HAL
open science

Stratégies de ré-analyse d'un essai de prévention du déclin des fonctions cognitives non concluant

Céline Bougel

▶ **To cite this version:**

Céline Bougel. Stratégies de ré-analyse d'un essai de prévention du déclin des fonctions cognitives non concluant. Médecine humaine et pathologie. Université Paul Sabatier - Toulouse III, 2020. Français. NNT : 2020TOU30192 . tel-03184809

HAL Id: tel-03184809

https://theses.hal.science/tel-03184809v1

Submitted on 29 Mar 2021

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

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Céline BOUGEL

Le 26 novembre 2020

Stratégies de ré-analyse d'un essai de prévention du déclin des fonctions cognitives non concluant.

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Epidémiologie**

Unité de recherche :
LEASP - Laboratoire d'Epidémiologie et Analyses en santé publique

Thèse dirigée par
Sandrine ANDRIEU et Nicolas SAVY

Jury
M. Jean-François DUPUY, Rapporteur
M. Pierre-Marie PREUX, Rapporteur
Mme Vanina BONGARD, Examinatrice
Mme Hélène JACQMIN-GADDA, Examinatrice
Mme Sandrine ANDRIEU, Directrice de thèse
M. Nicolas SAVY, Co-directeur de thèse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *26/11/2020* **par :**

Céline BOUGEL

Stratégies de ré-analyse d'un essai de prévention du déclin des fonctions cognitives non concluant

---

JURY

Sandrine Andrieu	Professeur d'Université	Directrice de thèse
Vanina Bongard	Fonction	Président du Jury
Hélène Jacqmin-Gadda	Fonction	Examinateur
Nicolas Savy	Maître de conférences	Co-directeur de thèse
Philippe Saint-Pierre	Maître de conférences	Co-directeur de thèse
Sébastien Déjean	Ingénieur de recherche	Co-directeur de thèse

---

École doctorale et spécialité :

MITT : Domaine Mathématiques : Épidémiologie

Unité de Recherche :

UMR 1027 INSERM - Université Toulouse III - Équipe Vieillissement - Axe MAINTAIN (MAintain Functions and INTrinsec capacities with Aging : Preventive and Personalized INterventional Research)

Directeur(s) de Thèse :

Sandrine Andrieu, *Nicolas* Savy et *Philippe* Saint Pierre

Rapporteurs :

Jean-François Dupuy et *Pierre-Marie* Preux

Remerciements

J'aimerais tout d'abord adresser mes remerciements aux encadrants de ce mémoire, Sandrine ANDRIEU, *ma directrice de thèse, Nicolas* SAVY *et Philippe* SAINT-PIERRE *mes co-directeurs. Je les remercie tous trois pour leur patience, leur disponibilité et surtout pour leurs judicieux conseils, qui ont contribué à alimenter ma réflexion. Un grand merci à Sandrine pour m'avoir fait confiance pour ce travail, pour avoir obtenu les financements nécessaires et pour m'avoir appris à communiquer scientifiquement. Vous savez combien vous m'avez apporté tout au long de ce travail, aussi bien sur le plan clinique que personnel. Nicolas et Philippe, vous étiez là pour me guider sur le plan statistique, pour écouter mes conclusions et me confirmer les prochaines analyses à mener. Je n'aurais jamais pu accomplir ce travail de recherche sans votre concours. Pour tout cela et bien plus encore, je vous remercie tous trois pour l'expérience enrichissante et pleine d'intérêt que vous m'avez fait vivre durant ces trente-six mois au sein de l'UMR1027 de l'INSERM et à l'IMT.*

Je tiens à remercier très chaleureusement l'Association Monégasque pour la recherche sur la maladie d'Alzheimer (AMPA), Harmonie Mutuelle et la Fondation de l'Avenir, la fondation Alzheimer Prevention in Occitania and Catalonia (APOC) ainsi que le projet INSPIRE (Fonds Européens de Développement Régional (FEDER)) pour avoir financé mes travaux. En dernier lieu, le groupe MAPT/DSA est remercié dans son intégralité.

Je désire également remercier toute les équipes de l'UMR1027 de Toulouse, en particulier l'équipe Vieillissement, pour m'avoir accueillie et fourni les outils nécessaires à ma recherche, des contacts professionnels ainsi qu'une aide sans condition. Je ne peux malheureusement pas tous vous citer, mais le cœur y est. Ceux avec qui j'ai été en contact se reconnaitront assurément.

Mes remerciements vont aussi à l'endroit des membres du jury pour avoir accepté juger ce travail de thèse. Vanina BONGARD, *je vous remercie d'avoir accepté de présider cette soutenance de thèse. Jean-François* DUPUY *et Pierre-Marie* PREUX, *je vous remercie d'avoir accepté porter un regard critique sur ce travail en tant que rapporteurs. Vos remarques m'ont permis de me rendre compte de la valeur de mon travail. Enfin, Hélène* JACQMIN-GADDA, *je vous remercie d'avoir accepté d'examiner ce travail de thèse lors de la soutenance.*

Je souhaite aussi remercier Sébastien DÉJEAN *d'avoir rejoins l'aventure en participant à nos réunions de travail sur son temps libre. Tu as parfaitement exercé ton rôle de canard en plastique et je ne t'en remercierai jamais assez. J'ai pu te parler très franchement durant tout ce temps, te faire part de toutes mes impressions sans cadre officiel. Tu m'as apporté un regard neuf et pragmatique extrêmement précieux. Tu m'as également appris à être plus factuelle, plus précise dans les termes que j'employais. Au final, tu étais comme un autre co-directeur de thèse.. Au même titre que Sébastien, je remercie Caroline* GIULIOLI *pour m'avoir tant aidé sur le plan clinique durant les deux premières années de ma thèse. Ton aide était primordiale pour moi et ton regard critique m'a aidé à progresser.*

J'ai trouvé une aide précieuse auprès de Julie AKKOYUN-FARINEZ. *Tes conseils réguliers et corrections concernant mon mémoire ont grandement facilité mon travail et m'ont aidé à prendre du recul. Tu as également toujours eu confiance en mes capacités et m'as aidé à conserver ma motivation dans les moments difficiles. Merci pour ton soutien quotidien et pour ton écoute attentive. J'adresse également mes sincères remerciements à mes parents pour avoir accepté de relire et critiquer mon mémoire avec un regard complètement externe à la recherche. Je remercie aussi Rémy* MAILLES, *mon conjoint, pour son soutien inconditionnel depuis plus de dix années déjà. Il a eu le courage d'endurer, en plus de notre quotidien, mes hauts et mes bas, mes excès de joie et mes moments de doute. Enfin, mes remerciements tout particuliers reviennent à Pascale* POMBOURCQ *qui, en plus de m'avoir apporté un soutien sans faille, m'a fait découvrir les biostatistiques et m'épaule depuis mon année de terminale.*

À tous ces intervenants qui, de près ou de loin, ont contribué à ces trois années de recherche, je présente, en ces quelques lignes, mes remerciements, mon respect, ma gratitude et ma reconnaissance.

Liste des abréviations

La liste ci-dessous développe les abréviations utilisées dans le mémoire, reprenant principalement des termes anglais.

ACTIVE	*Advanced Cognitive Training for Independant and Vital Elderly*
ADCS	*Alzheimer Disease Cooperative Study*
ADL	*Activities of Daily Living*
AFD	Analyse Factorielle Discriminante
AIC	*Akaïke Information Criterion*
ANOVA	Analyse de variance
ANSM	Agence Nationale de Sécurité du Médicament et des produits de santé
APICC	*Alzheimer's Prevention Initiative Cognitive Composite*
APOE	Apolipoprotéine E
BEP	Brevet d'études professionnel
BIC	*Bayesian Information Criterion*
BRSD	*Behavior Rating Scale for Dementia*
CAH	Classification Ascendante Hiérarchique
CAIDE	*Cardiovascular Risk Factors, Aging and Dementia*
CAP	Certificat d'aptitude professionnel
CART	*Classification And Regression Trees*
CDR	*Clinical Dementia Rating score*
CEP	Certificat d'études primaires
CNT	*Category Naming Test*
COWAT	*Controlled Word Association Test*
DHA	Acide docosahexaenoïque
DSA	*Data Sharing Alzheimer Group*
EM	*Expectation-Maximisation*
EMA	Agence Européenne des Médicaments
EPA	Acide eicosaentaenoïque
EVA	Échelles Analogiques Visuelles
FDA	*Food and Drug Administration*
FINGER	*Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability*
GDS	*Geriatric Depression Scale*
GEM	*The Ginko Evaluation of Memory*
HAMD	*Hamilton clinical rating scale for depression*
HATICE	*Healthy Ageing Through Internet Counselling in the Elderly*
HMM	*Hidden Markov Models*
IADL	*Instrumental Activities of Daily Living*
IC95%	Intervalle de Confiance à 95 %
IMC	Indice de Masse Corporelle
IM	Intervention Multidomaine
KmL	*K-means for longitudinal data*
kmlShape	*K-means for longitudinal data with shape respecting trajectories*
LIFE	*Lifestyle Interventions and Independance for Elders*
LOCF	*Last Observation Carried Forward*

MAPT	*Multidomain Alzheimer Preventive Trial*
MAR	*Missing At Random*
MCID	*Minimal clinically importance différence*
MMSE	*Mini-Mental State Examination*
OMS	*Organisation Mondiale de la Santé*
OR	*Odds Ratio*
PACC	*Preclinical Alzheimer Cognitive Composite*
PAQUID	Personnes Agées QUID
PAS	Pression artérielle systolique
PREADVISE	*Prevention of Alzheimer's Disease with Vitamine E and Selenium*
preDIVA	*Prevention of Dementia by Intensive Vascular care*
RLRI16	Rappel libre/ rappel indicé à 16 items
RF	*Random Forest*
RR	Risque relatif
SPINT MIND	*Systolic Blood Pressure Intervention Trial Memory and Cognition IN Decreased hypertension*
SPPB	*Short Physical Performance Battery*
TAU	*Tubulin associated unit*
TMT-A	Trail Making Test version A
TMT-B	Trail Making Test version B
WAIS-R	*Digit Symbol Substitution Test front the Revisited Wechsler code*

Table des matières

1 Introduction **1**

2 Essais cliniques et maladie d'Alzheimer **5**

1 Schéma général d'un essai clinique 6

1.1 Le schéma d'étude 6

1.2 Les critères d'éligibilité 7

1.3 Le(s) critère(s) de jugement 7

1.3.1 Critères qualitatifs 7

1.3.2 Critères quantitatifs 7

1.3.3 Critères composites 8

1.3.4 Critères longitudinaux 9

1.4 La comparabilité des groupes 9

1.5 L'analyse statistique 10

1.6 L'effet attendu et nombre de sujets nécessaire 11

2 Les essais de thérapie modificatrice et de prévention 11

2.1 Les essais de thérapie modificatrice 11

2.2 Les essais de prévention 12

3 Pathologie à l'étude : la maladie d'Alzheimer 14

3.1 Présentation de la maladie d'Alzheimer 14

3.2 Les principaux symptômes de la maladie 15

3.2.1 La mémoire 15

3.2.2 Le langage 16

3.2.3 L'attention 16

3.2.4 Les fonctions exécutives 16

3.2.5 Les fonctions visuo-spaciales 16

4 Les essais de prévention de la maladie d'Alzheimer 16

4.1 Généralités 17

4.2 Approches comparatives de grands essais de prévention 17

3 Le cas d'étude : l'essai Multidomain Alzheimer Prevention Trial 21
1 Présentation du protocole . . . 22
1.1 Schéma d'étude et critères d'éligibilité . . . 22
1.2 Objectif de l'essai . . . 23
1.3 Critères de jugement . . . 23
1.3.1 Critère de jugement principal . . . 24
1.3.2 Critères de jugement secondaires . . . 29
1.3.3 Critères de l'analyse post-hoc . . . 33
1.4 Plan d'analyses statistiques . . . 37
2 Principaux résultats de l'essai MAPT . . . 38
2.1 Population analysée dans MAPT . . . 38
2.2 Analyses statistiques des données de MAPT . . . 38

4 Le pré-traitement des données longitudinales 43
1 Généralités . . . 43
2 Populations sélectionnées pour les analyses post-hoc . . . 46
2.1 Population $\mathcal{P}_{tot}$ . . . 47
2.2 Population $\mathcal{P}_{comp}$ . . . 47
2.3 Population $\mathcal{P}_{mapt}$ . . . 50
3 Analyses exploratoires préliminaires : Exemple des données de MAPT . . . 50
3.1 Identification et étude des profils extrêmes d'évolution des fonctions cognitives . . . 50
3.1.1 Principe . . . 50
3.1.2 Comportements identifiés . . . 52
3.2 Définition et étude de catégories de variations des fonctions cognitives 54
3.2.1 Principe . . . 54
3.2.2 Catégories de variation obtenues . . . 55
4 Conclusion . . . 56

5 Outils de ré-analyse d'un essai de prévention non concluant 59
1 Techniques d'exploration des trajectoires de fonctions cognitives . . . 61
1.1 Le *clustering* fonctionnel pour la recherche de groupes homogènes . . 62
1.1.1 L'algorithme des centres mobiles . . . 62
1.1.2 L'algorithme de Classification Ascendante Hiérarchique . . . 71
1.1.3 La méthode de sémiologie graphique . . . 74
1.2 La classification fonctionnelle pour la recherche de groupes répondeurs 76
2 Techniques de modélisation des trajectoires de fonctions cognitives . . . 77

2.1 Méthode d'apprentissage par renforcement 77
2.1.1 Les règles de décision 78
2.1.2 Cas particulier du *Q-learning* 79
2.2 Méthodes markoviennes 80
2.2.1 Les modèles de Markov 82
2.2.2 Les modèles de Markov cachés 85

6 Application aux données de l'essai MAPT et recommandations 91
1 Résultats de l'exploration des trajectoires cognitives 92
1.1 Résultats du *clustering* fonctionnel 92
1.1.1 L'algorithme des centres mobiles 92
1.1.2 L'algorithme de Classification Ascendante Hiérarchique . . . 96
1.1.3 La méthode de sémiologie graphique 100
1.2 Résultats de la recherche de groupes répondeurs 103
1.2.1 Revue de la littérature 103
1.2.2 Interprétation statistique 105
1.2.3 Interprétation clinique 107
2 Résultats de la modélisation des trajectoires cognitives 109
2.1 Résultats des méthodes d'apprentissage statistique 109
2.1.1 Méthodes d'apprentissage "classiques" 109
2.1.2 Méthode d'apprentissage par renforcement 113
2.2 Résultats des méthodes markoviennes 114
2.2.1 Les modèles de Markov homogènes 114
2.2.2 Les modèles de Markov cachées 126

7 Conclusion et perspectives 135
1 Rappel de la problématique 135
2 Spécificités de notre approche 136
2.1 Les apports des approches d'exploration 136
2.2 Les apports des approches de modélisation 140
3 Perspectives de recherche 143

Références 171

Annexes 173
A.1 Synthèse des variables exploitées dans l'analyse post-hoc 174
A.2 Algorithmes d'estimation des paramètres d'un HMM 175

A.2.1 Apprentissage : Algorithme *forward-backward* 176
A.2.2 Évaluation : Algorithme *forward* 177
A.2.3 Optimisation : Algorithme *backward* 177

Chapitre 1

Introduction

Depuis de nombreuses années, les essais cliniques demeurent le *gold-standard* pour évaluer l'efficacité d'un traitement, y compris dans le domaine de la prévention. Dans le cas d'un essai non concluant, l'absence d'effet du traitement à l'épreuve est souvent mise en avant. Une alternative pourrait être que le traitement à l'épreuve est efficace, mais que la méthodologie déployée n'a pas permis de détecter cet effet. Dans ce travail, nous nous concentrons sur les essais de prévention. Ces derniers visent à démontrer l'effet protecteur d'une intervention sur le déclin d'une fonction, ce qui s'objective soit par repousser l'apparition d'un évènement, soit par retarder le déclin d'une fonction. Dans l'étude des fonctions cognitives, l'essai de prévention a pour objectif de retarder le déclin de la fonction. Ces essais possèdent certaines spécificités rendant l'étude d'un effet traitement potentiel beaucoup plus complexe à étudier que dans un cadre classique. En effet, ils sont caractérisés par un recueil de données répétées au long cours avec un éventuel effet d'apprentissage, une longue phase de latence avant de pouvoir détecter un effet, des sorties d'étude importantes et une hétérogénéité de la population d'étude. Nous faisons l'hypothèse que les méthodologies déployées ne sont pas adaptées, ce qui pourrait, en partie, expliquer l'absence d'effet constaté dans la grande majorité des essais de prévention du déclin des fonctions cognitives de la littérature. Ce constat est effectué malgré la concordance de nombreux travaux épidémiologiques quant à l'effet protecteur de certaines interventions.

La première spécificité des essais de prévention est que le critère de jugement principal, sur lequel repose l'évaluation de l'effet du traitement, se recueille de façon longitudinale [1]. À chaque sujet est associé soit une durée avant la survenue d'un évènement, soit une suite de valeurs du critère relevées à différents moments du suivi, soit une donnée fonctionnelle. La durée avant la survenue d'un évènement n'est pas étudiée dans le présent travail, seules les données quantitatives le sont. Chaque type de donnée peut être analysé par un ensemble de méthodes d'analyses statistiques spécifiques, comme celles traitant la dépendance des données répétées ou fonctionnelles au cours du temps. Pour ces dernières, nécessairement partiellement observées en pratique, des méthodes d'analyses de données répétées peuvent être employées. Dans ce cas, un paramètre représentatif de l'évolution de la donnée fonctionnelle (paramètre de synthèse) doit être défini. Nous proposons, entre autres, le taux d'accroissement entre les visites du suivi ou la conversion de la donnée en un état représentant différentes variations. Les méthodes d'analyses statistiques proposées par la suite traitent soit directement la donnée fonctionnelle, soit un des deux paramètres de synthèse. Le caractère longitudinal du critère de jugement principal implique également que les essais de prévention sont sensibles à la durée de suivi, en particulier pour la prévention à long

terme. Un long suivi augmente le nombre de perdus de vue et la variabilité des mesures dans le temps. La population des essais de prévention du déclin des fonctions cognitives est particulièrement âgée, ce qui signifie que leur risque de mortalité ou de (co)morbidités est très élevé. Le taux de perdus de vue est d'autant plus important chez les sujets dont les fonctions cognitives sont altérées [2]. Il n'y a pas de consensus quant aux meilleures méthodes statistiques pour analyser ces données, car aucune ne peut traiter parfaitement les données manquantes. Nous proposons des méthodes permettant d'en tenir compte, comme les approches markoviennes, l'apprentissage statistique (classique sur différentes périodes de suivi ou par renforcement sur l'intégralité du suivi) et les *k-means* longitudinaux basés sur la forme de la trajectoire. Ce caractère longitudinal du critère principal impacte directement l'effet attendu du traitement de l'essai de prévention. Il n'est pas attendu au début de l'essai mais tout au long du suivi, de façon d'autant plus importante que celui-ci est long. Il est donc progressif [3, 4] et potentiellement non linéaire. De plus, l'étude des fonctions cognitives montre généralement un effet d'apprentissage sur ces fonctions. Une grande diversité de types de critères quantitatifs continus ou de trajectoires existe et il est particulièrement difficile de faire des hypothèses pertinentes sur la courbe représentant cet effet particulier. Des modèles à effets mixtes sont souvent utilisés pour analyser ces données [5, 6]. Ils traitent les données répétées en les considérant comme dépendantes au cours du temps et tiennent compte de la variabilité liée aux individus. Ces modèles ne tiennent pas compte des effets tardifs du traitement ou de l'histoire du parcours du sujet. Nous proposons d'utiliser un algorithme d'apprentissage par renforcement, traitant les données fonctionnelles directement, pour tenir compte de ces effets particuliers [7]. Dans le cas spécifique des essais de prévention du déclin cognitif et/ou de la maladie d'Alzheimer, le niveau global des fonctions cognitives a été reconnu comme le marqueur de substitution de la démence [8], bien que très variable d'un sujet à l'autre. Des tests neurocognitifs spécifiques, produisant des scores cognitifs quantitatifs, peuvent être combinés entre eux pour fournir un niveau global des fonctions cognitives [9, 10]. Ce niveau global, dont de la combinaison de Z-scores forme un Z-score composite, constitue une mesure de substitution de la cognition standardisée, référencée et validée. Le critère de jugement utilisé dans ce présent travail appartient à cette catégorie de Z-score composite continu.

Une difficulté supplémentaire survient lorsque l'évaluation du résultat est soumise à l'incertitude de mesure. Cette dernière augmente la variabilité, complexifie les analyses statistiques et masque parfois une non-linéarité des trajectoires. Pour traiter cette particularité, nous proposons d'utiliser des approches markoviennes modélisant l'effet de covariables sur l'évolution des fonctions cognitives dans le temps. Ces dernières ne sont pas traitées comme données fonctionnelles mais sont assimilées à un des paramètres de synthèse évoqué précédemment : la succession d'états au cours du temps, chaque état représentant un niveau d'évolution sur une durée. Par ailleurs, une hypothèse sur la forme de la trajectoire bruitée est émise : nous supposons qu'il existe des sujets s'améliorant par rapport au groupe placebo. Ces sujets sont qualifiés, dans notre cas, de répondeurs au traitement et/ou à l'intervention testée. Nous définissons ce qu'est un répondeur dans notre contexte et nous les comparons aux non répondeurs.

Le dernier point spécifique aux essais de prévention réside en l'hétérogénéité de la population d'analyse. En effet, les sujets plutôt âgés participant à un essai de prévention présentent un historique personnel de nombreuses comorbidités. La variabilité des mesures est augmentée en fonction de la période de suivi. Nous savons, par ailleurs, que l'atteinte cérébrale est variable d'un individu à l'autre [11]. Pour tenir compte de cette hétérogénéité, nous proposons trois méthodes de *clustering* permettant de rechercher des sous-groupes de

sujets homogènes. La première traite directement les données fonctionnelles mais fait une hypothèse sur le nombre de groupes (*k-means* pour données longitudinales basés sur la forme des trajectoires). La suivante utilise un paramètre de synthèse de la donnée fonctionnelle mais définit le nombre de groupes a posteriori (Classification Ascendante Hiérarchique). La dernière est une méthode combinatoire utilisant un paramètre de synthèse de la donnée fonctionnelle et choisit le nombre de groupes a priori (sémiologie graphique).

En somme, ces trois spécificités des essais de prévention du déclin des fonctions cognitives (critère de jugement principal longitudinal, incertitude de mesure et hétérogénéité de la population) pourraient être appréhendées différemment au moment de l'analyse pour limiter les essais peu convaincants [12, 13]. Ces essais n'ont pas été analysés avec d'autres méthodes a posteriori, une fois le résultat de l'analyse principale définie dans le plan d'analyse statistique produit. L'objectif principal de ce travail consiste à ré-analyser les données d'un essai de prévention non concluant. Pour cela, nous avons proposé différentes méthodes statistiques répondant à cet objectif. Ces méthodes apportent, pour chaque spécificité de l'essai identifiée ci-dessus, des informations complémentaires qui seraient passées inaperçues avec les techniques d'analyses les plus couramment utilisées. Elles ne se veulent pas exhaustives et proposent des solutions pour pallier des problèmes méthodologiques spécifiques aux essais de prévention. Nous proposons également deux façons d'aborder un critère de jugement quantitatif continu.

L'objectif de ce travail est de ré-analyser les données d'un essai clinique afin de mieux identifier des groupes de sujets ayant des trajectoires d'évolution particulières ou susceptibles de bénéficier de l'intervention. Pour cela, nos objectifs spécifiques sont de :

1. Déterminer des sous-groupes de trajectoires homogènes (profils) dont l'évolution des fonctions cognitives était particulière ;
2. Déterminer des sous-groupes de sujets répondant au traitement et/ou à l'intervention ;
3. Déterminer si des sous-groupes de sujets avaient un effet traitement/intervention différent ;
4. Déterminer l'effet de covariables sur l'évolution des fonctions cognitives ;
5. Évaluer les facteurs influençant les transitions entre différents états renvoyant à des niveaux cognitifs ;

Nous avons appliqué des méthodes peu mobilisées en recherche clinique pour ré-analyser les données d'un essai de prévention (*Multidomain Alzheimer Preventive Trial* - MAPT). Ces analyses secondaires, sur données réelles, n'ont pas pour but de revenir a posteriori sur la conclusion initiale d'un essai clinique mais sont destinées à détecter de possibles effets passés inaperçus et à émettre, le cas échéant, quelques recommandations quant à l'utilisation et/ou l'adaptation des outils dans le contexte de ré-analyse de données longitudinales d'un essai de prévention non concluant.

Avant de répondre à ces objectifs, nous avons situé ces travaux dans le contexte des essais cliniques et des essais de prévention, et en particulier, des essais de prévention du déclin cognitif et/ou de la maladie d'Alzheimer (Chapitre 2).

Le cas d'étude de l'essai de prévention MAPT est plus précisément décrit dans le Chapitre 3. Le protocole est explicité, en particulier le critère de jugement principal. Il est composite, en ce sens qu'il moyenne quatre critères de mesures cognitives, et son analyse est longitudinale, car il est relevé cinq fois au cours de 36 mois de suivi. De plus, le sens clinique et le mode d'utilisation dans les analyses des variables d'études sont présentés. Dans un

second temps, la population d'étude de MAPT ainsi que les analyses effectuées dans l'essai MAPT sont décrites.

Le Chapitre 4 se consacre au pré-traitement des données longitudinales. Nous expliquons en quoi ces données sont particulières et les façons dont nous les avons traitées par la suite. Dans certains cas, nous avons considéré le Z-score composite longitudinal directement, tandis que dans d'autres cas, nous avons privilégié un paramètre synthétisant la donnée fonctionnelle. Il correspond aussi bien à un indicateur de variations, le taux d'accroissement, qu'à une succession d'intervalles de variations au cours du temps, les états. Ensuite, les différentes populations d'étude de ce travail sont décrites. La dernière partie de ce chapitre propose deux stratégies d'analyses exploratoires préliminaires des données de l'essai MAPT.

Les outils statistiques de ré-analyse d'un essai de prévention sont exposés dans le Chapitre 5. Les différentes méthodes adaptées au contexte des données longitudinales et des schémas d'évolution non nécessairement linéaires sont discutées. La stratégie d'analyse consiste à utiliser des méthodes exploratoires d'une part et des méthodes de modélisation d'autre part.

Pour la phase exploratoire, des méthodes de recherche de sous-groupes homogènes, pour lesquels la différence entre les paramètres observés dans les bras de l'essai serait significative, ont été sélectionnées [3]. Des méthodes non supervisées basées sur les distances entre les trajectoires ont été exploitées. En traitant directement la donnée fonctionnelle et précisant le nombre de sous-groupes supposés (*clusters*), les k-*means* longitudinaux permettent de diviser de façon homogène la population étudiée [14]. En synthétisant la donnée fonctionnelle, une alternative pertinente consiste à appliquer une Classification Ascendante Hiérarchique (CAH) [15]. Dans ce cas, le nombre de groupe n'est pas déterminé a priori, à défaut de perdre le traitement de la donnée en l'état. En dernier lieu, toujours en synthétisant la donnée fonctionnelle mais en faisant également un a priori sur le nombre de groupes, l'analyse de la sémiologie graphique a été utilisée [16]. Ces méthodes d'analyses sont regroupées dans la famille des méthodes de *clustering* fonctionnel [17]. Parallèlement, une méthode de classification fonctionnelle supervisée est également abordée [1] d'après la définition du statut répondeur.

La seconde partie du Chapitre 5 concerne la modélisation de la trajectoire des sujets au cours du suivi. Des méthodes d'apprentissage statistique, prenant en compte l'incertitude de mesure et traitant directement la donnée fonctionnelle, ont été exploitées. De plus, en considérant la trajectoire comme une succession d'états au cours du temps, des méthodes markoviennes ont été appliquées [18, 19]. Les états représentent, dans notre cas, des intervalles de variations sur une période de temps. Ces approches permettent donc de modéliser l'évolution des fonctions cognitives. En particulier, les modèles de Markov cachés (*Hidden Markov Models* - HMM [20]) ont été utilisés pour pallier l'incertitude de mesure de la donnée fonctionnelle.

L'ensemble de ces méthodes a été appliqué à des données réelles et leurs résultat apparaissent en Chapitre 6. Le Chapitre 7 résume et conclut sur ces analyses.

Chapitre 2

Essais cliniques et maladie d'Alzheimer

La première partie du Chapitre 2, en section 1, présente tous les éléments du *design* d'un essai clinique. Elle se décompose en 6 grandes sections. La section 1.1 décrit les trois schémas d'étude les plus utilisés et la section 1.2 les critères d'éligibilité. La section 1.3 discute les différents critères de jugements d'un essai. Le critère de jugement principal, essentiel pour évaluer l'effet attendu d'un essai clinique, doit être soigneusement choisi. Deux possibilités s'offrent aux concepteurs des études : des critères usuels ou des critères plus complexes tels que les critères composites (section 1.3.3). Les critères usuels contiennent soit des critères qualitatifs (section 1.3.1), correspondant par exemple au niveau de gravité d'une maladie, soit des critères quantitatifs (section 1.3.2), renvoyant entre autres à la mesure d'une variable biologique. Le second groupe de critères correspond à la réalisation d'une variable de synthèse combinant plusieurs paramètres. Les principaux inconvénients sont le poids à attribuer aux différentes composantes ainsi que la difficulté d'interprétation clinique. Ce type de critère intègre le critère de jugement de l'essai utilisé dans ce travail. Il a été relevé de façon longitudinale (section 1.3.4). Suite à quoi, les outils assurant la comparabilité des groupes sont dépeints en section 1.4. L'analyse statistique apparait en section 1.5 et l'effet attendu, associé au nombre de sujets nécessaire, vient en section 1.6.

La partie 2 du Chapitre 2 fait référence à d'autres types d'essais cliniques. L'objectif d'un essai clinique classique est d'étudier l'effet d'un traitement sur des symptômes cliniques. Seulement, les essais de thérapie modificatrice, présentés en section 2.1, en constituent un cas particulier. La différence entre ces essais réside dans leur objectif : les essais de thérapie modificatrice visent à modifier le cours évolutif de la maladie en intervenant sur le processus pathologique. En dernier lieu, les essais de prévention sont dépeints en section 2.2, avec leurs spécificités : critères de substitution avec un éventuel effet d'apprentissage, données longitudinales, longue phase de latence avant de pouvoir détecter un effet.

La pathologie à l'étude dans ce travail de thèse, à savoir la maladie d'Alzheimer, intervient en partie 3 du Chapitre 2. En premier lieu section 3.1, nous rappelons les mécanismes physiopathologiques de la maladie d'Alzheimer. Y figurent la formation des plaques séniles ainsi que la dégénérescence neurofibrillaire. De plus, les différentes phases de la maladie sont exposées. Dans un second temps, les principaux symptômes de la maladie (section 3.2). Nous n'avons présenté ici que les 5 symptômes nous intéressant particulièrement dans la suite du manuscrit par rapport à l'évolution de la pathologie, à savoir la mémoire (en 3.2.1),

le langage (3.2.2), l'attention (3.2.3), les fonctions exécutives (3.2.4) et les fonctions visuo-spatiales (3.2.5).

La dernière partie du chapitre, située en section 4, présente le lien entre la maladie d'Alzheimer et les essais de prévention. Pour cela, la section 4.1 reprend les principaux arguments démontrant que la maladie d'Alzheimer constitue une maladie de choix pour mettre en place un essai de prévention. Ensuite, la section 4.2 liste, de façon non exhaustive, des essais de prévention de la maladie d'Alzheimer ou du déclin des fonctions cognitives. Pour chacun d'eux, sont comparés dans un tableau les différents éléments du *design*, les méthodes d'analyses statistiques et les principaux résultats. L'intérêt de cette comparaison des grands essais de prévention consiste à déterminer leurs points communs et leurs différences, tant sur la conception que les résultats, montrant ainsi en quoi notre intérêt de ré-analyse d'un essai non concluant se justifie.

1 Schéma général d'un essai clinique

« Un essai clinique est une recherche biomédicale organisée et pratiquée sur l'Homme en vue du développement des connaissances biologiques ou médicales. Les essais cliniques portant sur les médicaments ont pour objectif, selon le cas, d'établir ou de vérifier certaines données pharmacocinétiques (modalités de l'absorption, de la distribution, du métabolisme et de l'excrétion du médicament), pharmacodynamiques (mécanisme d'action du médicament notamment) et thérapeutiques (efficacité et tolérance) d'un nouveau médicament ou d'une nouvelle façon d'utiliser un traitement connu », d'après l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM) [21].

Pour se faire, un (ou plusieurs) traitement(s) d'intérêt est (sont) mis en concurrence avec un autre traitement [22, 23]. Ce dernier est soit le traitement de référence, soit un placebo (voir le référentiel Européen ICH E10 [24] pour les recommandations du choix du groupe contrôle dans un essai clinique). Chaque groupe de participants recevant le même traitement constitue un bras de l'essai, ce qui signifie qu'il y a autant de bras d'essai que de traitement ou intervention étudiés parallèlement.

Dans les sections 1.1 à 1.6 ci-dessous sont listés les éléments constituant un essai clinique. Nous avons particulièrement insisté sur les critères de jugement en section 1.3, les analyses statistiques en section 1.5 et l'effet attendu selon le nombre de sujets nécessaires pour un essai clinique classique (voir section 1.6). Ces trois points sont importants pour la suite du manuscrit.

1.1 Le schéma d'étude

Le schéma d'étude désigne la façon dont la comparaison entre les différents traitements est effectuée [25]. Il dépend de l'objectif de recherche, scindé en un principal et un (ou plusieurs) objectif(s) secondaire(s). Selon sa formulation, l'essai est dit [26, 27] : de supériorité, d'équivalence ou de non infériorité. Bien sûr, il existe d'autres schémas d'études plus complexes mais ils ne sont pas abordés ici.

1.2 Les critères d'éligibilité

Le deuxième point pour construire un essai clinique consiste à définir la population d'étude. La population cible, permettant d'extrapoler les résultats et au sein de laquelle l'échantillon d'étude est sélectionné, doit être spécifiée. L'échantillon répond à une liste de critères préalablement définis, dit d'éligibilité [28, 29]. Parmi ces critères figurent les critères d'inclusion et de non inclusion.

1.3 Le(s) critère(s) de jugement

D'après la *Food and Drug Administration* (FDA) [30], l'efficacité d'un traitement se définit comme un bénéfice clinique pour le patient. Elle se démontre par des résultats cliniques comme une disparition des symptômes ou des effets sur des paramètres de substitution établis. Par ailleurs, le phénomène d'intérêt s'établit par une technique de mesure spécifique générant une mesure unique par groupe de traitement.

Le fondamental pour ce travail consiste à étudier les formes que peuvent revêtir les critères d'évaluation (*outcomes*) de l'effet du traitement ou de l'intervention étudiés par rapport à un traitement comparatif [28, 29]. Si plusieurs caractéristiques sont évaluées, il est nécessaire d'en choisir une principale et une (ou plusieurs) secondaire(s). Ce critère principal est choisi a priori lors de l'élaboration du protocole de façon à être le plus pertinent vis-à-vis de l'objectif de l'essai. Les critères qualitatifs (section 1.3.1) et quantitatifs (section 1.3.2) sont usuels. Les critères composites (section 1.3.3) sont plus complexes et regroupent, en une unique mesure, plusieurs composantes qu'il est pertinent d'étudier simultanément. Ils nous intéressent particulièrement dans le présent travail. Indépendamment de leur nature, les critères peuvent être évalués de manière soit transversale, soit longitudinale. Les différentes façons de prendre en compte la temporalité sont indiquées en section 1.3.4 et détaillées de façon approfondie en Chapitre 4.

1.3.1 Critères qualitatifs

La variable prendra des valeurs nommées catégories, profils, classes ou encore niveaux [31]. Ils peuvent également être issus de données quantitatives et correspondre au fait d'être inférieur ou supérieur à un certain seuil.

Les critères qualitatifs contiennent parfois plus de deux modalités. Ces dernières sont ordonnées ou non. Dans le premier cas, la quantité mesurée est la même pour chaque modalité et chacune d'elle correspond à un niveaux graduel [32]. Par exemple, ces critères englobent les échelles de sévérité ou les intervalles de valeurs d'une donnée quantitative.

1.3.2 Critères quantitatifs

D'après Cummings et al. [33], un marqueur biologique correspond à une caractéristique objectivement mesurée par une grandeur quantitative et reconnue comme un indicateur du processus biologique normal, du processus pathogène ou de la réponse pharmacologique à une intervention thérapeutique. Selon le guide de pharmacogénomique de la FDA [34], un biomarqueur valide se mesure d'après un système de tests analytiques dont les caractéris-

tiques de performance sont bien établies. Il doit provenir d'un cadre scientifique déterminé ou être associé à un ensemble de preuves physiologiques, toxicologiques, pharmacologiques ou cliniques.

Cette catégorie d'*outcome* intègre les marqueurs ou paramètres de substitution (*surrogate* [35, 36]). Si le bénéfice clinique escompté est considéré comme important [36], ces derniers ont pour vocation de se substituer à un critère clinique [30]. En remplaçant ce paramètre difficilement observable, ils sont censés prédire un bénéfice clinique sur la base de preuves épidémiologiques, thérapeutiques, physiologiques ou autres preuves scientifiques [33]. Ainsi, tous les marqueurs de substitution sont des marqueurs biologiques, mais tous les marqueurs biologiques ne peuvent pas constituer des critères de substitution.

Dans les essais de prévention, le marqueur de substitution reflète l'évolution de la pathologie étudiée [33]. L'effet d'un médicament se traduit par un effet sur cette progression [36].

1.3.3 Critères composites

Lorsque les preuves de l'efficacité d'un traitement sur des critères classiques (quantitatifs ou qualitatifs) sont difficiles à mettre en évidence [30], l'utilisation d'un critère composite est une alternative [37]. Il est analysé comme un critère unique, sans faire appel à des tests de comparaisons multiples [30]. Sa construction n'est ni standardisée ni consensuelle et s'adapte à chaque pathologie [30]. Dans tous les cas, un critère composite vise à refléter différents aspects de l'effet du traitement [33].

Lorsque le critère composite est quantitatif, il s'appuie sur une combinaison de biomarqueurs ou de mesures quantitatives. De la même façon qu'avec les critères qualitatifs, chaque composante vise à mesurer une partie du phénomène étudié, fournissant finalement un critère de niveau moyen. Dans l'étude de la maladie d'Alzheimer, différents critères sont manipulés en fonction de l'avancement de la pathologie [38]. La FDA soutient l'utilisation de critères composites cognitifs comme critères de jugement principaux dans ces essais de prévention [39]. Ainsi, comme le soulignent Cummings et al. [38], deux critères composites sont utilisés pour les stades précoces de la maladie. Le premier est le critère composite cognitif préclinique d'Alzheimer (*Preclinical Alzheimer Cognitive Composite* - PACC) [40]. Sa construction provient de tests cognitifs évaluant plusieurs domaines tels que la mémoire épisodique, les fonctions exécutives et l'orientation. Le score final du PACC est égal à la moyenne de 4 scores cognitifs. Le second critère est le score composite cognitif d'initiative de prévention de la maladie d'Alzheimer (*Alzheimer's Prevention Initiative Cognitive Composite* - APICC) [41]. Il est également construit à partir de cinq composantes correspondant à des domaines cognitifs distincts.

Le principal avantage de considérer un critère composite est d'augmenter la capacité à détecter l'effet de l'intervention [33, 41]. Cependant, ces critères soulèvent plusieurs problèmes. Leur principale limite, comme indiqué ci-dessus, est de ne pas être standardisés ou consensuels. Cela rend la comparaison entre les essais plus difficile.

Concernant les critères quantitatifs, il faut définir comment les combiner entre eux. Avoir une idée du sens et de la temporalité des variations de chacune des composantes est important. Cette connaissance aide à comprendre les variations du critère composite final [37]. Le dernier point négatif des critères composites est d'être plus complexes à interpréter, principalement lorsque l'effet de l'intervention n'est pas uniforme sur toutes ses

composantes [42]. Dans ce cas, la description précise de l'impact de facteurs explicatifs sur le score composite devient difficile, surtout si le nombre de composantes augmente [42].

1.3.4 Critères longitudinaux

Les critères longitudinaux se retrouvent dans trois cas de figure, mais un seul est abordé ici. Nous revenons sur les autres possibilités (donnée de survie et donnée fonctionnelle) dans le Chapitre 4. Ainsi, les données longitudinales sont assimilées à des mesures répétées dans le temps pour un même sujet, sur une période définie a priori. Cette dernière peut varier en fonction des études, tout comme la fréquence de mesure.

Le principal avantage de ce type de données est d'apporter des renseignements sur les changements du processus au cours du temps, ce qu'une étude transversale ne peut pas faire. Comme il s'agit de mesures répétées pour un même sujet, les changements observés sont intra-individuels et inter-individuels pour l'ensemble de la population. De plus, lorsque le nombre de points augmente, l'information recueillie est plus précise. Des courbes d'évolution du phénomène peuvent alors être étudiées. En revanche, le revers de répéter des mesures est que, dans certains cas, un effet apprentissage peut être observé. Cela signifie que plus le nombre de mesures sera élevé, dans une temporalité plus ou moins courte, plus le sujet saura ce qui l'attend. Il peut donc soit s'entraîner pour améliorer ses performances au prochain test, soit se souvenir des résultats précédents et les restituer. L'effet d'apprentissage des composantes cognitives du critère de jugement de la présente étude est évoqué, variable par variable, en section 1.3.1 du Chapitre 3.

Dans notre analyse d'un essai de prévention, le critère de jugement principal est composite, quantitatif (section 1.3.3) et mesuré de façon répétée au cours du suivi.

1.4 La comparabilité des groupes

Pour pouvoir mener à bien l'essai, il faut s'assurer que le protocole d'étude garantisse la comparabilité des groupes. La randomisation consiste à allouer par tirage au sort l'un des traitements de l'étude à un sujet [28, 43]. Ainsi, lorsqu'un essai est randomisé, les groupes constitués sont comparables en tout point en début de suivi, sauf pour le traitement. De plus, cette méthode respecte le principe d'ambivalence, ce qui signifie qu'un traitement n'est pas préféré par rapport à un autre. L'essai pourra donc conclure sur l'efficacité d'un traitement [43].

Cette comparabilité des groupes n'étant valable qu'à l'inclusion, des techniques complémentaires doivent être mises en place pour la maintenir au cours du suivi. Premièrement, le critère de jugement précédemment défini doit être mesuré de façon homogène et identique quel que soit le groupe de traitement [28, 29]. Les critères objectifs sont préférables. Ses techniques de mesure, les conditions de mesures et tout autre paramètre contrôlable sont définis dans le protocole d'étude. Deuxièmement, la mise en place de l'insu (ou essai en aveugle) est souvent utilisée [28, 29]. Elle consiste à faire en sorte qu'aucune personne ne modifie son comportement durant l'étude en rendant le traitement administré inconnu. Elle peut s'intégrer à différents niveaux : seulement au niveau du patient (simple insu), au niveau du patient et du médecin ou de l'équipe soignante (double insu) ou encore au niveau du patient, du médecin et du statisticien en charge de l'étude (triple insu). La simple et double insu évitent les défauts d'adhésion au traitement tandis que la triple garantie éga-

lement l'absence de choix d'une méthodologie d'analyse particulière au vue des données ou des résultats.

Lorsque la comparabilité des groupes est assurée, aussi bien au début qu'au cours du suivi, nous avons l'assurance que la différence finale observée sur le critère de jugement est l'effet du traitement seul, toutes choses étant égales par ailleurs.

1.5 L'analyse statistique

Le dernier point d'un essai est son plan d'analyses statistiques [28, 29]. Il consiste à définir la stratégie envisagée selon le référentiel ICH E9 décrivant les principes statistiques pour les essais cliniques [44], à savoir :

— les tests de statistique descriptive à utiliser selon les types de variables (Student, Chi-deux, log-rank,...) ;
— les méthodes d'ajustement adaptées au critère de jugement (régression linéaire, logistique, à risques proportionnels, ...) ;
— la population à sélectionner dans les données (comparaison des groupes issus de la randomisation en intention de traiter, gestion des valeurs manquantes, ...).

Pour ce qui est de l'interprétation des résultats, cela se fait en deux temps. Tout d'abord, ils sont interprétés d'un point de vue statistique [28]. Dans le premier cas de figure, l'hypothèse nulle du test statistique est rejetée et la puissance est suffisante. L'essai conclut qu'il y a une différence statistiquement significative entre les groupes, avec une certaine erreur contrôlée (généralement le risque α est de 5 %). Dans le second cas, il n'est pas possible de rejeter l'hypothèse nulle. En supposant que la puissance a posteriori soit vérifiée et suffisante (c'est-à-dire supérieure ou égale à la puissance a priori), l'essai conclut que la différence entre les groupes n'est pas statistiquement significative, au risque d'erreur de première espèce α près. Dans un second temps, l'interprétation clinique des résultats statistiques intervient, ce qui signifie que leur pertinence est évaluée [28, 29].

Un essai clinique est considéré comme positif lorsque les paramètres observés en fin de suivi diffèrent au sein des groupes de traitement, de façon significative sur le critère principal. Lorsque cette différence n'est pas statistiquement significative, aucune conclusion sur les effets de l'essai clinique ne peut être établie. Il est alors considéré comme non concluant.

Ce dernier point nous intéresse particulièrement puisque notre étude porte sur un essai de prévention non concluant. Nous avons donc exploré les différentes hypothèses expliquant en partie ce manque de significativité. Premièrement, les conditions d'applications du test statistique ne sont pas respectées [45], ce qui signifie que la stratégie d'analyse statistique n'est pas pleinement adaptée. Ensuite, l'échantillon étudié est trop hétérogène, de taille faible ou contenant un taux de données manquantes élevé (manque de puissance statistique [46], voir section 1.6). Nous pouvons également supposer que les paramètres ne sont pas mesurés au bon moment, de façon non homogène entre les bras ou encore qu'ils ne correspondent pas bien au phénomène d'intérêt (sensibilité ou spécificité insuffisante). L'importance de la conception préalable de l'essai prend ainsi tout son sens. Finalement, il est possible qu'il n'y ait pas d'effet.

1.6 L'effet attendu et nombre de sujets nécessaire

Les essais cliniques, tels qu'ils ont été construits initialement, ont pour objectif d'étudier l'effet d'un traitement sur un paramètre clinique bien défini. Ce dernier est identique dans tous les bras parallèles de l'essai. Généralement, il traduit un effet symptomatique [36]. Lorsqu'un sujet est inclus dans un essai clinique, parce qu'il possède une certaine pathologie, il espère qu'en fin de suivi ses symptômes disparaissent [36]. Que la pathologie apparaisse brutalement ou de façon chronique, l'objectif du traitement est de conduire à une absence de symptômes. L'exprimer ainsi parait certainement futile à cette étape du manuscrit, mais cela prendra tout son sens en section 2 lorsque nous présenterons d'autres types d'essais cliniques.

Cet effet dépend des choix méthodologiques cités précédemment, à savoir un critère de jugement bien défini, une méthodologie statistique et une puissance statistique adaptées. Cette dernière suppose que le calcul du nombre de sujets nécessaires pour mettre en évidence une différence entre les groupes a été correctement effectué. La taille de l'échantillon doit être suffisante pour détecter un effet et doit prendre en compte la durée de suivi, le nombre de bras de l'essai, le risque d'erreur de première espèce α souhaité, la puissance statistique 1-β, le nombre de perdus de vue estimé au cours du suivi ainsi que la nature de la différence (comparaison de moyennes, de proportions, ...).

2 Les essais de thérapie modificatrice et de prévention

À présent que le cadre d'un essai clinique classique a été défini en section 1, notre intérêt s'est porté sur certaines particularités. Elles donnent lieu à des essais différents tels que les essais de thérapie modificatrice (section 2.1) ou de prévention (section 2.2). Le présent travail s'inscrit dans le cadre d'un essai de prévention.

2.1 Les essais de thérapie modificatrice

Depuis quelques années et pour certaines pathologies chroniques dont la maladie d'Alzheimer, les essais dits *disease modifying therapy* ont été mis en place [3, 47]. Contrairement aux essais symptomatiques présentés en section 1, dont l'objectif était d'évaluer l'efficacité du traitements sur les symptômes de la maladie, les présents essais analysent l'impact de facteurs de progression d'une pathologie [48]. Les essais de thérapie modificatrice diffèrent des essais symptomatiques en ce sens qu'ils visent à modifier le cours évolutif d'une maladie en intervenant sur le processus pathologique sous-jacent [3].

La modification des maladies n'a pas été définie de manière rigoureuse et aucune définition consensuelle n'est disponible à l'heure actuelle [33]. Cependant, d'après la FDA, seules deux stratégies peuvent amener à une modification de maladie [48] : soit un bénéfice clinique est significativement démontré sur un biomarqueur, soit un effet durable sur l'évolution de la maladie est constaté. Selon l'Agence européenne des médicaments (EMA) [48], un traitement est considéré comme un modificateur d'une maladie s'il retarde les processus pathologiques ou physiopathologiques sous-jacents. En particulier, l'EMA s'accorde avec la FDA pour dire qu'un bénéfice clinique, établi sur un biomarqueur validé de manière adéquate et reflétant les principaux aspects physiopathologiques de la maladie, est suffisant

pour amener à une modification de maladie. Ce bénéfice se traduit par exemple par une diminution significative du taux de signes ou symptômes cliniques [48].

La principale différence entre les essais de thérapie modificatrice et les essais cliniques symptomatiques réside dans l'effet attendu de l'intervention testée. Plus précisément, une durée du suivi plus longue les caractérise par rapport aux essais symptomatiques [33, 47, 49]. Un essai de thérapie modificatrice a été défini comme une modification de l'histoire naturelle d'une maladie sur le long terme [3, 47]. La durée minimale recommandée s'élève à 12 mois et l'effet de modification du processus doit encore être observable plusieurs mois après l'intervention d'après la FDA, en particulier plus de 18 mois pour la maladie d'Alzheimer [3]. En moyenne, un tel essai s'étend sur une période de 18 mois [3, 49]. De même, ces essais nécessitent un plus grand nombre de sujets pour mettre en évidence un effet potentiel [47].

Concernant le (les) critère(s) de jugement, des critères cliniques longitudinaux classiques [3] ou le taux de progression de la maladie sont à considérer [33]. Par exemple, une diminution du taux de celui-ci entre 30 % et 50 % est admis comme raisonnable pour la maladie d'Alzheimer [3]. En alternative, des biomarqueurs de la maladie ou des critères de substitution (*surrogate*) sont utilisés pour spécifier le critère principal [3, 48, 49], toujours mesurés de façon longitudinale. En effet, selon la pathologie étudiée, séparer les effets symptomatiques des agents potentiels n'est pas toujours aisé. Certains critères d'évaluation sont influencés par les effets symptomatiques du traitement d'intérêt et ne sont pas en mesure de différencier cet effet de la modification du processus pathologique [49]. L'utilisation de critères classiques n'est donc pas toujours adaptée et le recours aux biomarqueurs comme critères de jugement se justifie si ces derniers ne subissent pas l'influence des effets symptomatiques des traitements.

Finalement, à propos des autres aspects du *design* de l'étude (critères d'éligibilité, comparabilité des groupes, analyse statistique), les essais de thérapie modificatrice ne diffèrent pas particulièrement des essais cliniques symptomatiques [3]. D'un point de vue statistique, Cummings et al. [48] ont listé quatre approches communes aux deux types d'essais : la mesure de la différence en fin d'essai entre les groupes de traitement [33], la durée avant apparition d'un évènement (par exemple, pour la maladie d'Alzheimer, une progression d'un indice de démence d'un score élevé à un score plus faible [33, 48]), l'augmentation de la différence entre les groupes de traitement au cours du temps et le changement de pente d'un critère quantitatif au cours du suivi.

2.2 Les essais de prévention

Les essais de prévention constituent un cas particulier des essais de thérapie modificatrice. Dans leur cas, la modification de la maladie se définit comme retarder la progression ou l'apparition de celle-ci en modifiant le processus pathologique. Ces essais s'intéressent à l'évaluation d'approches impactant le taux d'incidence ou de progression de la maladie. En fonction de la maladie, la majeure partie des sujets sélectionnés ne rencontrera pas l'évènement d'intérêt et ce malgré une longue liste de critères d'éligibilité [3, 47, 49]. Autrement dit, il est attendu que, chez certains individus seulement, une variation des paramètres associés à la pathologie soit significativement caractérisée. Ce sous-groupe n'est pas facilement identifiable a priori.

Comme dit précédemment, les essais de prévention possèdent certaines spécificités rendant leur conclusion difficile à étudier. Premièrement, ces études portent habituellement

sur les résultats longitudinaux [1] plutôt que sur la différence d'effet entre les groupes de traitement en fin d'étude [3, 33]. Cela signifie que chaque variable est mesurée plusieurs fois au cours du suivi pour un même sujet. Ces mesures étant interdépendantes, les choix méthodologiques statistiques doivent tenir compte de cette spécificité ainsi que des potentiels effets de seuil des variables. De plus, les études longitudinales en elles-mêmes ne sont pas conçues pour distinguer les effets symptomatiques ou modificateurs [33]. Il en découle que l'effet de l'intervention n'est pas attendu au début de l'essai mais il l'est à la fin du suivi ; et de façon d'autant plus important que le suivi est long. L'effet attendu de l'intervention est donc progressif [3, 4] et non nécessairement linéaire. Le dernier point, toujours lié au critère longitudinal, est que ces études sont sensibles à la durée du suivi [49], en particulier pour la prévention à long terme. Cette durée est d'autant plus importante que l'incidence de la maladie est faible, ou que le taux de progression est lent [3]. Cela signifie aussi que le nombre de perdus de vue au cours du suivi augmente, tout comme la probabilité de données manquantes intermittentes. En effet, les essais plus longs possèdent un taux d'attrition plus élevé ainsi qu'une plus grande variabilité des mesures dans le temps [33, 50]. De la même façon, nous avons déjà présenté dans le Chapitre 1 la spécificité de l'incertitude de mesure et de l'hétérogénéité de la population d'étude. Parmi les principaux problèmes au stade de la conception de l'essai figurent le choix de la période de suivi, du critère de jugement principal et de la méthode d'analyses statistiques.

Ainsi, puisque les méthodes d'analyses de ces différents types d'essais cliniques sont identiques malgré les spécificités évoquées ci-dessus, la question est de savoir si les méthodes statistiques utilisées sont réellement adaptées aux essais de prévention. Autrement dit, est-ce que les hypothèses associées aux méthodes classiquement employées sont bien vérifiées et prennent en considération la nature progressive de l'effet attendu au cours du suivi ? De même, est-ce que ces méthodes permettent de distinguer les sujets bénéficiant rapidement de l'intervention de ceux qui en bénéficient plus lentement ou pas du tout ?

Cette remise en question des méthodes statistiques a déjà été abordée dans un précédent travail [51]. Lorsque le critère de jugement principal de l'essai est un résultat binaire, évalué en fin du suivi, la méthode statistique empruntée est l'analyse de données de survie. Elle compare les probabilités de survie des différents groupes au moyen d'un test du log-rank. Dans le cadre d'un essai de prévention, l'hypothèse de proportionnalité des risques n'est pas respectée. Les conséquences sur la conclusion d'un essai de prévention ont été étudiées dans un article de Garès et al. [51]. Les auteurs ont démontré que l'analyse par un test du log-rank conduisait à la conclusion d'une absence d'effet statistiquement significatif. En revanche, l'analyse de ce même essai avec le test de Fleming-Harrington, permettant de passer outre l'hypothèse de proportionnalité des risques, concluait à un effet significatif. De plus, ce résultat était pertinent cliniquement, ce qui signifie que l'essai aurait été positif après application de ce choix méthodologique. Constater cette différence d'effet en fonction de l'outil utilisé a conduit Garès et al. [52], dans un autre travail, à développer des méthodes spécifiques. Elles tiennent compte de la caractéristique longitudinale des données d'un essai de prévention et fournissent une formule de re-calcul du nombre de sujets nécessaires.

Lorsque le critère de jugement principal de l'essai est de nature continue (voir section 1.3), l'analyse devient plus complexe. Effectivement, la mesure de la temporalité n'est plus directe mais découlant de mesures répétées dans le temps. De plus, l'éventail des analyses de survie à risques non proportionnels est très large et le test de Fleming-Harrington couvre la majeure partie des situations. Seulement, les résultats continus conduisent à une grande diversité de trajectoires et il est particulièrement difficile de faire des hypothèses

pertinentes sur la nature de la courbe. Des modèles à effets mixtes, des modèles mixtes à classes latentes ou quadratiques sont souvent exploités naturellement dans ce cas de figure [5, 6, 53, 54]. En fonction des hypothèses émises, ces modèles impliquent de faire des choix a priori que d'autres modèles ne requièrent pas.

3 Pathologie à l'étude : la maladie d'Alzheimer

3.1 Présentation de la maladie d'Alzheimer

En 2013, la prévalence mondiale de la démence, tous types confondus, était estimée à 36 millions de personnes atteintes. En 2015, cette prévalence représentait 46,8 millions de personnes dans le monde. En 2017, le nombre de sujets atteints constituait 5 % de la population Européenne de cette année là [55]. D'après l'Organisation Mondiale de la Santé, ce nombre culminait à plus de 50 millions en 2019. Des estimations ont montré que le nombre de sujets atteints de démence, dans le monde, devrait continuer d'augmenter et atteindre, en 2050, environ 115,4 millions de personnes [56].

La maladie d'Alzheimer est aujourd'hui la forme de démence la plus fréquemment rencontrée [57] et représente environ 70 % des démences toutes causes [58]. Elle est neurodégénérative, ce qui signifie que les altérations cérébrales sont progressives et irréversibles. Ces atteintes ne sont homogènes ni en terme de localisation dans les différentes zones du cerveau, ni en terme d'intensité [11]. La maladie d'Alzheimer est en particulier caractérisée par des anomalies physiopathologiques détectables plusieurs années avant les premiers symptômes cliniques [59]. Autrement dit, cette maladie possède une longue phase de latence [60]. Elle est caractérisée par des lésions cérébrales spécifiques, à l'image des plaques amyloïdes [61], et des dégénérescences neuro-fibrillaires [62]. Elles sont généralement constatées à l'occasion d'un examen post-mortem établissant avec certitude leur présence [63].

Cette pathologie évolue en plusieurs stades [57] et débute avec une longue phase de latence. Durant celle-ci, les processus pathologiques se mettent en place, y compris ceux concernant les plaques amyloïdes et la protéine TAU. La particularité de ce début de processus est d'être asymptomatique [64]. Lors d'un déficit cognitif léger, le sujet se plaint de troubles de la mémoire, d'oublis des mots courants ou d'endroit où se trouvent certains objets de la vie quotidienne.
Ensuite, vient la phase symptomatique de déclin lent et léger correspondant à un stade transitoire entre la phase précédente et la démence avérée [65]. Sa caractéristique principale est la plainte cognitive, cliniquement observée par une atteinte de la mémoire et de la concentration. Les sujets sont cependant toujours indépendants et les activités de la vie quotidienne maintenues. Cette phase inclut un autre stade caractérisé par un déclin modéré. Cette fois-ci, les symptômes sont plus spécifiques à la progression de la maladie. Parmi eux se trouvent l'oubli d'événements récents, l'altération de la capacité à résoudre des calculs mentaux. Le patient a des difficultés à exécuter des tâches complexes de la vie quotidienne. De plus, l'oubli de son propre passé, des sautes d'humeur ou un comportement d'effacement sont caractéristiques de ce stade.
Au stade du déficit cognitif modérément sévère, le sujet est moins autonome dans la réalisation des activités de la vie quotidienne (manger, aller aux toilettes, ...). De plus, la désorientation spatio-temporel est évidente et l'individu n'arrive plus à résoudre des problèmes simples.

Remarquons qu'avec ces symptômes s'installant progressivement, il n'est pas aisé de fournir une date exacte de début de chaque phase. Il n'existe aucun traitement curatif [57], mais des traitements sont disponibles pour améliorer temporairement les symptômes.

3.2 Les principaux symptômes de la maladie

La principale caractéristique clinique de la maladie d'Alzheimer réside dans une atteinte progressive des fonctions cognitives [57, 66]. Plusieurs études ont démontré que le déclin cognitif n'était ni homogène entre les sujets [57, 67], ni linéaire au cours du temps [57]. De plus, comme les éléments biologiques, ce déficit cognitif surviendrait plusieurs décennies avant le diagnostic. En effet, des épidémiologistes ont récemment publié une étude montrant que le déclin cognitif serait visible vers 45-49 ans [68].

Dans la présente section, nous présentons les principaux symptômes de la maladie d'Alzheimer. Cette liste n'est pas exhaustive.

3.2.1 La mémoire

La mémoire renvoie à une aptitude permettant d'enregistrer (encodage), de conserver (stockage) et de restituer une information donnée (récupération). Elle se compartimente en plusieurs domaines spécifiques [67]. La mémoire sensorielle enregistre toute nouvelle information pendant quelques centièmes de secondes [69]. Elle se scinde en mémoire à court terme et en mémoire à long terme [67]. La mémoire à court-terme concerne l'enregistrement temporaire d'une information (limitée à 7 ± 2 informations stockées) sur une durée de quelques secondes à quelques minutes. Nous utilisons puis oublions l'information en question (mémoire de travail). Cette dernière est particulièrement atteinte dans la maladie d'Alzheimer (amnésie) et s'observe par des lésions cérébrales frontales [63, 69]. Par contre, la mémoire à long terme, partie intégrante de la mémoire sensorielle, fait référence à la conservation durable voire définitive de l'information. Celle-ci a donc été encodée, stockée et peut être récupérée ultérieurement en faisant appel à des souvenirs récents ou anciens. Ces deux mémoires sont principalement utilisées lorsque nous stimulons notre ouïe ou l'acuité visuelle.

Plus spécifiquement encore, la mémoire à long terme est composée de la mémoire déclarative et de la mémoire non déclarative [67]. La première se divise en la mémoire sémantique (savoir général sur sa propre vie, comme sa couleur ou son plat préféré) et épisodique (évènements vécus et indexés dans leur contexte spatio-temporel et émotionnel, comme une réunion de famille) [67]. La mémoire non déclarative contient les effets d'amorçage et l'apprentissage de compétences [67]. La mémoire dite procédurale en est sa principale composante.

Notons qu'un sujet atteint de la maladie d'Alzheimer et présentant un déficit de la mémoire épisodique à long terme montrera des performances plus faibles en rappel libre, en rappel aidé d'indices ou encore en reconnaissance (agnosie) [67]. En effet, Shimamura et al. [70] ont démontré que les scores cognitifs diminuaient de façon significative pour la mémoire de travail et la mémoire de reconnaissance. De la même façon, lorsque la mémoire sémantique est évaluée, les sujets malades présentent des déficits dans les tâches de dénomination, de fluidité verbale, de vocabulaire et d'achèvement des phrases [67].

3.2.2 Le langage

Le langage constitue le moyen de communication humaine, passant par la parole et impliquant un mécanisme de compréhension. Le sujet malade va progressivement avoir des difficultés à communiquer et à comprendre ce qui est dit autour de lui. Comme indiqué en section 3.2, ce trouble du langage va même jusqu'à des difficultés d'expression dans le stade terminal (aphasie).

3.2.3 L'attention

L'attention définit la capacité à se focaliser sur une information et à la maintenir dans la conscience [63]. Elle peut être sélective, ce qui signifie qu'une seule source d'information est traitée à la fois, sans se laisser distraire. L'attention peut au contraire être divisée pour se répartir entre plusieurs tâches [63]. Enfin, elle s'exprime à différents degrés d'intensité (alerte, vigilance, etc). Cette fonction cognitive est indispensable dans les activités de la vie quotidienne.

3.2.4 Les fonctions exécutives

Avec la progression de la maladie, les fonctions exécutives sont également altérées [69]. Elles se mobilisent pour contrôler des comportements volontaires et réaliser des tâches complexes ou nouvelles. Elles sont aussi impliquées lors d'une situation routinière. Concrètement, ces fonctions renvoient à la planification, au raisonnement et à la concentration. Elles font appel aux processus d'inhibition, de flexibilité mentale et d'adaptabilité. Elles sont associées aux capacités de prendre des décisions et de résoudre des problèmes.

3.2.5 Les fonctions visuo-spaciales

De la même façon, la capacité d'un sujet malade à s'orienter dans l'espace, à percevoir les objets de son environnement et à les organiser en scène visuelle cohérente se dégrade avec la progression de la maladie. Cette fonction cognitive est associée à l'imagerie mentale, c'est-à-dire à la capacité à s'imaginer un objet et à le manipuler mentalement.

4 Les essais de prévention de la maladie d'Alzheimer

La dernière partie du Chapitre 2 présente le lien entre la maladie d'Alzheimer et les essais de prévention. La section 4.1 fait état des principaux argument de notre pathologie d'étude par rapport au contexte des essais de prévention évoqués précédemment. La section 4.2 compare, sous différents aspects du *design* d'étude, les grands essais de prévention de la maladie d'Alzheimer et/ou du déclin des fonctions cognitives. Ces comparaisons mettent en lumière notre objectif de proposer différentes méthodes statistiques de ré-analyse des résultats d'un essai de prévention non concluant.

4.1 Généralités

Si la maladie d'Alzheimer n'est pas curable [8, 71], il est cependant possible d'agir en amont de l'apparition des symptômes. Cela rend cette pathologie particulièrement adaptée à la mise en place des essais de prévention. En effet, il n'y a aucun moyen de prédire avec précision la vitesse de progression de la maladie à travers ses différents stades [57]. De plus, cette évolution est propre à chaque sujet.

Les essais de prévention de la maladie d'Alzheimer sont difficiles à concevoir pour plusieurs raisons [12, 72]. Tout d'abord, la phase asymptomatique de la pathologie survient des décennies avant les premiers symptômes (entre 10 et 20 ans pour les plaques séniles [73]), influençant le choix de la durée de suivi. Ensuite, aucun résultat directement lié à la démence n'est mesurable, mais il en existe des composites raisonnablement liés à son évolution [48] ou reconnus comme marqueurs de substitution. Le choix de la population d'étude est également important. Les sujets participant aux essais de prévention sont généralement des patients âgés de 65 ans et plus. Ils présentent déjà un historique personnel de nombreuses comorbidités. Parmi elles, se trouvent des maladies, s'exprimant encore au moment de l'inclusion dans l'essai ou non, qui ont été traitées ou non. Elles peuvent avoir des conséquences directes ou indirectes sur les fonctions cognitives et/ou le critère de jugement principal de l'essai. Cette accumulation de paramètres peut générer de l'hétérogénéité sur les critères mesurées dans l'essai, bien que la randomisation assure un certain contrôle. Si ces paramètres accumulés sont très liés à l'*outcome* d'intérêt et non contrôlés par la randomisation, alors la communauté scientifique peut se heurter à des difficultés d'interprétation. Il est également possible que l'ajustement statistique sur ces paramètres ne soit pas suffisant ou ne les considère pas tous, ce qui constitue un frein à l'interprétation correcte des résultats de l'essai.

4.2 Approches comparatives de grands essais de prévention

Nous présentons les principaux essais cliniques de prévention menés ces dernières années sur la maladie d'Alzheimer, ou plus généralement sur le déclin des fonctions cognitives. L'objectif est de dresser un panorama des types d'essais, des critères de jugement utilisés, des méthodes statistiques appliquées ainsi que des résultats obtenus. De façon à avoir une vue d'ensemble des essais présentés, le Tableau 2.1 reprend les caractéristiques principales de construction de ces derniers. Par la suite, nous décrivons les critères de jugement, les analyses statistiques menées ainsi que les principaux résultats. Le Tableau 2.2 résume ces derniers éléments d'analyse des essais. Il fournit ainsi un support synthétique et comparatif de ceux que nous présentons ici.

Le Tableau 2.1 dresse ainsi un panorama des principaux essais cliniques récents dans le domaine de l'étude de la maladie d'Alzheimer ou des fonctions cognitives. À la lumière de cette synthèse, nous constatons qu'ils ont certains points communs. Premièrement, ces essais étaient tous en adéquation avec la discussion menée en section 1.3 sur les différents critères de jugement. Leur durée de suivi était suffisante par rapport aux recommandations de la FDA ou de l'EMA. Le plus court d'entre eux durait 18 mois tandis que les deux plus long duraient 84 mois, soit 7 ans. Ensuite, le nombre de sujets sélectionnés lors des analyses était suffisant pour assurer la puissance statistique nécessaire à la détection d'un effet. La majorité d'entre eux avaient mis en concurrence classiquement deux bras parallèles.

Par ailleurs, ces essais différaient davantage sur leurs caractéristiques méthodologiques

Table 2.1 – Caractéristiques principales de la construction des différents essais.

Nom de l'essai	Références	Durée de suivi (en mois)	Sujets*	Bras**	Pays	Année de publication
GEM	[65, 74]	72	3069	2	États-Unis	2008
GuidAge	[9, 75, 76]	60	2820	2	France	2012
preDIVA	[77, 78]	72	3454	2	Pays-Bas	2016
PREADVISE	[79]	84	3786	4	États-Unis et Canada	2017
SPRINT MIND	[80, 81]	60	8563	2	États-Unis	2018
LIFE	[82, 83]	24	1476	2	États-Unis	2015
ACTIVE	[84]	24	2802	4	États-Unis	2002
FINGER	[85, 86]	24	1190	2	Finlande	2015
MAPT	[13, 87]	36	1680	4	France et Monaco	2017
HATICE	[88, 89, 90]	18	2398	2	Pays-Bas, Finlande et France	2019

* : correspond ici au nombre de sujets analysés.
** : correspond ici au nombre de bras parallèles d'intervention, incluant le bras placebo.

(Tableau 2.2). En terme de critère de jugement principal, 4 s'appuyaient sur un critère composite, 4 sur l'incidence (brute ou cumulée) de la démence et 1 sur des mesures brutes de tests cognitifs. De plus, deux d'entre eux considéraient un double critère de jugement principal, ce qui n'est pas très courant dans ce type d'essais cliniques. Par ailleurs, le nombre de mesures effectuées au cours du suivi des critères était variable. Notons cependant que pour l'essai preDIVA, nous avons indiqué 18 visites (1 visite tous les 4 mois) mais seules 6 d'entre elles seront considérées pour les analyses (restriction à 2 ans de suivi). Par contre, sur les 10 essais présentés, 6 d'entre eux ont appliqué des modèles de survie de Cox à risques proportionnels et 6 d'entre eux ont choisi des modèles pour données longitudinales (mixtes ou généralisé). Ces stratégies d'analyses font parties des méthodes classiquement employées pour l'analyse des résultats d'un essai clinique classique. Finalement, l'autre point de comparaison des essais est l'obtention (ou non) de résultats et la conclusion finale de l'essai. La majorité d'entre eux n'ont pas montré d'effet statistiquement significatif et étaient donc considérés comme des essais non concluants.

Ceci nous amène au présent travail de thèse. Nous avons fait l'hypothèse que la méthodologie d'analyses statistiques appliquée pour détecter un effait du traitement mis à l'épreuve pouvaient, en partie, expliquer l'absence d'effet constaté dans ces essais de prévention du déclin des fonctions cognitives. En particulier, d'après les spécificités intrinsèques aux essais de prévention, un effet potentiel d'un des traitements à l'étude aurait pu passer inaperçu en raison de la méthodologique utilisée. À ceci s'ajoute la supposition d'une construction de critère de jugement non adaptée au phénomène étudié. Nous avons vu que les résultats des essais dépendaient fortement de ce critère, de la méthode d'analyse et du nombre de sujets nécessaires. En dernier lieu, l'hypothèse que nous avons testé, en sus des deux précédentes, porte sur l'hétérogénéité de la population d'étude. Le bruit contenu dans les données peut masquer des sous-groupes de sujets pour lesquels l'effet attendu est significatif. Ainsi, l'ensemble de ces hypothèses ont été testées à travers les prochains chapitres du manuscrit et sur les données réelles de l'essai MAPT (voir Chapitre 3). Pour cela, nous avons proposé notamment un pré-traitement des données en Chapitre 4 et un panorama de

différentes méthodes de ré-analyse d'un essai de prévention en Chapitre 5.

Table 2.2 – Synthèse des caractéristiques méthodologiques des différents essais.

Nom de l'essai	*Outcome*	# Points	Modèles d'analyse	Résultats	Conclusion
GEM	Incidence démence	4	GEE + Cox proportionnels	*Outcome* identique	Non concluant
GuidAge	Quantitatif + Incidence démence	5	Mixtes + Cox proportionnels	-1 point *outcome* multiple le risque de démence par 3,5	Non concluant
preDIVA	Incidence cumulée démence + ALDS	18	Mixtes + Cox proportionnels	*Outcomes* identiques dans les bras	Non concluant
PREADVISE	Incidence démence	2	Cox proportionnels	*Outcome* identique	Non concluant
SPRINT MIND	Incidence démence probable	4	Cox proportionnels	Réduction incidence démence légère	Non concluant
LIFE	DSC + HVLT-R	2	ANCOVA + Cox proportionnels	*Outcome* identique	Non concluant
ACTIVE	Composite	4	Mixtes	Amélioration cognitive	Concluant
FINGER	Composite	2	Mixtes + Logistiques	Intervention multidomaine bénéfique	Concluant
MAPT	Composite	5	Mixtes	Aucune différence IM	Non concluant
HATICE	Composite	2	Linéaires généralisés	Intervention améliore	Concluant

est une abbréviation pour "nombre de"..

Chapitre 3

Le cas d'étude : l'essai Multidomain Alzheimer Prevention Trial

Le cas d'étude de cette thèse a été choisi en fonction des éléments de contexte fourni dans le Chapitre 2. Ce dernier détaille les spécificités du *design* d'un essai de prévention, dont les principaux points discutés sont les analyses statistiques appliquées et le critère de jugement de l'essai. De plus, nous avons focalisé notre attention sur la maladie d'Alzheimer car elle représente un domaine dans lequel la prévention se développe de façon importante et les chercheurs sont confrontés à des problèmes méthodologiques particuliers. Nous avons donc sélectionné un essai de prévention de la maladie d'Alzheimer.

Les données réelles de l'essai de prévention *Multidomain Alzheimer Preventive Trial* (MAPT [13, 87]) ont été ré-analysées grâce aux méthodes présentées en Chapitre 5. Les informations fournies ci-après ne sont pas exhaustives et ne concernent que celles nécessaires à la compréhension de la présente étude.

La section 1 reprend les éléments du *design* de l'étude. S'y trouvent ainsi les critères d'éligibilité (section 1.1) et l'objectif de l'essai (section 1.2). Ensuite, les critères de jugement sont abordés en section 1.3. En section 1.3.1 est présenté le critère de jugement principal de l'essai, de nature composite et longitudinale, ainsi que chacune de ses composantes en terme de signification et variation attendue par rapport au déclin des fonctions cognitives. Les critères de jugement secondaires de MAPT apparaissent en section 1.3.2 ainsi que les autres facteurs recueillis dans MAPT et pris en compte dans les analyses de ce présent travail (section 1.3.3). Ces trois sections constituent le référentiel des variables utilisées pour la ré-analyse de l'essai MAPT. La dernière partie de cette section présente le plan d'analyses statistiques (section 1.4).

Par ailleurs, la section 2 détaille la population d'étude d'une part (section 2.1), d'autre part les analyses statistiques et les principaux résultats obtenus dans l'article publié de MAPT (section 2.2). Ces analyses ont été menées préalablement au présent travail de recherche.

1 Présentation du protocole

Dans cette première partie, le protocole de l'essai MAPT est présenté. En particulier, les critères de l'étude ont été détaillés (section 1.3). En effet, le critère de jugement principal de l'essai étant composite, nous le présentons en section 1.3.1 ainsi que chacune de ses composantes. De plus, les critères de jugement secondaires sont abordés en section 1.3.2, ainsi que d'autres facteurs recueillis dans MAPT et utilisés par la suite dans nos analyses (section 1.3.3). Ce descriptif des variables est important pour comprendre les résultats des analyses présentés en Chapitre 6 ainsi que leur interprétation clinique.

1.1 Schéma d'étude et critères d'éligibilité

Le schéma d'étude de MAPT était un essai de supériorité randomisé, multicentrique (13 centres en France et Monaco), contrôlé versus placebo, en insu du traitement et comportant 4 groupes parallèles. Ces derniers comparaient trois groupes d'intervention et un contrôle. Ils étaient désignés comme suit : une intervention multidomaine (IM, détaillée ci-après) associée à la prise d'oméga 3 (Ω_3) ou à un placebo.

La période d'inclusion s'étendait du 30 mai 2008 au 24 février 2011. Les sujets éligibles à cet essai étaient à risque de développer la maladie d'Alzheimer. Les investigateurs d'étude ont obtenu le consentement éclairé écrit pour chacun des participants. Le protocole a été autorisé par le comité d'éthique de Toulouse (CPP SOOM III) [13]. L'essai a été enregistré sur Clinicaltrials.gov (NCT00672685). Plus précisément, les critères d'inclusion des sujets correspondaient à :

- Avoir 70 ans ou plus ;
- Vivre à domicile ;
- Avoir au moins l'un de ces critères :
  1. plainte mnésique exprimée à son médecin traitant ;
  2. limitation dans la réalisation d'une des activités instrumentale quotidiennes (*Instrumental Activities of Daily Living* - IADL [91]) ;
  3. vitesse de marche lente ($\leq$ 0,77 m/s ou plus de 5 secondes pour parcourir 4 mètres en marchant).
- Avoir un score au *Mini-Mental State Examination* (MMSE) $\geq$ 24 points ;
- Être en capacité de comprendre le protocole et la procédure de l'essai ;
- Être en capacité de donner son consentement écrit et éclairé et accepter les contraintes de l'étude ;
- Avoir assez de temps pour participer à l'IM ;
- Selon l'opinion de l'investigateur, être capable de suivre le traitement tout au long du suivi ;
- Avoir une couverture de sécurité sociale.

Les critères de non inclusion était très nombreux et sont détaillées dans les annexes de l'article de MAPT [13]. Nous ne citons que les principaux :

- Avoir été diagnostiqué comme ayant une maladie d'Alzheimer ;
- Avoir une détérioration cognitive (MMSE $<$ 24 points) ;
- Être dépendant dans au moins une des activités élémentaires quotidiennes (*Alzheimer Disease Cooperative Study - Activities of Daily Living Inventory - Prevention Instrument* ADCS-ADL-PI $<$ 6 points),

— Avoir pris des Ω_3 avant inclusion dans l'essai (dans les 6 derniers moins) ou à l'inclusion.

La comparabilité des groupes à l'inclusion a été assurée d'une part grâce à une randomisation équilibrée de ratio 1 :1 :1 :1 entre les bras d'intervention, par bloc de taille 8 et stratifiée par centre. D'autre part, l'essai a été mené en double insu du traitement (Ω_3 versus placebo) mais pas de l'IM. Les mesures des tests cognitifs ont été effectuées par des neurologues indépendants de l'essai et n'ayant pas connaissance du groupe d'allocation.

Durant les 2 premiers mois de l'essai, 12 sessions de groupe de l'IM ont été réalisées (phase intensive). Chaque session contenait 1 heure de stimulations mémorielles et de raisonnement, 45 minutes de conseils et démonstration d'exercices physique et 15 mn de conseils nutritionnels. L'IM contenait également des séances mensuelles de 1 h ainsi que trois consultations préventives de la maladie, administrées à l'inclusion puis annuellement (inclusion, 1 an et 2 ans). Par ailleurs, le traitement par Ω_3 consistait à prendre deux comprimés par jour. La dose journalière s'élevait à 800 mg d'acide docosahexaenoïque (DHA) et 225 mg d'acide eicosaentaenoïque (EPA).

Les données de MAPT ont été relevées à l'inclusion dans l'essai (T0/ *baseline*), puis à 6 (T6), 12 (T12), 24 (T24) et 36 mois (T36) de suivi. À chaque visite, les scores cognitifs des sujets pour chacune des composantes du Z-score composite ont été recueillis indépendamment (4 scores cognitifs distincts, détaillés en section 1.3.1.1, 1.3.1.2, 1.3.1.3 et 1.3.1.4). Des variables standard telles que le sexe, l'âge, la présence d'une symptomatologie dépressive ont été relevés. Les sujets ont été évalués selon d'autres performances cognitives que celles mesurées par le Z-score composite : le TMT-A et le TMT-B [92, 93], le *Controlled Word Association Test* (COWAT [94])). L'ensemble des variables disponibles pour les analyses est référencé et détaillé en section 1.3.2 pour les critères de jugement secondaires de l'essai MAPT et en section 1.3.3 pour les autres facteurs pris en compte dans notre analyse par rapport à celle de MAPT.

1.2 Objectif de l'essai

L'essai avait pour objectif de déterminer l'effet bénéfique d'un des traitements à l'épreuve par rapport au placebo. Les auteurs de MAPT ont a priori émis l'hypothèse que l'IM et Ω_3 auraient un effet protecteur sur les fonctions cognitives. De plus, la combinaison des deux traitements aurait un effet synergique, ce qui signifie que ces traitements agiraient entre eux pour renforcer leur effet respectif et créer un effet global protecteur sur les fonctions cognitives.

1.3 Critères de jugement

L'ensemble des critères présentés dans la présente section est résumé dans un tableau situé en Annexe A.1.

1.3.1 Critère de jugement principal

Le critère de jugement principal de l'essai résidait en un Z-score composite. Un score élevé indique un niveau cognitif global élevé [95]. La pertinence de l'utilisation d'un tel critère de jugement a déjà été étudiée par l'équipe 1 de l'UMR1027 de l'INSERM [9, 40, 95] et discutée précédemment en section 1.3. Retenons que l'avantage de tels éléments est de regrouper, dans un seul critère, plusieurs domaines d'intérêts. Cela augmente la quantité d'informations relatives à la pathologie étudiée tout en s'affranchissant de la correction des tests multiples [9]. Par contre, de tels critères peuvent diluer les résultats. Lorsque l'une des composante n'est pas fortement impactée par le traitement, l'effet de celui-ci sur le critère composite est amoindri. De la même façon, si les poids de la moyenne ne reflètent pas correctement le phénomène étudié, cela peut conduire à des résultats contradictoires entre deux composantes du critère, alors que les auteurs s'attendent à des résultats cohérents.

Dans ce cas présent, les domaines considérés comme pertinents à évaluer portaient sur la mémoire épisodique, l'orientation, les fonctions exécutives et le langage [9]. Ainsi, le critère de jugement principal de MAPT combinait les résultats de quatre tests cognitifs [9, 13, 40] :

- Le *free and total recall of the Free and Cued Selective Reminding test* [96, 97] (voir section 1.3.1.1),
- Les 10 premiers items du MMSE [98] pour l'orientation (voir section 1.3.1.2),
- Le *Wechsler Adult Intelligence Scale - Revisited* (WAIS-R [99]) présenté en section 1.3.1.3,
- Le *Category Naming Test* (CNT [94], détaillé en section 1.3.1.4).

Ces composantes ont été combinées par une moyenne non pondérée des Z-scores de chacune d'elle (variables centrées et réduites) [9]. En terme d'interprétation, un Z-score de -1 représente un écart-type en dessous de la moyenne initiale. Une diminution de 1 point sur ce score composite indique une diminution moyenne de 1 écart-type pour chacune des quatre composantes [13].

L'effet d'apprentissage potentiel du Z-score composite, par rapport aux connaissances cliniques, a été étudié. Le test de WAIS-R n'est pas connu pour être sensible à l'effet d'apprentissage. Le MMSE est soumis à un effet d'apprentissage de 0,05 points lors des 3 premières administrations du test puis il évolue de nouveau comme attendu, à savoir un déclin au cours du suivi [100]. La catégorie utilisée dans le CNT ne varie pas au cours du suivi de MAPT, donc il est raisonnable de supposer qu'un effet d'apprentissage peut se manifester sur cette composante. Deux listes du test Grober & Buschke ont été utilisées [13], alternant d'une visite à l'autre pour minimiser l'effet d'apprentissage de ce test. En somme, ce critère de jugement, composé des 4 mesures cognitives, était soumis à un effet d'apprentissage sur deux de ses composantes.

Dans toutes les analyses statistiques menées dans la présente étude, le Z-score composite, mesurant le niveau cognitif global des sujets, constitue la variable réponse, le critère d'intérêt principal (*outcome*). Il a été conservé en variable quantitative, sans aucune modification. Il est intéressant de garder à l'esprit le détail des composantes de ce score pour une meilleure interprétation clinique des résultats, mais cela n'aura pas réellement d'incidence sur les analyses. Ces dernières sont détaillées ci-après.

1.3.1.1 Composante 1 : *Free and total recall of the free and cued selective reminding test*

1.3.1.1.1 Description

Le test de Grober et Buschke (aussi appelé « rappel libre/rappel indicé à 16 items » (RL/RI-16)), évalue la mémoire épisodique verbale [96, 97]. C'est un test complet permettant d'observer différentes phases du processus mémoriel (encodage, stockage et récupération de l'information). L'examinateur dispose d'une liste de 16 items appartenant à des catégories différentes. Le score maximum est donc de 16 par épreuve.

Le test s'effectue en plusieurs temps, par lot de 4 mots. D'abord, l'examinateur présente les 4 mots. Il demande au sujet de les identifier en citant les catégories. Une fois les mots bien identifiés, il retire la fiche et demande un rappel immédiat des 4 mots en fournissant l'indice catégoriel. Le sujet dispose alors de 5 essais en cas d'erreur, mais les réponses ne sont pas comptabilisées lorsqu'elles ne sont pas trouvées du premier coup. Le score maximum pour ce rappel immédiat est de 16 points. Un compte à rebours de 100 à 80 (20 secondes maximum) est lancé et constitue une tâche distractive.
Ensuite vient le 1er essai, qui se divise en deux temps. D'abord un rappel libre / 16 points : le sujet dit tous les mots dont il se souvient, dans l'ordre qu'il souhaite. Les réponses exactes sont notées pendant 2 minutes. Pour les items non donnés en rappel libre, le sujet est aidé par l'indice catégoriel (rappel indicé / 16 points). En cas d'erreur, la réponse correcte est donnée mais n'est toujours pas comptabilisée.
Une tâche distractive de 20 secondes est de nouveau effectuée. Le $2^{i'eme}$ essai est identique au 1er (rappel libre + rappel indicé) et suivi d'une tâche distractive de 20 secondes. Le 3ème essai est identique aux 2 autres, si ce n'est que, lorsque le sujet ne se souvient pas, la bonne réponse ne lui est pas rappelée.
Après un délai de 20 minutes, occupé par des tâches distractives non verbales, l'examinateur procède au rappel différé. Le processus reste inchangé par rapport aux étapes précédents.

En pratique, la partie la plus informative du test est considérée, ce qui signifie que le rappel immédiat et le rappel différé ne sont pas comptabilisés. Les scores aux 3 essais libres et indicés sont cumulés : RL1 + RI1 + RL2 + RI2 + RL3 + RI3 = 48 points maximum. Avant de calculer le Z-score de cette variable pour l'introduire dans le Z-score composite de MAPT, un poids différent a été attribué au rappel libre par rapport à l'indicé. Le score utilisé correspondait à : (RL1 + RI1) + RL1 + (RL2 + RI2) + RL2 + (RL3 + RI3) + RL3 = 96 points [101]. Il est important de noter que les scores comptabilisés ne sont pas redondants. Le score total tient compte du rappel libre par définition mais les deux mesures sont cliniquement indépendantes. Sur le plan cognitif, la complémentarité de l'information a conduit les auteurs à les combiner [13]. En effet, un sujet ayant des déficits aux rappels totaux (libres et indicés) peut présenter un trouble d'encodage de l'information verbale en mémoire épisodique [101]. Les difficultés en rappel libre, si elles sont associées à des performances normales en rappels indicés, suggèrent la présence d'un trouble de récupération de l'information verbale. Enfin, un déficit à la tâche de reconnaissance sous-entend un trouble de vérification de l'information récupérée, qu'il soit au niveau de la consolidation ou de l'oubli de celle-ci.

1.3.1.1.2 Effet ou variation attendu

Parmi les déficits cognitifs et mnésiques permettant de mettre en évidence l'apparition d'une démence de type Alzheimer, la mémoire épisodique occupe une position de choix. Ce

système est impliqué dans l'encodage d'une information, son stockage et sa récupération dans un contexte spatio-temporel précis [102]. Un symptôme précoce de l'affectation et central dans le diagnostic de la maladie d'Alzheimer est la déficience de cette mémoire. C'est une condition nécessaire pour affirmer l'existence d'une démence.

Pour des sujets ayant la maladie d'Alzheimer, Giffard et al. [102] ont démontré que leurs performances en rappels libres ou indicés étaient plus faibles que celles de sujets non atteints par la maladie [101]. En particulier, si la partie encodage était davantage affectée que les autres, les faibles performances se ressentaient principalement au niveau du rappel libre [70]. Si la consolidation de l'information dans la phase de stockage était déficiente, alors les performances étaient plus faibles au niveau du rappel différé. Dans tous les cas, les performances du rappel indicé étaient inférieures chez les sujets atteints par la maladie comparé aux non atteints.

En ce qui concerne l'effet de l'âge, il est inversement corrélé au déclin de la mémoire épisodique. Dans une étude de 2008, Kalpouzos et al. [103] ont montré qu'une diminution des ressources cognitive liée à l'avance en âge conduisait au déclin des capacités associées à la mémoire épisodique. Seulement, les sujets âgés présentaient des performances soit similaires, soit meilleures que les sujets jeunes pour des tâches sortant du cadre d'expérimentation (ici envoyer une carte postale ou téléphoner à l'examinateur à une date précise). Ces tâches sont donc accomplies par le sujets à des moments déterminés précisément par l'examinateur et dans le cadre de leur vie quotidienne [104]. L'effet de l'âge n'était pas délétère mais plutôt avantageux dans ce type de tâches [105]. Ces résultats ont été confirmés par Rendell et Thomson [104], qui soulignaient que sur les 6 études dont ils font mention, 4 d'entre elles ont montré des résultats similaires. L'hypothèse associée à ce phénomène n'est pas la préservation ou la conservation des capacités acquises mais plutôt l'utilisation de stratégies compensatoires [103].

Néanmoins, Giffard et al. [102] mettaient en garde quant à l'utilisation isolée de ces performances. À elles seules, elles ne permettent pas d'affirmer ou de récuser l'apparition de la démence car les mesures ne sont ni assez sensibles ni assez spécifiques pour être utilisées de la sorte dans un diagnostic différentiel. En effet, McKhann et al. [106] listent également toute une panoplie de tests médicaux nécessaires à cela, tels que les examens cliniques et neurologiques, les méthodes électrophysiologiques (*computerized tomography*, *regional cerebral blood flow*, *positron emission tomography*, etc) [107].

1.3.1.2 Composante 2 : 10 items de l'orientation du MMSE

1.3.1.2.1 Description

Depuis 1975, il existe un test psychométrique permettant d'évaluer globalement les capacités cognitives des patients : le MMSE [98]. Il est très largement utilisé par les cliniciens et se présente sous la forme d'un questionnaire sur 30 items/points portant sur la réussite de 7 domaines distincts : l'orientation spatio-temporelle (10 points), le rappel immédiat de trois mots (3 points), l'attention et le calcul (5 points), le rappel différé des trois mots (3 points), le langage (8 points) et les praxies constructives (1 points).

Il s'agit d'un outil simple et rapide pour l'évaluation de l'efficience cognitive globale, le déclin cognitif, les troubles du langage, la mémoire épisodique et la mémoire de travail. Ce test n'a pas de durée de temps imposée, il s'effectue au rythme du sujet. Plus le score augmente, meilleures sont les fonctions cognitives [72, 108, 109].

Ce n'est toutefois qu'un test de dépistage d'une détérioration cognitive dont le résultat demande à être analysé en fonction de l'âge du patient, de son niveau culturel, de son niveau d'étude et du contexte clinique de l'étude [110]. Il n'est pas connu pour être assez sensible pour détecter des changements cognitifs très précoces [9]. Il ne permet pas, à lui seul, d'affirmer l'existence d'une démence ou de la récuser [98]. Notons tout de même que ce score au MMSE est connu pour être sensible à la détection d'une baisse de performances plusieurs années avant le diagnostique de démence [73].

Dans le cadre de la création du Z-score composite, seule la partie orientation de ce test (les 10 premiers items) a été considérée [111].

1.3.1.2.2 Effet ou variation attendu

D'un point de vue clinique, une étude de 1997 [112] a montré, à partir des données de la cohorte Personnes Âgées QUID (PAQUID) portant sur des sujets de 65 ans et plus, qu'avoir des performances au MMSE en dessous du premier quartile (score < 24) était associé de façon très forte à un risque de développement de la maladie d'Alzheimer à 3 ans de suivi. Ainsi, il est attendu que ce score diminue avec l'avance en âge [100] et d'autant plus chez les patients atteints de démence [110].

Ce test est connu pour être sensible au niveau d'éducation [110], aux troubles visuels et auditifs, mais surtout à la présence et au degré de démence [113, 114]. Il est également soumis à un effet d'apprentissage de 0,05 points lors des 3 premières administrations du test (dans l'étude, administration à 1 mois, 1 an, 2 ans), puis il évolue de nouveau comme sa tendance globale, à savoir décroissance sur le reste du suivi (dernières visites à 3 ans, 4 ans et 5 ans, toujours dans cette même étude) [100].

1.3.1.3 Composante 3 : *Digit Symbol Substitution Test*

1.3.1.3.1 Description

Le sous-test des codes de l'échelle de Wechsler est un test de raisonnement logique faisant appel aux mécanismes attentionnels [110]. Il évalue la vitesse de traitement de l'information et la mémoire de travail (système de rétention permettant la conservation et l'utilisation d'informations à court terme [102, 115], assimilée plus simplement à la mémoire à court terme). Il est également connu sous le nom de *Wechsler Adult Intelligence Scale - Revisited* (WAIS-R).

Le test est présenté sous la forme d'une table de correspondance entre neuf chiffres et neufs symboles, disponible sous les yeux du sujet durant l'intégralité du test. Sous cette grille se situe une longue table partiellement vide de sensiblement 100 cases. Les chiffres sont indiqués, dans un ordre aléatoire, mais aucun symbole n'est présent. Le sujet doit, dans un temps limité à 90 secondes, remplir la grille vide avec plus grand nombre de symboles possibles, tout en respectant la relation définie par le modèle. Les 7 premiers symboles équivalents aux 7 premiers chiffres de la grille vide sont préalablement remplis, avec l'examinateur, puis le chronomètre est déclenché et le sujet continue le test de façon autonome [99]. Au bout des 90 secondes [99], le chronomètre est arrêté et les symboles correctement reproduits et avec la bonne correspondance sont comptabilisés. Le score total comptabilise le nombre de cases correctement remplies [110].

1.3.1.3.2 Effet ou variation attendu

D'un point de vue clinique, ce score devrait augmenter selon le niveau d'éducation et diminuer lorsque le sujet avance en âge [110]. À l'instar de la mémoire épisodique, l'atteinte de la mémoire de travail est également une condition nécessaire pour affirmer l'existence d'une démence de type Alzheimer [115]. Ce fait est corroboré par plusieurs études récentes traitant aussi bien du vieillissement normal que pathologique. Giffard et al. [102] indiquaient que cette mémoire subit un effet délétère de la maladie d'Alzheimer. Ainsi, l'avance en âge du sujet impacte ses performances au WAIS-R.

Ce test est réalisable grâce à la capacité d'inhibition : le sujet ne retient pas directement de mémoire les symboles associés aux chiffres, il va plutôt faire des navettes entre la table vide et la pleine sans réellement retenir la correspondance. Comme indiqué dans l'article de Bherer et al. [115], cette capacité diminuait chez les patients au cours du temps. Le temps d'exécution du test devient donc plus long et le nombre de cases avec des symboles corrects diminue [63]. Ce ralentissement cognitif général expliquait plus spécifiquement le déclin des performances provenant de tests chronométrés [116, 117].

Le test des codes n'est pas connu pour être sensible à l'effet d'apprentissage, mais il l'est à la fatigue, au stress et à la fatigabilité.

1.3.1.4 Composante 4 : *Category Naming Test* (CNT)

1.3.1.4.1 Description

Le CNT résulte d'une mesure de fluence verbale catégorielle [94]. Ce test d'évocation lexicale sémantique évalue la production spontanée de mots appartenant à une même catégorie (exemple : animaux). Ce test est limité dans le temps à 2 minutes.

La consigne est la suivante : « Vous allez devoir me dire le plus de noms que vous connaissez dans la catégorie que je vais vous donner. Par exemple, pour la catégorie des fleurs, vous pouvez me dire : Rose, Dahlia, Violette ».

Les variables exploitables en pratique sont le nombre de mots générés ou le nombre de mots acceptés. Dans MAPT, les auteurs ont considéré le nombre de mots acceptés (à l'instar des autres scores qui ne comptabilisent que les réponses acceptables ou correctes).

1.3.1.4.2 Effet ou variation attendu

Analogiquement aux catégories précédentes, les performances des sujets dépendent de la présence de la maladie d'Alzheimer. En effet, d'après une étude de Pasquier et al. [118], les performances des sujets étaient plus faibles chez les patients atteints par la maladie que chez les non atteints, et ce pour les fluences catégorielles ou phonologiques (voir COWAT en section 1.3.2.1.2 pour les détails sur cette dernière). Les sujets témoins produisaient plus de mots que les cas. Ces résultats se retrouvent également dans les études de Henry et al. [119], de Morphy et al. [120] et de Nutter et al. [107]. Tous trois suggèrent que les performances de fluences catégorielles étaient plus faibles que les phonologiques, plus sensibles (100 % vs 89 % respectivement) et plus spécifiques (92,5 % vs 85 % resp.) [121].

1.3.2 Critères de jugement secondaires

Les critères de jugement secondaires concernaient les composantes individuelles du Z-score composite et les autres tests cognitifs de la section 1.3.2.1 : le score au MMSE total, le TMT-A et TMT-B (section 1.3.2.1.1) ainsi que le COWAT, présenté en section 1.3.2.1.2. En section 1.3.2.2 sont listés les facteurs cliniques tels que l'échelle d'évaluation de la démence (CDR, détaillé en section 1.3.2.2.1), la symptomatologie dépressive (*Geriatric Depression Scale* - GDS, présentée section 1.3.2.2.2), les échelles analogiques visuelles (EVA présentées en section 1.3.2.2.3), le niveau d'autonomie (ADCS, voir section 1.3.2.2.4), les performances physiques (SPPB, voir section 1.3.2.2.5) ainsi que le critère de Fried (voir section 1.3.2.2.6).

1.3.2.1 Scores cognitifs

1.3.2.1.1 Le score au *Trail Making Test*

Le *Trail Making Test* (TMT) existe en deux versions : A et B. Il quantifie les capacités de flexibilité mentale [92], avec principalement la version B, et la vitesse de traitement de l'information [93] dans le cadre d'une tâche simple (version A) ou d'une tâche plus complexe (version B). Ce test fait également appel au balayage visuel, à l'attention sélective et aux performances motrices [75].

Pour sa version A, le sujet effectue dans un premier temps un exemple avec l'examinateur. Il a une feuille devant lui contenant, dans des cercles, des chiffres allant de 1 à 8. Il doit les relier par ordre croissant. Durant toute l'épreuve, la pointe du crayon doit être maintenue sur la feuille. Si tel n'est pas le cas, le test est interrompu. Dans un second temps, le sujet passe au véritable test (numéros de 1 à 25). L'objectif consiste à l'exécuter le plus rapidement possible. Si le sujet fait une erreur, le chronomètre est arrêté, l'erreur est signalée puis le sujet reprend à l'endroit où il s'est trompé. Cette partie du test vise à évaluer la vitesse psychomotrice, l'attention et le balayage visuel [114].

Pour sa version B, il est nécessaire que le sujet connaisse l'alphabet. Sur la feuille se trouvent alors des numéros et des lettres majuscules (toujours au nombre de 25). Le sujet relie, toujours par ordre croissant, les numéros en alternant avec les lettres : 1- A- 2- B etc. Le crayon reste posé sur la feuille et le test chronométré. Cette version évalue la vitesse psychomotrice, l'attention et le balayage visuel, mais aussi le suivi mental (retenir la dernière lettre ou le numéro précédent) et le décalage de réglage (alternance chiffres et lettres) [114].

D'après les consignes de passation et de cotation de MAPT, tous les sujets ont réussi le test intégralement. Ils ont simplement été aidés ou non lors de l'exécution. En conséquence, il n'est pas intéressant de mesurer le score du sujets mais nous avons considéré le temps d'exécution du test, en secondes, comme variable de TMT [122]. Cliniquement, il est attendu que le temps d'exécution soit plus long lorsque le sujet est en difficulté ou présente un déclin cognitif.

Bherer et al. [115] soulignaient que la capacité d'alternance, utilisée dans des test cognitifs tels que le TMT (surtout version B), s'altérait chez la personne âgée. Cette capacité était davantage diminuée en présence de la maladie d'Alzheimer. En somme, il est attendu cliniquement que le temps d'exécution du test soit plus long avec l'avance en âge. Le temps d'exécution du test sera d'autant plus long chez les sujets âgés et déments.

Dans les analyses, les versions A et B ont été utilisées telles quelles. En revanche, pour

des modèles statistiques, l'information portée par ces deux variables est redondante. D'après les choix observés dans l'essai FINGER [86], une nouvelle variable aurait pu être considérée. Dans leurs analyses, les auteurs ont utilisé la différence entre les deux temps d'exécution, à savoir TMT-B – TMT-A [86]. Si celle-ci est petite, cela veut dire que le sujet n'a pas rencontré de difficulté d'exécution des tests que ce soit dans le cas simple ou complexe. Dans ce cas, le sujet est moins à risque de présenter un déclin de ses fonctions cognitives ou de développer une démence Alzheimer. Seulement, nous avons préféré utiliser une variable avec le moins de modification possible, ce qui implique que la version B a été privilégiée en modélisation.

1.3.2.1.2 Le *ContrOlled Word Association Test*

Le *ContrOlled Word Association Test* (COWAT) mesure la fluence verbale phonologique [94, 123]. Ce test d'évocation lexicale formelle évalue la production spontanée de mots commençant par une lettre définie. Ce test est limité dans le temps à 2 minutes.

La consigne est la suivante : « Vous allez devoir me dire le plus de mots français possible, soit des noms, soit des verbes, soit des adjectifs, etc., commençant par la lettre que je vais vous donner et ceci en deux minutes. Ne dites pas de mots de la même racine, de noms propres et ne vous répétez pas. Par exemple, avec la lettre L, vous pouvez me dire : Lune, Laver, Laborieuse ».

Dans MAPT, deux variables pour ce même test ont été relevées. Nous avions à disposition soit le nombre de mots générés, soit le nombre de mots acceptés. Le nombre de mots acceptés a été considéré comme score au test COWAT [122].

À la manière des autres tests cognitifs et du TMT (section 1.3.2.1.1), avoir une activité cognitive ou un entrainement cognitif diminue le risque de démence [72]. De plus, comme souligné en présentation du CNT (section 1.3.1.4), les performances des sujets étaient plus faibles chez ceux atteints par la maladie que chez les non atteints.

1.3.2.2 Facteurs cliniques

1.3.2.2.1 Le *Clinical Dementia Rating score*

Le *Clinical Dementia Rating score* (CDR) est une échelle d'évaluation du statut de démence [124, 125]. Il comprend plusieurs items tels que la mémoire, l'orientation, le jugement et la résolution des problèmes, les activités hors du domicile, les activités domestiques et loisirs et pour finir les soins personnels. Cette échelle vise à apprécier le degré de sévérité du déficit cognitif et fonctionnel induit par la maladie d'Alzheimer, indépendamment de l'hétérogénéité de la pathologie [76]. Elle est largement utilisée dans les essais cliniques portant sur la maladie d'Alzheimer [126].

La grille d'évaluation présente précisément chaque item. Une note les évalue. Le score total, obtenu par application d'un algorithme faisant intervenir les scores de chaque domaine [127], peut prendre les valeurs de 0, 0,5, 1, 2 ou 3 [128, 129]. Chaque valeur possible de ce score signifie [128, 130, 131] :

— 0 : absence de démence
— 0,5 : démence très légère ou douteuse
— 1 : démence légère
— 2 : démence modérée

— 3 : démence sévère

Pour une question de simplification des résultats et de leur exploitabilité en pratique, cette variable a été considérée comme binaire [128]. Pour la valeur à *baseline*, la dichotomisation n'avait aucune importante car seules 2 classes étaient présentes dans la population de MAPT. En revanche, au cours du suivi, l'absence de démence (score = 0) a été opposée aux autres types de classification de la démence.

1.3.2.2.2 La symptomatologie dépressive

La *Geriatric Depression Scale* (GDS) est un auto-questionnaire sur 15 items (score /15 points) permettant d'évaluer le degré de dépression chez les sujets âgés [132]. Comparativement à sa jumelle en 30 items [133], celle-ci évite les facteurs non spécifiques tels que la fatigue ou les pertes de concentration inhérentes à la personne âgée. Plus courte (entre 5 à 7 minutes pour être complète), elle fournit cependant une information similaire à la version en 30 items [133]. Elle est interprétable selon 3 seuils de valeurs :

— score total < 5 points : absence de dépression ;
— score total entre 5 et 9 points : dépression probable ;
— score total > 9 points : dépression hautement probable.

L'existence d'un lien entre la présence d'une dépression et maladie d'Alzheimer est relativement consensuelle dans la littérature. Concernant la nature du lien, la littérature n'est pas consensuelle [72, 86, 78, 87, 134, 135, 136, 137, 138, 139] :

— soit la dépression est un facteur de comorbidité, ce qui signifie que les sujets atteints développent une dépression au cours de leur maladie,
— soit la dépression annonce l'entrée dans un état de démence.

Cette variable est exploitable en continue, avec un haut score indiquant un état dépressif [109]. Dans notre étude, nous avons dichotomisé ce score, avec une valeur seuil choisie à 5, signifiant une absence de dépression (score < 5) contre les autres modalités possibles [140, 141].

1.3.2.2.3 Les échelles analogiques visuelles

Les échelles analogiques visuelles (EVA) évaluent l'intensité des troubles de la mémoire ressentis par le sujet (EVA-1 [72]) et la gêne ressentie dans les activités de la vie quotidienne (EVA-2 [87, 142]).

Le sujet dispose devant lui d'une échelle en ligne droite mesurant 10 cm dont les extrémités indiquent le sens de gradation (à gauche "Parfaitement" contre "Très mal" à droite pour EVA1 ; à gauche "Pas du tout gêné" versus "Extrêmement gêné" pour EVA2). Il doit tracer un trait vertical sur chaque échelle de façon à évaluer ses gênes. La mesure partant du bord gauche est effectuée en millimètres (mm). D'après l'article de Vellas et al. [87], plus le sujet se plaignait, c'est-à-dire plus la mesure de l'EVA était élevée, plus il était à risque de développer une démence de type Alzheimer ou d'avoir un déclin cognitif.

Pour les analyses statistiques, les deux variables EVA-1 et EVA-2 ont été considérées comme quantitatives.

1.3.2.2.4 Le niveau d'autonomie

Cette variable ADCS correspond aux activités élémentaires de la vie quotidienne par auto-questionnaire [87]. Son nom complet est *Alzheimer Disease Cooperative Study - Activities of Daily Living Inventory - Prevention Instrument*, abrégé en ADCS-ADL-PI. Cette échelle mesure les capacités fonctionnelles chez les sujets âgés dans le cadre des essais de prévention de la maladie d'Alzheimer. Elle est composée de 15 questions dont les cotations sont directement indiquées sur le cahier d'évaluation. Les valeurs du score ont une amplitude de 0 à 45 points. Un score maximal signifie avoir répondu, pour chaque question, par "aussi bien que d'habitude, sans difficulté", soit à un très bon niveau d'autonomie. Cliniquement, il est donc attendu que le score diminue en présence de démence ou de la maladie d'Alzheimer [143, 144]. La perte d'autonomie est précisément un des symptômes permettant de poser le diagnostic de la maladie.

Une autre échelle, apparentée à la première, est utilisée dans la littérature. L'échelle *Instrumental Activities of Daily Living* (IADL) mesure l'aptitude des sujets aux activités complexes de la vie quotidienne, adaptée aux sujets ambulatoires. Elle se décompose en 8 domaines différents : l'aptitude à utiliser le téléphone, à faire les courses, à la préparation des aliments, à l'entretien ménager, à la blanchisserie, aux moyens de transport, à la responsabilité à l'égard de son traitement et l'aptitude à manipuler l'argent. La cotation attribuée lors de l'entretien fournit un score compris entre 4 et 31 points. Tous les domaines n'ayant pas la même "valeur", ces résultats ont été recalculés selon les cotations fournies par Lawton et Brody [91]. Chaque domaine reçoit une évaluation binaire. Dans MAPT, le score correspondant aux IADL peut prendre les valeurs entre 0 et 8 points (somme des résultats) [145]. Un score élevé concorde avec un niveau d'autonomie élevé.

L'IADL n'est disponible qu'à *baseline* et en fin de suivi dans MAPT, aussi nous avons décidé de ne pas la considérer dans les analyses statistiques et de privilégier la variable ADCS [111]. Lorsque nous parlerons d'autonomie dans la suite de ce manuscrit, nous ferons référence à l'ADCS. Généralement, l'ADCS est utilisée comme une variable quantitative [146].

1.3.2.2.5 Le score de performances physiques

Le SPPB est un test de performances physiques. Il est composé de 3 grandes tâches : la vitesse de marche, se lever 5 fois d'une chaise (tâche chronométrée) et des tests d'équilibre de plus en plus difficiles [13]. Chaque mesure se voit attribuer un score catégoriel allant de 0 (incapacité à terminer le test) à 4 (meilleure performance). Le score final, composé de chaque sous-partie, correspond à la somme des différentes performances obtenues. Il donne un aperçu global du statut fonctionnel de la personne âgée [72, 86, 88, 87].

Les valeurs de cette variable quantitative s'étendent de 0 à 12 points. Avoir de faibles performances au SPPB, serait un facteur de risque [86, 88, 87] du déclin des fonctions cognitives. À l'inverse, pratiquer une activité physique régulière aurait un effet protecteur sur le développement de maladies telles qu'Alzheimer [72].

1.3.2.2.6 Le niveau de fragilité global

Le critère de Fried regroupe les critères d'évaluation de la fragilité. Il correspond à un syndrome multidimensionnel provoquant une diminution des réserves d'un sujet (physique,

énergétiques, cognitives, de santé et d'aptitude) [147]. Il a été proposé par Fried pour mesurer 5 grand domaines [87, 148, 149, 150, 151] :

1. la perte de poids involontaire : 4,5 kg ou plus perdu au cours de l'année précédente, ou, au cours du suivi, perte de 5 % du poids de l'année précédente ;
2. l'épuisement subjectif : faiblesse de l'énergie quantifiée à travers 2 questions ;
3. la force de la poignée de main : mesure effectuée à l'aide d'un dynamomètre ;
4. la vitesse de marche : chronomètre du temps de parcours d'une distance de 4 mètres et calcul de la vitesse de marche (évaluation de la lenteur) ;
5. l'activité physique : série de plusieurs questions permettant d'évaluer le niveau d'activité physique.

Chaque critère donne une évaluation finale binaire. La somme des critères conduit à un score de Fried compris entre 0 et 4 points [147]. Cette variable est catégorisée en degré de fragilité [149, 152] :

— Score de 0 : non fragile ;
— Score de 1 ou 2 : pré-fragile ;
— Score de 3 ou plus : fragile.

Dans le cadre de cette étude, nous avons construit une variable binaire en utilisant les définitions d'autres études [109, 125, 151, 153], c'est-à-dire en opposant le profil non fragile (score = 0) aux autres profils possibles.

Par construction du critère, chaque composante du score de Fried est connue pour être liée soit à une déficience cognitive, soit à la maladie d'Alzheimer [153]. Ainsi, plus le score de fragilité augmente, plus les performances du sujet sont faibles et donc plus il est à risque de développer la maladie d'Alzheimer [109, 151, 153].

1.3.3 Critères de l'analyse post-hoc

En dernier lieu, cette section décrit les autres facteurs pris en compte dans le présent travail de thèse. Pour MAPT, ils n'étaient ni critère principal ni critères secondaires, simplement des facteurs d'ajustement des analyses statistiques. Parmi cette liste de facteurs, nous trouvons l'âge (présenté en section 1.3.3.1), le sexe (voir section 1.3.3.2), le niveau d'études (voir section 1.3.3.3), le statut phénotypique de l'APOE4 (voir section 1.3.3.4). Nous trouvons également des facteurs de risques cardiovasculaires tels que l'hypertension (section 1.3.3.5), le cholestérol (section 1.3.3.6) et l'indice de masse corporelle (section 1.3.3.7), ainsi que le score de risque de la démence de type Alzheimer (section 1.3.3.8).

1.3.3.1 L'âge

L'âge est un facteur de risque qui doit être pris en compte dans les analyses statistiques. Il demeure un facteur de risque majeur de déclin cognitif [8, 12, 78, 108, 154, 155]. Lorsque la variable d'intérêt est le Z-score composite, l'âge doit apparaitre en tant que variable explicative du modèle.

1.3.3.2 Le sexe

Le sexe du sujet est un facteur d'ajustement dans de nombreuses études [78, 108, 125]. En effet, dans l'étude de Letenneur et al. [156], le sexe semblait être un facteur de risque

de développer une démence ou la maladie d'Alzheimer. Ce risque allait dans le sens que les femmes étaient plus sensibles que les hommes.

Pour les analyses statistiques, une interaction claire a été identifiée avec les 2 échelles de visualisation analogique (EVA) correspondant aux plaintes mnésiques [87, 122]. Ces 2 échelles sont présentées en détails en section 1.3.2.2.3. Les groupes cognitifs formés dans l'étude par rapport aux plaintes des sujets étaient significativement différents selon le sexe. Ce résultat s'observait aussi bien dans le groupe témoin que dans le groupe de sujets ayant des déficiences cognitives.

1.3.3.3 Le niveau d'études

Le niveau d'étude des sujets reflète des composantes plus difficilement mesurables telles qu'un environnement socio-culturel stimulant, une qualité de vie supérieure etc. Cette variable a été relevée en plusieurs modalités dans l'essai MAPT [109] :

— Aucun ou < Certificat d'études Primaires (CEP) ;
— CEP ;
— Brevet- Certificat d'aptitude professionnel (CAP) - Brevet d'études professionnel (BEP) ;
— Baccalauréat ;
— Post-Baccalauréat.

Le lien existant entre ce facteur et la maladie d'Alzheimer ou la survenue d'une démence est connu. Il indique qu'avoir un haut niveau d'étude serait un facteur protecteur de survenue de la démence [95, 157], en ce sens qu'il en retarderait les manifestations [73, 157]. En effet, les sujets ayant un haut niveau d'éducation ont une réserve cognitive [158] plus importante que les autres [159], ou un niveau cognitif global plus important [95]. Ces affirmations sont confirmées par Huang et al. [160] et Karp et al.[161].

Par contre, lorsque la démence est avérée, avoir un haut niveau d'étude "accélère" la progression de la maladie [73, 95], en ce sens que la date de diagnostic ne correspond pas à celle d'apparition de la démence. Les sujets, quels que soient leurs niveaux d'étude, cherchent à compenser les pertes associés à la maladie d'Alzheimer. La réserve cognitive aide le sujet de haut niveau d'étude à mettre en place davantage de mécanismes compensatoires. Il n'est pas évident de se rendre compte de la présence de la maladie puisque les symptômes sont camouflés : la date de diagnostic sera éloignée et antérieure à la date d'apparition de la démence. À l'inverse, pour un bas niveau d'étude ayant moins de ressources cognitives, la maladie va être diagnostiquée beaucoup plus rapidement : la date de diagnostic est très proche de la date d'apparition de la maladie. L'évolution de la maladie est similaire quel que soit le niveau d'étude, mais la fenêtre d'observation de l'évolution de la pathologie n'est pas identique, donnant l'impression d'une accélération chez les individus à haut niveau d'étude.

Pour faciliter le traitement de l'information et l'exploitabilité des résultats des modèles statistiques, cette variable a été dichotomisée. Une méthode similaire à celle de précédents articles a été utilisée [72, 108, 162] : la variable d'éducation, ou de niveau d'études, a été utilisée en tant que qualitative binaire, dont le seuil de coupure était de 6 années d'études (école maternelle et élémentaire). Autrement dit, dans notre cas, cela revient à considérer un facteur d'ajustement opposant un niveau d'études jusqu'au CEP versus les autres niveaux [73].

1.3.3.4 Le statut phénotypique

Le statut phénotypique des sujets fait uniquement référence au gène APOE. Chez l'humain, les plus courants allèles de l'APOE (ϵ_2, ϵ_3 ou ϵ_4 [163]) sont codés à partir d'un gène présent sur le long bras du chromosome 19 [164]. Ainsi, la protéine peut avoir plusieurs formes (polymorphe) de façon homozygote (2 fois le même allèle) ou hétérozygote (2 allèles différents), à savoir ϵ_2/ϵ_2, ϵ_2/ϵ_3, ϵ_3/ϵ_3, ϵ_2/ϵ_4, ϵ_3/ϵ_4 ou ϵ_4/ϵ_4 [165, 166]. La forme la plus répandue dans la population est la homozygote APOE-ϵ_3 : entre 50 à 90 % de la population la possède [163, 166, 167, 168] tandis que l'allèle APOE-ϵ_2 représente moins de 15 % et APOE-ϵ_4 entre 5 et 35 %.

Bien que moins fréquent dans la population, il a été prouvé que l'allèle ϵ_4 est lié à la maladie d'Alzheimer [165, 169, 170]. Il constitue donc un facteur de risque génétique de la démence, en particulier de la maladie d'Alzheimer. Il doit être pris en compte dans les analyses [72, 78, 108, 109, 154, 162]. En effet, la fréquence des allèles est totalement modifiée chez les sujets atteints de la maladie d'Alzheimer (autour de 40 % contre environ 15 % en population générale) [163, 167, 171]. Ces derniers seraient plus à risque d'avoir la maladie en portant l'allèle ϵ_4 comparé à ceux avant un phénotype ϵ_2/ϵ_3 ou ϵ_3 homozygote [171, 172].

Dans la base de données initiale de MAPT, cette variable est recueillie sous la forme suivante : E2, E2/E3, E2/E4, E3, E3/E4, E4. Pour nos analyses, le phénotype homo- ou hétérozygote E4 nous intéresse particulièrement, mais la distinction entre homo- et hétéro- n'est pas primordiale. Les modalités ont donc été regroupées de la façon indiquée dans le Tableau 3.1 [78, 173] :

Table 3.1 – Construction de la variable Phénotype en qualitative binaire.

Modalités	Correspondance
Sans risque / négatif	E2 + E2/E3 + E3
A risque / positif	E2/E4 + E3/E4 + E4

1.3.3.5 L'hypertension

Cette variable d'hypertension peut être construite à partir d'une valeur seuil de la Pression Artérielle Systolique (PAS) $\geq$ 160 mmHg ou avec un seuil de valeur de la Pression Artérielle Dyastolique $\geq$ 95 mmHg d'après [86, 88, 108].

Sachant que ces 2 variables sont disponibles dans la base de données de MAPT [72, 160], l'hypertension a été définie selon les valeurs de la PAS couchée au seuil de 140 mmHg, soit le stade 1 d'hypertension [86, 174]. Les sujets non hypertendus possédaient donc une PAS $\leq$ 140 mmHg.

1.3.3.6 Le cholestérol

La variable de cholestérolémie était implémentée en binaire dans MAPT [88, 163]. Nous ne l'avons donc pas modifié pour les analyses. D'après la littérature, le cholestérol est lié à de nombreux autres facteurs de risques évoqués précédemment.

Le niveau de cholestérol impacterait directement l'activité enzymatique de l'APP, particulièrement la production de la β-amyloïde. Un haut niveau de cholestérol (hypercholestérolémie) réduirait la production de APP, accélèrerait le dépôt de $A\beta$ et la formation

de plaques extra-cellulaires [175, 176, 177], perturberait l'oxygénation des tissus et la circulation sanguine [176]. Ainsi, de nombreuses études citées dans la revue de la littérature de Shobab et al. [175], ont montré qu'une forte concentration de cholestérol était liée à une augmentation de la sensibilité à la maladie d'Alzheimer et aux maladies cardio-vasculaires [163]. En particulier, l'hypercholestérolémie, définie par un taux de cholestérol supérieur ou égal à 240 mg/dL [178, 179], a été démontrée comme étant un facteur de risque de la maladie d'Alzheimer [167, 176, 179].

1.3.3.7 L'indice de masse corporelle

L'indice de masse corporelle (IMC) est calculé en divisant le poids du sujet (en kilogrammes) par sa taille au carré (en mètres). Plusieurs études ont utilisé l'IMC en tant que variable d'ajustement dans leurs modèles statistiques [86, 78, 88, 108, 154]. Dans leur étude, Xu et al. [180] ont démontré que les sujets en surpoids avaient un risque de démence supérieur aux autres sujets. Ce taux de démence augmentait lorsque l'IMC variait dans le même sens, que la variable soit de type continue (*Odds Ratio* (OR) de 1,08, Intervalle de Confiance à 95 % de IC95%=[1,03 ; 1,14]) ou qualitative (OR = 1,33 pour IMC = 26-30 $kg.m^{-2}$ et OR = 2,57 pour IMC > 30 $kg.m^{-2}$). Ces résultats tenaient compte de facteurs de confusions potentiels connus tels que l'âge, le niveau d'éducation et le sexe. Des résultats similaires sont observables dans les articles précédemment cités.

Par ailleurs, Nawab Qizilbash et al. [181] ont démontré, chez des sujets âgés en moyenne de 55 ans à l'inclusion et ayant un léger surpoids (IMC = 26,5 $kg.m^{-2}$), qu'avoir un IMC bas augmentait le risque de présenter une démence pour un âge plus avancé que dans l'étude de Xu et al. [180]. L'IMC était corrélé négativement à la démence sur le long terme alors qu'il était corrélé positivement sur le court terme. D'autres études ont montré que les sujets avaient un IMC plus faible lorsque la maladie est avérée [182, 183] car ils perdent du poids. Par ailleurs, les sujets de l'étude de Nawab Qizilbash et al. [181], dont l'IMC excédait 40 $kg.m^{-2}$ (obésité morbide), ont présenté un risque de démence plus faible de 29 % par rapport aux personnes de poids normal (18 $\leq$ IMC $\leq$ 25 $kg.m^{-2}$). Ce fait a été confirmé par [160, 184] et doit s'interpréter tout en gardant à l'esprit que les sujets en obésité morbide avaient un risque de décès prématuré plus important. L'obésité réduisait le risque de démence à un âge avancé, mais également les chances d'atteindre cet âge. En effet, l'obésité est liée à risque accru de développer de nombreux autres dysfonctionnements dans l'organisme, principalement de type cardiovasculaires.

L'obésité est une maladie provoquant une inflammation chronique de l'organisme depuis les tissus adipeux. Le pancréas, le foie et le cerveau sont les organes les plus touchés par cette inflammation. Or, dans le cerveau, cette dernière provoque la destruction des réseaux neuronaux (en particulier des épines dentritiques) et donc des capacités cognitives sur le long terme. C'est pourquoi la littérature s'accorde généralement sur le fait que l'obésité est facteur de risque de la démence, bien qu'il reste tout de même des discordances, à l'égal de Nawab Qizilbash et al. [181].

Plusieurs seuils existent [182, 185]. Généralement, le critère de dichotomie s'appuie sur une valeur définissant un seuil de surpoids. De façon à avoir une concordance entre les variables des analyses, nous avons utilisé comme seuil de dichotomie celui du score CAIDE (décrit ci-après en section 1.3.3.8). Le critère utilisé n'est alors pas le surpoids mais l'obésité, définie comme un IMC $\leq$ 30 $kg.m^{-2}$ pour les non obèses [86].

1.3.3.8 Le score *Cardiovascular Risk Factors, Aging and Dementia*

Le score *Cardiovascular Risk Factors, Aging and Dementia* (CAIDE) renvoie à une estimation du risque de survenue d'une démence de type Alzheimer. Elle se base sur les facteurs de risques connus tels que [186, 187] :

— Âge en 3 classes : < 47 ans, 47-53 ans et > 53 ans ;
— Sexe : 0 = femme, 1 = homme ;
— Éducation en 3 classes : 0-6 ans, 7-9 ans et ≥ 10 ans ;
— PAS en binaire ≤ 140 mmHg, > 140 mmHg ;
— IMC en binaire $\leq 30\ kg.m^{-2}$, $> 30\ kg.m^{-2}$;
— Cholestérol en binaire $\leq$ 6,5 mmol/L, $>$ 6,5 mmol/L ;
— Activité physique en binaire actif vs inactif.

Les composantes n'ayant pas le même poids, ce score varie de 0 à 15 [86].

Dans le descriptif de la population, ce score a été utilisé à part entière. Pour les analyses statistiques et modélisations, il est judicieux d'éviter les redondances dans le modèle. Nous avons donc choisi de considérer les composantes de façon indépendantes plutôt que d'inclure ce score dans les modèles.

1.4 Plan d'analyses statistiques

La différence de l'évolution du score entre les groupes d'intervention était évaluée à l'aide de modèles linéaires mixtes (procédure MIXED de SAS®).

Un premier modèle mixte (*mixed model repeated measures*) sans ajustement a été réalisé avec les effets fixes du groupe d'intervention, du temps considéré comme variable continue et en années ainsi que l'interaction entre le groupe d'intervention et le temps. Dans ce modèle les effets aléatoires pris en compte étaient l'intercept et le temps, si le test du maximum de vraisemblance restreinte était significatif pour le temps. La corrélation intra-centre a également été testée. Si le test du maximum de vraisemblance restreint n'était pas significatif, un modèle linéaire mixte à 2 niveaux prenant en compte uniquement la corrélation intra-sujet était envisagé. Différentes formes de trajectoires (quadratique, cubique) étaient testées. La méthode du maximum de vraisemblance (*Maximum Likelihood*) était utilisée pour tester les effets fixes du modèle et la méthode du maximum de vraisemblance restreint (*REstricted Maximum Likelihood*) pour tester les effets aléatoires. En somme, en considérant $Y(t)$ un marqueur définit en tout temps t, la loi a priori de l'effet aléatoire étant généralement supposée gaussienne, le modèle complet peut s'écrire de la façon suivante :

$$\begin{aligned}
Y_{ij} &= (\beta_0 + \alpha_{0i}) + (\beta_1 + \alpha_{1i})Tps_{ij} + (\beta_2 + \alpha_{2i})Tps_{ij} * Tps_{ij} + (\beta_3 + \alpha_{3i})Tps_{ij} * Tps_{ij} * Tps_{ij} \\
&+(\beta_4 + \alpha_{4i})Centre_i + \beta_5 Gpe_i + \beta_6 Gpe_i * Tps_{ij} + \beta_7 Gpe_i * Tps_{ij} * Tps_{ij} \\
&+\beta_8 Gpe_i * Tps_{ij} * Tps_{ij} * Tps_{ij} + e_{ij} \quad ,
\end{aligned}$$

avec Y_{ij} la variable réponse quantitative mesurant l'effet d'intervention sur le niveau global des fonctions cognitives au temps t_{ij}, i le numéro du sujet et j le temps de mesure, β les effets fixes des différentes variables, α le vecteur des effets aléatoires tels que $\alpha \sim N(0, B)$, B la matrice de variance-covariance et e_{ij} les erreurs de mesures supposées indépendantes des effets aléatoires tels que $e_{ij} \sim N(0, \sigma_e^2)$. La variable Tps_{ij} mesure l'effet du temps, Gpe_i l'effet du groupe d'intervention multidomaine (variable binaire) et $Centre_i$ l'effet du centre d'intervention. Enfin, un deuxième modèle mixte était envisagé pour prendre en compte les facteurs de confusion potentiels.

Un calcul du nombre de sujets nécessaires a été estimé. Pour un risque α de 1,25 % tenant compte du nombre de bras de l'étude et une puissance statistique de 80 %, 201 sujets dans chaque bras de l'essai étaient requis, soit un total de 804 sujets [87]. De plus, les auteurs ont considéré un taux d'attrition de 30 % sur les 36 mois de suivi, augmentant le nombre de sujets nécessaires à 1148, soit 287 par groupe.

2 Principaux résultats de l'essai MAPT

À présent que le protocole de l'essai MAPT est présenté, détaillons les principaux résultats obtenus. Dans un premier temps, nous présentons la population analysée lors de l'essai (section 2.1), puis nous résumons les analyses statistiques effectuées et les résultats en découlant (section 2.2). L'ensemble des éléments énoncés par la suite proviennent toujours de l'article de MAPT [13].

2.1 Population analysée dans MAPT

Concernant la population analysée dans l'essai, après observation des 480 sujets premiers sujets d'étude, une nouvelle estimation du taux d'attrition s'est avérée nécessaire car il était plus élevé que prévu. Un échantillon de 1680 sujets a finalement été constitué. Ce processus est décrit en Figure 3.1, de la phase d'éligibilité des sujets à l'échantillon final, avec détail des sujets contenus dans chaque groupe. Sur cette Figure 3.1 apparaissent, en avant-dernière ligne, une population de sujets per protocol. L'analyse de ces sujets a exclu toutes les violations majeures du protocole, en particulier les sujets ne remplissant pas les critères d'inclusion ou d'exclusion, n'ayant pas reçu au moins une dose d'Ω_3 ou n'ayant pas assisté à au moins une session d'IM, n'ayant pas reçu l'intervention qui leur avait été attribuée par la randomisation, ayant changé de groupe de traitement pendant le suivi, ou ayant pris des compléments d'acides gras polyinsaturés non étudiés pendant le suivi. Dans la dernière ligne de la Figure 3.1, la population qualifiée par MAPT de "en intention de traiter modifiée" est présentée. Elle correspond aux 1525 sujets randomisés dans les bras de l'essai dont au moins une valeur du critère de jugement principal est disponible après la visite d'inclusion.

Ainsi, la population de MAPT a été décrite selon les caractéristiques à l'inclusion (voir Figure 3.2). Nous avons repris le tableau de l'article de MAPT [13], sans tenir compte des recodages de variables proposés en sections 1.3.2 et 1.3.3. Ces informations ont indiqué qu'un biais de sélection était possible car le niveau d'étude de ces sujets était particulièrement élevé par rapport à la population générale de même âge.

2.2 Analyses statistiques des données de MAPT

Pour déterminer l'évolution du Z-score composite entre début et fin de suivi, des modèles linéaires mixtes pour données répétées dans le temps ont été implémentés [13]. Seuls les résultats du Z-score avant diagnostic de démence ont été pris en compte, les autres ont été censurés et considérés comme manquants. En conséquence, la gestion des données manquantes s'est faite sous l'hypothèse *Missing At Random* (MAR), ce qui signifie que les données manquantes au cours du suivi étaient considérées comme dépendantes des données

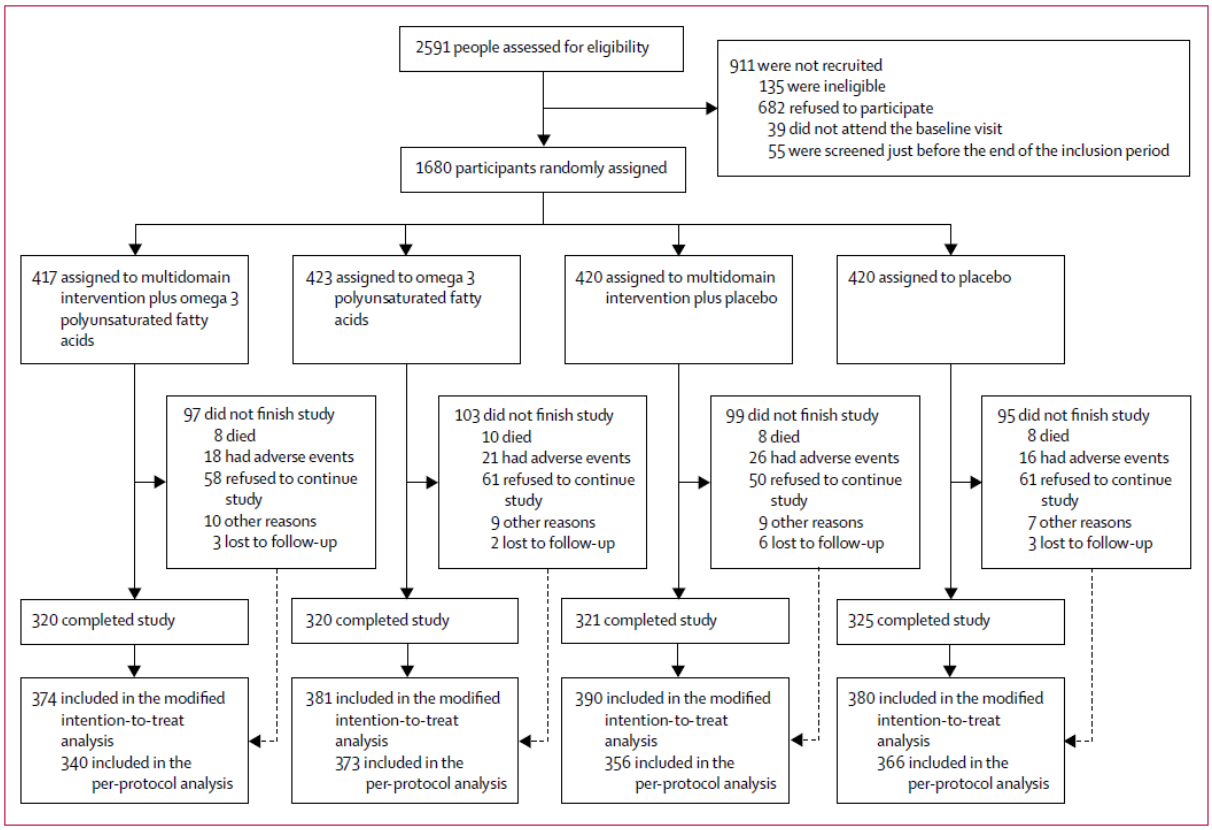

FIGURE 3.1 – Schéma du recrutement des sujets analysés dans l'essai MAPT et détail des populations d'analyses (per protocol et en intention de traiter modifiée). Source : [13].

précédentes. Le temps a été considéré en variable continue et des tests de maximum de vraisemblance ont déterminé la linéarité de la trajectoire moyenne. Pour cela, un effet temps au carré et temps au cube ont été testés dans le modèle.

Un intercept aléatoire sur le centre de recueil a été inséré dans le modèle, tout comme un intercept aléatoire pour l'individu. De plus, une pente aléatoire par individu a été ajoutée au modèle. Ces structures ont été évaluées par le test du maximum de vraisemblance restreint et introduites dans le modèle si les effets se révélaient significatifs. L'effet fixe principal portait sur les groupes d'intervention. Enfin, le dernier choix méthodologique des modèles mixtes a été de considérer une matrice de variance-covariance de type non structurée pour les effets aléatoires.

Les résultats des modèles n'indiquaient aucune différence statistiquement significative entre le groupe ayant bénéficié de l'IM et les autres groupes ($p = 0{,}142$). Le Z-score composite des sujets du groupe Ω_3+IM a en moyenne augmenté de 0,024 points au cours des 36 mois de suivi, tandis que celui du groupe contrôle a augmenté en moyenne de 0,069 points. De même, la différence entre le début et la fin de suivi pour les groupes Ω_3+pas d'IM ou IM+placebo n'était pas significative, comparativement au groupe contrôle. Dans cette population, l'intervention multidomaine, seule ou en combinaison avec l'administration d'Ω_3, n'a pas réduit significativement le déclin des fonctions cognitives au cours de 3 ans de suivi, comparativement au groupe contrôle. Par contre, des analyses en sous-groupe suggéraient que l'IM ralentirait le déclin cognitif (sous groupe selon le score CAIDE et ceux avec une tomographie par émission de positons (*Positons Emission Tomography scan*) amyloïde positif).

Avant de présenter le panorama des méthodes de ré-analyse des données d'un essai de prévention non concluant tel que celui-ci (voir Chapitre 5), nous proposons de revenir préalablement sur la notion de données longitudinales du critère de jugement principal en Chapitre 4. En effet, nous avons indiqué que les mesures répétées avaient pour avantage d'augmenter la précision de la mesure, de décrire l'évolution de paramètres au cours du temps, mais elles avaient pour inconvénient d'être soumises, dans certains cas, à un effet d'apprentissage. Dans le Chapitre 4, ces notions ne sont pas ré-évoquées, mais nous présentons une réflexion sur les différentes façons d'aborder des données longitudinales.

	Multidomain plus polyunsaturated fatty acids (n=374)	Multidomain plus placebo (n=390)	Polyunsaturated fatty acids (n=381)	Placebo (n=380)	Overall (n=1525)
Age, years	75·4 (4·4)	75·0 (4·1)	75·6 (4·7)	75·1 (4·3)	75·3 (4·4)
Female sex	229 (61%)	252 (65%)	245 (64%)	252 (66%)	978 (64%)
Education					
No diploma or primary school certificate	75 (20%)	81 (21%)	96 (26%)	82 (22%)	334 (22%)
Secondary education	145 (39%)	129 (33%)	120 (33%)	117 (31%)	511 (34%)
High school diploma	52 (14%)	56 (15%)	43 (12%)	67 (18%)	218 (15%)
University level	100 (27%)	120 (31%)	108 (29%)	110 (29%)	438 (29%)
APOE ε4	71 (23%)	76 (24%)	64 (21%)	76 (24%)	287 (23%)
Composite score (mean Z score)	−0·04 (0·68)	0·00 (0·71)	0·03 (0·63)	0·02 (0·66)	0·00 (0·67)
Mini Mental State Examination	28·15 (1·57)	28·05 (1·62)	28·14 (1·60)	28·10 (1·53)	28·11 (1·58)
Clinical dementia rating					
0	223 (60%)	224 (57%)	220 (58%)	220 (58%)	887 (58%)
0·5	151 (40%)	166 (43%)	160 (42%)	160 (42%)	637 (42%)
Free and Cued Selective Reminding Test					
Free recall*	27·33 (6·49)	27·41 (6·86)	28·24 (6·55)	27·55 (6·80)	27·63 (6·68)
Total recall*	45·18 (3·55)	45·33 (3·87)	45·62 (3·19)	45·04 (4·37)	45·29 (3·78)
Delayed free recall†	10·51 (2·90)	10·62 (2·81)	10·94 (2·80)	10·76 (2·99)	10·71 (2·88)
Delayed total recall†	15·40 (1·23)	15·43 (1·24)	15·48 (1·13)	15·37 (1·53)	15·42 (1·29)
Trail Making Test					
Part A	47·13 (16·53)	46·38 (17·19)	46·15 (16·15)	46·08 (17·58)	46·43 (16·86)
Part B	121·09 (59·89)	118·64 (57·51)	125·96 (68·78)	120·25 (58·80)	121·45 (61·37)
Verbal fluency					
Category Naming Test	25·45 (7·08)	26·26 (7·53)	26·01 (7·53)	26·29 (7·55)	26·00 (7·43)
Controlled Oral Word Association Test	19·35 (6·28)	19·88 (6·60)	19·56 (6·59)	20·21 (6·56)	19·75 (6·51)
DSST (WAIS—R coding)	37·42 (9·70)	38·57 (10·51)	37·18 (9·74)	38·36 (9·87)	37·89 (9·97)
Memory functioning‡	49·97 (16·44)	49·53 (16·56)	49·40 (17·12)	50·28 (17·35)	49·79 (16·86)
Consequences of everyday life‡	39·18 (22·73)	41·19 (23·38)	38·90 (22·92)	40·22 (23·02)	39·88 (23·01)
Geriatric Depression Scale	3·17 (2·47)	3·20 (2·58)	3·27 (2·67)	3·21 (2·70)	3·21 (2·60)
ADCSADLPI	39·49 (4·85)	39·58 (4·93)	39·80 (4·59)	39·95 (4·82)	39·70 (4·80)
Fried Frailty Criteria—gait speed	1·10 (0·26)	1·09 (0·27)	1·09 (0·26)	1·10 (0·26)	1·09 (0·26)
Short Physical Performance Battery	10·55 (1·72)	10·66 (1·72)	10·57 (1·52)	10·66 (1·55)	10·61 (1·63)
Fried Frailty Criteria					
Involuntary weight loss	15 (4%)	19 (5%)	23 (6%)	12 (3%)	69 (5%)
Exhaustion	60 (16%)	55 (14%)	59 (16%)	59 (16%)	233 (15%)
Weakness (handgrip strength)	81 (22%)	82 (22%)	92 (25%)	78 (21%)	333 (23%)
Slow gait speed	18 (5%)	20 (5%)	13 (3%)	9 (2%)	60 (4%)
Low physical activity	53 (14%)	49 (13%)	57 (15%)	48 (13%)	207 (14%)
Presence of ≥1 criteria	160 (45%)	165 (44%)	180 (49%)	160 (44%)	665 (45%)
Blood pressure (mm Hg)					
Systolic	142·79 (20·28)	141·16 (18·88)	140·42 (19·51)	139·74 (19·93)	141·02 (19·66)
Diastolic	79·83 (11·95)	78·18 (11·28)	78·27 (11·40)	79·00 (11·37)	78·81 (11·50)
Body-mass index	26·19 (4·34)	26·04 (3·92)	26·32 (4·07)	25·99 (3·84)	26·13 (4·04)

Data are mean (SD), or n (%), unless otherwise specified. Percentages were calculated on the basis of the number of participants for whom data were available for each criterion. DSST=Digit Symbol Substitution Test. WAIS—R=Wechsler Adult Intelligence Scale—Revised. ADCSADLPI=Alzheimer's Disease Cooperative Study Activities of Daily Living Prevention Instrument. *48 is the total possible score. †16 is the total possible score. ‡Measured on a visual analogue scale.

Table 1: **Baseline characteristics of the modified intention-to-treat population**

FIGURE 3.2 – Descriptif des caractéristiques relevées à baseline pour les sujets en intention de traiter modifiée de l'essai MAPT (N = 1525 sujets). Source : [13].

Chapitre 4

Le pré-traitement des données longitudinales

Le Chapitre 4 revient sur la nature des données à traiter dans ce travail. Faisant suite à ce qui a déjà été présenté dans les chapitres précédents, la section 1 repend les généralités sur ces données. En particulier, les données longitudinales sont au cœur du présent travail et nous proposons deux façons de les traiter : soit directement, soit à l'aide de paramètres représentatifs de leur évolution (paramètres de synthèse).

Ensuite, la section 2 décrit, au sein de la population de MAPT (voir Chapitre 3), les différentes populations analysées par les méthodes présentées au Chapitre 5 et dont les résultats sont indiqués dans le Chapitre 6. Découlant toutes de la même population initiale, elles correspondent à la restriction de certains individus dont le critère de jugement principal de l'essai est particulier. Pour cela, la population totale, sans aucune donnée manquante au cours du suivi sur le Z-score et celle avec au moins une visite après inclusion sont présentées respectivement en sections 2.1, 2.2 et 2.3.

En dernier lieu et avant d'exposer des outils de ré-analyse des données d'un essai clinique non concluant, des analyses préliminaires ont été menées en section 3. Elles visaient à vérifier l'adéquation entre les hypothèses cliniques sur les comportements attendus et les comportements réels. Deux analyses ont été effectuées : une recherche de sujets ayant une évolution des fonctions cognitives extrême en section 3.1 et une recherche de sujets appartenant à certaines catégories de variation des fonctions cognitives en section 3.2.

1 Généralités

Dans un premier temps, il est important de revenir sur la nature des données à traiter. Comme indiqué dans les deux précédents chapitres, le critère de jugement de MAPT est un Z-score composite longitudinal. En section 1.3.4 du Chapitre 2, nous n'avons indiqué qu'une façon de traiter l'aspect longitudinal : en les identifiant comme des mesures répétées dans le temps. En particulier, ces données permettent d'étudier l'évolution de paramètres au cours du temps, mais dans certains cas comme le notre, cette évolution est soumise à un effet d'apprentissage sur les fonctions cognitives. Dans le présent chapitre, nous ajoutons deux autres façons de traiter les données longitudinales.

Dans un premier temps, les données longitudinales les plus simples sont les données de survie. Le critère observé est le temps passé entre deux points de mesure : l'inclusion et la survenue de l'évènement. Nous mesurons ainsi directement une donnée correspondant à une durée, représentée en Figure 4.1.

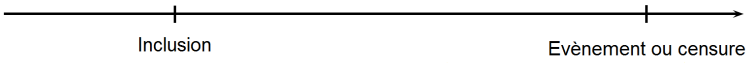

FIGURE 4.1 – Schéma simplifié d'une variable longitudinale simple : la donnée de survie.

Pour les mesures répétées dans le temps, abordées dans les chapitres précédents, c'est justement le fait de répéter la mesure qui intègre l'aspect longitudinal des données. Chaque observation contient ainsi un couple d'information : la durée séparant les mesures d'une part, c'est-à-dire la temporalité de la donnée ; la réponse à chaque instant de mesure d'autre part, soit la valeur de la donnée. Elles sont représentées en Figure 4.2 et correspondent au deuxième cas classiquement rencontré dans des essais cliniques ou de prévention.

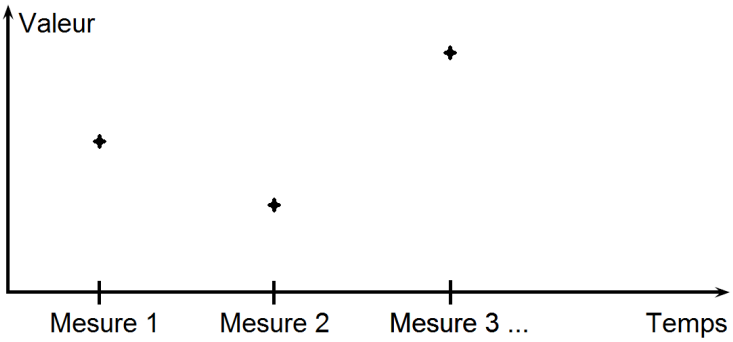

FIGURE 4.2 – Schéma simplifié d'une variable longitudinale classique : la mesure répétée dans le temps.

Nous introduisons ici une troisième et dernière manière d'aborder des données longitudinales : en les considérant comme des données fonctionnelles, représentées en Figure 4.3. Théoriquement, une variable aléatoire est qualifiée de fonctionnelle lorsque ses valeurs appartiennent à un espace de dimension infinie. Les observations sont ainsi mesurées en continu. Ces trajectoires peuvent être issues, par exemple, d'un espace fonctionnel tel que les applications continues de $\mathbb{R}$ dans $\mathbb{R}$. L'observation ici correspond donc à une fonction dépendante du temps.

L'intérêt de distinguer la nature des données longitudinales est de faciliter leur analyse. En effet, à chaque type de données est associé un ensemble de méthodes d'analyses statistiques spécifiques. Pour les données de survie par exemple, les modèles de survie constituent un ensemble de méthodes de choix. Parmi eux, le modèle de Cox à risques proportionnels est le plus utilisé. Dans ce travail, les données de survie ne sont pas exploitées. Pour les mesures répétées, les modèles de prédilection sont les modèles mixtes. Le modèle linéaire mixte a été utilisé dans le Chapitre 3 pour analyser le Z-score composite de MAPT. Nous avons discuté de leur utilité et indiqué qu'ils appartenaient aux méthodes classiques. Pour les données fonctionnelles, il existe également des méthodes d'analyses spécifiques telles que les analyses factorielles ou les méthodes d'interpolation. En pratique cependant, ces données sont nécessairement partiellement observées dans un nombre fini d'instants, ce qui les

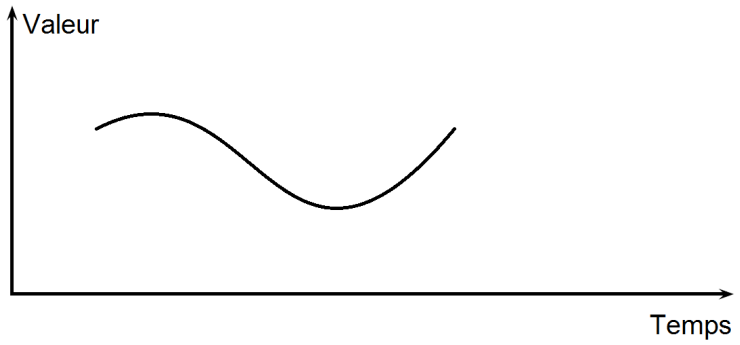

FIGURE 4.3 – Schéma simplifié d'une variable longitudinale théorique : la donnée fonctionnelle ou trajectoire.

rend plus compliquées à manipuler. De fait, la donnée fonctionnelle observée n'est en réalité qu'un format dégradé de la théorie, équivalent à une donnée répétée dans le temps. Plus l'échelle de temps sera petite, plus la donnée observée sera proche de la théorique. Une possibilité d'analyse consiste donc à se ramener à des techniques spécifiques aux mesures répétées, malgré la différence théorique entre les deux types de variables. Dans la cadre de ce travail, nous avons cherché comment adapter les données aux méthodes et vice versa de façon à exploiter d'autres méthodes d'analyse non spécifiques aux données fonctionnelles ou mesures répétées.

En effet, nous avons constaté qu'il est nécessaire de définir des paramètres représentatifs de l'évolution de la donnée fonctionnelle. Ces paramètres résumant l'information contenue dans une trajectoire sont appelés paramètres de synthèse ou indicateurs de variation. Une succession de ces indicateurs transforme une donnée fonctionnelle en mesure répétée dans le temps. Plusieurs paramètres de synthèse existent et les plus courants sont, par exemple, le minimum et le maximum. L'inconvénient de ces deux indicateurs, pour notre étude des fonctions cognitives, est que l'information qu'ils apportent est trop pauvre. Ils sont sensibles aux valeurs extrêmes, ce qui complique leur analyse si ces dernières sont aberrantes. De plus, les variations ne sont pas décrites. De la même façon, la moyenne, la médiane, les quartiles résument une tendance. En supposant que les valeurs extrêmes ne soient pas aberrantes, ils résument l'ensemble de la trajectoire par une valeur. Une nouvelle fois, les variations entre les différentes mesures ne sont pas prises en compte, ni l'amplitude séparant les mesures.

Ces indicateurs peuvent ensuite être combinés entre eux pour conduire à des paramètres de synthèse d'ordre supérieur. Par exemple, la différence entre le début et la fin de suivi conduit à la variation absolue. Elle peut provenir de la différence de valeurs directement ou de la différence d'indicateurs précédents (exemple du min et max). Le problème est qu'elle donne un ordre de grandeur de la variation, mais pas son importance. Une variation absolue de 50 avec une valeur initiale de 10 ou de 100 n'est pas équivalente dans certains domaines comme la mesure des fonctions cognitives. Pour cela, la variation relative peut être utilisée. Elle s'exprime soit par le coefficient multiplicateur (valeur finale / valeur initiale), soit par un taux de variation. Ce dernier est calculé en faisant la différence entre la valeur initiale et la valeur finale, divisée par la valeur initiale, le tout multiplié par 100. La variation relative est pertinente principalement lorsque la trajectoire est monotone. Or, nous sommes dans un cas de trajectoire bruitée, ce qui signifie que cette monotonie peut être masquée par le bruit. En somme, ces paramètres de synthèse ne sont pas suffisants pour résumer l'information qui nous intéresse.

L'inconvénient majeur de ces paramètres de synthèse, outre la perte d'informations engendrée par une utilisation indirecte de la donnée, réside dans leur calcul sur un intervalle donné, sans tenir compte de la taille de cet intervalle. Or, pour la mesure des fonctions cognitives, le temps séparant chaque mesure est primordial. Pour cela, le taux d'accroissement nous a paru pertinent. Cette quantité détermine la nature de la variation d'une fonction en approchant sa dérivée. Son signe indique le sens de variation : positif (respectivement négatif) signifie augmentation (resp. diminution) et nul signifie constant. Il serait donc plus précis pour résumer le comportement évolutif des fonctions cognitives, même si ces dernières sont bruitées. Son calcul découle de la différence entre les valeurs, divisée par le temps écoulé entre deux mesures consécutives :

$$Taux_{accroissement} = \frac{Valeur_{T_i} - Valeur_{T_{i-1}}}{T_i - T_{i-1}} \quad .$$

Plus la taille de l'intervalle considéré est petite, plus la mesure s'approche de la valeur réelle de la dérivée. Néanmoins, plus l'intervalle est petit, plus l'accroissement temps par temps est équivalent à la donnée fonctionnelle même. Les méthodes en sections 1.1.2 et 1.1.3 du Chapitre 5 utilisent le taux d'accroissement entre chaque visite comme critère de jugement principal. Les principaux avantages de ce paramètre de synthèse sont de tenir compte de la dépendance des valeurs au cours du temps (toujours une valeur commune entre deux intervalles) et de la chronologie des mesures.

Par ailleurs, le processus de transformation d'une donnée fonctionnelle vers une succession de valeurs discrètes, peut être abordé comme une succession d'états. En effet, les valeurs fonctionnelles peuvent être discrétisées en intervalles de temps ou de valeur. Chaque intervalle de valeur peut être interprété comme une catégorie de variation. Les changements de catégories de valeurs sur une période de temps donnée sont associés à un changement d'état. L'analyse de la section 3.2 en est un parfait exemple. Dans ce cas, trois intervalles de valeurs ont été définis : les valeurs hautes pour une augmentation du niveau global des fonctions cognitives, les basses pour un déclin et les autres pour une stabilité. Les états sont donc augmentation, déclin ou stabilité. Cette catégorisation découle des variations résumées par le taux d'accroissement. Ainsi, chaque état est associé à un type de variation et à une durée. Au final, la donnée fonctionnelle est équivalente à une succession d'états continus. Ce paramètre de synthèse est utilisé comme critère de jugement dans les sections 2.2.1 et 2.2.2 du Chapitre 5. En particulier, l'impact de covariables sur les changements d'états est analysé. Les états sont, dans le présent document, nommés paramètres de synthèse de façon à les distinguer de l'utilisation directe la donnée fonctionnelle. Ce qualificatif est, en réalité, un abus de langage dans le sens que, contrairement aux taux d'accroissement, résumant l'information de la donnée fonctionnelle sur un intervalle de temps donné, les états ne synthétisent rien et correspondent à une discrétisation de la donnée à un instant t.

2 Populations sélectionnées pour les analyses post-hoc

Dans la présente étude et comme souligné ci-dessus, nous avons analysé rétrospectivement les données réelles de l'essai MAPT présenté dans le Chapitre 3. Les participants inclus étaient les mêmes que ceux de MAPT, excepté un sujet. Ce dernier a retiré son consentement pour l'utilisation de ses données. Les critères d'éligibilité ont été conservés. Notre échantillon d'étude contenait ainsi 1679 sujets.

Le critère de jugement principal a été conservé. Le Z-score composite n'a pas été recalculé et correspond à celui fourni directement dans la base de l'essai. Nous avons également fait la requête de la base complète, ce qui signifie que, pour les sujets diagnostiqués déments au cours du suivi, leurs scores cognitifs après diagnostic n'ont pas été censurés. La démence est ainsi considérée comme une variable purement explicative et non comme un critère de censure. Ce choix était a priori préférable pour répondre à notre objectif de ré-analyse d'un essai de prévention non concluant, indépendamment de toute considération clinique de démence. L'information est plus complète. Les censures n'entraînant pas de biais, cette requête ne devrait pas impacter les résultats attendus de nos analyses. Par ailleurs, les données sont disponibles pour toute personne soumettant une requête détaillée au groupe *Data Sharing Alzheimer Group* (DSA – info.u1027-dsa@inserm.fr). Les données reçues ont été anonymisées.

Trois populations vont ainsi être exposées dans les sections ci-après. La première reprend l'ensemble de la population d'analyse, donc les 1679 sujets (section 2.1). Ensuite, la population la plus restrictive est présentée en section 2.2. Elle ne contient que les 1143 sujets ayant un suivi complet, soit aucune donnée manquante sur le critère de jugement principal sur les 36 mois de l'essai. Enfin, les 1525 sujets possédant au moins une valeur du critère de jugement principal après l'inclusion sont décrits en section 2.3. Dans tous les cas, les variables explicatives sont utilisées telles que présentées dans le Chapitre 3 en sections 1.3.1, 1.3.2 et 1.3.3, avec les recodages indiqués.

2.1 Population $\mathcal{P}_{tot}$

Dans un premier temps, l'ensemble de la population de MAPT a été décrite dans le Tableau 4.1 sur les caractéristiques relevées à l'inclusion. Les variables quantitatives sont présentées avec leur moyenne et écart-type, et les variables catégorielles avec les effectifs et pourcentages. Pour rappel, l'Annexe A.1 résume les différentes variables avec leur notation et leur signification. Cette population contient 1679 sujets dont les caractéristiques majoritaires sont un haut niveau d'études, de sexe féminin, non obèse, n'ayant pas de symptomatologie dépressive, non fragile, non porteurs de l'allèle ϵ_4 de l'APOE et sans hypercholestérolémie. De plus, le Z-score moyen à l'inclusion était nul, la moyenne d'âge était de 75,3 ans et le score d'autonomie était en moyenne de 40 sur 45 points.

2.2 Population $\mathcal{P}_{comp}$

La deuxième population analysée correspond aux sujets ayant un suivi complet. Tous les individus ayant des données manquantes sur le Z-score composite au cours de l'essai ont ainsi été supprimés. Le Tableau 4.2 montre l'impact de cette restriction de population sur les caractéristiques des sujets relevées à l'inclusion. Les 1143 sujets restants avaient pour caractéristiques majoritaires les mêmes que celles de la population $\mathcal{P}_{tot}$ de la section 2.1. En revanche, ces sujets avaient un Z-score moyen à l'inclusion plus élevé, valant 0,12, la moyenne d'âge diminuait à 74,9 ans et le score d'autonomie restait en moyenne de 40 sur 45 points.

TABLE 4.1 – Descriptif des caractéristiques à l'inclusion des 1679 sujets inclus dans l'essai MAPT.

Variables	Modalités	Moyenne Effectif	Écart-type Pourcentage	Manquant
Z-score		0,00	0,67	0
Age		75,34	4,42	0
TMT-A		46,53	17,13	3
TMT-B		122,6	63,44	36
COWAT		19,72	6,51	2
SPPB		10,58	1,65	32
EVA-1		49,95	16,98	0
EVA-2		39,86	23,06	0
Autonomie		39,68	4,80	14
CAIDE		7,61	2,02	58
Bras essai				0
	1 : Ω_3+IM	417	24,9	
	2 : Ω_3	422	25,1	
	3 : IM	420	25,0	
	4 : Placebo	420	25,0	
Etude				36
	1 : Haut	1272	77,4	
	0 : Bas	371	22,6	
Sexe				0
	1 : Hommes	592	35,3	
	0 : Femmes	1087	64,7	
CDR				1
	1 : Oui	706	42,1	
	0 : Non	972	57,9	
IMC				7
	1 : Oui	267	16,0	
	0 : Non	1405	84,0	
GDS				9
	1 : Oui	298	17,8	
	0 : Non	1372	82,2	
Fried				0
	1 : Oui	742	44,2	
	0 : Non	937	55,8	
Hypertension				12
	1 : Oui	703	42,1	
	0 : Non	964	57,9	
Phénotype				379
	1 : APOE4	299	33,0	
	0 : pas APOE4	1001	77,0	
Cholesterol				32
	1 : Oui	736	44,7	
	0 : Non	911	55,3	

Table 4.2 – Descriptif des caractéristiques à l'inclusion des 1143 sujets ayant un suivi complet sur les 36 mois de l'essai MAPT.

Variables	Modalités	Moyenne Effectif	Écart-type Pourcentage	Manquant
Z-score		0,12	0,59	0
Age		74,85	4,11	0
TMT-A		45,14	15,49	2
TMT-B		115,00	50,87	18
COWAT		20,12	6,28	1
SPPB		10,73	1,52	24
EVA-1		49,17	16,50	0
EVA-2		39,04	23,26	0
Autonomie		40,12	4,45	8
CAIDE		7,58	2,03	24
Bras essai				0
	1 : Ω_3+IM	286	25,0	
	2 : Ω_3	277	24,2	
	3 : IM	285	24,9	
	4 : Placebo	295	25,9	
Etude				14
	1 : Haut	884	78,3	
	0 : Bas	245	21,7	
Sexe				0
	1 : Hommes	402	35,2	
	0 : Femmes	741	64,8	
CDR				1
	1 : Oui	459	40,2	
	0 : Non	683	59,8	
IMC				2
	1 : Oui	183	16,0	
	0 : Non	958	84,0	
GDS				6
	1 : Oui	160	14,1	
	0 : Non	977	85,9	
Fried				0
	1 : Oui	474	41,5	
	0 : Non	669	58,5	
Hypertension				3
	1 : Oui	475	41,7	
	0 : Non	665	58,3	
Phénotype				114
	1 : APOE4	234	22,7	
	0 : pas APOE4	795	77,3	
Cholesterol				13
	1 : Oui	499	44,2	
	0 : Non	631	55,8	

2.3 Population $\mathcal{P}_{mapt}$

D'après les populations présentées en section 2.1 du Chapitre 3, nous avons constaté que certains sujets de l'essai MAPT avaient simplement une visite à *baseline*. Ces sujets n'ayant pas d'évolution de leurs fonctions cognitives par définition, ils ont été exclus de la plupart des analyses du Chapitre 6. Cette restriction a conduit aux 1525 sujets présentés et analysés dans MAPT. Les variables étant recodées par rapport au descriptif de MAPT et un sujet ayant retiré son consentement, il est important d'actualiser le descriptif de la population $\mathcal{P}_{mapt}$. Le Tableau 4.3 reprend les caractéristiques individuelles relevées à l'inclusion des sujets analysés. Pour les variables qualitatives, la même tendance que pour les populations $\mathcal{P}_{tot}$ (section 2.1) et $\mathcal{P}_{comp}$ (section 2.2) est observée. En revanche, le Z-score composite moyen valait 0,02, l'âge moyen des sujets était de 75,3 ans et le score d'autonomie moyen de 40 sur 45 points.

3 Analyses exploratoires préliminaires : Exemple des données de MAPT

La première étape de toute analyse statistique comporte un descriptif de la population d'étude, à l'instar de celui présenté dans la section 2 précédente. Dans le cas présent, nous aimerions décrire des profils d'évolution des fonctions cognitives. Les caractéristiques ont été présentées dans MAPT [13] en section 2.1 du Chapitre 3 et en section 2 ci-dessus. Les cliniciens menant un essai de prévention s'attendent à observer des comportements spécifiques en fonction des caractéristiques de la population étudiée. Deux analyses préliminaires ont été conduites afin d'étudier l'adéquation entre les hypothèses cliniques et les comportements observés dans notre étude.

La section 3.1 étudie deux profils extrêmes tandis que la section 3.2 caractérise plus généralement la tendance évolutive de la population. La particularité de ces deux analyses est qu'elles ne seront pas reprises dans le Chapitre 6 des résultats. Ces méthodes d'analyses statistiques descriptives s'appuyaient sur les connaissances cliniques a priori d'évolution des fonctions cognitives. Les informations obtenues sont directement fournies dans les sections correspondantes.

3.1 Identification et étude des profils extrêmes d'évolution des fonctions cognitives

3.1.1 Principe

Cette première exploration des données avait pour objectif d'étudier et de comptabiliser le nombre de sujets répondant à ce que nous avons qualifié de profils extrêmes. Ils ont été déterminés par rapport à l'évolution des valeurs du critère de jugement principal de l'essai MAPT. Dans notre cas, celui-ci était mesuré par une variable quantitative centrée et réduite (Z-score) correspondent au niveau cognitif global. De façon à tenir compte de l'évolution du niveau global et du temps écoulé entre chaque mesure, un indicateur de variation a été considéré : le taux d'accroissement décrit en section 1.

TABLE 4.3 – Descriptif des caractéristiques à l'inclusion des 1525 sujets ayant au moins une visite après inclusion dans l'essai MAPT.

Variables	Modalités	Moyenne Effectif	Écart-type Pourcentage	Manquant
Z-score		0,02	0,66	0
Age		75,25	4,38	0
TMT-A		46,43	16,86	2
TMT-B		121,50	61,37	30
COWAT		19,75	6,51	1
SPPB		10,61	1,63	30
EVA-1		49,79	16,86	0
EVA-2		38,76	23,01	0
Autonomie		39,70	4,80	14
CAIDE		7,61	2,02	43
Bras essai				0
	1 : Ω_3+IM	374	24,5	
	2 : Ω_3	381	25,0	
	3 : IM	390	25,6	
	4 : Placebo	380	24,9	
Etude				24
	1 : Haut	1167	77,7	
	0 : Bas	334	22,3	
Sexe				0
	1 : Hommes	547	35,3	
	0 : Femmes	978	64,1	
CDR				1
	1 : Oui	637	41,8	
	0 : Non	887	58,2	
IMC				4
	1 : Oui	241	15,8	
	0 : Non	1280	84,2	
GDS				7
	1 : Oui	261	17,2	
	0 : Non	1257	82,8	
Fried				0
	1 : Oui	655	43,6	
	0 : Non	860	56,4	
Hypertension				9
	1 : Oui	641	42,3	
	0 : Non	875	57,7	
Phénotype				284
	1 : APOE4	287	23,1	
	0 : pas APOE4	954	76,9	
Cholesterol				21
	1 : Oui	667	44,3	
	0 : Non	837	55,7	

Les deux profils extrêmes étaient les suivants : une amélioration stricte ou un déclin strict tout au long du suivi. Concrètement, cela signifie que le premier groupe contenait tous les sujets ayant des taux d'accroissement strictement positifs tout au long du suivi, au prorata de leur temps de suivi. Par opposition, le second groupe était constitué de ceux ayant des taux strictement négatifs tout au long du suivi.

D'après la composition de la population présentée en Chapitre 3, les sujets étaient à risque de montrer un déclin cognitif et plus particulièrement de développer la maladie d'Alzheimer. Il est cliniquement attendu que les sujets déclinant tout au long du suivi soient nombreux. En revanche, bien que l'intervention ait été conçue pour améliorer les fonctions cognitives, toujours par rapport aux caractéristiques des sujets inclus dans MAPT, le second groupe ne devrait pas contenir beaucoup d'individus.

3.1.2 Comportements identifiés

L'objectif de cette analyse étant avant tout exploratoire, nous avons étudié différentes périodes d'observation. Pour chacune d'elle, la population d'étude a été restreinte aux sujets ayant un suivi complet sur le critère de jugement principal. Par exemple, prenons un sujet venu à *baseline*, 6 mois de suivi et 24 mois de suivi. Il avait ainsi une valeur manquante à 12 mois de suivi. Il ne pouvait avoir un taux d'accroissement que sur la période entre T0-T6 mois. La valeur à 24 mois de suivi n'était pas prise en compte dans l'analyse. Le Tableau 4.4 ci-dessous résume les informations obtenues.

Table 4.4 – Description des profils extrêmes rencontrés dans l'essai MAPT au cours du suivi.

Période(1)	Sujets(2)	Augmentation(3)		Déclin(4)	
		Effectif	Pourcentage	Effectif	Pourcentage
T0-T6	1502	774	51,5	712	47,4
T0-T12	1368	333	24,3	139	10,2
T0-T24	1227	68	5,5	37	3,0
T0-T36	1143	9	0,8	10	0,9

(1) : Période d'observation définie en fonction des temps de mesures et pour laquelle les sujets ont un suivi complet : T0 = *baseline*, T6 = à 6 mois de suivi, T12 = 12 mois, T24 = 24 mois et T36 = fin de suivi ;
(2) : Nombre de sujets n'ayant aucune donnée manquante sur le critère de jugement pendant la période considérée ;
(3) : L'amélioration est définie comme une succession de taux d'accroissements strictement positifs tout au long du suivi ;
(4) : Le déclin est défini comme une succession de taux d'accroissements strictement négatifs tout au long du suivi.

Le Tableau 4.4 met en évidence plusieurs points intéressants. Pour commencer, des sujets répondaient aux deux profils extrêmes sur les différentes périodes d'observation. Plus le temps de suivi augmentait, plus leur proportion diminuait par rapport aux sujets encore présents dans l'étude. Au-delà de 12 mois de suivi, la proportion des sujets appartenant à ces profils extrêmes était pratiquement négligeable.

Sur les 6 premiers mois de l'essai, la population était pratiquement scindée en deux. Nous pouvons imaginer que les sujets augmentant leur niveau cognitif en début de suivi vont ensuite maintenir ou stabiliser ce dernier. Cela pourrait suggérer un effet bénéfique de l'intervention multidomaine. Cependant, la répartition des sujets ayant un profil extrême n'était pas statistiquement différente, aussi bien selon les quatre bras de l'essai (test du χ^2, p = 0,38) que selon le groupe d'IM (test du χ^2, p = 0,21). En revanche, à 6 mois de suivi, les sujets déclinants montraient une déficience en flexibilité mentale et en vitesse de

traitement de l'information, mesurées aussi bien par le TMT-A (p < 0,001) que le TMT-B (p < 0,001). Ces sujets étaient également moins autonomes (p < 0,001), exprimaient une plainte mémorielle plus importante (EVA-1, p < 0,001) tout comme leur gêne dans les activités quotidiennes (EVA-2, p = 0,035). Enfin, ces sujets ont perdu en moyenne un demi-point au MMSE (p < 0,001).

Concernant le reste du suivi, les principales caractéristiques statistiquement différentes sont recensées dans le Tableau 4.5. Il expose les effectifs, moyennes et écarts-types des différentes variables, par période d'observation et par profil extrême, ainsi que les p-values des tests associés. Ces derniers concernent les caractéristiques en fin de période d'observation, exceptés pour l'âge, les groupes de traitement et le groupe d'IM. Ces trois variables ont été relevées à l'inclusion. Notons $p_{periode}$ la p-value du test du χ^2 sur la période d'observation considérée, par exemple p_{T0-T36} pour l'intégralité du suivi. Aucune différence entre les groupes de traitement n'était identifiée ($p_{T0-T12} = 0{,}98$; $p_{T0-T24} = 0{,}85$; $p_{T0-T36} = 0{,}54$) au cours du suivi, tout comme pour le groupe d'intervention ($p_{T0-T12} = 0{,}84$; $p_{T0-T24} = 0{,}84$; $p_{T0-T36} = 0{,}65$).

Sur la période T0-T12 mois, par rapport à la période T0-T6 mois de suivi, l'ensemble des caractéristiques du Tableau 4.5 montraient une tendance analogue. En revanche, venaient s'ajouter des différences significatives sur le COWAT (ANOVA, p < 0,001) et le test des performances physiques (ANOVA, p < 0,001). Dans le premier cas, le groupe déclinant possédait une moyenne de 19 mots acceptés (écart-type de 7 mots), tandis que le groupe d'amélioration avait un score moyen de 21 mots (écart-type de 7 mots). Aucune donnée ne manquait sur cette variable. Dans le second cas, les performances physiques moyennes valaient respectivement de 10,7 points (± 1,7) pour 325 sujets s'améliorant contre 10,0 points (± 2,2) pour les 136 autres sujets. Pour la période T0-T24 mois, toujours par rapport à T0-T6, l'âge devenait significatif, dans le sens que les sujets déclinant tout au long du suivi étaient plus âgés. Le TMT-A n'était plus significatif, mais les autres caractéristiques le restaient. Finalement, sur l'intégralité du suivi, il s'avérait que les sujets déclinant différaient des autres sur l'âge et les caractéristiques relevées précédemment. En revanche, par rapport au début de suivi, la gêne ressentie dans les activités quotidiennes, l'autonomie et le TMT-A ne constituaient plus des facteurs différenciant les deux populations.

Ce qu'il faut retenir de cette classification des profils extrêmes est que, malgré une légère perte d'information due à la sélection des sujets et à l'utilisation d'un paramètre de synthèse de la donnée fonctionnelle, la population ne se comportait pas comme nous l'avions supposé a priori. Les profils extrêmes étaient d'autant moins présents en proportion que le suivi était long. Or, dans le cadre d'un déclin cognitif, il est attendu qu'ils soient d'autant plus présents parmi ceux participant encore à l'étude. Contrairement à ce que nous pensions, ces sujets n'étaient pas suffisamment nombreux pour générer de l'hétérogénéité dans les données et masquer une information importante qui aurait pu être détectée par les modèles mixtes. De plus, nous pensions avoir une majorité de sujets exprimant un déclin des fonctions cognitives au cours du suivi. Leur proportion était, en réalité, très faible.

3.2 Définition et étude de catégories de variations des fonctions cognitives

3.2.1 Principe

Pour cette seconde analyse préliminaire, nous avons mesuré des catégories de variations du critère de jugement principal. À l'instar de l'analyse en section 3.1, la population est scindée selon 3 types de variations : augmentation, déclin et stabilité. L'objectif consistait à créer une succession d'indicateurs de variation, non plus avec une valeur quantitative mais qualitative. Cette dernière, nommée "variation", se construit selon les valeurs réelles du critère de jugement principal. Ensuite, des profils d'évolution sont constitués selon l'ensemble des modalités recueillies pour un sujet donné. Puisque nous avons supposé que la population était hétérogène, procéder ainsi aiderait à maîtriser ce point tout en conservant le caractère longitudinal. En effet, la catégorie de variation est déterminée selon un intervalle autour d'une valeur, ce qui réduit le bruit des données.

La Figure 4.4 résume graphiquement la présente stratégie. Elle repose sur la notion d'intervalle autour d'une valeur. À chaque date de visite t, l'écart-type du critère de jugement a été calculé pour l'ensemble des sujets présents à cette date, notons le s_t. Nous avons construit un intervalle

$$I_t = [Z_t - bs_{t+1}; Z_t + as_{t+1}] \quad .$$

Si un individu possédait une valeur au temps $t+1$ comprise dans l'intervalle au temps t, c'est-à-dire si $Z_{t+1} \in I_t$, alors la modalité de la variable "variation" valait "stabilité". Si $Z_{t+1} < Z_t - bs_{t+1}$, la modalité prenait la valeur "déclin". Ce cas est représenté en Figure 4.4. De même, si $Z_{t+1} > Z_t + as_{t+1}$, la modalité devenait "augmentation". Autrement dit, les modalités de la variable "variation" ont été déterminées, en chaque temps, selon la valeur au temps précédent plus ou moins l'écart type au temps actuel. Notons que si cette dernière est manquante, nous considérons l'intervalle au temps $t-2$ et ainsi jusqu'à *baseline* si nécessaire.

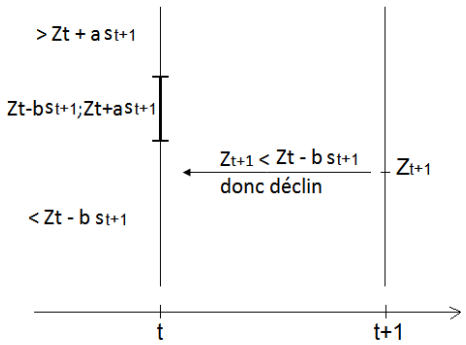

FIGURE 4.4 – Exemple, pour un sujet, de la définition de la catégorie de variations du critère de jugement principal quantitatif de l'essai de prévention MAPT. La valeur en t+1 est en-deçà de la borne inférieur de l'intervalle $I_t = [Z_t - bs_{t+1}; Z_t + as_{t+1}]$, impliquant que l'individu est dans la catégorie déclin à t+1.

Une fois ceci fait, un profil de variation a été construit. Nous avons identifié toutes les combinaisons des différentes modalités au cours du suivi. Chacune d'elle correspond à un profil d'évolution des fonctions cognitives. L'objectif principal était de déterminer la répartition des sujets dans les différents profils.

3.2.2 Catégories de variation obtenues

L'analyse a concerné les sujets ayant au moins une visite après inclusion, soit 1525 sujets (voir section 2.3). Pour commencer, nous avons choisi $a = b = 1$ dans le calcul de I_t. Tous les sujets ayant eu, à un moment donné, une augmentation au cours du suivi ont été fusionnés en un seul groupe. Ils étaient au nombre de 234, soit 15,3 % des sujets analysés. Ils correspondaient à un profil difficilement compréhensible a priori d'un point de vue clinique.

Ensuite, nous avons identifié 840 sujets, soit 55,1 % de la population analysée, avec un profil d'évolution des fonctions cognitives de type "stabilité" tout le suivi. Le Z-score composite de ces sujets était donc systématiquement dans l'intervalle I_t au cours du suivi. À ces sujets s'ajoutaient les profils de type "stabilité puis sortie d'étude". Cela signifiait qu'ils avaient des valeurs du Z-score composite dans l'intervalle I_t puis une succession de valeurs manquantes. La sortie d'étude pouvait avoir lieu à plus ou moins long terme au cours du suivi. Ils représentaient 274 sujets, soit 18,0 % de l'effectif. En somme, ces chiffres nous indiquaient que la population étudiée était constituée à 73,1 % de sujets ayant un profil de stabilité cognitive. Détecter ce type de variation, d'autant plus en majorité au sein de la population étudiée, était parfaitement en accord avec les informations recueillies en section 3.1.

Enfin, tous les autres sujets avaient un profil de type Stabilité-Déclin ou une alternance entre les deux. Ils correspondaient au total à 11,6 % des sujets. Ils incluaient également des données manquantes intermittentes ou dues à la sortie d'étude. Nous pouvions les distinguer en plusieurs groupes :

- Stabilité-Déclin : 72 sujets, soit 4,7 % ;
- Stabilité-Déclin-Stabilité : 97 sujets, soit 6,4 % ;
- Stabilité-Déclin-Stabilité-Déclin : 7 sujets, soit 0,5 % ;
- Stabilité-Déclin-Stabilité-Déclin-Stabilité : 1 sujet, soit 0,1 %.

Nous avons comparé ces profils sur différentes caractéristiques cliniques. Il en résultait que les sujets déclinant étaient plus âgés (ANOVA, $p < 0{,}001$), moins autonomes (ANOVA, $p < 0{,}001$) et plus nombreux à développer une démence au cours du suivi (test du χ^2, $p < 0{,}001$). En revanche, aucune différence entre ces trois groupes n'était notable en terme de répartition dans les 4 bras de l'essai (test du χ^2, $p = 0{,}571$) ou selon l'IM (test du χ^2, $p = 0{,}266$).

Dans un second temps, nous avons fait varier les valeurs de a et b entre 0 et 2 fois l'écart-type de la population à *baseline*, avec un pas de 0,1. Quelles que soient les valeurs prises par ces deux paramètres, le nombre de sujets ayant une augmentation au cours du suivi était constant. Ensuite, la répartition des profils pouvait être résumée ainsi :

1. $a > s_0$:
   - $b \in [0{,}0\,;\,0{,}4]$: Déclin majoritaire ;
   - $b \in [0{,}5\,;\,1{,}3]$: Stabilité majoritaire ;
2. $a \approx s_0$:
   - $b \in [0{,}0\,;\,0{,}1]$: Déclin majoritaire ;
   - $b \in [0{,}2\,;\,0{,}6]$: Augmentation majoritaire ;
   - $b \in [0{,}7\,;\,1{,}3]$: Stabilité majoritaire ;
3. $a < s_0$:
   - quel que soit b : Augmentation majoritaire.

En conclusion, nous avons dénombré 3 catégories d'évolution des fonctions cognitives au cours du suivi. Ces dernières étaient différents de celles identifiés en section 3.1 mais tout

de même très proches à la fois. En effet, nous avons distingué un groupe d'augmentation, un de déclin et un mal caractérisé, interprété comme une stabilité cognitive.

4 Conclusion

En somme, ce chapitre contient deux points importants pour la suite du mémoire. Pour commencer, les deux analyses d'exploration préalable des données de MAPT présentées dans ce chapitre étaient complémentaires. La première avait pour objectif d'identifier et d'étudier des sujets ayant un profil d'évolution des fonctions cognitives extrêmes. Ce qualificatif d'extrême opposait d'une part des sujets ayant un taux d'accroissement strictement négatif tout au long du suivi, équivalent à un déclin des fonctions cognitives ; d'autre part, ceux ayant un taux d'accroissement strictement positif sur l'intégralité du suivi, signe d'amélioration des fonctions cognitives. Il s'avère que les sujets déclinant montraient une déficience en capacités cognitives et plus particulièrement en flexibilité mentale. Ils étaient moins autonomes et ressentaient davantage une plainte mémorielle ou une gêne dans leurs activités quotidiennes par rapport aux autres sujets.
La seconde analyse préliminaire portait sur la définition et l'étude de catégories de variations des fonctions cognitives. Trois catégories ont été déterminées : une stabilité relative du Z-score composite, une diminution de celui-ci ou une augmentation de celui-ci. Cette analyse a montré que près de trois individus sur quatre pouvaient être considérés comme stables cognitivement, ce qui signifie que leur Z-score composite restait, d'une visite sur l'autre et pour $a = b = 1$, dans l'intervalle $I_t = [Z_t - s_{t+1}; Z_t + s_{t+1}]$. Pour les autres sujets, soit ils amélioraient leur Z-score, soit ils montraient un profil de déclin cognitif alternant des phases de stabilité et de déclin. Ces sujets déclinant cognitivement étaient plus âgés, moins autonomes et plus nombreux à développer une démence au cours du suivi.
En somme, ces deux analyses confirmaient toutes deux que la population ne se comportait pas comme attendu a priori d'un point de vue clinique. Non seulement peu de sujets déclinaient tout au long du suivi, mais surtout, la majorité de la population n'avait pas d'évolution du niveau global des fonctions cognitives au cours du suivi.

Par ailleurs, le second point important de ce chapitre porte sur la nature des données longitudinales. Trois formes de données longitudinales ont été identifiées et étaient soit des données de survies, soit des données répétées, soit des données factorielles. La méthode d'analyse de chacune d'elles est différente. Dans le cas de données factorielles, qui nous intéresse particulièrement, la donnée n'est jamais observée telle qu'elle en pratique. Même avec un intervalle de temps infiniment petit, le temps reste discrétisé. Une solution d'analyse de ces données consiste à synthétiser l'information de façon à obtenir une succession de données, donc des mesures répétées dans le temps. Le point clef des paramètres de synthèse est de tenir compte à la fois des valeurs réelles et du temps séparant les mesures. Pour cela, nous proposons d'utiliser le taux d'accroissement entre deux visites. Par ailleurs, pour tenir compte du bruit des données, nous proposons de convertir les données fonctionnelles en une succession d'états continus indiquant des catégories de variation. Nous rappelons que l'utilisation de tels critères conduit à une perte d'information comparativement à l'utilisation de la donnée en l'état.

Dans le Chapitre 5, nous proposons différents outils de ré-analyse des données d'un essai clinique. Ces outils traitent directement les données factorielles ou sont adaptées pour utiliser les paramètres de synthèse décrits ci-dessus. Dans un premier temps sont abordées

des techniques d'exploration des données (section 1). L'objectif de ces outils est de rechercher des groupes de sujets homogènes. Dans un second temps, des méthodes de modélisation des données sont exploitées (section 2). Cette fois-ci, l'objectif est de mesurer l'effet de covariables sur l'évolution des fonctions cognitives.

TABLE 4.5 – Récapitulatif des variables quantitatives significativement différentes entre les deux profils extrêmes identifiés dans l'essai MAPT, sur les différentes périodes de suivi.

Variables	Période	Augmentation			Déclin			P-value
		Effectif	Moyenne	Écart-type	Effectif	Moyenne	Écart-type	
Âge(1)								
	T0-T6	774	75,09	4,25	712	75,43	4,49	0,1440
	T0-T12	333	74,69	4,11	139	75,47	4,65	0,0715
	T0-T24	68	73,50	3,62	37	75,95	4,50	0,0031
	T0-T36	9	72,88	2,47	10	76,90	4,12	0,0265(2)
TMT-A(1)								
	T0-T6	772	44,12	19,59	710	47,66	19,20	<0,001
	T0-T12	332	41,38	16,59	138	46,86	22,58	0,0037
	T0-T24	67	39,07	13,10	37	44,24	17,49	0,0911
	T0-T36	9	38,55	17,73	10	50,70	15,90	0,1527(2)
TMT-B(1)								
	T0-T6	759	111,54	55,41	695	125,74	60,56	< 0,001
	T0-T12	327	103,18	49,36	135	127,44	71,06	< 0,001
	T0-T24	66	95,62	34,52	35	134,46	97,11	< 0,001
	T0-T36	9	93,67	34,16	9	160,00	74,26	0,0400(2)
EVA-1(1)								
	T0-T6	773	49,01	17,55	712	52,04	16,77	< 0,001
	T0-T12	333	45,64	17,01	139	51,99	17,96	< 0,001
	T0-T24	68	43,68	17,27	37	55,65	15,95	< 0,001
	T0-T36	9	41,89	20,26	10	64,00	15,16	<0,001(2)
EVA-2(1)								
	T0-T6	773	37,54	23,02	712	40,08	23,34	0,0350
	T0-T12	333	36,32	21,58	139	43,23	23,43	0,0021
	T0-T24	68	35,09	20,28	37	44,43	21,18	0,0286
	T0-T36	9	33,22	17,79	10	53,00	22,43	0,0789(2)
Autonomie(1)								
	T0-T6	767	40,69	4,45	706	39,89	5,06	< 0,001
	T0-T12	332	41,13	4,11	138	39,07	6,08	< 0,001
	T0-T24	68	41,63	3,42	36	38,19	6,35	< 0,001
	T0-T36	8	43,00	2,73	9	35,89	11,31	0,1404(2)
MMSE(1)								
	T0-T6	773	28,21	1,57	711	27,59	1,98	< 0,001
	T0-T12	333	28,48	1,54	139	27,12	2,57	< 0,001
	T0-T24	68	28,66	1,35	37	27,08	2,58	<0,001
	T0-T36	9	29,00	0,00	10	25,20	3,94	< 0,001(2)

(1) : Comparaison des valeurs moyennes par test de l'ANOVA ;
(2) : Comparaison des valeurs moyennes par test de Mann-Witney Wilcoxon.

Chapitre 5

Outils de ré-analyse d'un essai de prévention non concluant

Les méthodes présentées dans le présent chapitre diffèrent de celles utilisées pour l'analyse des essais de la section 4 du Chapitre 2 en ce sens qu'elles ne sont habituellement pas mobilisées pour analyser, de prime abord, les résultats d'un essai clinique. En revanche, pour ré-analyser un essai de prévention non concluant tel que celui de MAPT, dont la description figure en Chapitre 3, et dans un contexte tel que celui décrit jusqu'alors avec les spécificités liées aux essais de prévention ou à la façon de traiter une donnée longitudinale, elles trouvent leur utilité. En effet, elles ont été adaptées au contexte des données longitudinales, des schémas d'évolution non nécessairement linéaires et soumises à l'incertitude de mesure pour répondre à la problématique du présent travail de thèse.

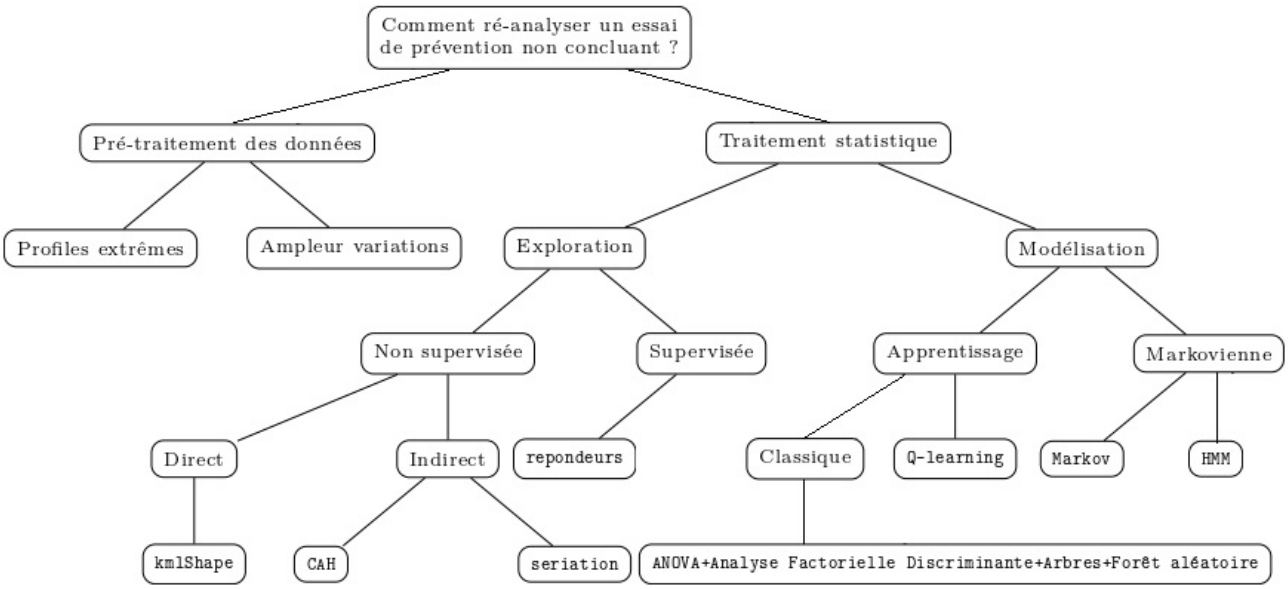

FIGURE 5.1 – Diagramme résumant l'ensemble des méthodes utilisées dans le présent travail pour ré-analyser les données de l'essai de prévention MAPT.

Deux catégories de méthodes de traitement statistique ont été distinguées dans ce chapitre, représentées en Figure 5.1, l'exploration et la modélisation des données. La première partie concerne l'exploration de trajectoires de fonctions cognitives. La stratégie d'analyse consiste à utiliser, dans un premier temps, des méthodes exploratoires non supervisées. Cela

signifie que nous ne cherchons pas à expliquer mais à observer une (ou des) variable(s). Les méthodes ont été regroupées dans la section 1.1 du *clustering* fonctionnel pour la recherche de groupes homogènes en terme d'évolution des fonctions cognitives. Trois stratégies distinctes ont été étudiées dans ce cadre, toutes basées sur les distances entre les trajectoires. La section 1.1.1 présente les k-*means*. En particulier, la section 1.1.1.1 rappelle les généralités sur le principe des k-*means* usuels, la section 1.1.1.2 se concentre sur les k-*means* longitudinaux et la section 1.1.1.3 s'intéresse aux k-*means* longitudinaux avec reconnaissance des formes des trajectoires. Nous nous sommes focalisés sur cette dernière extension de la méthode qui traite directement les données fonctionnelles (voir section 1 du Chapitre 4). Elle divise, de façon homogène, la population étudiée. Son avantage réside dans le traitement direct des données longitudinales fonctionnelles, la considération de l'allure de la trajectoire indépendamment d'un décalage dans le temps, mais son inconvénient provient de la nécessité à préciser le nombre de groupes à former. Pour éviter d'émettre cet a priori sur le nombre de groupes et à défaut de traiter directement les données fonctionnelles, une alternative pertinente consiste à appliquer une Classification Ascendante Hiérarchique (CAH - voir section 1.1.2). L'objectif de la méthode reste de constituer des groupes de sujets similaires. Dans notre cas, la méthode d'agrégation de Ward et le critère du coude pour l'inertie ont été appliqués pour déterminer le partitionnement optimal de la population. Le point fort de la CAH est sa facilité d'utilisation et de choix du nombre de groupe a posteriori, mais son point faible est de ne pas être spécifique aux données longitudinales. En conséquence, elle nécessite un pré-traitement des données, comme discuté dans la section 1.1.2. Ce pré-traitement consiste à définir un paramètre de synthèse de la donnée fonctionnelle (voir section 1 du Chapitre 4), dans notre cas un taux d'accroissement. La dernière méthode exploratoire non supervisée étudiée dans ce travail est la sémiologie graphique, présentée en section 1.1.3. Cette méthode d'analyse combinatoire réorganise une matrice de dissimilarités, par les lignes et/ou les colonnes, de façon à déterminer une structure en groupes homogènes. L'avantage de cette méthode est d'être visuelle et facile d'interprétation, mais ses inconvénients majeurs sont communs d'une part aux méthodes de la section 1.1.1, par le besoin de déterminer a priori le nombre de groupes, d'autre part à la méthode de la section 1.1.2, par l'utilisation d'un indicateur de variations pour s'adapter aux données longitudinales. Parallèlement aux méthodes non supervisées, une méthode de classification fonctionnelle supervisée est également abordée en section 1.2. Cette méthode vise à créer des groupes selon un critère défini a priori, de nature qualitatif, dont chaque modalité représente un groupe. Dans notre cas, la classification est dichotomique et correspond à un statut de répondeur ou non. C'est en cela que la méthode est supervisée, car la population est divisée selon une variable à expliquer. Cette méthode est rapide d'utilisation, traite directement des données fonctionnelles, mais la difficulté découle de la définition des modalités de la variable de groupe. Pour une vision d'ensemble de cette stratégie exploratoire, une synthèse des quatre méthodes est proposée dans le Chapitre 7.

Dans la seconde partie de ce chapitre, notre intérêt n'était plus orienté sur la description et la comparaison de groupes définis mais plutôt sur la mesure d'effets de covariables. Pour cela, trois méthodes de modélisation des données ont été étudiées et détaillées en section 2. Dans un premier temps, une méthode d'apprentissage statistique par renforcement a été appliquée, en section 2.1. Cette technique segmente un suivi en différentes étapes clefs nommées décisions. L'ensemble de toutes les décisions prises au cours du suivi conduit à un résultat potentiel. Le *Q-learning* vise à optimiser la règle de décision, étape par étape, à partir des données et de régressions linéaires successives. Sa particularité, en sus d'être indirecte en plusieurs étapes, consiste à considérer le temps en sens inverse. Les avantages de

la méthode sont de traiter directement avec des données longitudinales fonctionnelles et de tenir compte des effets à retardement d'un traitement. L'inconvénient est qu'elle s'applique difficilement si l'essai n'a pas été construit pour être analysée avec cette stratégie. Dans un second temps, les approches markoviennes ont été étudiées (section 2.2). Cela signifie que l'évolution des fonctions cognitives n'est plus observé en tant que trajectoire longitudinale mais comme une succession d'états au cours du temps (voir section 1 du Chapitre 4). La modélisation du processus multi-états consiste à identifier les facteurs affectant chaque transition possible entre les différents états. De façon naturelle, la modélisation nous mène vers une approche par modèles de Markov classiques présentés en section 2.2.1. L'avantage de cette méthode est d'être conçue pour analyser des mesures répétées dans le temps, mais l'inconvénient majeur est la définition d'un certain nombre de paramètres a priori (nombre d'états, probabilités initiales, matrice d'intensité de transition). Ensuite, l'objectif était de généraliser la méthode précédente en appliquant des modèles de Markov cachés (HMM - section 2.2.2) tenant compte de l'incertitude de mesure. Avec cette méthode, la construction du processus stochastique est double : d'une part un modèle de Markov tel que présenté en section 2.2.1 mais non connu, d'autre part un processus constitué de l'ensemble des observations. Ces modèles aident dans la détermination du nombre d'états régissant la partie Markovienne, mais le choix des paramètres de distribution initiale impacte directement les résultats obtenus.

1 Techniques d'exploration des trajectoires de fonctions cognitives

Dans un premier temps, nous avons porté notre attention sur l'exploration de trajectoires de fonctions cognitives. Comme indiqué dans la section 1.3.1 du Chapitre 2 ou la section 1.3.1 du Chapitre 3, l'étude de la maladie d'Alzheimer dans les essais cliniques utilise les fonctions cognitives comme marqueur de substitution de la démence. Ces dernières sont relevées de façon répétée au cours du suivi, mais proviennent en réalité d'une fonction dépendante du temps partiellement observée. L'hypothèse émise est qu'au sein d'une population participant aux essais cliniques, l'hétérogénéité de la population serait telle que certains sujets, ayant un profil d'évolution similaire, n'apparaitraient pas avec les méthodes d'analyse classiques. Le terme profil désigne mathématiquement une trajectoire, cliniquement une évolution caractéristique d'une population. Les sujets pourraient être regroupés selon leurs points communs de façon à définir des critères de prévention spécifiques à chaque population.

Nous présentons, dans les différentes sections ci-après, plusieurs méthodes d'identification de groupes de trajectoires cognitives similaires. Elles peuvent être divisées en deux classes. La section 1.1 présente des méthodes d'analyse non supervisée. Elles permettent de déduire la structure naturelle d'un ensemble de données à partir de différentes observations. En particulier, elles ne sont pas conçues pour expliquer une variable réponse. La section 1.2 en revanche décrit une méthode d'analyse de profils supervisée, en ce sens qu'une hypothèse forte sur la nature du profil de chaque groupe est émise a priori. Dans notre cas, les sujets ont été classés dans deux groupes connus. L'objectif de cette analyse est de déterminer les caractéristiques permettant de discriminer ces groupes. À terme, pour un ensemble de caractéristiques individuelles, nous aimerions prédire le groupe d'appartenance du sujet. La variable à expliquer est donc le groupe de répondeur.

1.1 Le *clustering* fonctionnel pour la recherche de groupes homogènes

Les méthodes de *clustering* de données fonctionnelles sont utilisées pour faciliter l'étude de populations d'effectifs importants en structurant un ensemble de données longitudinales [188]. Au sein de ces dernières, nous portons un intérêt particulier aux mesures répétées dans le temps ou aux données fonctionnelles. L'objectif principal de cette famille de méthodes est d'identifier des structures communes, de résumer des comportements, d'affecter de nouveaux individus à ces groupes et d'identifier les cas totalement atypiques [189, 190]. Aucune hypothèse sur la nature des trajectoires au sein de chaque groupe n'est nécessaire pour appliquer cet ensemble de méthodes ni sur l'incertitude liée aux mesures, constituant son avantage majeur [188, 191]. Le principal inconvénient provient de leur absence d'estimation de paramètres à interpréter, comme nous en fournissent les modèles de régression usuels [191]. Ces méthodes donnent lieu à une première approche purement descriptive des données.

Pour atteindre l'objectif commun des méthodes, qui est de rechercher des groupes homogènes au sein d'une population, plusieurs stratégies existent. Chaque façon de procéder possède ses propres conditions ou spécificités et conduit à des résultats différents. Dans cette première partie, quatre méthodes ont été sélectionnées pour être appliquées à notre contexte particulier. Pour commencer, nous avons sélectionné trois algorithmes de *clustering* fonctionnel : l'un est non hiérarchique (*k-means*) [14] et utilise une trajectoire observée comme critère de jugement ; l'autre est hiérarchique (CAH) [15] et le dernier est graphique [16]. Ces deux derniers algorithmes exploitent une succession de paramètres de synthèse de la donnée fonctionnelle (voir section 1 du Chapitre 4). Ensuite, en section 1.2, l'analyse des répondeurs est présentée. Cette analyse dichotomise la population selon un seuil de valeurs du Z-score composite défini d'après des critères cliniques. Elle traite ainsi directement les données fonctionnelles.

1.1.1 L'algorithme des centres mobiles

Parmi les méthodes d'analyse de données longitudinales, la méthode d'agrégation autour des centres mobiles a retenu notre attention [188, 191, 192, 193]. En effet, cet algorithme est conçu pour diviser une population en sous-groupes de sujets homogènes. Nous avons utilisé une extension des *k-means* spécifique aux données fonctionnelles, présentée en section 1.1.1.3. Aussi, nous rappelons en premier lieu l'algorithme classique des *k-means* en section 1.1.1.1, puis l'algorithme des *k-means* pour données longitudinales en section 1.1.1.2 pour enfin en arriver à l'extension d'intérêt. Comparativement aux deux autres, elle permet de tenir compte du décalage dans le temps et de la forme des trajectoires pour construire les groupes de sujets.

1.1.1.1 La méthode des *k-means* usuelle

De façon automatique, dynamique et non supervisée, cette méthode de partitionnement en k-moyennes (ou *k-means*) consiste à diviser une population donnée de taille n en k groupes, k étant fixé a priori [191, 194]. L'ensemble des informations disponibles concernant les individus est examiné par les différentes covariables explicatives. Chaque individu est un point de l'espace, composé de p coordonnées correspondant aux p variables observées. Les

groupes sont constitués selon un critère de proximité permettant d'identifier des individus possédant des caractéristiques similaires à partir des p variables quantitatives [191, 190]. Des variables qualitatives peuvent aussi être utilisées, mais ce ne sera pas notre cas. Cette méthode est particulièrement intéressante pour des bases de données de grande dimension que nous ne souhaitons pas réduire au préalable, comme avec une Analyse en Composantes Principales par exemple.

Cette méthode demande que le nombre de groupes souhaités k soit fixé initialement [192]. Ensuite, k points sont tirés aléatoirement parmi les données existantes [195]. Ils peuvent également, au même titre que le nombre de groupes, être choisi a priori selon les connaissances cliniques du phénomène étudié, mais ce ne sera pas notre cas. Ils correspondent aux centres provisoires (centroïdes, barycentres) des premiers groupes. La matrice des distances entre les individus et les centres est calculée. Chaque individu à classer va être attribué au centre provisoire le plus proche selon un critère de proximité. Lorsqu'il n'y a plus d'individu non attribué, les premiers groupes sont constitués et leurs centres de gravité recalculés, toujours provisoirement. Ces centres deviennent des points fictifs et se déplacent par rapport aux groupes formés à l'étape précédente et donc par rapport à leur valeur initiale. De façon récursive, les individus sont de nouveaux attribués aux centroïdes les plus proches. À chaque étape, une nouvelle partition des données est obtenue et les centroïdes recalculés avant ré-attribution des individus, d'où le qualificatif de centres mobiles [191]. L'algorithme se termine lorsque les centres des groupes (*clusters*) ne changent plus d'emplacement (voir l'algorithme simplifié 1). La Figure 5.2 illustre ces propos en montrant à gauche la population totale et progressivement vers la droite différents types de *clusters* en fonction du nombre k choisi, variant entre 2 et 4.

Algorithme simplifié 1 :
Choix a priori :

1. donnée = observation
2. k = le nombre de groupes souhaité
3. distance permettant d'évaluer la dissimilarité/similarité entre les individus

Initialisation : tirer aléatoirement k données ou choisir selon les connaissances cliniques k données
Tant Que les centroïdes sont modifiés **Faire** :
— Calcul des dissimilarités entre individus et les centroïdes
— Recherche du minimum des dissimilarités
— Attribuer chaque individu au centroïde le plus proche
— Actualiser le centroïde des groupes
Fin Tant Que
Partitionnement final : déterminé selon le nombre de groupes choisi.

Au sein de chaque groupe formé par les *k-means*, les individus ne sont pas ordonnés en sous-groupes imbriqués selon un critère de ressemblance, ce qui signifie que l'algorithme n'est pas hiérarchique.

Le critère d'optimisation de la partition utilise une mesure d'homogénéité de chaque *cluster* [193]. Dans le cas présent, la proximité entre les individus se quantifie par une distance dans un espace multidimensionnel. La distance usuelle dans l'espace des matrices

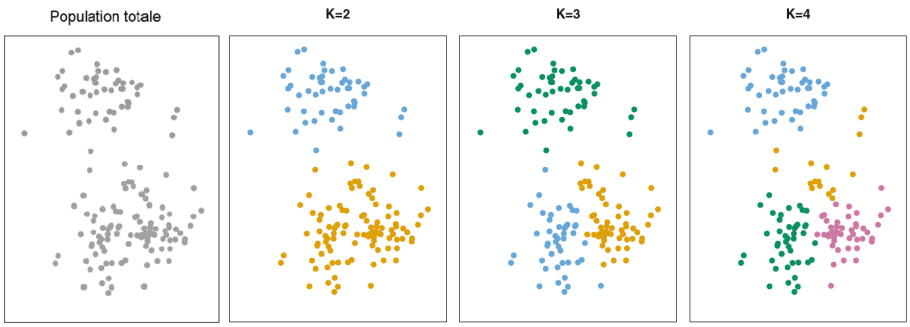

FIGURE 5.2 – Illustration de l'algorithme des *k-means* classique sur un jeu de données simulées de 150 individus [196]. À gauche, la population totale. Ensuite, les résultats de l'algorithme pour différentes valeurs du nombre de groupes k variant de 2 à 4. Les couleurs indiquent l'appartenance aux différents groupes.

carrées d'ordre n est la distance euclidienne [194, 197]. Elle se définit comme :

$$d(i, i') = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \quad ,$$

avec $x_{i'j}$ la valeur prise par l'individu i pour la variable j et $x_{i'j}$ celles de l'individu i'. L'objectif consiste à réduire la distance pour former des groupes sujets homogènes.

D'autres distances existent [193] et peuvent être sélectionnées : elles modifient le résultat final de l'algorithme mais pas son fonctionnement même. Cependant, le choix de la distance est complexe et souvent arbitraire mais également dépendant du type de données à analyser. Par exemple, la distance de Tchebychev, calculant la différence maximale des distances ou norme infinie, s'appuie sur très peu de valeurs et des séries triées. L'inconvénient de cette distance survient lorsque les valeurs extrêmes sont aberrantes. La distance de Mahalanobis est basée sur la corrélation et l'observation de moyennes, aussi appelée distance généralisée interpoint carré. Elle détermine la similarité entre une donnée connue et une inconnue, donnant un poids plus important aux composantes éloignées. Elle peut également comparer des groupes définis par une variable catégorielle, ce qui signifie que la distance de Mahalanobis intègre à la fois des données quantitatives et qualitatives. La distance de Manhattan, aussi dite *city-block*, s'intéresse au cheminement pour aller d'un point A à un point B selon un quadrillage. Elle utilise donc la distance en valeur absolue ou norme 1 [194]. Mesurant la différence moyenne entre deux individus, les valeurs extrêmes n'ont pas un poids important. De plus, elle traite les variables qualitatives. La dernière distance présentée ici est celle de Minkowski. Elle généralise de la distance Manhattan et euclidienne.

Cependant, sans recommandations cliniques particulières, nous avons considéré la distance euclidienne, calculée à partir de l'ensemble des valeurs quantitatives. Elle traite chaque composante d'un vecteur comme indépendant et de façon identique. Elle est connue pour tenir compte des variances des variables ainsi que de leur corrélation [190]. Elle accorde plus d'importance à la différence entre deux individus sur une variable ayant une forte variance que sur une variable avec faible variance. Lorsque les variables sont standardisées ou centrées réduites, cette distance correspond à un indice usuel de dissimilarité entre les variables. En effet, si $var(x)$ représente la variance de la variable x, $cov(x, y)$ la covariance entre x et y et $r(x, y)$ le coefficient de corrélation de Pearson entre x et y, alors la distance entre deux

points i et i' s'écrit :

$$d^2(i,i') = \sum_{j=1}^{p} \left(var(x_{ij}) + var(x_{i'j}) - 2cov(x_{ij}, x_{i'j})\right) \quad ,$$

$$d^2(i,i') = \sum_{j=1}^{p} \left(var(x_{ij}) + var(x_{i'j}) - 2r(x_{ij}, x_{i'j})\sqrt{var(x_{ij})}\sqrt{var(x_{i'j})}\right) \quad .$$

Dans le cas as particulier de variables centrées (espérance nulle) et réduites (variance égale à 1), $d^2(i,i') = 2\sum_{j=1}^{p}(1 - r(x_{ij}, x_{i'j}))$. Ainsi, à variances égales, les individus i et i' sont d'autant plus distants que les coefficients de corrélation de leurs variables quantitatives sont faibles en valeur absolue. La distance euclidienne peut également, dans d'autres cas que le nôtre, être étendue en présence de variables qualitatives ou dans un cas mixte (variables quantitatives et qualitatives) [190].

Finalement, lorsque l'algorithme n'est exécuté qu'une seule fois, la partition obtenue n'est pas nécessairement optimale [191, 195]. C'est pourquoi l'exécuter plusieurs fois en faisant varier les centres initiaux est recommandé pour évaluer la stabilité du résultat principalement.

1.1.1.2 La méthode des *k-means* longitudinale

Une extension de la méthode des *k-means* présentée à la section 1.1.1.1 a été développée pour s'appliquer au cas particulier de données longitudinales. Elle porte le nom de *k-means for longitudinal data* (`KmL`) [195, 198]. La donnée considérée en théorie est une donnée fonctionnelle, mais en pratique, cette trajectoire n'est observée qu'en un certain nombre de points. Cela revient à considérer une succession de valeurs mesurées à différents moments [197]. Ces mesures répétées sont notamment liées entre elles (dépendantes). L'extension de la méthode des *k-means* repose ainsi sur un algorithme similaire au précédent, si ce n'est que cette fois ci, le calcul d'une distance entre les trajectoires pour former les différents groupes est nécessaire (voir l'algorithme simplifié 2). Puisqu'aucune hypothèse sur la forme de ces dernières n'est émise et qu'aucune variable n'est à expliquer, cette analyse n'a pour visée que la description des données [195]. L'objectif consiste à définir une partition des individus en k groupes, k fixé a priori, dont les trajectoires sont proches relativement à la distance choisie (voir l'algorithme simplifié 2).

Dans ce contexte de données longitudinales, notons y_{it} la valeur de la variable quantitative d'intérêt Y pour un sujet i donné au temps t. Pour quantifier la proximité entre les sujets, l'algorithme `KmL` utilise, entre autres, une distance L^2, assimilée en pratique à une distance euclidienne entre deux trajectoires, définie alors comme [195] :

$$d(y_i, y_{i'}) = \sqrt{\sum_{t=1}^{T}(y_{it} - y_{i't})^2} \quad .$$

Les données longitudinales sont comparées en chaque temps du suivi, ce qui implique que deux trajectoires sont homogènes si leur distance est faible la plupart du temps. Deux trajectoires peuvent être proches dans l'ensemble mais avoir un point de mesure éloigné. Ce critère de proximité entre les sujets autorise une comparabilité des résultats avec d'autres méthodes d'analyses de données utilisant des matrices de dissimilarité. D'autres distances sont tout de même disponibles dans le package `KmL` [195], mais nous ne les détaillerons

pas ici. Parmi elles se trouvent celles exposées en section 1.1.1.1, telles que la distance de Manhattan ou de Minkowski.

Algorithme simplifié 2 :
Choix a priori :

1. donnée = trajectoire
2. k = le nombre de groupes souhaité
3. distance permettant d'évaluer la dissimilarité/similarité entre les trajectoires

Initialisation : tirer aléatoirement k données ou choisir selon les connaissances cliniques k données
Tant Que les centroïdes sont modifiés **Faire** :
— Calcul des dissimilarités entre trajectoires et les centroïdes
— Recherche du minimum des dissimilarités
— Attribuer chaque trajectoire au centroïde le plus proche
— Actualiser le centroïde des groupes
Fin Tant Que
Partitionnement final : déterminé selon le nombre de groupes choisi.

Une difficulté de la méthode usuelle, outre le choix de la distance, se trouve dans la définition du nombre optimal de groupes k. En l'absence de connaissances sur k, une possibilité consiste à exécuter plusieurs fois l'algorithme automatiquement (par défaut 20 fois [195]), en variant les conditions initiales et le nombre de *clusters*. Ensuite, le nombre k optimal minimisant ou maximisant un critère choisi est sélectionné. L'extension `KmL` met à disposition plusieurs critères évaluant la qualité des partitions. Le plus utilisé, implémenté par défaut dans le package `KmL`, est le critère non paramétrique de Calinski et Harabazt [199]. Il se calcule à partir du rapport des traces des matrices de covariance inter-*clusters* $M_{inter} = \sum_{m=1}^{k} n_m(\overline{y_m} - \overline{y})(\overline{y_m} - \overline{y})'$ et intra-*clusters* $M_{intra} = \sum_{m=1}^{k}\sum_{l=1}^{n_m}(y_{ml} - \overline{y_m})(y_{ml} - \overline{y_m})'$ [195]. Dans ces expressions, n_m représente le nombre de trajectoires au sein du cluster m, $\overline{y_m}$ la trajectoire moyenne de ce cluster, $\overline{y}$ la trajectoire moyenne de la population considérée. Notons également que v' correspond au vecteur v transposé. La trace de ces matrices résulte de la somme de leurs coefficients diagonaux. Celle de la matrice inter-*clusters* devient grande lorsque les groupes sont distincts et clairement séparés. Celle de la matrice intra-*clusters* devient petite lorsque les groupes formés sont homogènes. La partition optimale en k groupe minimise ainsi le critère défini comme [195, 199] :

$$C(k) = \frac{Trace(M_{inter})}{Trace(M_{intra})} \times \frac{n-k}{k-1} \quad .$$

Finalement, le dernier avantage de la méthode *k-means for longitudinal data* (`KmL`) renvoie à la prise en compte des données manquantes au cours du suivi [198]. La stratégie la plus simple exclue les individus ayant des données manquantes au cours de la trajectoire. Cette stratégie conduit parfois à une perte considérable d'information. Deux alternatives sont proposées par le package : imputer les données manquantes, c'est-à-dire attribuer une valeur à ces données [200], ou utiliser la distance euclidienne avec correction de Gower [195]. Les inconvénients de l'imputation est non seulement le choix de la méthode d'imputation, mais surtout la potentielle non adéquation de la valeur à la réalité clinique. Dans un cas nécessitant l'utilisation de cette méthode d'imputation, il est nécessaire de faire attention

aux choix effectuées. Pour la distance euclidienne avec correction de Gower, choisie par défaut par le package `Kml`, elle se calcule par :

$$d_{Gower}(y_i, y_{i'}) = \sqrt{\frac{t}{\sum_{k=1}^{t} w_{ii'k}} \sum_{k=1}^{t} (y_{ik} - y_{i'k})^2 . w_{ii'k}} \quad ,$$

avec $w_{ii'} = 0$ si y_{ik} ou $y_{i'k}$ ou les deux sont manquantes ; $w_{ii'} = 1$ sinon, ce qui nous ramène au cas de la distance euclidienne. Une fois le *clustering* terminé, le calcul des critères de qualité du partitionnement nécessite une imputation [195]. Autrement dit, le *clustering* peut être effectué sans imputation des données et avec la distance euclidienne corrigée, mais l'imputation devra tout de même avoir lieu pour déterminer le nombre optimal de *clusters* par le critère de Calinski et Harabazt [199]. Plusieurs méthodes sont proposées par le package de R, mais nous détaillerons celle de *CopyMean* dans la section 1.1.1.3 ci-après. Cette méthode s'appuie sur une imputation classique, qu'elle corrige ensuite par la moyenne des autres valeurs. Elle s'adapte donc aux variations au cours du temps.

1.1.1.3 La méthode des *k-means* longitudinale basée sur la forme des trajectoires

Dans la section 1.1.1.2 précédente, nous avons noté que le partitionnement des *k-means* longitudinaux s'effectuait selon une ressemblance à chaque temps de mesure (*clusters* par similarité locale [197]). Les trajectoires sont proches au même moment, mais la méthode ne tient pas compte du comportement global [197]. Deux trajectoires similaires ayant des amplitudes différentes devraient appartenir au même groupe (voir Figure 5.3 (c)). Seulement, le package `KmL` ne tient pas compte de la forme générale des trajectoires, de son allure globale. En effet, deux trajectoires de formes similaires, c'est-à-dire ayant des valeurs identiques décalées dans le temps, ne sont pas attribuées à un même groupe. Nous souhaiterions à présent partitionner les trajectoires non plus sur des similarités locales, comme avec `KmL`, mais aussi de façon indépendante d'un potentiel décalage dans le temps (voir Figure 5.3 (d)).

C'est précisément dans ce but que Génolini et al. [14] ont développé en 2016 une extension de leur précédent package `KmL` (*k-means for longitudinal data* - voir section 1.1.1.2). Toujours en considérant des données fonctionnelles directement et en utilisant le principe des *k-means* (voir l'algorithme simplifié 3), la méthode des *k-means* longitudinaux basées sur la forme des trajectoires (kmlShape [14]) utilise comme mesure de similarité la distance généralisée de Fréchet, sur laquelle nous revenons juste après. Cette méthode de partitionnement se qualifie de partitionnement silhouette. Elle est non hiérarchique, automatique à l'instar de celles présentées dans les sections précédentes et s'adapte à un ensemble volumineux de données, particulièrement dans le cas d'un nombre important d'individus. Elle est rapide d'utilisation et ne nécessite aucune hypothèse sur la forme des trajectoires attendues par groupe en fin de partitionnement [14]. Ainsi, l'objectif de la présente extension reste similaire à la méthode usuelle présentée en section 1.1.1.1, à savoir diviser la population en k sous-groupes homogènes, k défini a priori. Cependant, la spécificité de la présente méthode est de le faire indépendamment d'un décalage dans le temps et avec des distances adaptées à la fois aux données longitudinales et à la forme des trajectoires. Initialement, l'algorithme `kmlShape` a été construit pour analyser les données de cohorte.

Algorithme simplifié 3 :
Choix a priori :

1. donnée = trajectoire
2. k = le nombre de groupes souhaité
3. distance permettant d'évaluer la dissimilarité/similarité entre les trajectoires

Initialisation : tirer aléatoirement k données ou choisir selon les connaissances cliniques k données
Tant Que les centroïdes sont modifiés **Faire** :
— Recaler les trajectoires
— Calcul des dissimilarités entre trajectoires et les centroïdes
— Recherche du minimum des dissimilarités
— Attribuer chaque trajectoire au centroïde le plus proche
— Actualiser le centroïde des groupes
Fin Tant Que
Partitionnement final : déterminé selon le nombre de groupes choisi.

Afin de définir la proximité entre les sujets, l'algorithme `kmlShape` s'appuie sur la distance généralisée de Fréchet [14]. Cette dernière tient compte de la forme de la trajectoire et attribue une faible valeur si deux trajectoires sont proches en terme d'allure, de forme (*shape-respecting distance*, voir Figure 5.3) [14, 197]. Elle tend vers la distance euclidienne lorsque les séries temporelles sont longues, c'est-à-dire lorsque le nombre de mesures par sujets est élevé. Elle généralise à la fois la distance de Fréchet classique [201] et la distorsion temporelle dynamique (*Dynamic Time Warping*) [202]. Ces deux quantités, variantes l'une de l'autre, considèrent les données longitudinales comme indépendantes à chaque temps et ne sont donc pas adaptées en l'état pour traiter un problème de classification de courbes, d'où l'utilisation d'une version généralisée de la distance de Fréchet comme mesure de similarité.

La distance généralisée de Fréchet, de paramètre d'échelle de temps λ entre deux trajectoires P et Q, mesure la proximité entre des courbes continues [197]. Elle résulte d'une transformation affine $A : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \lambda x \\ y \end{pmatrix}$ de la distance de Fréchet classique. En notant o la composition de fonctions, α et β deux fonctions paramétrant les trajectoires respectives de P et Q, cela implique que [14] :

$$\begin{aligned} d_\lambda^F(P,Q) &= Inf_{(\alpha,\beta)} Max\left(Po\alpha oA, Qo\beta oA\right) \\ &= Inf_{(\alpha,\beta)} Max\left(d\left[\begin{pmatrix} \alpha(\lambda s) \\ P(\alpha(s)) \end{pmatrix}, \begin{pmatrix} \beta(\lambda s) \\ Q(\beta(s)) \end{pmatrix}\right]\right) \quad , \ s \in [0,t] \quad . \end{aligned}$$

Dit autrement, cette distance permet de faire un recalage de courbe, en re-paramétrant les trajectoires P et Q respectivement par α et β [14]. De façon à illustrer ce propos, reprenons l'exemple de Génolini et al. [14] dans lequel la distance de Fréchet est assimilée à une laisse représentant la longueur minimale séparant un maître de son chien lors d'une balade. Les deux individus marchent à des vitesses différentes le long de deux trajectoires. Chaque point de chaque trajectoire est associé au point le plus proche de l'autre trajectoire. La distance de Fréchet entre les deux courbes est égale au plus long lien séparant les deux trajectoires. Or ici, supposons que le chien précède son maitre de plusieurs minutes. L'échelle de temps de marche du chien est donc différente de celle du maitre, impliquant de reparamétrer les trajectoires dans une échelle de temps identique pour les comparer. Ainsi, la distance généralisée de Fréchet mesure la similarité entre les courbes décalées dans le temps, avec les

critères de la distance de Fréchet usuelle (max de la distance séparant les courbes).

Pour la construction des trajectoires moyennes des groupes (centres des *clusters*), la moyenne classique n'est pas adaptée à l'objectif du package. Les trajectoires en résultant sont moyennées, plus lissées (voir Figure 5.3 (f)). Ainsi, la moyenne de Fréchet est implémentée dans `kmlShape` [14]. Elle s'appuie sur la moyenne entre deux trajectoires $\mu_\lambda^F(P,Q) = \dfrac{Po\alpha oA + Qo\beta oA}{2}$, puis entre n trajectoires. Par définition, elle tient aussi compte de la silhouette des trajectoires présentes, d'où son nom de *shape-respecting mean*.

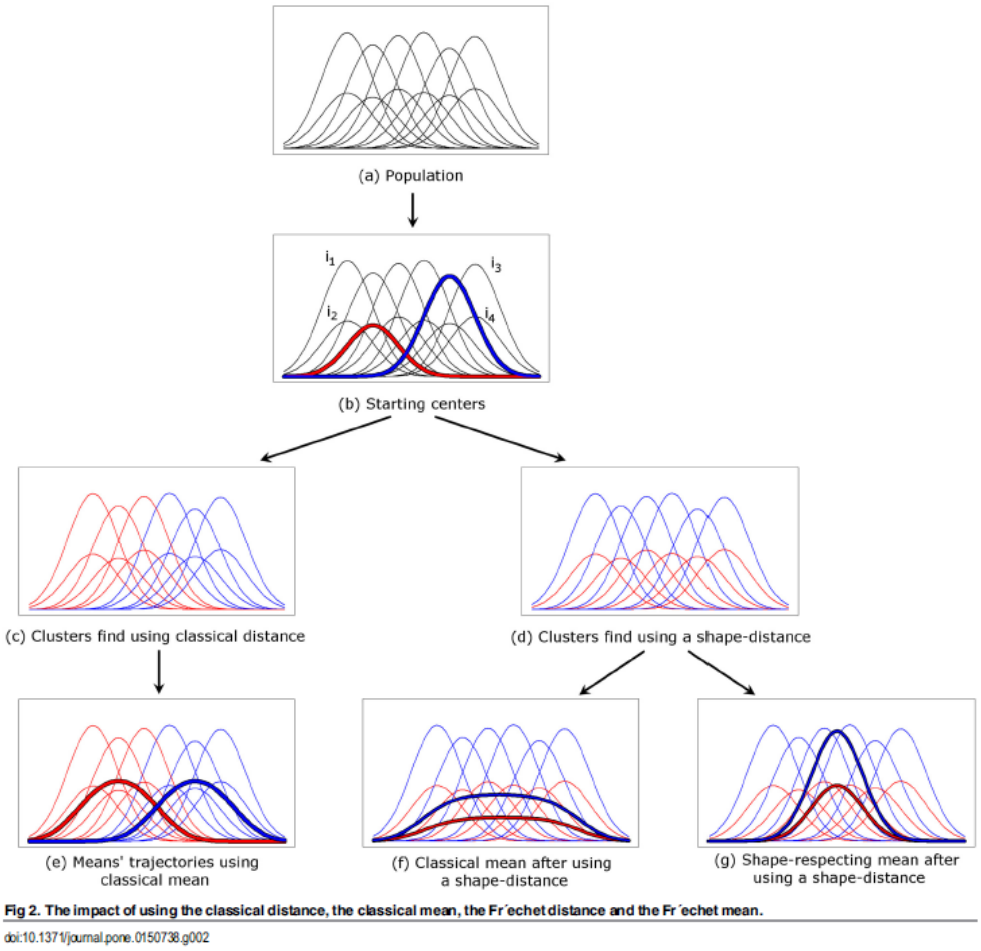

FIGURE 5.3 – Illustration de l'impact des distances classiques, de la moyenne classique, de la distance de Fréchet et de la moyenne de Fréchet [14].

La Figure 5.3, extraite de l'article de référence du package `kmlShape` [14], fournit un appui visuel à l'ensemble de ces propos. Rappelons que le principe de la méthode est identique à celles vues jusqu'à présent mais la mesure de similarité est modifiée. La Figure 5.3 est décomposée comme suit : (a) Représentation des trajectoires de l'ensemble de la population d'étude ; (b) Attribution des centres initiaux par tirage aléatoire de trajectoires (ici $k = 2$ *clusters*) ; (c) Formation des groupes avec des distances classiques, comme la distance euclidienne avec correction de Gower ; (d) Formation des groupes avec la distance généralisée de Fréchet respectant la forme de la trajectoire ; (e) Trajectoires moyennes des groupes avec la moyenne classique ; (f) Trajectoire moyenne classique après utilisation de la distance généralisée de Fréchet ; (g) Trajectoires moyennes des groupes avec distance respectant la forme de la trajectoire et moyenne des trajectoires. La Figure 5.3 (g) montre ainsi comment, en utilisant la distance généralisée de Fréchet et la moyenne de Fréchet, le package `kmlShape`

conduit à un *clustering* de courbes basé sur la forme de la trajectoire.

Ainsi, la Figure 5.3 illustre le fait que, si deux trajectoires sont tirées au hasard pour initialiser les algorithmes des *k-means*, les groupes formés par la méthode classique seront similaires sur la plupart des points [14]. En effet, ici une distribution gaussienne, de moyenne modérée et de variance importante, est tirée pour le *cluster* rouge. Pour le bleu, la moyenne est supérieure à la rouge et la variance inférieure. Avec la méthode classique, toutes les trajectoires ayant une moyenne autour de celle de la courbe rouge sont assignées au même *cluster* (Figure 5.3 (c)), de même pour les bleues. Ainsi, les sujets ayant la même courbe au même moment sont attribués au même groupe, de façon indépendante de la forme de leur trajectoire. Lorsque les moyennes classiques sont appliquées pour déterminer l'allure moyenne du groupe, nous constatons que les courbes de faible moyenne et variance élevée sont mélangées à celles de moyenne élevée et variance faible [14]. Lorsque la distance généralisée de Fréchet est utilisée (Figure 5.3 (d)), les deux groupes ne correspondent plus à la moyenne mais à la forme de la trajectoire [14]. Toutes celles avec une faible moyenne et variance importante appartiennent au même groupe, quel que soit leur date de réalisation, de même pour celles avec moyenne élevée et une variance faible. Ainsi, les groupes sont formés selon la forme, indépendamment du temps d'occurrence. Enfin, lorsque la moyenne de Fréchet est utilisée (Figure 5.3 (g)), la courbe moyenne des groupes correspond bien à une moyenne élevée et variance faible pour le groupe bleu, faible moyenne et variance élevée pour le groupe rouge. L'information extraite selon le package `KmL` ou selon le package `kmlShape` est donc complètement différente.

Une spécificité non négligeable des données observationnelles longitudinales est la présence et le traitement des données manquantes. La stratégie la plus simple mais la plus sélective, comme indiquée dans la section 1.1.1.2, consiste à ne pas les considérer. La deuxième stratégie est d'utiliser les distances adaptées à la gestion des données manquantes, qui se traduisent en pratique comme la distance euclidienne avec correction de Gower pour `KmL` [195] et la distance de Fréchet généralisée pour `kmlShape` [201]. La dernière option envisageable vise à imputer les données manquantes [200]. Une méthode d'imputation des données, proposée par le package, se nomme *CopyMean* [203, 204]. Elle utilise une version modifiée de l'interpolation linéaire tenant compte des variations au cours de la trajectoire. Pour chaque individu i, notons y_{ia} la trajectoire jusqu'au point a précédant la valeur manquante z et y_{ib} pour le point suivant cette valeur manquante z. Notons également ($\overline{y_{m1}}$, ..., $\overline{y_{mt}}$) les moyennes des trajectoires observées aux différents temps. La première étape de l'imputation consiste à implémenter l'interpolation linéaire, en reliant la dernière valeur disponible à la prochaine renseignée [203]. La seconde étape d'imputation se construit à partir de la trajectoire moyenne de tous les autres sujets disponibles au temps manquant [203]. Ensuite, *CopyMean* combine ces deux résultats pour fournir la valeur de l'imputation finale. Concrètement, cela se traduit par :

$$y_{iz} = \frac{y_{ib} - y_{ia}}{\overline{y_{mb}} - \overline{y_{ma}}} + y_{ia} \quad .$$

Lorsque les données sont manquantes en fin de suivi, c'est-à-dire qu'elles ne sont plus intermittentes mais monotones, selon la terminologie utilisée dans [204], cette stratégie ne peut être appliquée puisque y_{ib} n'existe pas. Cela signifie que l'interpolation linéaire n'est pas utilisable. La première étape d'interpolation est alors remplacée par le maintient de la dernière valeur disponible dans le temps (*Last Observation Carried Forward* (LOCF)). La seconde étape décrite précédemment s'appuyant sur la trajectoire moyenne des autres individus présents en fin de suivi reste inchangée. En conséquence, la nouvelle valeur d'imputation est

donnée par :

$$y_{iz} = \overline{y_{mz}} + (y_{ia} - \overline{y_{ma}}) \quad .$$

L'avantage majeur de cette méthode d'imputation proposée par le package `kmlShape`, outre la considération de données répétées comme dépendantes au cours du temps [197], réside dans l'absence d'hypothèse sur la forme de la trajectoire ou sur la distribution des *clusters*. De plus, considérer la distance généralisée de Fréchet comme mesure de similarité permet de comparer des trajectoires différentes : il n'est nécessaire ni d'avoir le même nombre de mesures par sujet, ni que les temps de mesures soient identiques d'un sujet à l'autre [14, 201].

Un inconvénient survient cependant lorsque la taille du jeu de données est grand car l'algorithme est gourmand en temps de calcul. En effet, le package propose trois options : utiliser le jeu directement, réduire le nombre d'individus pour l'initialisation des *clusters* ou réduire le nombre de points composant la trajectoire. Cette réduction de la taille du jeu de données conduit exactement au même résultat final mais réduit le temps de calcul [14]. Dans le premier cas, aucun inconvénient majeur. La réduction du nombre d'individus signifie qu'un certain nombre d'entre eux sont tirés au hasard dans la population d'étude pour initialiser l'algorithme. Ces sujets sont considérés, en raison du nombre limité de trajectoires typiques, comme ayant une trajectoire représentative d'un groupe de sujets. De plus, lorsque deux trajectoires sont très proches, la distance euclidienne et la distance généralisée de Fréchet se rapprochent, tout comme les moyennes de ces deux distances. Cela signifie que le calcul des groupes s'effectue, en terme de temps de calcul et de complexité, comme dans le cas des *k-means* longitudinaux. En ce qui concerne la réduction du jeu de données par la réduction du nombre de points de mesures, un constat similaire s'observe : le résultat obtenu est identique mais le temps de calcul est réduit [14].

Finalement, un autre inconvénient, indépendant de la taille du jeu de données, est qu'aucun critère de sélection du nombre optimal de groupes n'est vraiment pertinent [14]. Les critères usuels, tel que celui de Calinski et Harabazt [199] évoqué ci-dessus, s'appuient sur des distances classiques et n'ont pas été adaptés à la distance et à la moyenne de Fréchet. D'autant plus qu'en pratique, la réalité clinique des groupes obtenus est privilégiée [14]. En conséquence, il n'est pas utile de relancer plusieurs fois l'algorithme.

1.1.2 L'algorithme de Classification Ascendante Hiérarchique

Parmi les autres méthodes de *clustering* fonctionnel non supervisé, la classification ascendante hiérarchique (CAH) a également retenu notre attention [15, 188, 189, 205, 206]. La trajectoire est, dans le cas présent, résumée par des paramètres de synthèse (voir section 1 du Chapitre 4). De façon automatique, non supervisée et non paramétrique, elle consiste à diviser une population donnée en différentes classes ou sous-groupes. L'objectif est de regrouper un nombre n d'individus en un certain nombre k de groupes (*clusters*) homogènes à partir de p variables quantitatives [207]. Tout comme pour la méthode des *k-means* présentée en section 1.1.1, il est possible de se servir de variables qualitatives, mais ce ne sera pas notre cas ici. Les groupes obtenus par la CAH sont construits selon un critère de ressemblance (ou similarité), ce qui signifie que les sujets d'un même groupe possèdent des traits de caractère communs [206].

Initialement, la CAH considère comme unité de base l'ensemble des individus [15, 205]. Autrement dit, chaque sujet constitue son propre groupe de taille 1 (singleton) au début

de l'algorithme. Ensuite, selon des critères de dissimilarité (distance) [193, 206], les sujets sont agrégés deux à deux. C'est en cela que la méthode doit son nom d'ascendante. Plus les individus sont similaires, plus leur distance devient petite [15]. Les distances entre les nouveaux groupes formés et les sujets restants sont recalculées. De façon itérative, des groupes contenant de plus en plus de sujets se forment, par fusion des plus proches [15], avec calcul mis à jour à chaque étape des distances entre les groupes [207]. L'algorithme de classification se termine lorsque tous les individus sont agrégés en un unique groupe et que la structure de classification peut être définie [15] (voir l'algorithme simplifié 4). Elle est qualifiée de hiérarchique car elle fournit des classes de grande taille contenant elles-mêmes d'autres classes formées aux étapes précédentes [15].

Algorithme simplifié 4 :

Choix a priori :

1. distance évaluant la dissimilarité/similarité entre les individus (distance inter-individus)
2. critère d'agglomération (distance entre les groupes d'individus)

Initialisation : chaque individu constitue son propre groupe
Tant Que une seul groupe n'est pas formé **Faire** :
— Calcul des dissimilarités entre classes
— Recherche du minimum des dissimilarités
— Fusion des 2 classes les plus proches
Fin Tant Que
Partitionnement final : construction d'un arbre de classification (dendrogramme)

Comme pour la méthode des *k-means* présentée en section 1.1.1, plusieurs distances sont disponibles (euclidienne, Manhattan, Tchebychev, Mahalanobis, etc.) [193, 207]. La distance utilisée par défaut par la fonction `hclust` du package R `stats` pour l'analyse de données quantitatives est la distance euclidienne. Nous avons conservé le choix de cette distance par défaut pour les mêmes raisons que celles évoquées précédemment. De plus, tout comme pour la méthode des *k-means*, il est important de souligner que le choix de la distance ne modifie pas le fonctionnement de l'algorithme de la CAH.

À chaque étape de l'algorithme, la variabilité est mesurée par l'inertie I [193]. D'après le théorème de Huygens, pour une partition de n individus en K classes $C_1, ..., C_K$, avec d la distance euclidienne, on a :

$$\sum_{i=1}^{n} d^2(u_i, G) = \left[\sum_{k=1}^{K} \sum_{i \in C_k} d^2(i, G_k)\right] + \left[\sum_{k=1}^{K} n_k d^2(G_k, G)\right] \quad ,$$

où G est le centre d'inertie de tous les individus, G_k le centre d'inertie de la classe numéro k et n_k le nombre d'individus de la classe k. En d'autres termes, en notant I l'inertie, cette décomposition peut se lire comme [193] :

$$I(totale) = I(intra - classes) + I(inter - classes) \quad .$$

En conséquence, minimiser l'inertie intra-classes revient à maximiser l'inertie inter-classes, puisque l'inertie totale est constante pour un même jeu de données. Plus le nombre de classes augmente, plus la variabilité intra-classes diminue.

Lorsque la ressemblance entre les groupes d'individus est mesurée par un critère d'inertie (critère de partitionnement) et que la matrice de dissimilarité est calculée à partir de la

distance euclidienne, la méthode de CAH porte le nom de méthode d'agrégation de Ward [193, 208, 207]. Les sujets sont agrégés de manière à optimiser la qualité de la partition finale. À chaque pas de l'algorithme, l'indice de dissimilarité est égal au gain d'inertie inter-groupes résultant de la précédente agrégation [193]. L'objectif d'utiliser simultanément la distance euclidienne et le critère d'inertie est de réduire la perte d'inertie inter-groupe (à un nombre k donné de groupes, il reste $k(k-1)/2$ possibilités à examiner pour réduire les groupes à k-1 [199]).

Une fois l'algorithme appliqué aux données, un mode de représentation synthétique du résultat final consiste à construire un arbre hiérarchique ou dendrogramme [193] (voir Figure 5.4 dont chaque branche rassemble des sujets constituant un groupe [15]). Cet arbre binaire permet de visualiser les parentés entre les individus [193]. En effet, deux individus sont d'autant plus ressemblants que, pour aller de l'un à l'autre, il n'est pas nécessaire de remonter haut dans l'arbre. L'avantage est que l'arbre peut être tronqué à différentes hauteurs pour en visualiser les *clusters* obtenus après coupure de ses branches [193].

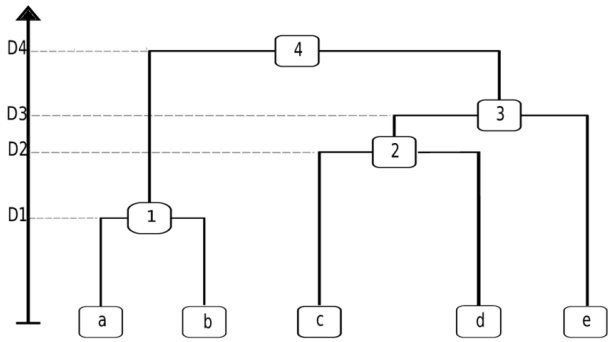

FIGURE 5.4 – Exemple de création d'un dendrogramme : les individus, représentés par les feuilles, sont indiqués par les lettres a, b, c, d et e. Ils sont successivement regroupées en *clusters* dont les différents nœuds sont numérotés de 1 à 4 [**?**].

Le choix du partitionnement, défini par la troncature de l'arbre hiérarchique, peut dépendre du domaine d'application (connaissance a priori du nombre de groupes souhaités) ou peut se baser sur des critères spécifiques (critère du coude, défini ci-après). Plus l'arbre est coupé vers les feuilles (les singletons des individus étudiés), plus la classification est fine et homogène. En effet, le partitionnement obtenu satisfait ces propriétés [15, 193] :

1. Toute classe est non vide ;
2. Tout sujet n'appartient qu'à une classe ;
3. À l'intérieur de chaque classe, les sujets sont homogènes (faible variabilité intra-groupe) ;
4. D'une classe à l'autre, les individus sont hétérogènes (forte variabilité inter-groupe).

Ainsi, si l'arbre est coupé trop haut, les individus auront peu de caractéristiques communes en ce sens qu'ils seront proches de la population totale. Si l'arbre est coupé trop bas, c'est l'effet inverse : les individus sont tellement différents qu'ils forment chacun leur propre groupe. L'arbre doit ainsi être tronqué de façon pertinente pour comparer les groupes obtenus. Graphiquement, pour choisir la partition optimale du dendrogramme, nous représentons l'écart d'inertie selon le nombre de groupes possibles [193]. Habituellement, pour assurer la qualité et l'optimalité de la partition finale, nous choisissons la partition avec la plus grande perte relative d'inertie (le plus grand saut) [194] (voir l'algorithme simplifié 4). La fonction `best.cutree` de l'extension `JLutils` du package `devtools` de R est implémentée pour aider l'utilisateur à faire ce choix à partir du critère du coude (plus grand saut d'inertie) [194, 209].

L'avantage majeur de cette méthode, par rapport à celle des *k-means* de la section 1.1.1, est que le nombre de groupes se choisit a posteriori [193]. Aucune hypothèse sur le nombre de *clusters* ou leur distribution n'est requise. Un autre avantage est sa reproductibilité. En effet, toute transformation monotone conduira au même arbre final [15]. Il n'est donc pas nécessaire de relancer plusieurs fois l'algorithme. Son principal inconvénient est que les variables servant à construire l'arbre hiérarchique sont considérées comme indépendantes [193]. Si tel n'est pas le cas, il est recommandé d'effectuer une analyse en composantes principales pour ensuite utiliser les premiers axes factoriels pour appliquer la CAH [31, 193]. L'objectif de cette manipulation est de transformer les données en des variables synthétiques et indépendantes afin d'appliquer la CAH. Cette technique n'a pas été utilisée par la suite car nous avons utilisé des paramètres de synthèse (voir section 1 du Chapitre 4).

Dans un contexte de données fonctionnelles, la CAH ne peut être directement appliquée. En effet, en plus de ne pas tenir compte de la dépendance entre les données, la méthode ne prends pas en compte le temps écoulé entre deux mesures. Pour rendre cette méthode pertinente au sein de notre stratégie de ré-analyse, un indicateur de variations de trajectoires doit être défini (voir section 1 du Chapitre 4). Divers indicateurs peuvent être envisagés, comme des coefficients de variation de la trajectoire (voir section 1 du Chapitre 4). Nous avons choisi, comme expliqué dans le Chapitre 4, de considérer le taux d'accroissement entre deux visites comme paramètre de synthèse de la donnée fonctionnelle.

En somme, la méthode présentée ici est particulièrement intéressante à appliquer lorsque nous n'avons pas d'a priori sur le phénomène étudié [193]. Elle diffère de la méthode des *k-means* présentée en section 1.1.1 en cinq points. Tout d'abord, le nombre k de *clusters* n'est pas défini a priori mais une fois la classification terminée, par l'utilisation du critère du coude. Ensuite, les centroïdes des *clusters* ne sont pas tirés aléatoirement mais définis selon l'unité de base qui est l'individu lui-même. La CAH n'a donc pas besoin d'une initialisation de l'algorithme par tirage aléatoire. Troisièmement, les groupes formés par la méthode des *k-means* incluent simplement des sujets proches selon le critère choisi, il n'y a pas de notion de hiérarchie des groupes. Ensuite, l'avantage pratique de la CAH est que ses résultats sont plus stables, en ce sens que relancer l'algorithme sur un même jeu de données, avec le même critère et la même distance, conduit à un résultat identique. Le dernier point souligne que la CAH n'est pas conçue pour traiter directement avec des données longitudinales fonctionnelles, d'où l'utilisation de l'indicateur de variation de trajectoire.

1.1.3 La méthode de sémiologie graphique

La dernière méthode d'analyse de données fonctionnelles abordée dans ce travail de recherche est la méthode de sémiologie graphique [16, 210, 211]. Cette méthode d'analyse combinatoire est connue sous le nom de graphiques de dissimilarités ou de sériation. De façon automatique, non supervisée et non paramétrique, elle utilise des techniques de séquençage pour réorganiser un jeu de données ou un dendrogramme (voir section 1.1.2). La sériation fait l'hypothèse de l'existence d'une structure inconnue entre les individus [210]. Elle correspond à une méthode de visualisation des données. Elle est parfois une alternative à la méthode présentée en section 1.1.2 [210]. Il y a quelques années, un logiciel effectuait ces manipulations [212] mais à présent, la méthode a été adaptée pour le logiciel R [16].

Pour sérier un tableau de données, tout comme pour les précédentes méthodes de *clustering* fonctionnel, il faut définir une matrice de dissimilarité (notée D) [16]. Cette dernière contient les distances entre les sujets. Les distances disponibles pour calculer les

similarités sont toujours les mêmes que celles décrites pour les *k-means* en section 1.1.1 ou la CAH en section 1.1.2. La distance usuelle reste, pour les mêmes raisons que précédemment, la distance euclidienne [16, 210].

Initialement, la méthode recode cette matrice D en couleurs [210, 211] (voir l'algorithme simplifié 5). Selon le nombre k de couleurs choisies avant de commencer, l'ensemble des valeurs de la matrice de dissimilarités est divisées en k classes. Chaque classe de valeurs se verra attribuer un carré de couleur [210], ou un carré avec une intensité de couleur si une échelle monochrome est choisie [212]. Ce principe reste similaire à celui utilisé usuellement pour les graphiques de matrices de corrélation [213, 214]. Dans un deuxième temps, la réorganisation linéaire a lieu. Les lignes et/ou les colonnes de la matrice D sont échangées pour mettre en évidence la structure des données avec les couleurs choisies [210]. Une méthode différente peut être utilisée pour les lignes et les colonnes pour plus de flexibilité [16]. Trouver l'organisation optimale implique d'évaluer toutes les solutions possibles. Visuellement, la permutation révèle des blocs de couleurs, assimilés aux *clusters* des sections 1.1.1 et 1.1.2. La similarité entre les différents *clusters* est reflétée par leur position les uns par rapport aux autres [210]. C'est en cela que la méthode fournit une représentation et un support visuel de la tendance évolutive des différents groupes, de façon globale mais également au sein de chaque groupe.

Algorithme simplifié 5 :
Choix a priori :
1. donnée = tableau de données
2. k un nombre de classes
3. distance permettant d'évaluer la dissimilarité/similarité entre les individus
4. fonction(s) de permutation des lignes et/ou des colonnes

Initialisation : diviser les valeurs de la donnée en k classes
Processus
— Attribuer une couleur à chaque classe de valeurs
— Permuter les lignes et/ou les colonnes selon la (les) fonction(s) choisie(s)
— Ajuster le nombre de classes si besoin et réitérer les étapes précédentes
Fin Processus Partitionnement final : graphique montrant la structure des données mise en évidence.

Plus formellement, notons ψ une permutation réorganisant les éléments de D en réordonnant simultanément ses lignes et ses colonnes [16, 210]. L'objectif de la sériation est de trouver une fonction de permutation ψ^+ (respectivement ψ^-) optimisant la valeur d'une fonction de perte notée L (resp. de mérite notée M), ce qui se traduit par [16] :

$$\psi^+ = \underset{\psi}{\operatorname{argmin}} L(\psi(D)) \quad ou \quad \psi^- = \underset{\psi}{\operatorname{argmax}} M(\psi(D)) \ .$$

Une matrice de dissimilarité symétrique représente un tableau de données (*two-way*) mais un seul ensemble d'individus (*one mode*) [16]. Ce cas nous intéresse particulièrement dans le cadre de cette thèse. La méthode peut aussi se généraliser à plusieurs dimensions (*k-way*) et plusieurs jeux de données (*k mode*) à réorganiser simultanément.

Au total, une trentaine de fonctions de permutations différentes sont proposées par le package `seriation` de R. Les fonctions de perte/mérite visent principalement à réduire le temps de calcul en fonction du problème d'intérêt. Une liste de ces fonctions est largement

présentée dans l'article de Hahsler et al. [16]. Les auteurs décrivent comment chaque critère est défini. Des tableaux récapitulatifs sont proposés, mentionnant également la catégorie du critère (perte ou mérite).

L'avantage majeur de cette méthode est que le nombre de couleurs s'adapte au nombre de groupes. Cela signifie que le modifier en fonction des groupes observés aide à mieux les définir. De plus, l'utilisateur peut constater de lui-même si ce nombre est correctement spécifié par rapport aux connaissances a priori du phénomène [210]. Par exemple, si les groupes demandés montrent une sous-structure particulière, cela peut signifier que le nombre de groupes souhaité a priori n'est pas adapté et devrait être augmenté. À l'inverse, si quelques sous-groupes sont représentés par trop de couleurs, elles ne seront pas distinctes les unes des autres, rendant le *clustering* peu visible. Dans ce cas, il est recommandé de réduire le nombre de groupes [210]. Le principal inconvénient de cette méthode est d'être utilisable seulement en cas d'un nombre réduit de facteurs à prendre en compte. En effet, de par sa nature combinatoire, le nombre de solutions augmente factoriellement en fonction de la taille de la population d'étude [210].

L'une des spécificités de l'utilisation de cette méthode dans le cadre de données longitudinales est que, à l'instar de la CAH présentée en section 1.1.2, un indicateur de variation de trajectoires doit être défini. La méthode n'est pas construite pour traiter directement des données fonctionnelles mais peut s'adapter aux données répétées dans le temps par le choix d'un paramètre de synthèse comme le taux d'accroissement (voir section 1 du Chapitre 4). L'autre point commun avec la CAH de la section 1.1.2 et les *k-means* de la section 1.1.1 réside dans le choix d'une distance pour construire la matrice de dissimilarité. Par contre, tout comme les *k-means* longitudinaux, cette méthode requiert de préciser le nombre de couleurs souhaité avant de commencer l'analyse. Elle émet donc un a priori sur le nombre de groupes à mettre en évidence.

1.2 La classification fonctionnelle pour la recherche de groupes répondeurs

Parmi les méthodes d'analyses de données fonctionnelles, celle de la classification supervisée selon un critère clinique a retenu notre attention. Cette classification n'utilise pas de méthode a proprement parlé mais s'appuie sur une classification dans différents groupes. Ces derniers étant connus a priori, la présente stratégie ressemble fortement à la recherche des profils extrêmes présentée en section 3.1 Chapitre 4. En traitant indirectement les données fonctionnelles partiellement observées, elle vise à créer des groupes selon un critère défini a priori. La donnée fonctionnelle n'est utilisée que pour la construction du critère de groupe. Dans notre cas, ce dernier correspond au statut de répondeur à l'intervention d'intérêt. Cela implique que, contrairement aux trois méthodes non supervisées précédentes, l'analyse peut avoir pour objectif d'expliquer les groupes de répondeurs. Il y a donc une variable à expliquer, le groupe répondeur, et des variables explicatives, les caractéristiques cliniques des sujets. Le point clé de cette analyse réside dans la définition de ce qu'est un répondeur.

Cette approche supervisée ne peut être mise en place qu'avec l'aide des connaissances du domaine d'application pour définir les critères cliniques applicables, les tailles des groupes souhaités, etc. En effet, une nouvelle variable catégorielle est construite d'après ces informations pour créer les groupes de répondeur ou non répondeurs. Chaque modalité de cette

nouvelle variable, construite dans notre cas par rapport à un seuil de valeurs du Z-score composite sur l'ensemble de ses mesures, représente un groupe de sujets pouvant être analysé seul ou comparé aux autres par des méthodes d'analyse usuelles. Les connaissances cliniques du domaine aident à choisir un seuil raisonnable et non arbitraire.

Ensuite, cette variable de groupes étant construite, deux stratégies d'analyses sont envisageables. La première consiste à décrire avec des tests usuels la population d'étude et les différents groupes formés par cette variable catégorielle. La seconde utilise des méthodes d'analyse telles que la régression logistique (simple ou multinomiale) ou des algorithmes d'apprentissage statistique classiques pour expliquer les groupes formés par la variable catégorielle. Pour notre étude, la première stratégie de description a été employée car notre objectif était, après celui de définir des sous-groupes de sujets homogènes, de quantifier le nombre de sujets appartenant à chaque groupe. Nous avons ainsi comparé les groupes entre eux, sans étudier les facteurs associés à la réponse à l'intervention.

2 Techniques de modélisation des trajectoires de fonctions cognitives

Dans un second temps, nous avons porté notre attention sur la modélisation des trajectoires de fonctions cognitives (voir Figure 5.1). En effet, pour ré-analyser les données d'un essai de prévention de la maladie d'Alzheimer, dont le critère de jugement principal est un Z-score composite longitudinal, deux stratégies ont été adoptées. La première concerne une méthode particulière d'apprentissage par renforcement détaillée en section 2.1. Elle traite directement avec les données répétées dans le temps. La seconde, regroupant les sections 2.2.1 et 2.2.2, porte sur les approches markoviennes. Dans ce cas, les trajectoires cognitives sont décrites par des paramètres de synthèses, ici une succession d'états.

2.1 Méthode d'apprentissage par renforcement

La première stratégie de cette section 2, repose sur l'apprentissage par renforcement. En effet, rappelons que les traitements médicaux résultent en pratique d'expériences, de telle sorte que si la dose est trop importante, le clinicien la diminue ; à l'inverse si la dose est trop faible, il l'augmente. Après quelques mois de traitement, un nouveau bilan a lieu pour modifier si besoin le traitement. Le clinicien apprends donc des résultats passés.

Dans le cas de notre étude, nous voulions utiliser l'apprentissage par renforcement pour modéliser le niveau cognitif global des sujets. Cela signifie que nous souhaitons apprendre la meilleure stratégie de traitement à adopter pour un sujet donné à partir de ses résultats cognitifs passés de façon à limiter la perte du niveau cognitif global. Des méthodes classiques telles que l'analyse de la variance, l'analyse factorielle discriminante, les arbres de régression et de classification ainsi que les forêts aléatoires ont été appliquées dans un premier temps. L'objectif de ces dernières était de rechercher des groupes selon le bras d'intervention ou de traitement de l'essai MAPT. Elles n'utilisaient donc pas la donnée fonctionnelle comme critère de jugement mais le bras de traitement de l'essai. Étant connues, ces quatre méthodes ne seront pas présentées dans ce chapitre mais les résultats se trouvent dans le Chapitre 6.

En revanche, la méthode du *Q-leaning* est expliquée en section 2.1.2. Pour cela, il est

nécessaire de définir préalablement ce qu'est une règle de décisions. La section 2.1.1 détaille la façon de segmenter les données longitudinales fonctionnelles de façon à définir une règle de décision. La section 2.1.2 explique comment le *Q-learning*, algorithme particulier d'apprentissage par renforcement, optimise cette règle de décision au cours du temps. L'avantage d'une telle méthode, en sus de traiter directement des données longitudinales et d'optimiser le résultat final, est qu'elle tient compte de particularités de la population d'étude visée par les essais de prévention. En effet, la section 4.1 du Chapitre 2 indiquait que les sujets de ces études présentent un historique personnel de nombreuses comorbidités. En utilisant des règles de décisions, cet historique est segmenté selon les différentes étapes, varie au cours du temps et est considéré comme une covariable explicative du modèle de chaque étape.

2.1.1 Les règles de décision

De façon schématique, les parcours de soins se résument en quelques grandes étapes. La Figure 5.5 présente le schéma d'une règle de décisions en deux étapes [7, 215]. L'axe des abscisses de la Figure 5.5 indique le temps de suivi d'un sujet lors d'une étude observationnelle longitudinale. La composante X_1 correspond aux covariables observées avant A_1. Par exemple, elle contient le sexe, l'âge, les antécédents médicaux, les traitements en cours d'un patient. À un moment précis, le clinicien choisit de prescrire un traitement parmi ceux autorisés dans l'essai. Il prend une première décision médicale A_1. Dans le cas d'un essai clinique, la décision A_1 est généralement la randomisation dans les bras de l'essai. À l'étape 1 de l'étude, l'histoire du patient H_1 est définie comme n'étant que X_1. Puis, le temps passe jusqu'à une deuxième prise de décision A_2. Entre A_1 et A_2 se trouvent potentiellement de nouvelles données X_2, faisant référence aux observations pré-traitement de l'étape 2 de l'étude. L'histoire du patient à l'étape 2 se définit comme étant $H_2 = (X_1, A_1, X_2)$. Le résultat obtenu Y à un instant donné constitue l'objet d'intérêt, ainsi que la règle de décision $d = (d_1, d_2)$ conduisant à ce résultat, avec $d_j : H_j \to A_j$, j $\in \{1, 2\}$. La règle de décision contient donc l'historique du sujet jusqu'en fin de suivi, c'est-à-dire ses données pré-traitement, ses variables entre les décisions et les décisions dans l'ordre chronologique. Ainsi, connaissant le parcours de soin d'un patient, l'objectif est d'optimiser les règles de décision menant à Y. Pour simplifier la théorie de la méthode, nous avons considéré qu'un parcours de soin ne pouvait être décrit que par deux prises de décision au cours du suivi, comme représenté sur la Figure 5.5. Ainsi, seules les règles à 2 décisions nous intéressent ici, bien que la méthode soit généralisable à des règles de décisions à n étapes.

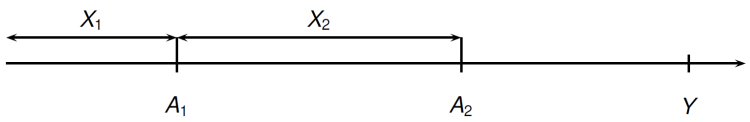

FIGURE 5.5 – Schéma des règles de décisions pour un patient donné, dans le cas de deux interventions médicales.

Pour obtenir ce résultat optimal, il est nécessaire d'estimer tous les résultats potentiels [7], c'est-à-dire ceux obtenus si telle règle de décision avait été suivie. En effet, une seule séquence est observée par sujet, les autres ne sont qu'hypothétiques. Le nombre de résultats possibles augmente en fonction du nombre de décisions. Ensuite, la prise de décision optimale en résulte. Dit autrement, les différentes règles de décision sont comparées entre elles pour

considérer la séquence de traitement ayant la plus grande valeur. Parmi les techniques d'estimation des règles de décision optimales, l'apprentissage par renforcement du *Q-learning* est présenté [216] (voir section 2.1.2 ci-dessous).

2.1.2 Cas particulier du *Q-learning*

Dans le cas d'un parcours de soin en deux décisions, les données longitudinales d'un patient sont déterminées par (X_1, A_1, X_2, A_2, Y). L'histoire est donnée par $H_1 \equiv X_1$ à l'étape 1 et $H_2 \equiv (X_1, A_1, X_2)$ à l'étape 2 de l'étude. Les données disponibles pour l'estimation des règles de décision optimales sont issues d'un échantillon aléatoire de N sujets.

Pour présenter la méthode, considérons qu'il n'y ait que deux traitements possibles à chaque étape. Cela signifie que $A_j \in \{-1, 1\}$, $j \in \{1, 2\}$ [7]. Ces traitements sont randomisés, conditionnellement à l'histoire du patient. L'étude peut avoir soit une seule récompense terminale (résultat principal Y), observée à la fin de l'étape 2 ; soit deux récompenses intermédiaires Y_1, Y_2 et le résultat final Y, observés après chaque étape. Le mot "récompense" est un terme spécifique à la théorie de l'apprentissage par renforcement, mais il correspond aussi bien à une issus positive que négative. En somme, Y est un cas particulier de la seconde configuration, avec $Y_1 = 0$ et $Y_2 = Y - Y_1$. Les deux règles de décision sont (d_1, d_2), avec $d_j(H_j) \in \{-1, 1\}$, $j \in \{1, 2\}$ [217]. Le *Q-learning* vise à déterminer la règle optimale $d^{opt} = (d_1^{opt}, d_2^{opt})$. Sa particularité réside dans un raisonnement en sens inverse du temps (*backward*) [7]. En effet, une régression linéaire sur la seconde étape d'intervention de l'étude est d'abord modélisée estimant le résultat potentiel de l'étape 2. Puis, avec une nouvelle régression à l'étape 1, la règle de décision optimale est obtenue. Le processus est expliqué plus en détails dans le paragraphe suivant. Il devient fondamental de souligner que le résultat Y de notre étude est de type quantitatif et représente le niveau cognitif global d'un sujet.

Plus formellement, puisque nous utilisons un algorithme d'apprentissage par renforcement, nous devons utiliser des fonctions de valeur. Dans notre cas, elles correspondent à des fonctions de qualité ou Q-fonctions. L'objectif est d'optimiser, à chaque étape, ces Q-fonctions. La Q-fonction à l'étape 2 est définie par [7] :

$$Q_2^{opt}(H_2, A_2) = \mathbb{E}\left[Y | H_2, A_2\right] \quad .$$

Cette espérance conditionnelle est modélisée par une régression linéaire. Quant à la fonction de qualité à l'étape 1, il est usuel de considérer le résultat potentiel $Y_1^* = \max_{a_2} Q_2^{opt}(H_2, a_2)$ maximisant l'étape 2. La Q-fonction à l'étape 1 se définit ainsi par :

$$Q_1^{opt}(H_1, A_1) = \mathbb{E}\left[Y_1^* | H_1, A_1\right] \ .$$

À l'instar de la fonction de qualité à l'étape 2, une régression linéaire modélise également cette espérance conditionnelle. Lorsque ces deux quantités sont connues, la règle de décision optimale découle de la relation suivante :

$$d_j^{opt}(h_j) = \arg\max_{a_j} Q_j^{opt}(h_j, a_j) \ , \ j \in \{1, 2\} \ .$$

En pratique, les Q-fonctions sont inconnues et doivent être estimées à partir des données observées. En considérant les modèles de régression linéaire, pour $j \in \{1, 2\}$, la règle de décision optimale Q_j^{opt} découle de la relation suivante :

$$Q_j^{opt}(H_j, A_j; \beta_j, \phi_j) = \beta_j^T H_{1j} + (\phi_j^T H_{2j}) A_j \ ,$$

avec H_{1j} et H_{2j} les deux vecteurs résumant les informations de l'histoire H_j du sujet j, aux étapes 1 et 2 de la règle de décision. En conséquence, H_{1j} correspond à l'effet principal de l'histoire (terme comprenant l'intercept de la régression) et H_{2j} désigne l'effet du traitement sur l'histoire (contenant aussi l'intercept de H_{1j}) ; β_j^T la transposée du vecteur des coefficients de régression. Les variables de H_{1j} sont qualifiées de prédictives, car elles sont utilisées pour accroître la précision des estimations, tandis que celles de H_{2j} sont dites prescriptives car elles permettent d'adapter les stratégies de traitement aux patients (variables d'adaptation [7]).

Pour résumer, l'algorithme du *Q-learning* consiste à [7] :

1. Étape 2 de la régression : $(\hat{\beta}_2, \hat{\phi}_2) = \arg\min_{\beta_2,\phi_2} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - Q_2^{opt}(H_{2i}, A_{2i}; \beta_2, \phi_2)\right)^2$;
2. Étape 1 résultat potentiel : $\hat{Y}_i^* = \max_{a_2} Q_2^{opt}(h_2, a_2; \hat{\beta}_2, \hat{\phi}_2)$, $i \in \{1, n\}$, n étant le nombre d'observations considérées ;
3. Étape 1 de la régression : $(\hat{\beta}_1, \hat{\phi}_1) = \arg\min_{\beta_1,\phi_1} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{Y}_i^* - Q_1^{opt}(H_{1i}, A_{1i}; \beta_1, \phi_1)\right)^2$.

Notons que, pour la seconde étape de l'algorithme du *Q-learning*, $\hat{Y}_i^*$ est un prédicteur de la variable aléatoire non observée $Y_i - Q_2^{opt}(H_{2i}, A_{2i}; \beta_2, \phi_2)$, avec $i \in \{1, ..., n\}$. Une fois les Q-fonctions estimées, trouver la valeur du résultat potentiel optimal s'obtient en optimisant les deux Q-fonctions. Effectivement, à l'étape j la règle de décision optimale vaut :

$$\hat{d}_j^{opt}(h_j) = \arg\max_{a_j} Q_j^{opt}(h_j, a_j; \hat{\beta}_j, \hat{\phi}_j) \quad , \; j \in \{1, 2\} \; .$$

Cette procédure indirecte en deux étapes peut paraître étrange de prime abord. Une procédure classique en une étape pour estimer le résultat optimal semble plus naturelle. Dans cette approche, nous pourrions modéliser l'espérance conditionnelle du résultat final Y par $\mathbb{E}(Y|X_1, A_1, X_2, A_2)$, puis exécuter une régression linéaire multivariée. La politique optimale serait estimée par :

$$(d_1^{opt}, d_2^{opt}) = \arg\max_{(a_1,a_2)} \mathbb{E}(Y|x_1, a_1, x_2, a_2) \; .$$

Seulement, la régression linéaire n'est pas utilisable en présence de biais potentiels provenant des études longitudinales, comme par exemple le biais de classement, car elle n'en tient pas compte. Ce dernier désigne une erreur systématique de mesure de l'exposition ou de l'évènement médical (souvent une maladie), provoquant une erreur de classement des sujets en malades/non malades ou en exposés/non exposés. De fait, le *Q-learning* est préférable à ces méthodes afin d'obtenir des résultats non biaisés. Il tient également compte des effets à retardement d'un traitement par la considération de l'histoire du sujet à chaque étape, avantage non négligeable pour l'analyse d'un parcours de soin.

2.2 Méthodes markoviennes

Dans cette seconde partie de la section 2 de la modélisation de trajectoires, l'évolution des fonctions cognitives n'est plus abordée comme un phénomène général, accompagnée d'une donnée fonctionnelle des fonctions cognitives. En effet, pour comprendre certains comportement cognitifs, une stratégie consiste à assimiler l'activité globale à différents états [18, 218, 219, 220]. Par exemple, ces états peuvent refléter l'apparition de symptômes cliniques au cours du suivi, des stades de la maladie ou des niveaux de marqueurs quantitatifs

[18, 218]. Leur signification dépend du phénomène étudié. Aussi, à chaque mesure de la donnée fonctionnelle partiellement observée est associé non plus une valeur quantitative, mais un état associé à un temps passé dans cet état. L'hypothèse est que ces états ont produit le niveau global de fonctions cognitives étudié. La donnée fonctionnelle, partiellement observée en pratique, est donc représentée par cette succession d'états, mesurés de façon répétée dans le temps et synthétisant l'information à un temps donné.

Ce changement de stratégie d'analyse découle de l'hypothèse suivante : nos approches de la section 1, dont objectif principal reposait sur l'exploration de groupes de trajectoires, présentaient l'inconvénient de considérer le déclin comme un phénomène lisse et continu. Or, l'évolution de la maladie d'Alzheimer se caractérise par plusieurs phases : asymptomatique, pré-déclin, déclin avéré (voir section 3 du Chapitre 2). Cet aspect n'est pas pris en compte par les méthodes précédemment utilisées pour identifier des sous-groupes de trajectoires. De plus, comme expliqué en section 3 du Chapitre 2, la date de survenue de la maladie ne peut être établie avec exactitude, puisque le processus de déclin cognitif s'installe progressivement. En conséquence, l'évolution des fonctions cognitives a été observée plus spécifiquement sur différentes périodes de suivi [18, 220]. Dans cette partie, nous considérons maintenant que la trajectoire globale du sujet correspond à un ensemble d'états, ou processus multi-états [19, 221], plutôt qu'à un ensemble de mesures quantitatives. Dans le cadre d'essais de prévention, les états et les covariables sont observés de façon intermittente au cours du suivi et l'information entre les visites est manquante [18, 19]. Pour chaque transition possible entre les différents états, l'objectif consiste à identifier les facteurs de risque affectant ces transitions en modélisant leur relation [19].

Un processus multi-états tel que nous l'avons décrit est stochastique [222]. En ayant fixé à l'avance le nombre d'états, le processus peut être modélisé par des chaînes de Markov [219]. Ces dernières offrent la possibilité d'analyser des séries temporelles tout en tenant compte de leurs contraintes de dépendance au cours du temps [223]. Les modèles de Markov permettent d'étudier différents comportements [224] et sont particulièrement intéressants pour modéliser la progression d'une maladie chronique telle que la maladie d'Alzheimer [221]. Ces modèles reposent sur l'hypothèse que l'information sur les états précédents est résumée par l'état présent (hypothèse de Markov), ce qui signifie que l'évolution du processus ne dépend que de l'état au temps t. Cette hypothèse implique que de tels modèles analysent les séries temporelles en utilisant des structures de dépendance locale plutôt que globale.

Deux stratégies sont appliquées dans cette section : une classique par modèles multi-états de Markov [222] (section 2.2.1) et une complémentaire par modèles de Markov cachés (*Hidden Markov Model* - HMM [20]) présentée en section 2.2.2.
D'une part, les chaînes de Markov classiques sont adaptées à notre objectif [219] et sont décrites en section 2.2.1. Elles modélisent l'évolution d'un processus multi-états dynamique au cours du temps. Les états sont ainsi définis préalablement, avec leur signification. Le processus étant connu, les transitions entre les états sont définies a priori. L'objectif d'un tel modèle est d'étudier les covariables influençant les transitions entre les états. D'après notre contexte, nous ne considérons que des modèles pour lesquels les états sont connus pour une série finie de temps.
D'autre part, les HMM semblent autant être adaptés à notre objectif de modélisation des trajectoires cognitives et sont référencés en section 2.2.2. La différence avec les modèles précédents réside dans la construction du modèle dont une partie est néanmoins commune. La chaîne de Markov n'est pas directement observée, non connue a priori, mais supposée existante et produisant les valeurs observées [20]. Les états sont donc cachés et liés aux

observations par une loi de probabilité. Ces états, mesurant une quantité non directement observée, tient compte d'une incertitude de mesure potentielle. Il faut également faire l'a priori d'un nombre d'états cachés [20]. Ces modèles sont flexibles sur l'intégration de covariables et la caractérisation de séries temporelles. Ils sont tout autant adaptés aux données continues qu'aux discrètes. En pratique, la documentation des HMM référence des études de séquences de certains types de comportements, pour les détecter et les décrire de façon précise [225]. Ils quantifient l'impact de facteurs environnementaux sur ces comportements.

2.2.1 Les modèles de Markov

En 1906, Andrey Andreyevich Markov introduit pour la première fois le concept d'une chaîne de Markov. Cette dernière, aussi appelée modèle de Markov observable, est un processus aléatoire composé de suites de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ décrivant l'évolution d'un système dynamique au cours du temps [220]. Les temps d'observations peuvent être discrets, c'est-à-dire fixés pour tous les individus, ou continus, variables d'un individu à l'autre. Les variables $(X_n)_{n\in\mathbb{N}}$ sont à valeurs dans un ensemble E appelé espace d'états. Ce dernier peut être continu (espace infini) ou discret (espace fini ou dénombrable) [226, 227]. Un modèle ayant un espace discret et un temps continu se qualifie de modèle multi-état de Markov [222]. C'est sur ce dernier que nous focaliserons notre attention dans cette section, car les temps d'observation des sujets de notre étude ne sont pas fixés à l'avance [220].

Au cours du temps t, le processus aléatoire évolue et décrit les mouvements entre les états possibles de E [228]. Ces mouvements portent le nom de transitions. Une intensité de transition en gouverne chacune d'elle [18, 220] et peut être représentée par un arbre déterministe de toutes les réalisations possibles. Graphiquement, le processus se représente à l'aide de cercles pour les états et de flèches unidirectionnelles indiquant les transitions possibles [229], à l'instar du schéma proposé en Figure 5.6. D'autres exemples de modèles types sont présentés dans l'exemple 1. À chaque temps, le processus peut passer d'un état à un autre ou rester dans son état actuel, avec une certaine probabilité [220, 229]. Les états sont alors qualifiés de transitoires lorsqu'ils sont parcourus au cours du processus et d'absorbants lorsque la chaîne ne peut plus changer d'état, c'est-à-dire lorsque la seule transition possible est de l'état vers lui-même [221].

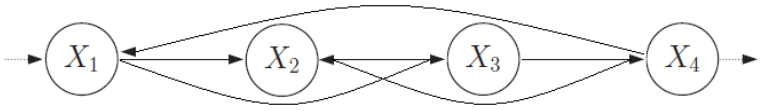

FIGURE 5.6 – Graphique dirigé d'un modèle de Markov observable, où X_t les états et les flèchent indiquent les transitions possibles entre les états.

Exemple 1 :
En considérant un modèle multi-états à temps continu avec 2 états et une seule transition possible [229], nous pouvons nous rapporter à un modèle de survie classique [219, 222]. Les deux états correspondent à l'observation de l'évènement (exemple, décès) et de la censure du temps d'évènement (exemple, vivant) [229]. De la même façon, les modèles à risques compétitifs ou de type sain-malade-décès (*illness-death*) sont modélisables via un modèle de Markov [219, 228, 229], dont les transitions autorisées sont de l'état sain vers malade, malade vers décès et sain vers décès. En fonction des maladies, une transition de malade vers sain est également envisageable.

Les intensités de transition représentent le risque instantané de passer de l'état $S(t) = i$ à l'état $S(t) = j \neq r$ [18, 222, 228] :

$$q_{ij}(t) = \lim_{\delta t \to 0} \frac{P(S(t+\delta t) = j | S(t) = i)}{\delta t} \quad .$$

Dans ce cas, $p_{ij}(t) = P(S(t+\delta t) = j|S(t) = i)$ est nommée probabilité de transition de l'état i au temps t vers l'état j au temps $t + \delta t$ [220, 230]. Ces probabilités sont indépendantes des états parcourus auparavant et dépendantes uniquement de l'état actuel du processus (hypothèse de Markov), ce qui implique que le processus est qualifié de sans mémoire [227]. Ainsi, un processus est qualifié de Markovien lorsque la connaissance de son état actuel est suffisante pour connaître son évolution future [227]. Cela se traduit par : $\forall k \in \mathbb{N}$, $\forall$ $(x_0, \ldots, x_{k+1}) \in E$ tels que $P(X_k = x_k, \ldots, X_0 = x_0) > 0$,

$$P(X_{k+1} = x_{k+1}|X_k = x_k, \ldots, X_0 = x_0) = P(X_{k+1} = x_{k+1}|X_k = x_k) \quad [18].$$

Par définition, les p_{ij} peuvent dépendre du temps $p_{ij}(t)$, de variables $z(t)$ variant au cours du temps ou de variables spécifiques aux individus, indépendantes du temps, expliquant les différences d'évolution du processus [19, 222, 229]. Ces derniers cas impliquent que les intensités de transitions individuelles possèdent une distribution déterminée par les covariables [19]. Lorsque les transitions ne dépendent pas du temps depuis l'entrée dans l'étude [221, 228], c'est-à-dire quand $\forall t \in \mathbb{N}$,

$$P(X_{t+1} = j|X_t = i) = P(X_1 = j|X_0 = i) \quad ,$$

la chaîne est qualifiée d'homogène [18, 227] ou de stationnaire [221]. Si elles ne sont pas constantes au cours du temps, la chaîne est dite non homogène [220, 228].

Les probabilités de transition d'un état à un autre, pour tous les états possibles, sont synthétisées par l'intermédiaire d'une matrice carrée de dimension le nombre d'états et dite « de transition », notée P [18]. De la même façon, la matrice Q synthétise les intensités de transitions. Chaque colonne de celles-ci correspond à un état. Chaque ligne contient, pour un état fixé, la probabilité de passer à un autre état pour la matrice P, l'intensité de transition d'un état à l'autre pour la matrice Q. Par exemple, la première ligne de, première colonne de la matrice P nous indique la probabilité, sachant que le processus est dans l'état 1, de rester dans l'état 1. Le coefficient sur la deuxième ligne, première colonne (p_{21}) renvoie à la probabilité, sachant que le processus était dans l'état 2, de passer à l'état 1. La principale propriété de ces deux matrices est d'être stochastique [20, 227, 231]. Cela signifie que tous leurs coefficients sont positifs ou nuls et leur somme par ligne vaut 1.

Afin de déterminer un modèle de Markov, il faut définir pour le processus concerné [229] :

— Les états possibles (connus a priori) ;
— Les probabilités initiales, c'est-à-dire dans quel état et avec quelles chances le processus commence ;
— La matrice de transition.

Pour estimer les effets de covariables sur les probabilités ou intensités de transition entre les états, la vraisemblance du modèle doit être calculée à partir de la matrice d'intensité de transition $P(t, t+\delta_t)$ [18, 222]. L'entrée (i, j) de $P(t, t+\delta_t)$, notée $p_{ij}(t, t+\delta_t)$, correspond à la probabilité d'être dans l'état j au temps $t+\delta_t$, sachant que le processus était dans l'état i au temps t [18]. Cette quantité est calculée d'après l'équation différentielle de Chapman-Kolmogorov [228, 232]. Si les intensités de transitions de la matrice Q sont constantes par intervalle $(t, t+\delta_t)$ (cas homogène), alors [18, 220, 222] elles peuvent être calculées, toujours

à partir de l'équation différentielle de Chapman-Kolmogorov, par la matrice M de terme général :

$$M = \exp(tQ) \approx \sum_{r=0}^{\infty} \frac{Q^r t^r}{r!} \quad ,$$

puis évaluées en ij. Dans le cas d'un modèle de Markov homogène, il est possible d'introduire des covariables, considerant un modèle à risques proportionnels de Cox pour les intensités de transitions. Ainsi, ces covariables sont intégrées dans le calcul de l'exponentiel de la formule précédente, à l'instar d'un modèle de survie classique, comme suit [19, 228, 230] :

$$Q(t, Z(t)) = \lambda_0 \exp(\beta^T Z(t)) \quad ,$$

avec λ_0 le risque instantané de base et β^T le vecteur des coefficients de régression [219, 220, 229]. Les paramètres du vecteur β s'obtiennent par maximisation de la vraisemblance [219]. Cette dernière étant le produit des probabilités de transitions, pour tous les individus r observés au temps s supposés distincts, nous déduisons [18, 222, 228] :

$$L(Q) = \prod_{r,s} L_{r,s} = \prod_{r,s} p_{S(t_{r,s}),S(t_{r,s+1})}(t_{r,s+1} - t_{r,s}) \quad .$$

Chaque composante $L_{r,s}$ est un élément de la matrice de transition P, évalué au temps $t = t_{r,s+1} - t_{r,s}$ [222]. Cette vraisemblance est maximisée par l'utilisation de la log-vraisemblance $\log(q_{ij})$ pour estimer les q_{ij} par des algorithmes d'optimisation standard [222]. Les variances sont calculées à partir de la matrice Hessienne de l'optimum.

Concernant la stratégie de sélection des variables pour le modèle, elle est similaire aux analyses de régression ou de survie [222]. Celle que nous avons appliquée aux données est la suivante. Dans un premier temps, le modèle sans covariable (à vide) a été estimé pour servir de comparaison. Ensuite, le seuil de conservation univarié est fixé, généralement à 20 %, pour sélectionner les variables pouvant avoir une influence dans le modèle multivarié. Ce seuil usuel de 20 % a été conservé pour nos analyses. Les interactions médicalement pertinentes entre les covariables ont ensuite été testées pour détecter la présence de facteurs modificateurs d'effet ou de confusion, avec un seuil de 20 % également. La sélection du modèle complet vient ensuite, en partant du modèle vide puis en rajoutant pas à pas des variables (stratégie manuelle ascendante). Trois critères ont été pris en considération pour valider l'ajout des variables : une baisse de l'AIC [31], un test du rapport de vraisemblance significatif au seuil de 5 % [18, 228, 230] et un modèle convergent avec une matrice Hessienne définie positive [222].

Finalement, le dernier point de la stratégie d'analyse consiste à vérifier l'adéquation du modèle multi-états. Pour cela, une approche communément utilisée compare graphiquement les prévalences observées des états avec les prévalences prédites par le modèle [222]. À chaque temps, ces quantités sont disponibles pour tous les sujets présents à ce moment-là [18]. Ces estimations sont soumises à l'hypothèse que la date de la maladie est connue. Pour effectuer une estimation précise, cela suppose que tous les sujets sont observés aux différents temps d'évaluation. Si tel n'est pas le cas, l'estimation suppose implicitement que les transitions n'ont lieux qu'aux moments observés, ce qui signifie que le résultat repose sur une approximation de la prévalence [222]. L'individu non effectivement observé au temps t est considéré comme étant resté dans le même état qu'au moment de sa dernière visite [18]. Plus formellement, il est également possible d'effectuer un test statistique. Le test de Pearson compare ainsi, pour chaque paire d'états possibles, les nombres de transitions observées et attendues. La comparaison s'effectue en chaque temps de suivi, pour chaque catégorie

des variables qualitatives et pour les intervalles de transition. Utilisant une distribution similaire à celle du χ^2, ce test implique de ne pas avoir de faibles effectifs dans son tableau de contingence [222].

À terme, le modèle de Markov aide à estimer la proportion de sujets dans chaque état et l'impact des covariables sur les probablités de transition entre les états. Un modèle Markovien permet également d'estimer le temps moyen passé dans un état donné pour l'ensemble de la population étudiée [18, 222]. Enfin, le temps moyen pour atteindre un état donné, sachant que le processus était dans un état particulier au départ, peut également être obtenu [18, 222]. Notons cependant que ces temps moyens ne dépendent ni de l'ordre des états parcourus, ni de leur nombre dans la séquence d'observation étudiée.

2.2.2 Les modèles de Markov cachés

La seconde stratégie markovienne abordée dans ce chapitre pour modéliser des trajectoires cognitives consiste à utiliser des modèles de Markov cachés (ou *Hidden Markov Model* (HMM)). Le cadre théorique utilise les mêmes concepts que celui des modèles de Markov classiques, présentées en section 2.2.1. En revanche, le modèle précédent était construit au préalable pour amener aux observations, ce qui implique que les états se définissaient a priori tout comme les probabilités initiales et celles de transition. Ces états étaient directement observés. Cette fois-ci, c'est l'inverse : l'objectif est de partir des données pour construire la chaîne de Markov correspondante, dont les états sont cachés. Les vrais états de la chaîne ne sont donc pas observés. En effet, il arrive parfois que l'observation ne soit pas directement issue des états mais soit une fonction des états [20]. Dans ce cas, l'ordre et le nombre d'états dans une séquence donnée peuvent avoir leur importance, en particulier pour l'étude des fonctions cognitives. Les modèles de Markov classiques, qualifiés d'observables puisque les états le sont directement, ne tiennent pas compte de ces points (voir section 2.2.1). De plus, les modèles classiques ne tiennent pas compte de l'erreur de mesure d'une donnée, alors que les HMM peuvent le faire via les états cachés. La question se pose alors de savoir, à partir de trajectoires cognitives individuelles, comment estimer la matrice de transition d'une chaîne de Markov non directement observée [227].

Le modèle de Markov caché correspond à une généralisation du modèle de Markov observable en ce sens qu'il correspond à une fonction aléatoire des états d'une chaîne de Markov [20]. Il appartient à la famille des méthodes d'inférence bayésienne de par ses états cachés [233] et est efficace en modélisation de données de séries temporelles, dont les effectifs sont petits mais les séquences de temps longues ; ou données longitudinales, dont les effectifs sont grands mais les séquences de temps courtes [225, 227, 234, 235]. En particulier, le modèle de Markov caché permet d'observer quelque chose qui n'est pas directement observable, via les états cachés, comme par exemple l'incertitude d'une mesure. Le modèle de Markov caché est reconnu pour être très flexible dans l'intégration de covariables, pour la segmentation des séries temporelles et s'adapte aussi bien aux données discrètes que continues [235].

Pour simplifier la présentation de ces modèles, nous avons divisés la présente section en différentes parties. Dans un premier temps, le cadre théorique est expliqué en section 2.2.2.1. Ensuite, nous présentons les méthodes d'estimation (section 2.2.2.2) et de validation du modèle (section 2.2.2.3). Dans la section 2.2.2.4 est abordé un point important de ce travail de recherche : le choix du nombre d'états cachés régissant le modèle de Markov caché. En dernier lieu, nous présentons en section 2.2.2.5 quelques exemples d'application des HMM présents dans la littérature et dont le cadre est similaire ou proche du notre.

2.2.2.1 Cadre théorique

L'hypothèse associée aux HMM indique que les données réelles $\{X_t : t \in 1, 2, ..., T\}$ sont supposées être issues d'un processus Markovien $\{C_t : t \in 1, 2, ..., T\}$ de paramètres inconnus, ayant sa dynamique propre et dépendant des réponses observées à chaque temps [225, 236]. Le point commun avec la méthode précédente est donc que la propriété de Markov s'applique sur la composante cachée du modèle, cela signifie que le processus aléatoire correspondant n'a pas de mémoire. Cela se traduit par [227, 235] : pour tout $t \in \mathbb{N}$,

$$P(C_t = c_t | C_1 = c_1, \ldots, C_{t-1} = c_{t-1}) = P(C_t = c_t | C_{t-1} = c_{t-1}) \quad .$$

Cette propriété permet de s'affranchir de l'hypothèse d'indépendance des données. Elle est connue sous le nom d'hypothèse de premier ordre de Markov [235]. Une chaîne cachée de second ordre signifierait que l'état au temps t dépend de l'état au temps $t-1$ et au temps $t-2$. Ce processus Markovien C_t ne pouvant être directement observable, d'où son nom de caché [235], il faut se rapporter aux réponses, aux données réelles observables, pour déterminer ses états [20]. Les réponses sont dites émises par les états selon une certaine loi de probabilité [235]. Autrement dit, l'idée de ce type de modèles consiste à considérer que le processus C_t soit caché et, par l'intermédiaire du vecteur d'observations X_t, l'objectif étant de déterminer les états de C_t [227].

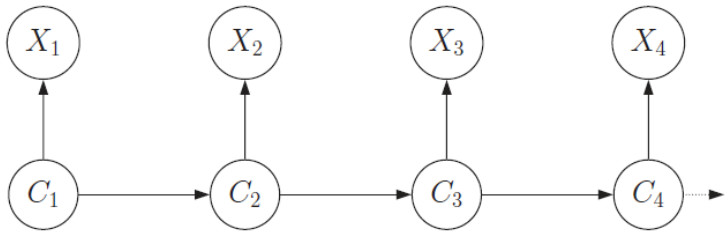

FIGURE 5.7 – Graphique dirigé d'un modèle de Markov caché, avec X_t les observations et C_t les états cachés [225].

En somme, le processus stochastique est double [227, 237] et est représenté en Figure 5.7. D'une part, il contient un processus C_t supposé existant et constitué d'un ensemble d'états, connectés entre eux par des probabilités de transitions inconnues. D'autre part, il est composé d'un processus X_t observable, constitué d'un ensemble de sorties ou observations. Chaque sortie est émise par un état caché selon des probabilités d'émission. Ces dernières sont issues d'une fonction de densité de probabilité [20, 235]. Dans le cas de données continues, les fonctions de densité de probabilité d'émission de chaque état peuvent être déterminées par une distribution de mélange [20]. Ce sera notre cas par la suite. Cette dernière lie les processus entre eux via la distributions de X_t [238], dépendante uniquement de l'état au temps présent C_t et non des états ou observations précédentes. L'hypothèse de Markov présentée en section 2.2.1 reste ainsi vérifiée et permet de tenir compte de la dépendance des données observées au cours du temps. Le processus X_t est un HMM à m-états lorsque la chaîne de Markov C_t est régie par m états. La conséquence immédiate du double processus implique que l'observation d'une séquence n'est plus directement liée à une unique suite d'états [239]. De la même façon, un état peut émettre plus d'une réponse.

Plus formellement [20, 240], une variable aléatoire observée constitue le premier élément d'un HMM. Lorsque cette variable peut être décrite via des modèles de mélanges, cela signifie que la population étudiée est supposée contenir plusieurs sous-populations homogènes [241]. Ils tiennent compte de l'hétérogénéité d'un paramètre au sein d'une population

en divisant une distribution globale en plusieurs sous-distributions correspondant aux différents sous-groupes [242]. Ils estiment, de façon paramétrique ou non, la distribution des variables aléatoires correspondant aux sous-populations. Dans le cas paramétrique, auquel nous avons porté un intérêt particulier, les paramètres de chaque sous-distribution identifiée (moyenne, variance et amplitude) résultent d'un processus stochastique et les observations sont supposées indépendantes [241, 242]. L'objectif d'un modèle de mélange est d'optimiser ces paramètres selon un critère de maximum de vraisemblance pour approcher le plus possible la distribution initiale de la population [242]. Pour un m-HMM, la suite observable $X = (x_1, \ldots, x_T)$ est la réalisation du processus aléatoire X_t résultant de m distributions de mélange [20], où les x_i prennent les valeurs dans l'ensemble V de M classes (les valeurs observées) [235].

Dans le cadre de l'analyse de données longitudinales [233], le processus générant les paramètres de la distribution de mélange au temps t est supposé provenir d'une chaîne de Markov homogène de distribution initiale π et de matrice de transition Γ. Cette chaîne constitue le second élément, la partie cachée, du HMM et c'est à partir d'elle que découlent les paramètres des sous-distributions [20]. Ainsi, la composante cachée du HMM concerne la réalisation $C = (c_1, \ldots, c_T)$ d'une suite de variables aléatoires C_t, où les c_i prennent les valeurs dans l'ensemble E de N classes correspondant aux N états. La dépendance des données au cours du temps est prise en compte par la propriété de Markov de la chaîne [235]. Notons donc que le processus X_t est indépendant conditionnellement à C_t. La distribution de X_t dépend uniquement de l'état C_t, soit le dernier état parcouru, et non des états ou observations précédent(e)s.

Les probabilités de transition d'un état à un autre s'écrivent $p_{i,j} = P(C_{t+1} = j|C_t = i)$, avec i et j les états possibles de la chaîne, t le temps [20, 235]. La chaîne est qualifiée d'homogène lorsque ces probabilités sont indépendantes du temps, comme nous l'avions indiqué en section 2.2.1. La matrice de transition contenant les probabilités de transition $p_{i,j}$ est ici notée $\Gamma(t)$. De la même façon, les probabilités marginales de C_t s'écrivent $u(t) = [P(C_t = 1), ..., P(C_t = m)]$, $t \in \mathbb{N}$ [20]. Ainsi, pour connaître la probabilité d'être dans un état donné au temps t, il suffit de multiplier les probabilités d'être dans les états au temps $t-1$ avec la matrice de transition : $u(t) = u(t-1)\Gamma$ [227, 235]. Avec ces notations, $u(1)$ correspond à la distribution initiale et δ au vecteur des probabilités marginales initiales de la chaîne de Markov. Si $\delta\Gamma = \delta$ et $\delta = u_1(t)$, alors la distribution est stationnaire [227, 235]. Lorsque la distribution initiale de la chaîne de Markov est égale à cette distribution stationnaire, cela implique que la chaîne est elle-même qualifiée de stationnaire. La chaîne possède donc la même distribution en tout temps $\forall t \in \mathbb{N}$, $u(t) = u(1)$, ce qui implique que la probabilité d'être dans un état reste identique au cours du temps.

Le lien entre la chaîne de Markov et le processus générant les observations X_t se fait grâce à la fonction de masse (ou fonction de densité) des probabilités [20, 235] :

$$\pi_i(x) = P(X_t = x|C_t = i) \quad ,$$

avec π_i la fonction de densité de probabilité du vecteur des observations X_t si la chaîne de Markov C_t est dans l'état i au temps t. Les π_i correspondent alors aux m-distributions état-dépendants du HMM. Dans notre cas, ces distributions proviennent de modèles de mélange.

En conséquence, les paramètres du modèle à estimer sont [20] :
— Les paramètres des sous-distributions, regroupés dans $\mathbf{P}$,
— Les probabilités initiales de la chaîne δ,

— Les probabilités de transition $p_{i,j}$ réunies dans la matrice de probabilités de transitions Γ.

En somme, un modèle de Markov caché peut être résumé aux éléments suivants :
— X_t un processus observé à m états formant l'ensemble d'états $E = e_1, \ldots, e_m$ produisant des observations x.
— C_t une chaîne de Markov non observée.
— m le nombre d'états de la chaîne de Markov correspondant aussi au nombre de composantes (distributions) considérées dans le modèle.
— T la longueur de la séquence d'observation de X (ou série temporelle).
— t un temps donné de la chaîne, $t < T$.
— i et j deux états données de la chaîne.
— p_{ij} les probabilités de transitions de l'état i vers l'état j.
— Γ la matrice des probabilités de transitions constituée est p_{ij}. En pratique, cette matrice n'est pas connue a priori, ce qui sera notre cas. Aussi, nous estimerons les probabilités de transition à partir des observations, en supposant que le temps entre les différentes transitions est identique au cours de l'étude (homogène).
— $u_i(t)$ les probabilités inconditionnelles.
— δ la distribution stationnaire d'une chaîne de Markov.
— $\pi_i(x)$ les distributions état-dépendantes, fonction de masse ou de densité.

2.2.2.2 Estimation des paramètres

L'estimation des paramètres d'un modèle de Markov caché résulte de la maximisation de la fonction de vraisemblance [225, 236, 235]. En particulier, la fonction de log-vraisemblance est utilisée [241]. Plus de détails sont fournis en Annexe A.2. En pratique, l'estimation des paramètres peut découler de trois types d'algorithmes d'inférence bayésienne différents [20, 225, 233, 237] : l'apprentissage, l'évaluation et l'optimisation. L'apprentissage d'une séquence d'observation utilise l'algorithme de Baum-Welch. Pour l'évaluation de la réalisation d'une séquence particulière, l'algorithme de maximisation numérique récursive, ou algorithme *forward*, est appliqué. En dernier lieu, l'optimisation, c'est-à-dire la détermination de la séquence d'états cachés la plus probable de générer les observations, est fondée sur l'algorithme *backward*, aussi appelé algorithme de Viterbi.

2.2.2.3 Validation du modèle

De façon à évaluer l'ajustement du modèle aux données, des critères de qualité d'ajustement doivent être définis. Pour les HMM, deux méthodes graphiques de diagnostic peuvent être utilisés. Lorsque le package `msm` du logiciel R est utilisé, comme c'était notre cas, ces éléments de validation sont fournis par défaut sur un même graphique [222].

Dans un premier temps, l'analyse des pseudo-résidus du modèle est essentielle [20, 243]. L'objectif est que ces derniers suivent une distribution normale centrée réduite. Ainsi, en observant l'évolution des résidus au cours du processus, il convient de vérifier qu'ils se situent globalement entre -1,96 et +1,96. De plus, l'analyse d'un diagramme quantile-quantile (*QQ-plot*) est complémentaire pour affirmer ou infirmer l'adéquation à une distribution normale [243].

Par ailleurs, les pseudo-résidus du modèle ne doivent pas être corrélés entre eux. Pour vérifier ce point, le corrélogramme de la fonction d'auto-corrélation est utilisé [235].

Il convient de s'assurer que les valeurs affichées se situent dans l'intervalle de confiance, délimité graphiquement par des lignes en pointillé. Autrement dit, le corrélogramme des pseudo-résidus du modèle ne doit montrer aucune structure.

En dernier lieu, il est possible de réaliser des simulations à partir des modèles estimés afin de vérifier la cohérence des résultats produits avec les données observées [243]. Si l'utilisateur possède un jeu de données dont les états sont connus, une validation croisée peut être envisagée pour déterminer l'adéquation du modèle. Dans ce cas, le taux de mauvaises identifications d'états après plusieurs simulations est un critère de choix. Cette méthode, qui aurait pu également être appliquée aux modèles de Markov de la section 2.2.1, n'a pas été utilisée dans notre cas car nous n'avions pas d'autre jeu de données dont les états sont connus.

2.2.2.4 Choix du nombre d'états cachés

Lorsque les HMM sont utilisés en analyse exploratoire, comme c'est le cas pour notre étude, les états ne sont, par définition, pas déterminés a priori. L'objectif est de sélectionner le nombre d'états cachés le plus adapté au phénomène étudié pour décrire les données. Il faut donc définir le nombre d'états ainsi que leur signification [20]. Pour se faire, deux options sont exploitables.

D'une part, des critères d'information théoriques classiques des modèles linéaires peuvent être utilisé [243]. Le premier est l'AIC [244], comme présenté en section 2.2.1. L'AIC le plus faible est retenu pour déterminer le nombre d'états optimal définissant ainsi l'architecture du modèle le plus adapté aux données [225, 245]. De plus, le critère d'information bayésien (*Bayesian Information Criterion* - BIC) peut être appliqué [235]. Le modèle optimal aura un BIC plus faible que les autres [246].

D'autre part, des tests statistiques peuvent être implémentés. Par exemple, comme présenté en section 2.2.1, le test du rapport de vraisemblance aide à déterminer l'apport de covariables expliquant le phénomène ou à comparer deux modèles entre eux [243].

Une fois le nombre d'états cachés spécifié, il est possible d'introduire des covariables dans le modèle. Celles-ci peuvent soit intervenir dans les probabilités de transitions d'un état à un autre, soit sur les probabilités initiales du modèle. Cela signifie que les probabilités initiales ou de transition sont une fonction des covariables [235]. En particulier, elles sont modélisées par une régression logistique binaire ou multinomiale.

2.2.2.5 Exemples d'utilisation dans la littérature

Cette approche par HMM a été utilisée dans la littérature pour analyser des données de la maladie d'Alzheimer ou pour étudier des fonctions cognitives. Les HMM sont particulièrement adaptés à l'étude des processus psychologiques ou cognitifs dans lesquels des états cognitifs qualitativement différents se développent au fil du temps [235]. Dans ce cas de ces pathologies, les fonctions cognitives en particulier sont soumises à l'incertitude de mesure qui n'est pas traitée par les méthodes classiques. En effet, une première étude s'est intéressée à des sujets atteints de la maladie d'Alzheimer [247]. Ils souhaitaient étudier l'évolution des symptômes psychotiques, tout en tenant compte simultanément de l'évolution des fonctions cognitives. Pour se faire, les symptômes psychotiques ont été mesurés par une échelle d'évaluation quantitative du comportement pour la démence *(Behavior Rating Scale for*

Dementia (BRSD)). Les états cachés ont été définis comme suit : jamais psychotique (état 1), pré-psychotique (état 2), psychose active (état 3) et psychose en rémission (état 4). Le score BRSD était distribué selon une loi de Poisson dont la moyenne dépendait de l'état caché.

De la même façon, Seltman et al. se sont intéressés au déclin cognitif dans la maladie d'Alzheimer [248], caractérisé par une variabilité de trajectoires importantes. Leur objectif était d'étudier, d'après les trajectoires des sujets au cours du suivi, l'évolution du déclin cognitif en même temps que le profil des symptômes psychotiques. Pour se faire, le déclin des fonctions cognitives a été modélisé par des régressions logistiques mixtes. Cela signifie qu'un effet aléatoire individuel a été ajouté au modèle pour tenir compte de la variabilité individuelle de l'allure des trajectoires cognitives. Les symptômes psychotiques observés ont été supposés issus d'un processus caché. Les 4 états de ce processus ont été définis a priori par : jamais psychotique, pré-psychotique, psychotique actif et psychotique en rémission. Les états avaient donc le même sens clinique que ceux de [247]. La rémission était définie ici par une phase relativement asymptomatique mais survenant après au moins une période psychotique active. Les symptômes psychotiques observés étaient supposés suivre une distribution de Poisson dont la moyenne dépendait de l'état caché. Des covariables génétiques et démographiques ont été ajoutées sur les probabilités initiales et de transition.

Finalement, une dernière analyse, dont le design est très proche de celui de notre étude, a retenu notre attention [238]. Leur objectif était de modéliser, de façon non supervisée, la progression de la maladie d'Alzheimer à partir d'un ensemble de biomarqueurs. Leur étude incluait 594 sujets dont les données ont été recueillies tous les 6 mois sur une période de 3 ans. Un modèle de Markov caché à 6 états a été utilisé pour déterminer les stades de l'évolution de la maladie (états cachés). Leur vecteur d'observation était multidimensionnel et sa distribution, pour chaque état, s'appuyait également sur des modèles de mélanges de gaussiennes. Les 4 composantes de ce vecteur étaient deux biomarqueurs ainsi que leur taux d'accroissement respectifs entre deux visites.

Chapitre 6

Application aux données de l'essai MAPT et recommandations

L'ensemble des méthodes présentées dans le Chapitre 5 ont été appliquées aux données réelles de l'essai de prévention MAPT, présenté dans le Chapitre 3. Ce chapitre 6 vise à montrer comment les méthodes ont été appliquées, quels étaient les choix effectués dans le contexte particulier évoqué jusqu'à présent et les résultats en découlant.

Pour rappel, un descriptif des caractéristiques des trois populations d'analyses a été effectué en section 2 du Chapitre 4. L'annexe A.1 propose un résumé de la section 1.3 du Chapitre 3, avec les noms des variables, leur signification ainsi que l'unité de mesure de chacune d'elle. Les paramètres de dispersion et de tendance centrale pour les variables quantitatives et les fréquences absolues et relatives pour les variables qualitatives y sont présentés. De plus, avant de détailler les outils de ré-analyse d'un essai de prévention non concluant, notons que des profils d'évolutions des fonctions cognitives ont été décrits préalablement en section 3 du Chapitre 4, confirmant que les trajectoires des fonctions cognitives n'étaient pas comme attendues cliniquement et devaient être analysées plus en détails.

L'organisation du chapitre est similaire à l'articulation du Chapitre 5. Dans un premier temps (section 1), les trajectoires ont été explorées avec les trois approches non supervisées. Pour cela, la section 1.1.1 fait état des résultats de la méthodes des *k-means* longitudinaux avec reconnaissance de la forme des trajectoires, la section 1.1.2 renvoie à la CAH et la section 1.1.3 à la sémiologie graphique. Par ailleurs, l'approche exploratoire supervisée par statut de répondeur se situe en section 1.2. Dans un second temps, ces mêmes trajectoires ont été modélisées par des algorithmes d'apprentissage statistique classiques (section 2.1). La section 2.1.1 présente les résultats d'approches classiques non détaillées en Chapitre 5, tandis que la section 2.1.2 résume les informations obtenues par le *Q-learning*. Les approches classiques avaient pour objectif de rechercher un effet du traitement ou de l'intervention multidomaine, expliqués par le Z-score composite. L'approche d'apprentissage par renforcement avait pour objectif de déterminer la règle de décision optimale conduisant au meilleur Z-score possible. Par ailleurs, les approches markoviennes ont été appliquées. Leurs résultats se trouvent en section 2.2, avec d'une part l'approche des modèles classiques dont les états sont directement observés (section 2.2.1), d'autre part l'approche des modèles de Markov dont les états sont cachés (section 2.2.2).

1 Résultats de l'exploration des trajectoires cognitives

Dans la première partie du chapitre sont abordées différentes techniques de ré-analyse d'un essai de prévention non concluant dans le but de rechercher des groupes de sujets homogènes. Ce processus peut avoir lieu sans a priori sur les groupes et sans variable à expliquer (section 1.1) ou en déterminant préalablement une variable à expliquer représentant la nature du comportement recherché (section 1.2). Dans les deux cas, la population d'analyse est divisée selon certaines caractéristiques, qu'il convient ensuite de décrire et d'analyser. Notons qu'aucune des 4 méthodes exploitées ne conduit à des estimations d'effets de covariables, simplement à une description de la différence entre les groupes en résultant.

1.1 Résultats du *clustering* fonctionnel

1.1.1 L'algorithme des centres mobiles

La première méthode de la stratégie de ré-analyse implémentait l'algorithme des centres mobiles, présentée en section 1.1.1 du Chapitre 5. Pour rappel, seule l'extension des *k-means* longitudinaux avec reconnaissance de la forme des trajectoires est utilisée ci-après (section 1.1.1.3 du Chapitre 5). Les motivations de son utilisation, les choix effectués dans notre contexte ainsi que les résultats statistiques et cliniques sont évoqués par la suite.

1.1.1.1 Interprétation statistique

D'après le *design* de l'essai MAPT, les sujets n'ont pas été inclus à une date unique. En effet, les sujets rentrent dans l'étude à des moments variables de l'évolution de leurs fonctions cognitives, le T0 correspondait ainsi à l'entrée dans l'essai clinique MAPT pour tous les sujets. Le temps définissant l'inclusion dans l'essai (*baseline*) correspondait à des dates différentes pour les sujets mais à un T0 voulant dire la même chose pour tous, c'est-à-dire l'entrée dans l'essai et donc la mise en route du traitement. Puisque les trajectoires ne commençaient pas toutes à la même date, le décalage dans le temps devait être considéré dans les analyses. La méthode des *k-means* longitudinaux reconnaissant la forme des trajectoire (package `kmlShape`) apparaissait comme appropriée à notre contexte. La totalité de la population d'étude ($\mathcal{P}_{tot}$ contenant 1679 sujets présentés en section 2.1 du Chapitre 4) a été considérée à cette étape et les données fonctionnelles des fonctions cognitives ont été utilisées directement, sans pré-traitement de la donnée.

Avant de commencer les analyses et dans le but d'avoir des tailles de groupes raisonnables, nous avons fixé une taille minimale de 10 % de la population (soit 168 sujets par groupe) [31]. De plus, la variable longitudinale d'intérêt considérée provenait du critère de jugement principal de l'essai MAPT, le Z-score composite.

Concernant les paramètres initiaux de la méthode, aucune hypothèse n'a été émise quant au nombre de groupes à déterminer. Plusieurs valeurs ont été testées. Le nombre retenu était de $k = 3$: le partitionnement en 2 groupes ne fournissait pas assez d'information, tandis que celui en 4 groupes conduisait à des effectifs de trop petite taille par rapport à nos critères. Pour initialiser l'algorithme des *k-means*, deux stratégies étaient proposées (voir section 1.1.1.3 du Chapitre 5). L'une considérait la population dans son ensemble et exploitait directement le package `kmlShape`. La seconde consistait à réduire le nombre de

trajectoires, c'est-à-dire à tirer aléatoirement un sous-ensemble considéré comme des trajectoires particulièrement représentatives de la population d'étude. Les autres trajectoires étaient assignées progressivement aux *clusters* dans les étapes suivantes de l'algorithme. La méthode a été appliquée avec et sans réduction du nombre de trajectoires. Un résultat identique a été obtenu par ces deux procédures. Ce constat était prévisible d'après la description du package, mais nous avons préféré le vérifier par nous-même. En effet, les deux stratégies tirent aléatoirement des trajectoires pour initialiser l'algorithme, soit dans l'ensemble de la population (sans réduction), soit dans un ensemble restreint de trajectoires (avec réduction). Par la suite, les *clusters* sont construits de la même façon et les trajectoires représentatives des groupes adaptées au cours des étapes. Il n'est donc pas surprenant d'obtenir un résultat identique, simplement une stratégie est plus rapide à exécuter que l'autre dans le cas de grandes dimensions. Les résultats ci-après correspondent à la méthode sans réduction du nombre de trajectoires, qui est préférable à notre sens car utilisant entièrement et directement la base de données. Rappelons que les groupes sont formés en appliquant l'algorithme de `kmlShape` et en utilisant la distance généralisée de Fréchet et la moyenne de Fréchet (voir section 1 du Chapitre 5).

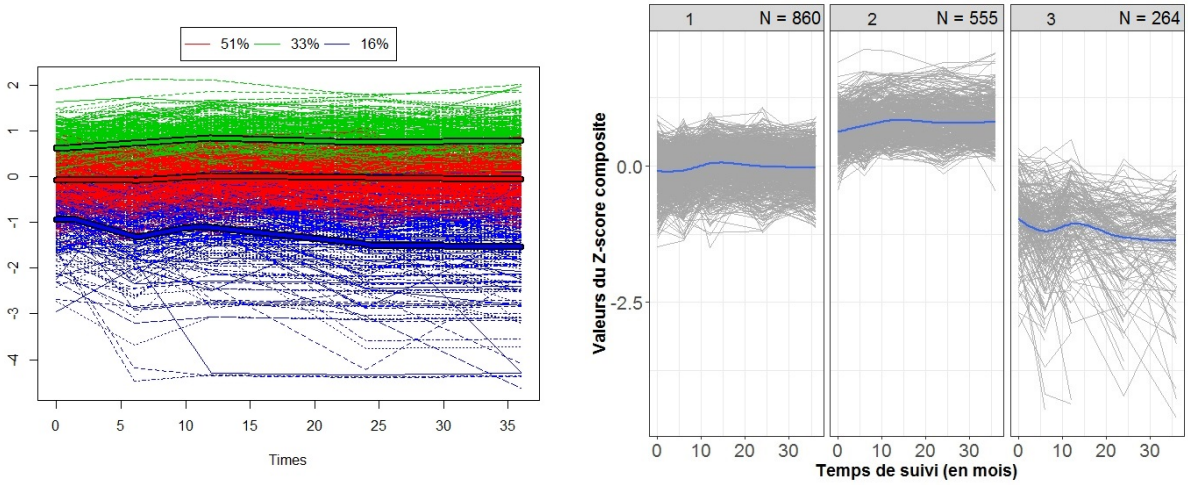

FIGURE 6.1 – *Spaghetti plot* des données observées pour les trois *clusters* définis par la méthode `kmlShape` pour les valeurs du Z-score composite, toutes trajectoires confondues à gauche et à droite avec séparation des trajectoires de chaque groupe identifié, MAPT (N = 1679 sujets).

La Figure 6.1 à gauche représente les 3 groupes obtenus en considérant cette stratégie d'analyse. L'ensemble des trajectoires individuelles du Z-score composite pour tous les sujets de l'essai y figurent. Les trajectoires moyennes identifiées après utilisation de l'algorithme `kmlShape` sont représentées en gras. Chaque couleur correspond à un *cluster*. En haut est indiqué le pourcentage de la population contenu dans chaque groupe. Afin de distinguer plus clairement les trajectoires moyennes et individuelles des groupes, nous les avons représentées séparément sur la Figure 6.1 à droite. Ces deux représentations fournissent exactement la même information, bien que la seconde soit visuellement plus explicite.

Le premier groupe, représenté en rouge, contenait la majorité de la population avec 51 % des sujets. Le deuxième, représenté en vert, en incluait 33 % et le groupe bleu les 16 % restant.

1.1.1.2 Interprétation clinique

D'un point de vue clinique, les profils d'évolution des fonctions cognitives obtenus étaient différents. La trajectoire moyenne du groupe majoritaire (groupe 1 sur la Figure 6.1) montrait une stabilité sur les 6 premiers mois, une amélioration sur les 6 suivants, un déclin léger sur la 2eme année de suivi et finalement, une stabilité en fin de suivi. Le deuxième groupe possédait un niveau cognitif moyen supérieur au premier à l'inclusion (trajectoire moyenne plus haute). Il était caractérisé par une amélioration des fonctions cognitives sur les 12 premiers mois, un déclin léger sur les 12 suivants et une stabilité sur la dernière période. Ces 2 groupes semblaient similaires selon l'évolution des fonctions cognitives sur l'intégralité du suivi, excepté pour la première année de suivi, dont l'évolution semble caractérisée par un niveau initial des trajectoires différent. Concernant le dernier groupe, il divergeait des autres en terme d'évolution. Sa trajectoire moyenne montrait un déclin des fonctions cognitives sur les 6 premiers mois, une amélioration légèrement sur les 6 suivants, de nouveau un déclin au cours de la 2eme année pour finir par se stabiliser (ou déclin très léger) sur la dernière année de suivi. Le niveau initial des sujets de ce dernier groupe était plus faible que celui des deux précédents.

Dans un second temps, les distinctions entre les groupes formés par le partitionnement de `kmlShape` ont été recherchées en examinant leurs caractéristiques cliniques. Le Tableau 6.1 reprend les caractéristiques observées à l'inclusion. Les divergences relevées allaient dans le sens attendu a priori, aucun résultat contradictoire n'a été relevé. La cohérence des différences cliniques nous a conforté dans l'idée que la méthode des *k-means* longitudinale avec reconnaissance de forme se révélait adaptée aux données. Cependant, contrairement à l'hypothèse émise en début d'analyse, selon laquelle les sous-population homogènes pouvaient différer selon les bras de l'essai, les sujets étaient répartis uniformément entre les 4 bras de l'essai (p =0,310, test du χ^2). Ce résultat était concordant avec ceux de l'essai MAPT car aucun effet intervention ou traitement n'avait été retrouvé sur les trois premières années de suivi [13]. De même, aucune différence n'avait été constatée dans les analyses préliminaires de la population (voir section 3 du Chapitre 4). La présente analyse ne montrait également aucune différence d'IMC à *baseline* (p = 0,137, test du χ^2), de cholestérolémie (p = 0,059, test du χ^2) ou d'hypertension (p = 0,427, test du χ^2). Par ailleurs, 92 % des sujets ayant été diagnostiqué comme ayant la maladie d'Alzheimer au cours du suivi (sujets déments) appartenaient au groupe 3 de la Figure 6.1 à droite, représenté en bleu ($p < 0{,}001$, test du χ^2). Ces sujets étaient au nombre de 5 dans le groupe 1 (0,58 % de l'effectif du groupe), 0 dans le groupe 2 et 59 dans le groupe 3 (22,35 % de l'effectif du groupe).

Le groupe de sujets présentant une évolution des fonctions cognitives différente des autres a particulièrement attiré notre attention. Son profil allait dans le sens d'un déclin au cours des trois années de suivi. Représenté en bleu sur la Figure 6.1 gauche et au numéro 3 sur la Figure 6.1 droite, il contenait 92,2 % des sujets déments de l'étude. Plus particulièrement, les individus allaient moins bien sur des facteurs de risques connus du déclin cognitif et de la maladie d'Alzheimer. Ils étaient significativement âgés de plus de 2 ans par rapport aux autres groupes et avaient un score de risques cardiovasculaires, vieillesse et démence (CAIDE) plus élevé. Ces sujets se plaignaient davantage d'un point de vue mnésique ou de la gêne ressentie dans les activités quotidiennes que les autres sujets (respectivement, l'écart minimum était de 7 mm et 9 mm environ). Leur niveau d'autonomie était plus faible de 3 points et la proportion de symptomatologie dépressive positive était plus importante dans ce groupe que dans les deux autres (28 %). Leurs scores cognitifs sur les 3 autres tests disponibles étaient également inférieurs et près de 39 % d'entre eux avaient un niveau

TABLE 6.1 – Récapitulatif des caractéristiques à *baseline* significativement différentes pour les 3 groupes identifiés pour le Z-score composite par la méthode des centres mobiles basée sur la forme des trajectoires (`kmlShape`), MAPT (N = 1679 sujets).

Variables	Modalités	Groupe 1	Groupe 2	Groupe 3	P-value
Age		75,49 (4,38)	74,00 (3,75)	77,64 (4,82)	< 0,001*
TMT-A		47,62 (14,93)	38,95 (11,93)	58,94 (23,72)	< 0,001*
TMT-B		125,24 (52,16)	92,63 (36,43)	181,76 (95,57)	< 0,001*
COWAT		18,55 (5,72)	23,38 (5,86)	15,82 (6,56)	< 0,001*
SPPB		10,55 (1,61)	10,97 (1,36)	9,89 (2,06)	< 0,001*
EVA-1		50,08 (16,56)	46,24 (15,55)	57,34 (18,72)	< 0,001*
EVA-2		40,01 (22,85)	35,13 (21,68)	49,30 (23,22)	< 0,001*
Autonomie		39,69 (4,50)	41,11 (3,85)	36,63 (6,00)	< 0,001*
CAIDE		7,79 (2,02)	7,09 (1,98)	8,15 (1,82)	< 0,001*
Etude					< 0,001
	1 : Haut	625 (74,14)	490 (90,24)	157 (61,09)	
	0 : Bas	218 (25,86)	53 (9,76)	100 (38,91)	
Sexe					< 0,001
	1 : Hommes	339 (39,42)	136 (24,50)	117 (44,32)	
	0 : Femmes	521 (60,58)	419 (75,50)	147 (55,68)	
CDR					< 0,001
	1 : Positif	370 (43,07)	150 (27,03)	186 (70,45)	
	0 : Négatif	489 (56,93)	405 (72,97)	78 (29,55)	
GDS					< 0,001
	1 : Oui	149 (17,41)	75 (13,61)	74 (28,14)	
	0 : Non	707 (82,59)	476 (86,39)	189 (71,86)	
Fried					< 0,001
	1 : Oui	363 (42,21)	206 (37,12)	173 (65,53)	
	0 : Non	497 (57,79)	349 (62,88)	91 (34,47)	
Phénotype					0,018
	1 : Oui	154 (23,12)	87 (19,73)	58 (30,05)	
	0 : Non	512 (76,88)	354 (80,27)	135 (69,95)	

Variables quantitatives, moyenne M (écart-type SD) ; Variables qualitatives, effectif N (pourcentage %).
* test de l'ANOVA ; test du χ^2

d'éducation considéré comme bas. La proportion d'individus portant l'allèle ϵ_4 du gène APOE, de façon homo- ou hétérozygote, était plus importante que dans les autres groupes (30 %). Ce groupe montrait un profil d'évolution des fonctions cognitives ayant une tendance similaire à celui construit en section 3.1 du Chapitre 4. Seulement, contrairement à cette analyse, des caractéristiques cliniques le différenciant ont été déterminées. La méthode des *k-means* pour données longitudinales et reconnaissant la forme de la trajectoire a ainsi complété les informations à notre disposition.

Inversement, les sujets du groupe 2, démontrant un score cognitif global plus stable au cours du suivi et un niveau moyen supérieur aux autres groupes, se caractérisaient par des résultats supérieurs sur ces mêmes facteurs initiaux. D'après ces informations, ils semblaient moins disposés à présenter une démence de type Alzheimer selon les facteurs de risques connus de la littérature. Cette information complète ainsi celle fournie par l'analyse préliminaire de la section 3 du Chapitre 4. En effet, un profil de variations "stable" avait été constitué sans pour autant être clairement défini en termes de caractéristiques cliniques.

Finalement, un point que nous aurions pu soulever est que les profils formés n'étaient pas aussi homogènes que ce que nous avions espéré. Ce constat est parfaitement en accord avec ce que nous avions remarqué à l'issue des analyses préliminaires de la population, à savoir que la population ne se comporte pas comme attendu cliniquement (voir section 3 du Chapitre 4). De la même façon, il n'était pas surprenant que les groupes ne soient pas différents selon les bras de l'essai ou l'IM, puisque l'essai MAPT n'a pas réussi à déterminer un effet groupe (essai non concluant). Ces résultats ont été confirmés par les quatres méthodes d'apprentissage statistiques présentées en section 2.1.1 du Chapitre 6 dont les résultats sont décrits ci-après. La concordance entre ces recherches de profils nous donne confiance en nos résultats. Le dernier point à mettre en évidence, pour la présente analyse, est que les variations du Z-score composite au cours du suivi semblaient trop faibles pour que l'algorithme de `kmlShape` les détecte correctement et constitue les groupes en fonction. Ici, les groupes semblaient formés sur le niveau initial du Z-score (comme les groupes 1 et 2). Une méthode alternative a été envisagée pour étudier, sous un autre angle, les profils d'évolution des fonctions cognitives (voir section 1.1.2 ci-après).

1.1.2 L'algorithme de Classification Ascendante Hiérarchique

1.1.2.1 Interprétation statistique

La deuxième méthode d'exploration des trajectoires cognitives provenait de la CAH. La distance euclidienne et la méthode d'agrégation de Ward (critère d'inertie) ont été appliquées aux données de MAPT.

De plus, pour rendre la méthode pertinente dans un contexte de données longitudinales, la CAH ne peut être appliquée telle quelle. Afin de tenir compte de l'aspect chronologique des données et du temps séparant les différences mesures, un indicateur de trajectoire a été construit. Les données fonctionnelles étaient donc transformées à l'aide d'un paramètre de synthèse. Les taux d'accroissement du Z-score composite entre chaque visite ont été calculés, à l'instar de ceux utilisés en section 3.1 du Chapitre 4, soit $p = 4$ taux au total. Le tableau de données utilisé pour la CAH contenait les 4 taux en colonnes et les individus en ligne. Il était donc de format similaire à celui de la section 1.1.1, ce qui signifie que les covariables utilisées pour former les groupes n'étaient pas indépendantes.

L'utilisation du paramètre de synthèse avait pour conséquence, outre une potentielle perte d'information car la donnée fonctionnelle était résumée sur une période donnée, de réduire la taille de l'échantillon d'étude aux sujets ayant un suivi complet. Cela signifie que les sujets ayant des données manquantes sur le Z-score composite ont été supprimées de l'analyse, réduisant la population à 1143 sujets (voir $\mathcal{P}_{comp}$ en section 2.2 du Chapitre 4). Ces sujets étaient donc présents du début à la fin du suivi, ce qui signifie que les perdus de vue n'ont pas été considérés dans l'analyse, malgré leur importance discutée dans les chapitres précédents.

Avant de s'intéresser aux résultats de la CAH, observons l'effet de la transformation des données brutes en taux d'accroissement. En Figure 6.2 sont représentées les données non traitées d'une part (à gauche) et les taux d'accroissement d'autre part. Nous avons constaté que la trajectoire moyenne (représentée en bleu) des données brutes était relativement constante au cours du suivi. De plus, les trajectoires individuelles étaient hétérogènes. En observant les taux d'accroissement au cours des différentes périodes de suivi, l'hétérogénéité semblait moindre, mis à part sur la période 2. Il semblerait donc qu'un changement se

produise entre 6 et 12 mois de suivi, se manifestant par une augmentation de l'étendue des valeurs des taux d'accroissement. De plus, l'évolution au cours du temps restait relativement constante. Les variations importantes observées en Figure 6.2 résultent principalement d'un effet d'optique dû à un changement d'échelle entre les deux graphiques. Une échelle identique ne peut être adoptée dans le cas présent car les taux d'accroissement seraient alors considérés comme nuls, donc aucune variation ne serait visible.

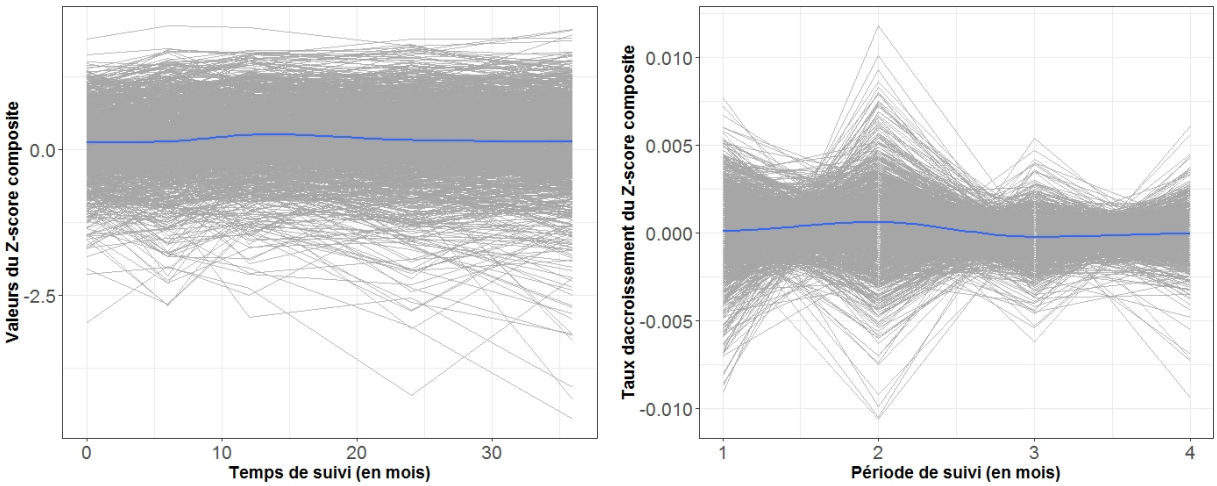

FIGURE 6.2 – Représentation des valeurs longitudinales du Z-score composite (données brutes à gauche) et des taux d'accroissement de ce dernier (à droite) pour la population analysée, MAPT (N = 1143 sujets). Période 1 = T0-T6 mois, période 2 = T6-T12, période 3 = T12-T24 et période 4 = T24-T36 mois de suivi.

D'après les choix évoqués pour la stratégie d'analyse, les résultats de la CAH ont conduit à un partitionnement optimal en 3 groupes. Le graphique d'inertie est présenté en Figure 6.3 ci-dessous. Quatre niveaux pouvaient être proposées pour élaguer l'arbre hiérarchique. En vert est indiqué le partitionnement en 2 groupes, en rouge en 3, en bleu en 4 et le dernier partitionnement raisonnable était celui en 6 groupes en violet. D'après la fonction `best.cutree` et le critère du coude, le partitionnement optimal était celui en 3 groupes. Une représentation graphique des courbes de ces 3 groupes est située en Figure 6.4. Son premier groupe, contenant 155 individus (14 % des sujets), son deuxième 505 sujets (44 %) et le dernier en contenait 483 (42 %). Notons que l'hétérogénéité des groupes est variable : le premier groupe est homogène tandis que le dernier semble être le plus hétérogène des trois.

1.1.2.2 Interprétation clinique

D'un point de vue clinique, la trajectoire moyenne du premier groupe de la Figure 6.4 était marquée par une diminution sur les 6 premiers mois de l'essai, une augmentation sur les 6 mois suivant, une diminution légère sur la deuxième année de suivi puis une stabilisation sur la dernière année. Concernant le groupe 2 de la Figure 6.4, la trajectoire moyenne était de type diminution légère sur les 6 premiers mois, augmentation, diminution et stabilisation sur la dernière période de suivi. Pour le dernier groupe de la Figure 6.4, la trajectoire moyenne commençait par une augmentation sur la première période, suivie d'une diminution et d'une stabilité sur les deux dernières années de suivi. En somme, aucun groupe n'était complètement stable d'un point de vue cognitif sur les 3 ans de suivi, bien qu'une stabilité soit notable sur la dernière année de suivi pour l'ensemble de la population étudiée. Ce

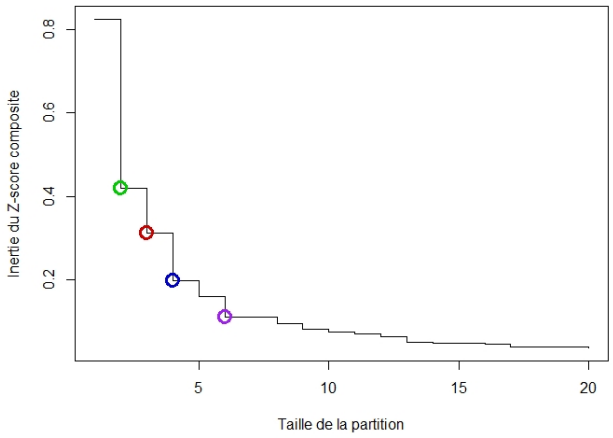

FIGURE 6.3 – Inertie associée au dendrogramme de la CAH à partir des 4 taux d'accroissement du Z-score composite en fonction du nombre de classes, MAPT (N = 1143 sujets). Sauts d'inertie identifiés à 2 groupes (en vert), à 3 groupes (en rouge), à 4 groupes (en bleu) et à 6 groupes (en violet).

résultat complète ainsi les analyses précédentes, dont deux profils semblaient relativement stables, tout en nuançant la notion de stabilité des fonctions cognitives. Cette stabilité était donc, dans cette analyse, synonyme de variations légères complémentaires (augmentation puis diminution ou vice versa) conduisant à un niveau constant à long terme.

D'après les caractéristiques cliniques relevées à *baseline*, très peu d'entre elles ressortaient comme significativement différentes entre les trois groupes (voir Tableau 6.2). En particulier, la population ne différait pas concernant sa répartition selon les bras de l'étude (p = 0,507, test du χ^2), ni selon la démence (1 sujet dans le groupe 1, 2 dans le groupe 2 et 2 dans le groupe 3). Les groupes avaient un niveau d'éducation équivalent (p = 0,627, test du χ^2), tout comme leur statut phénotypique (p = 0,210, test du χ^2), leur score CDR (p = 0,178, test du χ^2), leur sexe (p = 0,902, test du χ^2), leur symptomatologie dépressive (p = 0,314, test du χ^2), leur critère de Fried de fragilité (p = 0,301, test du χ^2), leur cholestérolémie (p = 0,095, test du χ^2) ou encore leur hypertension (p = 0,987, test du χ^2). Ces facteurs sont pourtant connus comme étant facteurs de risque de déclin cognitif ou de la survenue de la maladie d'Alzheimer (voir section 1.3 du Chapitre 3). Concernant les variables quantitatives, les tests de comparaisons appliqués provenaient d'ANOVA. Ces trois groupes avaient une moyenne d'âge similaire (p = 0,356), tout comme pour le score CAIDE (p = 0,301), la plainte mnésique (p = 0,187), le niveau d'autonomie (p = 0,737), les performances physiques (p = 0,330). Leur score au TMT-A était également proche (p = 0,080), à l'instar du score du test COWAT (p = 0,413).

Dans ces résultats, certains points étaient surprenant cliniquement. En effet, il est assez atypique d'avoir une plainte mnésique non significative alors que la gêne ressentie dans les activités quotidiennes l'est. Pour le TMT, dont la version B est plus complète que la A, il est attendu que les deux versions du test aillent de paire. Ce n'était pas le cas dans la présente étude (voir Tableau 6.2). Cependant, il est tout de même envisageable que l'un ressorte et pas l'autre d'après les domaines évalués par les tests (voir section 1.3 du Chapitre 3).

Par ailleurs, le choix du paramètre de synthèse de la donnée fonctionnelle sur une période de temps a restreint l'échantillon d'analyse aux sujets ayant un suivi complet pour

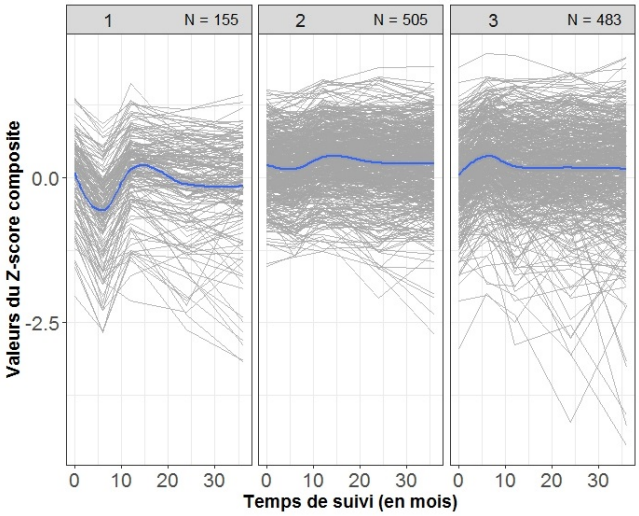

FIGURE 6.4 – *Spaghetti plot* des données observées pour un partitionnement en 3 groupes définis par l'inertie au dendrogramme de la CAH sur les taux d'accroissement du Z-score composite, MAPT (N = 1143 sujets).

TABLE 6.2 – Récapitulatif des caractéristiques à *baseline* significativement différentes pour les 3 groupes identifiés pour le Z-score composite par la méthode de la CAH, MAPT (N = 1143 sujets).

Variables	Modalités	Groupe 1	Groupe 2	Groupe 3	P-value
TMT-B		133,36 (70,07)	113,31 (46,98)	111,00 (46,17)	< 0,0001*
EVA-2		33,94 (22,19)	38,97 (22,19)	40,74 (24,38)	0,0065*
IMC					0,0097
	1 : Oui	12 (7,74)	89 (17,69)	82 (16,98)	
	0 : Non	143 (92,26)	414 (82,31)	401 (83,02)	

Variables quantitatives, moyenne M (écart-type SD) ; Variables qualitatives, effectif N (pourcentage %).
* test de l'ANOVA ; test du χ^2

le Z-score composite. Sur les 1143 sujets analysés, seulement 5 d'entre eux ont développé une démence au cours du suivi. Plus de 95 % des sujets déments avaient des valeurs manquantes sur le critère de jugement de l'étude, ce qui signifie qu'ils n'ont pas été considérés dans la présente analyse. Les sujets analysés par la CAH étaient donc sur-sélectionnés par rapport à la population générale. Dans l'essai GuidAge [9], les auteurs ont mis en évidence que les sujets présents jusqu'en fin de suivi sont généralement en meilleure santé et ont un niveau d'éducation plus élevé. Ils sont également caractérisés par un taux de déclin cognitif ou de démence plus faible que ceux ayant un suivi incomplet. Dans l'essai PREADVISE [79], les auteurs soulignaient le fait que la faible incidence de la démence au cours du suivi pourrait expliquer l'absence d'effet. C'est une hypothèse que nous pouvons renouveler au vue de nos résultats, bien qu'elle se traduise ici en terme de différence entre les groupes et non en terme d'effet.

1.1.3 La méthode de sémiologie graphique

1.1.3.1 Interprétation statistique

La dernière stratégie d'exploration non supervisée des trajectoires de fonctions cognitives employait l'analyse combinatoire de sémiologie graphique. Le premier choix que nous avons effectué pour l'appliquer était de restreindre la population d'étude uniquement aux sujets ayant un suivi complet (voir $\mathcal{P}_{comp}$ en section 2.2 du Chapitre 4). Ce choix, comme pour la CAH (voir section 1.1.2), découlait de l'utilisation d'un paramètre de synthèse de la donnée fonctionnelle. Ce dernier était d'ailleurs identique à celui choisi en section 1.1.2, à savoir le taux d'accroissement entre les visites. De plus, toujours pour assurer une comparabilité entre les méthodes, la matrice de dissimilarité entre les taux d'accroissement a été construite avec la distance euclidienne.

Par ailleurs, cette méthode nécessite d'émettre un a priori quant au nombre de groupes. Ce dernier a été défini selon les variations attendues cliniquement. Suite aux analyses préliminaires menées en section 3 du Chapitre 4, une échelle de couleur à trois niveaux ($k = 3$) a été adoptée, de type bleu-blanc-rouge [16]. Les variations des taux d'accroissement du Z-score composite ont ainsi été divisées en 3 classes. Ce choix était également motivé par le souhait de caricaturer les évolutions possibles.

Concernant la permutation des lignes et/ou des colonnes présentée en section 1.1.3 du Chapitre 5, deux types différents ont été sélectionnés respectivement pour les lignes et les colonnes. La permutation des lignes était de type *Spectral*, la méthode par défaut pour un tableau de données tel que le notre. Cela signifie que le critère d'optimisation de cette fonction de perte est la double somme des similarités (*2-Sum*). Elle quantifie la similarité entre les objets par le carré de la différence des rangs [16]. Pour être calculé, ce critère s'appuie sur la matrice des distances. Notons d_{ij}, $i \leq i$, $j \leq n$ la valeur de la dissimilarité entre deux objets i et j. Le critère de *2-Sum* multiplie les similarités entre objets de la façon suivante [249] :

$$L(D) = \sum_{i,j=1}^{p} s_{ij}(i-j)^2 \quad ,$$

avec $s_{ij} = \frac{1}{1+d_{ij}}$ représentant la similarité entre les objets i et j. L'objectif de la permutation est alors de minimiser ce critère.
Concernant les colonnes, nous avons imposé de conserver l'ordre chronologique des mesures, ce qui signifie l'absence de permutation entre elles [250]. Dans le cas des données de MAPT, la structure obtenue par sériation n'était dépendante que de la permutation des lignes de la matrice, donc la permutation des individus.

La Figure 6.5 représente notre jeu de données avant application de la permutation. Chaque colonne correspond à un taux d'accroissement du Z-score composite entre deux visites, les périodes associées étant indiquées en légende. Chaque ligne est équivalente à un sujet de l'étude avec ses données longitudinales. Sur le coté droit est quantifié l'amplitude des variations selon les différentes couleurs sélectionnées. La couleur rouge indiquait des variations allant de 0,0043 à 0,0118 ; la blanche de -0,0031 à 0,0043 et la bleue de -0,0106 à -0,0031. Mis à part sur les taux des 2 dernières années de suivi, pour lesquels la couleur prédominante était blanche (valeurs proches de 0, variations faibles), aucune structure des données n'est apparente avant permutation. Ce constat de stabilité en fin de suivi a déjà été noté avec la CAH en section 1.1.2.

Après avoir appliqué l'ensemble des choix spécifiques décrits ci-dessus, la permutation

a révélé une structure optimale en trois groupes. Ce résultat est représenté en Figure 6.6. Le premier groupe minoritaire, situé en haut de la Figure 6.6, correspondait à des sujets ayant une évolution de type rouge - bleu - mixte - mixte. Le deuxième groupe minoritaire, situé en bas, était son opposé avec des variations de type bleu - rouge - mixte - mixte. Le dernier, au milieu des deux précédents, englobe tous les autres sujets. Il était majoritaire avec un profil de variation globalement blanc. L'échelle des valeurs, indiquée à droite de l'image, était identique à celle de la Figure 6.5.

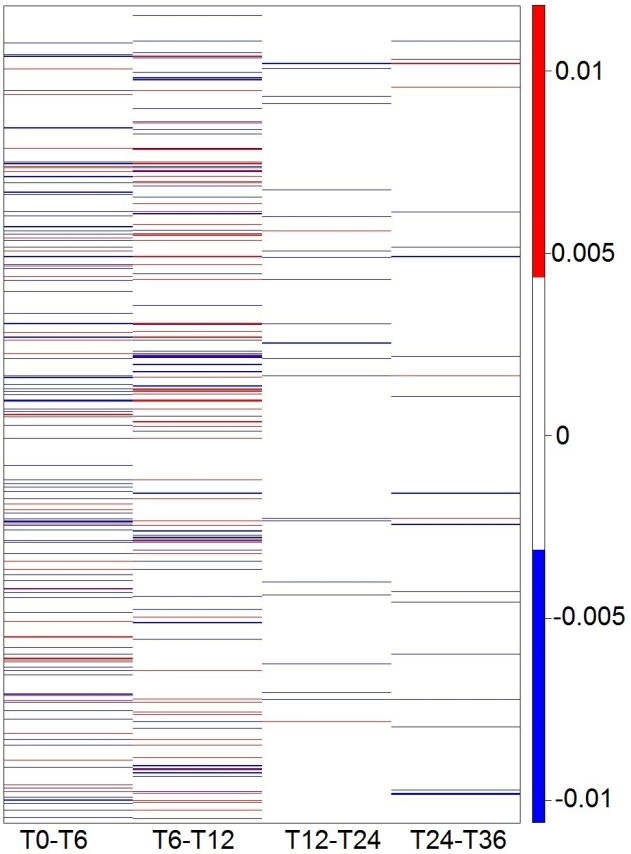

FIGURE 6.5 – Représentation des différentes couleurs sélectionnées pour la méthode de sériation avant permutation des lignes pour les taux d'accroissement du Z-score composite, MAPT (N = 1143 sujets). Rouge = augmentation ; blanc = constant ; bleu = diminution.

1.1.3.2 Interprétation clinique

Pour chaque profil apparent, le package de R n'indique pas si le sujet appartient ou non à un des profils identifiés. Ce package ne fait que la réorganisation de la base de données discutée plus tôt. Ainsi, nous avons extraits des sujets dont nous étions certains de leur appartenance aux profils décrits. Des sous-groupes de 100 sujets ont été sélectionnés pour décrire et comparer leurs caractéristiques cliniques. Leur nombre est arbitraire et nous semblait raisonnable pour avoir un aperçu de ce qu'il se passait au sein de cette population d'analyse. La sélection était manuelle : nous avons identifié, pour chaque profil, des zones homogènes sur le graphique. Nous avons mesuré la taille de ces zones, sachant que la taille totale du graphique correspondait à 1143 sujets. Avec un produit en croix, nous avons déterminé que les 100 premiers sujets de la matrice de données constituerait le premier profil, les 100 derniers le troisième profil, et pour les autres, nous avons sélectionné 100 sujets au niveau du milieu du graphique (indicés entre 500 et 600). Leurs trajectoires individuelles

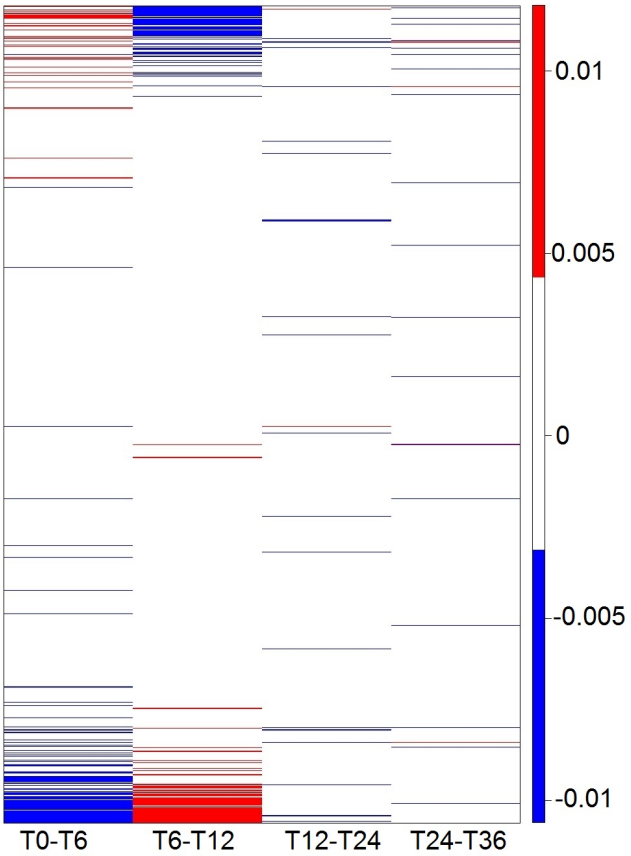

FIGURE 6.6 – Représentation des différentes couleurs sélectionnées pour la méthode de sériation après permutation des lignes pour les taux d'accroissement du Z-score composite, MAPT (N = 1143 sujets). Rouge = augmentation ; blanc = constant ; bleu = diminution.

sont représentées en Figure 6.7. Les groupes obtenus étaient plus homogènes qu'avec les deux précédentes méthodes.

En terme d'évolution, le premier groupe correspondait à des sujets dont le taux d'accroissement était positif sur les 6 premiers mois de l'étude, ce qui signifie qu'ils ont amélioré le niveau global de leurs fonctions cognitives (couleur rouge). Ensuite, sur la période T6-T12 mois, ces mêmes sujets déclinaient (couleur bleue). Sur l'année suivante, ils amélioraient de nouveaux leur niveau global, à un degré moindre qu'initialement. Lors de la dernière année, la tendance majoritaire était une stabilité des fonctions cognitives (couleur blanche). Concernant le deuxième groupe, le profil d'évolution étaient globalement stable ou constant sur les trois années de suivi, avec une légère amélioration entre T6 et T12 mois. Finalement, le dernier sous-groupe identifié montrait un déclin du niveau global des fonctions cognitives sur les 6 premiers mois de l'étude, suivi d'une augmentation sur la période T6-T12 mois. Au cours de la deuxième année de suivi, les fonctions cognitives déclinaient de nouveau, de façon moins importante que sur la première période de suivi. Finalement, les fonctions cognitives de ces sujets étaient stables sur la dernière année.

Cliniquement, les sujets étaient répartis de façon homogène dans les bras de l'essai (p = 0,978, test du χ^2) et très peu de caractéristiques apparaissaient comme statistiquement significatives. Cette fois-ci, le récapitulatif des différences statistiquement non significatives est présenté dans le Tableau 6.3. Le score au TMT-B était de 124,32 secondes (sd = 57,48) dans le groupe 1, de 106,32 s (sd = 39,98) pour le groupe 2 et 140,82 s (sd = 77,30) pour le groupe 3. Les sujets du groupe 3 avaient un temps d'exécution du TMT-B significativement

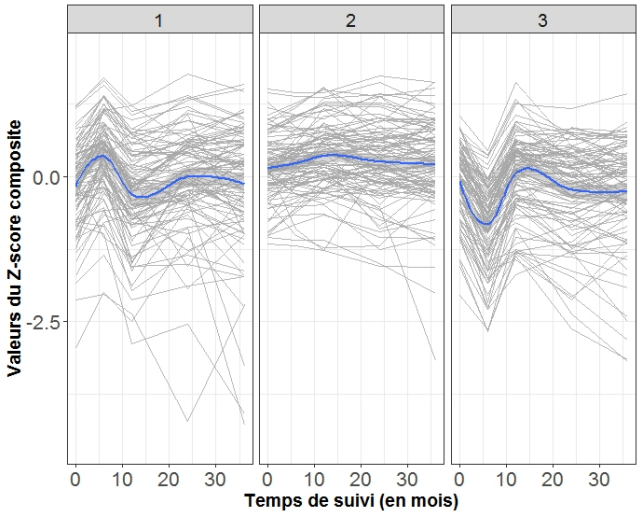

FIGURE 6.7 – *Spaghetti plot* des données observées pour les trois *clusters* définis par la méthode de sériation pour les taux d'accroissement du Z-score composite, MAPT (N = 300 sujets).

plus long que ceux des autres groupes (p = 0,001, test de l'ANOVA). Concernant l'IMC, la proportion des sujets en obésité était plus élevée dans le groupe 1 (21 %) que dans le groupe 2 (16 %) ou dans le groupe 3 (7 %). Cette disparité était significative (p = 0,018, test du χ^2).

Les résultats de cette analyse apportent donc de nouveaux profils d'évolution des fonctions cognitives. Elle complète les précédents résultats. Contrairement à ce que nous avions supposé initialement, ces groupes ne dépendaient pas du groupe d'intervention et n'étaient pas différents sur des critères cliniques pertinents. Ce constat est similaire à ce que nous avons relevé dans les précédentes analyses. Il n'était pas celui attendu d'un point de vue clinique, nous espérions relever davantage de différences entre les groupes. En revanche, ces derniers étaient plus homogènes en terme de variations inter-individuelle que dans les méthodes précédentes.

1.2 Résultats de la recherche de groupes répondeurs

Pour la dernière méthode exploratoire, nous avons adopté une approche supervisée. En effet, les groupes ont été formés selon un critère défini préalablement, critère que nous avons cherché à expliquer avec les caractéristiques relevées à l'inclusion. Avant de déterminer le statut de répondeur dans le cadre d'un essai de prévention du déclin des fonctions cognitives, une revue de la littérature a été effectuée pour avoir un aperçu de la définition de critères équivalents. Nous avons ciblé la recherche sur des critères de jugement quantitatifs.

1.2.1 Revue de la littérature

Une revue de la littérature permet de rapporter quelques exemples de l'utilisation de ce critère. En effet, dans leur revue systématique sur la dépression, Apaydin et al. [251] se sont appuyés sur des données provenant d'études antérieures. Leur objectif consistait à synthétiser les résultats d'efficacité de l'extrait de millepertuis, dont les vertus antidépressives

TABLE 6.3 – Récapitulatif des caractéristiques à *baseline* non significativement différentes pour les 3 groupes identifiés par la méthode de sériation pour les taux d'accroissement du Z-score composite, MAPT (N = 300 sujets).

Variables	Modalités	Groupe 1	Groupe 2	Groupe 3	P-value
Age		75,08 (4,07)	74,84 (4,05)	75,53 (4,19)	0,483*
TMT-A		46,61 (17,15)	44,07 (13,61)	49,06 (16,05)	0,082*
COWAT		20,12 (6,64)	20,53 (5,63)	18,95 (6,00)	0,166*
SPPB		10,80 (1,37)	10,31 (1,73)	10,50 (1,70)	0,102*
EVA-1		49,70 (17,60)	51,86 (16,02)	47,86 (16,81)	0,244*
EVA-2		41,35 (27,33)	42,11 (21,86)	36,73 (23,12)	0,237*
Autonomie		39,37 (5,69)	39,87 (4,27)	39,51 (5,05)	0,774*
CAIDE		7,88 (2,03)	7,40 (1,82)	7,54 (1,78)	0,190*
Etude					0,283
	1 : Haut	73 (73,74)	82 (82,00)	73 (73,74)	
	0 : Bas	26 (26,26)	18 (18,00)	26 (26,26)	
Sexe					0,227
	1 : Hommes	43 (43,00)	32 (32,00)	34 (34,00)	
	0 : Femmes	57 (57,00)	68 (68,00))	66 (66,00)	
CDR					0,074
	1 : Positif	45 (45,00)	36 (36,00)	52 (52,00)	
	0 : Négatif	55 (55,00)	64 (64,00)	48 (48,00)	
GDS					0,133
	1 : Oui	15 (15,15)	10 (10,00)	20 (20,20)	
	0 : Non	84 (84,85)	90 (90,00))	79 (79,80)	
Fried					0,286
	1 : Oui	46 (46,00)	39 (39,00)	50 (50,00)	
	0 : Non	54 (54,00)	61 (61,00)	50 (50,00)	
Hypertension					0,239
	1 : Oui	45 (45,00)	35 (35,00)	45 (45,45)	
	0 : Non	55 (55,00)	65 (65,00)	54 (54,55)	
Phénotype					0,738
	1 : Oui	21 (23,60)	24 (25,81)	25 (28,74)	
	0 : Non	68 (76,40)	69 (74,16)	62 (71(26))	
Cholesterol					0,515
	1 : Oui	45 (45,45)	50 (50,51)	42 (42,42)	
	0 : Non	54 (54,55)	49 (49,49)	57 (57,58)	

Variables quantitatives, moyenne M (écart-type SD) ; Variables qualitatives, effectif N (pourcentage %).
* test de l'Anova ; test du χ^2

sont connues. Les traitements de comparaison étaient un placebo, une dose de millepertuis différente ou un autre comparateur actif. Leur critère de jugement principal reposait sur l'échelle d'évaluation clinique de Hamilton pour la dépression (*Hamilton clinical rating scale for depression* - HAMD) ou toute autre échelle d'évaluation validée. Dans les publications sélectionnées, les auteurs faisaient état soit du nombre de patients en rémission, soit des taux de rechute. Pour quantifier la réponse au traitement, les répondeurs ont été définis comme des sujets ayant une perte de 50 % du score d'intérêt HAMD entre début et fin de suivi. Ils ont également utilisé la définition de la rémission (score HAMD inférieur ou égal à 7 ou 8 points). Paraschakis et al. [252] ont fait de même : à l'aide d'une revue systématique dans

le domaine de la dépression. Leur intérêt portait sur les changements de l'échelle de dépression de Hamilton. Les auteurs ont utilisé des classifications fournies dans les publications originales pour déterminer le statut répondeurs/non répondeurs.

D'autres études s'appuyant sur une comparaison avec un placebo ont attiré notre attention. La première portait sur les douleurs neuropathiques [253]. L'impact du traitement mis à l'épreuve sur la durée de la douleur a été étudié. Pour mesurer l'efficacité du traitement, le changement moyen de score de douleur en fin d'étude, par rapport au placebo, a été utilisé. Trois scores ont été analysés pour ce même objectif. Les auteurs avaient donc utilisé leur propre définition du statut répondeur. De la même façon, Katzberg et al. [254] ont porté leur attention sur la différence minimale importante sur le plan clinique (*Minimal clinically importance difference* (MCID)) dans le cadre du traitement sur la myasthénie grave. Le seuil usuel considérait une augmentation strictement supérieure à 2,5 unités, mais les essais cliniques utilisaient un seuil de 3,5 unités. Les auteurs se sont référés à quatre scores pour mesurer leur critère. L'évaluation de l'évolution du critère se basait sur la distribution du MCID calculée par l'erreur standard de la moyenne $SEM = SD \times (1 - \alpha_{Cronbach}\ )$ à partir des données du groupe placebo à 14 jours de suivi. Ils ont considéré la distribution et l'effet de taille, mesuré par $Esize = \dfrac{\mu_{14} - \mu_0}{SD\mu_0}$, avec $SD\mu_0$ l'écart-type à *baseline*, μ_0 la moyenne à *baseline* et μ_{14} la moyenne à 14 jours. Chacun des critères a permis de dichotomiser la population en répondeurs et non répondeurs.

1.2.2 Interprétation statistique

En somme, les définitions admises dans la littérature concernant cette notion de répondeur dépendent non seulement du domaine d'application de ce dernier, mais également des choix des auteurs. Ceux-ci proviennent principalement de critères cliniques découlant de la connaissance du domaine d'application. Aucun critère objectif n'a été défini jusqu'à présent. Pour tenir compte des définitions proposées dans la littérature ainsi que du contexte d'essais de prévention du déclin des fonctions cognitives, nous avons défini notre propre critère. Le taux d'accroissement a été considéré (voir section 1.1.2), à savoir résultant de la différence entre le début et la fin de suivi, divisé par le temps écoulé entre les deux. La notion de répondeur considéré s'appuyait sur cette quantité et une valeur seuil.

Plus précisément, la moyenne des taux d'accroissement a été calculée dans le groupe placebo, au prorata du temps de suivi. Chaque individu possède sa propre évolution et le vieillissement physiologique peut altérer les fonctions cognitives d'un sujet, même en l'absence de démence. De la même façon, il est possible qu'un sujet améliore ses fonctions cognitives pour plusieurs raisons (entrainement cognitif, etc). Les taux du groupe placebo ont servi de référence d'évolution sur les différentes périodes de suivi. Nous avons émis l'hypothèse qu'un sujet hors placebo avait réellement amélioré ses fonctions si son augmentation était supérieure à celle du placebo. Calculer ce taux d'accroissement moyen dans le groupe placebo permet de tenir compte de la variabilité au cours du temps. Ensuite, toujours pour chaque période de suivi, les taux d'accroissement pour tous les autres sujets ont été estimés, de façon individuelle. Ces valeurs ont été comparées à celles de la moyenne du groupe placebo sur la période de suivi correspondante. Un seuil de 20 % par rapport à la moyenne placebo a été déterminé comme critère de répondeur. Bien que subjectif, ce dernier reste cependant en accord avec les définitions proposées dans la littérature. Il regroupait les sujets répondeurs ou non répondeurs à l'intervention multidomaine et/ou au traitement par oméga-3. Le

processus de construction du critère de réponse se décrit sous forme d'algorithme simplifié 6.

Algorithme simplifié 6 :

Choix a priori :

1. seuil du répondeur (20 % par rapport au placebo)
2. critère d'évaluation (taux d'accroissement)

Initialisation : sélectionner la population d'étude (1525 sujets) puis séparer les sujets du groupe placebo et ceux hors placebo

Processus :

1. Calculer les temps de suivi des sujets des différents groupes définis à l'étape précédente (temps écoulé entre les mesures initiale et de fin de suivi) et les trier indépendamment du bras de l'essai :
   — 1 : sujets avec valeurs à T0 et T6 renseignées ;
   — 2 : sujets avec valeurs à T0 et T12 renseignées ;
   — 3 : sujets avec valeurs à T0 et T24 renseignées ;
   — 4 : sujets avec valeurs à T0 et T36 renseignées.
2. Calculer, uniquement pour les sujets placebo, la moyenne des taux d'accroissement pour chaque période définie à l'étape 1.
3. Calculer les taux d'accroissement individuels pour les sujets non placebo sur la période de suivi définie à l'étape 1.
4. Comparer les taux individuels avec la moyenne placebo, au seuil de 20 %, pour la période de suivi correspondante.

Fin Processus

Partitionnement final : une variable binaire a été construite et valait 1 si le sujet était répondeur, 0 sinon (dichotomie de la population).

Notons que cette méthode ne tient pas compte, par définition, des variations d'évolution au cours du suivi. Seules les variations entre début et fin de suivi ont été considérées. Une autre conséquence de ce choix était la taille de l'échantillon d'étude. Nous souhaitions considérer le plus grand échantillon possible, donc l'ensemble des sujets permettant de calculer le critère de répondeurs, au prorata de leur temps de participation. Autrement dit, l'analyse n'a pas été restreinte aux sujets ayant un suivi complet mais à ceux ayant une valeur à *baseline* et en fin de leur suivi (1525 sujets, dont 380 dans le groupe placebo). Seuls les sujets présents uniquement à *baseline* ont été écartés de l'analyse puisqu'ils ne pouvaient montrer une évolution de leurs fonctions cognitives.

Tous les sujets n'appartenant pas au groupe placebo et répondant aux critères listés ci-dessus ont été classés dans les 2 groupes prédéfinis. La population d'étude englobait 1145 sujets, dont 43 sujets ayant développé une démence au cours du suivi. Au total, 733 sujets étaient classés en répondeurs, soit 64 % de la population d'étude. Cela signifie qu'au cours du suivi, plus de la moitié des sujets ont amélioré leur niveau global des fonctions cognitives de 20 % ou plus par rapport à la variation moyenne du groupe placebo. Afin de distinguer les trajectoires individuelles et moyennes des deux groupes, nous avons représenté le *spaghetti plot* des données observées en Figure 6.8. Les sujets sont tous confondus, quelle que soit leur période de participation à l'étude et quel que soit le bras d'intervention.

Pour compléter cette information, les effectifs des deux groupes sont indiqués à chaque date de suivi dans le Tableau 6.4. Les sujets sont présentés indépendamment du bras d'in-

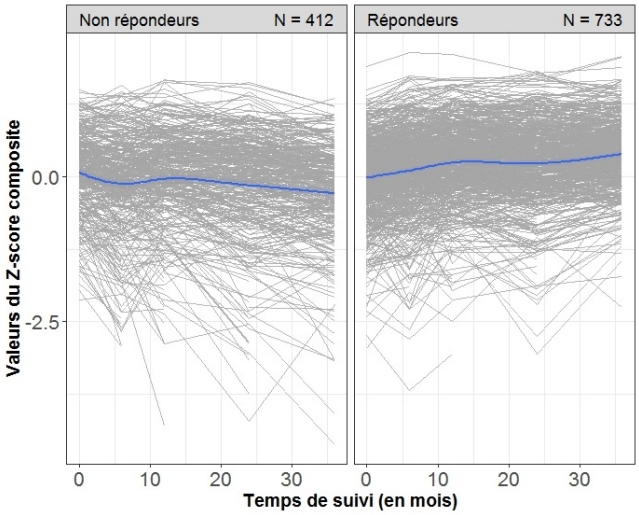

FIGURE 6.8 – *Spaghetti plot* des données observées pour les deux groupes définis par la classification dichotomique sur critère de répondeur, pour les taux d'accroissement du Z-score composite, MAPT (N = 1145 sujets).

tervention de l'essai.

1.2.3 Interprétation clinique

À l'instar des précédentes méthodes, les différents groupes obtenus ont été comparés entre eux selon les caractéristiques relevées à l'inclusion. Le Tableau 6.5 résume les différences significatives. Le premier groupe, correspondant aux non répondeurs, contenait une proportion de sujets déments plus importante que celui des répondeurs (p = 0,001, test du χ^2). Cette proportion représentait 60,5 % de l'ensemble des sujets déments présents dans cette analyse. De plus, leur niveau d'éducation était plus faible (p <= 0,001, test du χ^2) et ils étaient plus âgés (p < 0,001, test de l'ANOVA). Ils se plaignaient davantage de la gêne ressentie dans leurs activités quotidiennes (p = 0,01, test de l'ANOVA) et avaient un score au TMT-B plus long de 9,5 secondes (p = 0,01, test de l'ANOVA).

TABLE 6.4 – Description des 2 groupes identifiés par la classification par statut répondeur, pour les taux d'accroissement du Z-score composite de l'essai MAPT, au cours du suivi (N = 1145 sujets).

Période(1)	Répondeurs		Non Répondeurs	
	Effectif	Pourcentage	Effectif	Pourcentage
T0-T6	53	67,1	26	32,9
T0-T12	65	73,9	23	26,1
T0-T24	55	76,4	17	23,6
T0-T36	560	61,8	346	38,2

(1) : Période d'observation définie en fonction des temps de mesures : T0 = *baseline*, T6 = à 6 mois de suivi, T12 = 12 mois, T24 = 24 mois et T36 = fin de suivi.

Ces deux groupes de sujets ne différaient pas en terme de répartition dans les différents bras de l'essai (p = 0,181, test du χ^2), bien que la proportion de sujets ayant reçu l'IM était plus importante dans le groupe des répondeurs (63,4 % chez les non répondeurs contre

Table 6.5 – Récapitulatif des caractéristiques significativement différentes pour les 2 groupes identifiés par la classification par statut répondeur pour les taux d'accroissement du Z-score composite, MAPT (N = 1145 sujets).

Variables	Modalités	Groupe 1	Groupe 2	P-value
Age		75,91 (4,64)	74,98 (4,23)	<0,001*
TMT-B		127,94 (71,92)	118,46 (55,85)	0,014*
EVA-2		42,22 (22,78)	38,40 (23,05)	0,007*
Etude				<0,001
	1 : Haut	301 (73,24)	572 (80,11)	
	0 : Bas	110 (26,76)	142 (19,89))	
Cholesterol				<0,001
	1 : Oui	150 (37,22)	347 (47,80)	
	0 : Non	253 (62,78)	379 (52,20)	
Démence**				0,001
	1 : Oui	26 (6,31)	17 (2,32)	
	0 : Non	386 (93,69)	716 (97,68)	

Variables quantitatives, moyenne M (écart-type SD) ; Variables qualitatives, effectif N (pourcentage %).
* test de l'Anova ; test du χ^2.
Contrairement aux autres variables, la démence n'est pas relevée à *baseline*.

68,6 % chez les répondeurs). Ces deux groupes se ressemblaient également d'un point de vue du statut phénotypique (p = 0,853, test du χ^2), du sexe (p = 0,170, test du χ^2), du score CDR (p = 0,078, test du χ^2) ou encore de leur IMC (p = 0,152, test du χ^2). Aucun groupe ne se distinguait de l'autre pour la symptomatologie dépressive positive (p = 0,525, test du χ^2), pour le critère de fragilité de Fried (p = 0,067, test du χ^2) ou pour leur hypertension (p = 0,522, test du χ^2).

Par ailleurs, nombre de variables quantitatives ne différaient pas d'après le test de l'ANOVA. Le score CAIDE moyen n'était pas plus élevé dans un groupe par rapport à l'autre (p = 0,457). Le niveau de plainte mnésique restait similaire (p = 0,112) tandis que celui de la gêne ressentie dans les activités de la vie quotidienne était supérieur dans le groupe des non répondeurs (p = 0,007, test de l'ANOVA). De même, le temps d'exécution n'était pas significativement différent pour la version A du TMT (p = 0,512) alors qu'il augmentait d'environ 10 secondes pour sa version B (p = 0,014, test de l'ANOVA). Le nombre de mots acceptés pour le COWAT se valait entre les deux groupes (p = 0,398).

Soulignons tout de même que, par construction de la méthode, l'homogénéité des profils n'était pas assurée. En effet, le statut étant construit sur le taux d'accroissement entre début et fin de suivi, les variations au cours du suivi n'ont pas été prises en compte, ce qui signifie qu'il n'était pas garanti que les profils soient similaires au milieu du suivi. Seules les variations entre début et fin de suivi ont été quantifiées et comparées à celles du groupe placebo. En conséquence, nous avons pu passer à coté d'une période d'amélioration pour les suivi longs. Par exemple, un sujet s'étant amélioré puis déclinant et déclinant encore, pour une période totale de 2 ans de suivi, a peut être été classé en non répondeur. Cependant, puisque le critère est défini au prorata du temps de suivi, nous espérons avoir limité ce dernier point. Cette spécificité complique l'interprétation des résultats car, en plus d'écarter les variations au cours du suivi, tous les répondeurs sont présentés et confondus, quelle que soit la durée de suivi. Or, sur le plan clinique, un répondeur sur 6 mois de suivi n'a pas la même "valeur" qu'un répondeur sur 24 mois de suivi par exemple. Cette distinction n'a donc pas pu être abordée dans les présents résultats.

Enfin, dans les précédentes sections, nous n'avions pas distingué de groupes similaires aux répondeurs. Cette analyse complète donc l'information acquise jusqu'alors. Avec le critère d'une amélioration de 20 % par rapport à la moyenne placebo, nous avons défini un groupe contenant la majorité des déments (60,5 %) et dont le profil d'évolution des fonctions cognitives déclinait légèrement tout au long du suivi. Il ressemblait donc à celui identifié en section 1.1.1 avec la méthode des *k-means*. Ce constat était cohérent cliniquement et attendu implicitement dans la définition du répondeur. Le groupe des non répondeurs représentait 36 % de la population d'étude. La majorité des facteurs de risque de la démence n'était pas significativement différente entre ces groupes. La répartition selon les bras de l'essai était homogène.

2 Résultats de la modélisation des trajectoires cognitives

Dans la seconde partie du chapitre sont abordées deux techniques de modélisation des trajectoires cognitives. Le but recherché était l'estimation des effets de covariables dans un contexte d'apprentissage statistique d'une part (section 2.1), de méthodes markoviennes d'autre part (section 2.2). Dans le premier cas, cet effet est mesuré par des algorithmes "classiques" et d'apprentissage par renforcement, tandis que dans le second, il concerne les probabilités de transition entre les états. Ce processus markovien peut être modélisé en connaissant certains paramètres du phénomène, comme avec des modèles de Markov observables (section 2.2.1), ou dans le but de rechercher un nombre d'états optimal modélisant le phénomène et tenant compte de l'incertitude de mesure, à l'instar des modèles de Markov cachés (section 2.2.2). Dans les deux cas, la trajectoire des fonctions cognitives est assimilée à une succession d'états.

2.1 Résultats des méthodes d'apprentissage statistique

En premier lieu, des stratégies supervisées classiques ont été appliquées (section 2.1.1). L'objectif était d'expliquer, à l'instar des méthodes présentées en section 1 et par différentes covariables relevées dans MAPT, les 4 groupes définis par les bras d'intervention de l'essai. L'objectif secondaire visait à faire des prévisions pour de nouveaux individus, d'après les caractéristiques de ces groupes particuliers. Jusqu'à présent, nous n'avons jamais indiqué aux méthodes que la population était divisée en un groupe d'IM et un sans IM, tous deux en combinaison ou non avec le traitement. Cette variable de groupe permettait de décrire les profils identifiés uniquement. Pour en tenir compte, plusieurs méthodes classiques ont été utilisées successivement, pour complexifier pas à pas les analyses (sections 2.1.1.1 à 2.1.1.4). Par ailleurs, l'algorithme d'apprentissage par renforcement du *Q-learning* a été utilisé pour modéliser les trajectoires cognitives. Les difficultés rencontrés lors de l'application sont détaillées en section 2.1.2.

2.1.1 Méthodes d'apprentissage "classiques"

Parmi les approches classiques d'explication des bras d'intervention et/ou de traitement, l'analyse de la variance (ANOVA) est présentée en section 2.1.1.1. Cette approche

détermine si une association linéaire entre une variable explicative et la variable à expliquer existe. La section 2.1.1.2 présente l'analyse factorielle discriminante. Elle tient compte de plusieurs variables explicatives quantitatives à la fois. Ensuite, sont présentés les arbres de classification et de régression (section 2.1.1.3), permettant d'inclure des variables explicatives qualitatives, ainsi que les forêts aléatoires 2.1.1.4, qui généralisent les résultats des arbres.

2.1.1.1 Analyse de la variance

La première étape d'exploration supervisée consistait à utiliser l'ANOVA [255]. Ce modèle statistique comparait les moyennes du Z-score composite dans le groupe recevant l'intervention multidomaine (IM) contre celui sans IM.

À l'instar de l'analyse préliminaire décrite en section 3.1 du Chapitre 4, l'ANOVA a été appliquée sur les différentes périodes de suivi. L'analyse était ainsi restreinte aux sujets ayant un suivi complet pour le critère de jugement principal, sur les différentes périodes considérées. Ce nombre de sujets est identique à celui présenté dans le Tableau 4.4 de la section 3.1 du Chapitre 4. En considérant l'intégralité du suivi, nous revenions à la population $\mathcal{P}_{comp}$ présentée en section 2.2 du Chapitre 4. Dans un premier temps, l'approche paramétrique était pratiquée. Cela signifie que, pour chaque test, l'hypothèse de normalité se vérifiait graphiquement par des *QQ-plots* et l'hypothèse d'homoscédasticité via le test de Bartlett.

Sur chaque période d'observation, le taux d'accroissement du Z-score composite entre le début et la fin de la période a été calculé. La différence de taux moyens, selon le groupe d'intervention multidomaine, fournissait les p-values suivantes : $p_{T0-T6} = 0{,}052$, $p_{T0-T12} = 0{,}056$, $p_{T0-T24} = 0{,}060$ et $p_{T0-T36} = 0{,}041$. En somme, entre 0 et 24 mois de suivi, l'effet d'intervention était suggéré mais n'apparaissait pas comme statistiquement significatif. En revanche, la différence de taux moyens selon le groupe d'intervention était significative pour la période T0-T36 mois. Recevoir l'intervention impactait le taux d'accroissement moyen du critère de jugement principal de l'essai. Cet effet allait dans le sens que le taux était plus élevé dans le groupe bénéficiant de l'intervention. Notons que ces résultats ne tenaient pas compte des variations au cours du suivi, simplement de la valeur initiale et finale pour chaque période.

Dans cette situation présentant des p-values à la limite de la significativité, nous avons trouvé intéressant de confronter les résultats de l'ANOVA à ceux de la version non paramétrique. Cette dernière permet de s'affranchir des hypothèses précédentes. Le test de Kruskal Wallis a été implémenté. Cette fois-ci, les résultats étaient tous non significatifs. Les p-value valaient $p_{T0-T6} = 0{,}109$, $p_{T0-T12} = 0{,}066$, $p_{T0-T24} = 0{,}144$ et $p_{T0-T36} = 0{,}072$. L'effet précédent n'a donc pas été confirmé.

Par ailleurs, nous avons supposé que l'effet l'intervention devait être dissocié de l'effet traitement. Cela pourrait expliquer la non concordance des résultats des analyses précédentes. Nous avons mené une ANOVA à 2 facteurs, de façon à étudier l'intervention et le traitement séparément, ainsi que leur interaction. Les p-values associées aux différentes périodes sont résumées dans le Tableau 6.6 ci-dessous. Concernant l'intervention, les effets étaient similaires à l'analyse paramétrique. En revanche, aucun effet du traitement seul ou de l'interaction n'était significatif. Le test d'interaction n'était pas significatif pour les périodes étudiées.

TABLE 6.6 – Résumé des effets mesurés par l'ANOVA à deux facteurs, sur des taux d'accroissement entre début et fin de suivi, sur les différentes périodes possibles de l'essai MAPT.

Période(1)	Sujets(2)	Intervention		Traitement		Interaction
		Effectif(3)	P-value	Effectif(4)	P-value	P-value
T0-T6	1502	753	0,052	743	0,619	0,595
T0-T12	1368	692	0,056	678	0,761	0,810
T0-T24	1227	619	0,059	603	0,326	0,588
T0-T36	1143	571	0,041	563	0,932	0,732

(1) : Période d'observation définie en fonction des temps de mesures et pour laquelle les sujets ont un suivi complet : T0 = baseline, T6 = à 6 mois de suivi, T12 = 12 mois, T24 = 24 mois et T36 = fin de suivi ;
(2) : Nombre de sujets n'ayant aucune données manquantes sur le critère de jugement pendant la période considérée ;
(3) : Sujets bénéficiant de l'IM, indépendamment du groupe de traitement ;
(4) : Sujets bénéficiant du traitement par Ω_3, indépendament du groupe d'IM.

2.1.1.2 Analyse factorielle discriminante

La deuxième étape de cette stratégie de modélisation supervisée des groupes d'intervention consiste à prendre en compte plusieurs variables simultanément. L'hypothèse était que certaines d'entre elles influenceraient les effets à la limite de la significativité à l'étape précédente. Pour cela, une approche linéaire multivariée a étudié le lien entre le critère de jugement et les variables de groupes [256]. L'ensemble de la population $\mathcal{P}_{tot}$, présentée en section 2.1 du Chapitre 4 a été considérée et les variables explicatives (Z-score composite ou ses composantes) étaient considérées sans modification préalable.

Lorsqu'elle est utilisée pour décrire une population, l'analyse factorielle discriminante (AFD) n'émet aucune hypothèse a priori. Elle s'utilise principalement pour réduire les dimensions d'un jeu de données ou pour faire de la classification. Elle s'applique en présence d'une variable qualitative et de plusieurs variables quantitatives. Le second objectif de cette méthode peut consister à prédire l'appartenance d'un individu à un groupe prédéfini par la variable qualitative.

À l'image de la méthode décrite en section 2.1.1.1, nous souhaitions discriminer les groupes définis par les bras de l'essai. Quatre analyses ont été menées en parallèle :

— L'explication des 4 bras de l'essai selon
  1. le critère de jugement principal ;
  2. les composantes du critère de jugement principal ;
— L'explication du groupe d'IM selon
  1. le critère de jugement principal ;
  2. les composantes du critère de jugement principal.

La méthode avait un taux d'erreur de 27 % et 30 % pour la discrimination des bras de l'essai. Dans le premier cas, la seule information obtenue était qu'en prédiction, aucun sujet ne se trouvait dans le groupe placebo avec IM. Ce constat n'était pas confirmé par la deuxième analyse, avec les composantes individuelles du critère de jugement principal.

Concernant les groupes d'IM, le taux d'erreur était équivalent à pile ou face. Il valait 51 % pour la troisième analyse et 52 % pour la dernière.

En somme, contrairement à ce que nous avions supposé, les groupes n'ont pu être discriminés dans aucun des quatre cas explorés. Cela a conduit à deux dernières analyses, toujours avec les mêmes variables explicatives. Cette fois-ci, les groupes opposés étaient le

traitement par Ω_3 et l'IM contre le placebo et pas d'IM. Le taux d'erreur n'a pas été amélioré par cette manipulation et valait 53 % dans les deux cas. Comme pour les précédents cas, les groupes ne sont pas apparus comme clairement distincts.

2.1.1.3 Arbres de classification et de régression

Un des points négatifs de la méthode présentée en section 2.1.1.2, outre son incapacité à discriminer nos groupes, est de ne considérer que des variables quantitatives. Or, parmi les variables explicatives, certaines sont qualitatives. De façon à toutes les prendre en compte, les arbres de classification et de régression ont été envisagés [257] (*Classification And Regression Trees* - CART).

Rappelons que le principe de la méthode repose sur un partitionnement, de façon binaire et récursive, de l'ensemble de la population selon différentes variables explicatives. Ensuite, une sous-partition optimale est définie. Deux étapes sont nécessaires : la construction de l'arbre maximal puis son élagage pour construire l'arbre optimal. Plus l'arbre sera élagué haut, plus ses performances statistiques de prédiction seront amoindries. En effet, un arbre non élagué sous-ajuste les données et est soumis à un faible biais, ici le biais désigne le manque de précision du modèle, mais la variance est élevée, c'est-à-dire que l'erreur due aux fluctuations des données est importante. A contrario, un arbre élagué bas sur-ajuste les données. Les deux avantages de cette méthode sont sa facilité d'interprétation et sa large applicabilité. Les inconvénients majeurs proviennent de la variabilité du résultat en fonction de l'échantillon d'analyse et de son mode de fonctionnement. En effet, les choix successifs menant à la partition finale ne sont jamais remis en cause, ce qui signifie qu'une fois une variable choisie, elle ne peut plus être enlevée de l'arbre.

Trois analyses ont été investiguées. La première expliquait les 4 bras de l'essai. La seconde s'intéressait uniquement au groupe d'IM. La dernière portait sur l'opposition entre traitement par Ω_3 avec IM et placebo sans IM.

Dans les trois cas, nous avons obtenu un arbre de classification pour l'ensemble de la population $\mathcal{P}_{tot}$, présentée en section 2.1 du Chapitre 4. Seulement, les seuils proposés pour dichotomiser les variables n'avaient pas d'interprétation clinique. Il n'existait aucune configuration des variables nous amenant à une discrimination des groupes étudiés interprétable cliniquement. Ces arbres n'étaient pas exploitables en l'état et n'ont donc pas été élagués.

2.1.1.4 Forêt aléatoire

Finalement, la dernière étape de la stratégie d'apprentissage des trajectoires cognitives visait à appliquer l'algorithme des forêts aléatoires (*Random Forest* - RF). Il est fondé sur les arbres présentés en section 2.1.1.3. Sa particularité consiste à construire des arbres (CART) en introduisant des perturbations aléatoires et à combiner les arbres obtenus [258]. Le résultat obtenu est ainsi plus stable que celui des arbres. Cette méthode non paramétrique est très performante, aussi bien pour la prédiction qu'en terme de temps d'exécution.

Les *Random Forest* ont été exploitées de deux façons : la première avec les 4 bras de l'essai, la seconde avec le groupe d'IM. Dans les deux cas, l'ensemble de la population a été considérée ($\mathcal{P}_{tot}$, présentée en section 2.1 du Chapitre 4). Seulement, comme pour les précédentes méthodes, le taux d'erreur n'était pas acceptable. Il s'élevait à 75,4 % dans le premier cas, ce qui signifie que la prédiction du groupe pour tout nouveau sujet intégrant

l'étude avec la forêt proposée n'est trouvé qu'une fois sur quatre, et à 51,2 % dans le second. Suite à quoi, nous avons réitéré l'analyse avec l'opposition entre traitement par Ω_3 avec IM contre placebo sans IM. Cependant, le taux d'erreur n'a pas été amélioré par cette manipulation, bien qu'il se situait alors autour de 47,9 %.

2.1.2 Méthode d'apprentissage par renforcement

Dans cette analyse, la population a été restreinte aux sujets ayant plus de 2 visites au cours du suivi, soit sur 1525 sujets (voir $\mathcal{P}_{mapt}$, présentée en section 2.3 du Chapitre 4). Le Z-score composite en fin de suivi a directement été considéré, sans modification préalable. Dans un premier temps, les variables significatives en analyse univariée et les interactions médicalement pertinentes ont été testées selon une stratégie de sélection classique.

D'après la configuration de l'étude, nous avons présenté, en section 2.1.1 du Chapitre 5, des règles de décisions en deux étapes. Certains paramètres de la règle de décision apparaissaient comme évidents pour notre étude (voir Figure 5.5 de la section 2.1.1 du Chapitre 5). Pour commencer, la variable X_1 contenait toutes les données à *baseline*. Ensuite, Y correspondait à la valeur du critère de jugement principal à un moment donné. En particulier, il représentait la fin du suivi, la fin de la période intensive de l'intervention (6 mois) et la fin de la première année. La décision A_1 équivalait à la randomisation des sujets dans les différents bras de l'essai. De façon à avoir une décision binaire, nous avons considéré d'une part le groupe placebo sans IM ($A_{11} = 0$) versus les autres groupes ($A_{11} = 1$) ; d'autre part le groupe d'IM ($A_{12} = 1$) versus sans intervention ($A_{12} = 0$), indépendamment du traitement. La composante X_2 regroupait les mesures longitudinales entre A_1 et A_2.

La principale difficulté d'application de la méthode consistait à identifier la décision A_2. Dans l'essai étudié et après la randomisation, les patients étaient suivis au cours du temps, sans modification du traitement ou de l'intervention qui leur avait été attribué à *baseline*. Il n'y avait pas de seconde randomisation du bras de l'essai ou de quelconque décision impactant les fonctions cognitives des sujets. Le paramètre A_2 n'existait pas. Pour adapter les données à cette méthode, il aurait fallu créer une décision A_2. Par exemple, la randomisation dans les bras de l'essai a été décomposée en une décision pour le groupe d'intervention et l'autre dans le groupe de traitement. Nous aurions également pu évaluer l'adhérence au traitement ou la perte d'un certain pourcentage des fonctions cognitives. Le problème résultant concernait la temporalité de cette seconde décision et son sens clinique.

N'étant pas en capacité de définir artificiellement une décision cliniquement pertinente, nous nous sommes abstenus. Son impact sur les résultats était trop important pour être négligé. Les cas présentés ci-dessus, avec différentes valeurs de Y et la décision A_1, ont été modélisés. Seulement, l'application d'une méthode de *Q-learning* pour estimer la règle de décision optimale en une étape était équivalente à une régression linéaire classique. Or, cette dernière ne tient pas compte du caractère longitudinal des données, c'est pour cela que les modèles mixtes sont habituellement employés pour l'analyse des essais. Nous ne reviendrons pas sur les motifs de la non utilisation de ces modèles dans le cadre du présent travail. Ainsi, le seul intérêt du *Q-learning* était d'observer les répartitions optimales dans les différents groupes. Il était également intéressant de regarder le groupe prédit pour un sujet donné et certaines de ses caractéristiques. Au terme de cette analyse et malgré sa séduisante théorie, cette méthode n'était pas adaptable à nos données. Ces dernières n'ont pas été pensées pour cette utilisation, rendant impossible l'exploitation des résultats.

2.2 Résultats des méthodes markoviennes

Cette dernière section est divisée en deux, d'une part l'approche par modèles de Markov et d'autre part avec les modèles de Markov cachés. Dans les deux cas, excepté pour la section 2.2.1.2 dont la population contenait 19 sujets supplémentaires, la population d'étude correspondait aux sujets ayant au moins une visite post-*baseline*. La population $\mathcal{P}_{mapt}$ était présentée en section 2.3 du Chapitre 4. De plus, comme indiqué dans le chapitre précédent, les analyses markoviennes nécessitent de définir un état de santé à chaque visite. La donnée fonctionnelle quantitative initiale est ainsi redéfinie en une donnée fonctionnelle qualitative. La succession d'état est définie, en chaque temps t, à partir de différents seuils de valeurs du Z-score composite, faisant référence à des comportements cliniques particuliers.

2.2.1 Les modèles de Markov homogènes

De façon à modéliser les trajectoires cognitives, le package `msm` [222] de R a été utilisé. Il nécessite la transformation de la base de données dans un format dont les colonnes correspondent aux covariables et les lignes aux sujets et aux temps de mesure. D'après les données de MAPT et pour tenir compte des dates de visites variables d'un sujet à l'autre, le modèle de Markov à temps continu a été jugé plus pertinent pour notre étude. Plus particulièrement, des modèles de Markov homogènes d'ordre 1 ont été implémentés. Cela signifie que les transitions d'un état à un autre ne dépendaient pas du temps, tout comme les états ne dépendaient que de l'état précédent du système d'après l'hypothèse de Markov. Rappelons qu'une telle modélisation vise à explorer le risque d'évolution des fonctions cognitives et comment il est affecté par des covariables explicatives.

Plus particulièrement, le modèle de la section 2.2.1.1 oppose un état de déclin des fonctions cognitives à toute autre évolution possible. En section 2.2.1.2 est ajouté un troisième état pour les données manquantes. Une redéfinition d'un modèle à trois états, avec des variations strictes du Z-score composite, est implémenté en section 2.2.1.3. Pour cela, un état correspond à un score strictement égal au précédent, au autre à un score strictement supérieur et un dernier strictement inférieur. Finalement, le dernier modèle abordé est similaire au précédent, avec des variations plus souples (section 2.2.1.4). L'état intermédiaire correspond cette fois à un intervalle autour de la précédente valeur, c'est en cela que la définition est plus souple.

2.2.1.1 Modèle à 2 états

2.2.1.1.1 Interprétation statistique

Dans un premier temps, nous avons implémenté le modèle le plus simple possible, faisant fi des résultats de la section 1. L'hypothèse associée était que les observations pouvaient être modélisées par un modèle de Markov en 2 états (voir Figure 6.9). Le premier a été construit pour correspondre à des fonctions cognitives supérieures ou égales à celles mesurées lors de la dernière visite. Le second état renvoyait à un déclin strict des fonctions, toujours par rapport à la précédente valeur du Z-score composite. Le modèle de Markov se composait de degrés de sévérité ordonnés d'atteinte des fonctions cognitives. Cela signifie qu'il était préférable pour un sujet d'être dans l'état 1 que dans l'état 2. Cette conceptualisation du niveau cognitif est semblable à celle utilisée dans la section 3.1 du Chapitre 4 sur les profils extrêmes. Cela revient à opposer un comportement attendu cliniquement, le

déclin des fonctions cognitives, à tout autre comportement.

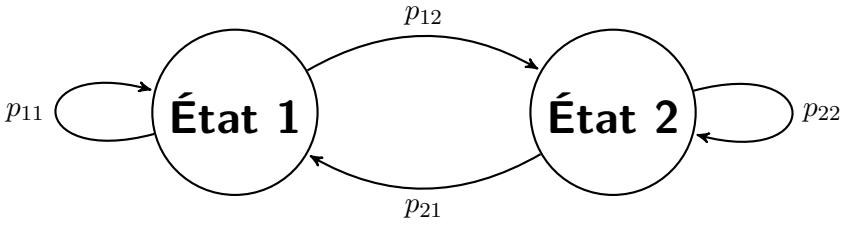

FIGURE 6.9 – Représentation d'un modèle de Markov observable à 2 états. Les p_{ij} correspondent aux probabilités de transitions pour aller de l'état i à l'état j.

D'après les critères d'inclusion de l'essai, les sujets de MAPT étaient supposés être à risque de décliner cognitivement ou de développer la maladie d'Alzheimer. Il était attendu qu'un déclin cognitif soit observé, plus ou moins tardivement, au cours du suivi. Nous avons considéré la population d'étude comme cognitivement stable à l'inclusion, ce qui implique que tous les sujets commençaient la chaîne dans l'état 1. Les probabilités initiales des différents états ont été choisies de la sorte :
— $\pi_1 = 1$
— $\pi_2 = 0$

Pour chaque visite après inclusion, les états ont été définis par rapport à la valeur précédente du Z-score composite. Cette valeur était individuelle de façon à tenir compte de l'évolution particulière de la maladie, qui est propre à chaque sujet et progressive au cours du temps, ce qui signifie qu'il n'était pas cohérent de définir un seuil unique pour l'ensemble de la population d'étude. Si cette valeur individuelle du Z-score composite était supérieure ou égale à la précédente, le sujet restait dans l'état 1. Sinon, le sujet changeait d'état vers le 2. En somme, l'état à la première visite après *baseline*, notée T_{1i} pour chaque sujet i, se résumait de façon binaire comme :
— $T_{1i} = 1$ si $Z_{1i} \geq Z_{0i}$
— $T_{1i} = 2$ si $Z_{1i} < Z_{0i}$
Les valeurs du Z-score composite en début de chaîne, pour les 1525 sujets analysés, conduisait à un population à 51,6 % dans l'état 1 et 46,9 % dans l'état 2. Les 23 sujets restants (soit 1,5 %) avaient un état manquant car leur Z-score composite l'était également à cette date. Un procédé identique a été appliqué pour définir les états des trois autres temps de mesure. Chaque sujet possédait une suite de 5 états au cours du suivi, un initial égal pour tous les sujets et un par date de visite après inclusion.

À ce stade, il était possible de décrire simplement l'ensemble des transitions observables entre les états dans la population pour l'intégralité du suivi. Elles sont indépendantes de toute covariable et purement descriptives. Le résumé des transitions observées, correspondant aux p_{ij} de la Figure 6.9, peut se faire sous forme de matrice :

$$M = \begin{pmatrix} 1528 & 1988 \\ 1215 & 563 \end{pmatrix} \quad .$$

La chaîne parcourait ainsi tous les états possibles, avec cependant une plus faible proportion de transitions de l'état 2 vers lui-même. La matrice M contenait les nombres de transitions observées d'un état vers un autre dans l'ensemble de la population, dont le nombre total vaut théoriquement 1525 x 5 = 7625.

Afin d'étudier l'impact de covariables sur les transitions d'un état à l'autre (p_{12} et p_{21} de la Figure 6.9), un modèle à risques proportionnels pour les intensités de transitions a été défini. La matrice Q_2 résume les transitions dans le modèle. Aucune contrainte clinique n'a été imposée sur celle-ci, ce qui signifie que toutes les transitions restaient possibles. Avec le package `msm`, les transitions autorisées sont indiquées par 1 et les autres par 0. La matrice du modèle sur les intensités de transitions choisie était donc :

$$Q_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

En dernier lieu, l'argument `gen.inits=TRUE` du package `msm` a été choisi à chaque étape de l'analyse. Il a pour fonction de générer aléatoirement les risques de base des modèles des intensités de transition autorisées par la matrice Q_2. Cela évitait d'émettre un a priori sur les intensités initiales.

Comme indiqué, en section 2.2.1 du Chapitre 5, le modèle sans covariable (modèle vide) a d'abord été construit sur les 1525 sujets. Son AIC s'élevait à 7442,34. Les covariables ont été ajoutées de façon spécifique pour chaque transition autorisée par Q_2, en appliquant la stratégie de sélection des variables décrite en section 2.2.1 du Chapitre 5, en commençant par la transition de l'état 1 vers 2. Le modèle homogène final avait un AIC de 7136,77. Il était donc préférable au modèle vide et se composait de 7 variables. Les résultats sont présentés dans le Tableau 6.7 et sur la Figure 6.10.

Table 6.7 – Récapitulatif des risques relatifs de transition du modèle homogène à 2 états (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets).

Variables	1 vers 1		1 vers 2		2 vers 1		2 vers 2	
Baseline	-0,246		0,246		0,263		-0,263	
	(-9,389	-0,006)	(0,004	9,389)	(0,007	10,055)	(-10,055	-0,007)
Etude			1,022		-			
			(0,893	1,170)				
Age			1,023		-			
			(1,010	1,036)				
CDR			1,069		-			
			(0,953	1,198)				
IMC			0,970		-			
			(0,837	1,125)				
GDS			1,001		-			
			(0,861	1,164)				
ADCS			1,002		-			
			(0,989	1,014)				
Hypertension			1,001		-			
			(0,893	1,122)				

Transition 1 vers 1 : de l'état non déclin vers lui-même ;
Transition 1 vers 2 (resp. 2 vers 1) : entre l'état non déclin vers déclin (resp. déclin vers non déclin) ;
Transition 2 vers 2 : de l'état déclin vers lui-même.

Chaque variable du modèle était significative au test du rapport de vraisemblance durant la stratégie de sélection des variables. Cette dernière a été choisie comme ascendante en l'absence de critères tels que la p-value. Elles apportaient donc toutes une information supplémentaire mais cela n'impliquait pas que leur effet était significatif. Au final, le seul effet statistiquement significatif était celui de l'âge. Les autres variables représentaient des facteurs de confusion dans la relation décrite par les modèles régissant les transitions.

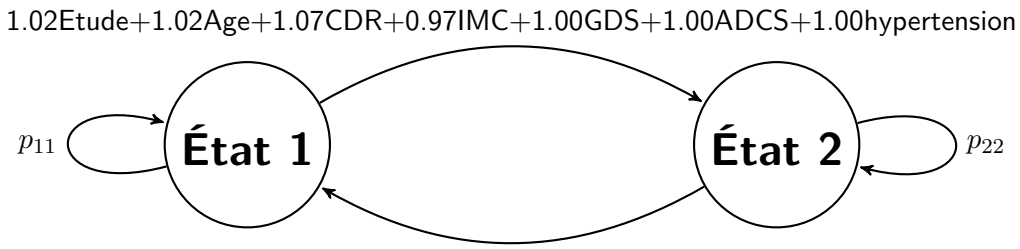

FIGURE 6.10 – Représentation du modèle de Markov homogène à 2 états après ajustement sur les covariables, MAPT (N = 1525 sujets).

2.2.1.1.2 Interprétation clinique

D'un point de vue clinique, revenons rapidement sur le nombre de transitions de la matrice M. Nous avons indiqué que peu de sujets étaient concernés par la transition de l'état 2 vers lui-même. Cela pourrait sous-entendre que le déclin cognitif n'était que passager dans cette population en ce sens qu'il se manifesterait de façon périodique mais pourrait être contrecarré. De plus, un effet bénéfique de l'intervention ou du traitement était envisageable. L'ajustement du modèle avec les différentes covariables permet de vérifier la véracité de ces premières impressions.

Les résultats du Tableau 6.7 ont été reproduits sur la Figure 6.10. Ils s'interprètent de façon classique, comme en analyse de survie. Autrement dit, les coefficients à *baseline* correspondaient aux risques relatifs ajusté sur les autres variables. Il s'interprète en sachant que les variables quantitatives sont égales à la valeur moyenne et que les variables qualitatives sont dans leur catégorie de référence. Dans le cas présent, ils représentaient les risques relatifs sachant que le sujet a 76 ans, un bas niveau d'études, un CDR nul, un IMC inférieur ou égal à 30 kg.m^{-2}, une symptomatologie dépressive négative, une autonomie de 40 points et pas d'hypertension.

Pour ce qui est de l'âge, il s'est avéré que plus un sujet avait un âge élevé, plus il était à risque d'effectuer une transition vers l'état de déclin des fonctions cognitives. Ce risque augmentait de 2,3 % par année supplémentaire (RR = 1,02, IC95% = [1,01 ; 1,04]).

Concernant les autres risques relatifs, ils n'étaient pas statistiquement significatifs. Cependant, si la tendance avait été significative, leur interprétation aurait été dans le sens attendu cliniquement (voir sections 1.3.2 et 1.3.3 du Chapitre 3 sur le référentiel des variables de MAPT). En effet, avoir un haut niveau d'étude augmentait le risque de décliner cognitivement. Rappelons que ce facteur retarde le diagnostic, mais une fois le déclin avéré, la maladie est généralement plus avancée chez les sujets de haut niveau d'études. De même, avoir un CDR non nul augmentait le risque de décliner, ce qui est cohérent par construction du critère. Le fait d'être obèse protègerait du risque de décliner cognitivement. Ce résultat est cohérent car l'IMC est inversement corrélé à la démence sur le long terme [180]. La présence d'une symptomatologie dépressive augmentait également le risque de décliner, tout comme l'hypertension. En revanche, une augmentation du score d'autonomie faisait accroître le risque de décliner. Ce dernier point est surprenant car augmenter son niveau d'autonomie devrait être associé à un moindre déclin cognitif.

Par ailleurs, le bras d'intervention ne comptait pas au nombre des facteurs d'ajustement du modèle, donc son effet n'était pas significatif. Ce facteur n'apportait ainsi aucune information sur le risque d'évolution des fonctions cognitives. La première impression au

regard du nombre de transitions de la matrice M était réfutée : le déclin cognitif pouvait être passager dans cette population, mais le traitement ou l'intervention n'impactaient pas le niveau cognitif global, tout au moins avec les états présentement définis.

En somme, cette analyse nous informe que les sujets à risque de décliner cognitivement (transition de l'état 1 vers l'état 2) étaient ceux dont l'âge augmentait. Ce résultat n'a rien de surprenant et est parfaitement démontré, aussi bien dans la littérature que dans les analyses de la section 1.

2.2.1.2 Modèle à 3 états, avec état manquant

2.2.1.2.1 Interprétation statistique

Dans la section 2.2.1.1, certains sujets ayant des données manquantes au cours du suivi ont été écartés des analyses. Cette spécificité des suivis longitudinaux n'a pas été prise en compte, seuls les états non manquants ont été considérés. Sur les 154 sujets concernés par ce cas de figure de données manquantes, 19 d'entre eux possédaient des dates de visite au cours du suivi. Les ajouter à la population d'étude augmentait le nombre de sujets analysés à 1544 et permettait d'étudier l'importance des perdus de vue au cours du suivi. Pour se faire, un troisième état, équivalent à une donnée manquante, a été construit. Il s'ajoutait aux deux états précédents présentés en section 2.2.1.1. Étudier l'impact de covariables sur les transitions menant à cet état particulier de manquant aideraient les cliniciens à comprendre les sorties d'études, qu'elles soient intermittentes ou définitives. L'objectif de considérer les transitions vers les deux autres états restait identique à l'analyse de la section 2.2.1.1.

Pour ce modèle de Markov homogène d'ordre 1, rappelons que l'état 1 correspondait à un niveau cognitif supérieur ou égal au précédent. L'état 2 renvoyait à un déclin strict du niveau global, toujours par rapport à la précédente valeur. Le dernier regroupait les données manquantes au cours du suivi. En somme, l'état à la première visite après *baseline*, notée T_{1i} pour chaque sujet i, se résumait comme :
— $T_{1i} = 1$ si $Z_{1i} \geq Z_{0i}$
— $T_{1i} = 2$ si $Z_{1i} < Z_{0i}$
— $T_{1i} = 3$ si Z_{1i} manquant
Un procédé identique a été appliqué pour définir les états des trois autres temps de mesure.

Concernant les probabilités initiales des différents états, l'hypothèse de la section 2.2.1.1 a été réitérée, conduisant à la distribution suivante :
— $\pi_1 = 1$
— $\pi_2 = 0$
— $\pi_3 = 0$

L'ensemble des transitions entre les états observables, dans la population et pour l'intégralité du suivi, se résumait par la matrice M ci-après :

$$M = \begin{pmatrix} 1589 & 1591 & 198 \\ 794 & 1320 & 211 \\ 34 & 45 & 394 \end{pmatrix} .$$

La chaîne passait par tous les états possibles, avec cependant une plus faible proportion de transitions en lien avec l'état 3.

De façon à étudier l'impact de covariables sur les modèles à risques proportionnels sur chaque transition autorisée, aucune contrainte clinique n'a été imposée sur une matrice des

risques Q_2, autorisant toutes les transitions possibles. Elle était définit par :

$$Q_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

En dernier lieu, l'argument `gen.inits=TRUE` du package `msm` a été utilisé à chaque étape de l'analyse pour générer aléatoirement les risques des intensités de transition initiales.

La stratégie d'analyse de la section 2.2.1 du Chapitre 5 a été appliquée. Le modèle à vide a été implémenté pour les 1544 sujets et possédait un AIC de 7723,50. Les covariables ont été ajoutées spécifiquement par transition, en considérant celle de l'état 1 vers 2 en premier lieu. Le modèle homogène final avait un AIC de 7236,55. Les 8 variables le composant apportaient une information importante pour expliquer le risque de transition. Les Tableaux 6.8 et 6.9 ainsi que la Figure 6.11 résument les résultats obtenus.

Finalement, seule la variable d'ajustement de la plainte mnésique avait un effet statistiquement significatif. Les autres variables représentaient des facteurs de confusion des modèles régissant les transitions.

2.2.1.2.2 Interprétation clinique

En s'intéressant particulièrement à la matrice de transition M, peu de sujets transitaient de l'état 3 vers l'état 1 ou du 3 vers 2. Cela signifie que lorsqu'une valeur était manquante au cours du suivi, la probabilité qu'elle reste manquante par la suite était plus grande que celle de transiter vers l'état de déclin cognitif, elle-même plus grande que de parcourir l'état des autres comportements d'évolution des fonctions cognitives. Dans le sens contraire, il y avait moins de sujets passant directement d'une amélioration cognitive à une donnée manquante ou d'un déclin à une donnée manquante. Dans l'article MAPT [13], les auteurs ont émis l'hypothèse que les données manquantes étaient aléatoires (MAR), ce qui semble être illustré par ces transitions.

Les résultats des Tableaux 6.8 et 6.9 s'interprètent comme en analyse de survie classique. Les coefficients à *baseline* correspondaient aux risques relatifs de transiter entre les états sachant que les variables quantitatives sont égales à la valeur moyenne. Dans le cas présent, ils s'interprétaient sachant que le sujet avait 76 ans, un bas niveau d'études, un CDR nul, un IMC inférieur ou égal à 30 kg.m^{-2}, une symptomatologie dépressive négative, une absence de plainte mémorielle et pas d'hypertension.

En particulier, ce modèle indiquait que plus la plainte mnésique ressentie par le sujet était importante, plus il était à risque d'effectuer une transition de l'état de déclin des fonctions cognitives vers des données manquantes. Ce risque augmentait de 11,2 % par millimètre de l'échelle EVA-1 supplémentaire (RR = 1,11, IC95% = [1,02 ; 1,21]).

Les autres risques relatifs n'étaient pas statistiquement significatifs. Il n'y avait donc aucun effet du bras de l'essai puisque cette variable n'a pas été intégrée au modèle, pour des raisons de non significativité. En revanche, pour les autres facteurs non significatifs du modèle, avoir un haut niveau d'éducation protégeait du déclin cognitif. De même, avoir un CDR non nul augmentait le risque de décliner tandis que l'obésité protégeait de ce risque. Ces trois résultats concordaient avec ceux constatés dans le modèle de la section 2.2.1.1, sans état manquant. L'hypertension augmentait le risque de transiter de l'état de déclin vers les autres comportements, assimilés à une amélioration ou stabilité des fonctions cognitives.

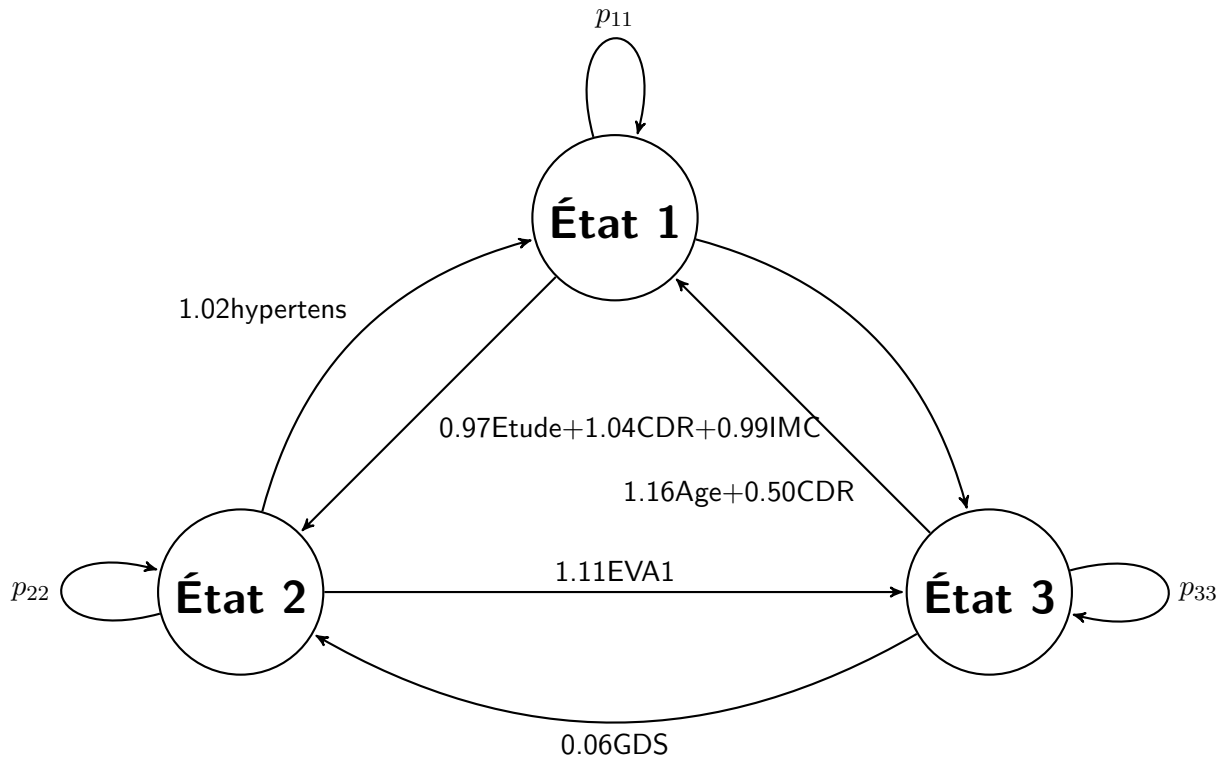

FIGURE 6.11 – Représentation du modèle de Markov homogène à 3 états (dont l'état 3 pour données manquantes) après ajustement sur les covariables, MAPT (N = 1544 sujets).

C'est en effet un facteur de risque du déclin cognitif, mais cela laisse supposer un déclin non linéaire, par paliers par exemple. Ce résultat pourrait donc être expliqué par des profils tels que ceux détectés dans en section 3.2 du Chapitre 4 par les analyses préliminaires sur l'ampleur des variations.

Par ailleurs, une difficulté d'interprétation est survenue pour les transitions de l'état manquant vers les autres états. En effet, la raison de ces données manquantes est inconnue. Toutes les causes sont confondues et considérées comme équivalentes. Le décès, les données intermittentes, les perdus de vue, les intolérances et autres causes ont été traitées de la même façon. Or, cliniquement, la nature des données manquantes est importante. Pour approfondir ces résultats et avoir une vision nette du phénomène étudié, il faudrait définir précisément un état par cause considérée. Dans le cas présent, nous avons simplement conclu que les transitions vers cet état de manquant renforçaient l'idée initiale d'une hétérogénéité des données, y compris au sein des données manquantes.

En somme, cette analyse montrait que les sujets à risque de présenter des données manquantes après une phase de déclin cognitif (transition de l'état 2 vers l'état 3) ressentaient une plainte mnésique plus importante que les autres.

2.2.1.3 Modèle à 3 états, avec variations strictes

2.2.1.3.1 Interprétation statistique

Par la suite, seuls les sujets ayant au moins 2 valeurs du Z-score composite au cours du suivi ont été considérés, soit 1525 sujets (voir $\mathcal{P}_{mapt}$ section 2.3 du Chapitre 4). De plus, la définition des états de degrés de sévérité d'atteinte des fonctions cognitives a été affinée.

L'hypothèse émise était que le modèle de Markov modélisé était constitué de 3 états de niveau cognitifs. Par rapport à l'analyse de la section 2.2.1.1, l'objectif consistait à dissocier pour mieux définir les comportements non attendus cliniquement dans différents états. Pour cela, l'état de déclin des fonctions cognitives était inchangé (état 3 de la Figure 6.12). Ensuite, un état de stabilité stricte a été défini (état 2 de la Figure 6.12). Il représentait une égalité stricte du niveau global des fonctions cognitives par rapport à la dernière valeur. Le dernier état renvoyait à une augmentation stricte des fonctions cognitives, toujours par rapport à la valeur précédente.

L'état à la première visite après *baseline* de cette chaîne de Markov homogène d'ordre 1, notée T_{1i} pour chaque sujet i, se résumait comme :
— $T_{1i} = 1$ si $Z_{1i} > Z_{0i}$
— $T_{1i} = 2$ si $Z_{1i} = Z_{0i}$
— $T_{1i} = 3$ si $Z_{1i} < Z_{0i}$
Après initialisation, la répartition de la population dans le premier état de la chaîne correspondait à 51,0 % dans l'état d'augmentation des fonctions cognitives, 46,9 % dans l'état de déclin, 0,6 % dans l'état de stabilité stricte et 1,5 % manquant. Un procédé identique a été appliqué pour définir les états des trois autres temps de mesure. La faible proportion de sujets ayant une état de stabilité cognitive n'était pas cohérente avec les analyses antérieures et peut traduire une définition non adaptée de l'état "strict".

L'hypothèse a priori de répartition initiale des sujets a été maintenue, ce qui signifie que les sujets étaient supposés être cognitivement stables en début de suivi, impliquant la distribution suivante :
— $\pi_1 = 0$
— $\pi_2 = 1$
— $\pi_3 = 0$

Par ailleurs, la matrice M des intensités de transitions observées se résumait comme suit :

$$M = \begin{pmatrix} 688 & 24 & 1263 \\ 805 & 11 & 725 \\ 1196 & 19 & 563 \end{pmatrix} \quad .$$

La chaîne parcourait tous les états possible, bien que le nombre de transitions vers l'état 2 soit plus faible que les autres.

Enfin, le dernier choix à faire a priori concerne la matrice Q_2 des risques relatifs du modèle. Aucune contrainte clinique n'a été imposée sur celle-ci mais une contrainte pratique : celle de la définition du modèle vide lorsque toutes les transitions étaient autorisées. Lorsque nous indiquions toutes les transitions possibles, la matrice hessienne n'était pas définie positive, donc le modèle ne convergeait pas. Nous avons testé toutes les transitions permettant d'obtenir un modèle vide convergeant tout en réduisant au minimum le nombre de transitions. Cela à conduit à la matrice Q_2 ci-dessous.

$$Q_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad .$$

L'argument `gen.inits=TRUE` du package `msm` a été utilisé à chaque étape de l'analyse pour générer aléatoirement les intensités de transition initiales.

En dernier lieu, la stratégie d'analyse de la section 2.2.1 du Chapitre 5 a été suivie. Le

modèle à vide a été implémenté pour les 1525 sujets et possédait un AIC de 7988,50. Les covariables ont été ajoutées spécifiquement par transition, en considérant celle de l'état 1 vers 2 en premier lieu. Le modèle homogène final, préférable à celui sans ajustement, avait un AIC de 7786,31. Seules 3 variables ont été retenues pour l'ajuster. Les Tableaux 6.8 et 6.11 ainsi que la Figure 6.12 résument les résultats obtenus.

Au final, seul l'ajustement sur l'âge du sujet avait un effet statistiquement significatif dans le présent modèle de Markov homogène d'ordre 1. Les autres variables représentaient des facteurs de confusion dans la relation modélisée par les intensités de transitions. Cependant, les intervalles de confiances de certains risques relatifs sous-entendaient que le modèle n'était pas cohérent et que les coefficients étaient difficilement estimables.

2.2.1.3.2 Interprétation clinique

En s'attardant sur la matrice des transitions M, remarquons que le nombre de transition de l'état de stabilité vers le déclin cognitif était important. Il était gênant de ne pas l'autoriser, donc que le modèle ne la considère pas possible. Le fait d'avoir très peu de transitions vers l'état 2 dans la matrice M, alors que les autres étaient en nombre raisonnable, suggérait certainement que les définitions des états n'étaient pas adaptées.

Par ailleurs, l'effet de l'âge sur la transition de l'état de déclin strict vers l'amélioration stricte était statistiquement significatif. Plus un sujet vieillissait, moins il était à risque d'effectuer une transition vers l'amélioration des fonctions cognitives. Ce risque diminuait de 2,21 % par année supplémentaire (RR = 0,98, IC95% = [0,97 ; 0,99]). Dit d'une autre façon, vieillir empêchait l'amélioration des fonctions. Ce résultat est concordant avec celui du modèle de la section 2.2.1.1 dans lequel l'âge était facteur de risque du déclin des fonctions cognitives.

Concernant les deux autres facteurs d'ajustement du modèle, aucun n'était significatif. Cependant, avoir un niveau d'études élevé était favorable à une stabilité cognitive et augmenter son niveau d'autonomie protégeait du déclin strict de ces fonctions. Ils n'étaient donc pas contradictoires avec les éléments fournis dans le référentiel des variables de MAPT (voir sections 1.3.2 et 1.3.3 du Chapitre 3).

En somme, ce modèle n'apportait pas vraiment de résultats supplémentaires par rapport à celui de la section 2.2.1.1, si ce n'est qu'il confirme que l'âge était un facteur de risque de décliner cognitivement. La définition des états pourrait expliquer, en partie, que le modèle n'apporte rien de nouveau.

2.2.1.4 Modèle à 3 états, avec définitions souples

2.2.1.4.1 Interprétation statistique

Suite aux précédents résultats de la section 2.2.1.3 et à leur pertinence clinique, les définitions des états ont été modifiées. En effet, il est pratiquement impossible d'avoir strictement le même score cognitif au cours d'un suivi, ce qui signifie que l'état de stabilité n'était pas réaliste. Pour tenir compte de la variabilité individuelle de la mesure cognitive, cet état a été défini à partir d'un intervalle symétrique autour de la précédente valeur. Plus précisément, l'état à la première visite après *baseline* de ce modèle de Markov homogène d'ordre 1, notée T_{1i} pour chaque sujet i prenait les valeurs suivantes :

— $T_{1i} = 1$ si $Z_{1i} > Z_{0i} + 30\%$

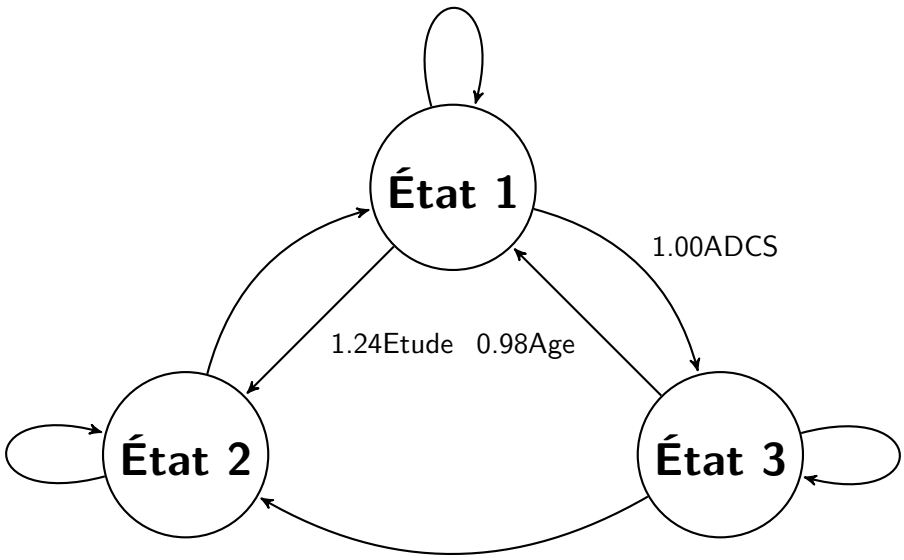

FIGURE 6.12 – Représentation du modèle de Markov homogène à 3 états de variations strictes après ajustement sur les covariables, MAPT (N = 1525 sujets).

— $T_{1i} = 3$ si $Z_{1i} < Z_{0i} - 30\%$
— $T_{1i} = 2$ sinon ($Z_{1i} \in Z_{0i} \pm 30\%$)

Le point clé de la définition des états repose sur le choix du pourcentage à considérer. N'ayant aucun a priori sur celui-ci, nous avons essayé plusieurs valeurs en commençant à 10 % avec un pas de 10 %. Le premier seuil fournissant un modèle convergeant s'élevait à 30 %, d'où la définition ci-dessus. Au final, ce modèle se rapproche de ce que nous aurions pu définir avec les résultats de la section 1 sur les analyses exploratoires.

Les probabilités de répartition initiales étaient identiques à celles du modèle précédent en section 2.2.1.3, soit :
— $\pi_1 = 0$
— $\pi_2 = 1$
— $\pi_3 = 0$

Concernant la matrice M des intensités de transitions observées, indépendamment de toute covariable, nous avions :

$$M = \begin{pmatrix} 543 & 374 & 786 \\ 878 & 715 & 740 \\ 921 & 76 & 261 \end{pmatrix} \quad .$$

La chaîne parcourait tous les états possibles, de façon plus équilibrée qu'avec les définitions de la section 2.2.1.3. Cela confirme que ces dernières n'étaient pas adaptées dans le sens où certaines transitions étaient peu présentes. Toutefois, le nombre de transitions vers l'état 3 vers l'état 2 était plus faible que les autres.

La modification des définitions des états, par rapport au modèle précédent, impactait la matrice Q_2 des risques relatifs modélisant les intensités de transition. Cette fois, la transition impossible du modèle de la section 2.2.1.3 était possible (de l'état 2 vers l'état 3). Aucune contrainte clinique n'a été imposée sur celle-ci et aucune contrainte pratique n'était nécessaire. Ainsi, la matrice Q_2 ne contenait que des 1. L'argument `gen.inits=TRUE` du pa-

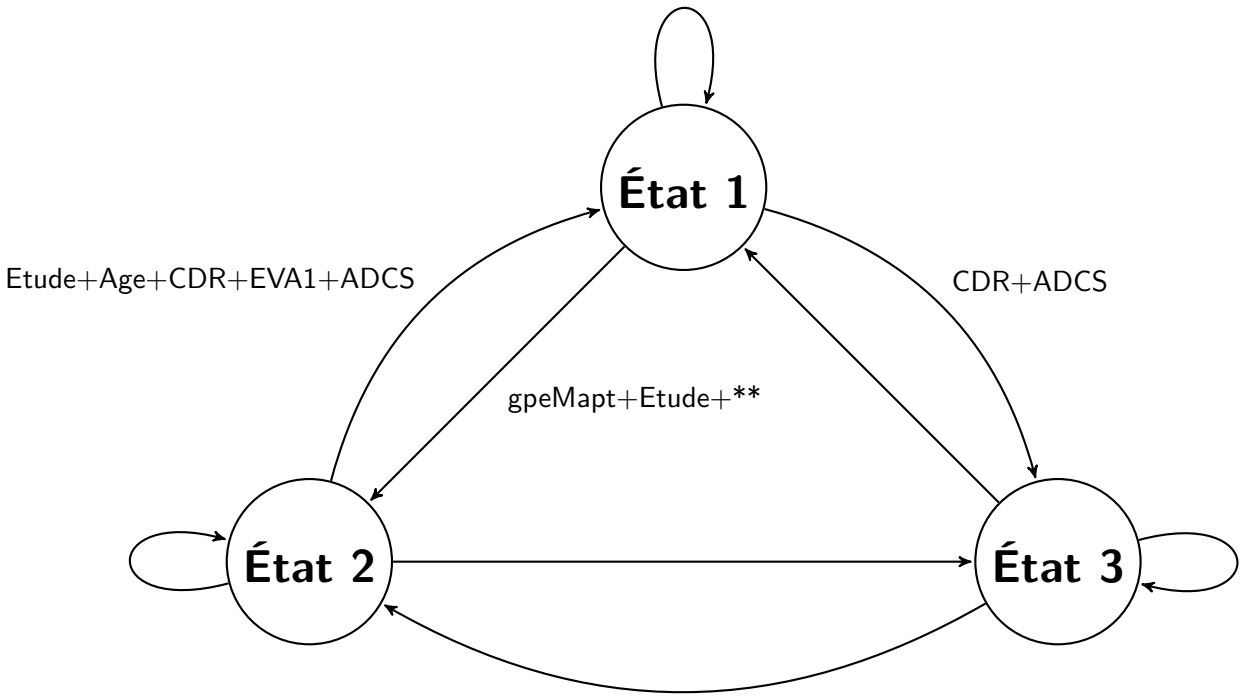

FIGURE 6.13 – Représentation du modèle de Markov homogène à 3 états de variations souples après ajustement sur les covariables, MAPT (N = 1525 sujets). ** : sexe+ Age+ BMI+ GDS+ EVA2+ Fried+ hyperT

ckage `msm` a été utilisé à chaque étape de l'analyse pour générer aléatoirement les intensités de transition initiales.

Suite à l'ensemble de ces choix et à la stratégie d'analyse présentée en section 2.2.1 du Chapitre 5, le modèle sans covariable a été construit sur les 1525 sujets. Son AIC s'élevait à 11371,36. Les covariables ont été ajoutées de façon spécifique pour chaque transition autorisée par Q_2, en commençant par la transition de l'état 1 vers 2. Le modèle homogène final avait un AIC de 10478,75. L'ajustement était donc préférable et le modèle contenait 16 variables. Les résultats sont présentés dans les Tableaux 6.12 et 6.13 ainsi que dans la Figure 6.13.

Finalement, le présent modèle contenait de nombreuses variables intervenant principalement sur les transitions entre les états 1 et 2. Les effets statistiquement significatifs concernaient 8 variables, à savoir le sexe, l'IMC, le critère de Fried, l'hypertension, le CDR, le niveau d'autonomie et la plainte mnésique.

2.2.1.4.2 Interprétation clinique

Tout d'abord, la matrices d'intensité Q_2, autorisant toutes les transitions possibles, était plus cohérente cliniquement que celle de la section 2.2.1.3. Les définitions des états étaient également plus adaptées à la réalité. Concernant le faible nombre de transitions de l'état de déclin des fonctions cognitives vers une stabilité relative, nous pouvons supposer que cet état concernait les quelques sujets identifiés en section 3.2 du Chapitre 4. Ces derniers, en faible proportion, évoluaient par alternances de phases de déclin et de stabilité cognitive.

Par ailleurs, 8 des effets présents dans le modèle étaient statistiquement significatifs.

Commençons par la transition de l'état d'amélioration des fonctions cognitives vers une stabilité relative de ces dernières. L'effet du groupe d'intervention multidomaine indiquait qu'un sujet bénéficiant de l'IM augmentait le risque de transiter par rapport à un sujet recevant le placebo (RR = 1,31, IC95% = [1,06 ; 1,62]). Le sexe indiquait qu'être un homme était protecteur de cette transition (RR = 0,62, IC95% = [0,51 ; 0,57]). En effet, ce facteur était protecteur du déclin des fonctions cognitives, ce qui est cohérent avec le sens de la transition. L'IMC montrait que l'obésité diminuait le risque changement d'état (RR = 0,74, IC95% = [0,59 ; 0,94]), tout comme le critère de fragilité de Fried (RR = 0,78, IC95% = [0,66 ; 0,93]). Ce dernier point est cohérent puisque la fragilité est un facteur de risque de déclin des fonctions cognitives. En dernier lieu, l'hypertension augmentait le risque relatif (RR = 1,18, IC95% = [1,00 ; 1,39]). Les trois derniers effets significatifs portaient sur la transition inverse, c'est-à-dire de la stabilité relative vers l'amélioration des fonctions cognitives. Une augmentation du niveau d'autonomie diminuait le risque de transiter (RR = 0,95, IC95% = [0,93 ; 0,98]). Ceci était cohérent si nous considérons que l'autonomie favorise la stabilité cognitive plutôt qu'une amélioration. En revanche, cette transition était favorisée par le score CDR, dans le sens qu'un sujet positif augmentait son niveau cognitif (RR = 1,88, IC95% = [1,44 ; 2,46]), et par la plainte mnésique (RR = 1,01, IC95% = [1,01 ; 1,02]). Ces deux derniers résultats étaient incohérents avec le sens attendu cliniquement.

Concernant les autres facteurs d'ajustement du modèle, commençons également par la transition de l'état 1 vers l'état 2. Le groupe de traitement par Ω_3, seul ou en combinaison avec l'IM, n'avait pas d'effet sur les transitions. Il avait tendance, non significativement, à diminuer le risque de transiter entre les états. Si cette tendance avait été significative, nous aurions peut être pu conclure, avec une causalité restant à prouver, que les Ω_3 protégeaient d'une perte des fonctions cognitives. Concernant le niveau d'étude, les deux effets indiquaient qu'il conduisait à une stabilité relative des fonctions cognitives. En effet, il augmentait le risque de transiter de l'amélioration cognitive vers la stabilité relative, mais il diminuait aussi celui de la stabilité relative vers l'amélioration. Pour ce qui est de l'âge, il intervenait sur cette transition et sur celle de l'état 2 vers l'état 1. D'une part, il avait un effet protecteur de la transition de l'état 1 vers l'état 2. Ceci est cohérent puisque l'âge est facteur de risque de déclin. D'autre part, l'avance en âge améliorait les fonctions cognitives, ce qui est surprenant cliniquement. En somme, l'effet de l'âge, sur les 2 transitions concernées, était soumis à discussion clinique. Pour la symptomatologie dépressive, il s'avérait qu'elle diminuait le risque de passer de l'état 1 vers l'état 2, tout comme la gêne ressentie dans les activités de la vie quotidienne. Les deux derniers effets non significatifs portaient sur la transition de l'état d'amélioration des fonctions cognitives directement vers celui de déclin. Ils concernaient d'une part le CDR, qui diminuait le risque de transiter par rapport aux sujets ayant un CDR nul, et d'autre part l'autonomie, qui augmentait ce risque. Ces deux résultats étaient surprenant cliniquement et n'étaient pas explicables.

En somme, ce modèle était plus adapté sur le plan statistique. L'ensemble de transitions étaient autorisées, les transitions observées relativement équilibrées entre les différents états et les intervalles de confiances des valeurs restreints. De plus, de nombreux facteurs expliquaient la transition de l'état d'amélioration des fonctions cognitives vers une stabilité relative et inversement. Toutefois, aucun effet significatif n'influençait les transitions en provenance ou vers l'état de déclin des fonctions cognitives.

2.2.2 Les modèles de Markov cachées

La dernière analyse du mémoire concerne la modélisation des trajectoires cognitives à partir des modèles de Markov cachés. Son objectif premier est la généralisation des résultats obtenus en section 2.2.1. En effet, le modèle est construit à partir des observations et non plus des connaissances a priori sur le phénomène. Cependant, cette méthode n'a pas été employée à des fins de classification supervisée des états mais dans le but de découvrir des classes inconnues ou latentes [238], avec potentiellement une mesure des fonctions cognitives bruitée. En effet, les HMM visent à regrouper ou séparer librement les différents états, à partir de données observées et en fonction de leur similarité statistique dans chaque groupe. Nous souhaitions principalement déterminer le nombre d'états le plus adapté pour décrire au mieux le phénomène étudié.

Pour faire cela, le nombre d'états optimal régissant la chaîne cachée a été recherché. Nous nous sommes heurtés à différents écueils lors de la mise en œuvre, conduisant à une absence de résultats concrets. Dans tous les cas, les données les plus simples et les plus restrictives, à savoir celles avec suivi complet (1143 sujets, voir $\mathcal{P}_{comp}$ en section 2.2 du Chapitre 4 pour leur description), ont été considérées dans chacune des analyses ci-après.

Dans un premier temps, un modèle sans covariable, dont la chaîne de Markov cachée serait définie par 2 états, a été implémenté à l'aide du package `momentuHMM` [259]. Le critère de jugement considéré était le Z-score composite. Préalablement à toute analyse par HMM, l'histogramme de la variable continue doit être observé. Le Z-score était manifestement gaussien. L'allure de la distribution a constitué notre premier écueil. En effet, dans un cas de chaîne cachée à 2 états, la distribution attendue est plutôt bimodale [235]. Or, dans un cas unimodal comme le nôtre, il n'était pas évident de construire intuitivement un mélange de deux gaussiennes, puisqu'une distribution par composante et par état est nécessaire pour implémenter le modèle [222]. Les paramètres de distribution initiale de chaque état ont ainsi été définis de façon subjective et sans estimation préalable. Par exemple, les valeurs moyennes, quantiles, minimum ou maximum de la distribution totale peuvent définir ceux des sous-distributions, mais nous avons privilégié dans un premier temps les quartiles Q_1 et Q_3. La distribution de l'état 1 était une gaussienne de moyenne Q_1 et d'écart-type $sd/2$, avec sd l'écart-type du Z-score composite global. Pour l'état 2, les paramètres de la gaussienne étaient respectivement Q_3 et $sd/2$. Le dernier choix effectué, pour sélectionner une période dont les dates de visite étaient séparées par une durée équivalente et relativement constante [243, 260], a restreint le suivi aux valeurs à T0, T6 mois et T12 mois. Plusieurs périodes étaient observables et l'ont été dans d'autres analyses, mais la première, considérée ici, portait sur la phase intensive de l'intervention multidomaine de l'essai MAPT. L'AIC du modèle de la première année de suivi s'élevait à 5620,4. Les probabilités de rester dans le même état s'avéraient égales à 1 tandis que celles de transiter étaient nulles. Deux hypothèses pouvaient expliquer cela : soit nous avions, malgré nos 1143 sujets, un problème de puissance, soit les paramètres initiaux n'étaient pas adaptés.

Pour tester la première hypothèse, le nombre de mesures par sujets a été augmenté. La période d'observation considérée s'étendait sur l'intégralité du suivi, avec des mesures annuelles, soit 4 mesures par sujet au lieu de 3 dans le précédent modèle. Le même constat a été observé pour la distribution, puisque l'*outcome* était identique, et les paramètres de distribution initiale ont été conservés. Ce modèle augmentait en terme d'AIC, de 5620,4 à 7769,7 avec l'ajout de la nouvelle mesure, mais les probabilités de transition restaient inchangées.

Nous avons donc supposé que le nombre de valeurs par sujet n'était pas encore assez important pour détecter les changements d'états. En effet, dans la littérature, nous avons cherché des informations à ce propos. Dans des études de séquences ADN, le nombre de points d'observation, ou la taille de la séquence d'observation, était au minimum de 14 [261]. Une autre étude du même domaine utilisait des séquences contenant 10 valeurs différences pour chaque jeu de données exploités [262]. Plus généralement, dans les études biologiques, les tailles de séquences sont de l'ordre de 2000 points [243] et peuvent aller jusqu'à 30000 points [244]. De fait, des valeurs artificielles ont été créées au cours du suivi par imputations successives. La méthode de *CopyMean*, présentée en section 1.1.1.3 du Chapitre 5, a été appliquée. Pour obtenir des valeurs tous les 6 mois, 2 valeurs supplémentaires ont été ajoutées, à T18 et T30 mois. Pour avoir un suivi tous les 3 mois et un nombre plus important de valeurs au cours du suivi, 6 imputations supplémentaires ont eu lieu. Sur les 13 valeurs du Z-score au cours du suivi, seules 38,5 % étaient réelles. Les résultats ont montré une amélioration des probabilités de transition : elles étaient moins faibles mais tendaient toujours vers 0. En revanche, le modèle n'étaient absolument pas pertinent ni cliniquement interprétable. Les résultats n'ont donc pas été exploités en l'état, ils ont simplement confirmé l'hypothèse d'une séquence d'observation trop courte pour détecter les transitions entre les états. Avec toutes les valeurs rajoutées, l'AIC du modèle était égal à 20665,8.

Pour tester la seconde hypothèse, les paramètres initiaux ont été choisis via l'utilisation de modèles de mélanges gaussiens (package `mixtools` de R déterminant le mélange optimal de distributions pour une variable [242]). L'avantage est qu'il suffit de rentrer le nombre de distributions souhaitées, une par état, soit 2 ici. L'inconvénient est que les résultats n'ont pas été améliorés en terme de transition. Le modèle suggérait donc qu'il n'existait pas de transition entre les états et son AIC valait 9485,5.

La dernière tentative de recherche résultait d'un mélange des deux hypothèses. D'une part, nous avons conservé l'optimisation des paramètres initiaux par modèles de mélange gaussiens. D'autre part, d'après le modèle suggéré par Sukkar et al. dans leur étude au *design* avoisinant le nôtre [238] ou encore par DeRuiter [263], la variable d'intérêt a été modifiée. Le Z-score composite n'est plus considéré directement mais comme dissocié. Ainsi, chacune des 4 composantes du Z-score a été considérée indépendamment des autres et associée à une distribution initiale qui lui est propre. Par contre, la composante du MMSE, ne portant que sur les 10 premiers items du tests dans le Z-score composite (voir section 1.3.1.2 du Chapitre 3), a été remplacée par le score total. Les autres variables correspondent à celles utilisées pour former le Z-score composite. Chaque distribution était de type gaussienne dans notre cas. Nous avions une distribution (composée de 2 paramètres) par état et pour chacune des composantes, soit 4 distributions par état, à fournir comme paramètres initiaux du modèle de Markov caché [244]. En terme de suivi, le plus long possible sans imputation a été observé, à savoir un suivi annuel. Trois modèles ont été implémentés, en faisant varier le nombre d'états cachés de 2 à 4, sans covariable explicative. Le modèle à 4 états n'a pas pu être exécuté car il était impossible de déterminer 4 distributions pour la composante du MMSE. D'après un article [227], cette non convergence n'était pas surprenante car les modèles de plus de 3 états ne convergent que si la séquence d'observations est longue. Le choix du nombre d'états cachés optimal balançait donc entre 2 et 3. L'AIC a été utilisé pour déterminer le meilleur modèle et valait respectivement 111375,3 et 109378,1. Celui à 3 états cachés a été retenu comme semblant le plus prometteur des deux.

Une fois le nombre d'états cachés déterminé, nous avons voulu ajouter des covariables explicatives des probabilités de transitions. Seuls les effets simples ont été estimés pour des

raisons de temps d'exécution (voir Tableau 6.14). Ce problème est récurrent dans les modèles de Markov cachés. En effet, Pirotta et al. [260] indiquaient que leur modèle mettait 18 h à tourner, pour 6 sujets et 10 000 itérations. Dedieu et al. [223] ont stipulé que le temps d'exécution des HMM était long, jusqu'à plusieurs jours, pour 100 échantillons de 1000 sujets, 3 covariables explicatives et 6 états cachés. Pour notre part, le Tableau 6.14 résume les temps d'exécution des modèles décrits ci-dessus. Il reprend également les principaux choix, à savoir considérer le Z-score ou ses composantes, le nombre d'états cachés, la période d'observation et le nombre de mesures correspondant ainsi que l'utilisation ou non de modèles de mélange gaussiens. Nous constatons qu'augmenter le nombre de mesures augmente le temps d'exécution. Utiliser des modèles de mélange n'influençait pas le temps d'exécution du modèle directement. Cela influençait le temps de préparation du modèle uniquement. Contrairement à ce que nous pensions a priori, dissocier le Z-score composite n'a pas outre-passé le problème de taille de séquence d'observation insuffisante. Ceci ajouté à la contrainte temporelle nous a empêché de franchir l'étape de l'analyse multivariée.

Table 6.14 – Temps d'exécution des différents modèles de Markov cachés implémentés durant notre stratégie, MAPT (N = 1143 sujets).

(1)	# états cachés	Période observée	# mesure par sujet	Modèles mélange	AIC	Temps exécution(2)	Temps en heure
1	2	T0-T12	3	non	5620,4	3937,2	1,1
1	2	T0-T36	4	non	7769,7	4519,8	1,3
1	2	T0-T36	13	non	20665,8	4551,0	1,3
1	2	T0-T36	4	oui	9485,5	3873,7	1,1
4	2	T0-T36	4	oui	111375,3	18462,8	5,1
4	3	T0-T36	4	oui	109378,1	57381,4	15,9

est une abréviation pour "nombre de" ;
(1) : Nombre de variables quantitatives servant de variable réponse, 1 pour le Z-score composite lui-même, 4 lorsqu'il est divisé en chacune de ses composantes ;
(2) : Temps total écoulé en secondes entre la soumission de l'expression et le retour du résultat pour le logiciel R.

En somme, ces analyses par modèles de Markov caché ont confirmé que le nombre d'états régissant au mieux le phénomène s'élevait à 3, ce qui concorde avec les résultats de la section 2.2.1.4. Notons également que la taille de la séquence d'observation restait insuffisante pour aboutir à des résultats, induisant que la méthode n'a pas été exploitée comme prévu initialement. À ceci s'est ajouté un temps d'exécution des HMM particulièrement long, compliquant les analyses.

TABLE 6.8 – Récapitulatif des risques relatifs de transition du modèle homogène à 3 états dont 1 pour données manquantes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1544 sujets). Partie 1

Variables	1 vers 1		1 vers 2		1 vers 3		2 vers 1		2 vers 2		2 vers 3	
Baseline	-0,035		0,035		0,000		0,025		-0,025		0,000	
	(-0,039	-0,032)	(0,032	0,039)	(0,000	0,000)	(0,022	0,028)	(-0,028	-0,022)	(0,000	0,001)
Etude			0,971									
			(0,852	1,106)								
Age												
CDR			1,039									
			(0,933	1,158)								
IMC			0,985									
			(0,855	1,136)								
GDS												
EVA1											1,112	
											(1,024	1,208)
Hypertension							1,018					
							(0,885	1,171)				

État 1 = amélioration des fonctions cognitives, état 2 = déclin et état 3 = donnée manquante.

TABLE 6.9 – Récapitulatif des risques relatifs de transition du modèle homogène à 3 états dont 1 pour données manquantes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1544 sujets). Partie 2

Variables	3 vers 1		3 vers 2		3 vers 3	
Baseline	0,015		0,038		-0,054	
	(0,001	0,268)	(0,007	0,219)	(-0,164	-0,018)
Etude						
Age	1,158					
	(0,836	1,604)				
CDR	0,495					
	(0,015	15,857)				
IMC						
GDS			0,058			
			(0,000	15,460)		
EVA1						
Hypertension						

État 1 = amélioration des fonctions cognitives, état 2 = déclin et état 3 = donnée manquante.

TABLE 6.10 – Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variation strictes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 1

Variables	1 vers 1	1 vers 2	1 vers 3	2 vers 1	2 vers 2
Baseline	-6,514	0,006	6,508	0,487	-0,487
	(-9,996e+138 -4,245e-138)	(0,000 89,240)	(3,188e-138 1,329e+139)	(0,015 15,360)	(-15,36 -0,015)
Etude		1,244			
		(0,048 32,270)			
Age					
ADCS			1,000		
			(0,988 1,012)		

État 1 = amélioration stricte des fonctions cognitives, état 2 = stabilité stricte et état 3 = déclin strict.

TABLE 6.11 – Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variation strictes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 2

Variables	3 vers 1	3 vers 2	3 vers 3
Baseline	6,808 (2.731e-138 1.697e+139)	0,004 (7,406e-09 2,440e+03)	-6,811 (-1,392e+139 -3.333e-138)
Etude			
Age	0,978 (0,965 0,991)		
ADCS			

État 1 = amélioration stricte des fonctions cognitives, état 2 = stabilité stricte et état 3 = déclin strict.

TABLE 6.12 – Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variations souples (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 1

Variables	1 vers 1		1 vers 2		1 vers 3		2 vers 1		2 vers 2		2 vers 3	
Baseline	-0.184		0.047		0.137		0.108		-0.119		0.011	
	(-0.347	-0.098)	(0.028	0.079)	(0.059	0.318)	(0.063	0.185)	(-0.185	-0.077)	(0.001	0.148)
IM			1.306									
			(1.055	1.616)								
Ω_3			0.929									
			(0.740	1.166)								
Ω_3 + IM			0.929									
			(0.744	1.162)								
Etude			1.221				0.423					
			(0.150	9.968)			(0.049	3.665)				
Sexe			0.621									
			(0.514	0.751)								
Age			0.971				1.024					
			(0.918	1.028)			(0.968	1.084)				
IMC			0.743									
			(0.590	0.935)								
GDS			0.899									
			(0.706	1.144)								
EVA-2			0.998									
			(0.994	1.002)								
Fried			0.780									
			(0.657	0.925)								
Hypertension			1.181									
			(1.002	1.393)								
CDR					0.982		1.882					
					(0.841	1.147)	(1.442	2.455)				
ADCS					1.002		0.954					
					(0.987	1.017)	(0.929	0.979)				
EVA-1							1.013					
							(1.005	1.021)				

État 1 = amélioration des fonctions cognitives, état 2 = stabilité souple et état 3 = déclin.

TABLE 6.13 – Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variations souples (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 2

Variables	3 vers 1		3 vers 2		3 vers 3	
Baseline	0,183		0,002		-0,184	
	(0,081	0,410)	(0,000	0,161)	(-0,413	-0,082)
IM						
Ω_3						
Ω_3 + IM						
Etude						
Sexe						
Age						
IMC						
GDS						
EVA-2						
Fried						
Hypertension						
CDR						
ADCS						
EVA-1						

État 1 = amélioration des fonctions cognitives, état 2 = stabilité souple et état 3 = déclin.

Chapitre 7

Conclusion et perspectives

1 Rappel de la problématique

Dans cette thèse, nous avons présenté les essais de prévention, en les distinguant des essais cliniques classiques par certaines spécificités. Dans le cas d'un essai non concluant, l'absence d'effet du traitement à l'épreuve est souvent mise en avant. Nous avons proposé de considérer, en alternative, que le traitement à l'épreuve pouvait être efficace mais que la méthodologie déployée n'a pas permis de détecter cet effet dans la grande majorité des essais de prévention du déclin des fonctions cognitives.

Nous avons identifié trois grandes spécificités des essais de prévention devant être prises en compte par la méthodologie d'analyse. En effet, ces essais sont caractérisés par des données répétées au cours du temps, potentiellement soumises à effet d'apprentissage sur les fonctions cognitives, modifiant l'effet attendu du traitement en cours d'essai. En somme, l'essai est sensible à la période de suivi. De plus, une hétérogénéité de la population d'étude est particulièrement notable. Pour chaque difficulté méthodologique, un ensemble de méthodes d'analyses spécifiques a été proposé pour optimiser la détection d'un effet potentiel. Par exemple, pour le caractère longitudinal de ces données, nous avons proposé deux stratégies : soit en utilisant directement la donnée, soit en exploitant des paramètres de synthèse résumant l'évolution de cette dernière. De façon à traiter les perdus de vue au cours du suivi, des approches markoviennes ou d'apprentissage statistique ont été mobilisées. Concernant l'incertitude de mesure, qui augmente la variabilité des résultats, nous avons proposé d'utiliser des approches markoviennes et les sujets considérés comme répondeurs au traitement et/ou à l'intervention. Enfin, pour l'hétérogénéité de la population d'analyse, trois méthodes de *clustering* permettant de rechercher des sous-groupes de sujets homogènes ont été explorées.

En somme, nous avons émis l'hypothèse que ces spécificités des essais de prévention du déclin des fonctions cognitives pouvaient expliquer, en partie, pourquoi la méthodologie d'analyse déployée n'était pas adaptée à leur traitement, conduisant ainsi la plupart d'entre à des résultats peu convaincants.

2 Spécificités de notre approche

Dans ce mémoire, nous avons proposé différentes méthodes statistiques de ré-analyse des résultats d'un essai de prévention non concluant. Nous souhaitions que chacune puisse détecter un effet qui serait passé inaperçu par les techniques d'analyses les plus couramment utilisées. Comme nous avons pu le voir dans le Chapitre 5, chacune possède ses avantages et inconvénients méthodologiques et sont synthétisées ici par le diagramme proposé en Figure 5.1 de ce chapitre. Il indique les différents outils exploités selon leur nature, l'objectif (exploratoire ou modélisation), les connaissances a priori ou non et selon un usage direct de la donnée fonctionnelle ou indirect par les paramètres de synthèse. Quelle que soit la méthode, il est important de garder à l'esprit que chacune possède ses propres conditions d'applications, différents paramètres à choisir ou à adapter et cela conduit à un résultat différent. Il est donc nécessaire d'être prudent quant à leur interprétation. Par contre, de l'information peut être extraite de chacune des analyses pour bénéficier soit directement pour les résultats de l'essai MAPT, soit pour la construction des futurs essais de prévention.

De la même façon, une synthèse des résultats apportés au chapitre précédent est proposée en section 2.1 pour l'approche exploratoire et en section 2.2 pour la modélisation des trajectoires cognitives. Des perspectives d'analyses pour compléter l'arsenal d'outils de ré-analyse sont indiquées en section 3.

2.1 Les apports des approches d'exploration

Au terme de la phase exploratoire des données de l'essai de prévention MAPT, plusieurs résultats importants doivent être soulignés. Dans un premier temps, les analyses préliminaires présentées en section 3 du Chapitre 4 visaient à explorer l'adéquation entre l'évolution des fonctions cognitives attendue cliniquement et la réalité au sein de la population de MAPT. Elles sont représentées sur la Figure 5.1 du Chapitre 5. Le constat reposait sur quatre points :

1. La population ne se comportait pas comme attendu cliniquement car plus de 70 % des sujets étaient stables cognitivement ;
2. Des profils extrêmes d'évolution des fonctions cognitives ont été détectés, en faible nombre. Quelques caractéristiques cliniques les distinguaient cependant ;
3. Trois types de variations des fonctions cognitives ont été mises en évidence : stabilité, déclin et amélioration ;
4. Aucune différence entre les groupes n'a été trouvée selon les bras de traitement par Ω_3 ou de l'IM.

D'autres analyses exploratoires non supervisées ont complété et renforcé ces informations (section 1 du Chapitre 6 et Figure 5.1 du Chapitre 5). Elles ont été comparées sur de nombreux points résumés dans le Tableau 7.1. Notre intérêt se porte particulièrement sur la ligne "Pertinence clinique", reprenant les principaux résultats.

Sur un échantillon de 1679 sujets, âgés de 70 ans ou plus et à risque de présenter une démence ou une pathologie de type Alzheimer au cours d'un suivi de 3 ans, les trois méthodes non supervisées présentées précédemment ont été appliquées pour former des groupes de trajectoires des fonctions cognitives. Le critère de jugement principal pour ces analyses était

soit un Z-score composite combinant les résultats de 4 tests cognitifs (analyse directe de la donnée fonctionnelle indiquée en Figure 5.1 du Chapitre 5), soit les taux d'accroissement du Z-score entre les différentes visites (analyses indirectes de la Figure 5.1 du Chapitre 5). Le package `kmlShape` (section 1.1.1 du Chapitre 6), implémentant la méthodes des *k-means* longitudinaux avec reconnaissance des groupes basée sur la forme de la trajectoire, a mis en évidence un groupe représentant 16 % de l'échantillon d'étude. Il était particulièrement intéressant car il contenait près de 92 % des sujets déments. De plus, il avait un profil d'évolution des fonctions cognitives bien spécifique en montrant un déclin au cours des 3 ans. Il se différenciait des deux autres groupes trouvés par de nombreux facteurs de risques de survenue de la démence, allant dans le sens attendu cliniquement.

Cependant, les sujets déments n'ont pas pu être trouvés dans les deux autres méthodes non supervisées (CAH et sémiologie graphique, respectivement en section 1.1.2 et 1.1.3 du Chapitre 6). En effet, des sujets avec suivi complet ont été utilisés dans les deux cas, ce qui impliquait qu'ils étaient venus à toutes les visites requises. Leur nombre était ainsi réduit à 1143 sujets, dont seulement 5 d'entre eux ont été considérés déments. Cette sélection de la population, résultant de l'incapacité des méthodes à traiter directement la donnée fonctionnelle et donc de la définition choisie pour les indicateurs de variation de trajectoire, a conduit à une population d'analyse différente de celle exploitée par `kmlShape`. À l'évidence, la composition des profils différait selon la méthode utilisée puisque les effectifs étaient différents [31]. Les résultats ne sont donc pas comparables directement entre les méthodes puisqu'ils ne portent ni sur les mêmes sujets, ni sur le même critère de jugement (analyses indirectes en Figure 5.1 du Chapitre 5). De plus, une précédente étude a montré que les sujets quittant l'étude avant la fin du suivi prévu sont soumis à un déclin des fonctions cognitives d'autant plus important qu'ils sortent de l'étude tôt [9]. Il n'est donc pas surprenant d'avoir observé une relative stabilité sur les deux dernières années de suivi avec la CAH et la sémiologie graphique. Ces résultats étaient concordants avec ceux évoqués ci-dessus. En outre, notons que les trois groupes, déterminés par chacune des méthodes exploratoires, différaient sur peu de caractéristiques pertinentes cliniquement mais semblaient se comporter d'une façon particulière entre l'inclusion et la première année de suivi d'une part et sur les 2 dernières années d'autre part. En sus de cela, aucun effet du bras de traitement n'a été relevé. Le dernier point de ces analyses indirectes d'exploration des trajectoires réside dans le choix du critère de jugement. Nous avons considéré les taux d'accroissement entre les visites, mais les résultats en sont dépendants. Il serait intéressant d'observer les groupes formés par d'autres critères cliniquement pertinents, tels qu'une combinaison de ceux évoqués en Chapitre 4.

De façon à compléter cette information et toujours dans l'optique de construire des profils d'évolution des fonctions cognitives, une dernière stratégie a été utilisée (exploration supervisée de la Figure 5.1 du Chapitre 5). Sur un échantillon de 1145 sujets, l'analyse supervisée par catégories de répondeurs a été exploitée en section 1.2 du Chapitre 6. Elle consistait à diviser en deux groupes la population, avec pour seuil une variation de 20 % par rapport au placebo sur la période d'observation considérée du taux d'accroissement. Concernant la pertinence clinique indiquée dans la dernière ligne du Tableau 7.1, nous avons constaté un résultat proche de celui obtenu par la méthode `kmlShape`, dans une moindre mesure. Plus de la moitié des sujets ont amélioré leur niveau global des fonctions cognitives et ont été considérés comme répondeurs. De plus, la proportion de sujets déments était d'environ 60 % dans le groupe des non répondeurs. Les caractéristiques cliniques distinguant les deux groupes allaient dans le sens attendu a priori. Ceci pourrait être résumé de façon simpliste en disant que les répondeurs se portaient mieux que les non répondeurs, dans le sens qu'ils

possédaient moins de facteurs de risque du déclin cognitif. Pour cette analyse cependant, nous avons indiqué que plusieurs définitions du répondeur existaient dans la littérature, tout comme il existe d'autres critères évaluant une différence d'évolution. Il serait intéressant d'explorer plus avant cette piste afin d'en observer les conséquences sur les résultats. Autre point important, nous avons constaté l'absence d'effet des bras d'intervention ou de l'IM.

Ce dernier constat est primordial pour plusieurs raisons. Tout d'abord, comme évoqué précédemment, il confirme que les modèles mixtes utilisés initialement pour analyser les données de MAPT n'ont pas fourni de résultats parce qu'il n'y avait pas de différence entre les bras de l'essai à détecter. Ensuite, il confirme que la stratégie d'analyse initiale était adaptée et n'avait pas pour objectif de modifier la conclusion de l'essai, bien que les méthodes exploratoires de ré-analyse d'un essai clinique non concluant aient toutes quatre fournis des informations complémentaires. Nos analyses confirmaient aussi que l'hétérogénéité des sujets, ou la variabilité du Z-score composite, était équivalente dans les 4 bras de l'essai MAPT. Initialement, nous avions supposé qu'il existait des sujets pour lesquels un effet aurait pu être détecté. Cette hypothèse n'a été vérifiée que sur l'existence de sous-groupes de sujets particuliers. En dernier lieu, ces analyses exploratoires ne suffisent pas à donner un avis sur la pertinence de la construction du Z-score composite. Les résultats ne sont pas suffisamment importants ou significatifs pour cela.

TABLE 7.1 – Résumé, comparaisons et recommandations pour l'utilisation des méthodes d'exploration des trajectoires cognitives appliquées dans ce mémoire.

	Modèles mixtes	*Clustering* fonctionnel		Sémiologie graphique	*Clustering* Hiérarchique	Répondeurs
Package R	nlme(1)	kml(2)	kmlShape	seriation	stats	-
Outcome	Z-score	Z-score	Z-score	Δ	Δ	Δ
Indicateurs	-	-	-	$\frac{T_{fin} - T_0}{Temps}$	$\frac{T_{fin} - T_0}{Temps}$	$\frac{T_{fin} - T_0}{Temps}$
♯ points	5	5	5	4 (entre visites)	4 (entre visites)	2 (entre T0 et T_{fin})
Nature données	Toutes	Toutes	Toutes	Suivi complet	Suivi complet	-
Réglage	-	♯ groupe	♯ groupe	♯ couleurs	♯ groupe	Seuil
A priori	Trajectoire linéaire	♯ groupe	3 groupes	3 couleurs	Non	Définition
Choix des groupes	Aucun	Calinski et H.	10% minimum	Aucun	Critère du coude	Aucun
Imputation	Non (MAR *)	Non (Grower)	Que si réduction (CopyMean)	-	-	-
Pertinence clinique	Comparaison ttt(3)	-	Déments détectés	Non	Non	Vont "mieux"
Alternatives	-	-	-	Indicateurs : min, max, variation cumulée, variations absolues, ...		

♯ : "nombre de"
Δ : "taux d'accroissement"
(1) Les modèles mixtes ont été utilisés dans l'article de référence de MAPT [13] et l'hypothèse des données manquantes aléatoires a été émise pour l'analyse.
(2) Cette méthode n'a pas été directement appliquée aux données de MAPT mais elle est présentée ici à titre de comparaison avec les autres méthodes
(3) abréviation locale pour groupes d'intervention ou bras de l'essai.

2.2 Les apports des approches de modélisation

À travers l'utilisation de stratégies de modélisation des données de l'essai de prévention MAPT, il est important d'insister sur certains résultats. Nous avons appliqué des méthodes d'apprentissage statistique (Figure 5.1 du Chapitre 5). Ces méthodes d'analyse supervisées classiques ont été explorées en section 2.1 du Chapitre 6. Quelle que soit celle appliquée, toutes quatre ont conduit à une unique information : l'impossibilité à détecter des différences selon le traitement de l'essai ou l'IM. Elles maintenaient le constat découlant des analyses des catégories de répondeurs. Le fait de n'obtenir aucun effet significatif ou cliniquement interprétable constituait un résultat en soit, suggérant que les modèles mixtes, utilisés dans l'article de MAPT, ne sont passés ni à côté d'un effet du traitement ni d'un effet de l'intervention. Ces analyses ont donc corroboré l'hypothèse que la méthode était adaptée et qu'elle n'avait pas pour objectif de revenir sur la conclusion de l'essai.

De plus, toujours en utilisation direct du Z-score composite (Figure 5.1 du Chapitre 5), nous avons abordé l'apprentissage statistique par renforcement, en utilisant le *Q-learning* en section 2.1.2 du Chapitre 6. L'avantage de cette méthode est de considérer les variables explicatives répétées dans le temps, à l'instar de l'*outcome*. Sa particularité réside dans un processus indirect en plusieurs étapes et dont le temps est considéré en sens inverse. Sur un échantillon de 1525 sujets, âgés de 70 ans ou plus et à risque de présenter un déclin cognitif ou une démence de type Alzheimer au cours d'un suivi de 3 ans, cette méthode n'a pas été exploitée. Nous étions dans l'incapacité de définir une seconde décision au cours du suivi. Le calcul des résultats potentiels était équivalent à une régression classique, donc non adapté à notre contexte.

Pour revenir sur notre hypothèse en suspens d'un critère non adapté, nous avons reconsidéré les trajectoires cognitives comme une succession d'états au cours du temps (Figure 5.1 du Chapitre 5). Cette nouvelle considération de la trajectoire n'utilisait pas le taux d'accroissement pour résumer l'information sur une période donnée mais un état par temps de mesure pour recoder l'information au temps t. Sur un échantillon de 1679 sujets, âgés de 70 ans ou plus et à risque de présenter un déclin cognitif ou une démence de type Alzheimer au cours d'un suivi de 3 ans, des modèles de Markov observables (états directement observés) ont été appliquées pour explorer le risque de transition entre les états (voir section 2.2.1 du Chapitre 6). Ces derniers ont été construits selon différents niveaux du critère de jugement principal de l'essai et selon des comportements attendus cliniquement. L'objectif de cette approche était d'appréhender l'ensemble de la trajectoire des fonctions cognitives à travers une succession d'états dans lequel se trouvait le sujet. Grâce aux modèles de Markov classiques, cet objectif se traduisait par la mise en évidence des effets différenciés selon les transitions autorisées [3]. Deux choix étaient envisageables, un avec l'effet de covariables différent sur chaque transition (structuré) et un avec effets identiques (réduit) [218]. Nous avons choisi de n'étudier que des modèles structurés de façon à approcher au mieux la réalité clinique. Un tableau de synthèse des différents modèles étudiés est proposé en Tableau 7.2.

Les deux premiers modèles opposaient le comportement attendu (déclin cognitif) avec les autres observés, à l'instar des analyses préliminaires menées en section 3.1 du Chapitre 4. Le second prenait en compte les données manquantes au cours du suivi comme un état indépendant (3 états - section 2.2.1.2 du Chapitre 6) tandis que le premier ne le faisait pas (2 états - section 2.2.1.1 du Chapitre 6). Sans donnée manquante, le modèle indiquait un effet significatif de l'âge : plus un sujet avançait en âge, plus il risquait de décliner cognitivement. Dans le deuxième cas, la plainte mnésique apparaissait comme facteur de risque

TABLE 7.2 – Synthèse des différents paramètres des quatre modèles de Markov observables étudiés pour les données de l'essai de prévention MAPT.

# Etats	Définition	# Sujets	Probabilité initiale	Répartition après initialisation	Transitions totales	Matrice intensités
2		1525				
				787		
	$1 : Z_{1i} \geq Z_{0i}$		$\pi_1 = 1$	715	$\begin{pmatrix} 1528 & 1988 \\ 1215 & 563 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$
	$2 : Z_{1i} < Z_{0i}$		$\pi_2 = 0$			
3		1544				
				787		
	$1 : Z_{1i} \geq Z_{0i}$		$\pi_1 = 1$	715	$\begin{pmatrix} 1589 & 1591 & 198 \\ 794 & 1320 & 211 \\ 34 & 45 & 394 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$
	$2 : Z_{1i} < Z_{0i}$		$\pi_2 = 0$	42		
	$3 : Z_{1i} = NA^*$		$\pi_3 = 0$			
3		1525				
				777		
	$1 : Z_{1i} > Z_{0i}$		$\pi_1 = 0$	10	$\begin{pmatrix} 688 & 24 & 1263 \\ 805 & 11 & 725 \\ 1196 & 19 & 563 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$
	$2 : Z_{1i} = Z_{0i}$		$\pi_2 = 1$	715		
	$3 : Z_{1i} < Z_{0i}$		$\pi_3 = 0$			
3		1525				
				704		
	$1 : Z_{1i} > Z_{0i} + 30\%$		$\pi_1 = 0$	270	$\begin{pmatrix} 543 & 374 & 786 \\ 878 & 715 & 740 \\ 921 & 76 & 261 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$
	$2 : Z_{1i} \in Z_{0i} \pm 30\%$		$\pi_2 = 1$	528		
	$3 : Z_{1i} < Z_{0i} - 30\%$		$\pi_3 = 0$			

est une abbréviation pour "nombre de" ;

* : NA signifie ici donnée manquante.

de présenter des données manquantes après une phase de déclin cognitif. Cependant, rappelons que des difficultés d'interprétation clinique quant à cet état de données manquantes ont été soulignées. Deux autres modèles ont été explorés, appuyés par les résultats obtenus à l'aide des analyses exploratoires de la section 1 du Chapitre 6. Ils visaient à définir 3 niveaux de fonctions cognitives, associés à 3 types d'évolutions. D'une part, les variations ont été considérées comme strictes (section 2.2.1.3 du Chapitre 6). Ce modèle n'autorisait pas toutes les transitions possibles et présentait des déséquilibres du nombre de transitions par rapport à l'état de stabilité stricte. Un effet de l'âge a été mis en évidence, dans le sens qu'il diminuait le risque d'amélioration des fonctions cognitives. D'autre part, les variations ont été assouplies et définies comme une stabilité de 30 % autour de la valeur précédente, une augmentation au-delà et une diminution en-deçà (voir section 2.2.1.4 du Chapitre 6). Les arguments statistiques et cliniques autorisaient toutes les transitions entre les états. La répartition de ces dernières, sur l'intégralité du suivi, était moins déséquilibrée que dans le modèle précédent. Au total, 8 effets statistiquement significatifs ont été mis en évidence, concentrés uniquement sur les transitions entre l'état d'amélioration des fonctions cognitives et celui de stabilité relative. Le modèle était également plus cohérent cliniquement sur la définition des états.

Ces quatre modèles partageaient une divergence majeure. En effet, leur définition des états étant différente, ils ne pouvaient pas être directement comparés entre eux, même en terme d'AIC pour déterminer le meilleur modèle [18]. Sur le plan théorique, ces modèles sont également critiquables à certains niveaux car ils sont sensibles à des paramètres tels que la définition des états (voir Tableau 7.2), l'hypothèse d'homogénéité, l'ordre de dépendance entre les états et les probabilités initiales. De plus, ils sont sensibles à la matrice d'intensité de transition (voir Tableau 7.2) et à l'initialisation de ces intensités. Dans notre cas, pour limiter les a priori sur le phénomène, un paramètre de génération aléatoire a été utilisé. Finalement, au cours de la stratégie de sélection des variables, les résultats sont soumis à plusieurs paramètres : la considération d'un modèle structuré ou réduit. L'ordre des transitions ainsi que l'ordre des variables testées avaient également leur importance car ils influencent directement la contribution à la vraisemblance du modèle (voir section 2.2.1 du Chapitre 5 sur la théorie du modèle).

Par ailleurs, des modèles de Markov cachés ont été employés. Nous les avons exploités pour rechercher le nombre d'états cachés décrivant au mieux les données. Des modèles de Markov cachés sans covariables sur les transitions (modèles à vide) ont été implémentés, pour différents nombres d'états cachés. Sans a priori sur ce nombre, nous l'avons fait varier de 2 à 4 états. Lorsque les HMM sont utilisés pour la recherche du nombre d'états cachés, comme c'était notre cas, les états de la chaîne n'ont pas de réelle signification. Leur interprétation, pour le modèle optimal, est déduite à partir de la description des caractéristiques des sujets appartenant aux états identifiés. L'information importante que nous avons extrait de cette analyse était que le nombre optimal d'états régissant la partie cachée du modèle était de 3. Nous avons ainsi conclu que ces 3 états avaient une signification similaire à celle définie en section 2.2.1.4 du Chapitre 6 avec les modèles de Markov observables.

3 Perspectives de recherche

Concernant les perspectives de ce travail, nous pouvons tout d'abord évoquer la possibilité d'utiliser d'autres méthodes exploratoires ou de modélisation des trajectoires cognitives. Le panorama présenté dans ce travail n'avait pas pour objectif d'être exhaustif, nous savons donc qu'il existe de nombreuses autres méthodes répondant à notre objectif. En particulier, soulignons que pour l'étude des profils répondeurs, seule une description et une comparaison des groupes ont été effectuées. Nous aurions ainsi pu mettre en œuvre l'autre stratégie que nous avons proposé, à savoir étudier les facteurs associés à la réponse à l'intervention. Dans ce cas, une définition du statut répondeur reste nécessaire pour appliquer des méthodes d'analyses de variables catégorielles, telle que la régression logistique.

Concernant les modèles de Markov classiques, certains points peuvent être approfondis. Les modèles que nous avons exploités étaient dotés de certaines similarités impactant directement leurs résultats. Pour commencer, la variable explicative d'apparition de la démence n'a pas été prise en compte dans les analyses suite à une incidence faible. Il est ainsi possible que ce facteur explique certaines transitions. Ensuite, nous avons systématiquement émis l'hypothèse que les sujets commençaient dans un état de non déclin ou de stabilité cognitive (voir Tableau 7.2). Cet a priori très fort n'est pas entièrement réaliste sur le plan clinique. Cela reviendrait à considérer que les fonctions cognitives ne déclinent pas avec l'âge ou les sujets comme n'ayant aucune autre comorbidité. Or, nous avons expliqué en section 2 du Chapitre 2 que la population d'étude était particulière, ce qui signifie que certains sujets peuvent être en phase de déclin des fonctions cognitives à l'inclusion, sans pour autant présenter une maladie d'Alzheimer. Les modèles précédents pourraient donc être adaptés avec des probabilités initiales sur chaque état, mais la difficulté réside en la détermination de leur valeur. En l'absence de connaissances cliniques, une possibilité consiste à utiliser des méthodes s'appuyant sur des techniques d'apprentissage statistique. Les modèles pourraient être initialisés comme tels puis leurs répartitions à l'état 1 seraient utilisés comme probabilités initiales des différents états. Dans tous les cas, ces dernières doivent être déterminées en gardant à l'esprit qu'elles influencent fortement les résultats.

Une autre similarité entre les modèles repose sur les hypothèses d'homogénéité et d'ordre 1 du modèle. Tout d'abord, l'ordre 1 signifie que l'état présent ne dépend que de l'état précédent. Pour des pathologies telles que la maladie d'Alzheimer, il n'y a aucune raison de croire que le prochain état du patient recevant un certain traitement dépend uniquement de l'état et du traitement qui précèdent, et non pas également de son histoire lointaine. Il aurait donc été intéressant d'augmenter l'ordre des chaînes de Markov afin de déterminer le nombre de mesures liées entre elles. Cette étude pourrait aussi bien se faire sur données réelles que sur données simulées. Ceci mis à part, l'homogénéité est également une hypothèse forte. Elle suppose que les intensités de transitions sont indépendantes du temps. Or, nous avons indiqué que les effets attendus dans les essais de prévention ne sont pas forcément linéaires et sont progressifs (voir section 2.2 du Chapitre 2). Les modèles non homogènes, autorisant des intensités de transitions non linéaires et dépendantes de la durée de suivi, pourraient donc être plus adaptés au phénomène étudié. Pour les modéliser, deux solutions existent : soit des modèles homogènes par morceaux [228], dont les résultats dépendant des seuils de coupure, soit des modèles non homogènes paramétriques [264, 265]. Le premier cas a particulièrement attiré notre attention d'après le *design* de l'essai MAPT. En effet, la phase intensive de l'intervention portant sur la période T0-T6 mois de suivi, nous avons supposé que les intensités de transitions pouvaient différer selon ce seuil de coupure. De la même façon, d'après les résultats exploratoires de la section 1, il semblait que les

fonctions cognitives évoluaient différemment entre la première année et les deux autres. Deux modèles homogènes par morceaux pourraient découler de ce constat : le premier en imaginant des changements de transition entre la première année et les deux suivantes, le second en utilisant les années comme seuils de coupures. De la même façon, des modèles semi-markoviens pourraient être appliqués aux données. Ils ont la particularité de considérer les transitions comme dépendantes du temps passé dans un état (temps de séjour dans l'état). Contrairement aux modèles markoviens, dont la distribution du temps de séjour est toujours exponentielle, la distribution des temps de séjours des modèles semi-markoviens peut être définie par un paramètre du modèle.

Pour aller plus loin dans les analyses des modèles de Markov cachés, et en supposant que les écueils rencontrés soient levés, les modèles multivariés devraient être analysés. En effet, ajouter des variables permet de construire un modèle avec une signification biologique plus cohérente et pertinente. Ces covariables peuvent avoir un effet sur les paramètres de distribution état-dépendantes ou sur les probabilités de transition de la chaîne de Markov cachée [266] (voir section 2.2.2 du Chapitre 5). Il serait donc intéressant de tester d'une part les effets sur les probabilités de transition, d'autre part sur les probabilités initiales, et dans un dernier temps sur l'ensemble du modèle. De la même façon, une étape suivante consisterait à prendre en considération des données manquantes. Pour notre exploration, nous n'avons considéré que les données complètes. Nous avons déjà évoqué précédemment les répercussions de ce choix, notamment en terme de sur-sélection des sujets. Les prendre correctement en considération mènerait à un modèle d'autant plus pertinent. Dans un dernier temps, les transitions avec covariables dépendantes du temps pourraient être explorées pour finaliser l'analyse [220, 230].

De façon plus générale sur ce travail, nous avions évoqué trois principales spécificités des essais de prévention. Le caractère longitudinal des données implique de savoir comment le traiter dans les analyses, mais surtout que la mesure du niveau cognitif global est hétérogène et soumis à l'incertitude de mesure. Cette variabilité des données a été abordée dans ce travail, sous différents angles, mais nous aurions pu l'analyser autrement. En effet, ces données sont principalement bruitées, ce qui signifie qu'il est difficile, sur une courte période de temps, de dissocier le bruit contenu dans les données d'un réel déclin des fonctions cognitives. Ce point aurait pu être investigué par des analyses sur données simulées. Différents bruits pourraient ainsi être générés et leur impact sur les données analysé.

À la lumière du présent manuscrit et contrairement à ce que Garès et al. [52, 51] avaient trouvé dans le cas de l'analyse d'un critère de jugement binaire, ce travail confirme que les modèles classiquement utilisés pour l'analyse du déclin des fonctions cognitives peuvent être adaptés pour détecter les effets d'une intervention. Le recours à des modèles alternatifs susceptibles de prendre en compte les *challenges* méthodologiques de ces essais doivent être poursuivis. D'autres perspectives peuvent être envisagées, comme le recours à un *design* particulier où une analyse statistique simple serait adaptée. Des essais adaptatifs, avec des périodes plus brèves d'intervention, pourraient être mobilisés.

Table des figures

3.1 Schéma du recrutement des sujets analysés dans l'essai MAPT et détail des populations d'analyses (per protocol et en intention de traiter modifiée). Source : [13]. 39

3.2 Descriptif des caractéristiques relevées à baseline pour les sujets en intention de traiter modifiée de l'essai MAPT (N = 1525 sujets). Source : [13]. 41

4.1 Schéma simplifié d'une variable longitudinale simple : la donnée de survie. . 44

4.2 Schéma simplifié d'une variable longitudinale classique : la mesure répétée dans le temps. 44

4.3 Schéma simplifié d'une variable longitudinale théorique : la donnée fonctionnelle ou trajectoire. 45

4.4 Exemple, pour un sujet, de la définition de la catégorie de variations du critère de jugement principal quantitatif de l'essai de prévention MAPT. La valeur en t+1 est en-deçà de la borne inférieur de l'intervalle $I_t = [Z_t - bs_{t+1}; Z_t + as_{t+1}]$, impliquant que l'individu est dans la catégorie déclin à t+1. 54

5.1 Diagramme résumant l'ensemble des méthodes utilisées dans le présent travail pour ré-analyser les données de l'essai de prévention MAPT. 59

5.2 Illustration de l'algorithme des *k-means* classique sur un jeu de données simulées de 150 individus [196]. À gauche, la population totale. Ensuite, les résultats de l'algorithme pour différentes valeurs du nombre de groupes k variant de 2 à 4. Les couleurs indiquent l'appartenance aux différents groupes. 64

5.3 Illustration de l'impact des distances classiques, de la moyenne classique, de la distance de Fréchet et de la moyenne de Fréchet [14]. 69

5.4 Exemple de création d'un dendrogramme : les individus, représentés par les feuilles, sont indiqués par les lettres a, b, c, d et e. Ils sont successivement regroupées en *clusters* dont les différents nœuds sont numérotés de 1 à 4 [**?**]. 73

5.5 Schéma des règles de décisions pour un patient donné, dans le cas de deux interventions médicales. 78

5.6 Graphique dirigé d'un modèle de Markov observable, où X_t les états et les flèchent indiquent les transitions possibles entre les états. 82

5.7 Graphique dirigé d'un modèle de Markov caché, avec X_t les observations et C_t les états cachés [225]. 86

6.1 *Spaghetti plot* des données observées pour les trois *clusters* définis par la méthode `kmlShape` pour les valeurs du Z-score composite, toutes trajectoires confondues à gauche et à droite avec séparation des trajectoires de chaque groupe identifié, MAPT (N = 1679 sujets). 93

6.2 Représentation des valeurs longitudinales du Z-score composite (données brutes à gauche) et des taux d'accroissement de ce dernier (à droite) pour la population analysée, MAPT (N = 1143 sujets). Période 1 = T0-T6 mois, période 2 = T6-T12, période 3 = T12-T24 et période 4 = T24-T36 mois de suivi. . . 97

6.3 Inertie associée au dendrogramme de la CAH à partir des 4 taux d'accroissement du Z-score composite en fonction du nombre de classes, MAPT (N = 1143 sujets). Sauts d'inertie identifiés à 2 groupes (en vert), à 3 groupes (en rouge), à 4 groupes (en bleu) et à 6 groupes (en violet). 98

6.4 *Spaghetti plot* des données observées pour un partitionnement en 3 groupes définis par l'inertie au dendrogramme de la CAH sur les taux d'accroissement du Z-score composite, MAPT (N = 1143 sujets). 99

6.5 Représentation des différentes couleurs sélectionnées pour la méthode de sériation avant permutation des lignes pour les taux d'accroissement du Z-score composite, MAPT (N = 1143 sujets). Rouge = augmentation ; blanc = constant ; bleu = diminution. 101

6.6 Représentation des différentes couleurs sélectionnées pour la méthode de sériation après permutation des lignes pour les taux d'accroissement du Z-score composite, MAPT (N = 1143 sujets). Rouge = augmentation ; blanc = constant ; bleu = diminution. 102

6.7 *Spaghetti plot* des données observées pour les trois *clusters* définis par la méthode de sériation pour les taux d'accroissement du Z-score composite, MAPT (N = 300 sujets). 103

6.8 *Spaghetti plot* des données observées pour les deux groupes définis par la classification dichotomique sur critère de répondeur, pour les taux d'accroissement du Z-score composite, MAPT (N = 1145 sujets). 107

6.9 Représentation d'un modèle de Markov observable à 2 états. Les p_{ij} correspondent aux probabilités de transitions pour aller de l'état i à l'état j. . . . 115

6.10 Représentation du modèle de Markov homogène à 2 états après ajustement sur les covariables, MAPT (N = 1525 sujets). 117

6.11 Représentation du modèle de Markov homogène à 3 états (dont l'état 3 pour données manquantes) après ajustement sur les covariables, MAPT (N = 1544 sujets). 120

6.12 Représentation du modèle de Markov homogène à 3 états de variations strictes après ajustement sur les covariables, MAPT (N = 1525 sujets). 123

6.13 Représentation du modèle de Markov homogène à 3 états de variations souples après ajustement sur les covariables, MAPT (N = 1525 sujets). ** : sexe+ Age+ BMI+ GDS+ EVA2+ Fried+ hyperT 124

Liste des tableaux

2.1 Caractéristiques principales de la construction des différents essais. 18

2.2 Synthèse des caractéristiques méthodologiques des différents essais. 20

3.1 Construction de la variable Phénotype en qualitative binaire. 35

4.1 Descriptif des caractéristiques à l'inclusion des 1679 sujets inclus dans l'essai MAPT. 48

4.2 Descriptif des caractéristiques à l'inclusion des 1143 sujets ayant un suivi complet sur les 36 mois de l'essai MAPT. 49

4.3 Descriptif des caractéristiques à l'inclusion des 1525 sujets ayant au moins une visite après inclusion dans l'essai MAPT. 51

4.4 Description des profils extrêmes rencontrés dans l'essai MAPT au cours du suivi. 52

4.5 Récapitulatif des variables quantitatives significativement différentes entre les deux profils extrêmes identifiés dans l'essai MAPT, sur les différentes périodes de suivi. 58

6.1 Récapitulatif des caractéristiques à *baseline* significativement différentes pour les 3 groupes identifiés pour le Z-score composite par la méthode des centres mobiles basée sur la forme des trajectoires (`kmlShape`), MAPT (N = 1679 sujets). 95

6.2 Récapitulatif des caractéristiques à *baseline* significativement différentes pour les 3 groupes identifiés pour le Z-score composite par la méthode de la CAH, MAPT (N = 1143 sujets). 99

6.3 Récapitulatif des caractéristiques à *baseline* non significativement différentes pour les 3 groupes identifiés par la méthode de sériation pour les taux d'accroissement du Z-score composite, MAPT (N = 300 sujets). 104

6.4 Description des 2 groupes identifiés par la classification par statut répondeur, pour les taux d'accroissement du Z-score composite de l'essai MAPT, au cours du suivi (N = 1145 sujets). 107

6.5 Récapitulatif des caractéristiques significativement différentes pour les 2 groupes identifiés par la classification par statut répondeur pour les taux d'accroissement du Z-score composite, MAPT (N = 1145 sujets). 108

6.6 Résumé des effets mesurés par l'ANOVA à deux facteurs, sur des taux d'accroissement entre début et fin de suivi, sur les différentes périodes possibles de l'essai MAPT. 111

6.7 Récapitulatif des risques relatifs de transition du modèle homogène à 2 états (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). 116

6.14 Temps d'exécution des différents modèles de Markov cachés implémentés durant notre stratégie, MAPT (N = 1143 sujets). 128

6.8 Récapitulatif des risques relatifs de transition du modèle homogène à 3 états dont 1 pour données manquantes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1544 sujets). Partie 1 129

6.9 Récapitulatif des risques relatifs de transition du modèle homogène à 3 états dont 1 pour données manquantes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1544 sujets). Partie 2 130

6.10 Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variation strictes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 1 131

6.11 Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variation strictes (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 2 132

6.12 Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variations souples (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 1 133

6.13 Récapitulatif des risques relatifs de transition du modèle homogène à 3 états de variations souples (et l'intervalle de confiance à 95 %) pour chaque covariable, MAPT (N = 1525 sujets). Partie 2 134

7.1 Résumé, comparaisons et recommandations pour l'utilisation des méthodes d'exploration des trajectoires cognitives appliquées dans ce mémoire. 139

7.2 Synthèse des différents paramètres des quatre modèles de Markov observables étudiés pour les données de l'essai de prévention MAPT. 141

8.1 Résumé des différentes variables quantitatives de l'essai MAPT exploitées dans les analyses post-hoc du présent mémoire. 174

8.2 Résumé des différentes variables qualitatives de l'essai MAPT exploitées dans les analyses post-hoc du présent mémoire. 175

Bibliographie

[1] Gong H, Xun X and Zhou Y. Profile clustering in clinical trials with longitudinal and functional data methods. *Journal of biopharmaceutical statistics* 2019 ; 29(3) : 541–557. DOI :10.1080/10543406.2019.1572614.

[2] Gardette V, Coley N, Toulza O et al. Attrition in geriatric research : how important is it and how should it be dealt with ? *The Journal of nutrition, health & aging* 2007 ; 11(3) : 265–271.

[3] Vellas B, Andrieu S, Sampaio C et al. Disease-modifying trials in alzheimer's disease : a european task force consensus. *The Lancet Neurology* 2007 ; 6(1) : 56–62. DOI : 10.1016/S1474-4422(06)70677-9.

[4] Vellas B, Bateman R, Blennow K et al. Endpoints for pre-dementia ad trials : a report from the eu/us/ctad task force. *The journal of prevention of Alzheimer's disease* 2015 ; 2(2) : 128–135. DOI :10.14283/jpad.2015.55.

[5] Thiébaut R and Jacqmin-Gadda H. Mixed models for longitudinal left-censored repeated measures. *Computer methods and programs in biomedicine* 2004 ; 74(3) : 255–260. DOI :10.1016/j.cmpb.2003.08.004.

[6] Proust-Lima C, Amieva H and Jacqmin-Gadda H. Analysis of multivariate mixed longitudinal data : a flexible latent process approach. *The British journal of mathematical and statistical psychology* 2013 ; 66(3) : 470–487. DOI :10.1111/bmsp.12000.

[7] Chakraborty B and Moodie EEM. *Statistical Methods for Dynamic Treatment Regimes : Reinforcement Learning, Causal Inference, and Personalized Medicine*. Springer, 2013. DOI :10.1007/978-1-4614-7428-9.

[8] Winblad B, Amouyel P, Andrieu S et al. Defeating alzheimer's disease and other dementias : a priority for european science and society. *The Lancet Neurology* 2016 ; 15(5) : 455–532. DOI :10.1016/S1474-4422(16)00062-4.

[9] Coley N, Gallini A, Ousset PJ et al. Evaluating the clinical relevance of a cognitive composite outcome measure : An analysis of 1414 participants from the 5-year guidage alzheimer's prevention trial. *Alzheimer's & Dementia* 2016 ; 12(12) : 1216–1225. DOI : 10.1016/j.jalz.2016.06.002.

[10] Vellas B, Andrieu S, Sampaio C et al. Endpoints for trials in alzheimer's disease : a european task force consensus. *The Lancet Neurology* 2008 ; 7(5) : 436–450. DOI : 10.1016/S1474-4422(08)70087-5.

[11] Crivello F, Tzourio-Mazoyer N, Tzourio C et al. Longitudinal assessment of global and regional rate of grey matter atrophy in 1,172 healthy older adults : modulation by sex and age. *PLoS One* 2014 ; 9(12) : e114478. DOI :10.1371/journal.pone.0114478.

[12] Andrieu S, Coley N, Lovestone S et al. Prevention of sporadic alzheimer's disease : lessons learned from clinical trials and future directions. *The Lancet Neurology* 2015 ; 14(9) : 926–944. DOI :10.1016/S1474-4422(15)00153-2.

[13] Andrieu S, Guyonnet S, Coley N et al. Effect of long-term omega 3 polyunsaturated fatty acid supplementation with or without multidomain intervention on cognitive function in elderly adults with memory complaints (mapt) : a randomised, placebo-controlled trial. *The Lancet Neurology* 2017 ; 16(5) : 377–389. DOI : 10.1016/S1474-4422(17)30040-6.

[14] Genolini C, Ecochard R, Benghezal M et al. kmlshape : An efficient method to cluster longitudinal data (time-series) according to their shapes. *PloS one* 2016 ; 11(6) : e0150738. DOI :10.1371/journal.pone.0150738.

[15] Johnson SC. Hierarchical clustering schemes. *Psychometrika* 1967 ; 32(3) : 241–254. DOI :10.1007/BF02289588. URL `https://EconPapers.repec.org/RePEc:spr:psycho:v:32:y:1967:i:3:p:241-254`.

[16] Hahsler M, Hornik K and Buchta C. Getting things in order : an introduction to the r package seriation. *Journal of Statistical Software* 2008 ; 25(3) : 1–34. DOI : 10.18637/jss.v025.i03. URL `http://www.jstatsoft.org/v25/i03/`.

[17] Lange JK, DiSegna ST, Yang W et al. Using cluster analysis to identify patient factors linked to differential functional gains after total knee arthroplasty. *The Journal of arthroplasty* 2019 ; 35. DOI :10.1016/j.arth.2019.08.039.

[18] Gentleman R, Lawless J, Lindsey J et al. Multi-state markov models for analysing incomplete disease history data with illustrations for hiv disease. *Statistics in medicine* 1994 ; 13(8) : 805–821. DOI :10.1002/sim.4780130803.

[19] Cook RJ, Yi GY, Lee KA et al. A conditional markov model for clustered progressive multistate processes under incomplete observation. *Biometrics* 2004 ; 60(2) : 436–443. DOI :10.1111/j.0006-341X.2004.00188.x.

[20] Rabiner LR. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* 1989 ; 77(2) : 257–286. DOI :10.1109/5.18626.

[21] ANSM. Qu'est ce qu'un essai clinique ?, 2020. URL `https://www.ansm.sante.fr/Activites/Essais-cliniques/Qu-est-ce-qu-un-essai-clinique/(offset)/2`. Last accessed 11 March 2020.

[22] Amoric M. Le premier essai clinique scientifique-james lind, londres, 1753. *Revue d'Orthopédie Dento-Faciale* 2012 ; 46(3) : 343–345. DOI :10.1051/odf/2012307.

[23] Lee JS, Ahn S, Lee KH et al. Consort 2010 statement : Updated guidelines for reporting parallel group randomised trials. *Epidemiology and health* 2014 ; 36(1). DOI : 10.4178/epih/e2014029.

[24] Guideline IHT. Choice of control group and related issues in clinical trials e10. *Choice E* 2000 ; 10.

[25] Jaillon P. L'essai clinique contrôlé randomisé. *Bulletin de l'Académie Nationale de Médecine* 2007 ; 191(4-5) : 739–758. DOI :10.1016/S0001-4079(19)33007-9.

[26] Jennison C and Turnbull BW. *Group sequential methods with applications to clinical trials*. Chapman and Hall/CRC, 1999.

[27] Hayes RJ and Moulton LH. *Cluster randomised trials, second edition*. Chapman and Hall/CRC, 2017. DOI :10.4324/9781315370286.

[28] Meinert CL. *Clinical Trials : Design, Conduct and Analysis*, *Monographs in Epidemiology and Biostatistics*, volume 39. Oxford University Press, USA, 2012. ISBN 978-0195387889.

[29] Pocock SJ. *Clinical trials : a practical approach*. John Wiley & Sons, 2013. ISBN 978-1118793916. DOI :10.1002/9781118793916.

[30] McKee AE, Farrell AT, Pazdur R et al. The role of the us food and drug administration review process : clinical trial endpoints in oncology. *The oncologist* 2010 ; 15 Suppl 1 : 13–8. DOI :10.1634/theoncologist.2010-S1-13.

[31] Creusier J and Biétry F. Analyse comparative des méthodes de classifications. *RIMHE : Revue Interdisciplinaire Management, Homme Entreprise* 2014 ; 10(1) : 105–123. DOI :10.3917/rimhe.010.0105.

[32] Jin Y, Hayes DF, Li L et al. Estrogen receptor genotypes influence hot flash prevalence and composite score before and after tamoxifen therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008 ; 26(36) : 5849–54. DOI :10.1200/JCO.2008.16.8377.

[33] Cummings JL. Challenges to demonstrating disease-modifying effects in alzheimer's disease clinical trials. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2006 ; 2(4) : 263–271. DOI :10.1016/j.jalz.2006.07.001.

[34] Goodsaid F and Frueh F. Biomarker qualification pilot process at the us food and drug administration. *The AAPS journal* 2007 ; 9(1) : E105–E108. DOI :10.1208/aapsj0901010.

[35] Fleming TR. Surrogate endpoints and fda's accelerated approval process. *Health affairs (Project Hope)* 2005 ; 24(1) : 67–78. DOI :10.1377/hlthaff.24.1.67.

[36] Katz R. Biomarkers and surrogate markers : an fda perspective. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics* 2004 ; 1(2) : 189–195. DOI :10.1602/neurorx.1.2.189.

[37] Bethel MA, Holman R, Haffner SM et al. Determining the most appropriate components for a composite clinical trial outcome. *American heart journal* 2008 ; 156(4) : 633–640. DOI :10.1016/j.ahj.2008.05.018.

[38] Cummings J, Ritter A and Zhong K. Clinical trials for disease-modifying therapies in alzheimer's disease : a primer, lessons learned, and a blueprint for the future. *Journal of Alzheimer's Disease* 2018 ; 64(s1) : S3–S22. DOI :10.3233/JAD-179901.

[39] Food, Administration D et al. Guidance for industry alzheimer's disease : Developing drugs for the treatment of early stage disease. draft guidance, 2013. us department of health and human services food and drug administration. *Center for Drug Evaluation and Research (CDER) http ://www fda gov/downloads/Drugs/GuidanceCompliance-RegulatoryInformation/Guidances/UCM338287 pdf Updated January* 2013 ; 16.

[40] Donohue MC, Sperling RA, Salmon DP et al. The preclinical alzheimer cognitive composite : measuring amyloid-related decline. *JAMA neurology* 2014 ; 71(8) : 961–970. DOI :10.1001/jamaneurol.2014.803.

[41] Ayutyanont N, Langbaum J, Hendrix S et al. The alzheimer's prevention initiative composite cognitive test score : sample size estimates for the evaluation of preclinical alzheimer's disease treatments in presenilin 1 e280a mutation carriers. *The Journal of clinical psychiatry* 2014 ; 75(6) : 652. DOI :10.4088/JCP.13m08927.

[42] Cordoba G, Schwartz L, Woloshin S et al. Definition, reporting, and interpretation of composite outcomes in clinical trials : systematic review. *BMJ (Clinical research ed)* 2010 ; 341 : c3920. DOI :10.1136/bmj.c3920.

[43] Stolberg HO, Norman G and Trop I. Randomized controlled trials. *American Journal of Roentgenology* 2004 ; 183(6) : 1539–1544.

[44] Phillips A and Haudiquet V. Ich e9 guideline 'statistical principles for clinical trials' : a case study. *Statistics in medicine* 2003 ; 22(1) : 1–11. DOI :10.1002/sim.1328.

[45] Robertson MC, Campbell AJ and Herbison P. Statistical analysis of efficacy in falls prevention trials. *The Journals of Gerontology Series A : Biological Sciences and Medical Sciences* 2005 ; 60(4) : 530–534. DOI :10.1093/gerona/60.4.530.

[46] Ancelle T. *Statistique, épidémiologie*. Editions Maloine, 2017. ISBN 978-2224035225.

[47] Xiong C, Van Belle G, Miller JP et al. Designing clinical trials to test disease-modifying agents : application to the treatment trials of alzheimer's disease. *Clinical trials (London, England)* 2011 ; 8(1) : 15–26. DOI :10.1177/1740774510392391.

[48] Cummings J and Fox N. Defining disease modifying therapy for alzheimer's disease. *The journal of prevention of Alzheimer's disease* 2017 ; 4(2) : 109–115. DOI :10.14283/jpad.2017.12.

[49] McGhee DJ, Ritchie CW, Zajicek JP et al. A review of clinical trial designs used to detect a disease-modifying effect of drug therapy in alzheimer's disease and parkinson's disease. *BMC neurology* 2016 ; 16(1) : 92. DOI :10.1186/s12883-016-0606-3.

[50] Brandt KD and Mazzuca SA. Lessons learned from nine clinical trials of disease-modifying osteoarthritis drugs. *Arthritis & Rheumatism* 2005 ; 52(11) : 3349–3359. DOI :10.1002/art.21409.

[51] Garès V, Andrieu S, Dupuy JF et al. On the fleming–harrington test for late effects in prevention randomized controlled trials. *Journal of Statistical Theory and Practice* 2017 ; 11(3) : 418–435. DOI :10.1080/15598608.2017.1295889.

[52] Garès V, Andrieu S, Dupuy JF et al. An omnibus test for several hazard alternatives in prevention randomized controlled clinical trials. *Statistics in medicine* 2015 ; 34(4) : 541–557. DOI :10.1002/sim.6366.

[53] Mura T, Proust-Lima C, Akbaraly T et al. Chronic use of benzodiazepines and latent cognitive decline in the elderly : results from the three-city study. *European Neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* 2013 ; 23(3) : 212–223. DOI :10.1016/j.euroneuro.2012.05.004.

[54] Muntoni F, Domingos J, Manzur AY et al. Categorising trajectories and individual item changes of the north star ambulatory assessment in patients with duchenne muscular dystrophy. *PloS one* 2019 ; 14(9) : e0221097. DOI :10.1371/journal.pone.0221097.

[55] Niu H, Álvarez-Álvarez I, Guillén-Grima F et al. Prevalence and incidence of alzheimer's disease in europe : A meta-analysis. *Neurologia* 2017 ; 32(8) : 523–532. DOI : 10.1016/j.nrl.2016.02.016.

[56] Prince M, Bryce R, Albanese E et al. The global prevalence of dementia : a systematic review and metaanalysis. *Alzheimer's & Dementia : the journal of the Alzheimer's Association* 2013 ; 9(1) : 63–75. DOI :10.1016/j.jalz.2012.11.007.

[57] Soto Martin ME. *Le déclin cognitif dans la maladie d'Alzheimer*. PhD Thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2014.

[58] Amieva H, Andrieu S, Berr C et al. Maladie d'alzheimer : enjeux scientifiques, médicaux et sociétaux. Research report, Institut national de la santé et de la recherche médicale(INSERM), 2007. URL `https://hal-lara.archives-ouvertes.fr/hal-01570630`.

[59] Nelson P, Head E, Schmitt F et al. Alzheimer's disease is not "brain aging" : neuropathological, genetic, and epidemiological human studies. *Acta neuropathologica* 2011 ; 121(5) : 571–587. DOI :10.1007/s00401-011-0826-y.

[60] HCSP. Prévention de la maladie d'alzheimer et des maladies apparentées (mama). *Avis et Rapports* 2017 ; .

[61] Rabinovici G, Gatsonis C, Apgar C et al. Association of amyloid positron emission tomography with subsequent change in clinical management among medicare beneficiaries with mild cognitive impairment or dementia. *JAMA* 2019 ; 321(13) : 1286–1294. DOI :10.1001/jama.2019.2000.

[62] Cleveland DW, Hwo SY and Kirschner MW. Purification of tau, a microtubule-associated protein that induces assembly of microtubules from purified tubulin. *Journal of molecular biology* 1977 ; 116(2) : 207–225. DOI :10.1016/0022-2836(77)90213-3.

[63] Harada CN, Love MCN and Triebel KL. Normal cognitive aging. *Clinics in geriatric medicine* 2013 ; 29(4) : 737–752. DOI :10.1016/j.cger.2013.07.002.

[64] Sperling R, Aisen P, Beckett L et al. Toward defining the preclinical stages of alzheimer's disease : Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2011 ; 7(3) : 280–292. DOI : 10.1016/j.jalz.2011.03.003.

[65] DeKosky S, Fitzpatrick A, Ives D et al. The ginkgo evaluation of memory (gem) study : design and baseline data of a randomized trial of ginkgo biloba extract in prevention of dementia. *Contemporary clinical trials* 2006 ; 27(3) : 238–253. DOI : 10.1016/j.cct.2006.02.007.

[66] Ardila A and Rosselli M. Neuropsychological characteristics of normal aging. *Developmental Neuropsychology - DEVELOP NEUROPSYCHOL* 1989 ; 5(4) : 307–320. DOI :10.1080/87565648909540441.

[67] Desgranges B, Eustache F, Rioux P et al. Memory disorders in alzheimer's disease and the organization of human memory. *Cortex ; a journal devoted to the study of the nervous system and behavior* 1996 ; 32(3) : 387–412. DOI :10.1016/S0010-9452(96) 80001-3.

[68] Singh-Manoux A, Kivimaki M, Glymour MM et al. Timing of onset of cognitive decline : results from whitehall ii prospective cohort study. *BMJ (Clinical research ed)* 2012 ; 344 : d7622. DOI :10.1136/bmj.d7622.

[69] Goldman-Rakic PS. *Circuitry of primate prefrontal cortex and regulation of behavior by representational memory*, volume 5. Wiley Online Library. ISBN 9780470650714, 1987. DOI :10.1002/cphy.cp010509.

[70] Shimamura AP and Jurica PJ. Memory interference effects and aging : Findings from a test of frontal lobe function. *Neuropsychology* 1994 ; 8(3) : 408–412. DOI : 10.1037/0894-4105.8.3.408.

[71] Feldman H and Gracon S. Alzheimer's disease : symptomatic drugs under development. *Clinical Diagnosis and Management of Alzheimer's disease* 1996 ; : 239–259.

[72] Coley N, Andrieu S, Gardette V et al. Dementia prevention : methodological explanations for inconsistent results. *Epidemiologic reviews* 2008 ; 30(1) : 35–66. DOI : 10.1093/epirev/mxn010.

[73] Amieva H, Jacqmin-Gadda H, Orgogozo JM et al. The 9 year cognitive decline before dementia of the alzheimer type : a prospective population-based study. *Brain* 2005 ; 128(5) : 1093–1101. DOI :10.1093/brain/awh451. URL `https://doi.org/10.1093/brain/awh451`. `https://academic.oup.com/brain/article-pdf/128/5/1093/780731/awh451.pdf`.

[74] DeKosky S, Williamson J, Fitzpatrick A et al. Ginkgo biloba for prevention of dementia : a randomized controlled trial. *JAMA : the journal of the American Medical Association* 2008 ; 300(19) : 2253–2262. DOI :10.1001/jama.2008.683.

[75] Andrieu S, Ousset PJ, Coley N et al. Guidage study : a 5-year double blind, randomised trial of egb 761 for the prevention of alzheimer's disease in elderly subjects with memory complaints. i. rationale, design and baseline data. *Current Alzheimer Research* 2008 ; 5(4) : 406–415. DOI :10.2174/156720508785132271.

[76] Vellas B, Coley N, Ousset PJ et al. Long-term use of standardised ginkgo biloba extract for the prevention of alzheimer's disease (guidage) : a randomised placebo-controlled trial. *The Lancet Neurology* 2012 ; 11(10) : 851–859. DOI :10.1016/S1474-4422(12) 70206-5.

[77] Richard E, Van den Heuvel E, van Charante EPM et al. Prevention of dementia by intensive vascular care (prediva) : a cluster-randomized trial in progress. *Alzheimer Disease & Associated Disorders* 2009 ; 23(3) : 198–204. DOI :10.1097/WAD. 0b013e31819783a4.

[78] Van Charante EPM, Richard E, Eurelings LS et al. Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (prediva) : a cluster-randomised controlled trial. *The Lancet* 2016 ; 388(10046) : 797–805. DOI :10.1016/S0140-6736(16) 30950-3.

[79] Kryscio RJ, Abner EL, Caban-Holt A et al. Association of antioxidant supplement use and dementia in the prevention of alzheimer's disease by vitamin e and selenium trial (preadvise). *JAMA neurology* 2017 ; 74(5) : 567–573. DOI :10.1001/jamaneurol. 2016.5778.

[80] Ambrosius W, Sink K, Foy C et al. The design and rationale of a multicenter clinical trial comparing two strategies for control of systolic blood pressure : the systolic blood pressure intervention trial (sprint). *Clinical Trials* 2014 ; 11(5) : 532–546. DOI : 10.1177/1740774514537404.

[81] Williamson J, Pajewski N, Auchus A et al. Effect of intensive vs standard blood pressure control on probable dementia : a randomized clinical trial. *JAMA : the journal of the American Medical Association* 2019 ; 321(6) : 553–561. DOI :10.1001/ jama.2018.21442.

[82] Fielding R, Rejeski J, Blair S et al. The lifestyle interventions and independence for elders study : design and methods. *Journals of Gerontology Series A : Biomedical Sciences and Medical Sciences* 2011 ; 66(11) : 1226–1237. DOI :10.1093/gerona/glr123.

[83] Sink K, Espeland M, Castro C et al. Effect of a 24-month physical activity intervention vs health education on cognitive outcomes in sedentary older adults : the life randomized trial. *JAMA : the journal of the American Medical Association* 2015 ; 314(8) : 781–790. DOI :10.1001/jama.2015.9617.

[84] Ball K, Berch D, Helmers K et al. Effects of cognitive training interventions with older adults : a randomized controlled trial. *JAMA : the journal of the American Medical Association* 2002 ; 288(18) : 2271–2281. DOI :10.1001/jama.288.18.2271.

[85] Ngandu T, Lehtisalo J, Levälahti E et al. Recruitment and baseline characteristics of participants in the finnish geriatric intervention study to prevent cognitive impairment and disability (finger)—a randomized controlled lifestyle trial. *International journal of environmental research and public health* 2014 ; 11(9) : 9345–9360. DOI :10.3390/ ijerph110909345.

[86] Ngandu T, Lehtisalo J, Solomon A et al. 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* 2015 ; 385(9984) : 2255–2263. DOI :10.1016/S0140-6736(15)60461-5.

[87] Vellas B, Carrié I, Guyonnet S et al. Mapt study : a multidomain approach for preventing alzheimer's disease : design and baseline data. *The journal of prevention of Alzheimer's disease (JPAD)* 2014 ; 1(1) : 13. DOI :10.14283/jpad.2014.34. URL `http://www.jpreventionalzheimer.com/all-issues.html?article=11`.

[88] Richard E, Jongstra S, Soininen H et al. Healthy ageing through internet counselling in the elderly : The hatice randomised controlled trial for the prevention of cardiovascular disease and cognitive impairment. *BMJ open* 2016 ; 6(6) : e010806. DOI :10.1136/bmjopen-2015-010806.

[89] Jongstra S, Beishuizen C, Andrieu S et al. Development and validation of an interactive internet platform for older people : the healthy ageing through internet counselling in the elderly study. *Telemedicine and e-Health* 2017 ; 23(2) : 96–104. DOI :10.1089/tmj.2016.0066.

[90] Richard E, Moll van Charante E, Hoevenaar-Blom M et al. Healthy ageing through internet counselling in the elderly (hatice) : a multinational, randomised controlled trial. *The Lancet Digital Health* 2019 ; 1(8) : e424–e434. DOI :10.1016/S2589-7500(19)30153-0.

[91] Lawton MP and Brody EM. Assessment of older people : self-maintaining and instrumental activities of daily living. *The gerontologist* 1969 ; 9(3_Part_1) : 179–186. DOI :10.1093/geront/9.3_Part_1.179.

[92] Army U. *Manual of Direction and Scoring Washington*. 1944.

[93] Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Perceptual and motor skills* 1958 ; 8(3) : 271–276. DOI :10.2466/pms.1958.8.3.271.

[94] Cardebat D, Doyon B, Puel M et al. Formal and semantic lexical evocation in normal subjects. performance and dynamics of production as a function of sex, age and educational level. *Acta Neurologica Belgica* 1990 ; 90(4) : 207–217. URL `http://europepmc.org/abstract/MED/2124031`.

[95] Wilson RS, Li Y, Aggarwal N et al. Education and the course of cognitive decline in alzheimer disease. *Neurology* 2004 ; 63(7) : 1198–1202. DOI :10.1212/01.WNL.0000140488.

[96] Grober E and Buschke H. Genuine memory deficits in dementia. *Developmental neuropsychology* 1987 ; 3(1) : 13–36. DOI :10.1080/87565648709540361.

[97] Pasquier F, Grymonprez L, Lebert F et al. Memory impairment differs in frontotemporal dementia and alzhemier's disease. *Neurocase* 2001 ; 7(2) : 161–171. DOI : 10.1093/neucas/7.2.161.

[98] Derouesné C, Poitreneau J, Hugonot L et al. Le mini-mental state examination (mmse) : un outil pratique pour l'évaluation de l'état til pratique pour l'évaluation cognitif des patients par le clinicien version française consensuelle. *La presse médicale* 1999 ; 948(21) : 1141–1148. DOI :PM-06-1999-28-21-0755-4982-101019-ART87.

[99] Wechsler D and De Lemos MM. *Wechsler adult intelligence scale-revised*. Harcourt Brace Jovanovich, 1981.

[100] Unger JM, van Belle G and Heyman A. Cross-sectional versus longitudinal estimates of cognitive change in nondemented older people : A cerad study. *Journal of the American Geriatrics Society* 1999 ; 47(5) : 559–563. DOI :10.1111/j.1532-5415.1999.tb02570.x.

[101] Amieva H, Carcaillon L, Rouze PLS et al. Cued and uncued memory tests : norms in elderly adults from the 3 cities epidemiological study. *Revue neurologique* 2007 ; 163(2) : 205–221. DOI :10.1016/s0035-3787(07)90392-6.

[102] Giffard B, Desgranges B and Eustache F. Le vieillissement de la mémoire : vieillissement normal et pathologique. *Gérontologie et société* 2001 ; 24(2) : 33–47. DOI :10.3917/gs.097.0033. URL `https://www.cairn.info/revue-gerontologie-et-societe1-2001-2-page-33.htm`.

[103] Kalpouzos G, Eustache F and Desgranges B. La mémoire prospective au cours du vieillissement : déclin ou préservation ? *NPG Neurologie-Psychiatrie-Gériatrie* 2008 ; 8(47) : 25–31. DOI :10.1016/j.npg.2008.06.003.

[104] Rendell PG and Thomson DM. Aging and prospective memory : Differences between naturalistic and laboratory tasks. *The Journals of Gerontology Series B : Psychological Sciences and Social Sciences* 1999 ; 54(4) : P256–P269. DOI :10.1093/geronb/54B.4.P256.

[105] Henry JD, MacLeod MS, Phillips LH et al. A meta-analytic review of prospective memory and aging. *Psychology and aging* 2004 ; 19(1) : 27–39. DOI : 10.1037/0882-7974.19.1.27.

[106] McKhann G, Drachman D, Folstein M et al. Clinical diagnosis of alzheimer's disease report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease. *Neurology* 1984 ; 77(7) : 939–939. DOI :10.1212/01.wnl.0000400650.92875.cf.

[107] Nutter-Upham KE, Saykin AJ, Rabin LA et al. Verbal fluency performance in amnestic mci and older adults with cognitive complaints. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 2008 ; 23(3) : 229–241. DOI :10.1016/j.acn.2008.01.005.

[108] Heude B, Ducimetière P and Berr C. Cognitive decline and fatty acid composition of erythrocyte membranes—the eva study. *The American journal of clinical nutrition* 2003 ; 77(4) : 803–808. DOI :10.1093/ajcn/77.4.803.

[109] Ávila-Funes JA, Amieva H, Barberger-Gateau P et al. Cognitive impairment improves the predictive validity of the phenotype of frailty for adverse health outcomes : the three-city study. *Journal of the American Geriatrics Society* 2009 ; 57(3) : 453–461. DOI :10.1111/j.1532-5415.2008.02136.x.

[110] Lechevallier-Michel N, Fabrigoule C, Lafont S et al. Normes pour le mmse, le test de rétention visuelle de benton, le set test d'isaacs, le sous-test des codes de la wais et le test de barrage de zazzo chez des sujets âgés de 70 ans et plus : données de la cohorte paquid. *Revue Neurologique* 2004 ; 160(11) : 1059–1070. DOI :10.1016/S0035-3787(04)71143-1.

[111] Wolinsky FD, Coe RM, Miller DK et al. Measurement of the global and functional dimensions of health status in the elderly. *Journal of Gerontology* 1984 ; 39(1) : 88–92. DOI :10.1093/geronj/39.1.88.

[112] Dartigues JF, Commenges D, Letenneur L et al. Cognitive predictors of dementia in elderly community residents. *Neuroepidemiology* 1997 ; 16(1) : 29–39. DOI :10.1159/000109668.

[113] Folstein MF, Folstein SE and McHugh PR. "mini-mental state" : a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research* 1975 ; 12(3) : 189–198. DOI :10.1016/0022-3956(75)90026-6.

[114] Mahurin RK, DeBettignies BH and Pirozzolo FJ. Structured assessment of independent living skills : preliminary report of a performance measure of functional abilities in dementia. *Journal of Gerontology* 1991 ; 46(2) : P58–P66. DOI : 10.1093/geronj/46.2.p58.

[115] Bherer L, Belleville S and Hudon C. Le déclin des fonctions exécutives au cours du vieillissement normal, dans la maladie d'alzheimer et dans la démence frontotemporale. *Psychologie & NeuroPsychiatrie du vieillissement* 2004 ; 2(3) : 181–189.

[116] McCrae RR, Arenberg D and Costa PT. Declines in divergent thinking with age : Cross-sectional, longitudinal, and cross-sequential analyses. *Psychology and Aging* 1987 ; 2(2) : 130–137. DOI :10.1037//0882-7974.2.2.130.

[117] Boone KB, Miller BL, Lesser IM et al. Performance on frontal lobe tests in healthy, older individuals. *Developmental Neuropsychology* 1990 ; 6(3) : 215–223. DOI :10.1080/87565649009540462.

[118] Pasquier F, Lebert F, Grymonprez L et al. Verbal fluency in dementia of frontal lobe type and dementia of alzheimer type. *Journal of Neurology, Neurosurgery & Psychiatry* 1995 ; 58(1) : 81–84. DOI :10.1136/jnnp.58.1.81.

[119] Henry JD, Crawford JR and Phillips LH. Verbal fluency performance in dementia of the alzheimer's type : a meta-analysis. *Neuropsychologia* 2004 ; 42(9) : 1212–1222. DOI :10.1016/j.neuropsychologia.2004.02.001.

[120] Murphy KJ, Rich JB and Troyer AK. Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of alzheimer's type dementia. *Journal of the International Neuropsychological Society : JINS* 2006 ; 12(4) : 570–574. DOI : 10.1017/S1355617706060590.

[121] Monsch AU, Bondi MW, Butters N et al. Comparisons of verbal fluency tasks in the detection of dementia of the alzheimer type. *Archives of neurology* 1993 ; 49(12) : 1253–1258. DOI :10.1001/archneur.1992.00530360051017.

[122] Gifford KA, Liu D, Lu Z et al. The source of cognitive complaints predicts diagnostic conversion differentially among nondemented older adults. *Alzheimer's & Dementia : the journal of the Alzheimer's Association* 2014 ; 10(3) : 319–327. DOI :10.1016/j.jalz.2013.02.007.

[123] Benton A and Hamsher K. Multilingual aphasia examination : Manual of instructions. *Iowa City : University of Iowa Google Scholar* 1976 ; .

[124] Hughes CP, Berg L, Danziger W et al. A new clinical scale for the staging of dementia. *The British journal of psychiatry : the journal of mental science* 1982 ; 140(6) : 566–572. DOI :10.1192/bjp.140.6.566.

[125] Delrieu J, Andrieu S, Pahor M et al. Neuropsychological profile of "cognitive frailty" subjects in mapt study. *The journal of prevention of Alzheimer's disease* 2016 ; 3(3) : 151–159. DOI :10.14283/jpad.2016.94.

[126] Schafer KA, Tractenberg RE, Sano M et al. Reliability of monitoring the clinical dementia rating in multicenter clinical trials. *Alzheimer disease and associated disorders* 2004 ; 18(4) : 219–222. URL `https://europepmc.org/articles/PMC4367865`.

[127] Morris JC. Clinical dementia rating : a reliable and valid diagnostic and staging measure for dementia of the alzheimer type. *International psychogeriatrics / IPA* 1997 ; 9 Suppl 1(S1) : 173–177. DOI :10.1017/S1041610297004870.

[128] Morris JC. The clinical dementia rating (cdr) : current version and scoring rules. *Neurology* 1993 ; 43 : 2412–2414. DOI :10.1212/WNL.43.11.2412-a.

[129] Chaves MLF, Camozzato AL, Godinho C et al. Validity of the clinical dementia rating scale for the detection and staging of dementia in brazilian patients. *Alzheimer Disease & Associated Disorders* 2007 ; 21(3) : 210–217. DOI :10.1097/WAD.0b013e31811ff2b4.

[130] Morris J, Storandt M, McKeel D et al. Cerebral amyloid deposition and diffuse plaques in“normal”aging evidence for presymptomatic and very mild alzheimer's disease. *Neurology* 1996 ; 46(3) : 707–719. DOI :10.1212/WNL.46.3.707.

[131] Berg L, McKeel DW, Miller JP et al. Neuropathological indexes of alzheimer's disease in demented and nondemented persons aged 80 years and older. *Archives of Neurology* 1993 ; 50(4) : 349–358. DOI :10.1001/archneur.1993.00540040011008.

[132] Sheikh JI and Yesavage JA. Geriatric depression scale (gds) : recent evidence and development of a shorter version. *Clinical Gerontologist : The Journal of Aging and Mental Health* 1986 ; 5 : 165–173. DOI :10.1300/j018v05n0109.

[133] Burke WJ, Roccaforte WH and Wengel SP. The short form of the geriatric depression scale : a comparison with the 30-item form. *Journal of geriatric psychiatry and neurology : Topics in geriatrics* 1991 ; 4(3) : 173–178. DOI :10.1177/089198879100400310.

[134] Jorm AF. History of depression as a risk factor for dementia : an updated review. *The Australian & New Zealand Journal of Psychiatry* 2002 ; 35(6) : 776–781. DOI : 10.1046/j.1440-1614.2001.00967.x.

[135] Ownby RL, Crocco E, Acevedo A et al. Depression and risk for alzheimer disease : systematic review, meta-analysis, and metaregression analysis. *Archives of general psychiatry* 2006 ; 63(5) : 530–538. DOI :10.1001/archpsyc.63.5.530.

[136] Da Silva J, Goncalves-Pereira M, Xavier M et al. Affective disorders and risk of developing dementia : systematic review. *The British journal of psychiatry : the journal of mental science* 2013 ; 202(3) : 177–186. DOI :10.1192/bjp.bp.111.101931.

[137] Bazin N and Bratu L. Depression in the elderly : prodroma or risk factor for dementia ? a critical review of the literature. *Gériatrie et Psychologie Neuropsychiatrie du Vieillissement* 2014 ; 12(3) : 289–297. DOI :10.1684/pnv.2014.0490.

[138] Hausner L, Froelich L, Gardette V et al. Regional variation on the presentation of alzheimer's disease patients in memory clinics within europe : data from the ictus study. *Journal of Alzheimer's Disease* 2010 ; 21(1) : 155–165. DOI : 10.3233/JAD-2010-091489.

[139] De Craen AJ, Heeren T and Gussekloo J. Accuracy of the 15-item geriatric depression scale (gds-15) in a community sample of the oldest old. *International journal of geriatric psychiatry* 2003 ; 18(1) : 63–66. DOI :10.1002/gps.773.

[140] D'Ath P, Katona P, Mullan E et al. Screening, detection and management of depression in elderly primary care attenders. i : The acceptability and performance of the 15 item geriatric depression scale (gds15) and the development of short versions. *Family practice* 1994 ; 11(3) : 260–266. DOI :10.1093/fampra/11.3.260.

[141] Nyunt MSZ, Fones C, Niti M et al. Criterion-based validity and reliability of the geriatric depression screening scale (gds-15) in a large validation sample of community-living asian older adults. *Aging and Mental Health* 2009 ; 13(3) : 376–382. DOI : 10.1080/13607860902861027.

[142] McNair D and Kahn R. Self-assessment of cognitive deficits. *Assessment in geriatric psychopharmacology* 1983 ; : 137–143.

[143] Galasko D, Bennett DA, Sano M et al. Adcs prevention instrument project : assessment of instrumental activities of daily living for community-dwelling elderly individuals in dementia prevention clinical trials. *Alzheimer Disease & Associated Disorders* 2006 ; 20 : S152–S169. DOI :10.1097/01.wad.0000213873.25053.2b.

[144] Doody RS, Gavrilova SI, Sano M et al. Effect of dimebon on cognition, activities of daily living, behaviour, and global function in patients with mild-to-moderate alzheimer's disease : a randomised, double-blind, placebo-controlled study. *The Lancet* 2008 ; 372(9634) : 207–215. DOI :10.1016/S0140-6736(08)61074-0.

[145] Galasko D, Bennett D, Sano M et al. An inventory to assess activities of daily living for clinical trials in alzheimer's disease. *Alzheimer disease and associated disorders* 1997 ; 11 Suppl 2 : S33—-S39. DOI :10.1097/00002093-199700112-00005.

[146] Kurz X, Scuvee-Moreau J, Rive B et al. A new approach to the qualitative evaluation of functional disability in dementia. *International journal of geriatric psychiatry* 2003 ; 18(11) : 1050–1055. DOI :10.1002/gps.1009.

[147] Rockwood K, Song X, MacKnight C et al. A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal (CMAJ)* 2005 ; 173(5) : 489–495. DOI :10.1503/cmaj.050051.

[148] Ensrud KE, Ewing SK, Taylor BC et al. Comparison of 2 frailty indexes for prediction of falls, disability, fractures, and death in older women. *Archives of internal medicine* 2008 ; 168(4) : 382–389. DOI :10.1001/archinternmed.2007.113. URL `https://doi.org/10.1001/archinternmed.2007.113`. `https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/414008/ioi70217_382_389.pdf`.

[149] Fried LP, Tangen CM, Walston J et al. Frailty in older adults : evidence for a phenotype. *The Journals of Gerontology Series A : Biological Sciences and Medical Sciences* 2001 ; 56(3) : M146–M157. DOI :10.1093/gerona/56.3.M146.

[150] Fried LP, Ferrucci L, Darer J et al. Untangling the concepts of disability, frailty, and comorbidity : implications for improved targeting and care. *The Journals of Gerontology Series A : Biological Sciences and Medical Sciences* 2004 ; 59(3) : M255–M263. DOI :10.1093/gerona/59.3.M255.

[151] Robertson DA, Savva GM and Kenny RA. Frailty and cognitive impairment—a review of the evidence and causal mechanisms. *Ageing research reviews* 2013 ; 12(4) : 840–851. DOI :10.1016/j.arr.2013.06.004.

[152] Cesari M, Gambassi G, Abellan van Kan G et al. The frailty phenotype and the frailty index : different instruments for different purposes. *Age and ageing* 2013 ; 43(1) : 10–12. DOI :10.1093/ageing/aft160.

[153] Buchman AS, Boyle PA, Wilson RS et al. Frailty is associated with incident alzheimer's disease and cognitive decline in the elderly. *Psychosomatic medicine* 2007 ; 69(5) : 483–489. DOI :10.1097/psy.0b013e318068de1d.

[154] Schaefer EJ, Bongard V, Beiser AS et al. Plasma phosphatidylcholine docosahexaenoic acid content and risk of dementia and alzheimer disease : the framingham heart study. *Archives of neurology* 2006 ; 63(11) : 1545–1550. DOI :10.1001/archneur.63.11.1545.

[155] Rocca W, Hofman A, Brayne C et al. Frequency and distribution of alzheimer's disease in europe : a collaborative study of 1980–1990 prevalence findings. *Annals of Neurology : Official Journal of the American Neurological Association and the Child Neurology Society* 1991 ; 30(3) : 381–390. DOI :10.1002/ana.410300310.

[156] Letenneur L, Gilleron V, Commenges D et al. Are sex and educational level independent predictors of dementia and alzheimer's disease ? incidence data from the paquid project. *Journal of Neurology, Neurosurgery & Psychiatry* 1999 ; 66(2) : 177–183. DOI :10.1136/jnnp.66.2.177.

[157] Katzman R. Education and the prevalence of dementia and alzheimer's disease. *Neurology* 1993 ; 43 : 13–20. DOI :10.1212/WNL.43.1_Part_1.13.

[158] Stern Y. Cognitive reserve and alzheimer disease. *Alzheimer Disease & Associated Disorders* 2006 ; 20 : S69–S74. DOI :10.1097/00002093-200607001-00010.

[159] Stern Y, Gurland B, Tatemichi TK et al. Influence of education and occupation on the incidence of alzheimer's disease. *JAMA : the journal of the American Medical Association* 1994 ; 271(13) : 1004–1010. DOI :10.1001/jama.1994.03510370056032.

[160] Huang Y and Mucke L. Alzheimer mechanisms and therapeutic strategies. *Cell* 2012 ; 148(6) : 1204–1222. DOI :10.1016/j.cell.2012.02.040.

[161] Karp A, Kåreholt I, Qiu C et al. Relation of education and occupation-based socioeconomic status to incident alzheimer's disease. *American journal of epidemiology* 2004 ; 159(2) : 175–183. DOI :10.1093/aje/kwh018. URL `https://doi.org/10.1093/aje/kwh018`. `https://academic.oup.com/aje/article-pdf/159/2/175/396130/kwh018.pdf`.

[162] Green RC, Schneider LS, Amato DA et al. Effect of tarenflurbil on cognitive decline and activities of daily living in patients with mild alzheimer disease : a randomized controlled trial. *JAMA : the journal of the American Medical Association* 2009 ; 302(23) : 2557–2564. DOI :10.1001/jama.2009.1866.

[163] Liu CC, Kanekiyo T, Xu H et al. Apolipoprotein e and alzheimer disease : risk, mechanisms and therapy. *Nature Reviews Neurology* 2013 ; 9(2) : 106. DOI :10.1038/nrneurol.2012.263.

[164] Harold D, Abraham R, Hollingworth P et al. Genome-wide association study identifies variants at clu and picalm associated with alzheimer's disease. *Nature genetics* 2009 ; 41(10) : 1088–1093. DOI :10.1038/ng.440.

[165] Poirier J, Bertrand P, Kogan S et al. Apolipoprotein e polymorphism and alzheimer's disease. *The Lancet* 1993 ; 342(8873) : 697–699. DOI :10.1016/0140-6736(93)91705-Q.

[166] Mahley RW and Rall Jr SC. Apolipoprotein e : far more than a lipid transport protein. *Annual review of genomics and human genetics* 2000 ; 1(1) : 507–537. DOI : 10.1146/annurev.genom.1.1.507.

[167] Bu G. Apolipoprotein e and its receptors in alzheimer's disease : pathways, pathogenesis and therapy. *Nature Reviews Neuroscience* 2009 ; 10(5) : 333–344. DOI : 10.1038/nrn2620.

[168] Eisenberg DT, Kuzawa CW and Hayes MG. Worldwide allele frequencies of the human apolipoprotein e gene : climate, local adaptations, and evolutionary history. *American Journal of Physical Anthropology* 2010 ; 143(1) : 100–111. DOI :10.1002/ajpa.21298.

[169] Wisniewski T and Frangione B. Apolipoprotein e : a pathological chaperone protein in patients with cerebral and systemic amyloid. *Neuroscience letters* 1992 ; 135(2) : 235–238. DOI :10.1016/0304-3940(92)90444-C.

[170] Poirier J, Delisle MC, Quirion R et al. Apolipoprotein e4 allele as a predictor of cholinergic deficits and treatment outcome in alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* 1995 ; 92(26) : 12260–12264. DOI :10.1073/pnas.92.26.12260.

[171] Farrer LA, Cupples LA, Haines JL et al. Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease : a meta-analysis. *JAMA : the journal of the American Medical Association* 1997 ; 278(16) : 1349–1356. DOI :10.1001/jama.278.16.1349.

[172] Corder EH, Saunders AM, Strittmatter WJ et al. Gene dose of apolipoprotein e type 4 allele and the risk of alzheimer's disease in late onset families. *Science* 1993 ; 261(5123) : 921–923. DOI :10.1126/science.8346443.

[173] Berr C, Dufouil C, Brousseau T et al. Early effect of apoe-e4 allele on cognitive results in a group of highly performing subjects : the eva study. *Neuroscience letters* 1996 ; 218(1) : 9–12. DOI :10.1016/0304-3940(96)13059-7.

[174] Chobanian AV, Bakris GL, Black HR et al. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure : the jnc 7 report. *ACC Current Journal Review* 2003 ; 12(19) : 31–32. DOI :10.1016/S1062-1458(03)00270-8.

[175] Shobab LA, Hsiung GYR and Feldman HH. Cholesterol in alzheimer's disease. *The Lancet Neurology* 2006 ; 4(12) : 841–852. DOI :10.1016/S1474-4422(05)70248-9.

[176] Ledesma MD and Dotti CG. Amyloid excess in alzheimer's disease : What is cholesterol to be blamed for ? *FEBS letters* 2006 ; 580(23) : 5525–5532. DOI : 10.1016/j.febslet.2006.06.038.

[177] Schönknecht P, Lütjohann D, Pantel J et al. Cerebrospinal fluid 24s-hydroxycholesterol is increased in patients with alzheimer's disease compared to healthy controls. *Neuroscience letters* 2002 ; 324(1) : 83–85. DOI :10.1016/S0304-3940(02)00164-7.

[178] Expert N. Panel on detection, evaluation and treatment of high blood cholesterol in adults. executive summary of the third report of the national cholesterol education program (ncep) expert panel on detection, evaluation and treatment of high blood cholesterol in adults (adult treatment panel iii). *JAMA* 2001 ; (285) : 2486–2497.

[179] Solomon A, Kivipelto M, Wolozin B et al. Midlife serum cholesterol and increased risk of alzheimer's and vascular dementia three decades later. *Dementia and geriatric cognitive disorders* 2009 ; 28(1) : 75–80. DOI :10.1159/000231980.

[180] Xu W, Atti A, Gatz M et al. Midlife overweight and obesity increase late-life dementia risk : a population-based twin study. *Neurology* 2011 ; 76(18) : 1568–1574. DOI : 10.1212/WNL.0b013e3182190d09.

[181] Qizilbash N, Gregson J, Johnson ME et al. Bmi and risk of dementia in two million people over two decades : a retrospective cohort study. *The lancet Diabetes & endocrinology* 2015 ; 3(6) : 431–436. DOI :10.1016/S2213-8587(15)00033-9.

[182] Nourhashemi F, Deschamps V, Larrieu S et al. Body mass index and incidence of dementia the paquid study. *Neurology* 2003 ; 60(1) : 117–119. DOI :10.1212/01.WNL.0000038910.46217.AA.

[183] Cronin-Stubbs D, Beckett LA, Scherr PA et al. Weight loss in people with alzheimer's disease : a prospective population based analysis. *Bmj (Clinical research ed)* 1997 ; 314(7075) : 178–179. DOI :10.1136/bmj.314.7075.178.

[184] Gustafson D, Rothenberg E, Blennow K et al. An 18-year follow-up of overweight and risk of alzheimer disease. *Archives of internal medicine* 2003 ; 163(13) : 1524–1528. DOI :10.1001/archinte.163.13.1524.

[185] Garry PJ and Vellas BJ. Practical and validated use of the mini nutritional assessment in geriatric evaluation. *Nutrition in clinical Care* 2002 ; 2(3) : 146–154. DOI :10.1046/j.1523-5408.1999.00108.x.

[186] Kivipelto M, Helkala EL, Hänninen T et al. Midlife vascular risk factors and late-life mild cognitive impairment a population-based study. *Neurology* 2001 ; 56(12) : 1683–1689. DOI :10.1212/WNL.56.12.1683.

[187] Kivipelto M, Ngandu T, Laatikainen T et al. Risk score for the prediction of dementia risk in 20 years among middle aged people : a longitudinal, population-based study. *The Lancet Neurology* 2006 ; 5(9) : 735–741. DOI :10.1016/S1474-4422(06)70537-3.

[188] Lebart L, Morineau A and Piron M. *Statistique exploratoire multidimensionnelle*, volume 3. Dunod Paris, 1995.

[189] Husson F, Lê S and Pagès J. *Analyse de données avec R*. Presses universitaires de Rennes, 2016.

[190] Qannari E, Vigneau E and Courcoux P. Une nouvelle distance entre variables. application en classification. *Revue de statistique appliquée* 1998 ; 46(2) : 21–32. URL `http://www.numdam.org/item/RSA_1998__46_2_21_0`.

[191] Tarpey T. Linear transformations and the k-means clustering algorithm : applications to clustering curves. *The American statistician* 2007 ; 61(1) : 34–40. DOI :10.1198/000313007X171016.

[192] MacQueen J et al. Some methods for classification and analysis of multivariate observations. In Cam LML and Neyman J (eds.) *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1. Oakland, CA, USA, University of California Press, pp. 281–297.

[193] Husson F, Lê S and Pagès J. *Exploratory multivariate analysis by example using R*. Chapman and Hall/CRC, 2011. DOI :10.1111/j.1751-5823.2011.00159_19.x.

[194] Madhulatha TS. An overview on clustering methods. *CoRR* 2012 ; abs/1205.1117. URL `http://arxiv.org/abs/1205.1117. 1205.1117`.

[195] Genolini C and Falissard B. Kml : k-means for longitudinal data. *Computational Statistics* 2010 ; 25(2) : 317–328. DOI :10.1007/s00180-009-0178-4.

[196] James G, Witten D, Hastie T et al. *An Introduction to Statistical Learning : With Applications in R*, volume 112. Springer, 2013. ISBN 9781461471370.

[197] Chouakria AD and Nagabhushan PN. Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification* 2007 ; 1(1) : 5–21. DOI :10.1007/s11634-006-0004-6.

[198] Genolini C and Falissard B. Kml : A package to cluster longitudinal data. *Computer Methods and Programs in Biomedicine* 2011 ; 104(3) : e112–e121. DOI :10.1016/j.cmpb.2011.05.008.

[199] Caliński T and Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 1974 ; 3(1) : 1–27. DOI :10.1080/03610927408827101.

[200] Van Buuren S. *Flexible Imputation of Missing Data, Second Edition*. Chapman and Hall/CRC, 2018. ISBN 9780429492259. DOI :10.1201/9780429492259.

[201] Fréchet MM. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)* 1906 ; 22(1) : 1–72. DOI :10.1007/BF03018603.

[202] Berndt DJ and Clifford J. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10. Seattle, WA, pp. 359–370.

[203] Genolini C, Écochard R and Jacqmin-Gadda H. Copy mean : a new method to impute intermittent missing values in longitudinal studies. *Open Journal of Statistics* 2013 ; 3(04) : 26–40. DOI :10.4236/ojs.2013.34A004.

[204] Genolini C, Lacombe A, Écochard R et al. Copymean : a new method to predict monotone missing values in longitudinal studies. *Computer methods and programs in biomedicine* 2016 ; 132 : 29–44. DOI :10.1016/j.cmpb.2016.04.010.

[205] Wilks DS. Cluster analysis. In *International geophysics*, volume 100. Elsevier, 2011. pp. 603–616. DOI :10.1016/B978-0-12-385022-5.00015-4.

[206] Allam A, Gumpeny RS et al. Analyzing microarray data of alzheimer's using cluster analysis to identify the biomarker genes. *International journal of Alzheimer's disease* 2012 ; 2012 : 649456. DOI :10.1155/2012/649456.

[207] Murtagh F and Legendre P. Ward's hierarchical agglomerative clustering method : which algorithms implement ward's criterion ? *Journal of classification* 2014 ; 31(3) : 274–295. DOI :10.1007/s00357-014-9161-z.

[208] Ward Jr JH. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 1963 ; 58(301) : 236–244. DOI :10.2307/2282967.

[209] Wickham H and Chang W. Devtools : Tools to make developing r packages easier. *R package version* 2016 ; 1(0).

[210] Hahsler M and Hornik K. Dissimilarity plots : A visual exploration tool for partitional clustering. *Journal of Computational and Graphical Statistics* 2011 ; 20(2) : 335–354. DOI :10.1198/jcgs.2010.09139. URL `https://doi.org/10.1198/jcgs.2010.09139`. `https://doi.org/10.1198/jcgs.2010.09139`.

[211] Hahsler M, Buchta C and Hornik K. *seriation : Infrastructure for Ordering Objects Using Seriation*, 2017. URL `https://CRAN.R-project.org/package=seriation`. R package version 1.2-2.

[212] Caraux G and Pinloche S. Permutmatrix : a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 2004 ; 21(7) : 1280–1281. DOI :10.1093/bioinformatics/bti141.

[213] Friendly M. Corrgrams : Exploratory displays for correlation matrices. *The American Statistician* 2002 ; 56(4) : 316–324. DOI :10.1198/000313002533.

[214] Hurley CB. Clustering visualizations of multidimensional data. *Journal of Computational and Graphical Statistics* 2004 ; 13(4) : 788–806. DOI :10.1198/106186004X12425.

[215] Laber EB, Lizotte DJ, Qian M et al. Dynamic treatment regimes : Technical challenges and applications. *Electronic journal of statistics* 2014 ; 8(1) : 1225–1272. DOI :10.1214/14-ejs920.

[216] Goldberg Y and Kosorok M. Q-learning with censored data. *Annals of statistics* 2012 ; 40(1) : 529–560. DOI :10.1214/12-AOS968.

[217] Schulte PJ, Tsiatis AA, Laber EB et al. Q-and a-learning methods for estimating optimal dynamic treatment regimes. *Statistical science : a review journal of the Institute of Mathematical Statistics* 2014 ; 29(4) : 640–661. DOI :10.1214/13-STS450.

[218] Alioum A, Leroy V, Commenges D et al. Effect of gender, age, transmission category, and antiretroviral therapy on the progression of human immunodeficiency virus infection using multistate markov models. *Epidemiology (Cambridge, Mass)* 1998 ; 9 : 605–612. DOI :10.2307/3702781. URL `http://www.jstor.org/stable/3702781`.

[219] Putter H, Fiocco M and Geskus RB. Tutorial in biostatistics : competing risks and multi-state models. *Statistics in medicine* 2007 ; 26(11) : 2389–2430. DOI :10.1002/sim.2712.

[220] Saint-Pierre P, Daures J, Godard P et al. The analysis of asthma control using homogeneous and non-homogeneous markov models. *FAR EAST JOURNAL OF THEORETICAL STATISTICS* 2005 ; 16(1) : 1–18.

[221] Gillaizeau F. *Development and validation of semi-Markovian multistate models for prognosis of patients with chronic diseases.* Theses, Université de Nantes, 2015. URL `https://tel.archives-ouvertes.fr/tel-01301317`.

[222] Jackson CH et al. Multi-state models for panel data : the msm package for r. *Journal of statistical software* 2011 ; 38(8) : 1–29. DOI :10.18637/jss.v038.i08. URL `http://www.jstatsoft.org/v38/i08/`.

[223] Dedieu D, Delpierre C, Gadat S et al. Mixed hidden markov model for heterogeneous longitudinal data with missingness and errors in the outcome variable. *Journal de la Société Française de Statistique* 2014 ; 105(1) : 73–98. URL `https://hal.archives-ouvertes.fr/hal-01864353`.

[224] Haccou P and Meelis E. *Statistical analysis of behavioural data : an approach based on time-structured models.* Oxford university press, 1992. ISBN 9780198546634.

[225] Zucchini W, MacDonald IL and Langrock R. *Hidden Markov models for time series : an introduction using R.* Chapman and Hall/CRC, 2017. ISBN 9781482253832. DOI : 10.1201/b20790.

[226] Le Gall JF. Intégration, probabilités et processus aléatoires. *Ecole Normale Supérieure de Paris* 2006 ; .

[227] Vouma Lekoundji JB. Modèles de markov cachés 2014 ; .

[228] Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C et al. Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research* 2009 ; 18(2) : 195–222. DOI :10.1177/0962280208092301.

[229] Andersen PK and Keiding N. Multi-state models for event history analysis. *Statistical methods in medical research* 2002 ; 11(2) : 91–115. DOI :10.1191/0962280202SM276ra.

[230] Sweeting M, Farewell V and De Angelis D. Multi-state markov models for disease progression in the presence of informative examination times : An application to hepatitis c. *Statistics in medicine* 2010 ; 29(11) : 1161–1174. DOI :10.1002/sim.3812.

[231] Delmas JF and Jourdain B. *Modèles Aléatoires*, volume 57. Springer, 2006. DOI : 10.1007/978-3-540-33284-8.

[232] Cox D and Miller H. *The theory of stochastic processes.* Routledge, 1965. ISBN 0412151707.

[233] Lapuyade-Lahorgue J. *Sur diverses extensions des chaînes de Markov cachées avec application au traitement des signaux radar.* PhD Thesis, 2008.

[234] Dupré X. *Contributions à la reconnaissance de l'écriture cursive à l'aide de modèles de Markov.* PhD Thesis, Thèse, Université René Descartes-Paris 5, 2003.

[235] Visser I. Seven things to remember about hidden markov models : A tutorial on markovian models for time series. *Journal of Mathematical Psychology* 2011 ; 55(6) : 403–415. DOI :10.1016/j.jmp.2011.08.002.

[236] Whoriskey K, Auger-Méthé M, Albertsen CM et al. A hidden markov movement model for rapidly identifying behavioral states from animal tracks. *Ecology and evolution* 2016 ; 7(7) : 2112–2121. DOI :10.1002/ece3.2795.

[237] Baum LE and Petrie T. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics* 1966 ; 37(6) : 1554–1563. DOI : 10.1214/aoms/1177699147.

[238] Sukkar R, Katz E, Zhang Y et al. Disease progression modeling using hidden markov models. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, pp. 2845–2848. DOI :10.1109/EMBC.2012.6346556.

[239] Viterbi A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory* 1967 ; 13(2) : 260–269.

[240] Brouard T, Slimane M, Asselin de Beauville JP et al. Apprentissage d'une chaîne de markov cachée. problèmes numériques liés à l'application à l'image. *Revue de statistique appliquée* 1998 ; 46(2) : 83–108. URL `http://www.numdam.org/item/RSA_1998__46_2_83_0`.

[241] Geoffrey McLachlan DP. *Finite Mixture Models*. 1 ed. Wiley Series in Probability and Statistics, Wiley-Interscience, 2000. ISBN 9780471006268,0471006262. DOI :10.1002/0471721182. URL `http://gen.lib.rus.ec/book/index.php?md5=1a32fda0f86b3542b75ff4a76ca9b600`.

[242] Benaglia T, Chauveau D, Hunter DR et al. mixtools : An R package for analyzing finite mixture models. *Journal of Statistical Software* 2009 ; 32(6) : 1–29. URL `http://www.jstatsoft.org/v32/i06/`.

[243] Patterson TA, Basson M, Bravington MV et al. Classifying movement behaviour in relation to environmental conditions using hidden markov models. *Journal of Animal Ecology* 2009 ; 78(6) : 1113–1123. DOI :10.1111/j.1365-2656.2009.01583.x.

[244] Bode NW and Seitz MJ. Using hidden markov models to characterise intermittent social behaviour in fish shoals. *The Science of Nature* 2018 ; 105(1-2) : 7. DOI : 10.1007/s00114-017-1534-9.

[245] Tong H. Determination of the order of a markov chain by akaike's information criterion. *Journal of applied probability* 1975 ; 12(3) : 488–497. DOI :10.1017/S0021900200048294.

[246] Li C and Biswas G. Applying the hidden markov model methodology for unsupervised learning of temporal data. *International Journal of Knowledge Based Intelligent Engineering Systems - KES Journal* 2002 ; 6(3) : 152–160.

[247] Durbin R, Eddy SR, Krogh A et al. *Biological sequence analysis : probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998. ISBN 9780521629713. DOI :10.1017/CBO9780511790492.

[248] Seltman HJ, Mitchell S and Sweet RA. A bayesian model of psychosis symptom trajectory in alzheimer's disease. *International journal of geriatric psychiatry* 2016 ; 31(2) : 204–210. DOI :10.1002/gps.4326.

[249] Barnard ST, Pothen A and Simon H. A spectral algorithm for envelope reduction of sparse matrices. *Numerical linear algebra with applications* 1995 ; 2(4) : 317–334. DOI :10.1002/nla.1680020402.

[250] Ding C and He X. Linearized cluster assignment via spectral ordering. In *Proceedings of the twenty-first international conference on Machine learning*. ICML '04, ACM, New York, NY, USA : Association for Computing Machinery. ISBN 1581138385, p. 30. DOI :10.1145/1015330.1015407. URL `https://doi.org/10.1145/1015330.1015407`.

[251] Apaydin EA, Maher AR, Shanman R et al. A systematic review of st. john's wort for major depressive disorder. *Systematic reviews* 2016 ; 5(1) : 148. DOI :10.1186/s13643-016-0325-2.

[252] Paraschakis A, Katsanos AH et al. Antidepressants for depression associated with traumatic brain injury : a meta-analytical study of randomised controlled trials. *East Asian archives of psychiatry : official journal of the Hong Kong College of Psychiatrists* 2017 ; 27(4) : 142–149. URL `http://europepmc.org/abstract/MED/29259144`.

[253] Pérez C, Latymer M, Almas M et al. Does duration of neuropathic pain impact the effectiveness of pregabalin ? *Pain practice : the official journal of World Institute of Pain* 2017 ; 17(4) : 470–479. DOI :10.1111/papr.12469.

[254] Katzberg HD, Barnett C, Merkies IS et al. Minimal clinically important difference in myasthenia gravis : outcomes from a randomized trial. *Muscle & nerve* 2014 ; 49(5) : 661–665. DOI :10.1002/mus.23988.

[255] Chambers JM, Freeny AE and Heiberger RM. Analysis of variance ; designed experiments. In *Statistical models in S*. Routledge, 2017. pp. 145–193. DOI : 10.1201/9780203738535-5.

[256] Venables WN and Ripley BD. *Modern Applied Statistics with S*. Fourth ed. Statistics and computing, New York : Springer, 2002. ISBN 0387954570,9780387954578,1441930086,9781441930088,0387217061,9780387217062. DOI :10.1007/978-0-387-21706-2. URL `http://www.stats.ox.ac.uk/pub/MASS4`. ISBN 0-387-95457-0.

[257] Breiman L, Friedman J, Olshen R et al. Classification and regression trees. *Cole Statistics/Probability Series* 1984 ; DOI :10.1002/cyto.990080516.

[258] Breiman L. Random forests. *Machine learning* 2001 ; 45(1) : 5–32. DOI :10.1023/A: 1010933404324.

[259] McClintock BT and Michelot T. momentuhmm : R package for generalized hidden markov models of animal movement. *Methods in Ecology and Evolution* 2018 ; 9(6) : 1518–1530. DOI :10.1111/2041-210X.12995.

[260] Pirotta E, Edwards EW, New L et al. Central place foragers and moving stimuli : A hidden-state model to discriminate the processes affecting movement. *Journal of Animal Ecology* 2018 ; 87(4) : 1116–1125.

[261] Narimatsu H and Kasai H. Duration and interval hidden markov model for sequential data analysis. In *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8. DOI :10.1109/IJCNN.2015.7280808.

[262] Blasiak S and Rangwala H. A hidden markov model variant for sequence classification. In *Twenty-Second International Joint Conference on Artificial Intelligence*. pp. 1192–1197. DOI :10.5591/978-1-57735-516-8/IJCAI11-203.

[263] DeRuiter SL, Langrock R, Skirbutas T et al. A multivariate mixed hidden markov model to analyze blue whale diving behaviour during controlled sound exposures. *arXiv preprint arXiv* 2016 ; 1602. DOI :10.1214/16-AOAS1008. `1602.06570`.

[264] Hubbard R and Zhou X. A comparison of non-homogeneous markov regression models with application to alzheimer's disease progression. *Journal of applied statistics* 2011 ; 38(10) : 2313–2326. DOI :10.1080/02664763.2010.547567.

[265] Chenand B and Zhou XH. Non-homogeneous markov process models with informative observations with an application to alzheimer's disease. *Biometrical Journal* 2011 ; 53(3) : 444–463. DOI :10.1002/bimj.201000122.

[266] Rafi S. *Chaînes de Markov cachées et séparation non supervisée de sources*. PhD Thesis, 2012.

[267] Baum LE, Petrie T, Soules G et al. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics* 1970 ; 41(1) : 164–171. DOI :10.1214/aoms/1177697196. URL `http://www.jstor.org/stable/2239727`.

[268] McLachlan G and Krishnan T. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007. ISBN 9780471201700. DOI :10.1002/9780470191613.

[269] Baum LE. An inequality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities III : Proceedings of the Third Symposium on Inequalities* 1972 ; 3(1) : 1–8.

[270] Krogh A, Larsson B, Von Heijne G et al. Predicting transmembrane protein topology with a hidden markov model : application to complete genomes. *Journal of molecular biology* 2001 ; 305(3) : 567–580. DOI :10.1006/jmbi.2000.4315. URL `https://doi.org/10.1006/jmbi.2000.4315`.

[271] MacDonald IL. Numerical maximisation of likelihood : A neglected alternative to em ? *International Statistical Review* 2014 ; 82(2) : 296–308. DOI :10.1111/insr.12041.

[272] Forney GD. The viterbi algorithm. *Proceedings of the IEEE* 1973 ; 61(3) : 268–278. DOI :10.1109/PROC.1973.9030.

Annexes du manuscrit

Table des matières des Annexes

A.1 Synthèse des variables exploitées dans l'analyse post-hoc 174
A.2 Algorithmes d'estimation des paramètres d'un HMM 175
A.2.1 Apprentissage : Algorithme *forward-backward* 176
A.2.2 Évaluation : Algorithme *forward* 177
A.2.3 Optimisation : Algorithme *backward* 177

A.1 Synthèse des variables exploitées dans l'analyse post-hoc

Le Tableau 8.1 ci-après résume l'ensemble des variables quantitatives présentées dans les sections 1.3.1 à 1.3.3 du Chapitre 3. Le Tableau 8.2 fais de même avec les variables qualitatives.

Table 8.1 – Résumé des différentes variables quantitatives de l'essai MAPT exploitées dans les analyses post-hoc du présent mémoire.

Notation	Étendue	Mesure	Facteur de risque	Unité de mesure
Z-score	-4,61 ; 2,12(1)	Critère de jugement principal mesurant le niveau cognitif global	Déclin	-
Age	70 et plus	Age du sujet	Haut	Année
TMT-A	13 ; 273(1)	Version A du TMT : tâche simple	Haut	Secondes
TMT-B	23 ; 660(1)	Version B du TMT : tâche complexe	Haut	Secondes
COWAT	1 ; 44(1)	Production spontanée de mots dans la catégorie	Faible	Mots acceptés
SPPB	0 ; 12	Performances physiques	Faible	Points
EVA-1	0 ; 100	Intensité des troubles de la mémoire	Haut	mm
EVA-2	0 ; 100	Gêne ressentie dans les activités de la vie quotidienne	Haut	mm
Autonomie	0 ; 45	Capacités fonctionnelles évaluant l'autonomie	Faible	Points
CAIDE	0 ; 15	Estimation du risque de survenue d'une démence de type Alzheimer	Haut	Points

(1) : étendue observée dans MAPT sur l'ensemble du suivi.

Table 8.2 – Résumé des différentes variables qualitatives de l'essai MAPT exploitées dans les analyses post-hoc du présent mémoire.

Notation	Nom long	Mesure	Facteur de risque	Unité de mesure
Bras essai		1 : Ω_3+IM	Tout sauf placebo	-
		2 : Ω_3		
		3 : IM		
		4 : Placebo		
Intervention	Intervention multi-domaine	0 : non	Non	-
		1 : oui		
Etude	Niveau d'études	0 : Bas, $\leq$ CEP	Bas	-
		1 : Haut, après CEP		
Sexe	Sexe du sujet	0 : Femme	Femme	-
		1 : Homme		
CDR	Echelle d'évaluation du statut de démence	0 : Négatif, score = 0	Positif	Points
		1 : Positif, score $\geq$ 0,5		
IMC	Indice de masse corporelle	0 : non obèse, $\leq$ 30	Oui	$kg.m^{-2}$
		1 : obèse, > 30		
GDS	Symptomatologie dépressive	0 : Absence, score < 5	Présence	Points
		1 : Présence, score $\geq$ 5		
Fried	Critère de Fried mesurant le niveau de fragilité global	0 : Non fragile, score = 0	Fragile	Points
		1 : Fragile, score > 0		
Hypertension	Pression artérielle systolique couchée	0 : Non, PAS $\leq$ 140	Oui	mmgH
		1 : Oui, PAS > 140		
Phénotype	Statut phénotypique selon APOE4	0 : sans APOE4	APOE4	-
		1 : avec APOE4		
Cholesterol	Mesure de la cholestérolémie	0 : non	Oui	mg/dL
		1 : oui		
Démence	Diagnostic de démence au cours du suivi	0 : non	Oui	-
		1 : oui		

A.2 Algorithmes d'estimation des paramètres d'un HMM

Comme indiqué en section 2.2.2.2 du Chapitre 5, l'estimation des paramètres d'un HMM résulte de la fonction de vraisemblance [225, 236]. En maximisant cette dernière,

une paramétrisation optimale du processus sous-jacent est obtenue [235]. Considérons un échantillon de T observations $x_1, ..., x_T$ supposé généré par un modèle de Markov caché. La fonction de vraisemblance de cet échantillon correspond à la distribution conjointe de toute les observations, ce qui implique qu'elle s'écrit comme :

$$L_T = \delta P(x_1)\Gamma P(x_2)...\Gamma P(x_T)1' = \alpha_T 1' \quad .$$

À partir de cette vraisemblance, les mises en pratique de cette méthode visent à résoudre trois types de problèmes présentés ci-dessous [20, 225, 237].

A.2.1 Apprentissage : Algorithme *forward-backward*

Une première façon d'optimiser cette vraisemblance dans une approche non supervisée est d'utiliser la méthode itérative *Expectation-Maximisation* (EM) [267] ou l'algorithme de Baum-Welch [20, 237].

En considérant les paramètres précédemment définis, l'objectif de l'algorithme EM est d'approcher la vraisemblance de façon itérative [268]. En effet, cette maximisation est souvent impossible à effectuer analytiquement pour les HMM. Dans cet algorithme, le logarithme de la vraisemblance a posteriori, c'est-à-dire une fois le processus caché estimé, est utilisé [241]. La procédure itérative consiste alors à initialiser les paramètres ; puis, à chaque étape, calculer l'espérance conditionnelle aux observations et estimer les paramètres par maximisation de cette espérance. Ce procédé est répété jusqu'à atteindre un critère d'arrêt choisi, adapté à chaque situation [242].

Ainsi, il en découle l'approche séquentielle de Baum-Welch, consistant à diviser l'émission de l'observation en deux parties : une de l'état initial ($t = 1$) jusqu'au temps t, avec l'état x_t au temps t et une autre de la fin du processus (temps T) jusqu'au temps $t + 1$, sachant qu'au temps t le processus était dans l'état x_t [227, 233]. Dans ce cas, cela revient à calculer la probabilité suivante :

$$P(X) = \sum_{i=1}^{m} P(X1, ...X_t|C_t = i)P(X_{t+1}, ..., X_T|C_t = i) = \sum_{i=1}^{m} \alpha_t(i)\beta_t(i) \quad ,$$

avec $\alpha_t(i)$ la probabilité d'émettre le début de l'observation et β_i celle de la fin d'observation. Le calcul de α s'appuie avec un t croissant (*forward*) tandis que le calcul de β est réalisé avec t décroissant (*backward*) [267, 269]. Cet algorithme converge vers un maximum local de la vraisemblance [270].

L'objectif étant de maximiser cette probabilité, l'algorithme est construit pour estimer les paramètres du modèle petit à petit, hormis le nombre d'états cachés qui reste constant après initialisation [20]. L'algorithme commence par sélectionner un ensemble initial de paramètres. Il calcule par itération les étapes du processus, qu'il répète jusqu'à un critère de convergence. Généralement, ce critère de fin d'algorithme est défini comme la convergence de la log-vraisemblance, c'est-à-dire lorsque la différence de valeur entre deux itérations est de l'ordre de 10^{-6} [246]. À chaque étape de l'apprentissage, la probabilité d'émission des observations doit être améliorée [227]. C'est à ce moment-là que l'algorithme EM est utilisé.

En pratique, étant donné une séquence d'observations, la méthode apprend à ajuster les paramètres du modèle dans le but de maximiser la probabilité décrite ci-dessus. L'objectif final est d'adapter au mieux le HMM au processus observé.

A.2.2 Évaluation : Algorithme *forward*

La deuxième façon d'optimiser la vraisemblance du modèle consiste à utiliser une méthode de maximisation numérique récursive ou algorithme *forward* [235]. L'avantage de cette méthode est d'être facilement applicable avec les logiciels actuels et peu coûteuse en temps de calcul [271].

Au temps t, le processus est dans l'état c_t avec une certaine probabilité déterminée par Γ. Il génère le symbole x_t avec la probabilité $P(x_t)$. Le temps avance du temps t au temps $t+1$. Le processus effectue une transition vers l'état c_{t+1} à partir de l'état c_t avec la probabilité déterminée par Γ, générant à son tour le symbole x_{t+1} avec une probabilité $P(x_{t+1})$). Le processus se répète jusqu'à la fin de la séquence ($t = T$). Ainsi, par itération nous obtenons :

$$\alpha_T = \delta P(x_1)\Gamma P(x_2)...\Gamma P(x_T) = \delta P(x_1) \prod_{s=2}^{T} P(x_s) \quad .$$

Il en découle directement que $\alpha_1 = \delta P(x_1)$ et $\alpha_t = \alpha_{t-1}\Gamma P(x_t)$, $\forall t \geq 2$.

Le nombre d'opérations nécessaires à l'optimisation de la vraisemblance par l'algorithme *forward* est de l'ordre de Tm^2 et les éléments α_t portent de nom de *forward* probabilités [271].

L'inconvénient majeur de cette méthode, par définition, réside dans le fait que α_t soit un produit de probabilités. Cela signifie que sa valeur décroit lorsque t augmente, jusqu'à être arrondie à 0. Une solution à ce problème est de considérer non plus la vraisemblance mais le logarithme de celle-ci [270], comme indiqué ci-dessus. Un autre inconvénient, commun à la méthode précédente, est la présence de plusieurs maxima locaux au lieu d'un maximum global. Modifier les valeurs initiales et relancer l'algorithme aide à savoir si le maximum global est atteint. Les valeurs d'initiation jouant un rôle important, elles doivent être correctement choisies. Pour les paramètres de distribution, une première approche consiste à considérer les valeurs moyennes et de dispersion de la distribution utilisée.

En pratique, cet algorithme calcule la probabilité d'une séquence particulière à partir d'un modèle donné [20]. Ce problème d'évaluation est aussi connu sous le nom de *scoring*. Dans le cas de plusieurs modèles compétitifs, cette démarche conduit à sélectionner le modèle optimal générant la séquence d'observations choisie.

A.2.3 Optimisation : Algorithme *backward*

Le dernier point, toujours connaissant le modèle associé et la séquence d'observation [227], est de trouver la séquence d'états cachés la plus probable ayant conduit à la génération de la séquence d'observation [20]. Cette démarche d'optimisation consiste à décoder les états coïncidant le mieux avec la séquence d'observations étudiée, pour une chaîne de Markov à temps discret en particulier [239]. Portant également le nom de *decoding*, ce problème se résolve en s'appuyant sur l'algorithme de Viterbi ou *backward* [233, 239].

L'objectif est de calculer la séquence S_T^* la plus probable de se produire. Cette séquence, aussi appelée séquence a posteriori, n'est déterminable qu'une fois le modèle entièrement spécifié, d'où son nom. Les probabilités nécessaires à cette procédure découlent de l'algorithme *forward-backward*. L'algorithme *backward* est construit exactement de la même façon que l'algorithme *forward*, sauf que les observations sont considérées à l'inverse du temps, avec un t décroissant [227]. Pour maximiser le critère d'optimisation global, la probabilité d'être dans

un état $c_t = i$ au temps t, sachant les observations de la séquence X et le modèle (noté λ) se définit par :

$$\xi_t(i)P(C_t = i|X, \lambda) \quad .$$

Avec les probabilités de l'algorithme de Baum-Welch, ξ_t peut s'exprimer de la façon suivante [20] :

$$\xi_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{m} \alpha_t(i)\beta_t(i)} \quad ,$$

avec α_t l'observation partielle de la séquence jusqu'au temps t et β_t l'observation partielle du temps $t+1$ à la fin de la séquence. Au final,

$$S_t^* = \underset{1 \leq i \leq m}{\text{argmax}} \xi_t(i) \quad \forall 1 \leq t \leq T \quad .$$

Par ailleurs, plutôt que d'optimiser toutes les séquences d'état possibles dont le nombre augmente lorsque T augmente, une solution consiste à maximiser la séquence temps par temps. Cette démarche porte le nom d'algorithme de Viterbi [272] ou décodage local car les états sont choisis en utilisant la maximisation locale pour chaque t séparément [235]. Dans ce cas, nous devons définir [20] :

$$\nu_t(i) = \max_{x_1, x_2, ..., x_{t-1}} P(X_2, X_2, ..., X_t|C_t = i) \quad .$$

Nous en déduisons :

$$\nu_{t+1}(j) = [\max_i \nu_t(i)p_{ij}]b_j(X_{t+1}) \quad ,$$

avec $b_j(O_{t+1})$ correspondant à la probabilité d'observation selon la distribution de l'état j. L'ensemble des maxima trouvés est regroupé dans un vecteur $\psi_t(i)$. Par conséquent, l'algorithme se transforme en $\nu_1(i) = \delta b_j(X_1)$ et $\psi_1(i) = 0$. Par récursion, les $\nu_t(j)$ et $\psi_t(j) = \underset{1 \leq i \leq m}{argmax}[\nu_{t-1}(i)p_{ij}]$ sont calculés pour chaque temps de la séquence. Finalement, les maxima sont atteints et l'algorithme se termine. La séquence optimale est obtenue par :

$$S_T^* = \underset{1 \leq i \leq m}{argmax} \nu_T(i) \quad .$$

Notons ainsi que la principale différence avec l'algorithme *forward*, outre un temps décroissant, résulte de la maximisation sur les états précédents. Elle remplace la somme de l'algorithme *forward*. Le nombre d'opérations nécessaires à l'optimisation de la vraisemblance par l'algorithme *backward* est de l'ordre de Tm^2 et les éléments β_t sont appelés les *backward* probabilités.

Résumé

Contexte : Les essais cliniques demeurent le gold standard pour évaluer l'efficacité d'un traitement, y compris dans le domaine de la prévention. Un essai non concluant se traduit souvent par l'absence d'effet du traitement, alors qu'une méthodologie non optimale peut être en cause. Dans ce travail, nous abordons les essais de prévention, dont l'objectif est de retarder le déclin des fonctions cognitives car les résultats des grands essais dans ce domaine sont peu convaincants. Nous avons fait l'hypothèse que les spécificités de ces essais (données répétées, éventuel effet d'apprentissage, longue phase de latence avant de pouvoir détecter un effet, incertitude de mesure, population hétérogène. . .) pourraient être mieux appréhendées au moment de l'analyse.

Méthodes : Pour chaque difficulté méthodologique, des méthodes d'analyse spécifiques ont été proposées pour optimiser la détection d'un effet potentiel. Le caractère longitudinal des données est un aspect fondamental que nous prenons en compte dans les analyses, soit en utilisant la trajectoire, soit en utilisant des paramètres représentatifs de son évolution (taux d'accroissement, transition entre états). L'hétérogénéité de la population est explorée par des méthodes de clustering fonctionnel basées sur la forme de la trajectoire ou des méthodes nécessitant un paramètre de synthèse (classification ascendante hiérarchique, sémiologie graphique). L'incertitude de la mesure a été abordée par l'analyse de sujets répondeurs au traitement ou par des modèles de Markov cachés. Une méthode d'apprentissage statistique par renforcement a été utilisée pour traiter l'effet retardé du traitement préventif.

Résultats : Différentes méthodes d'analyses ont été appliquées aux données réelles de l'essai Multidomain Alzheimer Preventive Trial. Les analyses ont indiqué que la population n'évoluait pas comme attendu cliniquement mais avait un niveau cognitif global stable sur les 3 ans de suivi. La composition des profils d'évolution différait selon la méthode utilisée. Les k-means pour données longitudinales avec reconnaissance de la forme de la trajectoire et l'analyse des répondeurs ont mis en évidence un groupe déclinant au cours du suivi. La classification ascendance hiérarchique et la sémiologie graphique ont toutes deux confirmé que la majorité des sujets de l'étude étaient stables cognitivement.

Conclusion : Dans le cadre d'un essai de prévention, les données peuvent être ré-analysées par des méthodes non mobilisées habituellement. Des étapes préalables de prétraitement des données s'avèrent nécessaires pour certaines analyses. Nous avons mobilisé des méthodes qui nous paraissaient appropriées pour mieux prendre en compte certaines limites méthodologiques, le but n'étant pas d'atteindre l'exhaustivité ni de revenir sur la conclusion de l'essai. Au total, nous n'avons pas identifié une méthode à utiliser préférentiellement dans un essai de prévention car chacune des méthodes a pu répondre à une limite mais pas à l'ensemble des limites méthodologiques. D'autres travaux sont nécessaires pour répondre à l'ensemble des particularités méthodologiques.

Mots clés : essai de prévention, cognition, donnée longitudinale, paramètre de synthèse, analyses statistiques.

Abstract

Context : Clinical trials remain the gold standard for evaluating the effectiveness of a treatment, including in the area of prevention. An inconclusive trial often results in no treatment effect, while a non-optimal methodology may be involved. In this work, we discuss prevention trials, whose objective is to delay cognitive decline because the results of large trials in this area are unconvincing. We hypothesized that the specificities of these trials (repeated data, possible learning effect, long latency phase before being able to detect an effect, measurement uncertainty, heterogeneous population...) could be better understood at the time of analysis.

Methods : For each methodological difficulty, specific analysis methods have been proposed to optimize the detection of a potential effect. The longitudinal nature of the data is a fundamental aspect that we take into account in the analyses, either by using the trajectory or by using parameters representative of its evolution (rate of change, transition between states). The heterogeneity of the population is explored by functional clustering methods based on the shape of the trajectory or methods requiring a synthesis parameter (hierarchical ascending classification, graphical semiology). Measurement uncertainty has been addressed by the analysis of responders subjects to the treatment or by hidden Markov models. A statistical reinforcement learning method was used to address the delayed effect of preventive treatment.

Results : Different methods of analysis were applied to the real data from the Multidomain Alzheimer Preventive Trial. Analyses indicated that the population did not evolve as clinically expected but had stable global cognitive levels during the 3-years follow-up. The composition of the evolution profiles differed according to the method used. The k-means for longitudinal data with trajectory shape recognition and responders analysis revealed a declining group during follow-up. Both hierarchical ascendancy classification and graphical semiology confirmed that the majority of subjects in the study were cognitively stable.

Conclusion : In a prevention trial, the data can be re-analysed by methods not usually used. Some analyses may require pre-processing of the data. We have used methods that we felt were appropriate to better take into account certain methodological limitations, the aim being neither to achieve completeness nor to revisit the conclusion of the trial. In total, we did not identify a method to be used preferentially in a prevention trial because each method may have addressed one but not all methodological limitations. Further work is required to address all methodological particularities.

Key words : prevention trial, cognition, longitudinal data, synthesis parameter, statistical analysis.