; 0.173]	0.545	
	Volume of WMH	0.150	[0.025; 0.275]	0.018				
Treated, uncontrolled	AD biomarkers	-0.122	[-0.335; 0.092]	0.264	-0.109	[-0.322; 0.104]	0.318	
	Neurodegeneration	0.124	[0.001; 0.247]	0.048	0.117	[-0.002; 0.235]	0.053	
	Cognition	-0.091	[-0.233; 0.050]	0.206	-0.147	[-0.283; -0.010]	0.035	

Note: Estimates (β and 95% Confidence Interval) are reported for one standard deviation increase of the pathway variable. They correspond to the mean difference compared to participants free of hypertension. Abbreviations: AD, Alzheimer's disease; SD, standard deviation; WMH, white matter hyperintensities.

^aDirect effect is the coefficient in the direct regression between the explanatory variable and the outcome. Total effect is the combination of the coefficients involved in all the existing paths between the explanatory variable and the outcome (Figure 1).

TABLE 3 Estimates of direct and total effects of observed and latent variables in the structural part of the model: The MEMENTO cohort, 2011–2014

From	To	Direct effects ^a			Total effects ^a		
		β , SD	95% CI	P-value	β , SD	95% CI	P-value
Volume of WMH	AD neuropathology	0.087	[0.012; 0.161]	0.023			
	Neurodegeneration	0.111	[0.060; 0.163]	<0.001	0.131	[0.085; 0.176]	<0.001
	Cognition	-0.031	[-0.084; 0.022]	0.251	-0.135	[-0.179; -0.091]	<0.001
AD pathology	Neurodegeneration	0.221	[0.107; 0.336]	<0.001			
	Cognition	-0.226	[-0.370; -0.082]	0.002	-0.369	[-0.492; -0.246]	<0.001
Neurodegeneration	Cognition	-0.646	[-0.883; -0.409]	<0.001			

Note: Estimates (β and 95% confidence interval) are reported for one standard deviation increase of the pathway variable. They correspond to the mean effect for one standard deviation increase of the explanatory variable.

Abbreviations: AD, Alzheimer's disease; SD, standard deviation; WMH, white matter hyperintensities.

^aDirect effect is the coefficient in the direct regression between the explanatory variable and the outcome. Total effect is the combination of the coefficients involved in all the existing paths between the explanatory variable and the outcome (Figure 1).

ence group (MD = 0.124 SD, 95% CI = [0.001; 0.247]). No hypertension status was significantly associated with AD pathology.

The "treated, uncontrolled" group was the only group globally associated with lower cognition than in the reference group (MD = -0.147 SD, 95% CI = [-0.283; -0.010]). This association was mainly carried out by higher neurodegeneration: once adjusted for neurodegeneration, AD pathology and WML volumes, there was no direct effect of any hypertension exposure on cognition (MD = -0.091 SD, 95% CI = [-0.233; 0.050], $P = 0.21$ for "treated, uncontrolled" vs. reference group). Sensitivity analyses did not suggest pathway modification according to sex and APOE (Table S4, Table S5 in supporting information; overall likelihood ratio tests for interactions $P = 0.545$ and $P = 0.223$, respectively). Few modulations were observed though for the "treated, uncontrolled" (Table S6 in supporting information) group according to age. The greater volume of WMH for "treated, uncontrolled" group compared to the reference was carried by participants aged above 71.9 years old (MD = 0.273, 95% CI = [0.107; 0.439] vs. MD = -0.022 95% CI = [-0.143; 0.099] for participants aged 71.9 years old and less).

4 | DISCUSSION

In a cross-sectional analysis of a large clinical cohort including participants with either isolated cognitive complaints or mild cognitive impairment, we report that the deleterious effect of uncontrolled hypertension on cognitive performance is mainly mediated through markers of neurodegeneration and WML load whereas AD biomarkers (amyloid, p-tau) do not seem to have a substantial influence. Interestingly after adjusting on both confounding factors and dementia biomarkers, cognition of participants "untreated, high BP" or "treated, controlled" did not differ from cognition of the reference group (no hypertension). We adjusted our model for overweight (BMI) and diabetes, which are often comorbidities with hypertension. Further adjusted for hypercholesterolemia did not change our results (data not shown).

The "untreated, high BP" group is likely to gather heterogeneous profiles, that is, some participants with recent high BP (and therefore untreated), others with high BP due to white coat effect and this could explain that it is not associated with lower cognition and greater neurodegeneration. It has already been reported that "untreated, high BP" individuals have higher WML volume than "untreated, normal BP" individuals and it is also possible that the impact on cognition could be seen later in time.^{17,35}

The observation that neither cognition, nor dementia biomarkers, differed between "treated, normal bp" and "untreated, normal BP" groups is in favor of effective treatment and control of hypertension to be protective for cognition.^{36–38} These results indicate that identifying people with uncontrolled hypertension and intervening against hypertension might be a key to prevent the occurrence of dementia.

Several neuroimaging studies have consistently reported effects of hypertension on brain structure and function despite large heterogeneity. Some were restricted to small and selected samples.^{39–41} Various definitions of hypertension were considered (either binary

[HBP or treated vs. no hypertension] or continuous SBP/DBP measures, or by taking into account antihypertensive treatment status and blood pressure control). Age at hypertension assessment varied from mid-life⁴² to late life. In the Risk Development in Young Adults (CARDIA) Study, white matter integrity (fractional anisotropy [FA]) was investigated in relation to hypertension control in 698 community-dwelling adults (mean age 50 years). They reported no difference in mean FA for participants who were hypertensive but not taking antihypertensive medication, and participants whose hypertension was controlled at normotensive levels through antihypertensives compared to participants who were normotensive. In contrast, mean FA was lower in subjects on antihypertensive drugs but whose blood pressure remained high (i.e., uncontrolled). These results are in line with ours, pointing out that uncontrolled hypertension, even in middle-aged individuals, has an impact on brain structure. However in this report the mediating effect of FA on the hypertension/cognition association was not reported.

The Systolic Blood Pressure Intervention Trial (SPRINT) tested the effect of intensive SBP control (SBP target < 120 mm Hg) versus a standard SBP treatment goal (SBP < 140 mm Hg) total in 9361 randomized participants aged 50 years or older, with SBP 130 to 180 mm Hg at screening visit and increased cardiovascular risk.⁴³⁻⁴⁵ Cognitive results from secondary analysis of the trial indicated a lower rate of mild cognitive impairment with intensive SBP control, with an inconclusive effect on probable dementia.⁴⁴ Our findings are consistent with this result as insufficient BP control may be associated with mild cognitive impairment.

Our study is the first to investigate the mediating role of multimodal biomarkers of dementia on the association between hypertension control and cognition. We did not show that AD biomarkers mediated this association. These findings indicate that uncontrolled hypertension may impact brain structure regardless of AD neuropathological changes or before facilitating its development.⁴⁶

Another hypothesis is that high BP may induce a chronic neuroinflammation reaction of the brain, which subsequently contributes to brain structural atrophy by releasing neurotoxic immune mediators.⁴⁷ Accumulating evidence has suggested that chronic neuroinflammation is increasingly emerging as an important pathological factor in the development and progression of AD. According to this theory, it is apparent that abnormal A β deposition can activate microglia and astrocytes, trigger an innate immune response, and subsequently release inflammatory mediators, which contribute to disease development and progression.⁴⁸ Thus chronic inflammation could be a pathophysiological mechanism that supports our findings about association among inadequate control of blood pressure, neurodegeneration, and lower cognition.

There were also some limitations in our study. First, for some biomarkers, particularly AD biomarkers, the rate of missing data was high (85% for CSF biomarkers, 71% for amyloid PET biomarkers). We were, however, able to perform the analysis on the total sample using the full information maximum likelihood estimation (FIML) under the realistic missing at random (MAR) hypothesis, which states that the probability of missing data does not depend on unob-

served data.³² It seemed indeed plausible that systematic differences between observed and missing values for AD biomarkers can be entirely explained by other observed variables (three markers of cognition, four markers for neurodegeneration and covariates including age, sex, APOE, education, and cardiovascular risk factor). Also, as MEMENTO cohort recruitment is clinic based, the observed findings may not fully translate in the general older population. On the other hand, undertaking a population-based study with such clinical and subclinical investigations would be challenging, and would also lead to a selection in participation. Selection occurs in both clinical and population-based cohorts and in MEMENTO we have undertaken to minimize it. Finally, by considering a cross-sectional study, the temporal relationship among hypertension status, dementia biomarkers, and cognition were not taken into account, and causal interpretations require cautiousness. The findings thus warrant further exploration in longitudinal studies.

The study has also strengths. A wide range of biomarkers was acquired in a highly standardized setting on >2000 participants allowing a multi-dimensional assessment of brain aging and pathology biomarkers. Using SEM, all this information was simultaneously leveraged in a mediation analysis of the hypertension-cognition association, offering a unique insight on underlying mechanisms.

To conclude, we showed that uncontrolled hypertension is associated with worse brain structure and function. It is possible that these individuals (a quarter of our sample) may not be adhering to their treatment, may not be receiving adequate treatment, or may be phenotypically disposed to high blood pressure. Strategies aiming at optimizing BP control in these individuals could contribute to promoting a healthier brain and dementia prevention.

ACKNOWLEDGMENTS

The MEMENTO cohort is funded by the Fondation Plan Alzheimer (Alzheimer Plan 2008-2012), and the French Ministry of Research (MESRI, DGRI) through the Plan Maladies Neurodégénératives (2014-2019). This work was also supported by CIC 1401-EC, Bordeaux University Hospital (CHU Bordeaux, sponsor of the cohort), Inserm, and the University of Bordeaux. This work received funding from the French National Research Agency (ANR) as part of the Investment for the Future Programme ANR-18-RHUS-0002. The MEMENTO cohort has received funding support from AVID, GE Healthcare, and FUJIREBIO through private-public partnerships. The Insight-PreAD substudy was promoted by INSERM in collaboration with the Institut du Cerveau et de la Moelle Epinière, Institut Hospitalo-Universitaire, and Pfizer and has received support within the "Investissement d'Avenir" (ANR-10-AIHU-06) program. The funders had no role in study design; in data collection, analysis, and interpretation; or in writing of report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

CONFLICTS OF INTEREST

The authors report no conflicts of interest. Author disclosures are available in the supporting information.

REFERENCES

- Collaborators GBDDF. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Heal.* 2022;396(10248):413-446. [https://doi.org/10.1016/S2468-2667\(21\)00249-8](https://doi.org/10.1016/S2468-2667(21)00249-8)
- Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet.* 2020; 396(10248): 413-446. [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6)
- Yu JT, Xu W, Tan CC, et al. Evidence-based prevention of Alzheimer's disease: systematic review and meta-analysis of 243 observational prospective studies and 153 randomised controlled trials. *J Neurol Neurosurg Psychiatry.* 2020; 91(11): 1201-1209. <https://doi.org/10.1136/jnnp-2019-321913>
- Takeda S, Rakugi H, Morishita R. Roles of vascular risk factors in the pathogenesis of dementia. *Hypertens Res.* 2020; 43(3): 162-167. <https://doi.org/10.1038/s41440-019-0357-9>
- Iturria-Medina Y, Sotero RC, Toussaint PJ, Mateos-Perez JM, Evans AC. Alzheimer's Disease Neuroimaging I. Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nat Commun.* 2016; 7: 11934. <https://doi.org/10.1038/ncomms11934>
- Emdin CA, Rothwell PM, Salimi-Khorshidi G, et al. Blood Pressure and risk of vascular dementia: evidence from a primary care registry and a cohort study of transient ischemic attack and stroke. *Stroke.* 2016; 47(6): 1429-1435. <https://doi.org/10.1161/STROKEAHA.116.012658>
- Burke GL, Hughes TM. Arterial changes connecting hypertension to Alzheimer's disease and related dementias. *JACC Cardiovasc Imaging.* 2021; 14(1): 186-188. <https://doi.org/10.1016/j.jcmg.2020.10.018>
- Collaborators GBDS. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019;18(5):439-458. [https://doi.org/10.1016/S1474-4422\(19\)30034-1](https://doi.org/10.1016/S1474-4422(19)30034-1)
- Launer LJ, Hughes T, Yu B, et al. Lowering midlife levels of systolic blood pressure as a public health strategy to reduce late-life dementia: perspective from the Honolulu Heart Program/Honolulu Asia Aging Study. *Hypertension.* 2010; 55(6): 1352-1359. <https://doi.org/10.1161/HYPERTENSIONAHA.109.147389>
- Launer LJ, Ross GW, Petrovitch H, et al. Midlife blood pressure and dementia: the Honolulu-Asia aging study. *Neurobiol Aging.* 2000; 21(1): 49-55. [https://doi.org/10.1016/s0197-4580\(00\)00096-8](https://doi.org/10.1016/s0197-4580(00)00096-8)
- Whitmer RA, Sidney S, Selby J, Johnston SC, Yaffe K. Midlife cardiovascular risk factors and risk of dementia in late life. *Neurology.* 2005; 64(2): 277-281. <https://doi.org/10.1212/01.WNL.0000149519.47454.F2>
- Meng XF, Yu JT, Wang HF, et al. Midlife vascular risk factors and the risk of Alzheimer's disease: a systematic review and meta-analysis. *J Alzheimers Dis.* 2014; 42(4): 1295-1310. <https://doi.org/10.3233/JAD-140954>
- Power MC, Weuve J, Gagne JJ, McQueen MB, Viswanathan A, Blacker D. The association between blood pressure and incident Alzheimer disease: a systematic review and meta-analysis. *Epidemiology.* 2011; 22(5): 646-659. <https://doi.org/10.1097/EDE.0b013e31822708b5>
- Proitsi P. Disentangling the complex relationship between hypertension and dementia. *Biol Psychiatry.* 2021; 89(8): 742-744. <https://doi.org/10.1016/j.biopsych.2021.02.005>
- de Leeuw FE, de Groot JC, Oudkerk M, et al. Hypertension and cerebral white matter lesions in a prospective cohort study. *Brain.* 2002; 125: 765-772. <https://doi.org/10.1093/brain/awf077>
- Dufouil C, Chalmers J, Coskun O, et al. Effects of blood pressure lowering on cerebral white matter hyperintensities in patients with stroke: the PROGRESS (Perindopril Protection Against Recurrent Stroke Study) Magnetic Resonance Imaging Substudy. *Circulation.* 2005; 112(11): 1644-1650. <https://doi.org/10.1161/CIRCULATIONAHA.104.501163>
- Godin O, Tzourio C, Maillard P, Mazoyer B, Dufouil C. Antihypertensive treatment and change in blood pressure are associated with the progression of white matter lesion volumes: the Three-City (3C)-Dijon Magnetic Resonance Imaging Study. *Circulation.* 2011; 123(3): 266-273. <https://doi.org/10.1161/CIRCULATIONAHA.110.961052>
- Jochimsen HM, Muller M, Visseren FL, et al. Blood pressure and progression of brain atrophy: the SMART-MR Study. *JAMA Neurol.* 2013; 70(8): 1046-1053. <https://doi.org/10.1001/jamaneurol.2013.217>
- Jeon SY, Byun MS, Yi D, et al. Influence of hypertension on brain amyloid deposition and Alzheimer's disease signature neurodegeneration. *Neurobiol Aging.* 2019; 75: 62-70. <https://doi.org/10.1016/j.neurobiolaging.2018.11.001>
- Godin O, Maillard P, Crivello F, et al. Association of white-matter lesions with brain atrophy markers: the three-city Dijon MRI study. *Cerebrovasc Dis.* 2009; 28(2): 177-184. <https://doi.org/10.1159/000226117>
- Dufouil C, Dubois B, Vellas B, et al. Cognitive and imaging markers in non-demented subjects attending a memory clinic: study design and baseline findings of the MEMENTO cohort. *Alzheimers Res Ther.* 2017; 9(1): 67. <https://doi.org/10.1186/s13195-017-0288-0>
- Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. *Neurology.* 1988; 38(6): 900-903. <https://doi.org/10.1212/wnl.38.6.900>
- Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol.* 2004; 19(2): 203-214. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8)
- Operto G, Chupin M, Batrancourt B, et al. CATI: a large distributed infrastructure for the neuroimaging of cohorts. *Neuroinformatics.* 2016; 14(3): 253-264. <https://doi.org/10.1007/s12021-016-9295-8>
- Ashburner J, Friston KJ. Unified segmentation. *Neuroimage.* 2005; 26(3): 839-851. <https://doi.org/10.1016/j.neuroimage.2005.02.018>
- Chupin M, Hammers A, Liu RS, et al. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *Neuroimage.* 2009; 46(3): 749-761. <https://doi.org/10.1016/j.neuroimage.2009.02.013>
- Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage.* 2006; 31(3): 968-980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- Samaille T, Fillon L, Cuingnet R, et al. Contrast-based fully automatic segmentation of white matter hyperintensities: method and validation. *PLoS One.* 2012; 7(11): e48953. <https://doi.org/10.1371/journal.pone.0048953>
- Habert MO, Marie S, Bertin H, et al. Optimization of brain PET imaging for a multicentre trial: the French CATI experience. *EJNMMI Phys.* 2016; 3(1). <https://doi.org/10.1186/s40658-016-0141-8>
- Buchert R, Wilke F, Chakrabarti B, et al. Adjusted scaling of FDG positron emission tomography images for statistical evaluation in patients with suspected Alzheimer's disease. *J Neuroimaging.* 2005; 15(4): 348-355. <https://doi.org/10.1177/1051228405280169>
- Toussaint PJ, Perlberg V, Bellec P, et al. Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. *Neuroimage.* 2012; 63(2): 936-946. <https://doi.org/10.1016/j.neuroimage.2012.03.091>
- Little RJ, Rubin DB. *Statistical Analysis with Missing Data.* 2nd ed. Hoboken (N.J.): Wiley-Interscience; 2002.
- Brosseau-Liard PE, Savalei V. Adjusting incremental fit indices for non-normality. *Multivariate Behav Res.* 2014; 49(5): 460-470. <https://doi.org/10.1080/00273171.2014.933697>
- Rosseel Y. lavaan: an R package for structural equation modeling. *J Stat Softw.* 2012; 48: 1-36. <https://doi.org/10.18637/jss.v048.i02> SE-Articles

35. Dufouil C, de Kersaint-Gilly A, Besancon V, et al. Longitudinal study of blood pressure and white matter hyperintensities: the EVA MRI Cohort. *Neurology*. 2001; 56(7): 921-926. <https://doi.org/10.1212/wnl.56.7.921>
36. Law CSW, Yeong KY. Repurposing antihypertensive drugs for the management of Alzheimer's disease. *Curr Med Chem*. 2021; 28(9): 1716-1730. <https://doi.org/10.2174/0929867327666200312114223>
37. den Brok M, van Dalen JW, Abdulrahman H, et al. Antihypertensive medication classes and the risk of dementia: a systematic review and network meta-analysis. *J Am Med Dir Assoc*. 2021; 22(7): 1386-1395. <https://doi.org/10.1016/j.jamda.2020.12.019>. e15.
38. Cunningham EL, Todd SA, Passmore P, Bullock R, McGuinness B. Pharmacological treatment of hypertension in people without prior cerebrovascular disease for the prevention of cognitive impairment and dementia. *Cochrane Database Syst Rev*. 2021; 5: CD004034. <https://doi.org/10.1002/14651858.CD004034.pub4>
39. Wang Q, He C, Zhu Y, Zhang Q, Zhang Z, Xie C. Cortical atrophy mediates the accumulating effects of vascular risk factors on cognitive decline in the Alzheimer's disease spectrum. *Aging (Albany NY)*. 2020; 12(14): 15058-15076. [10.18632/aging.103573](https://doi.org/10.18632/aging.103573)
40. Garcia-Alberca JM, Mendoza S, Gris E, Royo JL, Cruz-Gamero JM, Garcia-Casares N. White matter lesions and temporal atrophy are associated with cognitive and neuropsychiatric symptoms in patients with hypertension and Alzheimer's disease. *Int J Geriatr Psychiatry*. 2020; 35(11): 1292-1300. <https://doi.org/10.1002/gps.5366>
41. Manolio TA, Olson J, Longstreth WT. Hypertension and cognitive function: pathophysiologic effects of hypertension on the brain. *Curr Hypertens Rep*. 2003; 5(3): 255-261. <https://doi.org/10.1007/s11906-003-0029-6>
42. Haight T, Nick Bryan R, Erus G, et al. White matter microstructure, white matter lesions, and hypertension: an examination of early surrogate markers of vascular-related brain change in midlife. *Neuroimage Clin*. 2018; 18: 753-761. <https://doi.org/10.1016/j.nicl.2018.02.032>
43. Group SR, Wright Jr JT, Williamson JD, et al. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med*. 2015; 373(22): 2103-2116. <https://doi.org/10.1056/NEJMoa1511939>
44. Group SMI for the SR. Williamson JD, Pajewski NM, Pajewski NM, et al. Group SMI for the SR. Effect of intensive vs standard blood pressure control on probable dementia: a randomized clinical trial. *JAMA*. 2019; 321(6): 553-561. <https://doi.org/10.1001/jama.2018.21442>
45. Group TSMI for the SR. Nasrallah IM, Pajewski NM, Auchus AP, et al. Group TSMI for the SR. Association of Intensive vs Standard Blood Pressure Control With Cerebral White Matter Lesions. *JAMA*. 2019; 322(6): 524-534. <https://doi.org/10.1001/jama.2019.10551>
46. Zlokovic BV. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci*. 2011; 12(12): 723-738. <https://doi.org/10.1038/nrn3114>
47. Corlier F, Hafzalla G, Faskowitz J, et al. Systemic inflammation as a predictor of brain aging: contributions of physical activity, metabolic risk, and genetic risk. *Neuroimage*. 2018; 172: 118-129. <https://doi.org/10.1016/j.neuroimage.2017.12.027>
48. Heneka MT, Carson MJ, El Khoury J, et al. Neuroinflammation in Alzheimer's disease. *Lancet Neurol*. 2015; 14(4): 388-405. [https://doi.org/10.1016/S1474-4422\(15\)70016-5](https://doi.org/10.1016/S1474-4422(15)70016-5)

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lespinasse J, Chêne G, Mangin J-F, et al. Associations among hypertension, dementia biomarkers, and cognition: The MEMENTO cohort. *Alzheimer's Dement*. 2022;00-00. <https://doi.org/10.1002/alz.12866>

4.3 Compléments

4.3.1 Résultats supplémentaires de l'article

Tableau 4. e-Table 1. Distribution of missing data for each variable of interest by hypertension status – N=2048, The MEMENTO cohort, 2011-2014

Variable	Untreated, Normal blood pressure	Untreated, High blood pressure	Treated, Controlled	Treated, Uncontrolled
Age	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Sex	0 (0%)	0 (0%)	0 (0%)	0 (0%)
APOE genotype	33 (5%)	22 (5.8%)	28 (5.3%)	18 (3.8%)
Years of education	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)
Body Mass Index	6 (0.9%)	2 (0.5%)	1 (0.2%)	4 (0.8%)
Cardiovascular history	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Diabetes	0 (0%)	0 (0%)	0 (0%)	0 (0%)
volume of WMH	69 (10.4%)	34 (9%)	40 (7.5%)	37 (7.8%)
CSF p-tau	561 (84.2%)	321 (84.7%)	456 (86%)	406 (85.8%)
CSF A β ₄₂ / A β ₄₀	563 (84.5%)	322 (85%)	457 (86.2%)	408 (86.3%)
SUVr PET Amyloid	488 (73.3%)	255 (67.3%)	387 (73%)	341 (72.1%)
Cortical Thickness	43 (6.5%)	25 (6.6%)	30 (5.7%)	28 (5.9%)
Volume of Hippocampus	51 (7.7%)	31 (8.2%)	35 (6.6%)	39 (8.2%)
Brain Parenchymal Fraction	38 (5.7%)	25 (6.6%)	25 (4.7%)	22 (4.7%)
SUVr PET FDG	284 (42.6%)	153 (40.4%)	221 (41.7%)	203 (42.9%)
Free Reminding Test	1 (0.2%)	5 (1.3%)	5 (0.9%)	5 (1.1%)
Trail Making Test B	16 (2.4%)	14 (3.7%)	21 (4%)	28 (5.9%)
Verbal Fluency	12 (1.8%)	6 (1.6%)	7 (1.3%)	10 (2.1%)

APOE = apolipoprotein E gene, WMH = white matter hyperintensities, CSF = cerebrospinal fluid, p-tau = phosphorylated tau, A β = amyloid beta, SUVr = standard uptake value ratio, PET = positron emission tomography, FDG = fluorodeoxyglucose.

Tableau 5. e-Table 2. Estimates of direct and total effects of covariates on dementia biomarkers and cognition computed from structural equation model - The MEMENTO cohort, 2011-2014

Exposure	Pathway variable	Direct effects			Total effects		
		β , sd	95% CI	P-value	β , sd	95% CI	P-value
Age (10 years)		0.313	[0.260 ; 0.365]	<0.001			
Male		0.112	[0.019 ; 0.205]	0.018			
> 12 years of study		-0.014	[-0.104 ; 0.077]	0.763			
APOE4 carrier	WMH Volume	0.052	[-0.045 ; 0.148]	0.296			
BMI (1 unit)		-0.009	[-0.020 ; 0.001]	0.072			
CV history		0.164	[0.018 ; 0.310]	0.027			
Diabetic		0.037	[-0.103 ; 0.178]	0.603			
Age (10 years)		0.418	[0.335 ; 0.502]	<0.001	0.445	[0.365 ; 0.526]	<0.001
Male		0.127	[-0.014 ; 0.267]	0.077	0.136	[-0.004 ; 0.277]	0.057
> 12 years of study	AD neuropathology	-0.159	[-0.299 ; -0.018]	0.027	-0.160	[-0.301 ; -0.019]	0.026
APOE4 carrier		0.849	[0.691 ; 1.007]	<0.001	0.853	[0.696 ; 1.011]	<0.001
BMI (1 unit)		-0.035	[-0.053 ; -0.018]	<0.001	-0.036	[-0.053 ; -0.019]	<0.001
CV history		0.111	[-0.130 ; 0.352]	0.365	0.126	[-0.116 ; 0.367]	0.307

Exposure	Pathway variable	Direct effects			Total effects		
		β , sd	95% CI	P-value	β , sd	95% CI	P-value
Diabetic		0.143	[-0.140 ; 0.426]	0.322	0.146	[-0.137 ; 0.430]	0.312
Age (10 years)		0.642	[0.571 ; 0.713]	<0.001	0.775	[0.698 ; 0.852]	<0.001
Male		0.507	[0.418 ; 0.596]	<0.001	0.550	[0.464 ; 0.636]	<0.001
> 12 years of study		-0.021	[-0.105 ; 0.062]	0.618	-0.058	[-0.138 ; 0.022]	0.156
APOE4 carrier	Neurodegeneration	0.032	[-0.100 ; 0.164]	0.635	0.227	[0.127 ; 0.326]	<0.001
BMI (1 unit)		0.023	[0.011 ; 0.034]	<0.001	0.014	[0.003 ; 0.025]	0.016
CV history		0.105	[-0.030 ; 0.239]	0.127	0.151	[0.022 ; 0.280]	0.022
Diabetic		0.352	[0.181 ; 0.523]	<0.001	0.389	[0.230 ; 0.547]	<0.001
Age (10 years)		0.196	[0.009 ; 0.382]	0.039	-0.416	[-0.471 ; -0.360]	<0.001
Male		0.226	[0.063 ; 0.389]	0.007	-0.164	[-0.271 ; -0.056]	0.003
> 12 years of study		0.603	[0.503 ; 0.703]	<0.001	0.677	[0.584 ; 0.771]	<0.001
APOE4 carrier	Cognition	0.004	[-0.136 ; 0.144]	0.960	-0.337	[-0.447 ; -0.228]	<0.001
BMI (1 unit)		0.001	[-0.013 ; 0.016]	0.871	0.001	[-0.011 ; 0.012]	0.890
CV history		0.011	[-0.140 ; 0.163]	0.884	-0.120	[-0.259 ; 0.020]	0.093
Diabetic		0.058	[-0.138 ; 0.255]	0.561	-0.227	[-0.375 ; -0.079]	0.003

APOE4 = epsilon 4 allele of the apolipoprotein E gene, BMI = body mass index, WMH = white matter hyperintensities, CV = cardiovascular, AD = Alzheimer's disease.

Estimates (β and 95% Confidence Interval) are reported for one standard deviation increase of the pathway variable.

Tableau 6. e-Table 3. Estimates of the factor loadings and residual variances of the observed variables for each latent variable – N=2048, The MEMENTO cohort, 2011-2014

Latent variable	Observed variable	Loading*		Residual variance**	
		λ , sd	[95% CI]	ϵ	[95% CI]
AD biomarkers	CSF A β ₄₂ / A β ₄₀	0.823	[0.744 ; 0.902]	0.323	[0.193 ; 0.452]
	CSF p-tau	0.703	[0.621 ; 0.785]	0.506	[0.391 ; 0.621]
	SUVr Amyloid PET	0.861	[0.783 ; 0.939]	0.259	[0.125 ; 0.392]
Neurodegeneration	Cortical Thickness	0.508	[0.468 ; 0.548]	0.742	[0.702 ; 0.782]
	Volume of Hippocampus	0.676	[0.562 ; 0.789]	0.543	[0.390 ; 0.697]
	Brain Parenchymal Fraction	0.745	[0.623 ; 0.866]	0.445	[0.265 ; 0.626]
	SUVr FDG PET	0.612	[0.560 ; 0.664]	0.626	[0.563 ; 0.689]
Cognition	Free Reminding Test	0.714	[0.674 ; 0.753]	0.491	[0.434 ; 0.547]
	Verbal Fluency	0.636	[0.597 ; 0.675]	0.596	[0.546 ; 0.645]
	Trail Making Test B	0.609	[0.565 ; 0.654]	0.629	[0.575 ; 0.683]

AD = Alzheimer' disease, CSF = cerebrospinal fluid, A β = protein amyloid β , p-tau = phosphorylated tau, SUVr = standard uptake value ratio, PET = positron emission tomography, FDG = fluorodeoxyglucose.

* Estimated loadings are given for 1 standard deviation increase of latent variable relative to 1 standard deviation of observed variable.

** Residual variances are given in percentage of variance (ϵ) of the observed variable. The percentage of explained variance of each dimension can be deduced as 1 – ϵ .

Tableau 7. e-Table 4. Estimates of associations between hypertension status on dementia biomarkers and cognition according to sex. The MEMENTO cohort, 2011-2014

Exposure	Pathway variable	β , sd	CI (95%)	P-value
Treated, Controlled & Female	WMH volume	0.010	[-0.109 ; 0.130]	0.866

Exposure	Pathway variable	β , sd	CI (95%)	P-value
Treated, Controlled & Male		0.062	[-0.108 ; 0.233]	0.472
Treated, Uncontrolled & Female		0.098	[-0.049 ; 0.245]	0.192
Treated, Uncontrolled & Male		0.228	[0.023 ; 0.433]	0.029
Untreated, HBP & Female		0.205	[0.021 ; 0.390]	0.029
Untreated, HBP & Male		0.176	[-0.047 ; 0.400]	0.122
Treated, Controlled & Female		-0.066	[-0.322 ; 0.190]	0.613
Treated, Controlled & Male		0.099	[-0.168 ; 0.366]	0.466
Treated, Uncontrolled & Female	AD Biomarkers	-0.238	[-0.512 ; 0.035]	0.088
Treated, Uncontrolled & Male		0.020	[-0.278 ; 0.317]	0.896
Untreated, HBP & Female		-0.115	[-0.395 ; 0.165]	0.421
Untreated, HBP & Male		0.130	[-0.186 ; 0.446]	0.420
Treated, Controlled & Female		-0.092	[-0.220 ; 0.035]	0.157
Treated, Controlled & Male		0.089	[-0.095 ; 0.273]	0.343
Treated, Uncontrolled & Female	Neurodegeneration	0.191	[0.033 ; 0.349]	0.018
Treated, Uncontrolled & Male		0.055	[-0.140 ; 0.250]	0.582
Untreated, HBP & Female		-0.051	[-0.191 ; 0.090]	0.477
Untreated, HBP & Male		0.045	[-0.143 ; 0.233]	0.639
Treated, Controlled & Female		-0.054	[-0.232 ; 0.124]	0.551
Treated, Controlled & Male		0.159	[-0.040 ; 0.359]	0.117
Treated, Uncontrolled & Female	Cognition	-0.080	[-0.262 ; 0.101]	0.386
Treated, Uncontrolled & Male		-0.092	[-0.300 ; 0.116]	0.384
Untreated, HBP & Female		-0.037	[-0.214 ; 0.140]	0.682
Untreated, HBP & Male		0.078	[-0.137 ; 0.292]	0.477

HBP = high blood pressure, WMH = white matter hyperintensities, CSF = cerebrospinal fluid, AD = Alzheimer's disease. Mean effects are relatives to 1 standard deviation of the pathway variable

Tableau 8. e-Table 5. Estimates of associations between hypertension status on dementia biomarkers and cognition according to APOE4 status. The MEMENTO cohort, 2011-2014

Exposure	Pathway variable	β , sd	CI (95%)	P-value
Treated, Controlled & APOE4 carrier		-0.026	[-0.217 ; 0.164]	0.787
Treated, Controlled & APOE4 non carrier		0.051	[-0.065 ; 0.168]	0.388
Treated, Uncontrolled & APOE4 carrier	WMH volume	0.101	[-0.107 ; 0.310]	0.340
Treated, Uncontrolled & APOE4 non carrier		0.168	[0.023 ; 0.313]	0.023
Untreated, HBP & APOE4 carrier		0.268	[-0.013 ; 0.549]	0.061
Untreated, HBP & APOE4 non carrier		0.158	[-0.004 ; 0.320]	0.056
Treated, Controlled & APOE4 carrier		-0.075	[-0.446 ; 0.295]	0.690
Treated, Controlled & APOE4 non carrier		0.031	[-0.178 ; 0.240]	0.771
Treated, Uncontrolled & APOE4 carrier	AD Biomarkers	-0.112	[-0.520 ; 0.297]	0.592
Treated, Uncontrolled & APOE4 non carrier		-0.137	[-0.366 ; 0.093]	0.243
Untreated, HBP & APOE4 carrier		-0.141	[-0.535 ; 0.253]	0.482
Untreated, HBP & APOE4 non carrier		0.030	[-0.221 ; 0.281]	0.817
Treated, Controlled & APOE4 carrier		0.151	[-0.058 ; 0.359]	0.156
Treated, Controlled & APOE4 non carrier		-0.094	[-0.218 ; 0.029]	0.134
Treated, Uncontrolled & APOE4 carrier	Neurodegeneration	0.126	[-0.115 ; 0.367]	0.306
Treated, Uncontrolled & APOE4 non carrier		0.123	[-0.016 ; 0.262]	0.083
Untreated, HBP & APOE4 carrier		0.053	[-0.166 ; 0.272]	0.636
Untreated, HBP & APOE4 non carrier		-0.049	[-0.181 ; 0.083]	0.469
Treated, Controlled & APOE4 carrier		-0.128	[-0.367 ; 0.112]	0.298
Treated, Controlled & APOE4 non carrier	Cognition	0.095	[-0.067 ; 0.258]	0.250
Treated, Uncontrolled & APOE4 carrier		-0.179	[-0.424 ; 0.066]	0.151
Treated, Uncontrolled & APOE4 non carrier		-0.056	[-0.219 ; 0.107]	0.501

Exposure	Pathway variable	β , sd	CI (95%)	P-value
Untreated, HBP & APOE4 carrier			-0.071[-0.312 ; 0.169]	0.561
Untreated, HBP & APOE4 non carrier			0.036[-0.131 ; 0.203]	0.670

HBP = high blood pressure, WMH = white matter hyperintensities, CSF = cerebrospinal fluid, AD = Alzheimer's disease. Mean effects are relatives to 1 standard deviation of the pathway variable

Tableau 9. e-Table 6. Estimates of associations between hypertension status on dementia biomarkers and cognition according to age. The MEMENTO cohort, 2011-2014

Exposure	Pathway variable	β , sd	CI (95%)	P-value
Treated, Controlled & age>71.9			0.069[-0.095 ; 0.233]	0.409
Treated, Controlled & Aged <= 71.9 (median age)			0.021[-0.082 ; 0.124]	0.692
Treated, Uncontrolled & age>71.9	WMH volume		0.273[0.107 ; 0.439]	0.001
Treated, Uncontrolled & Aged <= 71.9 (median age)			-0.022[-0.143 ; 0.099]	0.719
Untreated, HBP & age>71.9			0.272[0.075 ; 0.470]	0.007
Untreated, HBP & Aged <= 71.9 (median age)			0.133[-0.064 ; 0.331]	0.185
Treated, Controlled & age>71.9			0.083[-0.173 ; 0.339]	0.525
Treated, Controlled & Aged <= 71.9 (median age)			-0.097[-0.318 ; 0.124]	0.388
Treated, Uncontrolled & age>71.9	AD Biomarkers		-0.130[-0.392 ; 0.131]	0.329
Treated, Uncontrolled & Aged <= 71.9 (median age)			-0.146[-0.408 ; 0.116]	0.274
Untreated, HBP & age>71.9			-0.127[-0.391 ; 0.137]	0.347
Untreated, HBP & Aged <= 71.9 (median age)			0.142[-0.161 ; 0.444]	0.358
Treated, Controlled & age>71.9			-0.014[-0.153 ; 0.126]	0.849
Treated, Controlled & Aged <= 71.9 (median age)			-0.027[-0.167 ; 0.112]	0.702
Treated, Uncontrolled & age>71.9	Neurodegeneration		0.145[0.001 ; 0.288]	0.048
Treated, Uncontrolled & Aged <= 71.9 (median age)			0.096[-0.101 ; 0.294]	0.340
Untreated, HBP & age>71.9			-0.033[-0.186 ; 0.121]	0.677
Untreated, HBP & Aged <= 71.9 (median age)			-0.005[-0.149 ; 0.139]	0.943
Treated, Controlled & age>71.9			0.078[-0.103 ; 0.258]	0.399
Treated, Controlled & Aged <= 71.9 (median age)			-0.010[-0.173 ; 0.154]	0.906
Treated, Uncontrolled & age>71.9	Cognition		-0.060[-0.230 ; 0.110]	0.490
Treated, Uncontrolled & Aged <= 71.9 (median age)			-0.130[-0.330 ; 0.070]	0.202
Untreated, HBP & age>71.9			0.009[-0.168 ; 0.186]	0.918
Untreated, HBP & Aged <= 71.9 (median age)			0.021[-0.165 ; 0.207]	0.826

HBP = high blood pressure, WMH = white matter hyperintensities, CSF = cerebrospinal fluid, AD = Alzheimer's disease.

Mean effects are relatives to 1 standard deviation of the pathway variable

4.3.2 Spécification du modèle avec R et le package *lavaan*

Les données transversales utilisées pour un SEM classique sont structurées de sorte que chaque ligne corresponde à une observation différente, dans notre cas chaque ligne correspond à un participant de MEMENTO et chaque colonne est une variable observée.

On utilise la syntaxe spécifique du package *lavaan* pour spécifier le modèle sous la forme d'une variable de type caractère. Les variables latentes de neuropathologie de type Alzheimer, de neurodégénérescence et de cognition, respectivement `lv_NC`, `lv_ND`, et `lv_AD` sont

définies par leurs équations de mesure avec l'opérateur \approx et les variables manifestes correspondantes :

- RISCTOTRL, TMTB_TAUX et FLU_ANIM qui sont respectivement les scores de la somme des rappels libres du RLRI-16, le nombre de déplacements corrects par seconde aux TMT-B et le nombre de noms d'animaux cités.
- mean_thick, HIPPO_VOL_TIV, BPF_TRANS_M0 et SUVR_ADNI_Tr_M0 qui sont respectivement l'épaisseur moyenne du cortex, le volume de l'hippocampe (par rapport au volume intracrânien), la fraction du parenchyme et le SUVR TEP-FDG moyen pour plusieurs régions d'intérêt.
- PTAU, AB4240 et SUVR_moyen qui sont respectivement le niveau de protéine p-tau dans le LCR, le rapport des protéines $A\beta_{42}/A\beta_{40}$ dans le LCR et le SUVR moyen de la TEP amyloïde.

#Variables latentes

```
lv_NC ≈ RISCTOTRL + TMTB_TAUX + FLU_ANIM
lv_ND ≈ mean_thick + HIPPO_VOL_TIV + BPF_TRANS_M0 + SUVR_ADNI_Tr_M0
lv_AD ≈ PTAU + AB4240 + SUVR_moyen
```

Les autres équations de régression qui définissent la partie structurelle du modèle sont indiquées par l'opérateur \sim . La maladie des petits vaisseaux est représentée par le volume des HSB (HSB_VOL) et les régressions sont ajustées sur l'âge (AGE_CONS), le sexe (SEX), la présence de l'allèle $\epsilon 4$ (APOE_eps4_dm), un niveau d'étude > 12 ans (NIVETUD_dm1), l'indice de masse corporelle (IMC), la présence d'antécédents cardio-vasculaire (ATCD_CV_dm) et la présence d'un diabète (CM_DIAB_dm). Enfin l'exposition aux catégories d'hypertension est représentée par EXPO_cat_dm1, EXPO_cat_dm2 et EXPO_cat_dm3 contre la référence (les participants sans hypertension).

On peut attribuer à chaque coefficient un nom à l'aide de l'opérateur * pour faciliter l'analyse de médiation qui inclut le calcul des effets totaux et leurs intervalles de confiance respectifs (par exemple mv_nc est le nom attribué au coefficient qui associe la cognition au volume des HSB).

#Régressions

```
lv_NC ~ mv_nc*HSB_VOL + ad_nc*lv_AD + nd_nc*lv_ND +
sex_nc*SEX_dm + age_nc*AGE_CONS + apo_nc*APOE_eps4_dm + etu1_nc*NIVETUD_dm1 +
exp1_nc*EXPO_cat_dm1 + exp2_nc*EXPO_cat_dm2 + exp3_nc*EXPO_cat_dm3 +
bmi_nc*IMC + atcdcv_nc*ATCD_CV_dm + diab_nc*CM_DIAB_dm

lv_ND ~ mv_nd*HSB_VOL + ad_nd*lv_AD +
sex_nd*SEX_dm + age_nd*AGE_CONS + apo_nd*APOE_eps4_dm + etu1_nd*NIVETUD_dm1 +
```

```
exp1_nd*EXPO_cat_dm1 + exp2_nd*EXPO_cat_dm2 + exp3_nd*EXPO_cat_dm3 +
bmi_nd*IMC + atcdcv_nd*ATCD_CV_dm + diab_nd*CM_DIAB_dm
```

```
lv_AD ~ mv_ad*HSB_VOL +
sex_ad*SEX_dm + age_ad*AGE_CONS + apo_ad*APOE_eps4_dm +
etu1_ad*NIVETUD_dm1 +
exp1_ad*EXPO_cat_dm1 + exp2_ad*EXPO_cat_dm2 + exp3_ad*EXPO_cat_dm3 +
bmi_ad*IMC + atcdcv_ad*ATCD_CV_dm + diab_ad*CM_DIAB_dm
```

```
HSB_VOL ~ sex_mv*SEX_dm + age_mv*AGE_CONS + apo_mv*APOE_eps4_dm +
etu1_mv*NIVETUD_dm1 +
exp1_mv*EXPO_cat_dm1 + exp2_mv*EXPO_cat_dm2 + exp3_mv*EXPO_cat_dm3 +
bmi_mv*IMC + atcdcv_mv*ATCD_CV_dm + diab_mv*CM_DIAB_dm
```

Pour le calcul des effets totaux on utilise les noms précédemment attribués aux différents coefficients et l'opérateur `:=`. Par exemple les lignes suivantes détaillent le calcul des effets totaux pour les catégories d'exposition à l'hypertension sur les variables latentes `lv_AD`, `lv_ND` et `lv_NC`.

```
# Mediation
total_exp1_ad := exp1_ad + exp1_mv*mv_ad
total_exp2_ad := exp2_ad + exp2_mv*mv_ad
total_exp3_ad := exp3_ad + exp3_mv*mv_ad

total_exp1_nd := exp1_nd + exp1_ad*ad_nd + exp1_mv*mv_nd + exp1_mv*mv_ad*ad_nd
total_exp2_nd := exp2_nd + exp2_ad*ad_nd + exp2_mv*mv_nd + exp2_mv*mv_ad*ad_nd
total_exp3_nd := exp3_nd + exp3_ad*ad_nd + exp3_mv*mv_nd + exp3_mv*mv_ad*ad_nd

total_exp1_nc := exp1_nc + exp1_nd*nd_nc + exp1_ad*ad_nc + exp1_ad*ad_nd*nd_nc + exp1_mv*mv_nc +
exp1_mv*mv_nd*nd_nc + exp1_mv*mv_ad*ad_nc + exp1_mv*mv_ad*ad_nd*nd_nc
total_exp2_nc := exp2_nc + exp2_nd*nd_nc + exp2_ad*ad_nc + exp2_ad*ad_nd*nd_nc + exp2_mv*mv_nc +
exp2_mv*mv_nd*nd_nc + exp2_mv*mv_ad*ad_nc + exp2_mv*mv_ad*ad_nd*nd_nc
total_exp3_nc := exp3_nc + exp3_nd*nd_nc + exp3_ad*ad_nc + exp3_ad*ad_nd*nd_nc + exp3_mv*mv_nc +
exp3_mv*mv_nd*nd_nc + exp3_mv*mv_ad*ad_nc + exp3_mv*mv_ad*ad_nd*nd_nc
```

Puis des effets totaux du volume des HSB sur la neurodégénérescence et la cognition ainsi que l'effet total de la neuropathologie de type Alzheimer sur la cognition.

```
total_mv_nd := mv_nd + mv_ad*ad_nd
total_mv_nc := mv_nc + mv_nd*nd_nc + mv_ad*ad_nc + mv_ad*ad_nd*nd_nc
total_ad_nc := ad_nc + ad_nd*nd_nc
```

4.3.3 Estimation du modèle avec *R* et le package *lavaan*

L'estimation d'un SEM avec le package *lavaan* se fait par la fonction `sem()` :

```
fit_1 <- sem(model = model_1, data = data_SEM_1, std.lv = T, estimator = "MLR", missing = "fiml.x")
```

Où `model_1` est une unique chaîne de caractères pour spécifier le modèle et `data` le jeu de données. Dans notre cas nous avons utilisé l'option `std.lv = TRUE` pour identifier le modèle en fixant les variances résiduelles des variables latentes à 1, `missing = 'fiml.x'` pour utiliser l'estimation par maximum de vraisemblance sous l'hypothèse de données MAR, et `estimator = 'MLR'` pour utiliser une correction des variances et de la statistique de test robuste à la non-normalité.

Les estimations et indicateurs de performances du SEM sont obtenus par la commande `summary(fit_1, std.noxt = T)` où `fit_1` est l'objet qui résulte de la fonction `sem()` et `std.noxt = TRUE` permet de standardiser les coefficients estimés à l'exception des variables observées exogènes, soit les catégories d'exposition à l'hypertension et les variables d'ajustement dans le cas de notre modèle. Ainsi les résultats sont plus facilement interprétables puisque les coefficients standardisés correspondent à l'effet de l'exposition (oui vs non) et sont comparables entre les variables latentes et le volume des HSB (puisque proportionnels aux écart-types respectifs).

4.4 Conclusions

Les résultats mettent en évidence l'importance de la prise en charge et du contrôle de l'hypertension. Nous avons en effet montré que l'hypertension non contrôlée est associée à une dégradation de la structure cérébrale et des performances cognitives inférieures. Ces résultats apportent de nouvelles perspectives sur les mécanismes qui sous-tendent l'association entre l'hypertension et la démence. Les stratégies visant à optimiser le contrôle de la pression artérielle chez les patients pourraient contribuer au maintien du cerveau en bonne santé et à la prévention de la démence.

L'étude des chemins d'association de la relation entre l'hypertension et la cognition a montré que l'hypertension n'était pas associée aux marqueurs in vivo de la maladie d'Alzheimer mais uniquement à la maladie des petits vaisseaux et à la neurodégénérescence. Hormis les catégories d'exposition à l'hypertension, la maladie des petits vaisseaux n'était pas non plus associée aux marqueurs in vivo de la maladie d'Alzheimer. Ces résultats ne sont pas en faveur d'une interaction mais plutôt d'un rôle distinct des facteurs de risques vasculaires. Cette étude apporte une preuve supplémentaire de la contribution vasculaire aux troubles cognitifs et à la démence.

Une limite de ces travaux est qu'ils reposent sur l'analyse transversale des données recueillies à l'inclusion dans l'étude. L'analyse ne prend donc pas en compte la temporalité des relations entre l'exposition à l'hypertension, les biomarqueurs du LCR, les biomarqueurs d'imagerie cérébrale et les évaluations neuropsychologiques. Le sens des relations définies par le SEM sont des hypothèses qui s'appuient sur les connaissances actuelles du sujet dans

la littérature scientifique. Les notions de causalité doivent alors être interprétées avec prudence.

Avant d'étudier les relations temporelles entre les différents marqueurs de progression et les chemins d'association de l'exposition à l'hypertension, il est important de comprendre la temporalité des marqueurs de progression. C'est l'objectif des modèles de progression de la maladie qui sont développés dans la partie suivante.

5 MODELE DE PROGRESSION DE LA MAMA : APPLICATIONS A LA COHORTE MEMENTO

5.1 Résumé

La MAMA se caractérise par des atteintes multiples qui incluent l'accumulation pathologique de protéines dans le cerveau, la mort des neurones avec une atrophie cérébrale, ainsi que le déclin cognitif progressif. Cependant la séquence des atteintes au cours de l'histoire naturelle de la maladie est insuffisamment comprise. Nous avons vu que modéliser l'évolution des marqueurs de MAMA pour décrire l'histoire naturelle de la maladie induit des difficultés statistiques. En effet les échelles de temps habituelles dans le cadre d'une étude de cohorte (le temps depuis l'inclusion dans l'étude et l'âge chronologique), sont inappropriées et le temps jusqu'au diagnostic clinique n'est le plus souvent disponible uniquement pour un petit sous-échantillon des populations d'étude. Les modèles de progression de la maladie proposent comme solution de définir le temps de la maladie par une variable latente.

Pour ces travaux de recherche nous avons utilisé une approche avec un modèle mixte multivarié qui réaligne les trajectoires individuelles à court terme dans le temps latent de la maladie pour décrire les trajectoires à long terme des marqueurs étudiés. Contrairement à la littérature existante, notre méthodologie exploite les données sur l'information du diagnostic clinique comme une référence partiellement observée pour guider l'estimation du temps latent de la maladie. Le modèle était estimé avec le logiciel Stan par inférence bayésienne.

Nous avons appliqué la méthodologie dans l'étude MEMENTO à un échantillon d'analyse de 2186 participants avec un suivi intensif de 5 ans pour analyser les mesures répétées de 12 marqueurs incluant les concentrations en protéines $A\beta_{42}$, p-tau et t-tau dans le LCR, les mesures d'épaisseurs corticales de 3 régions d'intérêts ainsi que le volume de l'hippocampe, le métabolisme du glucose (SUVr TEP-FDG), et les tests neuropsychologiques évaluant la mémoire (RLRI-16), le langage (fluence sémantique) et la fonction exécutive (TMT-A). Pour étudier l'implication des facteurs de risques vasculaires et plus particulièrement la progression de la maladie des petits vaisseaux, nous avons aussi inclus le volume des HSB.

Le temps latent de la maladie estimé s'est étendu sur trente ans avant le diagnostic clinique. Nous avons considéré le profil d'une femme âgée de 70 ans avec un niveau d'études

élevé (> 12 ans) et porteuse de l'*APOE4* pour décrire la chronologie des marqueurs. Les marqueurs p-tau et t-tau dans le LCR précédaient de 5 ans les marqueurs d'atrophie cérébrale et de 10 ans la dégradation des scores au tests cognitifs. Le volume des HSB avait une progression plus rapide que les autres marqueurs, et concomitante aux marqueurs d'atrophie du cerveau. Nos résultats ont montré que la chronologie de la progression des marqueurs était différente en fonction du niveau d'éducation et du statut *APOE4*. Le premier marqueur était p-tau principalement chez les participants avec un niveau d'études élevé : la progression du trouble de la mémoire intervenait des années plus tard parmi les participants avec un niveau d'étude élevé (> 12 ans) alors qu'elle intervenait simultanément avec le niveau de p-tau et l'atrophie de l'hippocampe parmi les participants avec un niveau d'études plus faible (\leq 12 ans). Le niveau d' $A\beta_{42}$ était le marqueur le plus variable dans la séquence : il intervenait après l'atrophie de l'hippocampe et l'altération de la mémoire parmi les participants non porteurs de l'*APOE4* avec un faible niveau d'études, alors qu'il intervenait de façon concomitante à l'atrophie de l'hippocampe parmi les participants porteurs de l'*APOE4*.

En utilisant les mesures recueillies dans le cadre du suivi intensif de l'étude MEMENTO, nous avons pu identifier le temps de maladie latent pour prendre en compte l'hétérogénéité inter-individuelle à l'entrée dans l'étude et décrire l'histoire naturelle de la maladie au cours des années précédant le diagnostic clinique de la démence.

5.2 Article 2 : « Disease progression model anchored around clinical diagnosis in longitudinal cohorts: example of Alzheimer's disease and related dementia »

RESEARCH

Disease progression model anchored around clinical diagnosis in longitudinal cohorts: example of Alzheimer's disease and related dementia

Jeremie Lespinasse PharmD MSc^{1,2}, Carole Dufouil PhD^{1,2†}, Cecile Proust-Lima PhD^{1*†} and the Memento study group

* Correspondence:

cecile.proust-lima@u-bordeaux.fr

¹Inserm research center Bordeaux

Population Health, Bordeaux

School of Public Health, CIC

1401-EC, Bordeaux University,

Bordeaux, France

Full list of author information is

available at the end of the article

[†]co-last authors

Abstract

Background: Alzheimer's disease and related dementia (ADRD) are characterized by multiple and progressive anatomo-clinical changes including accumulation of abnormal proteins in the brain, brain atrophy and cognitive dysfunctions. Understanding the sequence and timing of these changes is of primary importance to gain insight into the disease natural history and ultimately allow earlier diagnosis. Yet, modeling their change over disease course from cohort data is challenging as the usual timescales (time since inclusion, chronological age) are inappropriate and time-to-clinical diagnosis is available on very small subsets of participants with short follow-ups prior to diagnosis. One solution to circumvent this challenge is to define the disease time as a latent variable.

Methods: We developed a multivariate mixed model approach that realigns individual trajectories into the latent disease time to describe disease progression. In contrast with the existing literature, our methodology exploits the clinical diagnosis information as a partially observed and approximate reference to guide the estimation of the latent disease time. The model estimation was carried out in the Bayesian Framework using Stan. We applied the methodology to the Memento study, a French multicentric clinic-based cohort of 2186 participants with 5-year intensive follow-up. Repeated measures of 12 ADRD markers stemmed from cerebrospinal fluid (CSF), brain imaging and cognitive tests were analyzed.

Results: The estimated latent disease time spanned over twenty years before the clinical diagnosis. Considering the profile of a woman aged 70 with a high level of education and *APOE4* carrier (the main genetic risk factor for ADRD), CSF markers of tau proteins accumulation preceded markers of brain atrophy by 5 years and cognitive decline by 10 years. However we observed that individual characteristics could substantially modify the sequence and timing of these changes, in particular for CSF level of $A\beta_{42}$.

Conclusion: By leveraging the available clinical diagnosis timing information, our disease progression model does not only realign trajectories into the most homogeneous way. It accounts for the inherent residual inter-individual variability in dementia progression to describe the long-term anatomo-clinical degradations according to the years preceding clinical diagnosis, and to provide clinically meaningful information on the sequence of events.

Trial registration: clinicaltrials.gov, NCT01926249. Registered on 16 August 2013

Keywords: multivariate mixed model; disease progression model; latent time; neurodegenerative disease; Alzheimer's disease

1 Background

Alzheimer's disease and related dementia (ADRD) are characterized by progressive changes in multiple anatomic-clinical domains including decline in one or several cognitive functions (such as memory, language and executive function) leading to clinical dementia with functional dependency and death [1]. Alzheimer's disease (AD) neuropathology is identified by the abnormal accumulation of proteins that form amyloid plaques and tau neurofilaments in the brain [2]. It is well established that vascular neuropathology also contribute to cognitive impairment and dementia [3]. Especially small vessel disease (causing white matter lesions and silent brain infarcts) could double the risk of clinical dementia [4, 5]. Progressive atrophy of some brain regions due to the death of neurons was also highlighted, more specifically of the hippocampus [6] or of the medial temporal lobe [7]. A decade ago, an hypothetical model of disease progression was proposed [8] to temporally order these progressive changes. It postulates that amyloid and tau are involved in cellular mechanisms of protein deposition that would induce later neuronal dysfunction and atrophy of the brain structures. Decline in cognitive functions then appears as a result of the loss of neural tissues. Several studies found evidence supporting the initiating role of the amyloid protein on these pathological changes [9, 10, 11], but these results were obtained among study participants at distinct clinical stages and no study provided a clear understanding of the anatomic-clinical changes over the entire natural history of the disease. In addition, the initial model completely ignored vascular contribution to cognitive impairment.

Modeling disease progression from cohort studies is statistically straightforward in many diseases using the mixed model theory for instance [12, 13], but it faces a fundamental statistical challenge in ADRD. Indeed, since neuropathological changes likely occur 15 to 25 years before any clinical diagnosis can be reached [10, 14], confirming the sequence and timing of the neuropathological degradation would require a follow-up of more than 20 years before diagnosis of clinical dementia which is only possible in population-based cohorts recruiting persons before middle age. However markers of neuropathological changes are mainly collected in clinical cohorts in which repeated measures of the most recent brain magnetic resonance imaging (MRI), brain positron emission tomography scanner (PET scan) and cerebrospinal fluid (CSF) derived biomarkers of ADRD before clinical diagnosis can be available. This is the case with the Memento cohort, a french nationwide clinic based study with 2323 participants followed up for 5 years, that gathers clinical examination, amyloid and tau biomarkers from cerebrospinal fluid, multiple brain images from MRI and PET scans (amyloid and glucose), and a neuropsychological tests battery. As shown in Figure 1 (C), modeling trajectories of the different markers according to the time to diagnosis would substantially reduce the sample size to demented participants but also limit the analysis only to the ultimate stages of the disease since most of participants in clinical ADRD cohorts are not followed-up for more than 5 years (e.g. [15, 16]). Alternative fixed timescales, such as time since inclusion or chronological age (Figure 1 (A) and (B), respectively), do not solve this temporal challenge for describing disease progression. Indeed, time since inclusion does not have any biological meaning, covers only a short period and is very heterogeneous

because participants are included at different clinical stages. Although much more relevant in research on ADRDs and age-related disorders, chronological age still induces too much inter-individual heterogeneity as people do not age similarly and ADRD onset may arise at various ages.

In the absence of a relevant completely-observed timescale, latent disease progression models have been developed with the aim to directly retrieve the unobserved disease time from the data. These data-driven methods usually consist in re-aligning the participants trajectories according to the unobserved disease time by assuming that participants experience overall the same disease progression. After a first methodology proposed by Jedynak *et al.* [17] to estimate a continuous disease time and describe the long-term progression of the biomarkers, many approaches have been developed and improved. Re-alignment of the individual trajectories into the disease progression time scale is managed using time-warping functions, either through an individual-specific time-shift [18, 19, 20, 21, 22, 23], through an individual time-shift combined with individual rate of progression [17, 24, 25, 26, 27] or through the definition of an exponential progression score [28]. Initially estimated individual by individual [17], most techniques now estimate the time-warping functions using random effects, thus entering the framework of nonlinear mixed models [19, 20, 21, 22, 23, 27]. Beyond the use of time-warping functions, the different models also varied according to the type of information considered (biological samples, brain images and/or neuropsychological evaluations) and according to the specification of the trajectories with sigmoid or exponential functions applied to Gaussian markers in their natural scales or after percentile transformations combined with linear mixed models.

Despite the rise of disease progression models based on a latent disease time, none of the techniques directly considered the partially-observed information provided by the clinical dementia diagnosis. Yet, when based on a clinical expertise, diagnosis represents a landmark in the natural history of the disease that could help anchor the definition of the latent disease time along the actual ADRD process. We thus propose in this work a latent disease time model that directly incorporates the partially-observed diagnosis information to re-align the individual trajectories along the ADRD disease time. This Disease Progression Anchored Model (DPAM) is an extension of the latent time joint mixed effect model (LTJMM) developed by Li *et al.* (2017) [19] and estimated in the Bayesian framework using Stan. The methodology was applied to describe the progression of 12 biomarkers including markers of AD neuropathology, small vessel disease, brain atrophy and cognitive functioning in the French clinic-based Memento study.

2 Methods

2.1 The Memento Cohort

The Memento cohort is a clinic-based study that recruited consecutively 2323 participants between April 2011 and June 2014 within the French national network of university-based memory clinics (Centres de Mémoires de Ressources et de Recherche [CMRR]). Participants were followed-up every 6 to 12 months during 5

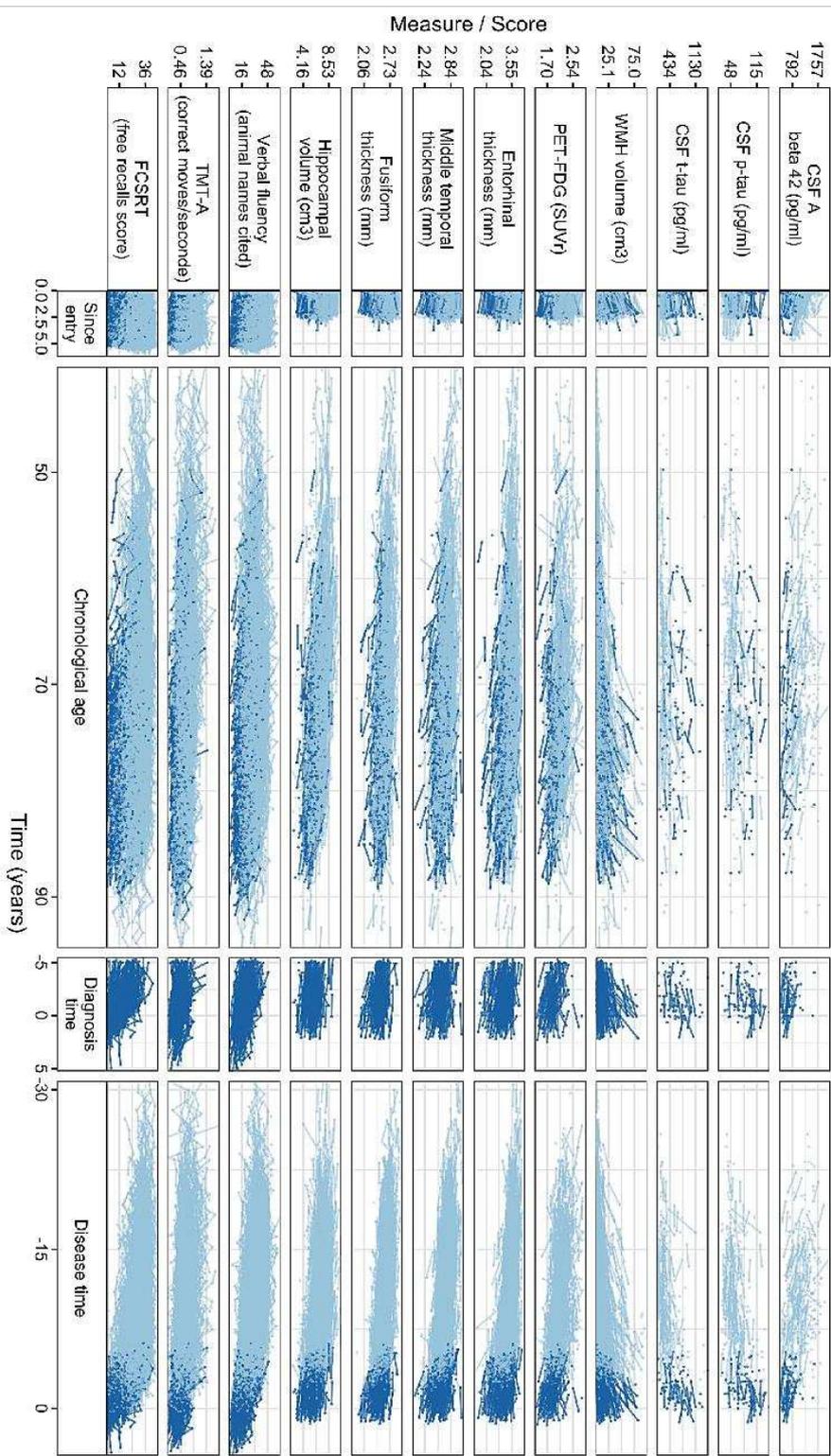


Figure 1: Individual trajectories of 12 markers in the Memento Cohort, France, 2011-2019 (2186 participants, 286 incident cases of dementia) according to 4 different timescales: (A) time in the study (or follow-up time), (B) age, (C) time to diagnosis (available only for the incident cases) and (D) the latent disease time estimated using the disease progression anchored model.

years. Inclusion criteria required that participants were not demented, had a clinical dementia rating (CDR) ≤ 0.5 , and performing 1 standard deviation worse than the subject's own age, sex and educational-level group mean in one or more cognitive functions (through neuropsychological tests performed within 6 months preceding the screening phase). Participants with isolated subjective complaints were also eligible if aged 60 years or older. This study was performed in accordance with the Declaration of Helsinki. All participants provided written informed consent. The MEMENTO cohort protocol was approved by an ethics committee ("Comité de Protection des Personnes Sud-Ouest et Outre Mer III"; approval number 2010-A01394-35) and was registered in ClinicalTrials.gov (Identifier: NCT01926249).

The study protocol is described in details in [16]. Every year, participants underwent a clinical evaluation that included an extensive battery of neuropsychological tests. Suspected cases of dementia during follow-up were reviewed by an independent expert committee and final clinical dementia diagnoses were established. At inclusion and at 24 months follow-up, all the patients were invited to undergo cerebral MRI, a 18F-fluorodeoxyglucose PET (FDG-PET) brain scan, and to have a lumbar puncture. The analytical sample consisted of the 2186 participants with at least one measure for one biomarker during the follow-up and without missing information on confounding variables including age, sex, education years and *APOE* status (apolipoproteine E gene).

2.2 Markers of ADRD

Repeated measures of markers of AD neuropathology, small vessel disease, brain atrophy and cognitive functioning were analyzed for a total of 12 markers of brain aging.

Biomarkers of AD neuropathology

The three markers of AD neuropathology were the amyloid- β 42 peptide ($A\beta$ 42), total tau (t-tau), and phosphorylated tau (p-tau181) measured from CSF using the standardized commercially available INNOTEST sandwich enzyme-linked immunosorbent assay (Fujirebio, Ghent, Belgium).

Biomarkers of brain atrophy

The markers of brain atrophy were the cortical thicknesses in three regions associated with ADRD progression (middle temporal, entorhinal, fusiform)[29] and the hippocampus volume measured from MRI T1-weighted with FreeSurfer[30] and SACHA[31], respectively. Hippocampus volume was relative to the total intracranial volume. Decreasing glucose metabolism as marker of neuron loss was measured by the mean FDG-PET uptake in AD-specific regions expressed as standard uptake value ratios (SUVr) [32, 33].

Biomarker of small vessel disease

Small vessel disease is mainly characterized by white matter lesions. Acquisitions of specific MRI 2D-T2 FLAIR sequences allowed to assess these lesions by the white matter hyperintensities (WMH) absolute volume from hyperintense signals using automated and validated method developed by the CATI [34].

Cognitive Assessment

Three key cognitive functions commonly impaired with the disease progression [35] were included:

- episodic memory with the sum score of the 3 free recalls from the free and cued selective reminding test (FCSRT), a french adaptation of the Grober and Buschke test [36].
- semantic verbal fluency (number of animals cited in 120 seconds)[37].
- executive functions with the number of correct moves per seconds at the trail making test A (TMT-A) [38].

2.3 Disease progression anchored model

Let's consider K markers measured repeatedly over time. They are denoted Y_{ijk} for the value of marker k ($k = 1, \dots, K$) for subject i ($i = 1, \dots, N$) at time t_{ijk} , with j the occasion ($j = 1, \dots, n_{ik}$). Time t is the fully observed timescale: age or time since entry in the study in our case. The DPAM is defined in three steps: (i) define the latent disease time from the observed timescale, (ii) define a comparable scale for all the markers, (iii) define the multivariate mixed model for the marker trajectories according to the latent disease time. We describe each step below.

Latent disease time definition

A latent disease time can be generically defined as an individual-specific monotonic function $s_i(t)$ of the observed time t . In our framework, the latent disease time corresponds to the actual time since clinical diagnosis had it been made in continuous time. By denoting T_i^* the actual unobserved time of diagnosis, the latent disease time is defined as:

$$s_i(t) = t - T_i^* \tag{1}$$

This definition assumes there is no distortion of time between t and s . The time in the disease s is a shift of the observed time t so that it is anchored to the actual time of diagnosis: $s_i(T_i^*) = 0$. We assume that the actual time of diagnosis is a latent variable with generic distribution \mathcal{D} . In the main analysis, we considered for instance a lognormal distribution: $\ln(T_i^*) \sim \mathcal{N}(\mu_T, \sigma_T^2)$. Without additional properties, the definition of this latent time shift is very standard and unrelated to the prior knowledge we may have about the time to clinical dementia.

In cohorts that focus on ADRD risk factors and natural history, only a part of the participants is diagnosed with clinical dementia during the follow-up, either because some participants dropped out the study or died clinical dementia diagnosis free or were free of clinical dementia diagnosis by the end of the planned follow-up. However, whether a participant was diagnosed with clinical dementia or not, valuable information can be leveraged to anchor the latent time T_i^* to the actual time of diagnosis.

Prior knowledge on the diagnosis

The clinical stage of the participants at the time of diagnosis is relatively homogeneous since clinical dementia is diagnosed by an independent committee of experts in the context of clinic based cohorts with intensive follow-up. Thus, time of clinical dementia diagnosis provides a reliable anchor time.

Let D_i be the indicator that the participant had a confirmed diagnosis of clinical dementia during follow-up. For participants diagnosed with clinical dementia ($D_i = 1$) the observed diagnosis time T_i^{diag} is likely in the neighbourhood of the actual unobserved time T_i^* . For participants who dropped out free of clinical dementia ($D_i = 0$), the time at the last clinical evaluation T_i^{last} is likely to be smaller than the actual unobserved disease time T_i^* . This was translated into the following constraints:

$$\begin{aligned} T_i^* &> T_i^{\text{last}} - \epsilon_L \text{ for } D_i = 0 \\ T_i^{\text{diag}} - \epsilon_L &< T_i^* < T_i^{\text{diag}} + \epsilon_U \text{ for } D_i = 1 \end{aligned}$$

where ϵ_L and ϵ_U are fixed scalars translating the lack of accuracy around the clinical evaluation and the diagnosis of clinical dementia. They have to be determined according to the study protocol and frequency of clinical evaluations.

Severity scale and comparison of markers

Independently from timescale definition, describing and comparing the sequence and speed of degradation across markers induces an additional challenge. Each marker has its own scale, and some, such as psychometric tests, are not necessarily Gaussian. Following previous works [18, 19, 20], we transformed the raw markers data Y_{ijk} into:

(i) percentiles $P_{ijk} = F_{Y_k}(Y_{ijk})$ ($P_{ijk} \in [0, 1]$) using the empirical cumulative distribution function F_Y of each marker. This allowed to define a severity scale from 0 (minimum value) to 1 (maximum value) on which to compare the sequence of markers' impairments. Note that markers were flipped when necessary, so that higher values systematically indicated higher impairment (0 = best condition observed and 1 = worst condition observed). Since the empirical cumulative distribution function highly depends on sample characteristics, it may not translate equi-distributed levels of impairments. For instance, in the Memento cohort, most patients remain at a very early clinical stage so that moderate to advanced stages would only correspond to the highest percentiles of the distribution. As suggested by other authors [19], this was corrected by preliminarily applying individual weights (w_i) to the marker measures according to the clinical stage of the participant. This percentile transformation can be obtained by the weighted cumulative density function $P_{ijk} = F_{Y_k}^w(Y_{ijk}, w_i)$ implemented in the Hmisc R package [39].

(ii) normalized values $\tilde{Y}_{ijk} = \Phi^{-1}(P_{ijk})$ using the inverse of the Gaussian cumulative distribution function Φ^{-1} . This allowed to apply multivariate linear mixed models for normal dependent variables \tilde{Y} and make predictions into the percentile scale using the inverse transformation $P_{ijk} = \Phi(\tilde{Y}_{ijk})$.

Multivariate linear mixed effects model

We described the marker trajectories in the normalized scale \tilde{Y}_{ijk} according to the latent disease time $s_i(t_{ijk})$ using the following multivariate linear mixed model:

$$\tilde{Y}_{ijk} = \mathbf{F}(s_i(t_{ijk}))^\top \boldsymbol{\beta}_k + \mathbf{X}_i(t_{ijk})\boldsymbol{\gamma}_k + \mathbf{F}(s_i(t_{ijk}))^\top \mathbf{u}_{ik} + \varepsilon_{ijk} \quad (2)$$

where \mathbf{F} is a basis of time functions defining the shape of the trajectory according to the disease time. Associated with β_k , it gives the mean trajectory of normalized marker k (for the reference profile of covariates). $\mathbf{X}_i(t_{ijk})$ are adjustment covariates associated with fixed effects γ_k and ε_{ijk} are the independent Gaussian error of measurement with marker-specific variance $\sigma_{\varepsilon_k}^2$. Finally \mathbf{u}_{ik} are the individual-and-marker-specific random effects defining the individual departure from the marker-mean trajectory. We assumed $\mathbf{u}_{ik} \sim \mathcal{N}(0, \mathbf{B}_k)$ with \mathbf{B}_k an unstructured variance covariance matrix. Random effects and errors are assumed independent. In addition, we assumed that the markers-specific random deviations were independent across markers so that the latent time-shift captured the inter-markers correlation.

DPAM specification for the Memento Cohort

In the application to the Memento Cohort, we considered a linear marker-specific trajectory ($\mathbf{F}(s) = (1, s)^\top$) and constrained $\beta_k \geq 0$ to impose a mean degradation over time for all the biomarkers. We also included a random slope only for the neuropsychological tests. This was to prevent any numerical non-identifiability issues for MRI and CSF markers where a maximum of two measures was collected. In addition, we considered as adjustment covariates: age, sex, years of education and *APOE4* status. Finally, we added an indicator of first visit for the neuropsychological tests to correct for the first passing effect [40]. Finally, we set the lack of accuracy around clinical evaluation to $\epsilon_L = \epsilon_U = 1.5$ years in the main analysis given clinical dementia diagnoses were performed every 6 months.

The individual weights used to define the percentiles of the marker severity scales were derived from four clinical stages according to the CDR-SB (CDR sum of the boxes) score at entry in the study: CDR-SB = 0, N = 784, $w_i = 2.80$; CDR-SB = 0.5, N = 794, $w_i = 2.74$; CDR-SB = 1, N = 323, $w_i = 6.81$; and CDR-SB > 1, N = 285, $w_i = 7.57$.

2.4 Estimation procedure

The estimation of our disease progression model was done in the Bayesian framework using Hamiltonian Monte Carlo No-U-turn sampling algorithm (HMC-NUTS) [41] to approximate the posterior distribution of the parameters with Markov Chain Monte Carlo (MCMC). We used Stan software (version 2.20.0) [42, 43] through the CmdStan interface with parallel computations on both the chains and the individuals. A commented version of our program, freely adapted from LTJMM [19], is available at https://github.com/jrmie/dpm_anchored.

Prior distributions

We considered standard weakly-informative priors for the multivariate mixed model parameters in equation (2) with for all $k = 1, \dots, K$: each element of β_k and γ_k following $\mathcal{N}(0, 10^2)$ (with β_k imposed to be positive), and σ_{ε_k} and the variances of the random-effects \mathbf{u}_{ik} following half-Cauchy(0, 2.5). For the latent disease time, we assumed the following distribution to incorporate the ϵ_L constraint and allow for negative $T_i^* : \ln(T_i^* + \epsilon_L) \sim \mathcal{N}(\mu_{T\epsilon}, \sigma_{T\epsilon})$ with $\mu_{T\epsilon} \sim \mathcal{N}(10, 10^2)$, and $\sigma_{T\epsilon} \sim \text{half-Cauchy}(0, 2.5)$.

Posterior summaries

We ran 4 chains of 6000 iterations burn-in and 2000 iterations for sampling, and we retained 1 iteration every 4 iterations to avoid auto-correlation issues from consecutive samples. Thus we approximated the posterior distribution with $D=2000$ iterations (500 by chain) and reported posterior means and 95% confidence intervals (CI95%) of the parameters.

Diagnostic checks

Diagnostic tools of Stan were used to evaluate the estimation procedure: convergence of the MCMC with the Gelman and Rubin [44] potential scale reduction statistic \hat{R} which compares variances between and within chains and effective sample size ratio (ESS) [45] which estimates sample size without any auto-correlation. These indicators were considered as satisfied if $\hat{R} < 1.05$ and $ESS/D \geq 0.5$ for all parameters of the model.

Sensitivity analyses

We assessed the influence of our definition of the latent disease time and the associated constraints in sensitivity analyses. Specifically, we compared our DPAM using the actual dementia time defined according to a lognormal distribution and the use of prior information on the observed diagnoses times with the non-anchored disease progression model specification in which T_i^* followed a Gaussian distribution without any constraint. We also evaluated the stability of the results when considering weaker constraints ($\epsilon_L = \epsilon_U = 3$ year) to guide the estimation of the disease time. The comparison was based on the residual root mean squared error (RMSE).

2.5 Predictions of the biomarkers' mean trajectories in the severity scale

A central output of this methodology is the description of biomarkers mean trajectories according to latent disease time s . Let define $P_{ik}(s_i(t_{ijk})) = P_{ijk}$ and $\tilde{Y}_{ik}(s_i(t_{ijk})) = \tilde{Y}_{ijk}$. The mean trajectory of biomarker k for a covariate profile \mathbf{x} (independent of time for simplicity) according to latent disease time s in the severity scale is:

$$\mathbb{E}(P_{ik}(s)|_{X_i(t)=\mathbf{x}}) = \int \Phi(\tilde{y}) f_{\tilde{Y}(s)|_{X_i(t)=\mathbf{x}}}(\tilde{y}) d\tilde{y} \quad (3)$$

where $f_{\tilde{Y}(s)|_{X_i(t)=\mathbf{x}}}$ is the density function of $\tilde{Y}(s)$ given $X_i(t) = \mathbf{x}$. At each iteration d of the MCMC, this integral can be approximated by the Monte Carlo technique as $\mathbb{E}(P_{ik}(s)|_{X_i(t)=\mathbf{x}}) \approx \hat{P}_{ik}(s, \mathbf{x}) = \frac{1}{M} \sum_{m=1}^M \Phi(\tilde{y}_m)$ where \tilde{y}_m is randomly drawn from $\mathcal{N}\left(\mathbf{F}(s)\beta_k^{(d)} + \mathbf{x}\gamma_k^{(d)}, \mathbf{F}(s)^\top B_k^{(d)} \mathbf{F}(s) + \sigma_{\epsilon_k}^2\right)$ where $^{(d)}$ indicates the value of the parameters at the MCMC iteration d . We considered $M = 2000$.

The mean and its CI95% over the iterations are retained to describe the mean biomarker trajectory $\hat{P}_{ik}(s, \mathbf{x})$ over latent disease time for covariate profile \mathbf{x} between $s = -30$ years before dementia to $s = 5$ years after.

3 Results

Participants in the analytical sample (N=2186) were aged 70.9 years (SD=8.7) on average, 61.7% were women, 39.2% had more than 12 years education and 29.9% carried at least 1 allele $\epsilon 4$ of *APOE* gene (*APOE4*, a major genetic risk factor for ADRD). Additional description of the analytical sample at inclusion in the cohort is reported in Table 1. During the 5-year follow-up, 284 participants were newly diagnosed with dementia.

A description of the distribution markers at baseline is reported in Table 1 and individual trajectories displayed in Figure 1 (A,B,C). Almost all the participants had a least one cognitive measure at baseline, 2047 participants had volumetric MRI, 1236 had PET-FDG and 342 had CSF biomarkers. The number of repeated measures varied also substantially from an average of approximately 1.5 for CSF, MRI and FDG-PET to 4.6 for cognitive measures among participants who completed at least one measurement of the marker of interest during follow-up.

3.1 Estimated latent disease time

Individual estimated times of actual clinical dementia T_i^* were used to display the estimated time to actual clinical dementia from entry in the cohort $s_i(0)$ (Figure 2). For instance, $s_i(0) = -3$ corresponds to an estimated actual time of clinical dementia of 3 years after entering the cohort. Participants entered the cohort on

Table 1 Descriptive characteristics of the analysis sample at inclusion and over follow-up, Memento cohort, France, 2011-2019 (N=2186)

Variable	Inclusion in the study			Follow-up	
	Mean/n	SD/%	N	Repeated measures	N*
Individuals characteristics					
Age (years)	70.9	8.7	2186		
Female	1349	61.7%	2186		
>12 years education	856	39.2%	2186		
<i>APOE4</i> carrier	654	29.9%	2186		
CDR sum of the boxes					
= 0	784	35.9%	2186		
= 0.5	794	36.3%	2186		
= 1	323	14.8%	2186		
> 1	285	13.0%	2186		
Markers of AD (CSF)					
t-tau (pg/ml)	376.0	264.8	341	1.3	395
p-tau (pg/ml)	62.6	29.5	342	1.3	396
$A\beta_{42}$ (pg/ml)	1096.2	411.2	342	1.3	396
Brain imaging markers					
WMH volume (mm ³)	9.9	13.2	1993	1.7	2056
Volume of hippocampus (cm ³)	6.5	1.3	2047	1.7	2048
Entorhinal cortex (mm)	3.3	0.4	2041	1.7	2057
Middle temporal cortex (mm)	2.7	0.1	2047	1.7	2064
Fusiform cortex (mm)	2.6	0.1	2047	1.7	2064
PET-FDG (SUVr)	2.1	0.3	1236	1.5	1400
Cognitive tests					
Verbal Fluency (animals cited)	28.3	8.8	2147	4.6	2164
TMT-A (good moves/sec)	0.6	0.2	2165	4.6	2184
Free recalls FCSRT (sum score)	26.1	8.3	2169	4.5	2176

SD = standard deviation, *APOE4* = allele $\epsilon 4$ of the apolipoprotein E gene, CDR = clinical dementia rating, AD = Alzheimer's disease, CSF = cerebrospinal fluid, t-tau = total tau, p-tau = phosphorylated tau, $A\beta_{42}$ = isoform 42 of protein amyloid, WMH = white matter hyperintensities, PET = positron emission tomography, FDG = fluorodeoxyglucose, SUVr = standard uptake value ratio, TMT-A = trail making test A, FCSRT = free and cued selective reminding test

average 10.3 years before the estimated actual clinical dementia onset with a range from 0.74 to 30.8 years. Among incident dementia cases, the time to dementia onset was between 0.74 and 6.15 years. These times to dementia were very close to the observed clinical dementia diagnoses with an inaccuracy ranging from 1.01 years prior to the estimated time and 0.59 years after the estimated time (while the constraints allowed up to ± 1.5 years).

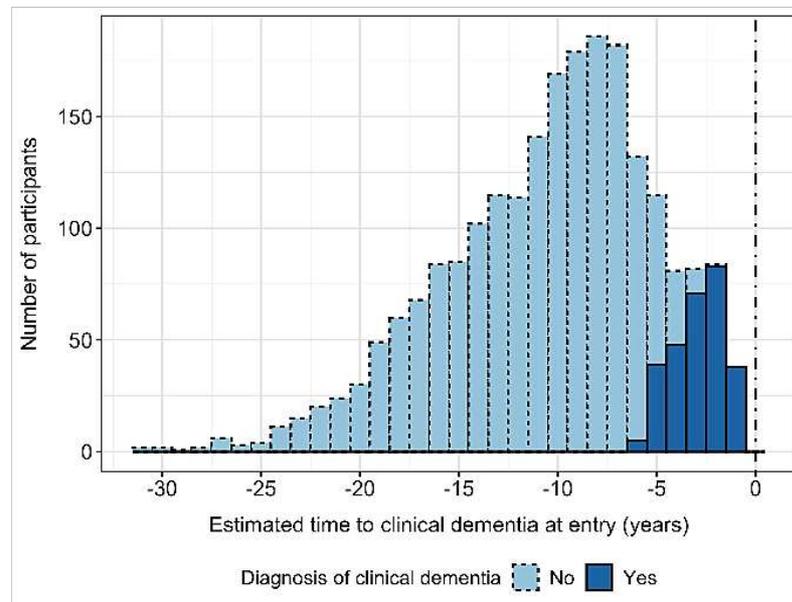


Figure 2: Posterior distribution of the estimated individual time to actual dementia at entry in the Memento Cohort (France, 2011-2019, N=2186) according to the last dementia diagnosis status.

3.2 Covariates association with the biomarker levels

Figure 3 displayed the mean and corresponding CI95% of the covariates associations with each biomarker. All coefficients γ_k are reported in standard deviation (SD) of the considered marker and adjusted for the other covariates. Age was significantly associated with worse levels for all markers.

The association with sex showed a substantially greater severity for men in memory domain (mean difference (MD)=0.33, 95%CI=[0.26, 0.40]), hippocampus volume (MD=0.37, 95%CI=[0.29, 0.44]) and FDG-PET (MD = 0.39, 95%CI = [0.30, 0.48]), and to a lesser extent on amyloid level (CSF $A\beta_{42}$) (MD=0.28, 95%CI=[0.12, 0.45]), WMH (MD=0.14, 95%CI=[0.05, 0.22]) and fusiform (MD=0.14, 95%CI=[0.06, 0.21]) and middle temporal (MD=0.14, 95%CI = [0.07, 0.22]) cortical thicknesses. Women tended to have higher level of p-tau (MD=-0.18, 95%CI=[-0.36, 0.00]).

High education (>12 years) was related to substantially better scores at cognitive tests: memory (MD=-0.40, 95%CI=[-0.46, -0.434]), language (MD=-0.48, 95%CI=[-0.55, -0.42]) and executive function (MD=-0.29, 95%CI=[-0.36, -0.23]); and slightly associated with lower degradation in FDG-PET (MD=-0.13, 95%CI=[-0.22, -0.05])

and Hippocampus volumes (MD=-0.11, 95%CI=[-0.19, -0.04]) but it was not related to cortical thicknesses, WMH or CSF markers.

APOE4 carriers had generally more severe degradation with larger differences for $A\beta_{42}$ (MD=0.60, 95%CI=[0.45, 0.74]), p-tau (MD=0.39, 95%CI=[0.25, 0.54]) and t-tau (MD=0.44, 95%CI=[0.29, 0.59]) in CSF.

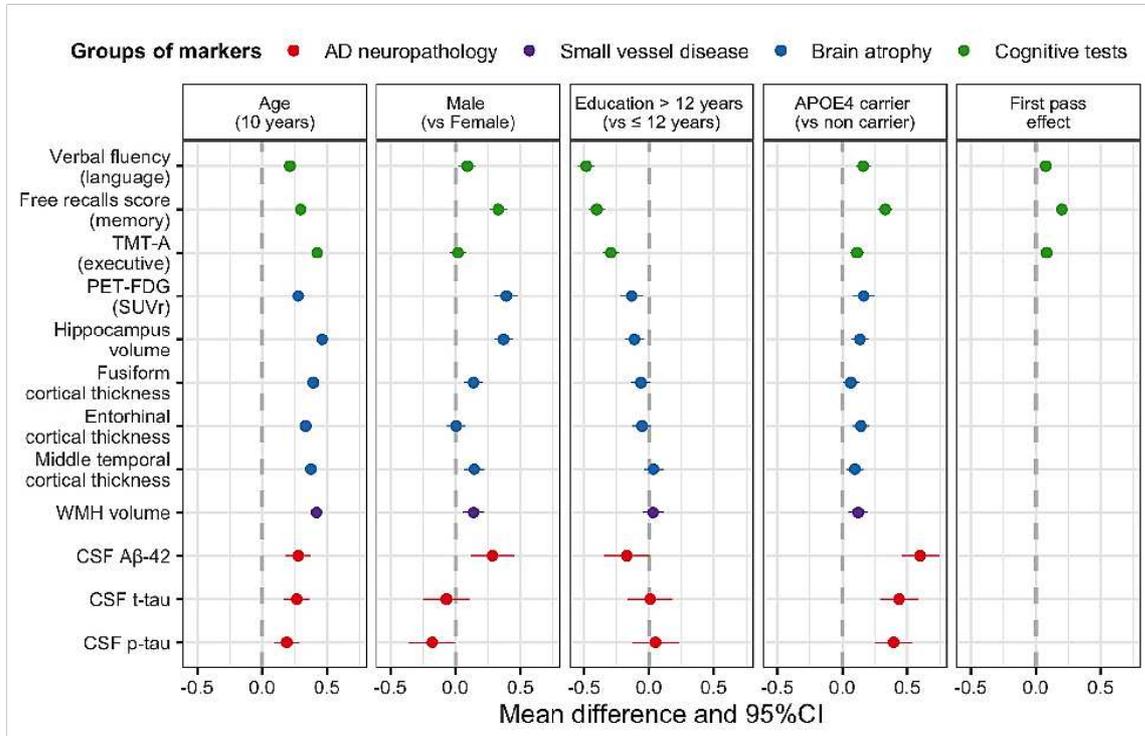


Figure 3: Estimated association of age, sex, education, *APOE4* status, and first practice effect with each of the 12 biomarkers in the normalized scale, the Memento Cohort, France, 2011-2019 (N=2186).

3.3 Trajectories of the markers in the latent disease time

Figure 4 displays the averaged trajectories of markers' progressions, between 30 years prior to clinical dementia to 5 years after, for a typical participant: woman of 70 years old, with more than 12 years of education and carrier of *APOE4* gene. For better clarity CI95% are not shown.

Marker severities

Thirty years prior to clinical dementia, all the markers were on average at low levels of severity (below 25%) with the highest levels for total tau and p-tau (23% (95%CI=[0.14,0.32]) and 30% (95%CI=[0.20,0.40]), respectively). In comparison, the 30% severity was reached for $A\beta_{42}$, WMH volume, volumetric MRI markers (cortical thicknesses, hippocampus volume, FDG-PET), memory and executive functioning about 10 to 12 years later (that is 18-20 years prior to clinical dementia) and about 15 years later for verbal fluency. Total tau and p-tau remained the more

impaired markers at all times although they degraded more slowly. WMH showed the fastest degradations with cortical thicknesses.

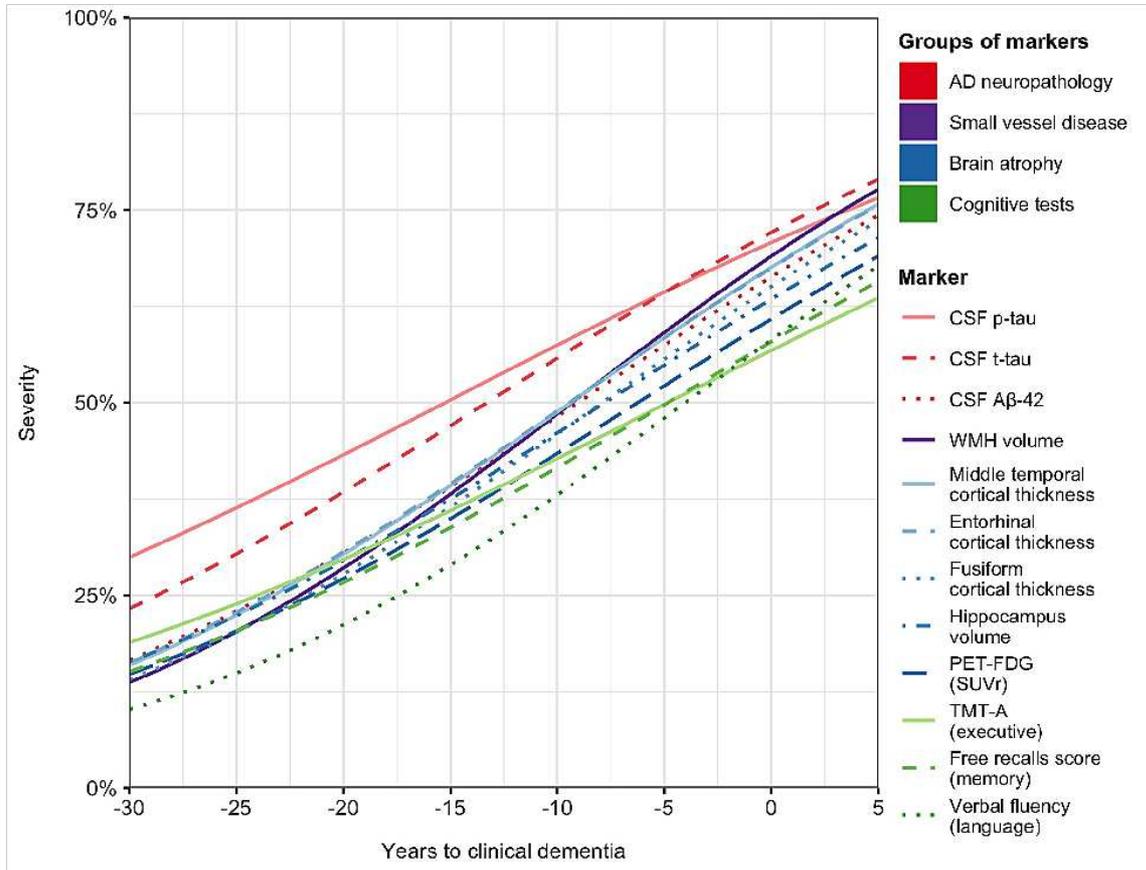


Figure 4: Mean trajectories of the 12 biomarkers of progression in the percentile scale according to latent disease time for a women of 70 years old, with more than 12 years of education and *APOE4* carrier, the Memento Cohort, France, 2011-2019 (N=2186).

Order of marker changes

Figure 5 summarizes the order in which the markers reach 50% severity with the respective uncertainties (95%CI) in the time scale of the disease (years before clinical dementia) for the same typical covariate profile as previously. According to the weighted severity scale the 50th percentile may give an indication of the entry into moderate severity. This level was first reached by p-tau and total-tau 15.4 (95%CI=[10.6, 20.3]) and 13.4 (95%CI=[9.5,17.3]) years before clinical dementia. About 5 years later the moderate severity was reached by A β 42 along with WMH volume, and cortical thicknesses of the middle temporal and entorhinal regions: respectively 9.1 (95%CI=[5.8, 12.3]), 9.3 (95%CI=[7.7, 11.0]), and 9.4 (95%CI=[7.8, 11.1]) years before clinical dementia. Then fusiform cortical thickness, hippocampus atrophy and glucose metabolism followed 1.5 to 3 years later with moderate severity reached 7.9 (95%CI=[6.3,9.5]), 7.8 (95%CI=[5.9,9.5]) and

6.2 (95%CI=[4.2,8.2]) years before clinical dementia, respectively. Finally, cognitive tests reached the moderate severity about 10 years after the p-tau in CSF. That was 4.2 (95%CI=[2.5,6.0]), 4.8 (95%CI=[3.0,6.7]) and 4.9 (95%CI=[2.9,6.9]) years before clinical dementia for language, memory and executive function, respectively

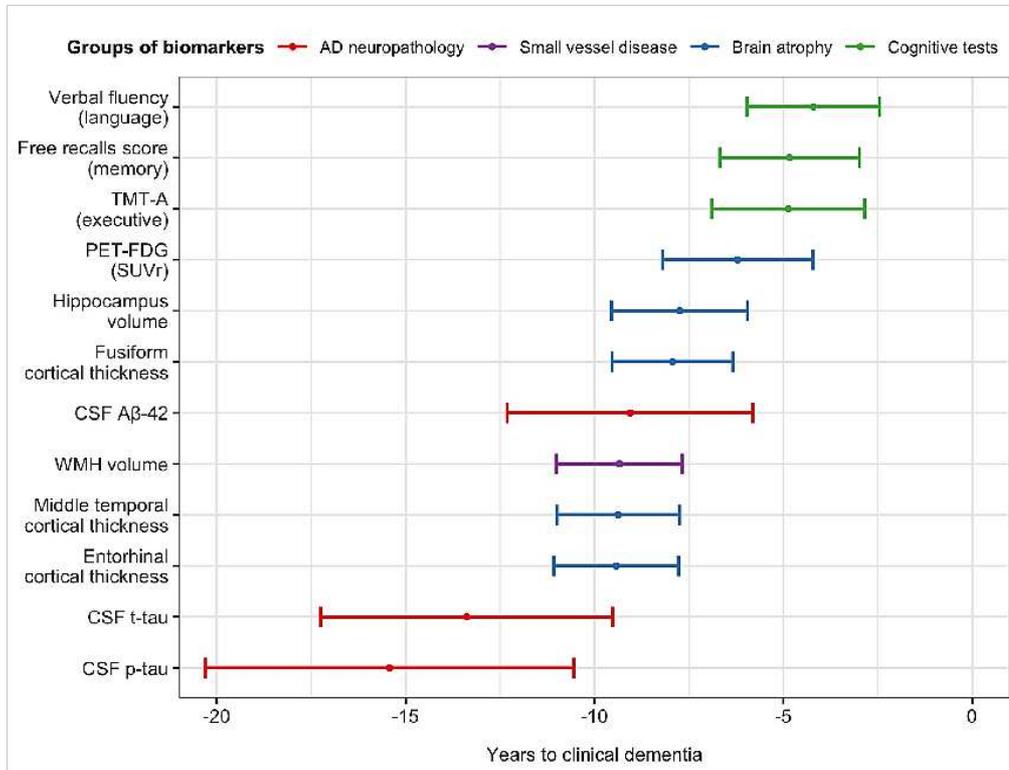


Figure 5: Ordering sequence and uncertainty (95%CI) of the biomarkers reaching the moderate severity for a women of 70 years old, with more than 12 years of education and *APOE4* carrier, the Memento Cohort, France, 2011-2019 (N=2186).

Covariate profiles

Because of the differential effect of the covariates on the markers, the degradation sequence differed according to the profile of the participant. To give a better sense of the heterogeneity of the sequence, we displayed in Figure 6 the averaged trajectories (along with CI95%) of 4 landmark biomarkers (p-tau level, A β ₄₂ level, hippocampus volume and memory test score) according to education years and *APOE4* status.

The anteriority of p-tau degradation was found mainly among the high education groups. Memory impairment progressed years later (among highly educated profiles) or contemporaneous (among less educated profiles) with p-tau level and hippocampus atrophy. A β ₄₂ level was the most variable marker in the sequence. It reached moderate severity level years later hippocampus atrophy and even after memory impairment in the profile *APOE4* non carriers and low education while the

degradation of $A\beta_{42}$ marker was at about the same time as the one of hippocampus atrophy for $APOE_4$ carriers.

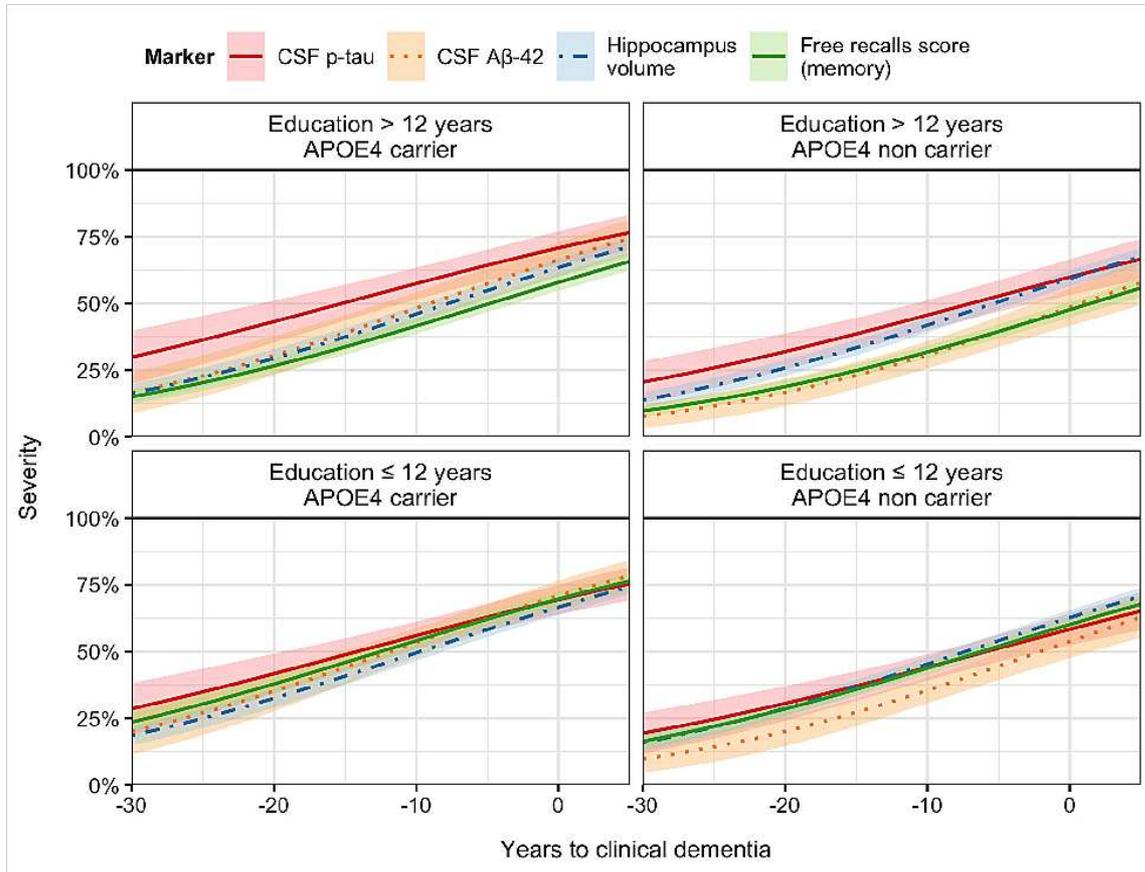


Figure 6: Average trajectories of 4 markers progression ($A\beta_{42}$, p-Tau, volume of hippocampus and FCSRT) according to latent disease time in the percentile scale for the 4 covariate profiles (education and $APOE_4$), the Memento Cohort, France, 2011-2019 (N=2186).

3.4 Sensitivity analyses

The fit to the data was unchanged when considering a larger uncertainty around the observed clinical diagnoses with $\epsilon_L = \epsilon_U = 3$ years (RMSE=0.0858 for both models, see e-Figure 1 in supplementary materials for RMSE per marker), and the results remained virtually the same. We also compared our DPAM that assumes a log-normal distribution for the latent disease time and anchors the latent disease time along the clinical dementia diagnosis with a non-anchored disease progression model (similar to the LTJMM methodology [19]) in which the latent disease time definition was completely data-driven and the latent time shift distribution was assumed as normal. In this non-anchored model, the latent disease time is centered on the average stage of the analytical sample at entry in the cohort. The non-anchored approach performed very similarly as our DPAM with RMSE = 0.0825 and RMSE

= 0.0858, respectively. The slight gain in RMSE of the non-anchored model was due to a slightly better fit of the neuropsychological data (see e-Figure 1 in the supplementary materials for a RMSE separated by marker). In this model, the estimated latent disease times of individuals diagnosed with clinical dementia were very far from the actual time to clinical diagnosis with a span over 15 years (e-Figure 2 in the supplementary materials). Indeed, as completely data-driven, the latent time shift was determined as the one homogenizing at most the data, and it was more influenced by the neuropsychological markers than MRI and CSF markers as the former brought much more information with more repeated measures. This underlines the importance of anchoring the model to realign the trajectories in link with the patient staging rather than only the inter-marker correlation.

4 Discussion

We developed a disease progression model to describe the markers' trajectories of the anatomic-clinical dimensions identified in AD/DRD progression towards clinical dementia. Using the intensive follow-up data of the French Memento Cohort, we identified a large variability in the patients staging at study entry with an estimated time to actual clinical dementia spanning over 30 years. The sequence of markers progression substantially varied according to education and *APOE4* status. However, we consistently identified p-tau as the first marker showing a pathological progression years before the onset of structural damage visible at brain imaging. Moreover, white matter hyperintensities, occurring concomitantly to regional brain atrophies, seemed to progress faster than other markers.

Compared to the rich literature on disease progression modelling using latent disease times ([17, 20, 19, 28, 21]), our adopted approach goes one step further. As previous works, we defined the latent disease time as an individual latent time shift shared by the disease markers and estimated it from the data. However, we also leveraged prior information on the clinical diagnoses to guide the estimation of the latent time-shift towards an actual clinical dementia diagnosis. As shown in the sensitivity analyses, without this prior information, the latent time-shift may over homogenize the trajectories of the markers. In contrast, anchoring the definition of latent time shift around the observed clinical diagnosis made it possible to realign markers trajectories around the actual time of clinical dementia (which is an important step in the clinical management of patients) while preserving the inherent heterogeneity in disease progression across individuals. In contrast with the literature, we also chose a lognormal distribution for the latent time shift (rather than a normal distribution) to allow for a longer tail in the time shift distribution, in agreement with real life data. This is in line with time-to-event analyses where lognormal distribution is usually preferred over normal distribution [46].

We applied our DPAM to the data of a large and well-defined clinical cohort in France, the Memento study which includes highly monitored and repeated data from cerebrospinal fluid biosample, brain imaging and cognitive data. Estimates of individual disease times to clinical dementia at entry in the cohort extended over decades, a consistent result with previous reported study on duration of AD/DRD

clinical stages [14]. Sequence and timing of markers along the natural history of the disease were partially consistent with the theoretical model of Jack *et al.* [8] although we found major differences in the sequence and timing of the markers according to the individual characteristics. Our findings supported that CSF p-tau showed increasing severity years before the degradation of glucose metabolism and brain atrophy on neuroimaging. Structural brain changes also preceded worsening of cognitive function. While amyloid deposit is widely considered as the initial cause of Alzheimer's disease [47], timing of CSF A β 42 were unclear as moderate severity was reached later than for CSF p-tau, a consistent result with previous disease progression model [48]. Indeed, timing of amyloid degradation substantially varied according to *APOE* ϵ status, thus contributing to the discussion challenging the plausibility of the amyloid cascade hypothesis [49]. Our results were also consistent with the small vessel disease contribution to cognitive impairment and dementia [3], suggesting that volume of white matter hyperintensities is the most rapidly deteriorating marker and contemporaneous with the degradation of cortical thicknesses years before cognitive impairment.

As any disease progression model, our approach relies on parametric assumptions. First, we considered a linear trajectory after a transformation of the markers into normal framework in agreement with previous works [19]. This has the advantage of reducing the nonlinear issues to a linear mixed model while still translating the sigmoid shape of markers severity hypothesized in the Jack model. Alternative sigmoid models usually model a generalized logistic trajectory [17] directly in the severity scale. It could be interesting in future works to compare the two approaches on the same data. Second, to distinguish the inter-marker correlation due to the disease staging from the intra-marker correlation, we assumed that the latent time shift captured all the correlation shared across markers and considered that marker-specific random effects were independent between markers. This assumption could be relaxed by allowing some correlation between subsets of markers, for instance MRI-derived markers or neuropsychological tests. This could be an extension of our work. Finally, we accounted for differences across covariate profiles through a global effect on each marker's severity. Although of interest, considering interactions between individual characteristics and rate of marker degradation would substantially complicate the model and the estimation procedure due to the higher number of additional parameters to estimate. Other progression models considered a covariate effect on the disease time [21, 23] rather than on each marker measure separately. This may be interesting for exploring covariates that may delay the progression towards dementia. However, as found in our application, some covariates may differentially modulate markers' trajectories and the sequence of markers' degradation. This was the case for years of education that showed large differences only in the neuropsychological tests, as an illustration of the concept of cognitive resilience [50]. Finally, by using the mixed model theory, our results are robust to missing at random (MAR) data and dropout. Given the number of markers included in the analysis, it is plausible that dropout and missing visits may be predictable by the markers and information considered.

5 Conclusion

Disease progression models are attractive methods to characterize the complete natural history of a disease when observed time is not relevant as it is the case for ADRD. We were able to shift individual trajectories into a latent disease time scale that extends over 30 years prior to dementia clinical diagnosis and describe the trajectories of markers' degradation. In contrast to previous findings, we used information from actual diagnosis time of clinical dementia to estimate the latent time underlying the long term progression of the markers. Differences with previous works on ADRD progression may come from the unique use of the Memento cohort data. Indeed, most disease progression models in ADRD were applied to the data from the Alzheimer's Disease Neuroimaging Initiative study (ADNI) [17, 19, 20, 28, 27, 23]. Given the major role of individual characteristics, applications on more diverse and less selected cohorts such as the Memento cohort are essential for ADRD research.

Appendix

Funding

The MEMENTO cohort is funded by the Fondation Plan Alzheimer (Alzheimer Plan 2008-2012), and the French Ministry of Research (MESRI, DGRI) through the Plan Maladies Neurodégénératives (2014-2019). This work was also supported by CIC 1401-EC, Bordeaux University Hospital (CHU Bordeaux, sponsor of the cohort), Inserm, and the University of Bordeaux. This work received funding from the French National Research Agency (ANR) as part of the Investment for the Future Programme ANR-18-RHUS-0002. The MEMENTO cohort has received funding support from AVID, GE Healthcare, and FUJIREBIO through private-public partnerships. The Insight-PreAD substudy was promoted by INSERM in collaboration with the Institut du Cerveau et de la Moelle Epinière, Institut Hospitalo-Universitaire, and Pfizer and has received support within the "Investissement d'Avenir" (ANR-10-AIHU-06) program. The funders had no role in study design, in data collection, analysis, and interpretation, or in writing of report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Abbreviations

$A\beta_{12}$ = amyloid beta 42, ADRD = Alzheimer's disease and related dementia, *APOE4* = allele $\epsilon 4$ of the apolipoprotein E gene, CDR = clinical dementia rate, CDR-SB = clinical dementia rate sum of the boxes, CI = confidence interval, CSF = cerebrospinal fluid, DPM = disease progression model, DPAM = disease progression anchored model, FCSRT = free and cued reminding test, FDG = fluorodeoxyglucose, LTJMM = latent time joint mixed model, MD = mean difference, MRI = magnetic resonance imaging, PET = positron emission tomography, p-Tau = phosphorylated tau, RMSE = root mean square error, SUVR = standard uptake value ratio, TMT-A = trail making test A, t-Tau = total tau, WMH = white matter hyperintensities.

Availability of data and materials

Memento data access request is available via the Dementia Platform UK Data Access appliance form (<https://portal.dementiasplatform.uk/Apply>) or via the Memento Secretariat (sophie.lamarque@u-bordeaux.fr). The R and Stan scripts used for the model specification are available at https://github.com/jrmie/dpm_anchored.

Ethics approval and consent to participate

This study was performed in accordance with the guidelines of the Declaration of Helsinki. The Memento study protocol has been approved by the local ethics committee ("Comité de Protection des Personnes Sud-Ouest et Outre Mer III"; approval number 2010-A01394-35). All participants provided written informed consent.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Inserm research center Bordeaux Population Health, Bordeaux School of Public Health, CIC 1401-EC, Bordeaux University, Bordeaux, France. ²Pôle de santé publique, Centre Hospitalier Universitaire (CHU) de Bordeaux, Bordeaux, France.

References

- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*. 2011;7(3):263-269.
- Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*. 1991;82(4):239-259.

3. Zlokovic BV, Gottesman RF, Bernstein KE, Seshadri S, McKee A, Snyder H, et al. Vascular contributions to cognitive impairment and dementia (VCID): A report from the 2018 National Heart, Lung, and Blood Institute and National Institute of Neurological Disorders and Stroke Workshop. *Alzheimer's and Dementia*. 2020;16:1714–1733.
4. Azarpazhooh MR, Avan A, Cipriano LE, Munoz DG, Sposato LA, Hachinski V. Concomitant vascular and neurodegenerative pathologies double the risk of dementia. *Alzheimer's and Dementia*. 2018;14(2):148–156.
5. Bos D, Wolters FJ, Darweesh SKL, Vernooij MW, de Wolf F, Ikram MA, et al. Cerebral small vessel disease and the risk of dementia: A systematic review and meta-analysis of population-based evidence. *Alzheimer's and Dementia*. 2018;14(11):1482–1492.
6. Schröder J, Pantel J. Neuroimaging of hippocampal atrophy in early recognition of Alzheimer's disease - a critical appraisal after two decades of research. *Psychiatry Research - Neuroimaging*. 2016;247:71–78.
7. Berron D, van Westen D, Ossenkoppele R, Strandberg O, Hansson O. Medial temporal lobe connectivity and its associations with cognition in early Alzheimer's disease. *Brain*. 2020 4;143:1233–1248. Available from: <https://doi.org/10.1093/brain/awaa068>.
8. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*. 2010;9(1):119–128.
9. Jack CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*. 2013;12(2):207–216.
10. Bateman RJ, Xiong C, Benzinger TLS, Fagan AM, Goate A, Fox NC, et al. Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. *New England Journal of Medicine*. 2012;367(9):795–804.
11. Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study. *The Lancet Neurology*. 2013;12(4):357–367.
12. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982 dec;38(4):963–974.
13. Lindstrom MJ, Bates DM. Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics*. 1990;46(3):673–687.
14. Vermunt L, Sikkes SAM, van den Hout A, Handels R, Bos I, van der Flier WM, et al. Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. *Alzheimer's and Dementia*. 2019;15(7):888–898.
15. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI) Clinical characterization; 2010.
16. Dufouil C, Dubois B, Vellas B, Pasquier F, Blanc F, Hugon J, et al. Cognitive and imaging markers in non-demented subjects attending a memory clinic: Study design and baseline findings of the MEMENTO cohort. *Alzheimer's Research and Therapy*. 2017;9(1):1–13.
17. Jedynak BM, Lang A, Liu B, Katz E, Zhang Y, Wyman BT, et al. A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort. *NeuroImage*. 2012;63(3):1478–1486.
18. Donohue MC, Jacqmin-Gadda H, Le Goff M, Thomas RG, Raman R, Gamst AC, et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's and Dementia*. 2014;10(5):S400–S410.
19. Li D, Iddi S, Thompson WK, Donohue MC. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Statistical Methods in Medical Research*. 2017;28(3):835–845.
20. Lorenzi M, Filippone M, Frisoni GB, Alexander DC, Ourselin S. Probabilistic disease progression modeling to characterize diagnostic uncertainty: Application to staging and prediction in Alzheimer's disease. *NeuroImage*. 2017;190(August 2017):56–68.
21. Raket LL. Statistical Disease Progression Modeling in Alzheimer Disease. *Frontiers in Big Data*. 2020;3(August):1–18.
22. Garbarino S, Lorenzi M. Modeling and Inference of Spatio-Temporal Protein Dynamics Across Brain Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2019;11492 LNCS:57–69.
23. Kühnel L, Berger AK, Markussen B, Raket LL. Simultaneous modeling of Alzheimer's disease progression via multiple cognitive scales. *Statistics in Medicine*. 2021;40(14):3251–3266.
24. Bilgel M, Prince JL, Wong DF, Resnick SM, Jedynak BM. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *NeuroImage*. 2016;134:658–670.
25. Marinescu RV, Eshaghi A, Lorenzi M, Young AL, Oxtoby NP, Garbarino S, et al. A vertex clustering model for disease progression: Application to cortical thickness images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2017;10265 LNCS:134–145.
26. Schiratti JB, Allasonnière S, Colliot O, Durrleman S. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *Journal of Machine Learning Research*. 2017;18:1–33.
27. Koval I, Bône A, Louis M, Lartigue T, Bottani S, Marcoux A, et al. AD Course Map charts Alzheimer's disease progression. *Scientific Reports*. 2021;11(1):1–16.
28. Bilgel M, Jedynak BM. Predicting time to dementia using a quantitative template of disease progression. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*. 2019;11:205–215.
29. Desikan RS, Cabral HJ, Fischl B, Guttman CRG, Blacker D, Hyman BT, et al. Temporoparietal MR Imaging Measures of Atrophy in Subjects with Mild Cognitive Impairment That Predict Subsequent Diagnosis of Alzheimer Disease. *American Journal of Neuroradiology*. 2009 3;30:532.
30. Fischl B, Kowne AVD, Destrieux C, Halgren E, Ségonne F, Salat DH, et al. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*. 2004 1;14:11–22.
31. Chupin M, Hammers A, Liu RS, Colliot O, Burdett J, Bardinet E, et al. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *NeuroImage*. 2009;46:749–761.

32. Buchert R, Wilke F, Chakrabarti B, Martin B, Brenner W, Mester J, et al. Adjusted scaling of FDG positron emission tomography images for statistical evaluation in patients with suspected Alzheimer's disease. *J Neuroimaging*. 2005;15(4):348–355.
33. Habert MO, Marie S, Bertin H, Reynal M, Martini JB, Diallo M, et al. Optimization of brain PET imaging for a multicentre trial: the French CATI experience. *EJNMMI Phys*. 2016;3:6.
34. Samaille T, Fillon L, Cuingnet R, Jouvent E, Chabriat H, Dormont D, et al. Contrast-based fully automatic segmentation of white matter hyperintensities: method and validation. *PLoS One*. 2012;7(11):e48953.
35. Weintraub S, Wicklund AH, Salmon DP. The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*. 2012;2.
36. Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. *Neurology*. 1988;38(6):900–903.
37. Thurstone LL. Psychophysical analysis. *American Journal of Psychology*. 1987 9;100:587.
38. Tombaugh TN. Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*. 2004;19(2):203–214.
39. Harrell Jr FE. Hmisc: Harrell Miscellaneous; 2021. R package version 4.6-0. Available from: <https://CRAN.R-project.org/package=Hmisc>.
40. Vivot A, Power MC, Glymour MM, Mayeda ER, Benitez A, Spiro A, et al. Jump, Hop, or Skip: Modeling Practice Effects in Studies of Determinants of Cognitive Change in Older Adults. *American Journal of Epidemiology*. 2016;183(4):302–314.
41. Hoffman MD, Gelman A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*. 2014;15:1593–1623.
42. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. *Journal of statistical software*. 2017;76(1).
43. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual; 2022. Version 2.30. Available from: <https://mc-stan.org/>.
44. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*. 1992;7(4):457–472.
45. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed.; 2013.
46. Jacqumin-Gadda H, Commenges D, Dartigues JF. Random Change-point Model for Joint Modeling of Cognitive Decline and Dementia. *Biometrics*. 2006 3;62:254–260. Available from: <https://doi.org/10.1111/j.1541-0420.2005.00443.x>.
47. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia*. 2018;14(4):535–562.
48. Li D, Iddi S, Thompson WK, Rafii MS, Aisen PS, Donohue MC. Bayesian latent time joint mixed-effects model of progression in the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*. 2018;10:657–668.
49. Frisoni GB, Altomare D, Thal DR, Ribaldi F, van der Kant R, Ossenkoppele R, et al. The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. *Nature Reviews Neuroscience*. 2022;23(1):53–66.
50. Arenaza-Urquijo EM, Vemuri P. Resistance vs resilience to Alzheimer disease. *Neurology*. 2018;90:695–703.

5.3 Compléments

5.3.1 Résultats supplémentaires de l'article

La figure ci-dessous montre la distribution des estimations du temps latent individuel obtenues à partir du modèle non-contraint : une approche similaire au modèle LTJMM (Li et al. 2017) où l'information du temps de diagnostic n'est pas utilisée. Les estimations sont complètement dépendantes de la population d'analyse puisque $T_{dem} = 0$ correspond à l'état de progression moyen des participants dans l'histoire naturelle de la maladie à l'entrée dans l'étude. On peut voir que le temps de maladie des participants avec un diagnostic de démence au cours du suivi est plus avancé. Par exemple $T_{dem} = 5$ signifie que le participant se situe à un stade de progression dans l'histoire naturelle de la maladie 5 ans plus tard que la moyenne des participants. Cependant le modèle n'apporte aucune information sur le temps de diagnostic pour faire une interprétation clinique de ce résultat.

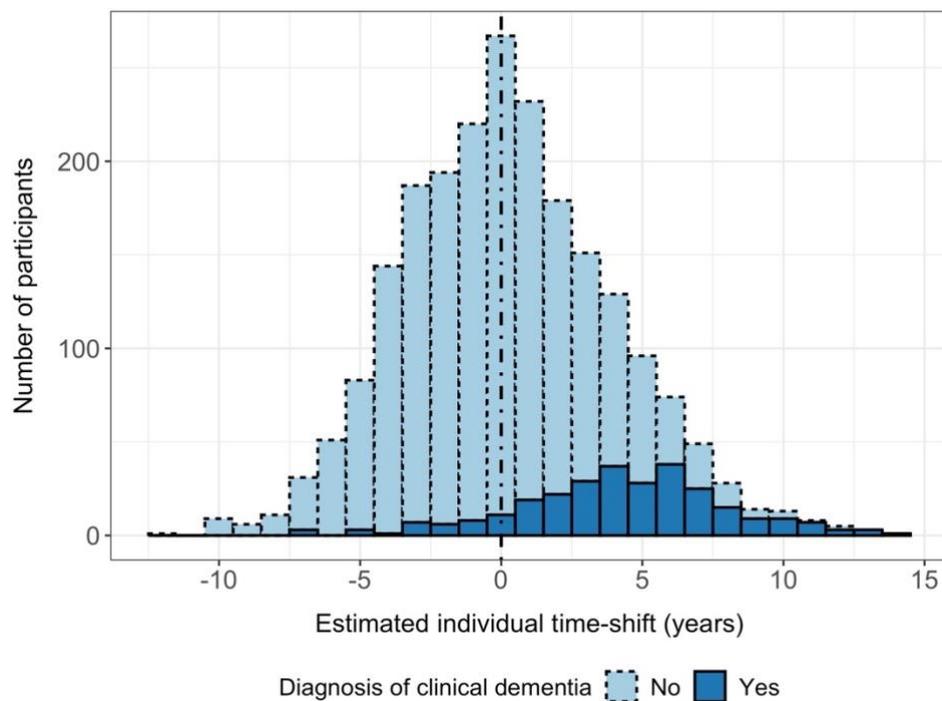


Figure 28. Distribution postérieure des temps de recalage individuels estimés à partir d'un modèle de progression non-contraint. Cohorte MEMENTO 2011-2019 (N=2186)

La figure ci-dessous représente les RMSE pour chaque marqueur du modèle non-contraint, du modèle contraint sur le temps de diagnostic associé à une incertitude de 3 ans et du modèle contraint sur le temps de diagnostic associé à une incertitude de 1,5 ans qui est

le modèle de l'analyse principale présentée dans l'article. Les 3 modèles présentent des performances similaires puisque les erreurs résiduelles varient peu.

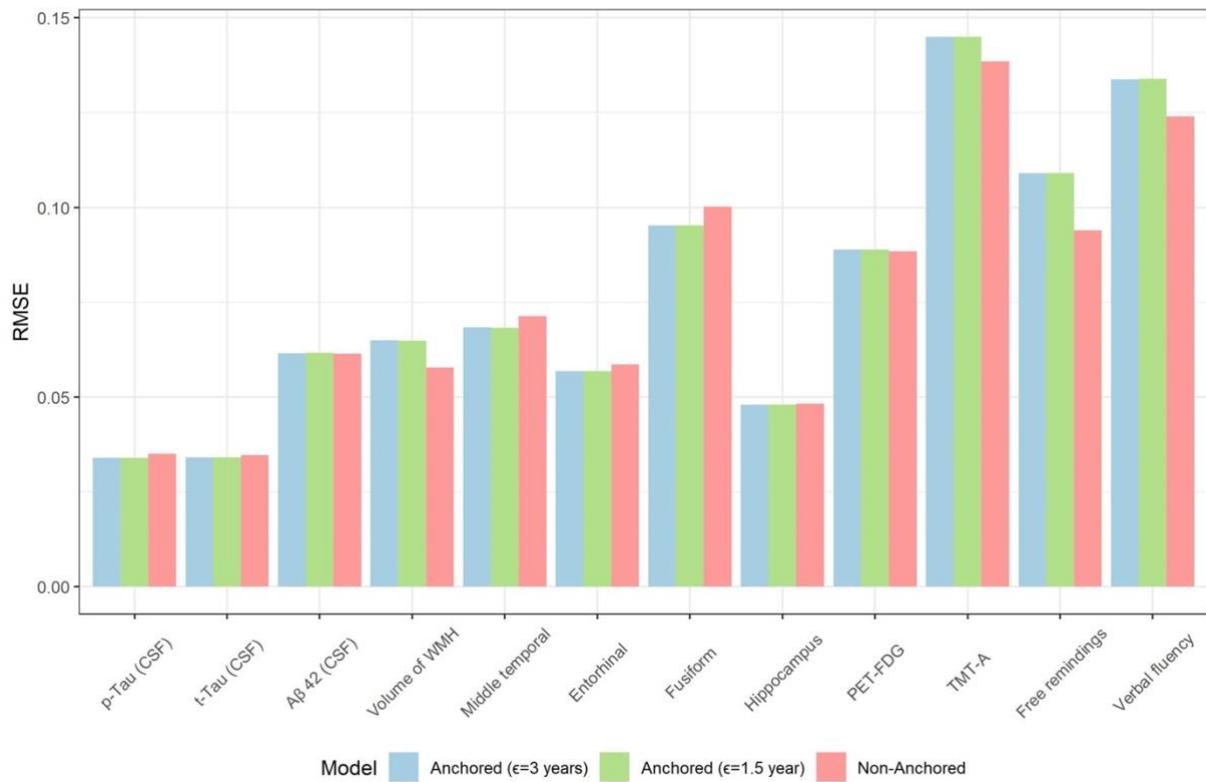


Figure 29. Residual mean square error per marker for the main disease progression anchored model (DPAM) with constraint of 1.5 years around the observed clinical diagnosis, the DPAM with a weaker constraint of 3 years, and a non-anchored disease progression.

5.3.2 Développement par étape du modèle de progression autour du diagnostic

Les travaux de recherche sur le développement d'un modèle de progression de la MAMA et son application aux données de la cohorte MEMENTO ont débuté par la compréhension du code utilisé pour la spécification du modèle LTJMM (« latent time joint mixed model ») (eq. 2.12), le modèle mixte à temps latent conjoint développé par (Li et al. 2017).

Modification du modèle LTJMM

Après une prise en main de Stan, le premier objectif a été de modifier le code pour :

(1) Supposer que les effets aléatoires entre les marqueurs soient indépendants. De cette façon le modèle distingue bien la variabilité inter-individuelle commune entre les marqueurs (le temps latent de progression) de la variabilité inter-individuelle spécifique aux marqueurs (intercepts et pentes aléatoires). Les effets aléatoires sont toujours distribués

selon une distribution multivariée normale mais avec une matrice de variance covariance diagonale par bloc.

(2) Relaxer la contrainte sur la somme des intercepts aléatoires $\sum_{k=1}^{(K-1)}(u_{0ik}) = -u_{0iK}$, prévue par (Li et al. 2017) pour s'assurer que le modèle soit identifiable. La contrainte sur les effets aléatoires était justifiée par les travaux de (Kneip and Gasser 1988) sur la convergence des modèles de régression auto-modélisés (SEMOR). Nous avons comparé les modèles avec et sans contraintes pour confirmer que nous n'en avons pas besoin.

Cependant les temps de calcul étaient considérables, entre 3 et 5 jours pouvaient être nécessaires pour échantillonner un nombre suffisant d'itérations pour la stabilisation des chaînes MCMC sur leur distribution stationnaire.

Calculs parallèles de la distribution postérieure des paramètres

La deuxième étape du développement s'est concentrée sur l'implémentation du calcul parallèle de la distribution postérieure à chaque itération (voir la partie 5.3.3.). Par la suite, la construction de la majorité des chaînes MCMC ne nécessitait plus qu'une dizaine d'heures de calcul tout en utilisant un nombre d'itérations de 2 à 4 fois plus important. Pour valider l'implémentation du calcul parallèle, nous avons pris soin de vérifier que les estimations des paramètres obtenues sur données simulées étaient identiques avec et sans les calculs parallèles.

Modèles avec mélange de distributions

Le modèle développé fait l'hypothèse que tous les participants connaissent la même dynamique de progression de la maladie, et notamment que tous les participants évolueront vers la démence s'ils sont observés pendant une durée suffisamment longue. Cette hypothèse semble très forte mais elle est similaire à l'hypothèse des modèles à risque proportionnels classiquement utilisée dans les méthodes d'analyses de données de survie.

Nous avons essayé de relaxer l'hypothèse de temps latent Gaussien en étudiant des modèles avec un mélange de distributions gaussiennes pour le temps latent. La spécification et l'estimation des modèles ont été réalisées avec la fonction `multlcmm()` du package R « `lcmm` » (Proust-Lima, Philipps, and Liqueur 2015) dans un premier temps puis avec l'implémentation d'un processus de Dirichlet dans le code de Stan dans un second temps. Les estimations obtenues par l'approche fréquentiste (`lcmm`) et l'approche bayésienne (Stan) étaient

similaires et les résultats n'ont pas permis de conclure une amélioration du modèle. Notamment en termes de signification clinique des résultats, d'amélioration de l'ajustement aux données et de la capacité de discrimination des participants en fonction des groupes identifiés.

Ajout de l'information partielle sur le temps du diagnostic

Les résultats de ces analyses ont fait l'objet de plusieurs présentations orales où les discussions ont encouragé le développement d'une autre approche originale pour définir le temps de maladie latent par rapport aux autres travaux de la littérature. D'une part le temps de maladie était considéré comme totalement inconnu, et d'autre part l'étude MEMENTO dispose d'un suivi clinique intensif avec un comité d'experts pour l'évaluation des cas de démence. Nous avons donc considéré l'information clinique du temps de diagnostic suffisamment fiable pour ne pas considérer le temps de maladie comme totalement inconnu. Nous avons utilisé le temps de diagnostic comme point d'ancrage dans l'estimation du temps latent de façon à décrire la dynamique de progression des marqueurs avec une signification clinique plus pertinente. Nous avons aussi introduit la contrainte nommée C ou ϵ_L qui traduit l'incertitude autour du temps de diagnostic (voir la partie « priori knowledge on the diagnosis » de l'article 2).

Distribution log-normale du temps de maladie

La distribution du temps latent de maladie est caractérisée pour un nombre important de participants en amont du diagnostic et un nombre très faible après, ce qui occasionne une distribution asymétrique. Nous avons constaté qu'une loi gaussienne n'était pas la plus adaptée à cette distribution et le modèle a été modifié pour supposer un a priori $P(\theta)$ selon une distribution log-normale du temps de maladie.

5.3.3 Parallélisation des calculs de la distribution postérieur $P(\theta|X)$

L'algorithme HMC-NUTS a des avantages certains pour explorer efficacement toutes les dimensions de l'espace des paramètres mais en contrepartie les temps de calcul s'allongent considérablement avec la complexité du modèle. La documentation du logiciel fournit des explications sur une fonctionnalité de parallélisation du calcul qui peut notamment être utilisé dans le cadre de modèles mixtes ou hiérarchiques (Stan Development Team

2022a). A chaque itération on peut réaliser un calcul parallèle de la distribution postérieure des paramètres pour construire plus rapidement la chaîne MCMC.

Paramétrage du compilateur de Stan

L'implémentation du calcul parallèle est possible dans un système « multithread » c'est-à-dire des processeurs avec la capacité d'exécuter simultanément plusieurs tâches, c'est le cas sur la plupart des ordinateurs récents. Il est nécessaire d'ajouter les lignes `CXXFLAGS += -DSTAN_THREADS` et `CXXFLAGS += -pthread` au fichier « local » du compilateur.

La deuxième possibilité est d'utiliser un environnement multiprocesseur équipé d'un outil MPI (« Message Passing Interface ») tel que OpenMPI, c'est le cas des serveurs de calcul dédiés à cette utilisation. Dans ce cas, il est nécessaire d'ajouter les lignes `STAN_MPI=true` et `CXX=mpicxx` au fichier « local » du compilateur.

Organisation des données

Dans le cas de modèles mixtes multivariés par exemple il est possible de diviser le jeu de données en regroupant les observations de chaque individu. Soit un jeu de données qui comprend les observations Y_{ij} de N individus à j_i occasions ($i = 1, \dots, N$ et $j = 1, \dots, j_i$) et un modèle qui comprend p paramètres partagés par l'ensemble des individus et q_i paramètres spécifiques à chaque individu i . La fonction `map_rect` implémentée dans Stan permet de distribuer les groupes de données sur les différentes unités de calcul. Dans notre cas un groupe correspond à un individu i . La fonction `map_rect(f, phi, theta, xr, xi)` est définie de la façon suivante :

- Un vecteur `phi` qui contient les p paramètres partagés par l'ensemble des individus, ce qui inclut les paramètres de variances et les paramètres fixes du modèle.
- Une structure de données `theta` de dimensions $N \times Q$ qui contient pour chaque individu les q_i paramètres individuels, ce qui inclut les paramètres aléatoires du modèle ; où Q est le nombre maximum de paramètres individuels parmi l'ensemble des individus.
- Une structure de données `xr` de dimensions $N \times J^{réel}$ qui contient les $j_i^{réel}$ observations pour chaque individu i qui appartiennent à l'ensemble des réels. Avec $J^{réel} = \max(j_i^{réel})$.

- Une structure `xi` de dimensions $N \times J^{entier}$ qui contient les j_i^{entier} observations pour chaque individu i qui appartient à l'ensemble des entiers. Avec $J^{entier} = \max(j_i^{entier})$.
- Une fonction `f` qui spécifie l'équation du modèle à partir des objets `phi`, `theta`, `xr` et `xi` pour incrémenter la fonction de probabilité logarithmique par les contributions de chaque individu.

De cette manière les données peuvent être divisées en N groupes de j_i lignes. Le calcul du logarithme de la fonction de probabilité postérieure $\log(P(\theta|X))$ est alors divisé entre les différentes unités de calcul en utilisant la somme des logarithmes des contributions individuelles à la vraisemblance $\log(P(X|\theta))$ par la ligne de code suivante :

```
target += sum(map_rect(f, phi, theta, xr, xi));
```

Simulations

L'estimation du gain de temps obtenu par le calcul parallèle dépend de nombreux paramètres : le nombre d'unités de calcul mais aussi de la complexité du modèle, en particulier le nombre de paramètres de variance-covariance et les contraintes spécifiées.

La simulation suivante a été réalisée en générant des observations à partir du modèle LTJMM pour 5 mesures répétées de 3 marqueurs chez 100 individus. En utilisant le calcul parallèle et en augmentant le nombre de processeurs on observe que le temps de calcul peut être divisé par 10 dans le cas de cette simulation.

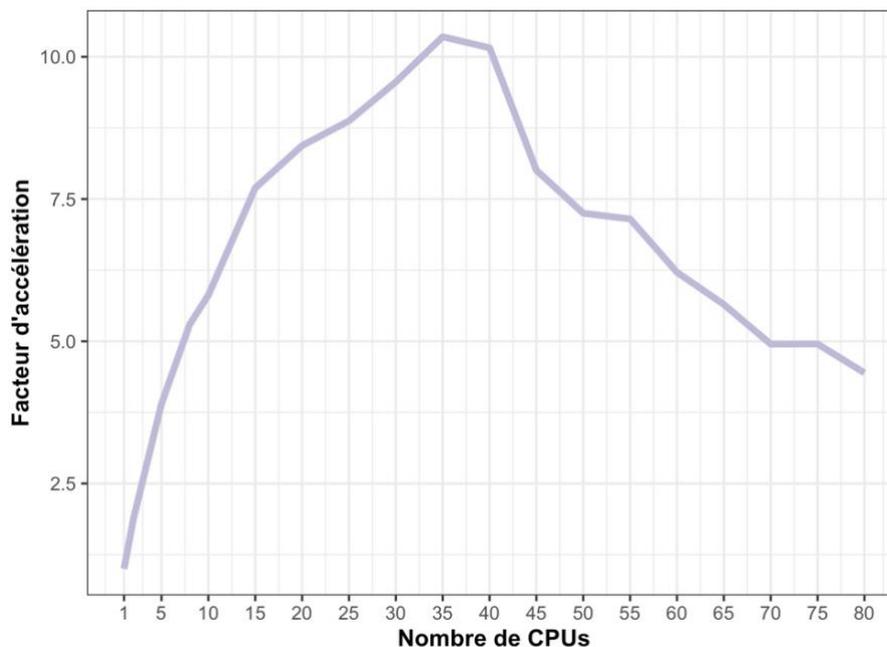


Figure 30. Facteur d'accélération de l'échantillonnage HMC-NUTS (Stan) avec calcul parallèle en fonction du nombre de processeurs à partir d'un jeu de données simulé.

Il existe un nombre optimal de processeurs car il arrive un moment où les temps d'échange des informations entre les unités de calcul prennent plus de temps que le calcul lui-même.

5.3.4 Spécification du modèle avec Stan

Cette partie détaille le code utilisé pour spécifier le modèle principal de l'article 2 dans le but d'être plus explicite sur les particularités de *Stan* et permettre de répliquer les analyses. Une partie du code utilisé pour les analyses est accessible dans le répertoire https://github.com/jrmie/dpm_anchored. Le fichier « *dpm.stan* » comprend le script du modèle *Stan* qui se compose de 7 blocs de code que nous allons définir.

Le block `data` permet d'identifier et de vérifier les données en entrée du programme (nombre et type de données par exemple).

```
data{
  int<lower = 0> n_obs;           // nombre d'observations
  int<lower = 0, upper = n_obs> n_sub; // nombre d'individus N
  int<lower = 0, upper = n_obs> n_out; // nombre de marqueurs K
  int<lower = 0> n_x;           // nombre de covariables (incluant l'intercept)
  matrix[n_obs, n_x] x;       // valeurs des covariables  $x_{ijk}$ 
  vector[n_obs] y;           // valeurs des marqueurs  $Y_{ijk}$ 
  vector[n_obs] time;        // temps  $t_{ijk}$ 
  int<lower = 1, upper = n_out> outcome[n_obs]; // identifiant k du marqueur
  int<lower = 1, upper = n_sub> subject[n_obs]; // identifiant i de l'individu
  int<lower = 1, upper = n_obs> sub_n_obs[n_sub];
  // nombre d'observations  $J_i$  pour chaque individu
  int<lower = 0, upper = n_sub> n_ndem; // nombre d'individus sans démenche
  vector[n_sub] dem;          // indicateur de diagnostic de démenche
  vector[n_sub] t_evt;        // temps d'évènement  $T^{\text{diag}}$  ou  $T^{\text{last}}$ 
  int id_slp[n_out];         // indicateur de pente aléatoire pour le marqueur k
  real C;                    // valeur de incertitude C autour du temps d'évènement
}
```

Le block `transformed data` permet d'organiser et de transformer les données. C'est ici que nous pouvons regrouper les données d'un même individu pour le calcul parallèle des contributions individuelles à la vraisemblance.

```
transformed data{
  vector[n_sub] L_eta; // contrainte inférieure individuelle pour eta_T_dem
  vector[n_sub] U_eta; // contrainte supérieure individuelle pour eta_T_dem

  real xr[n_sub, (n_x+2)*max(sub_n_obs) + 1];
  // données (réels) des N individus [ $x_{jk}, Y_{jk}, t_{jk}, T^{\text{diag}}/T^{\text{last}}$ ]
  int n_obs_out[n_sub, 3 + max(sub_n_obs)];

  // données (entiers) des N individus [ $J_i, K, n_x, k_j$ ]

  int end;
  int pos = 1;
```

```

for (i_sub in 1:n_sub) {          // boucle sur l'ensemble des observations

  // Cette partie regroupe les données par individu dans les structures xr et
  n_obs_out
  int n = sub_n_obs[i_sub];
  int diff = max(sub_n_obs)-n;
  int fill_int = 0;
  real fill_real = 0;
  end = pos + n - 1;
  n_obs_out[i_sub] = append_array(append_array({n, n_out, n_x},
                                             outcome[pos:end]),
                                  rep_array(fill_int, diff));
  xr[i_sub] = append_array(append_array({t_evt[i_sub]},
                                       to_array_1d(append_col(x[pos:end:],
                                                             append_col(y[pos:end],
                                                             time[pos:end])))),
                           rep_array(fill_real,
                                     diff*(n_x+2)));

  pos += n;

  // Cette partie définit les limites inférieure et supérieure de eta_T_dem en
  // fonction du type d'évènement. eta_T_dem est dans l'échelle logarithmique
  L_eta[i_sub] = log(t_evt[i_sub]+0.0001); // .0001 évite une valeur infinie
  if (dem[i_sub] == 1){
    U_eta[i_sub] = log(t_evt[i_sub]+(2*C));
  } else {
    U_eta[i_sub] = log(100);
    // 100 ans est utilisé comme limite supérieure pour les individus sans
    // diagnostic de démence afin de réduire l'espace des paramètres
  }
}
}
}

```

Le bloc `parameters` permet de définir les paramètres initiaux sur lesquels les valeurs à partir des lois de distribution a priori seront échantillonnées.

```

parameters{
  matrix[n_out, n_x] beta;          // coefficients (intercepts + covariables)
  vector<lower = 0>[n_out] gamma;    // coefficients pentes fixes
  vector<lower = 0>[n_out] sigma_e;  // variances des erreurs de mesures

  vector<lower = 0>[n_out] sigma_u0; // variances des intercepts aléatoires
  vector<lower = 0>[sum(id_slp)] sigma_u1; // variances des pentes aléatoires
  matrix[n_sub, n_out] eta_u0;      // intercepts aléatoires (standardisés)
  matrix[n_sub, sum(id_slp)] eta_u1; // pentes aléatoires (standardisées)

  real<lower = 0> mu_t_dem;           // moyenne de  $T_i^*$ 
  real<lower = 0> sigma_t_dem;       // variance de  $T_i^*$ 
  vector<lower = 0, upper = 1>[n_sub] eta_t_dem_raw;
  // recalage  $T_i^{**}$  avant transformation pour obtenir  $T_i^*$ 
}

```

Le bloc `transformed parameters` permet par exemple de transformer ou de générer de nouveaux paramètres à partir des paramètres initiaux et des données.

```

transformed parameters{
  vector[n_out*n_x + 2*n_out] phi;
  // vecteur de paramètres au niveau populationnel [beta + gamma + sigma_e]
  vector[2*n_out + 1] theta[n_sub];
}

```

```

// N vecteurs de paramètres au niveau individuel [z_u0 + z_u1 + T_dem_i]

matrix[n_sub, n_out] z_u0;          // intercepts aléatoires
matrix[n_sub, n_out] z_u1;          // pentes aléatoires

vector<offset = mu_t_dem, multiplier = sigma_t_dem>[n_sub] eta_t_dem = L_eta
  + (U_eta-L_eta) .* eta_t_dem_raw;
// eta_t_dem = recalage temporel individuel dans l'échelle log(temps)
vector[n_sub] T_dem_i = exp(eta_t_dem) - C;
// T_dem_i = recalage temporel individuel dans l'échelle du temps

{ // 1ère boucle locale pour fixer les pentes aléatoires non estimée
int j_slp = 1;
for (i in 1:n_out){
  z_u0[,i] = eta_u0[,i]*sigma_u0[i];
  if (id_slp[i] == 0){
    z_u1[,i] = rep_vector(0, n_sub);
  } else {
    z_u1[,i] = eta_u1[,j_slp]*sigma_u1[j_slp];
    j_slp += 1;
  }
}
}

{ // 2ème boucle locale pour organiser les paramètres pour le calcul parallèle
phi[1:n_out*n_x] = to_vector(beta);
phi[n_out*n_x+1 : (n_x+1)*n_out] = gamma;
phi[(n_x+1)*n_out+1 : (n_x+2)*n_out] = sigma_e;
for (t in 1:n_sub) {
  theta[t] = append_col(z_u0[t,], append_col(z_u1[t,], T_dem_i[t]));
}
}
}

```

Le bloc `model` définit toutes les valeurs qui vont incrémenter le calcul du logarithme du postérieur $P(\theta^*|X)$ avec la vraisemblance $P(X|\theta^*)$ et les lois de distribution a priori $P(\theta^*)$

```

model{
  // Lois de distribution à priori
  to_vector(beta) ~ normal(0, 10);
  for (i in 1:n_out) {
    gamma[i] ~ normal(0, 10);
    sigma_e[i] ~ cauchy(0, 2.5);
    sigma_u0[i] ~ cauchy(0, 2.5);
    to_vector(eta_u0[,i]) ~ normal(0, 1);
  }

  for (i in 1:sum(id_slp)) {
    sigma_u1[i] ~ cauchy(0, 2.5);
    to_vector(eta_u1[,i]) ~ normal(0, 1);
  }

  mu_t_dem ~ normal(10, 10);
  sigma_t_dem ~ normal(0, 2.5);
  eta_t_dem ~ normal(mu_t_dem, sigma_t_dem); //implique que T_dem_i ~ lognormal()

  // calcul de la log-vraisemblance
  target += sum(map_rect(map_ranef, phi, theta, xr, n_obs_out));
  // calcul parallèle avec map_rect() et la fonction utilisateur map_ranef()
}

```

Le bloc `generated quantities` permet de calculer une quantité à la fin de chaque itération à partir des données et des valeurs θ^* échantillonnées. Nous avons défini ici la log-vraisemblance et les valeurs prédites individuelles.

```
generated quantities{
  vector[n_obs] log_lik;
  vector[n_obs] preds;

  // Calcul de la log-vraisemblance  $\log(P(X|\theta^*))$ 
  for (i in 1:n_obs) {
    log_lik[i] = normal_lpdf(y[i] | dot_product(x[i], beta[outcome[i]])
      + gamma[outcome[i]] * (time[i]-T_dem_i[subject[i]])
      + z_u0[subject[i], outcome[i]]
      + z_u1[subject[i], outcome[i]]*(time[i]-T_dem_i[subject[i]]),
      sigma_e[outcome[i]]);
  }

  // Calcul des valeurs prédites  $\bar{Y}^l$ 
  preds[i] = dot_product(x[i], beta[outcome[i]])
    + gamma[outcome[i]] * (time[i]-T_dem_i[subject[i]])
    + z_u0[subject[i], outcome[i]]
    + z_u1[subject[i], outcome[i]]*(time[i]-T_dem_i[subject[i]]);
}
}
```

Le bloc `functions` permet de définir les fonctions utilisateurs nécessaires au programme. Ici une seule fonction est déclarée `map_ranef` pour le calcul parallèle des log-vraisemblance individuelles. Ce bloc doit apparaître en premier dans le script avant le bloc `data`.

```
functions{
  vector map_ranef(vector phi, vector theta, real[] xr, int[] n_obs_out) {

    # données -----
    int n_obs_ = n_obs_out[1];           // nombre d'observations j
    int n_out_ = n_obs_out[2];          // nombre de marqueurs K
    int n_x_ = n_obs_out[3];           // nombre de covariables

    int outcome_[n_obs_] = n_obs_out[4:(3+n_obs_)]; // identifiant k du marqueur
    matrix[n_obs_, n_x_] x_ = to_matrix(xr[2:(1+n_x_*n_obs_)], n_obs_, n_x_);
    // covariables  $x_{jk}$ 
    vector[n_obs_] y_ = to_vector(xr[(1+n_x_*n_obs_+1):(1+n_obs_*(n_x_+1))]);
    // observations  $Y_{jk}$ 
    real time_[n_obs_] = xr[(1+n_obs_*(n_x_+1)+1):(1+n_obs_*(n_x_+2))];
    // temps  $t_{jk}$ 
    real t_evt_ = xr[1];                // temps d'évènement  $T^{\text{diag}}$  ou  $T^{\text{last}}$ 

    # paramètres populationnels -----
    matrix[n_out_, n_x_] beta_ = to_matrix(phi[1:(n_out_*n_x_)], n_out_, n_x_);
    // paramètre
    vector[n_out_] gamma_ = phi[(n_out_*n_x_+1) : ((n_x_+1)*n_out_)];
    vector[n_out_] sigma_e_ = phi[((n_x_+1)*n_out_+1) : ((n_x_+2)*n_out_)];

    # paramètres individuels -----
    vector[n_out_] z_u0_ = to_vector(theta[1:n_out_]);
    // intercepts aléatoires de l'individu i
    vector[n_out_] z_u1_ = to_vector(theta[(n_out_+1):(2*n_out_)]);
    // pentes aléatoires de l'individu i
  }
}
```

```

real T_dem_i = theta[(2*n_out_)+1]; // temps de recalage temporel  $T_i^*$ 

# modèle -----
vector[n_obs_] mu; // moyenne du prédicteur
vector[n_obs_] sigma_e_aug; // variance de l'erreur de mesure
real ll; // vraisemblance individuelle

for (i in 1:n_obs_) {
  mu[i] = dot_product(x_[i], beta_[outcome_[i]])
    + gamma_[outcome_[i]] * (time_[i]-T_dem_i_)
    + z_u0_[outcome_[i]]
    + z_u1_[outcome_[i]]*(time_[i]-T_dem_i_);
  // outcome and individual specific random slope
  sigma_e_aug[i] = sigma_e_[outcome_[i]];
}
ll = normal_lpdf(y_ | mu, sigma_e_aug);

// calcul du logarithme de la densité de probabilité des observations  $Y_{jk}$ 
// pour l'individu i sachant  $\theta^*$ 
return(ll);
}
}

```

5.4 Conclusions

Le modèle que nous avons développé dans le cadre de ces travaux propose une méthodologie originale pour réaligner les trajectoires individuelles des marqueurs dans l'échelle du temps de la maladie. Les modèles de progression de la maladie sont des méthodes intéressantes pour décrire l'histoire naturelle complète d'une maladie lorsque le temps observé n'est pas pertinent. En comparaison aux autres travaux de la littérature ils ne se contentent pas d'homogénéiser le plus possible la variabilité inter-individuelle, ils utilisent l'information partiellement observée du temps de diagnostic pour estimer le temps latent de maladie, ainsi nous avons pu décrire les trajectoires des marqueurs et la séquence des évènements en fonction des années précédant le diagnostic clinique de démence.

Dans notre application à la MAMA, nous avons pu décrire les trajectoires des marqueurs dans une échelle de temps latent de la maladie qui s'étend sur 30 ans avant le diagnostic de démence clinique. D'autres analyses sont nécessaires pour mieux caractériser les trajectoires des marqueurs à l'aide de méthodes plus flexibles, pour étudier les sous-types possibles de la maladie, ainsi que les relations complexes entre les principaux facteurs de risque et l'histoire naturelle de la maladie. Une perspective de ce travail à terme est d'identifier les individus à un stade précoce, des années avant les symptômes cognitifs et la démence clinique.

6 DISCUSSION GENERALE

6.1 Résumé des travaux et analyses complémentaires

Nous avons pu montrer à l'aide d'un modèle à équations structurelles que l'association entre l'hypertension artérielle et le niveau de cognition, chez les participants de l'étude MEMENTO, était indépendante de la neuropathologie de type Alzheimer mais était en partie expliquée par la maladie des petits vaisseaux et l'atrophie du cerveau. De plus les participants avec une hypertension artérielle contrôlée par la prise d'un traitement antihypertenseur ne présentaient pas de différence significative sur le niveau de cognition comparativement aux participants sans hypertension. Les analyses complémentaires ne montraient pas de modifications significatives des chemins d'associations entre les hommes et les femmes ainsi qu'en fonction du statut *APOE*. Mais l'association entre l'hypertension non-contrôlée et le volume plus important d'HSB était observée uniquement chez les plus âgés.

Nous avons aussi développé un modèle de progression en fonction du temps latent de la maladie qui nous a permis de prendre en compte l'hétérogénéité inter-individuelle des données longitudinales pour décrire les trajectoires des marqueurs au cours de l'histoire naturelle de la maladie jusqu'au diagnostic clinique de la démence. Le plus souvent les protéines tau dans le LCR étaient les premières à changer et la dégradation des scores cognitifs arrivait en dernier. Le niveau d'études et le génotype *APOE4* exerçaient une influence importante sur la séquence des changements et plus particulièrement sur la diminution du niveau de protéine $A\beta_{42}$ dans le LCR. Les analyses complémentaires ont permis de confirmer l'intérêt de prendre en compte l'information partielle sur le diagnostic de démence clinique pour estimer le temps latent de la maladie.

6.2 Association entre l'hypertension artérielle, la maladie des petits vaisseaux, les lésions de type Alzheimer, l'atrophie du cerveau et la cognition

Dans cette thèse nous avons étudié l'association entre l'hypertension artérielle et les marqueurs de MAMA avec un SEM. Les SEM représentent des outils statistiques puissants pour étudier les interrelations d'un ensemble de variables et analyser les chemins d'associations. Ils permettent une grande liberté de spécifications d'hypothèses et de notions

conceptuelles par l'ajout de variables latentes. Nous avons pu ainsi étudier les voies d'actions de l'hypertension dans sa relation avec la fonction cognitive. Les SEM sont des outils bien adaptés à l'étude des facteurs de risque d'une maladie.

Les résultats reposent cependant sur de nombreuses hypothèses et doivent être interprétés avec précaution. En particulier l'interprétation causale des associations qui peut difficilement être établie à partir de données transversales. Pour mieux comprendre les relations causales entre l'exposition à l'hypertension et les marqueurs de progression de maladie, une perspective est d'explorer la temporalité de ces marqueurs dans le cadre d'une analyse longitudinale. Cela pourrait être réalisé par des modèles à équations structurelles dynamiques (Voelkle et al. 2012; Driver and Voelkle 2018; Tihomir Asparouhov, Hamaker, and Muthén 2018; Taddé et al. 2020) qui ont été développés récemment pour étendre les modèles SEM aux données répétées. Ce type d'approche permettrait par exemple d'étudier plus précisément les voies d'actions de l'hypertension artérielle. Par exemple nos résultats ont montré que l'association entre l'hypertension et la fonction cognitive est en partie expliquée par la maladie des petits vaisseaux. Avec une approche dynamique on pourrait étudier le rôle de l'hypertension artérielle sur le déclin cognitif et la vitesse de progression de la maladie des petits vaisseaux.

6.3 Modèle de progression de la MAMA : applications à la cohorte MEMENTO

Les travaux sur les modèles de progression ont permis de développer une approche innovante pour décrire l'histoire naturelle de la maladie. Nous avons pu étudier les trajectoires des différents marqueurs et la chronologie des changements. Nos résultats étaient en partie cohérents avec le modèle théorique de la dynamique de progression des marqueurs (Jack et al. 2010) : la chronologie des atteintes était très dépendante des caractéristiques individuelles. Les résultats de (Young et al. 2014), qui ont utilisé une approche événementielle (partie 2.4.3), dépendaient aussi des caractéristiques individuelles, en particulier l'incertitude pour l' $A\beta_{42}$ dans le LCR. En effet, leurs résultats étaient cohérents avec le modèle théorique et l'hypothèse de la cascade amyloïde uniquement chez les participants porteurs de l'*APOE4*.

Plusieurs limites et améliorations à notre modèle peuvent être suggérées. Nous avons vu l'importance de bien distinguer la variabilité inter-individuelle partagée par les marqueurs

de la variabilité inter-individuelle spécifique aux différents marqueurs pour capturer le temps latent de la maladie. La définition du temps latent pourrait être biaisée et influencée par la corrélation entre les marqueurs étudiés. Par exemple si on étudie 6 scores à des tests neuropsychologiques particulièrement corrélés entre eux et seulement 2 marqueurs d'atrophie du cerveau avec une corrélation moindre : le temps latent risque d'être principalement défini par la corrélation des scores. Ce trait latent ne sera pas nécessairement représentatif de celui que l'on cherche à estimer et qui est partagé par l'ensemble des changements anatomiques et cliniques observés au cours de la progression de la MAMA. Une solution serait de spécifier une matrice variances-covariances par blocs où les effets aléatoires seraient distribués selon une distribution multivariée propre à chaque groupe de biomarqueur.

Le rôle des caractéristiques individuelles sur les trajectoires des différents marqueurs était évalué par un effet global qui ne permet pas d'apprécier l'impact, en particulier du niveau d'études et de l'*APOE4* sur les vitesses de dégradation des marqueurs. Pour étudier l'effet des principaux facteurs de risque et facteurs protecteurs, on pourrait étendre notre modèle de progression avec des coefficients d'interaction entre les caractéristiques individuelles et les trajectoires des différents marqueurs ou bien stratifier les analyses pour comparer la chronologie de la dégradation des différents marqueurs entre sous-groupes de participants.

Enfin les mesures répétées dans l'étude MEMENTO comprenaient 1 ou 2 mesures à l'exception des tests neuropsychologiques. Lorsque d'autres données seront disponibles, il sera utile de reproduire les analyses pour obtenir des estimations potentiellement plus précises du temps latent de maladie, avec en particulier les acquisitions IRM réalisées 4 ans après l'inclusion dans l'étude. Il serait également intéressant de comparer l'approche par les percentiles que nous choisis pour l'échelle de sévérité des marqueurs avec l'utilisation d'une fonction logistique pour obtenir une trajectoire sigmoïde (Jedynak et al. 2012). On pourrait aussi envisager des trajectoires plus souples notamment par des approches non paramétriques.

6.4 Extrapolation des résultats à la population générale

Les données de la cohorte MEMENTO sont recueillies dans un contexte clinique et portent sur un échantillon qui n'est pas représentatif de la population générale, les résultats

des travaux de la thèse ne sont donc pas généralisables sans discuter des biais associés à la sélection des participants.

Dans les travaux qui ont permis d'estimer la durée des différents stades cliniques (présentés en introduction partie 2.4.2) les auteurs ont mis en évidence que les estimations de durée de la maladie étaient plus courtes pour les participants de cohortes cliniques par rapport aux participants de cohortes en population générale. Cette différence pourrait être due au fait que les participants des études cliniques sont recrutés à des stades de progression plus avancés (Vermunt et al. 2019). Similairement il a été montré que les participants avec une plainte cognitive subjective ont une probabilité plus importante d'évoluer vers un trouble cognitif léger lorsqu'ils sont recrutés dans le cadre d'une cohorte clinique comparativement aux participants recrutés en population générale (Snitz et al. 2018). Enfin une précédente analyse incluant les données de l'étude MEMENTO et portant sur l'analyse de l'association entre sexe/genre et l'évolution des scores cognitifs recueillis dans 8 cohortes retrouvait des résultats différents entre les cohortes cliniques et celles en population générale (Rouanet et al. 2022).

Cependant, les cohortes cliniques permettent de recueillir de nombreuses mesures répétées de données multimodales en particulier les biomarqueurs d'imagerie cérébrale et de neuropathologie de type Alzheimer qui sont des données complexes à recueillir en population générale pour étudier l'histoire naturelle de la maladie. Les auteurs des publications scientifiques dans le domaine des modèles de progression de la maladie utilisent quasi exclusivement les données de la cohorte ADNI, hors le protocole de l'étude implique le recrutement de participants dans une population très sélectionnée : par exemple pour ADNI1 la population d'étude était constituée de 229 participants sans trouble cognitifs, 398 participants avec un trouble cognitif léger et 192 participants avec un diagnostic de démence (Petersen et al. 2010). La contribution des travaux de la thèse dans le domaine des modèles de progression de maladie apporte des résultats sur une population moins sélectionnée et plus homogène. Les données de la cohorte Memento sont probablement plus représentatives de la population générale que les données de l'étude ADNI.

Dans le cadre de notre modèle de progression de la maladie une perspective intéressante serait de comparer les résultats obtenus avec les données de la cohorte Memento aux résultats qu'on obtiendrait avec les données de la cohorte ADNI. Une perspective à plus long terme est le développement de biomarqueurs sanguins actuellement

à l'étude (Teunissen et al. 2022). S'ils s'avèrent utiles pour étudier la progression de la maladie, ils pourraient permettre d'appliquer des modèles de progression de la maladie dans des cohortes en population générale qui ont souvent l'avantage d'être plus larges et moins sélectionnées.

6.5 Implications des facteurs de risques vasculaires

Les résultats des travaux de la thèse sur l'implication de l'hypertension et de la maladie des petits vaisseaux dans l'histoire naturelle des démences sont en faveur d'une contribution au déclin cognitif distincte du rôle des lésions de type Alzheimer. Nous avons trouvé que l'association entre l'hypertension artérielle lorsqu'elle n'est pas contrôlée et le volume des HSB plus important concernait les participants les plus âgés, ce résultat pourrait s'expliquer par l'augmentation de la prévalence des lésions vasculaires cérébrales en fonction de l'âge (Nelson et al. 2011).

Nous avons considéré la neuropathologie de type Alzheimer par la quantification des protéines amyloïde au scanner TEP amyloïde et par le niveau de protéines p-tau et $A\beta_{42}/A\beta_{40}$ dans le LCR. Les résultats du modèle de progression sur l'incertitude de la séquence chronologique de $A\beta_{42}$ comparativement aux protéines p-tau et t-tau (qui étaient principalement les premiers marqueurs dégradés au cours du temps de la maladie) pose question sur la pertinence de regrouper les protéines tau et amyloïde dans la variable latente neuropathologie de type Alzheimer telle qu'elle a été définie. De plus une précédente étude a mis en évidence l'absence d'association entre l'hypertension artérielle et le niveau d' $A\beta_{42}$ mais une association significative entre l'hypertension artérielle et les niveaux de protéines p-tau et t-tau (I. Bos et al. 2019).

La recherche sur la contribution vasculaire au trouble cognitif et à la démence a pour objectif de comprendre, prévenir et traiter les effets délétères des neuropathologies vasculaires sur le cerveau et les fonctions cognitives (Zlokovic et al. 2020). Le mécanisme d'action de la maladie des petits vaisseaux sur le déclin cognitif est insuffisamment compris mais plusieurs études suggèrent que l'atteinte de l'intégrité des artérioles, des veinules et des capillaires ne permet pas de maintenir un débit sanguin cérébral suffisant et une barrière hémato-encéphalique fonctionnelle pour répondre aux besoins des tissus du cerveau (Alber et al. 2019).

La prévention des facteurs de risque vasculaires reste la principale stratégie pour éviter ou traiter les conséquences cognitives des anomalies vasculaires. Nous avons vu précédemment que plusieurs études ont montré le rôle délétère de l'hypertension artérielle au cours de la vie lorsque qu'elle n'est pas bien contrôlée par la prise d'un traitement antihypertenseur. Mais la prévention des facteurs de risque vasculaires comprend aussi un ensemble de mesures hygiéno-diététiques. L'étude FINGER (Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability) est un essai clinique interventionnel randomisé dont les résultats suggèrent qu'une alimentation équilibrée, l'exercice physique, l'exercice mental et un contrôle effectif de l'hypertension artérielle pendant 2 ans permet de maintenir ou améliorer la fonction cognitive (Ngandu et al. 2015). Cependant cet essai clinique n'a pas montré d'effets significatifs sur le volume des HSB (Stephen et al. 2019).

Nous avons étudié la maladie des petits vaisseaux par l'analyse du volume des HSB uniquement. D'autres analyses permettraient d'étudier le rôle de l'hypertension artérielle avec les autres marqueurs de la maladie des petits vaisseaux comme les lacunes, les micro-saignements, les infarctus silencieux et les espaces périvasculaires. Par ailleurs l'hypertension artérielle n'est pas le seul facteur de risque vasculaire, par exemple : le diabète, le surpoids, le tabagisme, l'inactivité physique, on pourrait analyser simultanément les chemins d'association d'un ensemble de facteurs de risque avec la cognition et les marqueurs de MAMA.

6.6 Conclusion générale

Les travaux de recherche réalisés dans le cadre de cette thèse ont permis d'étudier l'histoire naturelle des démences et l'implication des facteurs de risques vasculaires par des approches multimodales en s'appuyant sur des méthodes statistiques originales pour analyser les données de marqueurs biologiques, d'imagerie du cerveau et de mesures neuropsychologiques qui ont été recueillies dans le cadre de la cohorte clinique MEMENTO.

D'une part, nous avons pu contribuer à mieux comprendre les chemins d'association dans la relation entre la cognition et l'hypertension artérielle. Ainsi que l'importance du contrôle de l'hypertension artérielle qui pourrait contribuer à la prévention de la démence.

D'autre part, nous avons pu décrire les trajectoires à long terme des marqueurs de MAMA pendant les années précédant le diagnostic clinique de la démence, soulignant

l'incertitude sur la séquence chronologique des changements anatomiques et cliniques et le rôle des caractéristiques individuelles sur l'histoire naturelle de la maladie.

7 REFERENCES

- Alber, Jessica, Suvarna Alladi, Hee Joon Bae, David A. Barton, Laurel A. Beckett, Joanne M. Bell, Sara E. Berman, et al. 2019. "White Matter Hyperintensities in Vascular Contributions to Cognitive Impairment and Dementia (VCID): Knowledge Gaps and Opportunities." *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 5: 107–17. <https://doi.org/10.1016/j.trci.2019.02.001>.
- Alpérovitch, Annick, P. Amouyel, J. F. Dartigues, P. Ducimetière, B. Mazoyer, K. Ritchie, C. Tzourio, et al. 2003. "Vascular Factors and Risk of Dementia: Design of the Three-City Study and Baseline Characteristics of the Study Population." *Neuroepidemiology* 22 (6): 316–25. <https://doi.org/10.1159/000072920>.
- American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders - Text Revision (DSM-IV-TR)*. American Psychiatric Association.
- . 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-V*. American psychiatric association.
- Archetti, Damiano, Silvia Ingala, Vikram Venkatraghavan, Viktor Wottschel, Alexandra L. Young, M. Bellio, Esther E. Bron, et al. 2019. "Multi-Study Validation of Data-Driven Disease Progression Models to Characterize Evolution of Biomarkers in Alzheimer's Disease." *NeuroImage: Clinical* 24 (January). <https://doi.org/10.1016/j.nicl.2019.101954>.
- Arenaza-Urquijo, Eider M., and Prashanthi Vemuri. 2018. "Resistance vs Resilience to Alzheimer Disease." *Neurology* 90 (15): 695–703. <https://doi.org/10.1212/WNL.0000000000005303>.
- Aronow, Wilbert S, Jerome L Fleg, Carl J Pepine, Nancy T Artinian, George Bakris, Alan S Brown, Keith C Ferdinand, et al. 2011. "ACCF/AHA 2011 Expert Consensus Document on Hypertension in the Elderly: A Report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents Developed in Collaboration With the American Academy of Neurology, American." *Journal of the American College of Cardiology* 57 (20): 2037–2114. <https://doi.org/https://doi.org/10.1016/j.jacc.2011.01.008>.
- Artbuckle, J L. 2019. "Amos (Version 26.0)." Chicago : IBM SPSS.
- Ashburner, John, and Karl J. Friston. 2005. "Unified Segmentation." *NeuroImage* 26 (3): 839–51. <https://doi.org/10.1016/j.neuroimage.2005.02.018>.
- Asparouhov, T, and B Muthén. 2005. "Multivariate Statistical Modeling with Survey Data." *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, 14–16.
- Asparouhov, Tihomir, Ellen L. Hamaker, and Bengt Muthén. 2018. "Dynamic Structural Equation Models." *Structural Equation Modeling* 25 (3): 359–88. <https://doi.org/10.1080/10705511.2017.1406803>.

- Bateman, Randall J., Chengjie Xiong, Tammie L.S. Benzinger, Anne M. Fagan, Alison Goate, Nick C. Fox, Daniel S. Marcus, et al. 2012. "Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease." *New England Journal of Medicine* 367 (9): 795–804. <https://doi.org/10.1056/nejmoa1202753>.
- Bayes, Thomas, and Richard Price. 1763. "LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S." *Philosophical Transactions of the Royal Society of London* 53 (January): 370–418. <https://doi.org/10.1098/rstl.1763.0053>.
- Beach, Thomas G., Sarah E. Monsell, Leslie E. Phillips, and Walter Kukull. 2012. "Accuracy of the Clinical Diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010." *Journal of Neuropathology and Experimental Neurology* 71 (4): 266–73. <https://doi.org/10.1097/NEN.0b013e31824b211b>.
- Bentler, P. M. 1990. "Comparative Fit Indexes in Structural Models." *Psychological Bulletin* 107 (2): 238.
- . 1995. "EQS Structural Equations Program Manual." Multivariate software Encino, CA.
- Bentler, P. M., and Douglas G. Bonett. 1980. "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures." *Psychological Bulletin* 88 (3). <https://doi.org/10.1037/0033-2909.88.3.588>.
- Betancourt, Michael. 2017. "A Conceptual Introduction to Hamiltonian Monte Carlo," January. <http://arxiv.org/abs/1701.02434>.
- Bilgel, Murat, and Bruno M. Jedynek. 2019. "Predicting Time to Dementia Using a Quantitative Template of Disease Progression." *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring* 11: 205–15. <https://doi.org/10.1016/j.dadm.2019.01.005>.
- Bilgel, Murat, Rebecca L. Kosciak, Yang An, Jerry L. Prince, Susan M. Resnick, Sterling C. Johnson, and Bruno M. Jedynek. 2017. "Temporal Order of Alzheimer's Disease-Related Cognitive Marker Changes in BLSA and WRAP Longitudinal Studies." *Journal of Alzheimer's Disease* 59 (4): 1335–47. <https://doi.org/10.3233/JAD-170448>.
- Bilgel, Murat, Jerry L. Prince, Dean F. Wong, Susan M. Resnick, and Bruno M. Jedynek. 2016. "A Multivariate Nonlinear Mixed Effects Model for Longitudinal Image Analysis: Application to Amyloid Imaging." *NeuroImage* 134: 658–70. <https://doi.org/10.1016/j.neuroimage.2016.04.001>.
- Borenstein, Amy R, and James A Mortimer. 2016a. "Clinical Appearance, Progression, and Classification." In *Alzheimer's Disease Life Course Perspectives on Risk Reduction*, 7–23. San Diego: Academic Press.
- . 2016b. "The 'First' Case." In *Alzheimer's Disease Life Course Perspectives on Risk Reduction*, 3–

5. San Diego: Academic Press.

- Bos, Daniel, Frank J. Wolters, Sirwan K.L. Darweesh, Meike W. Vernooij, Frank de Wolf, M. Arfan Ikram, and Albert Hofman. 2018. "Cerebral Small Vessel Disease and the Risk of Dementia: A Systematic Review and Meta-Analysis of Population-Based Evidence." *Alzheimer's and Dementia* 14 (11): 1482–92. <https://doi.org/10.1016/j.jalz.2018.04.007>.
- Bos, Isabelle, Stephanie J.B. Vos, Suzanne E. Schindler, Jason Hassenstab, Chengjie Xiong, Elizabeth Grant, Frans Verhey, John C. Morris, Pieter Jelle Visser, and Anne M. Fagan. 2019. "Vascular Risk Factors Are Associated with Longitudinal Changes in Cerebrospinal Fluid Tau Markers and Cognition in Preclinical Alzheimer's Disease." *Alzheimer's and Dementia* 15 (9): 1149–59. <https://doi.org/10.1016/j.jalz.2019.04.015>.
- Boyle, Patricia A., Lei Yu, Robert S. Wilson, Sue E. Leurgans, Julie A. Schneider, and David A. Bennett. 2018. "Person-Specific Contribution of Neuropathologies to Cognitive Loss in Old Age." *Annals of Neurology* 83 (1): 74–83. <https://doi.org/10.1002/ana.25123>.
- Braak, H, and E Braak. 1991. "Neuropathological Staging of Alzheimer-Related Changes." *Acta Neuropathologica* 82 (4): 239–59. <https://doi.org/10.1007/BF00308809>.
- Brewer, Sarah. 2011. *The Human Body: A Visual Guide to Human Anatomy*. Quercus.
- Browne, Michael W., and Robert Cudeck. 1992. "Alternative Ways of Assessing Model Fit." *Sociological Methods & Research* 21 (2). <https://doi.org/10.1177/0049124192021002005>.
- Buchert, R, F Wilke, B Chakrabarti, B Martin, W Brenner, J Mester, and M Clausen. 2005. "Adjusted Scaling of FDG Positron Emission Tomography Images for Statistical Evaluation in Patients with Suspected Alzheimer's Disease." *J Neuroimaging* 15 (4): 348–55. <https://doi.org/10.1177/1051228405280169>.
- Caroli, A, and G B Frisoni. 2010. "The Dynamics of Alzheimer's Disease Biomarkers in the Alzheimer's Disease Neuroimaging Initiative Cohort." *Neurobiology of Aging* 31 (8): 1263–74. <https://doi.org/10.1016/j.neurobiolaging.2010.04.024>.
- Cheong, JeeWon, and David P. MacKinnon. 2012. "Mediation/Indirect Effects in Structural Equation Modeling." In *Handbook of Structural Equation Modeling.*, edited by Rick H. Hoyle, 417–35. New York, NY, US: The Guilford Press.
- Chib, Siddhartha, and Edward Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *American Statistician* 49 (4): 327. <https://doi.org/10.2307/2684568>.
- Chui, Helena C., and Liliana Ramirez-Gomez. 2015. "Clinical and Imaging Features of Mixed Alzheimer and Vascular Pathologies." *Alzheimer's Research and Therapy*. BioMed Central Ltd. <https://doi.org/10.1186/s13195-015-0104-7>.
- Chupin, M., A. Hammers, R. S.N. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux. 2009. "Automatic Segmentation of the Hippocampus and the Amygdala Driven by

- Hybrid Constraints: Method and Validation.” *NeuroImage* 46 (3): 749–61. <https://doi.org/10.1016/j.neuroimage.2009.02.013>.
- Clark, Christopher M., Julie A. Schneider, Barry J. Bedell, Thomas G. Beach, Warren B. Bilker, Mark A. Mintun, Michael J. Pontecorvo, et al. 2011. “Use of Florbetapir-PET for Imaging β -Amyloid Pathology.” *Jama* 305 (3): 275–83. <https://doi.org/10.1001/jama.2010.2008>.
- Commenges, Daniel, and H el ene Jacqmin-Gadda. 2015. *Mod eles Biostatistiques Pour l’ pid miologie*. De Boeck Sup erieur.
- Corbo, R. M., and R. Scacchp. 1999. “Apolipoprotein E (APOE) Allele Distribution in the World. Is APOE*4 a ‘thrifty’ Allele?” *Annals of Human Genetics* 63 (4): 301–10. <https://doi.org/10.1017/S0003480099007575>.
- Cronbach, Lee J, and Paul E Meehl. 1955. “Construct Validity in Psychological Tests.” *Psychological Bulletin* 52 (4): 281–302. <https://doi.org/10.1037/h0040957>.
- Debette, St ephanie, and H. S. Markus. 2010. “The Clinical Importance of White Matter Hyperintensities on Brain Magnetic Resonance Imaging: Systematic Review and Meta-Analysis.” *BMJ (Online)* 341 (7767): 288. <https://doi.org/10.1136/bmj.c3666>.
- Debette, St ephanie, Sabrina Schilling, Marie-Gabrielle Duperron, Susanna C Larsson, and Hugh S Markus. 2019. “Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-Analysis.” *JAMA Neurology* 76 (1): 81–94. <https://doi.org/10.1001/jamaneurol.2018.3122>.
- Debette, St ephanie, S. Seshadri, A. Beiser, R. Au, J. J. Himali, C. Palumbo, P. A. Wolf, and C. DeCarli. 2011. “Midlife Vascular Risk Factor Exposure Accelerates Structural Brain Aging and Cognitive Decline.” *Neurology* 77 (5): 461–68. <https://doi.org/10.1212/WNL.0b013e318227b227>.
- DeCarli, Charles, Joseph Massaro, Danielle Harvey, John Hald, Mats Tullberg, Rhoda Au, Alexa Beiser, Ralph D’Agostino, and Philip A. Wolf. 2005. “Measures of Brain Morphology and Infarction in the Framingham Heart Study: Establishing What Is Normal.” *Neurobiology of Aging* 26 (4): 491–510. <https://doi.org/10.1016/j.neurobiolaging.2004.05.004>.
- Desikan, R S, F Segonne, B Fischl, B T Quinn, B C Dickerson, D Blacker, R L Buckner, et al. 2006. “An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans into Gyral Based Regions of Interest.” *Neuroimage* 31 (3): 968–80. <https://doi.org/10.1016/j.neuroimage.2006.01.021>.
- Dickerson, Bradford C., Akram Bakkour, David H. Salat, Eric Feczko, Jenni Pacheco, Douglas N. Greve, Fran Grodstein, et al. 2009. “The Cortical Signature of Alzheimer’s Disease: Regionally Specific Cortical Thinning Relates to Symptom Severity in Very Mild to Mild AD Dementia and Is Detectable in Asymptomatic Amyloid-Positive Individuals.” *Cerebral Cortex* 19 (3): 497–510. <https://doi.org/10.1093/cercor/bhn113>.

- Ding, Jie, Kendra L. Davis-Plourde, Sanaz Sedaghat, Phillip J. Tully, Wanmei Wang, Caroline Phillips, Matthew P. Pase, et al. 2019. "Antihypertensive Medications and Risk for Incident Dementia and Alzheimer's Disease: A Meta-Analysis of Individual Participant Data from Prospective Cohort Studies." *The Lancet Neurology* 19 (1). [https://doi.org/10.1016/S1474-4422\(19\)30393-X](https://doi.org/10.1016/S1474-4422(19)30393-X).
- Donohue, Michael C., Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G. Thomas, Rema Raman, Anthony C. Gamst, Laurel A. Beckett, et al. 2014. "Estimating Long-Term Multivariate Progression from Short-Term Data." *Alzheimer's and Dementia* 10 (5): S400–410. <https://doi.org/10.1016/j.jalz.2013.10.003>.
- Driver, Charles C, and Manuel C Voelkle. 2018. "Hierarchical Bayesian Continuous Time Dynamic Modeling." *Psychological Methods* 23 (4): 774–99. <https://doi.org/10.1037/met0000168>.
- Duane, Simon, A D Kennedy, Brian J Pendleton, and Duncan Roweth. 1987. "Hybrid Monte Carlo." *Physics Letters B* 195 (2): 216–22. [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X).
- Dubois, Bruno, Howard H. Feldman, Claudia Jacova, Steven T. DeKosky, Pascale Barberger-Gateau, Jeffrey Cummings, André Delacourte, et al. 2007. "Research Criteria for the Diagnosis of Alzheimer's Disease: Revising the NINCDS-ADRDA Criteria." *Lancet Neurology* 6 (8): 734–46. [https://doi.org/10.1016/S1474-4422\(07\)70178-3](https://doi.org/10.1016/S1474-4422(07)70178-3).
- Duering, Marco, Ruthger Righart, Frank Arne Wollenweber, Vera Zietemann, Benno Gesierich, and Martin Dichgans. 2015. "Acute Infarcts Cause Focal Thinning in Remote Cortex via Degeneration of Connecting Fiber Tracts." *Neurology* 84 (16): 1685–92. <https://doi.org/10.1212/WNL.0000000000001502>.
- Dufouil, C., John Chalmers, Oguzhan Coskun, Véronique Besançon, Marie Germaine Bousser, Pierre Guillon, Stephen MacMahon, et al. 2005. "Effects of Blood Pressure Lowering on Cerebral White Matter Hyperintensities in Patients with Stroke: The PROGRESS (Perindopril Protection Against Recurrent Stroke Study) Magnetic Resonance Imaging Substudy." *Circulation* 112 (11): 1644–50. <https://doi.org/10.1161/CIRCULATIONAHA.104.501163>.
- Dufouil, C., A. De Kersaint-Gilly, V. Besançon, C. Levy, E. Auffray, L. Brunnereau, A. Alperovitch, and C. Tzourio. 2001. "Longitudinal Study of Blood Pressure and White Matter Hyperintensities: The EVA MRI Cohort." *Neurology* 56 (7): 921–26. <https://doi.org/10.1212/WNL.56.7.921>.
- Dufouil, C, Bruno Dubois, Bruno Vellas, Florence Pasquier, Frédéric Blanc, Jacques Hugon, Olivier Hanon, et al. 2017. "Cognitive and Imaging Markers in Non-Demented Subjects Attending a Memory Clinic: Study Design and Baseline Findings of the MEMENTO Cohort." *Alzheimer's Research and Therapy* 9 (1): 1–13. <https://doi.org/10.1186/s13195-017-0288-0>.
- Duyckaerts, Charles, Benoît Delatour, and Marie Claude Potier. 2009. "Classification and Basic Pathology of Alzheimer Disease." *Acta Neuropathologica* 118 (1): 5–36. <https://doi.org/10.1007/s00401-009-0532-1>.

- Eisenberg, Dan T.A., Christopher W. Kuzawa, and M. Geoffrey Hayes. 2010. "Worldwide Allele Frequencies of the Human Apolipoprotein E Gene: Climate, Local Adaptations, and Evolutionary History." *American Journal of Physical Anthropology* 143 (1): 100–111. <https://doi.org/10.1002/ajpa.21298>.
- Ellis, Kathryn A., Ashley I. Bush, David Darby, Daniela De Fazio, Jonathan Foster, Peter Hudson, Nicola T. Lautenschlager, et al. 2009. "The Australian Imaging, Biomarkers and Lifestyle (AIBL) Study of Aging: Methodology and Baseline Characteristics of 1112 Individuals Recruited for a Longitudinal Study of Alzheimer's Disease." *International Psychogeriatrics* 21 (4): 672–87. <https://doi.org/10.1017/S1041610209009405>.
- Enders, Craig K. 2006. "Analyzing Structural Equation Models with Missing Data." In *Structural Equation Modeling: A Second Course*, edited by Gregory R. Hancock and Ralph O. Mueller, 2:315–44.
- Epelbaum, Stephane, Vincent Bouteloup, Jean F. Mangin, Valentina La Corte, Raffaella Migliaccio, Hugo Bertin, Marie O. Habert, et al. 2018. "Neural Correlates of Episodic Memory in the Memento Cohort." *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 4 (January): 224–33. <https://doi.org/10.1016/j.trci.2018.03.010>.
- Ewers, M, K Buerger, S J Teipel, P Scheltens, J Schröder, R P Zinkowski, F H Bouwman, et al. 2007. "Multicenter Assessment of CSF-Phosphorylated Tau for the Prediction of Conversion of MCI." *Neurology* 69 (24): 2205. <https://doi.org/10.1212/01.wnl.0000286944.22262.ff>.
- Finkbeiner, Carl. 1979. "Estimation for the Multiple Factor Model When Data Are Missing." *Psychometrika* 44 (4): 409–20. <https://doi.org/10.1007/BF02296204>.
- Fischer, Daniela, Marco D. Mukrasch, Jacek Biernat, Stefan Bibow, Martin Blackledge, Christian Griesinger, Eckhard Mandelkow, and Markus Zweckstetter. 2009. "Conformational Changes Specific for Pseudophosphorylation at Serine 262 Selectively Impair Binding of Tau to Microtubules." *Biochemistry* 48 (42): 10047–55. <https://doi.org/10.1021/bi901090m>.
- Fischl, Bruce, André Van Der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H. Salat, Evelina Busa, et al. 2004. "Automatically Parcellating the Human Cerebral Cortex." *Cerebral Cortex* 14 (1): 11–22. <https://doi.org/10.1093/cercor/bhg087>.
- Fonteiijn, Hubert M., Matthew J. Clarkson, Marc Modat, Josephine Barnes, Manja Lehmann, Sebastien Ourselin, Nick C. Fox, and Daniel C. Alexander. 2011. "An Event-Based Disease Progression Model and Its Application to Familial Alzheimer's Disease." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6801 LNCS: 748–59. https://doi.org/10.1007/978-3-642-22092-0_61.
- Gakidou, Emmanuela, Ashkan Afshin, Amanuel Alemu Abajobir, Kalkidan Hassen Abate, Cristiana Abbafati, Kaja M. Abbas, Foad Abd-Allah, et al. 2017. "Global, Regional, and National Comparative Risk Assessment of 84 Behavioural, Environmental and Occupational, and Metabolic Risks or

- Clusters of Risks, 1990-2016: A Systematic Analysis for the Global Burden of Disease Study 2016.” *The Lancet* 390 (10100): 1345–1422. [https://doi.org/10.1016/S0140-6736\(17\)32366-8](https://doi.org/10.1016/S0140-6736(17)32366-8).
- Garbarino, Sara, and Marco Lorenzi. 2021. “Investigating Hypotheses of Neurodegeneration by Learning Dynamical Systems of Protein Propagation in the Brain.” *NeuroImage* 235 (April): 117980. <https://doi.org/10.1016/j.neuroimage.2021.117980>.
- Gatz, Margaret, Chandra A. Reynolds, Laura Fratiglioni, Boo Johansson, James A. Mortimer, Stig Berg, Amy Fiske, and Nancy L. Pedersen. 2006. “Role of Genes and Environments for Explaining Alzheimer Disease.” *Archives of General Psychiatry* 63 (2): 168–74. <https://doi.org/10.1001/archpsyc.63.2.168>.
- Gauthier S, Rosa-Neto P, Morais JA, & Webster C. 2021. “World Alzheimer Report 2021: Journey through the Diagnosis of Dementia.” London, England.
- Gelman, Andrew, and Donald B Rubin. 1992. “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science* 7 (4): 457–72. <https://doi.org/10.1214/ss/1177011136>.
- Geman, S, and D Geman. 1984. “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6 (6): 721–41. <https://doi.org/10.1109/TPAMI.1984.4767596>.
- Genin, E., D. Hannequin, D. Wallon, K. Sleegers, M. Hiltunen, O. Combarros, M. J. Bullido, et al. 2011. “APOE and Alzheimer Disease: A Major Gene with Semi-Dominant Inheritance.” *Molecular Psychiatry* 16 (9): 903–7. <https://doi.org/10.1038/mp.2011.52>.
- Godin, Ophélie, Christophe Tzourio, Pauline Maillard, Bernard Mazoyer, and Carole Dufouil. 2011. “Antihypertensive Treatment and Change in Blood Pressure Are Associated with the Progression of White Matter Lesion Volumes: The Three-City (3C)-Dijon Magnetic Resonance Imaging Study.” *Circulation* 123 (3): 266–73. <https://doi.org/10.1161/CIRCULATIONAHA.110.961052>.
- Gosche, K M, J A Mortimer, C D Smith, W R Markesbery, and D A Snowdon. 2002. “Hippocampal Volume as an Index of Alzheimer Neuropathology.” *Neurology (Ovid)* 58 (10): 1476–82.
- Gouw, Alida A., Alexandra Seewann, Wiesje M. Van Der Flier, Frederik Barkhof, Annemieke M. Rozemuller, Philip Scheltens, and Jeroen J.G. Geurts. 2011. “Heterogeneity of Small Vessel Disease: A Systematic Review of MRI and Histopathology Correlations.” *Journal of Neurology, Neurosurgery and Psychiatry* 82 (2): 126–35. <https://doi.org/10.1136/jnnp.2009.204685>.
- Grasset, Leslie, Carol Brayne, Pierre Joly, H el ene Jacqmin-Gadda, Karine Peres, Alexandra Foubert-Samier, Jean Fran ois Dartigues, and Catherine Helmer. 2016. “Trends in Dementia Incidence: Evolution over a 10-Year Period in France.” *Alzheimer’s and Dementia* 12 (3): 272–80. <https://doi.org/10.1016/j.jalz.2015.11.001>.
- Greenland, Sander, Judea Pearl, and James M. Robins. 1999. “Causal Diagrams for Epidemiologic Research.” *Epidemiology* 10 (1): 37–48.

- Grober, E, H Buschke, H Crystal, S Bang, and R Dresner. 1988. "Screening for Dementia by Memory Testing." *Neurology* 38 (6): 900–903. <https://doi.org/10.1212/wnl.38.6.900>.
- Guerrero, R., A. Schmidt-Richberg, C. Ledig, T. Tong, R. Wolz, and D. Rueckert. 2016. "Instantiated Mixed Effects Modeling of Alzheimer's Disease Markers." *NeuroImage* 142: 113–25. <https://doi.org/10.1016/j.neuroimage.2016.06.049>.
- Habert, M O, S Marie, H Bertin, M Reynal, J B Martini, M Diallo, A Kas, Cati Nuclear Medicine network, and R Trebossen. 2016. "Optimization of Brain PET Imaging for a Multicentre Trial: The French CATI Experience." *EJNMMI Phys* 3 (1): 6. <https://doi.org/10.1186/s40658-016-0141-8>.
- Hachinski, Vladimir C, Paul Potter, and Harold Merskey. 1987. "Leuko-Araiosis." *Archives of Neurology* 44 (1): 21–23. <https://doi.org/10.1001/archneur.1987.00520130013009>.
- Hansson, Oskar. 2021. "Biomarkers for Neurodegenerative Diseases." *Nature Medicine* 27 (6): 954–63. <https://doi.org/10.1038/s41591-021-01382-x>.
- Hansson, Oskar, Sylvain Lehmann, Markus Otto, Henrik Zetterberg, and Piotr Lewczuk. 2019. "Advantages and Disadvantages of the Use of the CSF Amyloid β (A β) 42/40 Ratio in the Diagnosis of Alzheimer's Disease." *Alzheimer's Research & Therapy* 11 (1). <https://doi.org/10.1186/s13195-019-0485-0>.
- HAS - Service des bonnes pratiques professionnelles. 2016. "Prise En Charge de l'hypertension Artérielle de l'adulte." Saint-Denis La Plaine.
- Hastings, W K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57 (1): 97–109. <https://doi.org/10.1093/biomet/57.1.97>.
- Hayduk, Leslie. 2018. "Review Essay on Rex B. Kline's Principles and Practice of Structural Equation Modeling: Encouraging a Fifth Edition." *Canadian Studies in Population* 45 (3–4): 154. <https://doi.org/10.25336/csp29397>.
- Hoffman, Matthew D., and Andrew Gelman. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15: 1593–1623.
- Holst, Klaus Kähler, and Esben Budtz-Jørgensen. 2013. "Linear Latent Variable Models: The Lava-Package." *Computational Statistics* 28 (4): 1385–1452. <https://doi.org/10.1007/s00180-012-0344-y>.
- Hoyle, Rick H. 2012. "Model Specification in Structural Equation Modeling." In *Handbook of Structural Equation Modeling.*, edited by Rick H. Hoyle, 126–44. New York, NY, US: The Guilford Press.
- Hu, Li Tze, and Peter M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives." *Structural Equation Modeling* 6 (1): 1–55. <https://doi.org/10.1080/10705519909540118>.
- Iadecola, Costantino, Kristine Yaffe, José Biller, Lisa C. Bratzke, Frank M. Faraci, Philip B. Gorelick, Martha Gulati, et al. 2016. "Impact of Hypertension on Cognitive Function: A Scientific Statement

- from the American Heart Association." *Hypertension* 68 (6): e67–94. <https://doi.org/10.1161/HYP.0000000000000053>.
- Jack, Clifford R., David A. Bennett, Kaj Blennow, Maria C. Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M. Holtzman, et al. 2018. "NIA-AA Research Framework: Toward a Biological Definition of Alzheimer's Disease." *Alzheimer's and Dementia* 14 (4): 535–62. <https://doi.org/10.1016/j.jalz.2018.02.018>.
- Jack, Clifford R., David S. Knopman, William J. Jagust, Ronald C. Petersen, Michael W. Weiner, Paul S. Aisen, Leslie M. Shaw, et al. 2013. "Tracking Pathophysiological Processes in Alzheimer's Disease: An Updated Hypothetical Model of Dynamic Biomarkers." *The Lancet Neurology* 12 (2): 207–16. [https://doi.org/10.1016/S1474-4422\(12\)70291-0](https://doi.org/10.1016/S1474-4422(12)70291-0).
- Jack, Clifford R., David S. Knopman, William J. Jagust, Leslie M. Shaw, Paul S. Aisen, Michael W. Weiner, Ronald C. Petersen, and John Q. Trojanowski. 2010. "Hypothetical Model of Dynamic Biomarkers of the Alzheimer's Pathological Cascade." *The Lancet Neurology* 9 (1): 119–28. [https://doi.org/10.1016/S1474-4422\(09\)70299-6](https://doi.org/10.1016/S1474-4422(09)70299-6).
- Jack, Clifford R., Val J. Lowe, Stephen D. Weigand, Heather J. Wiste, Matthew L. Senjem, David S. Knopman, Maria M. Shiung, et al. 2009. "Serial PIB and MRI in Normal, Mild Cognitive Impairment and Alzheimer's Disease: Implications for Sequence of Pathological Events in Alzheimer's Disease." *Brain* 132 (5): 1355–65. <https://doi.org/10.1093/brain/awp062>.
- Jack, Clifford R., Prashanthi Vemuri, Heather J. Wiste, Stephen D. Weigand, Timothy G. Lesnick, Val Lowe, Kejal Kantarci, et al. 2012. "Shapes of the Trajectories of 5 Major Biomarkers of Alzheimer Disease." *Archives of Neurology* 69 (7): 856–67. <https://doi.org/10.1001/archneurol.2011.3405>.
- Jansen, Iris E., Jeanne E. Savage, Kyoko Watanabe, Julien Bryois, Dylan M. Williams, Stacy Steinberg, Julia Sealock, et al. 2019. "Genome-Wide Meta-Analysis Identifies New Loci and Functional Pathways Influencing Alzheimer's Disease Risk." *Nature Genetics* 51 (3): 404–13. <https://doi.org/10.1038/s41588-018-0311-9>.
- Jedynak, Bruno M., Andrew Lang, Bo Liu, Elyse Katz, Yanwei Zhang, Bradley T. Wyman, David Raunig, C. Pierre Jedynak, Brian Caffo, and Jerry L. Prince. 2012. "A Computational Neurodegenerative Disease Progression Score: Method and Results with the Alzheimer's Disease Neuroimaging Initiative Cohort." *NeuroImage* 63 (3): 1478–86. <https://doi.org/10.1016/j.neuroimage.2012.07.059>.
- Johnson, Jeremy C.S., Laura McWhirter, Chris J.D. Hardy, Sebastian J. Crutch, Charles R. Marshall, Catherine J. Mummery, Jonathan D. Rohrer, et al. 2021. "Suspecting Dementia: Canaries, Chameleons and Zebras." *Practical Neurology* 21 (4): 300–312. <https://doi.org/10.1136/practneurol-2021-003019>.
- Johnson, Keith A., Nick C. Fox, Reisa A. Sperling, and William E. Klunk. 2012. "Brain Imaging in Alzheimer

- Disease." *Cold Spring Harbor Perspectives in Medicine* 2 (4): a006213. <https://doi.org/10.1101/cshperspect.a006213>.
- Jones, Richard H., and Francis Boadi-Boateng. 1991. "Unequally Spaced Longitudinal Data with AR(1) Serial Correlation." *Biometrics* 47 (1). <https://doi.org/10.2307/2532504>.
- Jöreskog, K G. 1969. "A General Approach to Confirmatory Maximum Likelihood Factor Analysis." *Psychometrika* 34 (2): 183–202. <https://doi.org/10.1007/BF02289343>.
- Jöreskog, K G, and D Sörbom. 1979. "Advances in Factor Analysis and Structural Equation Models, Abt." *Books, Cambridge, Massachusetts*.
- Jöreskog, K G, and Dag Sörbom. 2018. "LISREL 10 for Windows." Skokie (IL): Scientific Software International, Inc.
- Kapasi, Alifiya, Charles DeCarli, and Julie A. Schneider. 2017. "Impact of Multiple Pathologies on the Threshold for Clinically Overt Dementia." *Acta Neuropathologica* 134 (2): 171–86. <https://doi.org/10.1007/s00401-017-1717-7>.
- Karanth, Shama, Peter T. Nelson, Yuriko Katsumata, Richard J. Kryscio, Frederick A. Schmitt, David W. Fardo, Matthew D. Cykowski, Gregory A. Jicha, Linda J. Van Eldik, and Erin L. Abner. 2020. "Prevalence and Clinical Phenotype of Quadruple Misfolded Proteins in Older Adults." *JAMA Neurology* 77 (10): 1299–1307. <https://doi.org/10.1001/jamaneurol.2020.1741>.
- Kline, Rex B. 2018. "Response to Leslie Hayduk's Review of Principles and Practice of Structural Equation Modeling, 1 4th Edition." *Canadian Studies in Population* 45 (3–4): 188–95. <https://doi.org/10.25336/csp29418>.
- Kline, Rex B. 2012. "Assumptions in Structural Equation Modeling." In *Handbook of Structural Equation Modeling.*, edited by Rick H. Hoyle, 111–25. New York, NY, US: The Guilford Press.
- Kneip, Alois, and Theo Gasser. 1988. "Convergence and Consistency Results for Self-Modeling Nonlinear Regression." *The Annals of Statistics* 16 (1): 82–112. <https://doi.org/10.1214/aos/1176350692>.
- Koval, Igor, Alexandre Bône, Maxime Louis, Thomas Lartigue, Simona Bottani, Arnaud Marcoux, Jorge Samper-González, et al. 2021. "AD Course Map Charts Alzheimer's Disease Progression." *Scientific Reports* 11 (1): 1–16. <https://doi.org/10.1038/s41598-021-87434-1>.
- Kühnel, Line, Anna Karin Berger, Bo Markussen, and Lars L. Raket. 2021. "Simultaneous Modeling of Alzheimer's Disease Progression via Multiple Cognitive Scales." *Statistics in Medicine* 40 (14): 3251–66. <https://doi.org/10.1002/sim.8932>.
- Laird, N M, and J H Ware. 1982. "Random-Effects Models for Longitudinal Data." *Biometrics* 38 (4): 963–74. <https://doi.org/10.2307/2529876>.
- Landau, Susan M., Ming Lu, Abhinay D. Joshi, Michael Pontecorvo, Mark A. Mintun, John Q. Trojanowski, Leslie M. Shaw, and William J. Jagust. 2013. "Comparing Positron Emission

- Tomography Imaging and Cerebrospinal Fluid Measurements of β -Amyloid." *Annals of Neurology* 74 (6): 826–36. <https://doi.org/10.1002/ana.23908>.
- Launer, L J, K Masaki, H Petrovitch, D Foley, and R J Havlik. 1995. "The Association between Midlife Blood Pressure Levels and Late-Life Cognitive Function. The Honolulu-Asia Aging Study." *JAMA* 274 (23): 1846–51.
- Lee, Sven J. van der, Frank J. Wolters, M. Kamran Ikram, Albert Hofman, M. Arfan Ikram, Najaf Amin, and Cornelia M. van Duijn. 2018. "The Effect of APOE and Other Common Genetic Variants on the Onset of Alzheimer's Disease and Dementia: A Community-Based Cohort Study." *The Lancet Neurology* 17 (5): 434–44. [https://doi.org/10.1016/S1474-4422\(18\)30053-X](https://doi.org/10.1016/S1474-4422(18)30053-X).
- Lei, Pui-Wa, and Qiong Wu. 2012. "Estimation in SEM." In *Handbook of Structural Equation Modeling.*, edited by Rick H. Hoyle, 164–80. New York, NY, US: The Guilford Press.
- Li, Dan, Samuel Iddi, Wesley K. Thompson, and Michael C. Donohue. 2017. "Bayesian Latent Time Joint Mixed Effect Models for Multicohort Longitudinal Data." *Statistical Methods in Medical Research* 28 (3): 835–45. <https://doi.org/10.1177/0962280217737566>.
- Li, Dan, Samuel Iddi, Wesley K. Thompson, Michael S. Rafii, Paul S. Aisen, and Michael C. Donohue. 2018. "Bayesian Latent Time Joint Mixed-Effects Model of Progression in the Alzheimer's Disease Neuroimaging Initiative." *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring* 10: 657–68. <https://doi.org/10.1016/j.dadm.2018.07.008>.
- Lindstrom, Mary J, and Douglas M Bates. 1990. "Nonlinear Mixed Effects Models for Repeated Measures Data." *Biometrics* 46 (3): 673–87.
- Little, Roderick J A, and Donald B Rubin. 2019. *Statistical Analysis with Missing Data. 3rd Edition.* Vol. 793. John Wiley & Sons.
- Livingston, Gill, Andrew Sommerlad, Vasiliki Orgeta, Sergi G Costafreda, Jonathan Huntley, David Ames, Clive Ballard, et al. 2017. "Dementia Prevention, Intervention, and Care." *Lancet (London, England)* 390 (10113): 2673–2734. [https://doi.org/10.1016/S0140-6736\(17\)31363-6](https://doi.org/10.1016/S0140-6736(17)31363-6).
- Lorenzi, Marco, Maurizio Filippone, Giovanni B. Frisoni, Daniel C. Alexander, and Sebastien Ourselin. 2017. "Probabilistic Disease Progression Modeling to Characterize Diagnostic Uncertainty: Application to Staging and Prediction in Alzheimer's Disease." *NeuroImage* 190 (August 2017): 56–68. <https://doi.org/10.1016/j.neuroimage.2017.08.059>.
- Lumsden, Amanda L., Anwar Mulugeta, Ang Zhou, and Elina Hyppönen. 2020. "Apolipoprotein E (APOE) Genotype-Associated Disease Risks: A Phenome-Wide, Registry-Based, Case-Control Study Utilising the UK Biobank." *EBioMedicine* 59. <https://doi.org/10.1016/j.ebiom.2020.102954>.
- Luo, Jingqin, Folasade Agboola, Elizabeth Grant, John C Morris, Colin L Masters, Marilyn S Albert, Sterling C Johnson, et al. 2022. "Accelerated Longitudinal Changes and Ordering of Alzheimer Disease Biomarkers across the Adult Lifespan." *Brain* awac238.

<https://doi.org/10.1093/brain/awac238>.

- Mahmood, Syed S., Daniel Levy, Ramachandran S. Vasan, and Thomas J. Wang. 2014. "The Framingham Heart Study and the Epidemiology of Cardiovascular Disease: A Historical Perspective." *The Lancet* 383 (9921): 999–1008. [https://doi.org/10.1016/S0140-6736\(13\)61752-3](https://doi.org/10.1016/S0140-6736(13)61752-3).
- Marinescu, Răzvan Valentin, Arman Eshaghi, Marco Lorenzi, Alexandra L. Young, Neil P. Oxtoby, Sara Garbarino, Timothy J. Shakespeare, Sebastian J. Crutch, and Daniel C. Alexander. 2017. "A Vertex Clustering Model for Disease Progression: Application to Cortical Thickness Images." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10265 LNCS: 134–45. https://doi.org/10.1007/978-3-319-59050-9_11.
- Mattsson, Niklas, Michael Schöll, Olof Strandberg, Ruben Smith, Sebastian Palmqvist, Philip S Insel, Douglas Hägerström, et al. 2017. "18 F-AV-1451 and CSF T-tau and P-tau as Biomarkers in Alzheimer's Disease." *EMBO Molecular Medicine* 9 (9): 1212–23. <https://doi.org/10.15252/emmm.201707809>.
- Mcardle, J Jack. 1980. "Causal Modeling Applied to Psychonomic Systems Simulation." *Behavior Research Methods & Instrumentation*. Vol. 12.
- McKhann, Guy M., David Drachman, Marshall Folstein, Robert Katzman, Donald Price, and Emanuel M Stadlan. 1984. "Clinical Diagnosis of Alzheimer's Disease." *Neurology* 34 (7): 939 LP – 939. <https://doi.org/10.1212/WNL.34.7.939>.
- McKhann, Guy M., David S. Knopman, Howard Chertkow, Bradley T. Hyman, Clifford R. Jack, Claudia H. Kawas, William E. Klunk, et al. 2011. "The Diagnosis of Dementia Due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease." *Alzheimer's and Dementia* 7 (3): 263–69. <https://doi.org/10.1016/j.jalz.2011.03.005>.
- Merkle, Edgar C, and Yves Rosseel. 2018. "Blavaan: Bayesian Structural Equation Models via Parameter Expansion." *Journal of Statistical Software* 85 (4): 1–30. <https://doi.org/10.18637/jss.v085.i04>.
- Metropolis, Nicholas, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *The Journal of Chemical Physics* 21 (6): 1087–92. <https://doi.org/10.1063/1.1699114>.
- Meyers, John, and Kelly Meyers. 1995. "Rey Complex Figure Test under Four Different Administration Procedures." *The Clinical Neuropsychologist* 9 (1): 63–67. <https://doi.org/10.1080/13854049508402059>.
- Mills, Katherine T, Joshua D Bundy, Tanika N Kelly, Jennifer E Reed, Patricia M Kearney, Kristi Reynolds, Jing Chen, and Jiang He. 2016. "Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries." *Circulation* 134 (6): 441–50. <https://doi.org/10.1161/CIRCULATIONAHA.115.018912>.

- Möller, H.-J., and M. B. Graeber. 1998. "The Case Described by Alois Alzheimer in 1911." *European Archives of Psychiatry and Clinical Neuroscience* 248 (3): 111–22. <https://doi.org/10.1007/s004060050027>.
- Morris, John C. 1993. "The Clinical Dementia Rating (Cdr): Current Version and Scoring Rules." *Neurology* 43 (11): 2412–14. <https://doi.org/10.1212/wnl.43.11.2412-a>.
- Muthén, Linda K, and Bengt Muthen. 2017. *Mplus User's Guide. Eighth Edition*. Los Angeles (CA): Muthén & Muthén.
- Nelson, Peter T., Elizabeth Head, Frederick A. Schmitt, Paulina R. Davis, Janna H. Neltner, Gregory A. Jicha, Erin L. Abner, et al. 2011. "Alzheimer's Disease Is Not 'Brain Aging': Neuropathological, Genetic, and Epidemiological Human Studies." *Acta Neuropathologica*. <https://doi.org/10.1007/s00401-011-0826-y>.
- Ngandu, Tiia, Jenni Lehtisalo, Alina Solomon, Esko Levälähti, Satu Ahtiluoto, Riitta Antikainen, Lars Bäckman, et al. 2015. "A 2 Year Multidomain Intervention of Diet, Exercise, Cognitive Training, and Vascular Risk Monitoring versus Control to Prevent Cognitive Decline in at-Risk Elderly People (FINGER): A Randomised Controlled Trial." *The Lancet* 385 (9984): 2255–63. [https://doi.org/10.1016/S0140-6736\(15\)60461-5](https://doi.org/10.1016/S0140-6736(15)60461-5).
- Nichols, Emma, Jaimie D. Steinmetz, Stein Emil Vollset, Kai Fukutaki, Julian Chalek, Foad Abd-Allah, Amir Abdoli, et al. 2022. "Estimation of the Global Prevalence of Dementia in 2019 and Forecasted Prevalence in 2050: An Analysis for the Global Burden of Disease Study 2019." *The Lancet Public Health* 7 (2): e105–25. [https://doi.org/10.1016/S2468-2667\(21\)00249-8](https://doi.org/10.1016/S2468-2667(21)00249-8).
- O'Bryant, Sid E, Stephen C Waring, C Munro Cullum, James Hall, Laura Lacritz, Paul J Massman, Philip J Lupo, Joan S Reisch, Rachelle Doody, and Texas Alzheimer's Research Consortium. 2008. "Staging Dementia Using Clinical Dementia Rating Scale Sum of Boxes Scores: A Texas Alzheimer's Research Consortium Study." *Archives of Neurology* 65 (8): 1091–95. <https://doi.org/10.1001/archneur.65.8.1091>.
- Olsson, Bob, Ronald Lautner, Ulf Andreasson, Annika Öhrfelt, Erik Portelius, Maria Bjerke, Mikko Hölttä, et al. 2016. "CSF and Blood Biomarkers for the Diagnosis of Alzheimer's Disease: A Systematic Review and Meta-Analysis." *The Lancet Neurology* 15 (7): 673–84. [https://doi.org/10.1016/S1474-4422\(16\)00070-3](https://doi.org/10.1016/S1474-4422(16)00070-3).
- Palmqvist, Sebastian, Henrik Zetterberg, Kaj Blennow, Susanna Vestberg, Ulf Andreasson, David J. Brooks, Rikard Owenius, et al. 2014. "Accuracy of Brain Amyloid Detection in Clinical Practice Using Cerebrospinal Fluid β -Amyloid 42: A Cross-Validation Study against Amyloid Positron Emission Tomography." *JAMA Neurology* 71 (10): 1282–89. <https://doi.org/10.1001/jamaneurol.2014.1358>.
- Pantoni, Leonardo. 2010. "Cerebral Small Vessel Disease: From Pathogenesis and Clinical

- Characteristics to Therapeutic Challenges.” *The Lancet Neurology* 9 (7): 689–701. [https://doi.org/10.1016/S1474-4422\(10\)70104-6](https://doi.org/10.1016/S1474-4422(10)70104-6).
- Pantoni, Leonardo, Julio H. Garcia, and Gregory G. Brown. 1996. “Vascular Pathology in Three Cases of Progressive Cognitive Deterioration.” *Journal of the Neurological Sciences* 135 (2): 131–39. [https://doi.org/10.1016/0022-510X\(95\)00273-5](https://doi.org/10.1016/0022-510X(95)00273-5).
- Perrine, Anne-Laure, Lecoffre Camille, Blacher Jacques, and Valérie Olié. 2018. “L’hypertension Artérielle En France : Prévalence, Traitement et Contrôle En 2015 et Évolutions Depuis 2006.” *Bull Epidémiol Hebd* 10: 170–79.
- Petersen, R C, P S Aisen, L A Beckett, M C Donohue, A C Gamst, D J Harvey, C R Jack, et al. 2010. “Alzheimer’s Disease Neuroimaging Initiative (ADNI) Clinical Characterization.” *Neurology* 74 (3): 201–9. <https://doi.org/10.1212/WNL.0b013e3181cb3e25>.
- Prince, Martin, Cleusa Ferri P, Renata Sousa, Jim Jackson, Emiliano Albanese, Wagner S Ribeiro, and Mina Honyashiki. 2009. “World Alzheimer Report 2009: The Global Prevalence of Dementia.” *Alzheimer’s Disease International*, 1–96.
- Proust-Lima, Cécile, Viviane Philipps, and Benoit Liqueur. 2015. “Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package Lcmm.” <https://doi.org/10.18637/jss.v078.i02>.
- Raket, Lars Lau. 2020. “Statistical Disease Progression Modeling in Alzheimer Disease.” *Frontiers in Big Data* 3 (August): 1–18. <https://doi.org/10.3389/fdata.2020.00024>.
- Rinne, Juha O., Dean F. Wong, David A. Wolk, Ville Leinonen, Steven E. Arnold, Chris Buckley, Adrian Smith, et al. 2012. “Flutemetamol PET Imaging and Cortical Biopsy Histopathology for Fibrillar Amyloid β Detection in Living Subjects with Normal Pressure Hydrocephalus: Pooled Analysis of Four Studies.” *Acta Neuropathologica* 124 (6): 833–45. <https://doi.org/10.1007/s00401-012-1051-z>.
- Rizvi, Batool, Atul Narkhede, Briana S. Last, Mariana Budge, Giuseppe Tosto, Jennifer J. Manly, Nicole Schupf, Richard Mayeux, and Adam M. Brickman. 2018. “The Effect of White Matter Hyperintensities on Cognition Is Mediated by Cortical Atrophy.” *Neurobiology of Aging* 64: 25–32. <https://doi.org/10.1016/j.neurobiolaging.2017.12.006>.
- Roberts, Rosebud O., Yonas E. Geda, David S. Knopman, Ruth H. Cha, V. Shane Pankratz, Bradley F. Boeve, Robert J. Ivnik, Eric G. Tangalos, Ronald C. Petersen, and Walter A. Rocca. 2008. “The Mayo Clinic Study of Aging: Design and Sampling, Participation, Baseline Measures and Sample Characteristics.” *Neuroepidemiology* 30 (1): 58–69. <https://doi.org/10.1159/000115751>.
- Robinson, John L., Edward B. Lee, Sharon X. Xie, Lior Rennert, Eunran Suh, Colin Bredenberg, Carrie Caswell, et al. 2018. “Neurodegenerative Disease Concomitant Proteinopathies Are Prevalent, Age-Related and APOE4-Associated.” *Brain* 141 (7): 2181–93.

<https://doi.org/10.1093/brain/awy146>.

- Rosseel, Yves. 2012. "Lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48 (2 SE-Articles): 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Rouanet, Anaïs, Justina Avila-Rieger, Aline Dugravot, Jeremie Lespinasse, Rachel Stuckwisch, Richard Merrick, Emma Anderson, et al. 2022. "How Selection Over Time Contributes to the Inconsistency of the Association Between Sex/Gender and Cognitive Decline Across Cognitive Aging Cohorts." *American Journal of Epidemiology* 191 (3): 441–52. <https://doi.org/10.1093/aje/kwab227>.
- Sabuncu, Mert R, Rahul S Desikan, Jorge Sepulcre, Boon Thye T Yeo, Hesheng Liu, Nicholas J Schmansky, Martin Reuter, et al. 2011. "The Dynamics of Cortical and Hippocampal Atrophy in Alzheimer Disease." *Archives of Neurology* 68 (8): 1040–48. <https://doi.org/10.1001/archneurol.2011.167>.
- Samaille, T, L Fillon, R Cuingnet, E Jouvent, H Chabriat, D Dormont, O Colliot, and M Chupin. 2012. "Contrast-Based Fully Automatic Segmentation of White Matter Hyperintensities: Method and Validation." *PLoS One* 7 (11): e48953. <https://doi.org/10.1371/journal.pone.0048953>.
- Satizabal, Claudia L., Alexa S. Beiser, Vincent Chouraki, Geneviève Chêne, Carole Dufouil, and Sudha Seshadri. 2016. "Incidence of Dementia over Three Decades in the Framingham Heart Study." *New England Journal of Medicine* 374 (6): 523–32. <https://doi.org/10.1056/nejmoa1504327>.
- Satorra, Albert, and Pete M Bentler. 1994. "Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis." In *Latent Variables Analysis: Applications for Developmental Research*, edited by Alexander von Eye and Clifford C Clogg, 399–419. Sage Publications, Inc.
- Schiratti, Jean Baptiste, Stéphanie Allasonnière, Olivier Colliot, and Stanley Durrleman. 2017. "A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations." *Journal of Machine Learning Research* 18: 1–33.
- Schuff, Norbert, Duygu Tosun, Philip S Insel, Gloria C Chiang, Diana Truran, Paul S Aisen, Clifford R Jack, and Michael W Weiner. 2012. "Nonlinear Time Course of Brain Volume Loss in Cognitively Normal and Impaired Elders." *Neurobiology of Aging* 33 (5): 845–55. <https://doi.org/10.1016/j.neurobiolaging.2010.07.012>.
- Scull, Andrew. 2010. "Psychiatry and the Social Sciences, 1940–2009." *History of Political Economy* 42 (December): 25–52. <https://doi.org/10.1215/00182702-2009-071>.
- Selkoe, Dennis J, and John Hardy. 2016. "The Amyloid Hypothesis of Alzheimer's Disease at 25 Years." *EMBO Molecular Medicine* 8 (6): 595–608. <https://doi.org/10.15252/emmm.201606210>.
- Silverberg, Nina, Cerise Elliott, Laurie Ryan, Eliezer Masliah, and Richard Hodes. 2018. "NIA Commentary on the NIA-AA Research Framework: Towards a Biological Definition of Alzheimer's Disease." *Alzheimer's & Dementia* 14 (4): 576–78. <https://doi.org/10.1016/j.jalz.2018.03.004>.
- Sluijmer, Jasper D, Femke H Bouwman, Hugo Vrenken, Marinus A Blankenstein, Frederik Barkhof, Wiesje M van der Flier, and Philip Scheltens. 2010. "Whole-Brain Atrophy Rate and CSF Biomarker

- Levels in MCI and AD: A Longitudinal Study.” *Neurobiology of Aging* 31 (5): 758–64. <https://doi.org/10.1016/j.neurobiolaging.2008.06.016>.
- Smailagic, Nadja, Louise Lafortune, Sarah Kelly, Chris Hyde, and Carol Brayne. 2018. “18F-FDG PET for Prediction of Conversion to Alzheimer’s Disease Dementia in People with Mild Cognitive Impairment: An Updated Systematic Review of Test Accuracy.” *Journal of Alzheimer’s Disease* 64 (4): 1175–94. <https://doi.org/10.3233/JAD-171125>.
- Smith, Eric E., Svetlana Egorova, Deborah Blacker, Ronald J. Killiany, Alona Muzikansky, Bradford C. Dickerson, Rudolph E. Tanzi, Marilyn S. Albert, Steven M. Greenberg, and Charles R.G. Guttmann. 2008. “Magnetic Resonance Imaging White Matter Hyperintensities and Brain Volume in the Prediction of Mild Cognitive Impairment and Dementia.” *Archives of Neurology* 65 (1): 94–100. <https://doi.org/10.1001/archneurol.2007.23>.
- Snitz, Beth E, Tianxiu Wang, Yona Keich Cloonan, Erin Jacobsen, Chung-Chou H Chang, Tiffany F Hughes, M Ilyas Kamboh, and Mary Ganguli. 2018. “Risk of Progression from Subjective Cognitive Decline to Mild Cognitive Impairment: The Role of Study Setting.” *Alzheimer’s & Dementia* 14 (6): 734–42. <https://doi.org/10.1016/j.jalz.2017.12.003>.
- Spirtes, Peter, Thomas Richardson, Christopher Meek, Richard Scheines, and Clark Glymour. 1998. “Using Path Diagrams as a Structural Equation Modeling Tool.” *Sociological Methods & Research* 27 (2): 182–225. <https://doi.org/10.1177/0049124198027002003>.
- Stan Development Team. 2022a. “Parallelization.” In *Stan User’s Guide*.
- . 2022b. “Posterior Analysis.” In *Stan Reference Manual*.
- . 2022c. “Stan Modeling Language Users Guide and Reference Manual.” <https://mc-stan.org>.
- Steiger, James H., and J. C. Lind. 1980. “Statistically Based Tests for the Number of Common Factors.” In *The Annual Meeting of the Psychometric Society*.
- Stephen, Ruth, Yawu Liu, Tiia Ngandu, Riitta Antikainen, Juha Hulkkonen, Juha Koikkalainen, Nina Kemppainen, et al. 2019. “Brain Volumes and Cortical Thickness on MRI in the Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER).” *Alzheimer’s Research & Therapy* 11 (1): 53. <https://doi.org/10.1186/s13195-019-0506-z>.
- Stern, Yaakov. 2012. “Cognitive Reserve in Ageing and Alzheimer’s Disease.” *The Lancet Neurology*. [https://doi.org/10.1016/S1474-4422\(12\)70191-6](https://doi.org/10.1016/S1474-4422(12)70191-6).
- Sweeney, Melanie D., Axel Montagne, Abhay P. Sagare, Daniel A. Nation, Lon S. Schneider, Helena C. Chui, Michael G. Harrington, et al. 2019. “Vascular Dysfunction—The Disregarded Partner of Alzheimer’s Disease.” *Alzheimer’s and Dementia* 15 (1): 158–67. <https://doi.org/10.1016/j.jalz.2018.07.222>.
- Taddé, Bachirou O, H el ene Jacqmin-Gadda, Jean-Fran ois Dartigues, Daniel Commenges, and C ecile Proust-Lima. 2020. “Dynamic Modeling of Multivariate Dimensions and Their Temporal

- Relationships Using Latent Processes: Application to Alzheimer's Disease." *Biometrics* 76 (3): 886–99. <https://doi.org/https://doi.org/10.1111/biom.13168>.
- Telgte, Annemieke Ter, Esther M.C. Van Leijsen, Kim Wiegertjes, Catharina J.M. Klijn, Anil M. Tuladhar, and Frank Erik De Leeuw. 2018. "Cerebral Small Vessel Disease: From a Focal to a Global Perspective." *Nature Reviews Neurology* 14 (7): 387–98. <https://doi.org/10.1038/s41582-018-0014-y>.
- Teunissen, Charlotte E., Inge M.W. Verberk, Elisabeth H. Thijssen, Lisa Vermunt, Oskar Hansson, Henrik Zetterberg, Wiesje M. van der Flier, Michelle M. Mielke, and Marta del Campo. 2022. "Blood-Based Biomarkers for Alzheimer's Disease: Towards Clinical Implementation." *The Lancet Neurology* 21 (1): 66–77. [https://doi.org/10.1016/S1474-4422\(21\)00361-6](https://doi.org/10.1016/S1474-4422(21)00361-6).
- Thurstone, L L. 1987. "Psychophysical Analysis." *American Journal of Psychology* 100 (3–4): 587.
- Tombaugh, Tom N. 2004. "Trail Making Test A and B: Normative Data Stratified by Age and Education." *Archives of Clinical Neuropsychology* 19 (2): 203–14. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8).
- Tomlinson, B E, G Blessed, and M Roth. 1968. "Observations on the Brains of Non-Demented Old People." *Journal of the Neurological Sciences* 7 (2): 331–56. [https://doi.org/https://doi.org/10.1016/0022-510X\(68\)90154-8](https://doi.org/https://doi.org/10.1016/0022-510X(68)90154-8).
- . 1970. "Observations on the Brains of Demented Old People." *Journal of the Neurological Sciences* 11 (3): 205–42. [https://doi.org/https://doi.org/10.1016/0022-510X\(70\)90063-8](https://doi.org/https://doi.org/10.1016/0022-510X(70)90063-8).
- Tucker, Ledyard R., and Charles Lewis. 1973. "A Reliability Coefficient for Maximum Likelihood Factor Analysis." *Psychometrika* 38 (1). <https://doi.org/10.1007/BF02291170>.
- United Nations Department of Economic and Social Affairs, Population Division. 2022. "World Population Prospects 2022: Summary of Results."
- Vermunt, Lisa, Sietske A.M. Sikkes, Ardo van den Hout, Ron Handels, Isabelle Bos, Wiesje M. van der Flier, Silke Kern, et al. 2019. "Duration of Preclinical, Prodromal, and Dementia Stages of Alzheimer's Disease in Relation to Age, Sex, and APOE Genotype." *Alzheimer's and Dementia* 15 (7): 888–98. <https://doi.org/10.1016/j.jalz.2019.04.001>.
- Voelkle, Manuel C, Johan H L Oud, Eldad Davidov, and Peter Schmidt. 2012. "An SEM Approach to Continuous Time Modeling of Panel Data: Relating Authoritarianism and Anomia." *Psychological Methods* 17 (2): 176–92. <https://doi.org/10.1037/a0027543>.
- Wardlaw, Joanna M., Colin Smith, and Martin Dichgans. 2019. "Small Vessel Disease: Mechanisms and Clinical Implications." *The Lancet Neurology* 18 (7): 684–96. [https://doi.org/10.1016/S1474-4422\(19\)30079-1](https://doi.org/10.1016/S1474-4422(19)30079-1).
- Weintraub, Sandra, Alissa H. Wicklund, and David P. Salmon. 2012. "The Neuropsychological Profile of Alzheimer Disease." *Cold Spring Harbor Perspectives in Medicine* 2 (4).

<https://doi.org/10.1101/cshperspect.a006171>.

- Williams, Bryan, Giuseppe Mancina, Wilko Spiering, Enrico Agabiti Rosei, Michel Azizi, Michel Burnier, Denis L. Clement, et al. 2018. "2018 ESC/ESH Guidelines for The management of Arterial Hypertension." *European Heart Journal*. Oxford University Press. <https://doi.org/10.1093/eurheartj/ehy339>.
- Williamson, Jeff D., Nicholas M. Pajewski, Alexander P. Auchus, R. Nick Bryan, Gordon Chelune, Alfred K. Cheung, Maryjo L. Cleveland, et al. 2019. "Effect of Intensive vs Standard Blood Pressure Control on Probable Dementia: A Randomized Clinical Trial." *JAMA - Journal of the American Medical Association* 321 (6): 553–61. <https://doi.org/10.1001/jama.2018.21442>.
- Wolz, Robin, Valterri Julkunen, Juha Koikkalainen, Eini Niskanen, Dong Ping Zhang, Daniel Rueckert, Hilka Soininen, and Jyrki Lötjönen. 2011. "Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease." *PLoS ONE* 6 (10): 1–9. <https://doi.org/10.1371/journal.pone.0025446>.
- World Health Organization. 2001. "Biomarkers In Risk Assessment: Validity And Validation." *Environmental Health Criteria*. Geneva.
- Wu, Yu Tzu, Alexa S. Beiser, Monique M.B. Breteler, Laura Fratiglioni, Catherine Helmer, Hugh C. Hendrie, Hiroyuki Honda, et al. 2017. "The Changing Prevalence and Incidence of Dementia over Time-Current Evidence." *Nature Reviews Neurology* 13 (6): 327–39. <https://doi.org/10.1038/nrneurol.2017.63>.
- Young, Alexandra L., Razvan V. Marinescu, Neil P. Oxtoby, Martina Bocchetta, Keir Yong, Nicholas C. Firth, David M. Cash, et al. 2018. "Uncovering the Heterogeneity and Temporal Complexity of Neurodegenerative Diseases with Subtype and Stage Inference." *Nature Communications* 9 (1): 1–16. <https://doi.org/10.1038/s41467-018-05892-0>.
- Young, Alexandra L., Neil P. Oxtoby, Pankaj Daga, David M. Cash, Nick C. Fox, Sebastien Ourselin, Jonathan M. Schott, and Daniel C. Alexander. 2014. "A Data-Driven Model of Biomarker Changes in Sporadic Alzheimer's Disease." *Brain* 137 (9): 2564–77. <https://doi.org/10.1093/brain/awu176>.
- Yuan, Ke-Hai, and Peter M Bentler. 1998. "Normal Theory Based Test Statistics in Structural Equation Modelling." *British Journal of Mathematical and Statistical Psychology* 51 (2): 289–309. <https://doi.org/10.1111/j.2044-8317.1998.tb00682.x>.
- Zlokovic, Berislav V., Rebecca F. Gottesman, Kenneth E. Bernstein, Sudha Seshadri, Ann McKee, Heather Snyder, Steven M. Greenberg, et al. 2020. "Vascular Contributions to Cognitive Impairment and Dementia (VCID): A Report from the 2018 National Heart, Lung, and Blood Institute and National Institute of Neurological Disorders and Stroke Workshop." *Alzheimer's and Dementia* 16 (12): 1714–33. <https://doi.org/10.1002/alz.12157>.

