



**Université Lille 2**  
**Droit et Santé**

UNIVERSITE LILLE 2 DROIT ET SANTE  
**FACULTE DE MEDECINE HENRI WAREMBOURG**

Année : 2015

THESE POUR LE DIPLOME D'ETAT  
DE DOCTEUR EN MEDECINE

**Évaluation de l'apport de deux scores de risque génétique pour la  
prédiction du risque de maladie d'Alzheimer dans l'étude des 3 Cités**

Présentée et soutenue publiquement le 24 mars 2015 à 18h  
au Pôle Formation

**Par Fleur Maury**

---

**JURY**

**Président :**

**Monsieur le Professeur Jean-Louis Salomez**

**Assesseurs :**

**Monsieur le Professeur Philippe Amouyel**

**Monsieur le Professeur Régis Beuscart**

**Directeur de Thèse :**

**Madame le Docteur Florence Richard**

---

## Liste des abréviations

ADGC	Alzheimer's Disease Genetics Consortium
APOE	Apolipoprotéine E
ATCD	Antécédent
AUC	Aire sous la courbe ROC (Area under the ROC curve)
CCPPRB	Comité consultatif pour la protection des personnes participant à la recherche biomédicale
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CHR	Chromosome
CHRU	Centre hospitalier régional universitaire
CV	Cardio-vasculaire
DSM IV	Manuel diagnostique et statistique des troubles mentaux (Diagnostic and Statistical Manual of Mental Disorders) 4ème édition
EADI	European Alzheimer's Disease Initiative
FAR	Fréquence de l'allèle de référence
GERAD	Genetic and Environmental Risk in Alzheimer's Disease (consortium)
GRS	Score de risque génétique (Genetic risk score)
GRS-19	Score de risque génétique comprenant les 19 SNPs avec $p \leq 5 \times 10^{-8}$ dans la méta-analyse IGAP
GRS-32	Score de risque génétique comprenant les 32 SNPs avec $p \leq 5 \times 10^{-6}$ dans la méta-analyse IGAP
GWAS	Etude d'association pangénomique (Genome-Wide Association Study)
HR	Hazard Ratio
IGAP	International Genomics of Alzheimer's Project
IC 95 %	Intervalle de confiance à 95 %
IDI	Integrated discrimination improvement
IMC	Indice de masse corporelle
KM	Kaplan Meier
LCR	Liquide céphalo-rachidien
MA	Maladie d'Alzheimer
MCI	mild cognitive impairment
MMSE	Mini Mental State Examination
NINCDS-ADRDA	National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association
NRI	Net reclassification improvement
OR	Odds Ratio
QI	Qualité d'imputation
ROC	Receiver Operating Characteristic
SNP	Single Nucleotide Polymorphism
TEP scan	Tomographie par Émission de Positrons

# Table des matières

<b>RÉSUMÉ.....</b>	<b>1</b>
<b>INTRODUCTION.....</b>	<b>3</b>
I.Maladie d'Alzheimer.....	3
A.Démences.....	3
B.Diagnostic de la MA.....	4
C.Épidémiologie.....	6
II.Facteurs de susceptibilité génétique de la forme tardive de maladie d'Alzheimer.....	7
A.1993, un allèle majeur identifié : APOE-ε4.....	7
B.2009 à 2011, identification de 9 loci associés avec la MA.....	8
C.2013, doublement du nombre de loci découverts.....	10
III.Prédiction de risque de MA, place de la génétique.....	11
A.Prédiction du risque de MA.....	11
B.Place des facteurs de susceptibilité génétique identifiés par les GWAS dans la prédiction du risque de MA.....	13
C.Objectif de la thèse.....	13
<b>MATÉRIELS ET MÉTHODES.....</b>	<b>15</b>
I.Population d'étude : étude des 3 Cités (3C).....	15
A.Population et suivi.....	15
B.Critère de jugement principal.....	16
C.Variables socio-clinico-biologiques.....	17
D.Données génétiques.....	18
II.Sélection des participants pour l'analyse.....	18
III.Calcul des scores de risque génétique.....	19
A.Deux GRSs : 19 et 32 SNPs.....	19
B.Méthode de calcul.....	20
IV.Analyses statistiques.....	20
A.Analyses descriptives de la population.....	20
B.Association des SNPs avec la MA.....	20
C.Association des GRSs avec la MA.....	21
D.Mesures d'amélioration de la prédiction par les GRSs.....	21
E.Analyses complémentaires.....	25
F.Logiciel utilisé.....	25
<b>RÉSULTATS.....</b>	<b>27</b>
I.Analyses descriptives.....	27
A.Population d'étude.....	27
B.Risque observé de MA.....	28

C.Score de risque génétique.....	29
II.Mesures d'association.....	29
A.Analyse individuelle des SNPs.....	29
B.GRSs.....	31
C.APOE-ε4.....	31
III.Interaction APOE-ε4 x GRS.....	31
IV.Mesures d'amélioration de la prédiction.....	32
A.Apport des GRSs.....	32
B.Apport d'APOE-ε4.....	33
C.Modèle de base comprenant les facteurs de risque cardio-vasculaire.....	34
<b>DISCUSSION.....</b>	<b>37</b>
I.Résultats principaux.....	37
II.Comparaison à la littérature.....	37
III.Apport des 13 SNPs suggestifs.....	38
IV.Confirmation du rôle d'APOE-ε4.....	38
V.Choix de la méthode de calcul des GRSs.....	39
VI.Utilité du cNRI et de l'IDI pour mesurer les performances d'un nouveau prédicteur....	40
VII.Forces et faiblesses de l'étude.....	41
VIII.Perspectives.....	42
<b>CONCLUSION.....</b>	<b>44</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES.....</b>	<b>45</b>
<b>ANNEXES.....</b>	<b>49</b>
Annexe 1. Critères diagnostiques de démence selon le DSM-IV.....	49
Annexe 2. Critères diagnostiques de la maladie d'Alzheimer NINCDS-ADRDA.....	50
Annexe 3. Loci identifiés par la méta-analyse IGAP.....	51
Annexe 4. Synthèse bibliographique.....	52
Annexe 5. Caractéristiques à l'inclusion des 9 294 participants de l'étude 3C.....	53
Annexe 6. Calibration des modèles.....	54
Annexe 7. Amélioration de la prédiction induite par l'ajout des GRSs à un modèle de base incluant les facteurs de risque cardio-vasculaire.....	55
Annexe 8. Méthodes de pondération d'un GRS utilisant la dose allélique.....	57
Annexe 9. Valeurs de référence pour l'interprétation des mesures d'amélioration de la prédiction.....	58
Annexe 10. Comparaison entre les participants inclus et exclus de l'analyse.....	59

## RÉSUMÉ

**Contexte** : Pour lutter contre le fardeau croissant que représente la maladie d'Alzheimer (MA) pour la société, une stratégie consiste à agir tôt, en amont des manifestations cliniques, et à construire des outils simples et performants pour prédire le risque de développer la maladie. L'identification récente de plusieurs nouveaux facteurs de susceptibilité génétique, en plus du gène de l'apolipoprotéine E (APOE), représente une piste pour le développement de ces outils. L'objectif de ce travail était donc d'évaluer l'amélioration de la prédiction du risque de MA lors de l'ajout de ces facteurs de susceptibilité génétique à un modèle comprenant les principaux facteurs de risque de MA connus.

**Méthode** : L'analyse a été réalisée sur 5 922 participants âgés de 65 ans et plus de l'étude des 3 Cités. A partir des résultats de la dernière méta-analyse d'étude d'association pangénomique (GWAS), un score de risque génétique (GRS) composé des 19 loci associés à la MA (GRS-19), et un GRS comprenant également 13 loci supplémentaires (GRS-32) ont été calculés. L'association avec le risque de MA a été estimée par l'Hazard Ratio (HR) issu d'un modèle de Cox. L'amélioration de la prédiction induite par l'ajout d'un GRS à un modèle prédictif comprenant l'âge, le sexe, le niveau d'éducation, l'allèle APOE- $\epsilon$ 4, le centre et la langue maternelle, a été évaluée à l'aide de la différence de C-index ( $\Delta$ C-index), du cNRI (continuous net reclassification improvement) et de l'IDI (integrated discrimination improvement).

**Résultats** : Après un suivi médian de 7,6 ans, 459 participants ont développé une MA. Les deux GRSs étaient significativement associés à la MA ( $HR_{GRS-19}$  [intervalle de confiance à 95%] = 1,05 [1,01 ; 1,09] et  $HR_{GRS-32}$  = 1,06 [1,03 ; 1,09], pour un allèle à risque supplémentaire). Cette association se traduisait par une amélioration faible et non significative de la prédiction du risque de développer une MA :  $\Delta C$ -index = 0,00263 [-0,00117 ; 0,00642], cNRI = 0,0941 [-0,00176 ; 0,216] et IDI = 0,00186 [-0,000274 ; 0,00686]. Les résultats semblaient néanmoins meilleurs avec le GRS-32.

**Conclusion** : Si les facteurs de susceptibilité génétique récemment identifiés par les GWAS ont ouvert de nouveaux axes de recherche pour l'exploration des mécanismes impliqués dans la MA, leur apport conjoint pour la prédiction de risque est encore insuffisant pour pouvoir les intégrer au sein d'une équation de risque.

## INTRODUCTION

### I. Maladie d'Alzheimer

Les démences représentent un problème majeur de santé publique de par leur impact sur le patient et leur entourage, leur retentissement économique sur la société, et leur prévalence croissante. La maladie d'Alzheimer (MA), forme la plus fréquente de démence, fut identifiée en 1907 par le psychiatre allemand éponyme Aloïs Alzheimer qui en décrivit les lésions anatomo-pathologiques caractéristiques.

#### A. Démences

##### 1. Définition

L'Organisation Mondiale de la Santé décrit la démence comme « un syndrome, généralement chronique ou évolutif, dans lequel on observe une altération de la fonction cognitive (capacité d'effectuer des opérations de pensée), plus importante que celle que l'on pourrait attendre du vieillissement normal. Elle affecte la mémoire, le raisonnement, l'orientation, la compréhension, le calcul, la capacité d'apprentissage, le langage et le jugement. La conscience n'est pas touchée. Une détérioration du contrôle émotionnel, du comportement social ou de la motivation accompagne souvent, et parfois précède, les troubles des fonctions cognitives. La démence est causée par un ensemble de maladies et de traumatismes qui affectent principalement ou dans une moindre mesure le cerveau » [2,3]. Les critères diagnostiques généralement utilisés sont ceux du DSM IV (manuel diagnostique et statistique des troubles mentaux 4ème édition) (**Annexe 1**).

## 2. Types de démences

Plusieurs types de démence sont identifiables selon la symptomatologie ou le processus physiopathologique (**Tableau 1**) : la maladie d'Alzheimer est la plus fréquente (50-75%), suivie par la démence vasculaire, la démence à corps de Lewy et la démence fronto-temporale.

**Tableau 1. Types de démences [1].**

Type de démence	Symptômes caractéristiques	Neuropathologie	Proportion
<b>Maladie d'Alzheimer (MA)</b>	Troubles de la mémoire Apathie et dépression Survenue progressive	Plaques amyloïdes Dégénérescences neurofibrillaires	<b>50-75%</b>
<b>Démence vasculaire</b>	Similaire à la MA Mémoire moins altérée Fluctuation de l'humeur Fragilité physique Survenue par à-coup	Maladie cérébro-vasculaire Infarctus isolés ou plus diffus	<b>20-30%</b>
<b>Démence à corps de Lewy</b>	Fluctuations importantes des fonctions cognitives Hallucinations visuelles Syndrome parkinsonien	Corps de Lewy corticaux (alpha-synucléine)	<b>&lt; 5%</b>
<b>Démence fronto-temporale</b>	Modifications de la personnalité, de l'humeur Désinhibition Troubles du langage	Pathologie limitée aux lobes frontaux et temporaux	<b>5-10%</b>

## B. Diagnostic de la MA

### 1. Présentation clinique

La MA est une maladie liée à l'âge. Son diagnostic repose principalement sur la clinique : l'histoire de la maladie (son évolution), l'interrogatoire de l'entourage, l'examen physique et l'exploration neuropsychologique et comportementale. Il est posé devant un tableau clinique de démence présentant les caractéristiques suivantes :

- Une symptomatologie d'apparition progressive (mois, année)



- Une atteinte cognitive dominée initialement par des troubles de la mémoire portant d'abord sur les faits récents puis anciens, à laquelle s'associent progressivement des troubles phasiques, praxiques ou gnosiques.

## 2. Critères diagnostiques

Plusieurs critères ont été proposés pour aider et standardiser le diagnostic de MA. Parmi ceux-ci, les critères du National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) sont les plus utilisés (**Annexe 2**) [6]. Ils permettent de porter un diagnostic de « MA possible » ou « MA probable ». Le diagnostic de certitude nécessite l'examen anatomo-pathologique post-mortem du cerveau, et la visualisation dans les régions hippocampiques et corticales des deux lésions caractéristiques de la MA : plaques amyloïdes (dépôts extra-cellulaires de substance amyloïde) et dégénérescences neurofibrillaires. Ces critères ont été récemment mis à jour, pour tenir compte de l'évolution des connaissances sur les autres causes neurodégénératives de démence, l'existence de formes atypiques de maladie d'Alzheimer et l'utilisation potentielle de biomarqueurs [7]. Ces derniers sont de deux types :

- les marqueurs amyloïdes quantifiant les dépôts amyloïdes dans le cerveau (Pittsburgh compound B (PiB) TEP scan) et la concentration de peptides A $\beta$ 42 (composant principal des plaques amyloïdes) dans le liquide céphalo-rachidien (LCR), et
- les marqueurs de la mort neuronale : 18-F FDG TEP scan, IRM cérébrale et dosage dans le LCR des protéines tau et tau phosphorylée.

Ces critères ne sont toutefois pas utilisés de façon systématique dans le diagnostic de routine.

## C. Épidémiologie

### 1. Prévalence et incidence

En 2007, 850 000 personnes étaient atteintes de MA en France, soit environ 6 % des personnes âgées de plus de 65 ans [4]. Au niveau mondial, le rapport 2013 d'*Alzheimer's disease international* [5] estimait que 44 millions de personnes souffraient de démence en 2013, avec des prévisions dépassant les 130 millions de personnes en 2050. Ce rapport précise en outre que deux tiers d'entre eux vivent dans les pays pauvres ou en voie de développement. Le vieillissement de la population attendu dans ces pays fait de la MA un problème mondial de santé publique.

### 2. Formes de MA

La MA possède un important déterminisme génétique. Dans 1 à 2% des cas, la MA se transmet selon un mode autosomique dominant. Ces formes de MA présentent généralement un début précoce (forme précoce) et sont causées par des mutations rares, à pénétrance complète. Pour 40% de ces formes familiales, des mutations ont été identifiées dans trois gènes : APP, PSEN1 et PSEN2. L'implication de ces gènes dans le métabolisme et l'agrégation du peptide amyloïde  $\beta$  a placé ce dernier au centre des recherches sur le processus physiopathologique de la MA [8].

Dans 98% des cas, la MA ne se transmet pas selon les lois de Mendel, mais plusieurs éléments suggèrent néanmoins une forte composante génétique : un risque deux fois plus important chez les apparentés au premier degré de cas de MA [9], et une forte héritabilité, estimée entre 60 et 80 % par des études sur des paires de jumeaux [10]. Contrairement à la forme précoce, son mode de transmission est plus complexe et relèverait de l'interaction de nombreux variants génétiques relativement fréquents, à faible pénétrance, avec des facteurs environnementaux. Ces formes de MA sont dites multifactorielles et surviennent généralement après 65

ans (forme tardive).

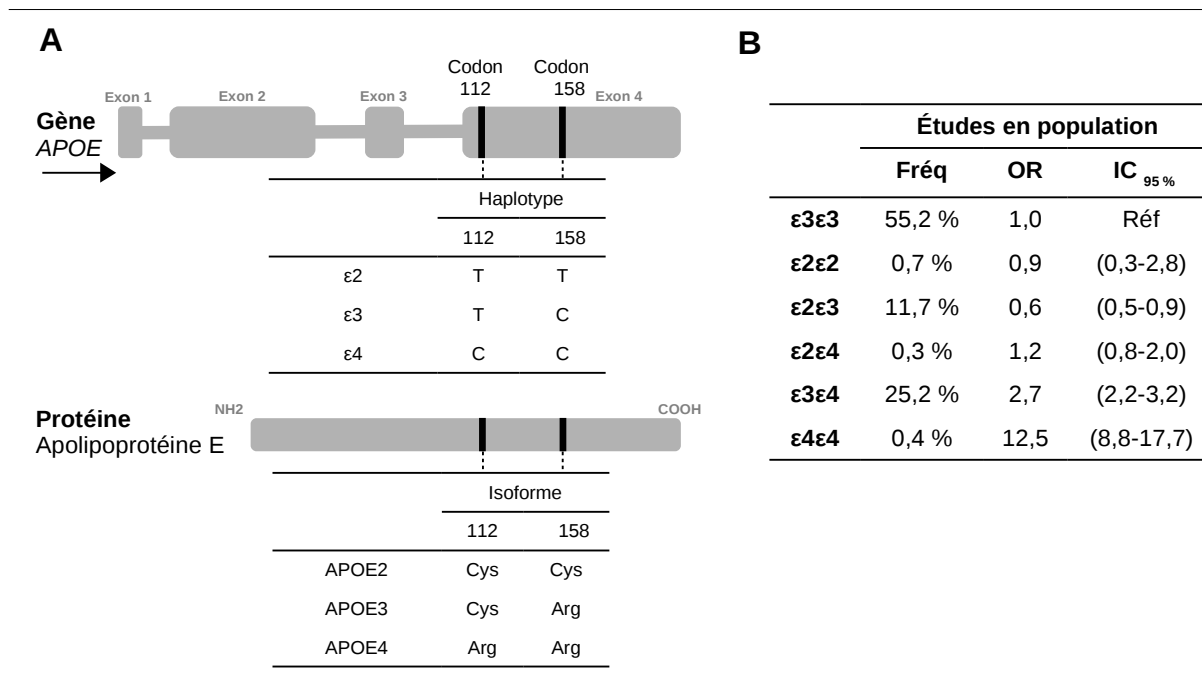
Le travail réalisé dans cette thèse portait uniquement sur la forme tardive pour laquelle l'identification des facteurs génétiques est un enjeu important, tant pour la compréhension des processus physiopathologiques, que pour l'identification de biomarqueurs ou la mise en place de traitement. Pour la suite de la thèse, « maladie d'Alzheimer » désignera « forme tardive de maladie d'Alzheimer ».

## **II. Facteurs de susceptibilité génétique de la forme tardive de maladie d'Alzheimer**

Puisqu'il existe un consensus sur la présence d'une forte composante génétique dans la MA, des efforts considérables ont été déployés au cours des vingt dernières années pour en caractériser ses déterminants. Si l'allèle APOE- $\epsilon$ 4 fut rapidement incriminé, il a fallu attendre l'émergence des technologies d'analyse à haut débit du génome et la constitution de grandes cohortes pour compléter, en partie, le tableau génétique de la MA.

### **A. 1993, un allèle majeur identifié : APOE- $\epsilon$ 4**

Le gène *APOE* fut le premier gène mis en cause dans la forme tardive de la MA. Situé sur le chromosome 19, il code pour l'apolipoprotéine E, un transporteur de lipide essentiel à l'entretien et au remodelage des membranes neuronales. Deux variations génétiques fréquentes, responsables de modifications de la chaîne protéique, définissent 3 haplotypes :  $\epsilon$ 2,  $\epsilon$ 3 et  $\epsilon$ 4. En découlent 3 isoformes d'apolipoprotéine E, dont les différences de configuration spatiale influencent son activité : APOE2, APOE3 et APOE4 [20] (**Figure 1**).



**Figure 1. Gène APOE.**

La représentation A est inspirée de la figure 1 de [11] ; le tableau B est issu des résultats de [12] sur des populations de descendance européenne.

En 1993, l'allèle APOE-ε4 fut identifié comme facteur de risque de MA [13,14]. Par rapport aux individus ε3/ε3, le phénotype de plus fréquent dans les populations de descendance européenne, les porteurs d'un allèle ε4 présentent un risque de MA multiplié par 2 à 3, et les porteurs de deux allèles ε4 ont un risque multiplié par plus de 12 [12,15]. La présence d'ε4 n'est toutefois ni une condition suffisante ni une condition nécessaire au développement de la maladie [16]. Plus marginalement, l'allèle ε2, le moins fréquent, semblerait quant à lui avoir un effet protecteur [17,18]. Compte tenu de l'impact important de l'allèle ε4 sur le risque de développer la maladie, APOE est considéré comme le gène majeur de la MA [19].

### B. 2009 à 2011, identification de 9 loci associés avec la MA

Après la découverte du gène APOE, de nombreuses études d'association, fondées sur une approche gène candidat (étude d'un nombre réduit de gènes ciblés

par des hypothèses physiopathologiques) ont été menées, à la recherche de nouveaux facteurs de risque génétique. Malheureusement, en dehors de la confirmation du rôle de l'allèle APOE- $\epsilon$ 4, les résultats n'étaient pas répliqués (faux positifs et faux négatifs). La principale raison de cette inconsistance était l'utilisation d'échantillons de taille insuffisante pour mettre en évidence de faibles effets (puissance statistique insuffisante).

En 2005, les premières études d'association pangénomique (GWAS pour genome-wide association study) ont pu être réalisées dans la MA grâce aux avancées technologiques. Les GWAS présentent l'avantage d'analyser simultanément de très nombreuses variations génétiques réparties sur l'ensemble du génome, sans hypothèse a priori. Pour la grande majorité d'entre elles, des variations d'une seule paire de bases, appelées SNPs (Single-Nucleotide Polymorphism) sont utilisées. L'étude de l'association de ces SNPs avec la maladie permet de cibler des régions chromosomiques (loci). Malgré cette approche très large du génome, les premières GWAS sur la MA ont été décevantes. Elles n'ont pas permis l'identification robuste de nouveaux facteurs génétiques, les tailles d'échantillons restant probablement insuffisantes [16]. Pendant plus d'une décennie, APOE- $\epsilon$ 4 est donc resté le seul facteur de susceptibilité génétique incontestable de la MA.

Entre 2009 et 2011, la création de quatre grands consortia européens et américains ont permis de mettre en évidence 9 loci à proximité des gènes CLU, PICALM, BIN1, CR1, MS4A6A, ABCA7, CD2AP, EPHA1 et CD33 (**Tableau 2**) [21-25]. Contrairement à APOE- $\epsilon$ 4, les associations retrouvées étaient faibles (OR compris entre 0.85 et 1.23), expliquant sans doute la difficulté de leur identification dans de petites populations.

**Tableau 2. Principales GWAS réalisées sur des populations de descendance européenne.**

Étude	Année	Consortium	Taille d'échantillon		Gènes identifiés en dehors de la région d'APOE
			Découverte	Réplication	
[21]	2009	EADI (European Alzheimer's Disease Initiative)	2 032 cas de MA 5 328 contrôles	3 978 cas de MA 3 297 contrôles	CLU, CR1
[22]	2009	GERAD (Genetic and Environmental Risk in Alzheimer's Disease)	3 941 cas de MA 7 848 contrôles	2 023 cas de MA 2 340 contrôles	CLU, PICALM
[23]	2010	CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology)	3 006 cas de MA 4 642 contrôles	6 505 cas de MA 13 532 contrôles	CLU, PICALM, BIN1
[24]	2011	GERAD+	8 309 cas de MA 7 366 contrôles	4 896 cas de MA 4 903 contrôles 8 286 cas de MA 21 258 contrôles	CLU, PICALM, CR1, BIN1, MS4A6A/MS4A4E, ABCA7 CD2AP, EPHA1, CD33
[25]	2011	ADGC (Alzheimer's Disease Genetics Consortium)	6 688 cas de MA 13 685 contrôles	3 531 cas de MA 3 565 contrôles	CLU, PICALM, CR1, BIN1, MS4A4A
[26]	2013	IGAP (International Genomics of Alzheimer's Project)	17 008 cas de MA 37 646 contrôles	8 572 cas de MA 11 312 contrôles	CLU, PICALM, BIN1, CR1, MS4A6A, ABCA7, CD2AP, EPHA1, INPP5D, MEF2C, HLA-DRB5/HLA-DRB1, ZCWPW1, NME8, PTK2B, CELF1, SORL1, SLC24A4/RIN3, FERMT2, CASS4

### C. 2013, doublement du nombre de loci découverts

Enfin, en 2013, la constitution du consortium IGAP (International Genomics of Alzheimer's Project) a permis de mettre en place la plus grande GWAS sur la MA à ce jour [26]. En combinant les populations des quatre consortia précédents, soit un total de 74046 individus, cette méga-méta-analyse identifia 19 SNPs significativement associés à la MA avec une valeur de  $p < 5 \times 10^{-8}$  (**Tableau 2**) : 8 était déjà connus (CD33 n'étant pas répliqué) et 11 étaient nouveaux. De plus, cette étude rapportait 13 loci prometteurs, appelés suggestifs, dont la valeur de  $p$  était comprise entre  $5 \times 10^{-8}$  et  $1 \times 10^{-6}$  (**Annexe 3**).

Ces facteurs de susceptibilité génétique ont ouvert de nouveaux axes de recherche sur les processus physiopathologiques de la MA et l'élaboration de nouvelles thérapeutiques. Ils pourraient par ailleurs être utilisés pour prédire le risque de développer une MA et mettre en place des mesures de prévention ciblées.

### **III. Prédiction de risque de MA, place de la génétique**

#### **A. Prédiction du risque de MA**

##### **1. Utilité de la prédiction du risque**

Retarder l'apparition de la MA d'un an, réduirait de 11 % le nombre de cas de MA chez les individus de plus de 60 ans [27]. Or, environ un tiers des cas de MA dans le monde serait attribuable à l'association de sept facteurs de risque modifiables qui sont, par ordre d'importance, le niveau d'éducation, le tabagisme, l'activité physique, la dépression, l'hypertension, le diabète et l'obésité [30]. Sachant qu'aujourd'hui il n'existe pas de traitement curatif de la MA, la prévention primaire, dont l'objectif est de prévenir la survenue de la maladie, représenterait donc un levier important pour réduire sa prévalence. En outre, puisque les facteurs de risque cardio-vasculaire semblent impliqués dans la survenue de la MA [34], les mêmes mesures de prévention seraient, au moins en partie, utiles pour la prévention du risque de MA et d'événement cardio-vasculaire.

Dans ce contexte, une prédiction fiable du risque permettrait la mise en place de mesures préventives ciblées. Elle participerait à la constitution de groupes à haut risque de développer la maladie, utilisable en recherche clinique, en vue d'identifier des traitements précoces retardant l'évolution de la maladie. Dans cette optique, les équations de risque, ou score de risque sont fréquemment utilisés.

## 2. Equations de risque

Les équations de risque visent à définir une probabilité de survenue d'une maladie pour un individu, à partir de ses facteurs de risque, qu'ils soient constitutionnels (âge, sexe ou hérédité) ou modifiables. Autrement dit, un nombre réduit de facteurs de risque connus, et facilement mesurables, sont pondérés et agrégés au sein d'une équation pour définir un risque global. Plusieurs niveaux de risque, impliquant une prise en charge préventive ou thérapeutique différente, peuvent ensuite être établis. Cet outil permet d'une part au clinicien de disposer d'une aide à la décision pour la prise en charge de ses patients, et d'autre part, il permet à la population générale de bénéficier d'une information simple et facilement compréhensible sur ces facteurs de risque. Le score de Framingham, très largement utilisé en clinique, [28,29] est l'un des scores de risque les plus connus. Il définit 4 classes de risque d'événement cardio-vasculaire dans les 10 ans, en fonction de l'âge, du sexe, du cholestérol, du tabagisme, de la présence d'un diabète et de la pression artérielle.

Basé sur ce principe, plusieurs scores de risque pour la démence furent proposés [31,32]. Pour l'évaluation du risque de MA en particulier, Reitz et al [33] construisirent en 2010 un score incluant l'âge, le sexe, l'allèle APOE- $\epsilon$ 4, le niveau d'éducation ainsi que d'autres facteurs de risque cardio-vasculaire tels que le diabète, l'hypertension, le tabagisme et le cholestérol. Cet outil reste cependant à valider.

Que ce soit pour la démence ou la MA, les scores de risque proposés incluaient l'allèle APOE- $\epsilon$ 4, reconnu pour être un facteur prédictif important [19,35]. En revanche la place des autres facteurs de susceptibilité génétique au sein de ces scores reste à définir.



## **B. Place des facteurs de susceptibilité génétique identifiés par les GWAS dans la prédiction du risque de MA**

### **1. Constitution de scores de risque génétique**

Les facteurs de susceptibilité récemment identifiés dans la MA sont nombreux et leur taille d'effet est modeste. Une stratégie généralement employée pour évaluer globalement leur effet consiste donc à les agréger au sein d'un score de risque génétique (GRS). Une fois résumée par une seule variable, la valeur ajoutée conjointe de ces facteurs de susceptibilité génétique par rapport aux autres facteurs de risque (âge, niveau d'éducation, APOE- $\epsilon$ 4, etc) peut être évaluée plus facilement. En dehors de la MA, cette approche a été utilisée dans d'autres maladies complexes comme le diabète [36], la sclérose en plaque [37] ou les maladies cardio-vasculaires [38,39].

### **2. Les GRSs pour la MA dans la littérature**

Trois articles ont déjà évalué la valeur ajoutée d'un GRS pour la prédiction du risque de MA [40-42]. Ces derniers ne mettaient pas en évidence d'apport prédictif significatif des GRSs (**Annexe 4**). Toutefois, Les GRSs utilisés ne prenaient pas en compte les nouveaux variants identifiés par la méta-analyse IGAP. Il était donc intéressant de réévaluer ces GRSs, actualisés avec les nouveaux loci découverts.

## **C. Objectif de la thèse**

L'objectif principal de la thèse était donc d'évaluer les performances de 2 scores de risque génétique basés sur le jeu le plus récent de variants génétiques associés à la MA (méta-analyse du consortium IGAP [26]). Les performances ont été évaluées d'une part, par la mesure de l'association des GRSs avec la MA, d'autre part par l'amélioration de la prédiction du risque de MA par rapport aux principaux facteurs de risque connus, dont APOE- $\epsilon$ 4.

En objectif secondaire, pour conforter le rôle prédictif de l'allèle APOE- $\epsilon$ 4, et à

titre de comparaison avec les GRSs, l'amélioration de la prédiction liée à l'allèle APOE- $\epsilon$ 4 a été estimée par rapport aux principaux facteurs de risque connus (hors GRS).

## MATÉRIELS ET MÉTHODES

### I. Population d'étude : étude des 3 Cités (3C)

Ce travail a été réalisé à partir des données d'une grande cohorte prospective Française : l'étude des 3 Cités (3C) [43]. Construite pour évaluer les facteurs de risque vasculaire intervenant dans le développement des accidents vasculaires cérébraux et des démences, l'étude 3C a permis le recueil de données (socio-démographique, clinique, psychologique, biologique, d'imagerie, etc) sur plus de neuf mille personnes issues de 3 grandes villes Françaises.

#### A. Population et suivi

L'échantillon initial de l'étude a été constitué entre mars 1999 et mars 2001 par tirage au sort sur les listes électorales des villes de Bordeaux, Montpellier et Dijon. Les critères d'inclusions étaient les suivants :

- Avoir un âge supérieur ou égal à 65 ans,
- Être non institutionnalisé,
- Être inscrit sur les listes électorales de l'une des 3 communes.
- Avoir donné son consentement pour faire partie de l'étude (un formulaire de consentement écrit a été recueilli).

Neuf mille deux cent quatre-vingt quatorze individus ont été inclus. Les données ont été recueillies via un questionnaire standardisé couplé à une entrevue avec un neuropsychologue ou une infirmière entraînée. Les caractéristiques socio-démographiques des participants de l'étude à leur inclusion sont présentées en **Annexe 5**. Si le patient y consentait, un échantillon de sang était prélevé. Le recueil

prospectif des données a été réalisé à 2, 4, 7 et 10 ans de suivi.

Les investigations cliniques ont été conduites en suivant les principes énoncés dans la déclaration d'Helsinki. Le protocole de 3C a été validé par le comité consultatif pour la protection des personnes participant à la recherche biomédicale (CCPPRB) du CHRU du Kremlin-Bicêtre.

## **B. Critère de jugement principal**

À chaque suivi, les démences incidentes étaient activement recherchées via une procédure en 3 étapes :

1. La réalisation d'une batterie de tests neuropsychologiques ;
2. Un examen clinique par un neurologue. Tous les participants de Bordeaux et Montpellier ont pu en bénéficier. Dans le centre de Dijon, seul ceux potentiellement déments sur la base des tests neuropsychologiques (Mini Mental State Examination (MMSE) et test d'Isaac) ont été examinés par le neurologue ;
3. Une revue des cas par un comité de neurologues indépendants, pour obtenir un consensus sur le diagnostic et sur l'étiologie de la démence, selon les critères du DSM IV et du NINCDS-ADRDA.

Lorsqu'une démence était diagnostiquée, la date de survenue retenue était la date intermédiaire entre le suivi où le diagnostic était posé et le suivi précédent.

Le critère de jugement principal était la survenue d'une MA chez les participants initialement non déments. Les MA « possibles », les MA « probables », et les démences dites « mixtes » (démences associant une composante cérébrovasculaire et une composante neurodégénérative de type Alzheimer) ont été considérées comme des MA incidentes.

### C. Variables socio-clinico-biologiques

Les variables socio-clinico-biologiques classiquement décrites comme facteurs de risque ou comme facteurs de confusion ont été retenues pour l'analyse. Elles sont détaillées dans le **Tableau 3**. Dépression et IMC (indice de masse corporelle) n'ont pas été utilisées en raison du grand nombre de données manquantes (>10%).

**Tableau 3. Descriptif des variables socio-clinico-biologiques à l'inclusion retenues pour l'analyse.**

Variable	Type	Description
<b>Age</b>	Continue	Âge en année
<b>Sexe</b>	Catégorielle	2 classes : <ul style="list-style-type: none"> <li>• Homme</li> <li>• Femme</li> </ul>
<b>Centre</b>	Catégorielle	3 classes : <ul style="list-style-type: none"> <li>• Dijon</li> <li>• Bordeaux</li> <li>• Montpellier</li> </ul>
<b>Niveau d'éducation</b>	Catégorielle	4 classes : <ul style="list-style-type: none"> <li>• Faible : aucune étude ou certificat d'étude primaire</li> <li>• Moyen-faible : niveau BEPC</li> <li>• Moyen élevé : niveau baccalauréat ou équivalent, CAP ou BEP</li> <li>• Élevé : études supérieures</li> </ul>
<b>Langue maternelle</b>	Catégorielle	2 classes : <ul style="list-style-type: none"> <li>• Français</li> <li>• Autre</li> </ul>
<b>Fumeur</b>	Binaire	Présence d'un tabagisme actif
<b>Hypertension</b>	Binaire	Présence d'une hypertension (pression artérielle systolique $\geq$ 140 mmHg ou pression artérielle diastolique $\geq$ 90 mmHg), ou prise de traitement anti-hypertenseur
<b>Hypercholestérolémie</b>	Binaire	Présence d'une hypercholestérolémie ( $\geq$ 6,20 mmol/L) ou prise d'un traitement hypocholestérolémiant
<b>Diabète</b>	Binaire	Présence d'une hyperglycémie ( $\geq$ 7,0 mmol/L) ou prise de traitement anti-diabétique
<b>ATCD cardio-vasculaire</b>	Binaire	Présence d'un antécédent cardio-vasculaire (infarctus du myocarde, accident vasculaire cérébral, angor ou artérite des membres inférieurs)

## D. Données génétiques

Les échantillons d'ADN disponibles de 3C ont été transférés au Centre National de Génotypage (CNG ; [www.cng.fr](http://www.cng.fr)) pour y être génotypés. Les échantillons de qualité suffisante ont été génotypés grâce à une puce Illumina Human610-Quad [21]. Une série d'imputations génétiques basées sur le panel de référence 1000 Genomes a été réalisée : après une étape de pré-phasage par le logiciel SHAPEIT [44] l'imputation proprement dite a été opérée par le logiciel IMPUTE 2 [45]. Les SNPs étaient pris en compte dans le processus d'imputation si toutes les conditions suivantes étaient réunies : call rate (proportion de données non manquantes)  $\geq 98\%$ , une valeur de  $p > 1 \times 10^{-6}$  au test d'équilibre de Hardy-Weinberg et une fréquence d'allèle mineur (allèle le moins fréquent)  $\geq 1\%$ .

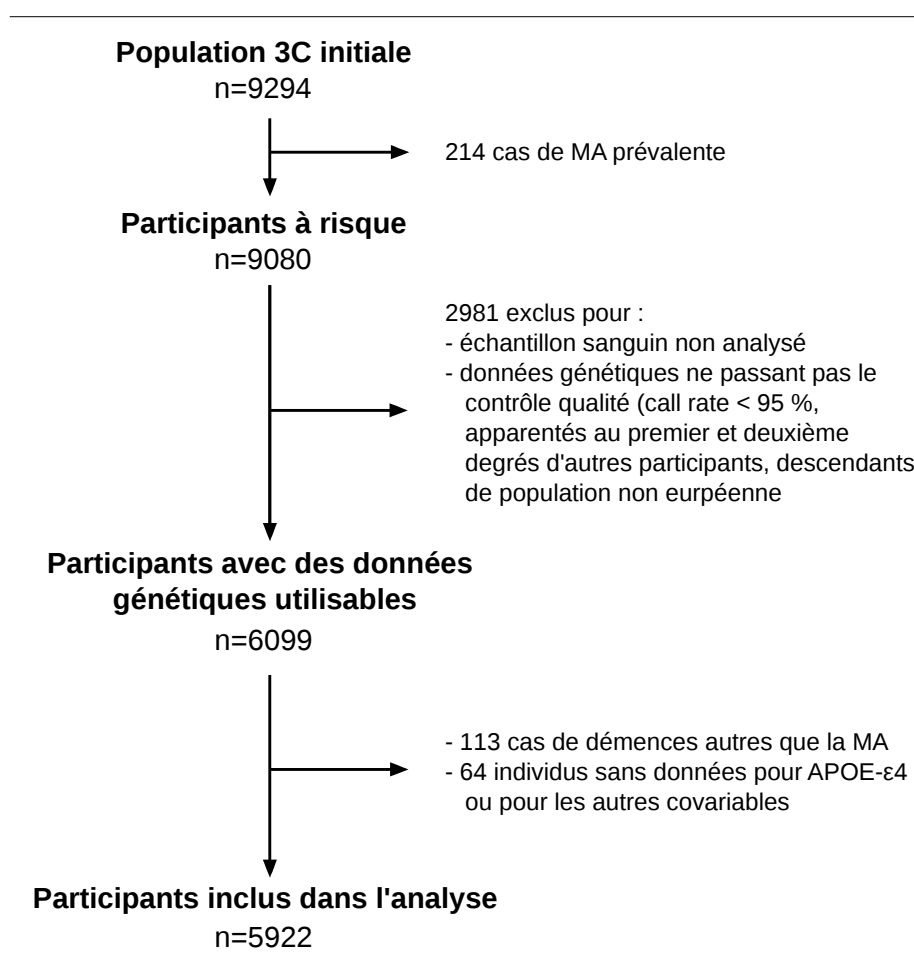
Parmi le grand nombre de SNPs obtenus, les trente deux SNPs identifiés par la méta-analyse IGAP [26] (19 SNPs associés à la MA et 13 SNPs suggestifs) ont été utilisés pour calculer le score de risque génétique. À partir des probabilités génotypiques obtenues grâce à l'imputation, les doses alléliques pour l'allèle à risque (nombre d'allèle augmentant le risque, compris entre 0 et 2), ont été calculées.

Le génotypage d'APOE- $\epsilon$ , habituellement difficile via les puces de génotypage, a été réalisé grâce à la technologie TaqMan [46]. La variable binaire « présence ou non d'au moins un allèle- $\epsilon 4$  » a été construite pour l'analyse.

## II. Sélection des participants pour l'analyse

Sur les 9 080 non déments à l'inclusion, 2 981 ont été exclus de l'analyse faute de donnée génétique (refus de prélèvement sanguin par le participant, échantillon ou données génétiques de mauvaise qualité). Les démences incidentes autres que la MA (n=113) et les participants présentant des données manquantes pour les covariables (n=64) ont également été exclus. L'échantillon final à analyser comportait

donc 5922 participants (**Figure 2**).



*Figure 2. Diagramme de flux.*

### III. Calcul des scores de risque génétique

#### A. Deux GRSs : 19 et 32 SNPs

En fonction du nombre de SNPs inclus dans le calcul, 2 scores ont été calculés. Le premier se basait uniquement sur les 19 SNPs associés à la MA dans la méta-analyse IGAP, dont la valeur de p était inférieure à  $5 \times 10^{-8}$  (GRS-19) (**Annexe 3**). Le second score intégrait également les 13 SNPs suggestifs de IGAP (**Annexe 3**), définis par une valeur de p comprise entre  $5 \times 10^{-8}$  et  $1 \times 10^{-6}$ , pour un total de 32 SNPs (GRS-32).

## B. Méthode de calcul

Les GRSs ont été calculés comme la somme des  $I$  doses d'allèles à risque  $G_i$ , pondérées par le logarithme des ORs respectifs  $OR_i$  (1). Pour en faciliter l'interprétation, une transformation a été appliquée afin qu'un point du score corresponde à une augmentation moyenne d'un allèle à risque (2). La valeur du GRS-19 était donc comprise entre 0 et 38, celle du GRS-32, entre 0 et 64.

$$GRS = \sum_{i=1}^I (\log(OR_i) \times G_i) \times \frac{I}{\sum_{i=1}^I \log(OR_i)}$$

(1)

(2)

## IV. Analyses statistiques

### A. Analyses descriptives de la population

Les caractéristiques à l'inclusion de la population d'analyse ont été résumées avec les effectifs et les pourcentages des variables catégorielles, et avec la moyenne et l'écart-type des variables quantitatives. Ces caractéristiques ont été comparées entre les MA incidentes et les non déments, à l'aide d'un test de Student pour comparer les moyennes et d'un test du Chi2 pour comparer les pourcentages. De la même manière, les caractéristiques initiales ont été comparées entre les individus inclus et les individus exclus de l'analyse, en vue d'évaluer le risque de biais de sélection induit par l'exclusion d'un nombre important de participants.

Les incidences cumulées (risques observés) en fonction du temps de suivi ont été estimées selon la méthode de Kaplan Meier (KM).

### B. Association des SNPs avec la MA

Les associations entre les 32 SNPs d'IGAP et la survenue d'une MA ont été



testées individuellement dans un modèle de Cox ajusté sur l'âge, le sexe, le centre, le niveau d'éducation, la langue maternelle et la présence d'APOE-ε4.

### **C. Association des GRSs avec la MA**

L'association entre chaque GRS (GRS-19 et GRS-32) et la survenue d'une MA a été estimée à l'aide des Hazard Ratios (HR) issus d'un modèle de Cox ajusté sur l'âge, le sexe, le centre, le niveau d'éducation, la langue maternelle et la présence APOE-ε4. Les conditions de validité du modèle ont été vérifiées à l'aide du test des résidus de Schoenfeld.

### **D. Mesures d'amélioration de la prédiction par les GRSs**

#### **1. Principe général**

La valeur ajoutée d'un GRS à une équation de risque visant à prédire la MA a été évaluée par la comparaison du pouvoir discriminant de 2 modèles prédictifs (modèle de Cox) : avec ou sans le GRS. Un modèle de base, incluant l'âge, le sexe, le centre, la langue maternelle, le niveau d'éducation et la présence d'APOE-ε4, a donc été comparé à un nouveau modèle comprenant également l'un des deux GRSs. Trois paramètres mesurant l'amélioration de la discrimination ont été calculés : la différence du C-index entre les 2 modèles (**ΔC-index**), le NRI continu (continuous Net Reclassification Improvement : **cNRI**) et l'**IDI** (Integrated Discrimination Improvement).

S'agissant d'une étude de cohorte, la prédiction de risque de MA devait être calculée au bout d'une période de suivi définie, les données de suivi étant alors censurées à ce délai. Pour cette étude, une période de 8 ans a été choisie car elle représentait un bon compromis entre un nombre suffisant de cas incidents de MA et un nombre de perdus de vue acceptable.

En résumé, l'amélioration de la prédiction du risque de développer une MA,

dans les 8 ans après l'inclusion, a été estimée à l'aide du  $\Delta C$ -index, du cNRI et de l'IDI.

## 2. C-index et différence de C-index

Le C-index ou C-statistique mesure la capacité discriminante du risque prédit par un modèle de Cox . Il peut être considéré comme une extension de l'aire sous la courbe Receiver Operating Characteristic (AUC) dans le cas où les données sont censurées. Tout comme l'AUC, ce paramètre est compris entre 0,5 et 1. Plus le C-index est proche de 1, plus la discrimination est bonne. Il s'interprète comme suit : pour un individu malade et un individu non malade tirés au sort, le C-index est la probabilité que l'individu malade ait un risque prédit, sur une période donnée, supérieur à celui du non malade. Plusieurs méthodes de calcul du C-index ont été proposées dans la littérature [47-50]. Celle décrite par Uno et al [50] a été utilisée dans cette étude. Le  $\Delta C$ -index représente l'amélioration du C-index par l'ajout du GRS dans le modèle. Il se calcule simplement comme la différence entre le C-index du nouveau modèle (avec GRS) et celui du modèle de base (sans GRS).

## 3. Continuous NRI (cNRI) et IDI

Le NRI et l'IDI mesurent également l'amélioration de la discrimination permise par l'ajout d'un nouveau marqueur à un modèle prédictif. Ils ont été récemment proposés par Pencina, d'Agostino et al [51] en complément de la différence de C-index ou d'AUC. Les méthodes de calcul de ces paramètres sont détaillées dans les **Figure 3** et **Figure 4**, dans le cas simple où les données ne sont pas censurées.

Le cNRI quantifie la capacité d'un marqueur à augmenter la probabilité (risque) prédite chez les individus qui vont devenir malades et à diminuer la probabilité prédite chez ceux qui ne le deviendront pas. Il est la somme entre :

- chez les individus qui développeront la maladie : la proportion d'entre-eux dont la probabilité prédite augmente, moins la proportion d'entre-eux dont la

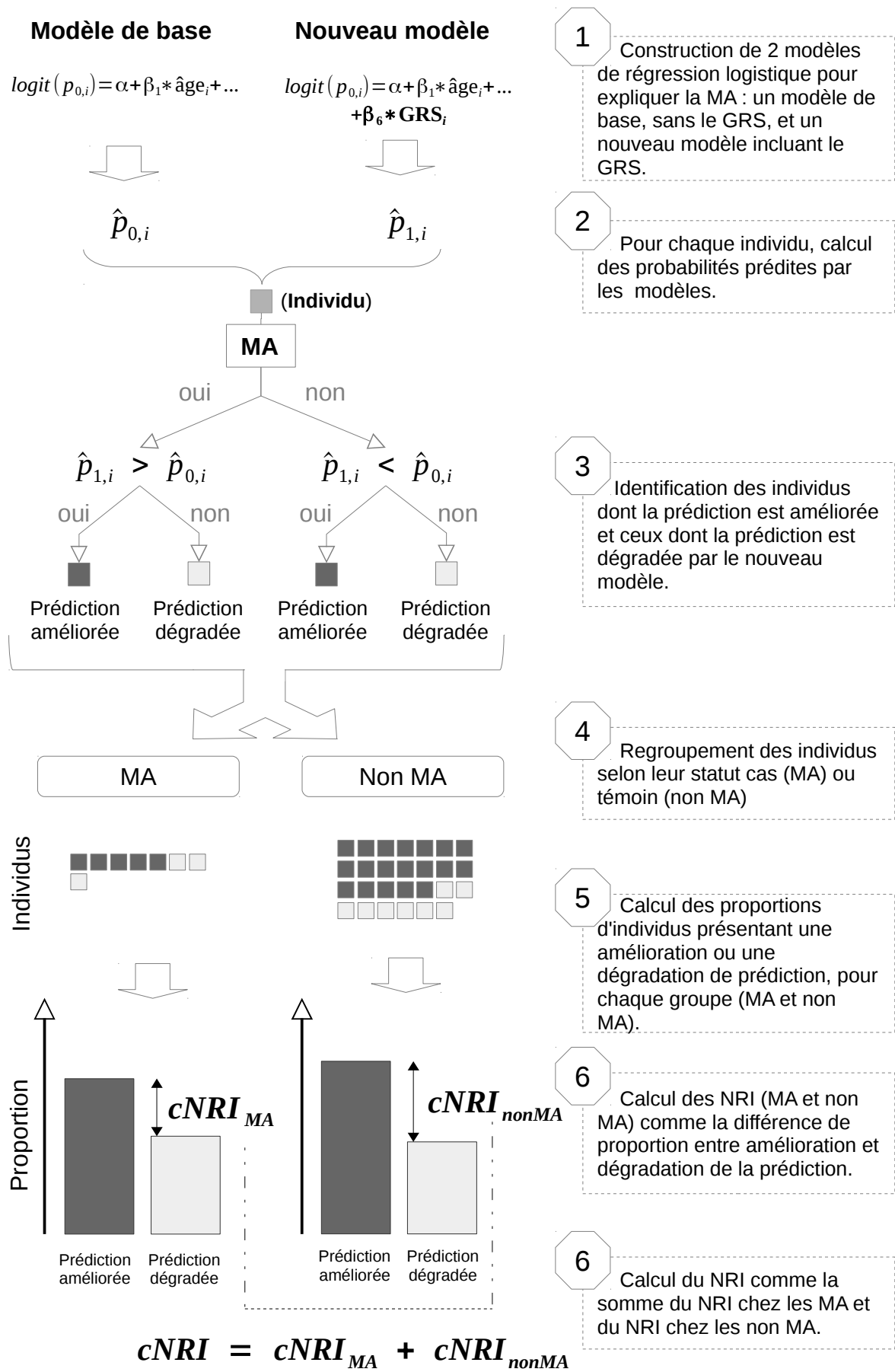


Figure 3. Méthode de calcul du NRI continu en l'absence de donnée censurée.

probabilité prédite diminue ; et

- chez les individus qui ne développeront pas la maladie : la proportion d'entre-eux dont la probabilité prédite diminue, moins la proportion d'entre-eux dont la probabilité prédite augmente.

Le cNRI est compris entre 0 et 2. Le cNRI divisé par deux s'interprète comme l'amélioration moyenne de la discrimination.

Tout comme le  $\Delta C$ -index, l'IDI est une différence d'une mesure de discrimination entre le modèle de base et le nouveau modèle. Il s'agit en effet de la différence entre l'écart de risque moyen prédit chez les malades et celui chez les non malades. Il peut

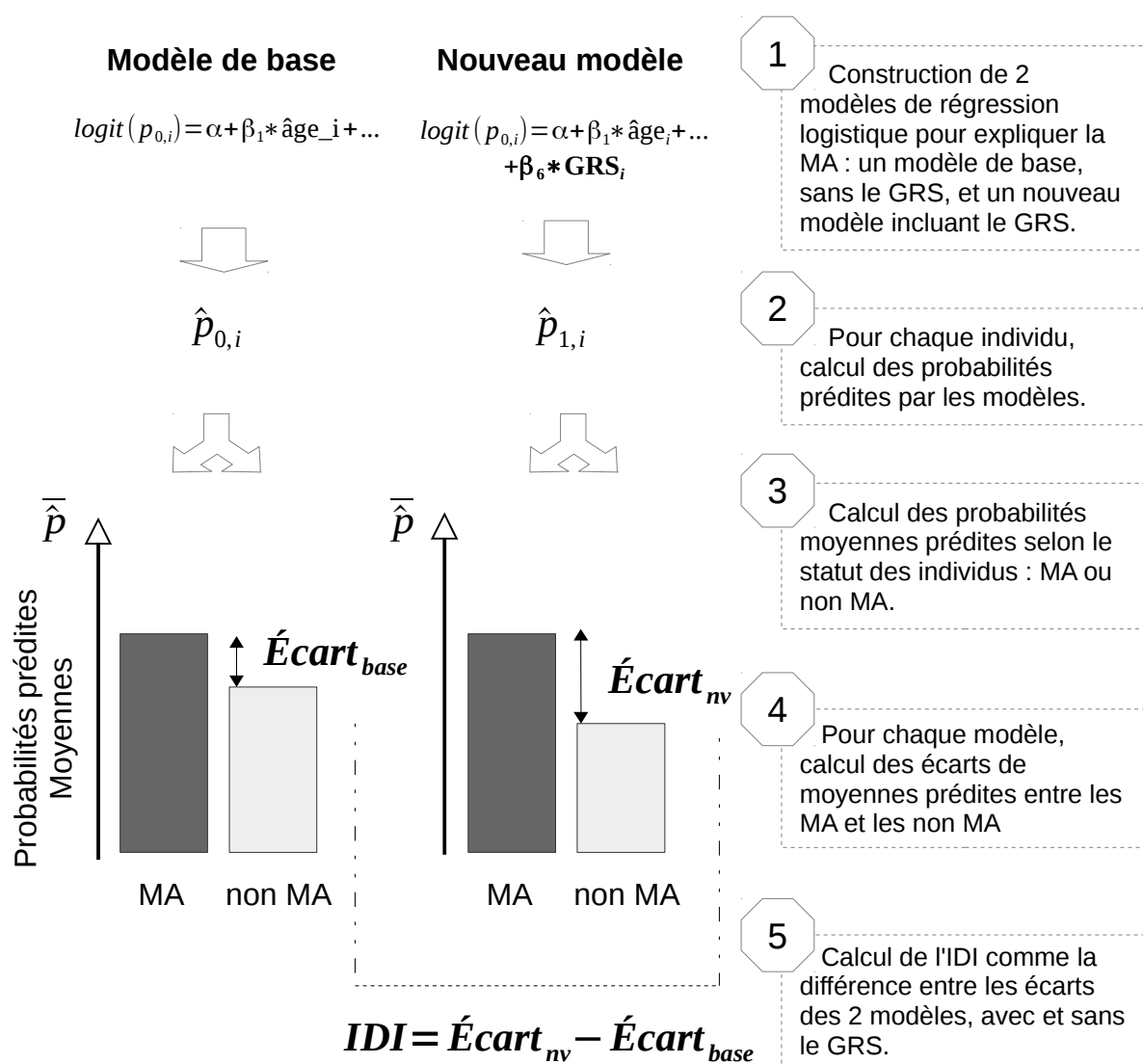


Figure 4. Méthode de calcul de l'IDI en l'absence de donnée censurée.

également être vu comme la différence des coefficients de détermination des modèles ( $R^2$ ) [52], ou comme l'augmentation moyenne de la sensibilité à spécificité constante. Sa valeur est comprise entre 0 et 1.

L'extension du NRI et de l'IDI pour tenir compte des données censurées proposée par Pencina et al [53], puis développé par Uno et al [54], a été utilisée pour l'analyse. Les intervalles de confiance à 95% ont été estimés par bootstrap de 1000 itérations. Pour conclure à une amélioration statistiquement significative, les 3 paramètres devaient être positifs, avec un intervalle de confiance à 95 % strictement supérieur à 0. La bonne calibration des modèles, autorisant l'interprétation de ces paramètres a été évaluée graphiquement par la représentation du risque prédit en fonction du risque observé.

### **E. Analyses complémentaires**

Pour explorer si la valeur ajoutée du GRS était meilleure chez les porteurs d'APOE- $\epsilon$ 4, le terme d'interaction APOE- $\epsilon$ 4 x GRS a été testé. Les 3 paramètres d'amélioration de la prédiction ont également été estimés séparément chez les porteurs et les non porteurs d'APOE- $\epsilon$ 4.

Par ailleurs, afin de confirmer et d'évaluer l'utilité d'APOE- $\epsilon$ 4 pour la prédiction du risque de MA, l'amélioration de la discrimination liée à APOE- $\epsilon$ 4 a été calculée, à partir d'un modèle de base comprenant l'âge, le sexe, le centre, le niveau d'éducation et la langue maternelle.

Dans une analyse de sensibilité, les analyses ont été répétées en ajoutant les facteurs de risque cardio-vasculaire dans le modèle de base : hypertension, diabète, hypercholestérolémie, tabac et antécédents personnels cardio-vasculaires.

### **F. Logiciel utilisé**

Les analyses statistiques ont été réalisées à l'aide du logiciel R version 3.0 [55]

et des bibliothèques « survival », « survIDINRI » et « survC1 ».

## RÉSULTATS

### I. Analyses descriptives

#### A. Population d'étude

Les caractéristiques à l'inclusion des participants sont présentées dans le

**Tableau 4.**

*Tableau 4. Descriptif de la population analysée.*

	Population totale N = 5922	Non MA N = 5463	MA incidentes N = 459	P *
<b>Age</b> (année)	74,2 ± 5,5	73,9 ± 5,4	77,7 ± 5,5	<0,001
<b>Femme</b>	3598 (60,8%)	3309 (60,6%)	289 (63,0%)	0,9
<b>Centre</b>				<0,001
Bordeaux	1441 (24,3%)	1291 (23,6%)	150 (32,7%)	
Dijon	3867 (65,3%)	3588 (65,7%)	279 (60,8%)	
Montpellier	614 (10,4%)	584 (10,7%)	30 (6,5%)	
<b>Niveau d'éducation</b>				<0,001
Bas	1998 (33,7%)	1781 (32,6%)	217 (47,3%)	
Moyen-bas	937 (15,8%)	890 (16,3%)	47 (10,2%)	
Moyen-haut	1592 (26,9%)	1495 (27,4%)	97 (21,1%)	
Haut	1395 (23,6%)	1297 (23,7%)	98 (21,4%)	
<b>Langue maternelle : français</b>	5776 (97,5%)	5336 (97,7%)	440 (95,9%)	<0,001
<b>MMSE</b>	27,4 ± 1,9	27,5 ± 1,9	26,3 ± 2,0	<0,001
<b>Diabète</b>	4653 (78,6%)	497 (9,1%)	69 (15,0%)	<0,001
<b>Hypertension</b>	3387 (57,2%)	4279 (78,3%)	374 (81,5%)	0,1
<b>Hypercholestérolémie</b>	325 (5,5%)	3125 (57,2%)	262 (57,1%)	0,9
<b>Tabagisme actuel</b>	561 (9,5%)	310 (5,7%)	15 (3,3%)	0,03
<b>ATCD de maladie CV</b>	561 (9,5%)	492 (9,0%)	69 (15,0%)	<0,001
<b>Porteur d'APOE-ε4</b>	1210 (20,2 %)	1174 (21,5%)	138 (30%)	<0,001
<b>GRS-19</b>	19,6 ± 2,6	19,6 ± 2,6	19,9 ± 2,6	0,002
<b>GRS-32</b>	25,8 ± 2,9	25,8 ± 2,9	26,2 ± 2,9	0,006

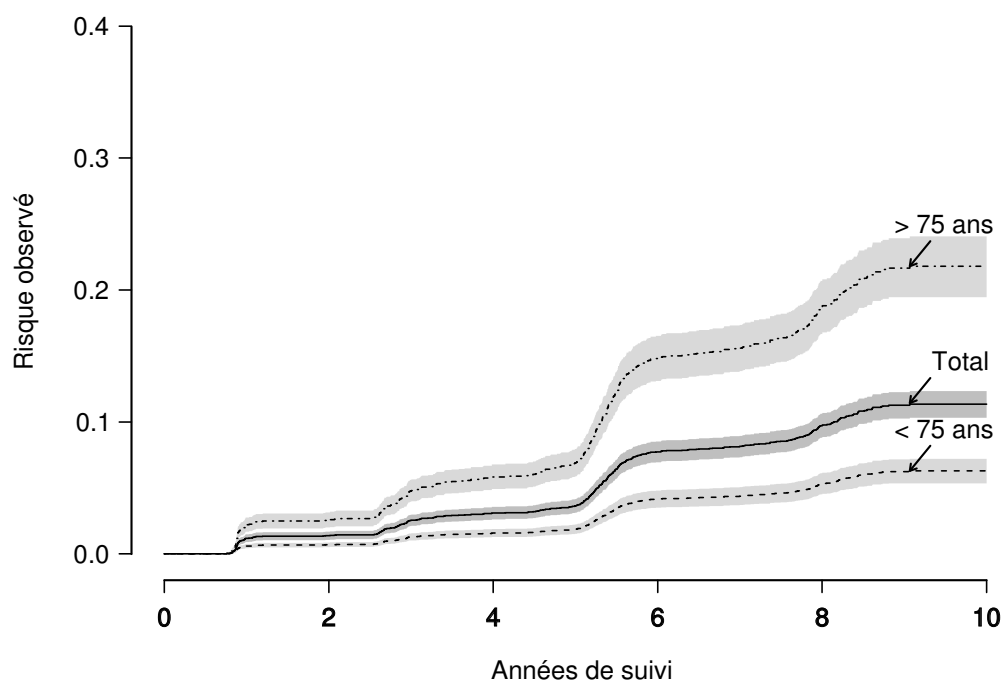
Tableau présentant la moyenne ± écart-type des variables quantitatives et le nombre d'individus (%) des variables binaires ou catégorielles.

\* valeur de p issue de la comparaison entre MA et non MA, estimée par un test de Student ou un test du Chi2.

Plus de la moitié des participants était suivie à Dijon (65,3 %). L'âge moyen était de  $74,2 \pm 5,5$  ans avec un sexe ratio de 0.64. Vingt pour cent d'entre-eux était porteurs d'au moins un allèle APOE- $\epsilon 4$ . Comme attendu, la proportion de sujets porteurs d'au moins un allèle APOE- $\epsilon 4$  était significativement plus importante chez les sujets ayant développé une MA comparée à celle des non déments.

## B. Risque observé de MA

A la date de point, 459 participants ont présenté une MA. La médiane de suivi était de 7,6 ans. La **Figure 5** représente la courbe de risque estimée par la méthode de Kaplan Meier. À 8 ans de suivi, le risque de démence était de 9,7 % : 417 participants avaient développé une MA, 2 818 n'avaient pas présenté de démence, et



Individu à risque	5922	5014	4052	3736	2818	66
Nb cumulé de MA	0	75	160	349	417	459
Risque de MA	0%	1,3%	3,1%	7,7%	9,7%	11,3%

**Figure 5. Risque observé de développer une MA en fonction du temps de suivi et de l'âge d'inclusion, calculé par la méthode Kaplan Meier.**

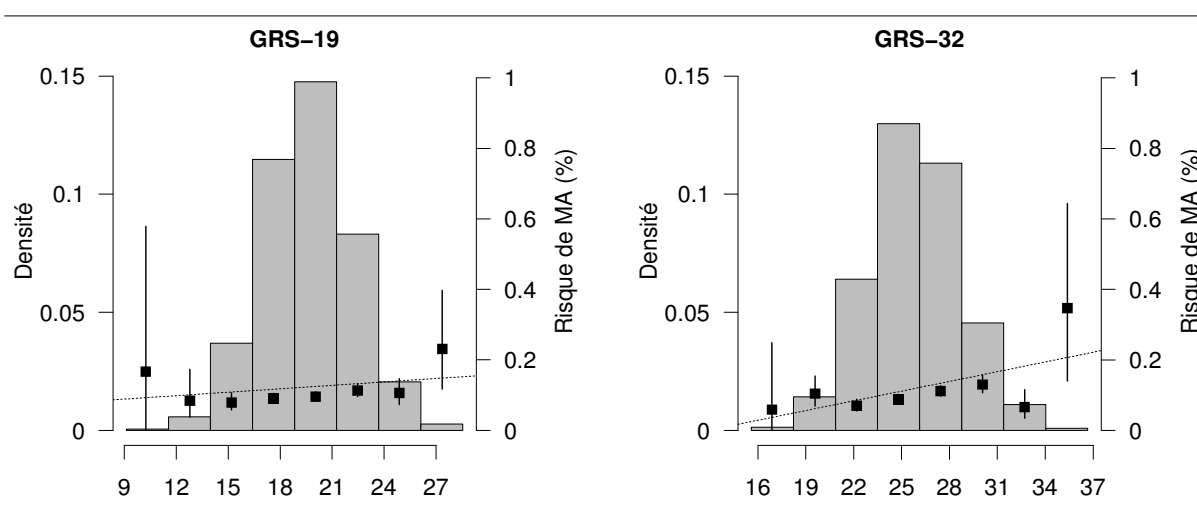
*NB : Le suivi théorique à 10 ans était en pratique réalisé entre 9 et 10 ans.*



2 687 étaient perdus de vue ou décédés.

### C. Score de risque génétique

En moyenne le GRS-19 était de 19.6 ( $\pm 2,6$ ) et le GRS-32 était de 25,8 ( $\pm 2,9$ ) (**Tableau 3**). Les distributions des GRSs ainsi que les risques de MA observés à 8 ans en fonction des valeurs de GRS sont présentés dans la **Figure 6**.



**Figure 6. Distributions des GRSs et risques observés de MA par tranches de GRS.** Les histogrammes représentent les distributions des GRSs. Les points représentent le risque observé de développer une MA dans les 8 ans (et leurs intervalles de confiance à 95 %), estimés par la méthode de Kaplan Meier.

## II. Mesures d'association

### A. Analyse individuelle des SNPs

Estimés dans des modèles de Cox séparés, les HRs de 25 des 32 SNPs révèlent une augmentation du risque pour un allèle à risque supplémentaire ( $HR > 1$ ), en accord avec les résultats d'IGAP. Néanmoins, aucune association n'était statistiquement significative, que ce soit en utilisant un seuil de valeur de p conservateur de 0,0015 (correction de Bonferroni), ou avec un seuil de 0,05 (**Tableau 5**).

**Tableau 5. Hazard Ratios (HR) des 19 variants estimés séparément à l'aide d'un modèle de Cox.**

SNP	Gène le plus proche <sup>§</sup>	3C			IGAP	
		FAR*	HR (95% CI)	p	OR (95% CI)	p
<b>1) SNPs associés avec la MA</b>						
rs6656401	CR1	0.18	1.03 (0.83-1.29)	0.76	1.18 (1.14-1.22)	5.69x10 <sup>-24</sup>
rs35349669	INPP5D	0.46	0.88 (0.74-1.04)	0.12	1.08 (1.05-1.11)	3.17x10 <sup>-08</sup>
rs6733839	BIN1	0.38	1.14 (0.96-1.35)	0.13	1.22 (1.18-1.25)	6.94x10 <sup>-44</sup>
rs190982	MEF2C	0.61	1.13 (0.95-1.35)	0.15	1.08 (1.05-1.11)	3.23x10 <sup>-08</sup>
rs10948363	CD2AP	0.28	1.06 (0.89-1.27)	0.52	1.10 (1.07-1.13)	5.20x10 <sup>-11</sup>
rs77372866	HLA-DRB5/DRB1	0.25	0.97 (0.80-1.27)	0.76	1.11 (1.08-1.15)	2.94x10 <sup>-12</sup>
rs11771145	EPHA1	0.66	1.13 (0.80-1.18)	0.17	1.11 (1.08-1.14)	1.12x10 <sup>-13</sup>
rs1476679	ZCWPW1	0.71	1.00 (0.84-1.19)	0.99	1.09 (1.06-1.12)	5.58x10 <sup>-10</sup>
rs2718058	NME8	0.61	0.99 (0.84-1.17)	0.94	1.08 (1.05-1.11)	4.76x10 <sup>-09</sup>
rs28834970	PTK2B	0.37	1.07 (0.90-1.27)	0.45	1.10 (1.08-1.13)	7.37x10 <sup>-14</sup>
rs9331896	CLU	0.62	1.07 (0.90-1.27)	0.46	1.16 (1.13-1.19)	2.77x10 <sup>-25</sup>
rs10792832	PICALM	0.64	0.89 (0.76-1.06)	0.20	1.15 (1.12-1.18)	9.32x10 <sup>-26</sup>
rs10838725	CELF1	0.36	1.03 (0.87-1.22)	0.75	1.08 (1.05-1.11)	1.12x10 <sup>-08</sup>
rs11218343	SORL1	0.96	1.51 (0.93-2.44)	0.10	1.30 (1.22-1.39)	9.73x10 <sup>-15</sup>
rs983392	MS4A6A	0.58	1.06 (0.90-1.26)	0.47	1.11 (1.09-1.14)	6.14x10 <sup>-16</sup>
rs10498633	SLC24A4/RIN3	0.79	1.16 (0.95-1.41)	0.16	1.10 (1.06-1.13)	5.54x10 <sup>-09</sup>
rs17125944	FERMT2	0.07	1.09 (0.80-1.49)	0.60	1.14 (1.09-1.19)	7.95x10 <sup>-09</sup>
rs4147929	ABCA7	0.19	1.16 (0.95-1.42)	0.15	1.15 (1.11-1.19)	1.06x10 <sup>-15</sup>
rs7274581	CASS4	0.91	1.12 (0.83-1.51)	0.45	1.14 (1.09-1.20)	2.46x10 <sup>-08</sup>
<b>2) SNPs suggestifs</b>						
rs6678275	None	0.17	1.04 (0.84-1.30)	0.72	1.09 (1.05-1.13)	3.42x10 <sup>-07</sup>
rs6448799	HS3ST1	0.30	1.08 (0.90-1.30)	0.42	1.08 (1.05-1.11)	6.61x10 <sup>-08</sup>
rs72807343	SOSTM1	0.02	1.19 (0.62-2.28)	0.61	1.35 (1.20-1.52)	7.40x10 <sup>-07</sup>
rs9381040	TREML2	0.70	1.19 (1.00-1.43)	0.06	1.07 (1.04-1.10)	6.35x10 <sup>-07</sup>
rs7818382	NDUFAF6	0.46	1.16 (0.99-1.36)	0.07	1.07 (1.04-1.10)	8.00x10 <sup>-08</sup>
rs7920721	ECHDC3	0.40	0.95 (0.80-1.13)	0.58	1.07 (1.04-1.10)	2.89x10 <sup>-07</sup>
rs10751667	AP2A2	0.65	1.04 (0.87-1.24)	0.66	1.07 (1.04-1.10)	6.28x10 <sup>-07</sup>
rs7295246	ADAMST20	0.40	0.91 (0.77-1.08)	0.27	1.07 (1.04-1.10)	2.96x10 <sup>-07</sup>
rs2337406	IGH@	0.87	1.01 (0.79-1.29)	0.95	1.14 (1.09-1.21)	2.72x10 <sup>-07</sup>
rs74615166	TRIP4	0.02	1.12 (0.64-1.96)	0.69	1.29 (1.17-1.42)	4.29x10 <sup>-07</sup>
rs8035452	SPPL2A	0.66	1.00 (0.84-1.19)	0.97	1.07 (1.04-1.10)	3.18x10 <sup>-07</sup>
17:61538148	ACE	0.02	0.79 (0.41-1.54)	0.49	1.34 (1.20-1.50)	3.08x10 <sup>-07</sup>
rs7225151	SCIMP	0.12	1.01 (0.79-1.29)	0.92	1.10 (1.06-1.15)	3.66x10 <sup>-07</sup>

Les HRs sont ajustés sur l'âge, le sexe, le centre, le niveau d'éducation, la langue maternelle et la présence d'un allèle APOE-ε4.

\* Fréquence de l'allèle de référence (ici l'allèle de référence est l'allèle à risque).

§ Gène situé à une proximité de ± 100 kb du SNP.

## B. GRSs

Le GRS-19 et le GRS32 étaient significativement associés à une augmentation du risque de MA. Les HRs ajustés sur l'âge, le sexe, le centre, le niveau d'éducation, la langue maternelle et APOE-ε4 sont présentés dans le **Tableau 6**. Pour un point supplémentaire du GRS (un allèle à risque), le risque de MA était augmenté de 5% pour le GRS-19 et 6 % pour le GRS-32. Les HR [IC à 95%] pour l'augmentation de une déviation standard étaient de 1,15 [1,05 ; 1,26] pour le GRS-19 et de 1,24 [1,12 ; 1,33] pour le GRS-32. Les HRs par tertile de GRS sont représentés dans la **Figure 7**.

**Tableau 6. Hazard Ratios des GRSs et d'APOE-ε4.**

	Modèle de base + GRS-19		Modèle de base + GRS-32	
	HR	IC 95 %	HR	IC 95 %
<b>GRS</b>	1,05	(1,01 ; 1,09)	1,06	(1,03 ; 1,09)
<b>APOE-ε4</b>	2,18	(1,78 ; 2,66)	2,20	(1,80 ; 2,70)

*Modèle de base : modèle incluant l'âge, le sexe, le centre, la langue maternelle et APOE-ε4.*

## C. APOE-ε4

APOE-ε4 était fortement associé à la MA, avec un risque multiplié par 2 chez les participants possédant au moins un allèle APOE-ε4 (**Tableau 6**).

## III. Interaction APOE-ε4 x GRS

Que ce soit pour le GRS-19 ou le GRS-32, l'interaction avec APOE-ε4 n'était pas statistiquement significative : valeur de  $p = 0.7$  pour le GRS-19 et valeur de  $p = 0.9$  pour le GRS-32. Chez les non porteurs d'APOE-ε4, l'HR [95%CI] était de 1,06 [1,01;1,11] pour la GRS-19 et 1,06 [1,02;1,10] pour le GRS-32. Chez les porteurs,

l'HR [95%CI] était de 1,04 [0,97;1,11] pour le GRS-19 et 1,05 [0,99;1,12] pour le GRS-32. Les résultats par tertile de GRS au sein de chacune des strates (porteurs et non porteurs d'APOE-ε4) sont présentés dans la **Figure 7**.

## IV. Mesures d'amélioration de la prédiction

### A. Apport des GRSs

Le  $\Delta$ C-index, le cNRI et l'IDI obtenus par l'addition des GRSs au modèle de base sont présentés dans le **Tableau 7**. Aucun des 3 paramètres n'était significativement différent de 0 avec le GRS-19. Le C-index du modèle de base était de 0,760 et l'ajout du GRS-19 montrait une augmentation de 0,0026 du C-index. Une amélioration de la prédiction était observée chez 4,8 % des cas incidents et 4,6 % des non déments pour un cNRI de 0,094. L'IDI était d'environ 0,0019.

**Tableau 7. Amélioration de la prédiction induite par l'ajout des GRSs à un modèle de base.**

	GRS-19		GRS-32	
	Valeur	IC 95 %	Valeur	IC 95 %
<b><math>\Delta</math>C-index</b>	0,00263	(-0,00117 ; 0,00642)	0,00471	(-0,000547 ; 0,00997)
Modèle de base*	0,760	(0,741 ; 0,780)	0,760	(0,741 ; 0,780)
Modèle de base + GRS	0,763	(0,743 ; 0,783)	0,765	(0,745 ; 0,785)
<b>cNRI</b>	0,0941	(-0,00176 ; 0,216)	<b>0,201</b>	<b>(0,0973 ; 0,302)</b>
Chez les MA	0,0480	(-0,0186 ; 0,131)	0,109	(0,0322 ; 0,175)
Chez les non MA	0,0461	(-0,00441 ; 0,0996)	0,0923	(0,0366 ; 0,149)
<b>IDI</b>	0,00186	(-0,000274 ; 0,00686)	<b>0,00358</b>	<b>(0,000268 ; 0,00998)</b>

*Modèle de base : modèle incluant l'âge, le sexe, le centre, la langue maternelle et APOE-ε4.*

Les résultats du GRS-32 semblaient légèrement meilleurs, avec notamment un cNRI de 0,20, soit une amélioration de la prédiction d'environ 10 % chez les cas incidents de MA et chez les non déments, et un l'IDI de 0,0036. Le cNRI et l'IDI

étaient, tout deux significativement différents de 0. La calibration des modèles était bonne (**Annexe 6**).

Les résultats obtenus séparément chez les porteurs et les non porteurs de

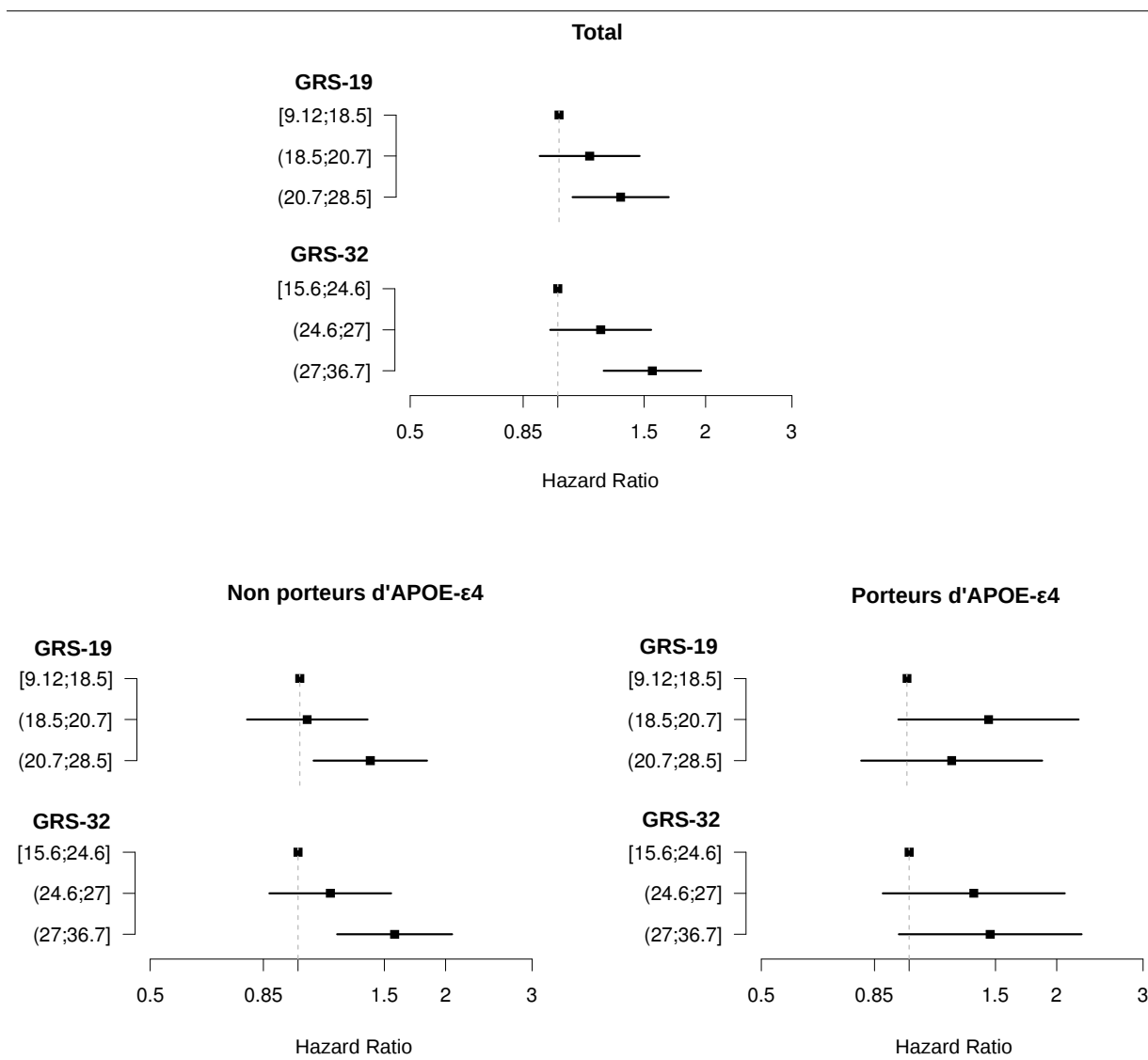


Figure 7. Hazard Ratios par tertiles de GRS sur l'ensemble de l'échantillon puis séparément chez les porteurs et les non porteurs d'APOE-ε4.

l'allèle APOE-ε4 sont présentés dans le **Tableau 8**.

### B. Apport d'APOE-ε4

L'amélioration de la prédiction liée à APOE-ε4 par rapport à un modèle incluant l'âge, le sexe, le centre, le niveau d'éducation et la langue maternelle, montrait une

augmentation significative des 3 paramètres (**Tableau 9**). Le  $\Delta$ C-index était de 0,0087 et l'IDI de 0,013 et le cNRI de 0,20.

### **C. Modèle de base comprenant les facteurs de risque cardio-vasculaire**

Lorsque les facteurs de risque cardio-vasculaire étaient ajoutés au modèle de base, le C-index de base était de 0,769. Les paramètres d'amélioration de la discrimination ne semblaient pas différents (**Annexe 7**).

Tableau 8. Amélioration de la prédiction induite par l'ajout des GRSs, mesurée séparément chez les porteurs et les non porteurs d'APOE-ε4.

	GRS-19				GRS-32			
	Non porteurs d'APOE-ε4		Porteurs d'APOE-ε4		Non porteurs d'APOE-ε4		Porteurs d'APOE-ε4	
	Valeur	IC 95 %	Valeur	IC 95 %	Valeur	IC 95 %	Valeur	IC 95 %
<b>ΔC-index</b>	0,00168	(-0,00225 ; 0,0056)	0,00155	(-0,00542 ; 0,00852)	0,0041	(-0,00151 ; 0,00972)	0,00444	(-0,00442 ; 0,0133)
Modèle de base	0,771	(0,744 ; 0,799)	0,726	(0,679 ; 0,773)	0,771	(0,744 ; 0,799)	0,726	(0,679 ; 0,773)
Modèle de base + GRS	0,773	(0,745 ; 0,801)	0,727	(0,682 ; 0,773)	0,775	(0,749 ; 0,802)	0,730	(0,683 ; 0,777)
<b>cNRI</b>	0,103	(-0,0224 ; 0,224)	0,140	(-0,0868 ; 0,378)	<b>0,193</b>	<b>(0,0623 ; 0,328)</b>	0,162	(-0,0548 ; 0,393)
Chez les MA	0,0614	(-0,0286 ; 0,149)	0,058	(-0,0904 ; 0,213)	0,104	(0,0133 ; 0,200)	0,0647	(-0,0773 ; 0,235)
Chez les non MA	0,0414	(-0,0184 ; 0,105)	0,0819	(-0,0612 ; 0,198)	0,0894	(0,0236 ; 0,151)	0,0971	(-0,0183 ; 0,199)
<b>IDI</b>	0,00290	(-0,000328 ; 0,0108)	0,000534	(-0,00144 ; 0,0122)	0,00395	(-0,000123 ; 0,0121)	0,00330	(-0,000872 ; 0,0181)

Modèle de base : modèle incluant l'âge, le sexe, le centre, le niveau d'éducation et la langue maternelle.

**Tableau 9. Amélioration de la discrimination induite par l'ajout d'APOE-ε4 à un modèle de base.**

	APOE-ε4	
	Valeur	IC 95 %
<b>Δ C-index</b>	<b>0,00874</b>	<b>(0,000131 ; 0,0176)</b>
Modèle de base	0,752	(0,729 ; 0,774)
Modèle de base + APOE-ε4	0,760	(0,739 ; 0,783)
<b>cNRI</b>	<b>0,198</b>	<b>(0,107 ; 0,285)</b>
Chez les MA	-0,429	(-0,514 ; -0,348)
Chez les non MA	0,627	(0,597 ; 0,656)
<b>IDI</b>	<b>0,0133</b>	<b>(0,00475 ; 0,0248)</b>

*Modèle de base : modèle incluant l'âge, le sexe, le centre, le niveau d'éducation et la langue maternelle.*



## DISCUSSION

### I. Résultats principaux

Dans cette cohorte prospective de 5 922 participants âgés de 65 ans et plus, deux scores de risque génétique, basés sur les derniers résultats de GWAS [26], étaient significativement associés à une augmentation du risque de MA. En moyenne, la présence d'un allèle à risque supplémentaire augmentait le risque de 5 à 6 %. L'ajout des GRSs à un modèle prédictif incluant les principaux facteurs de risque (dont l'allèle APOE- $\epsilon$ 4), ne montrait en revanche qu'une faible amélioration de la prédiction du risque de MA, indépendante de la présence de l'allèle APOE- $\epsilon$ 4. Néanmoins, les résultats étaient plus encourageants pour le GRS-32 que pour le GRS-19.

### II. Comparaison à la littérature

Dans une étude de 2012 recherchant les relations entre un GRS composé des 9 loci et les composantes de la cognition, Verhaaren et al [41] mettaient déjà en évidence une association entre le GRS et la MA : HR [95 % CI] de 1,14 [1,03-1,27] pour une augmentation d'une déviation standard du GRS, après ajustement sur l'âge, le sexe et la présence d'APOE- $\epsilon$ 4. La différence de C-index observée ne mettait pas en évidence d'amélioration statistiquement significative de la discrimination (**Annexe 3**). L'actualisation du GRS avec les nouveaux SNPs identifiés par le consortium IGAP offrait donc une légère amélioration de la force d'association et un bénéfice prédictif plus prometteur, notamment avec le GRS-32. D'autres études [40,42,56] dont les résultats sont plus difficilement comparables en

raison du schéma de l'étude ou du phénotype étudié, sont présentées dans l'**Annexe 3**. Si la majorité d'entre elles montrait une association avec la maladie, aucune amélioration significative de la prédiction, basée sur la différence de C-index ou d'AUC, n'était mise en évidence.

Lorsque ces performances étaient comparées à celles des GRSs construits sur le même principe pour d'autres maladies complexes comme le diabète [36,58], les maladies cardio-vasculaires [38,59-62] ou l'insuffisance rénale [63], les résultats étaient du même ordre de grandeur. Des GRSs plus performants ont cependant été observés pour la prédiction de la fibrillation ventriculaire [39].

### **III. Apport des 13 SNPs suggestifs**

Les meilleurs résultats observés avec le GRS-32 par rapport au GRS-19 impliquent la présence d'informations pertinentes pour la prédiction au sein des 13 SNPs suggestifs. Bien qu'ils ne dépassent pas le seuil GWAS de significativité ( $5 \times 10^{-8}$ ), plusieurs éléments suggèrent en effet qu'il s'agirait de vrais signaux : rs9381040, localisé près de TREML2, est également proche du gène TREM2, un gène récemment impliqué dans la MA [64,65] ; rs74615166 situé dans le locus du gène TRIP4 a été répliqué dans une étude récente [66].

### **IV. Confirmation du rôle d'APOE-ε4**

Les performances observées avec l'allèle APOE-ε4, confirment les conclusions de précédentes études [19,35] : APOE-ε4 est un facteur de risque essentiel pour la prédiction du risque qui a sa place au sein d'équations de risque. Il est toutefois important de rappeler que, selon la HAS, le génotypage d'APOE n'est recommandé, à l'heure actuelle, ni comme test de dépistage ni comme test diagnostique complémentaire de la MA [67].

## V. Choix de la méthode de calcul des GRSs

Plusieurs méthodes de calcul de GRS ont été décrites pour évaluer le rôle de la génétique dans les pathologies complexes [57]. Les plus fréquemment rencontrées combinent les SNPs sous forme de doses alléliques comme dans ce travail. Plus rarement, le calcul se base directement sur le génotype (AA, Aa ou aa).

Lorsque la dose allélique est utilisée, le calcul du GRS consiste en la somme des doses ( $G_i$ ), pondérée par un facteur représentant la taille de l'effet du SNPs ( $w_i$ ) :

$$GRS = \sum_{i=1}^I w_i G_i$$

Selon la pondération  $w_i$ , plusieurs méthodes sont possibles : sans pondération (comptage simple), avec une pondération par la taille de l'effet des SNPs issues de la littérature (somme pondérée par le logarithme de l'OR), ou avec une pondération estimée dans un modèle de régression dédiée. La description, les avantages et les inconvénients de chacune d'elles sont décrits dans l'**Annexe 8**.

Toutefois, l'utilisation de la dose allélique se fonde sur l'hypothèse forte que le risque augmente proportionnellement au nombre d'allèles (modèle additif). Or, cette hypothèse ne se vérifie pas en cas d'effet dominant (sur-risque équivalent chez les homozygotes et les hétérozygotes) ou récessif (sur-risque uniquement chez les homozygotes). Dans cette situation, il est donc préférable d'utiliser le génotype pour construire un score plus précis. Cette méthode est cependant plus consommatrice de degrés de liberté lors de la construction des modèles (deux variables par SNP au lieu d'une) et tous les logiciels employés pour les GWAS ne permettent pas de l'utiliser.

Dans ce travail, la méthode par somme des doses alléliques pondérées par les tailles d'effet issues de la littérature paraissait la plus adaptée puisque :

- Le calcul du score de risque génétique se basait sur les résultats de la plus large méta-analyse de GWAS dans le champ de la MA, permettant des estimations précises des ORs ;
- Ces estimations se basaient sur la dose allélique et non sur le génotype complet ;
- Cette méthode étaient la plus fréquemment retrouvée dans la littérature, notamment dans les précédentes études de GRSs dans la MA ;
- Une seule population était à notre disposition. Sa taille et son faible nombre de MA incidentes ne permettaient pas de la diviser en 2 populations : une pour la construction du GRS basée sur un modèle de régression, et une autre pour son évaluation.

Par ailleurs, l'observation d'une association entre les GRSs et la MA alors qu'aucun des SNPs utilisés pour le calcul du GRS n'était significativement associé à la MA suggère que la méthode d'agrégation employée était pertinente.

## **VI. Utilité du cNRI et de l'IDI pour mesurer les performances d'un nouveau prédicteur**

Une mesure clé pour évaluer l'utilité en clinique des équations de risque, est sa capacité à discriminer ou séparer ceux qui développeront la maladie de ceux qui ne la développeront pas. Plusieurs paramètres mesurant la discrimination ont été proposés [73], mais l'AUC/C-index est de loin le plus populaire. Par conséquent, lorsqu'il s'agit de mesurer l'amélioration de la discrimination obtenue par l'ajout d'un nouveau prédicteur dans un modèle de prédiction de risque, une approche naturelle et intuitive consiste à faire la différence d'AUC/C-index. Cette mesure présente cependant un certain nombre de limites [74]. Généralement faible et conservatrice, elle pourrait écarter à tort des marqueurs pourtant réellement importants. En outre,

elle n'est pas comparable d'une étude à l'autre si le modèle de base n'est pas le même car elle dépend fortement de la discrimination du modèle de base (plus il est élevé moins la différence est importante). Ces éléments ont conduit Pencina, d'Agostino et al [51] à proposer deux nouvelles mesures : le cNRI et l'IDI. Si l'utilité de ces dernières fut controversée au moment de leur publication [68,69], elles semblent aujourd'hui avoir été adoptées par les chercheurs et les cliniciens, notamment dans le domaine des maladies cardio-vasculaires. Pencina et al précisent que le  $\Delta C$ -index, le cNRI et l'IDI offrent des informations complémentaires [70-72]. Ces mesures doivent donc être rapportées ensembles pour obtenir une évaluation globale de la valeur ajoutée du nouveau modèle par rapport au modèle de base.

Une difficulté liée à l'utilisation de ces mesures est l'évaluation de l'amplitude de l'amélioration de la prédiction qu'elles représentent. Pour faciliter leur interprétation, Pencina et al [71] ont produit une série de résultats basés sur des données simulées. Les valeurs théoriques de  $\Delta C$ -index, cNRI et IDI y sont calculées pour trois tailles d'effet : faible, moyenne et élevée. L'**Annexe 9** présente un extrait de ces valeurs adaptées au contexte de l'étude. Elles permettent d'attribuer une taille d'effet faible au GRS-32 et moyenne à l'allèle APOE- $\epsilon 4$ .

Toutefois, il n'existe pas de règle d'inférence précise qui serait fonction de ces trois mesures et qui permettrait de conclure ou non à un apport significatif. Dans ce travail, nous avons fait un choix plutôt conservateur de ne conclure à une amélioration significative que si les 3 intervalles de confiance à 95 % étaient strictement supérieurs à 0.

## VII. Forces et faiblesses de l'étude

Cette étude présente plusieurs points forts. Tout d'abord, elle utilise le dernier

jeu de SNPs associés à la MA et identifiés par la plus grande méta-analyse de GWAS sur la MA à ce jour. Ensuite, les SNPs suggestifs ont été utilisés pour construire un deuxième score moins restrictif. Leurs performances ont été testées dans une large cohorte prospective de presque six mille participants et plusieurs paramètres complémentaires ont été calculés pour évaluer de manière robuste la valeur ajoutée de ces GRSs. Enfin, le schéma de l'étude respectait au mieux les critères GRIPS (Genetic Risk Prediction Studies Statement) [78] pour la rédaction d'article sur l'apport de la génétique dans la prédiction de risque.

Cette étude présente plusieurs limites. La principale relève du fait que l'étude 3C était incluse dans la méta-analyse IGAP. Même si elle ne représentait que 7 % de l'ensemble des participants inclus dans la méta-analyse, il existe un risque de surestimation des résultats. D'autre part, la présence d'une grande majorité de SNPs issus du processus d'imputation et non du génotypage lui-même peut-être responsable d'une perte de précision dans le calcul du score. Ce risque est néanmoins partiellement couvert par l'excellente qualité de l'imputation. Enfin, 35 % des participants de l'étude 3C n'ont pas pu être conservés pour l'analyse, principalement en raison de l'absence de données génétiques. Ces individus différaient légèrement des participants utilisés dans les analyses (**Annexe 10**), ce qui a pu limiter la validité externe de ce travail. En outre, la population de l'étude 3C est une population de personnes âgées initialement non institutionnalisées et globalement en meilleure santé que la population générale âgée de 65 ans et plus.

## VIII. Perspectives

Si dans ce travail les GRSs ont été construits pour faire de la prédiction de risque, ils peuvent également être utilisés pour explorer les mécanismes physiopathologiques de la MA, en recherchant par exemple des associations avec

des endophénotypes (phénotype intermédiaire entre génétique et manifestation clinique de la maladie). Dans cette optique, des études antérieures à IGAP ont mis en évidence des associations avec le volume hippocampique, l'épaisseur du cortex cérébral et le dosage d'A $\beta_{1-42}$  et de tau dans le LCR [75,76] (**Annexe 3**). Il serait donc intéressant de poursuivre ces recherches en utilisant les SNPs d'IGAP, un travail déjà commencé par Chauhan et al sur les traits IRM de la MA [77].

D'autre part, plusieurs éléments amènent à penser que les performances d'un GRS tel qu'il est construit dans ce travail pourraient être améliorées. D'une part, les variants inclus dans le score pourraient ne pas être les variants causaux mais des variants en déséquilibre de liaison avec un variant causal. Les analyses de séquençage de l'ADN pourraient permettre l'identification plus fine de ces variants génétiques causaux. D'autre part, l'incorporation d'interactions gène-gène ou gène-environnement identifiées représenterait également un levier d'amélioration des performances d'un GRS.

Enfin, il pourrait être intéressant de comparer l'apport des facteurs de susceptibilité génétique pour la prédiction de risque à celui d'une variable « antécédent familial de MA ». Cette analyse n'a malheureusement pas pu être réalisée dans 3C faute de données suffisantes.

## CONCLUSION

Dans le cadre du développement d'outils pour prédire le risque de MA et prévenir son développement, deux GRSs basés sur les résultats de la méta-analyse IGAP ont été créés. Ces derniers étaient significativement associés avec la MA. En revanche, ils n'amélioraient pas suffisamment la prédiction du risque pour conclure à une réelle valeur ajoutée par rapport au risque déjà obtenu avec les principaux facteurs de risque connus, dont l'allèle APOE- $\epsilon$ 4. Ainsi, si les loci identifiés par les GWAS ont ouvert de nouveaux axes de recherche pour l'exploration des mécanismes impliqués dans la MA, ils n'ont pas encore leur place au sein d'équations de risque.



## RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Prince M, Jackson J and Alzheimer's Disease International . World Alzheimer Report 2009: The Global Prevalence of Dementia. {Alzheimer's Disease International}, 2009.
- [2] World Health Organization and others . The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization, 1992.
- [3] World Health Organization and Alzheimer's Disease International . Dementia: A Public Health Priority. World Health Organization, 2012.
- [4] Institut national de la santé et de la recherche médicale . Maladie d'Alzheimer: enjeux scientifiques, médicaux et sociétaux. Inserm, 2007.
- [5] International A. D. World Alzheimer Reports 2013: The Global Impact of Dementia 2013-2050. 2013.
- [6] McKhann G, Drachman D, Folstein M, Katzman R, Price D and Stadlan E. M. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939-944.
- [7] McKhann G. M, Knopman D. S, Chertkow H, Hyman B. T, Jack C. R, Kawas C. H *et al*. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263-269.
- [8] Hardy J. A and Higgins G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science (80- )*. 1992;256(5054):184-185.
- [9] Lautenschlager N. T, Cupples L. A, Rao V. S, Auerbach S. A, Becker R, Burke J *et al*. Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: What is in store for the oldest old? *Neurology*. 1996;46(3):641-650.
- [10] Gatz M, Reynolds C. A, Fratiglioni L, Johansson B, Mortimer J. A, Berg S *et al*. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168-174.
- [11] Liu C.-C, Liu C.-C, Kanekiyo T, Xu H and Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9(2):106-118.
- [12] Farrer L. A, Cupples L. A, Haines J. L, Hyman B, Kukull W. A, Mayeux R *et al*. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA*. 1997;278(16):1349-1356.
- [13] Saunders A. M, Schmechel K, Breitner J. C, Benson M. D, Brown W. T, Goldfarb L *et al*. Apolipoprotein E epsilon 4 allele distributions in late-onset Alzheimer's disease and in other amyloid-forming diseases. *Lancet*. 1993;342(8873):710-711.
- [14] Strittmatter W. J, Saunders A. M, Schmechel D, Pericak-Vance M, Enghild J, Salvesen G. S *et al*. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A*. 1993;90(5):1977-1981.
- [15] Bertram L, McQueen M. B, Mullin K, Blacker D and Tanzi R. E. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet*. 2007;39(1):17-23.
- [16] Bettens K, Sleegers K and Van Broeckhoven C. Genetic insights in Alzheimer's disease. *Lancet Neurol*. 2013;12(1):92-104.
- [17] Corder E. H, Saunders A. M, Risch N. J, Strittmatter W. J, Schmechel D. E, Gaskell P *et al*. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet*. 1994;7(2):180-184.
- [18] Royston M. C, Mann D, Pickering-Brown S, Owen F, Perry R, Raghavan R *et al*. Apolipoprotein E epsilon 2 allele promotes longevity and protects patients with Down's syndrome from dementia. *Neuroreport*. 1994;5(18):2583-2585.
- [19] Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O *et al*. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry*. 2011;16(9):903-907.
- [20] Mahley R. W. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science*. 1988;240(4852):622-630.
- [21] Lambert J.-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M *et al*. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1094-1099.

- [22] Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere M. L *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet.* 2009;41(10):1088-1093.
- [23] Seshadri S, Fitzpatrick A. L, Ikram M. A, DeStefano A. L, Gudnason V, Boada M *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA.* 2010;303(18):1832-1840.
- [24] Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J.-C, Carrasquillo M. M *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet.* 2011;43(5):429-435.
- [25] Naj A. C, Jun G, Beecham G. W, Wang L.-S, Vardarajan B. N, Buross J *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet.* 2011;43(5):436-441.
- [26] Lambert J. C, Ibrahim-Verbaas C. A, Harold D, Naj A. C, Sims R, Bellenguez C *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45(12):1452-1458.
- [27] Ziegler-Graham K, Brookmeyer R, Johnson E and Arrighi H. M. Worldwide variation in the doubling time of Alzheimer's disease incidence rates. *Alzheimers Dement.* 2008;4(5):316-323.
- [28] Wilson P. W, D'Agostino R. B, Levy D, Belanger A. M, Silbershatz H and Kannel W. B. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998;97(18):1837-1847.
- [29] D'Agostino R. B, Vasan R. S, Pencina M. J, Wolf P. A, Cobain M, Massaro J. M *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation.* 2008;117(6):743-753.
- [30] Norton S, Matthews F. E, Barnes D. E, Yaffe K and Brayne C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol.* 2014;13(8):788-794.
- [31] Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H and Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol.* 2006;5(9):735-741.
- [32] Barnes D. E, Covinsky K. E, Whitmer R. A, Kuller L. H, Lopez O. L and Yaffe K. Predicting risk of dementia in older adults: The late-life dementia risk index. *Neurology.* 2009;73(3):173-179.
- [33] Reitz C, Tang M.-X, Schupf N, Manly J. J, Mayeux R and Luchsinger J. A. A summary risk score for the prediction of Alzheimer disease in elderly persons. *Arch Neurol.* 2010;67(7):835-841.
- [34] de Bruijn R. F.A. G and Ikram M. A. Cardiovascular risk factors and future risk of Alzheimer's disease. *BMC Med.* 2014;12:130.
- [35] Ohara T, Ninomiya T, Kubo M, Hirakawa Y, Doi Y, Hata J *et al.* Apolipoprotein genotype for prediction of Alzheimer's disease in older Japanese: the Hisayama Study. *J Am Geriatr Soc.* 2011;59(6):1074-1079.
- [36] Meigs J. B, Shrader P, Sullivan L. M, McAteer J. B, Fox C. S, Dupuis J *et al.* Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med.* 2008;359(21):2208-2219.
- [37] Jager P. L. D, Chibnik L. B, Cui J, Reischl J, Lehr S, Simon K. C *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol.* 2009;8(12):1111-1119.
- [38] Paynter N. P, Chasman D. I, Paré G, Buring J. E, Cook N. R, Miletich J. P *et al.* Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA.* 2010;303(7):631-637.
- [39] Everett B. M, Cook N. R, Conen D, Chasman D. I, Ridker P. M and Albert C. M. Novel genetic markers improve measures of atrial fibrillation risk prediction. *Eur Heart J.* 2013;34(29):2243-2251.
- [40] Ebbert M. T. W, Ridge P. G, Wilson A. R, Sharp A. R, Bailey M, Norton M. C *et al.* Population-based Analysis of Alzheimer's Disease Risk Alleles Implicates Genetic Interactions. *Biol Psychiatry.* 2013.
- [41] Verhaaren B. F. J, Vernooij M. W, Koudstaal P. J, Uitterlinden A. G, van Duijn C. M, Hofman A *et al.* Alzheimer's disease genes and cognition in the nondemented general population. *Biol Psychiatry.* 2013;73(5):429-434.
- [42] Carrasquillo M. M, Crook J. E, Pedraza O, Thomas C. S, Pankratz V. S, Allen M *et al.* Late-onset Alzheimer's risk variants in memory decline, incident mild cognitive impairment, and Alzheimer's disease. *Neurobiol Aging.* 2015;36(1):60-67.
- [43] 3C. Study Group . Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology.* 2003;22(6):316-325.
- [44] Delaneau O, Zagury J.-F and Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 2013;10(1):5-6.
- [45] Howie B. N, Donnelly P and Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.
- [46] Dufouil C, Richard F, Fiévet N, Dartigues J. F, Ritchie K, Tzourio C *et al.* APOE genotype, cholesterol level, lipid-lowering treatment, and dementia: the Three-City Study *Neurology.* 2005;64(9):1531-8.

- [47] Harrell F. E, Lee K. L and Mark D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4):361-387.
- [48] Pencina M. J and D'Agostino R. B. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med.* 2004;23(13):2109-2123.
- [49] Chambless L. E and Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med.* 2006;25(20):3474-3486.
- [50] Uno H, Cai T, Pencina M. J, D'Agostino R. B and Wei L. J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* 2011;30(10):1105-1117.
- [51] Pencina M. J, D'Agostino R. B, D'Agostino R. B and Vasan R. S. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008;27(2):157-72; discussion 207-12.
- [52] Pepe M. S, Feng Z and Gu J. W. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., *Statistics in Medicine* (DOI: 10.1002/sim.2929). *Stat Med.* 2008;27(2):173-181.
- [53] Pencina M. J, D'Agostino R. B and Steyerberg E. W. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011;30(1):11-21.
- [54] Uno H, Tian L, Cai T, Kohane I. S and Wei L. J. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med.* 2013;32(14):2430-2442.
- [55] R Core Team . R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2013.
- [56] Rodríguez-Rodríguez E, Sánchez-Juan P, Vázquez-Higuera J. L, Mateo I, Pozueta A, Berciano J *et al.* Genetic risk score predicting accelerated progression from mild cognitive impairment to Alzheimer's disease. *J Neural Transm.* 2012.
- [57] Che R and Motsinger-Reif A. A. A new explained-variance based genetic risk score for predictive modeling of disease risk. *Stat Appl Genet Mol Biol.* 2012;11(4):Article 15.
- [58] Raynor L. A, Pankow J. S, Duncan B. B, Schmidt M. I, Hoogeveen R. C, Pereira M. A *et al.* Novel risk factors and the prediction of type 2 diabetes in the Atherosclerosis Risk in Communities (ARIC) study. *Diabetes Care.* 2013;36(1):70-76.
- [59] Ripatti S, Tikkanen E, Orho-Melander M, Havulinna A. S, Silander K, Sharma A *et al.* A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet.* 2010;376(9750):1393-1400.
- [60] Brautbar A, Pompeii L. A, Dehghan A, Ngwa J. S, Nambi V, Virani S. S *et al.* A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis.* 2012;223(2):421-426.
- [61] Hughes M. F, Saarela O, Stritzke J, Kee F, Silander K, Klopp N *et al.* Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. *PLoS ONE.* 2012;7(7):e40922.
- [62] Thanassoulis G, Peloso G. M, Pencina M. J, Hoffmann U, Fox C. S, Cupples L. A *et al.* A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circ Cardiovasc Genet.* 2012;5(1):113-121.
- [63] O'Seaghda C. M, Yang Q, Wu H, Hwang S.-J and Fox C. S. Performance of a genetic risk score for CKD stage 3 in the general population. *Am J Kidney Dis.* 2012;59(1):19-24.
- [64] Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med.* 2013;368(2):117-127.
- [65] Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson P. V, Snaedal J *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med.* 2013;368(2):107-116.
- [66] IGAP . Follow-up of loci from the International Genomics of Alzheimer's Disease Project identifies TRIP4 as a novel susceptibility gene. *Transl Psychiatry.* 2014;4:e358.
- [67] Haute Autorité de Santé . Recommandation de bonne pratique : Maladie d'Alzheimer et maladies apparentées : diagnostic et prise en charge. decembre-2011.
- [68] Cook N. R. Clinically relevant measures of fit? A note of caution. *Am J Epidemiol.* 2012;176(6):488-491.
- [69] Pencina M. J, D'Agostino R. B, Demler O. V, Janssens A. C.J. W and Greenland P. Pencina et al. respond to "The incremental value of new markers" and "Clinically relevant measures? A note of caution". *Am J Epidemiol.* 2012;176(6):492-494.
- [70] Pencina M. J, D'Agostino R. B and Demler O. V. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. *Stat Med.* 2012;31(2):101-113.
- [71] Pencina M. J, D'Agostino R. B, Pencina K. M, Janssens A. C.J. W and Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* 2012;176(6):473-481.

- [72] Steyerberg E. W, Pencina M. J, Lingsma H. F, Kattan M. W, Vickers A. J and Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest*. 2012;42(2):216-228.
- [73] D'Agostino R, Griffith J, Schmid C and Terrin N. Measures for evaluating model performance . 1997:253-258.
- [74] Cook N. Use and misuse of the receiver operating characteristic curve in risk prediction *Circulation*. 2007;115(7):928-935.
- [75] Biffi A, Anderson C. D, Desikan R. S, Sabuncu M, Cortellini L, Schmansky N *et al*. Genetic variation and neuroimaging measures in Alzheimer disease. *Arch Neurol*. 2010;67(6):677-685.
- [76] Sabuncu M. R, Buckner R. L, Smoller J. W, Lee P. H, Fischl B, Sperling R. A *et al*. The association between a polygenic Alzheimer score and cortical thickness in clinically normal subjects. *Cereb Cortex*. 2012;22(11):2653-2661.
- [77] Chauhan G, Adams H. H. H, Bis J. C, Weinstein G, Yu L, Töglhofer A. M *et al*. Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. *Neurobiol Aging*. 2015.
- [78] Janssens A. C.J. W, Ioannidis J. P. A, Bedrosian S, Boffetta P, Dolan S. M, Dowling N *et al*. Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration. *Eur J Hum Genet*. 2011;19(5):18 p preceding 494.

## ANNEXES

### Annexe 1. Critères diagnostiques de démence selon le DSM-IV

- A. Apparition de déficits cognitifs multiples, comme en témoignent à la fois :
  - I. Une altération de la mémoire (altération de la capacité à apprendre des informations nouvelles ou à se rappeler les informations apprises antérieurement) ;
  - II. Une (ou plusieurs) des perturbations cognitives suivantes :
    - a. Aphasie (perturbation du langage)
    - b. Apraxie (altération de la capacité à réaliser une activité motrice malgré des fonctions motrices intactes)
    - c. Agnosie (impossibilité de reconnaître ou d'identifier des objets malgré des fonctions sensorielles intactes)
    - d. Perturbation des fonctions exécutives (faire des projets, organiser, ordonner dans le temps, avoir une pensée abstraite).
- B. Les déficits cognitifs des critères AI et AII sont tous les deux à l'origine d'une altération significative du fonctionnement social ou professionnel et représentent un déclin significatif par rapport au niveau de fonctionnement antérieur.
- C. Les déficits cognitifs des critères AI et AII ne sont pas dus :
  - I. A d'autres affections du système nerveux central qui peuvent entraîner des déficits progressifs de la mémoire et du fonctionnement cognitif (par exemple : maladie cérébro-vasculaire, maladie de Parkinson, maladie de Huntington, hématome sous-dural, hydrocéphalie à pression normale, tumeur cérébrale) ;
  - II. A des affections générales pouvant entraîner une démence (par exemple : hypothyroïdie, carence en vitamine B12 ou en folates, pellagre, hypercalcémie, neurosyphilis, infection par le VIH) ;
  - III. A des affections induites par une substance.
- D. Les déficits ne surviennent pas de façon exclusive au cours de l'évolution d'un délirium
- E. La perturbation n'est pas expliquée par un trouble trouble dépressif majeur ou une schizophrénie

## **Annexe 2. Critères diagnostiques de la maladie d'Alzheimer NINCDS-ADRDA**

### **Critères de maladie d'Alzheimer probable :**

- syndrome démentiel établi sur des bases cliniques et documenté par le Mini-Mental State Examination, le Blessed Dementia Scale ou tout autre test équivalent et confirmé par des preuves neuropsychologiques
- déficit d'au moins deux fonctions cognitives
- altérations progressives de la mémoire et des autres fonctions cognitives
- absence de trouble de conscience
- survenue entre 40 et 90 ans, le plus souvent au-delà de 65 ans
- en l'absence de désordres systémiques ou d'une autre maladie cérébrale pouvant rendre compte par eux-mêmes, des déficits mnésiques et cognitifs progressifs

### **Ce diagnostic de maladie d'Alzheimer probable est renforcé par :**

- la détérioration progressive des fonctions telles que le langage (aphasie), les habilités motrices (apraxie) et perceptives (agnosie)
- la perturbation des activités de la vie quotidienne et la présence de troubles du comportement
- une histoire familiale de troubles similaires surtout si confirmés histologiquement
- le résultat aux examens standards suivants :
  - normalité du liquide céphalo-rachidien
  - EEG normal ou siège de perturbations non spécifiques comme la présence d'ondes lentes
  - présence d'atrophie cérébrale d'aggravation progressive

### **Autres caractéristiques cliniques compatibles avec le diagnostic de maladie d'Alzheimer probable après exclusion d'autres causes :**

- période de plateaux au cours de l'évolution
- présence de symptômes tels que dépression, insomnie, incontinence, idées délirantes, illusions, hallucinations, réactions de catastrophe, désordres sexuels et perte de poids. Des anomalies neurologiques sont possibles surtout aux stades évolués de la maladie, notamment des signes moteurs tels qu'une hypertonie, des myoclonies ou des troubles de la marche.
- crises comitiales aux stades tardifs
- scanner cérébral normal pour l'âge

### **Signes rendant le diagnostic de maladie d'Alzheimer probable incertain ou improbable :**

- début brutal
- déficit neurologique focal tel que hémiparésie, hypoesthésie, déficit du champ visuel, incoordination motrice à un stade précoce
- crises convulsives ou troubles de la marche en tout début de maladie

### **Le diagnostic clinique de la maladie d'Alzheimer possible :**

- peut être porté sur la base du syndrome démentiel, en l'absence d'autre désordre neurologique, psychiatrique ou systémique susceptible de causer une démence, en présence de variante dans la survenue, la présentation ou le cours de la maladie;
- peut être porté en présence d'une seconde maladie systémique ou cérébrale susceptible de produire un syndrome démentiel mais qui n'est pas considérée comme la cause de cette démence ;
- et pourrait être utilisé en recherche clinique quand un déficit cognitif sévère progressif est identifié en l'absence d'autre cause identifiable.

### **Les critères pour le diagnostic de maladie d'Alzheimer certaine sont :**

- les critères cliniques de la maladie d'Alzheimer probable ;
- et la preuve histologique apportée par la biopsie ou l'autopsie.

### Annexe 3. Loci identifiés par la méta-analyse IGAP

Loci identifiés par la méta-analyse IGAP [26].

SNP	Gène le plus proche <sup>§</sup>	QI <sup>†</sup> dans 3C	CHR <sup>‡</sup>	Allèle de référence/ autre	OR IGAP (IC 95%)	P IGAP
<b>1) SNPs associés avec la MA (valeur de <math>p &lt; 5 \times 10^{-8}</math>)</b>						
rs6656401*	CR1	0.95	1	A/G	1.18 (1.14-1.22)	$5.69 \times 10^{-24}$
rs35349669*	INPP5D	0.96	2	T/C	1.08 (1.05-1.11)	$3.17 \times 10^{-08}$
rs6733839*	BIN1	0.92	2	T/C	1.22 (1.18-1.25)	$6.94 \times 10^{-44}$
rs190982*	MEF2C	0.92	5	A/G	1.08 (1.05-1.11)	$3.23 \times 10^{-08}$
rs10948363*	CD2AP	1.00	6	G/A	1.10 (1.07-1.13)	$5.20 \times 10^{-11}$
rs77372866*	HLA-DRB5/DRB1	1.00	6	C/A	1.11 (1.08-1.15)	$2.94 \times 10^{-12}$
rs11771145	EPHA1	1.00	7	G/A	1.11 (1.08-1.14)	$1.12 \times 10^{-13}$
rs1476679*	ZCWPW1	0.99	7	T/C	1.09 (1.06-1.12)	$5.58 \times 10^{-10}$
rs2718058	NME8	1.00	7	A/G	1.08 (1.05-1.11)	$4.76 \times 10^{-09}$
rs28834970*	PTK2B	0.99	8	C/T	1.10 (1.08-1.13)	$7.37 \times 10^{-14}$
rs9331896*	CLU	0.94	8	T/C	1.16 (1.13-1.19)	$2.77 \times 10^{-25}$
rs10792832*	PICALM	1.00	11	G/A	1.15 (1.12-1.18)	$9.32 \times 10^{-26}$
rs10838725*	CELF1	1.00	11	C/T	1.08 (1.05-1.11)	$1.12 \times 10^{-08}$
rs11218343*	SORL1	1.00	11	T/C	1.30 (1.22-1.39)	$9.73 \times 10^{-15}$
rs983392*	MS4A6A	0.98	11	A/G	1.11 (1.09-1.14)	$6.14 \times 10^{-16}$
rs10498633	SLC24A4/RIN3	1.00	14	G/T	1.10 (1.06-1.13)	$5.54 \times 10^{-09}$
rs17125944	FERMT2	1.00	14	C/T	1.14 (1.09-1.19)	$7.95 \times 10^{-09}$
rs4147929*	ABCA7	0.99	19	A/G	1.15 (1.11-1.19)	$1.06 \times 10^{-15}$
rs7274581*	CASS4	0.97	20	T/C	1.14 (1.09-1.20)	$2.46 \times 10^{-08}$
<b>2) SNPs suggestifs (valeur de <math>p</math> comprise entre <math>5 \times 10^{-8}</math> et <math>10^{-6}</math>)</b>						
rs6678275*	None	1.00	1	C/G	1.09 (1.05-1.13)	$3.42 \times 10^{-07}$
rs6448799*	HS3ST1	0.96	4	T/C	1.08 (1.05-1.11)	$6.61 \times 10^{-08}$
rs72807343*	SOSTM1	0.90	5	T/C	1.35 (1.20-1.52)	$7.40 \times 10^{-07}$
rs9381040*	TREML2	0.99	6	C/T	1.07 (1.04-1.10)	$6.35 \times 10^{-07}$
rs7818382*	NDUFAF6	0.99	8	T/C	1.07 (1.04-1.10)	$8.00 \times 10^{-08}$
rs7920721	ECHDC3	1.00	10	G/A	1.07 (1.04-1.10)	$2.89 \times 10^{-07}$
rs10751667*	AP2A2	0.98	11	T/A	1.07 (1.04-1.10)	$6.28 \times 10^{-07}$
rs7295246*	ADAMST20	0.97	12	G/T	1.07 (1.04-1.10)	$2.96 \times 10^{-07}$
rs2337406*	IGH@	0.99	14	C/T	1.14 (1.09-1.21)	$2.72 \times 10^{-07}$
rs74615166*	TRIP4	0.94	15	C/T	1.29 (1.17-1.42)	$4.29 \times 10^{-07}$
rs8035452*	SPPL2A	1.00	15	T/C	1.07 (1.04-1.10)	$3.18 \times 10^{-07}$
17:61538148*	ACE	0.80	17	A/G	1.34 (1.20-1.50)	$3.08 \times 10^{-07}$
rs7225151	SCIMP	1.00	17	A/G	1.10 (1.06-1.15)	$3.66 \times 10^{-07}$

\* SNP imputé dans 3C.

§ Gène situé à une proximité de  $\pm 100$  kb du SNP.

† Qualité de l'imputation rapportée par la variable Info du logiciel IMPUTE2.

‡ Chromosome.

## Annexe 4. Synthèse bibliographique

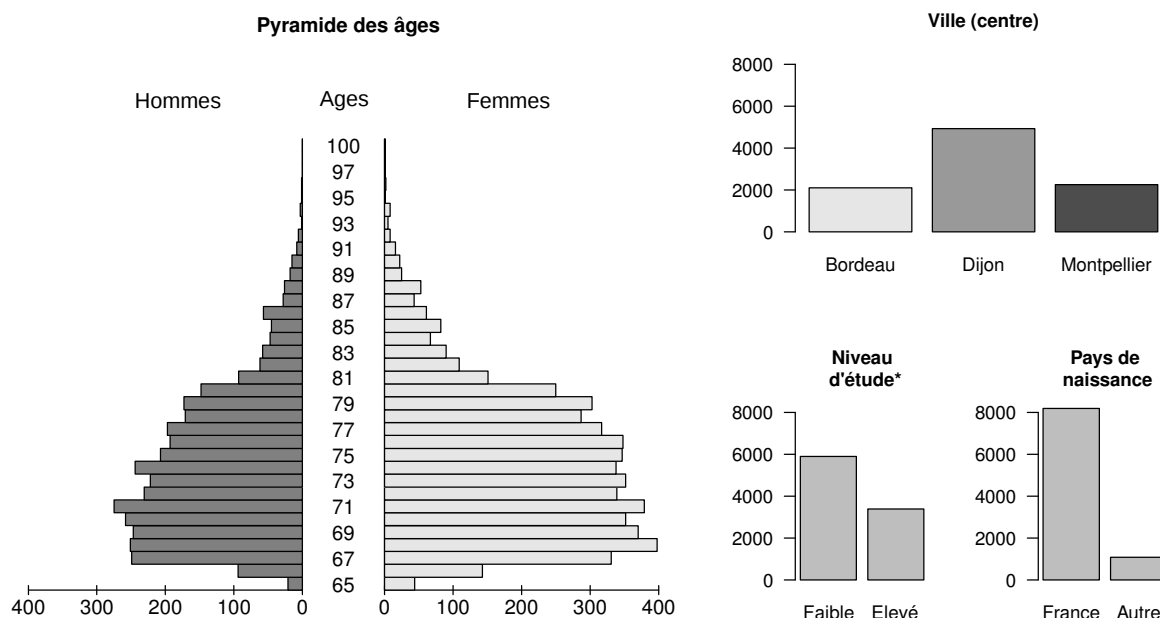
### Synthèse bibliographique

Publication	Type d'étude	Calcul du GRS	Résultats		
			Association du GRS avec la MA	Amélioration de la prédiction du risque de MA	Autres résultats
Avril 2014 [42]	Cohorte de 2674 participants avec un de suivi médian de 5 ans	9 loci [24] Somme pondérée par l'OR	Pas d'association avec la MA/MCI : HR = 1,03 (0,95-1,13) pour un allèle supplémentaire	Pas d'amélioration par rapport à un modèle de base incluant l'âge, le sexe, le niveau d'éducation et l'allèle APOE-ε4 : C-index 0,685 → 0,685	Non corrélé à une diminution des fonctions cognitives mesurées par le LMDR (The 30-minute delayed recall scores)
2013 [40]	Cohorte 2419 individus (The Cache Country Study on Memory Health and Agging)	9 loci [15] Somme pondérée par l'OR		Pas d'amélioration significative par rapport à modèle incluant l'age, le sexe et l'allèle APOE-ε4 AUC 0,783 → 0,782	
Avril 2012 [41]	Cohorte de 5507 participants avec un suivi à 10 ans (Rotterdam)	10 loci [15] Somme pondérée par l'OR	Association avec la MA : HR = 1,14 (1,03-1,27) pour 1 déviation standard	Pas d'amélioration significative par rapport à un modèle incluant l'age, le sexe et l'allèle APOE-ε4 AUC 0,8148 → 0,8159	Pas d'association avec les fonctions cognitives en dehors d'une légère association avec la mémoire
Nov 2012 [56]	Cohorte 288 participants MCI* Moyenne de suivi de 2.2 ans	8 loci [15] Somme pondérée par l'OR	Conversion de MCI* vers MA : OR = 1,32 (0,57-3,06) HR = 2.06 (1.07-3.98) pour le 3ème tertile vs 1er tertile		
Nov 2012 [76]	Cas (n=100) / témoins (n=104) (ADNI : Alzheimer's Disease Neuroimaging Initiative)	26 loci avec une $p < 10^{-5}$ de [22] Somme pondérée par l'OR	Association avec la MA : OR = 1.55 pour 1 déviation standard		- Association avec l'épaisseur du cortex cérébral - Association avec $A\beta_{1-42}$ dans le LCR - Pas d'association avec tau dans le LCR
Juin 2010 [75]	Cas (MCI=346, MA=171) / témoins (n=205) (ADNI)	5 loci [21,22]	Association avec le statut clinique (MCI* puis MA) OR = 1,14 (1,04-1,25) pour une augmentation de 1 quartile		Association avec le volume hippocampique, volume amygdalien, épaisseur du cortex entorhinal, épaisseur du cortex temporal.

\* MCI, « mild cognitive impairment », stade clinique dans lequel les troubles cognitifs peuvent être objectivés mais n'ont pas de retentissement sur les activités de la vie quotidienne.



## Annexe 5. Caractéristiques à l'inclusion des 9 294 participants de l'étude 3C

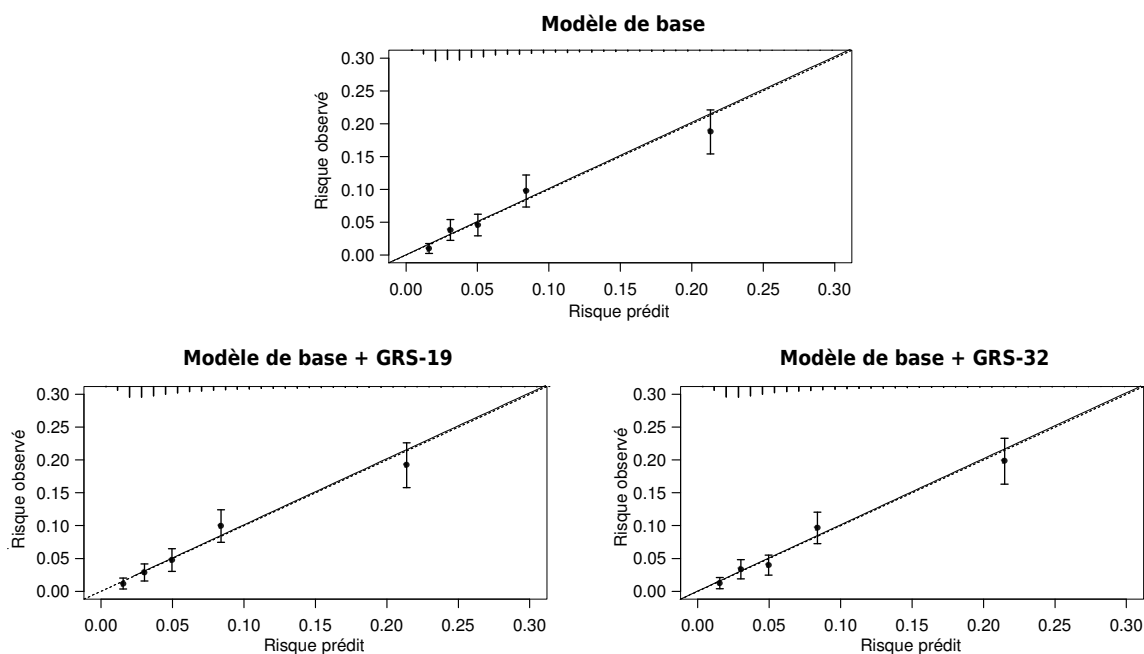


### Caractéristiques générales de la population de l'étude 3C.

\* Elevé : niveau baccalauréat ou études supérieures ; Faible : aucune étude, certificat d'études primaires, CAP, BEP ou BEPC.

La présence d'une distribution parallèle de l'âge chez les hommes et chez les femmes reflète le fait que la plupart du temps, mari et femme ont accepté de participer. Les couples constituent ainsi 2/3 des participants. Par rapport à la population générale des plus de 65 ans, on remarque une sous-représentation des femmes et des personnes très âgées (seulement 6 % de plus de 85 ans au lieu de 12 %), un niveau supérieur d'étude, des catégories socio-économiques plus élevées (90 % des femmes ont eu une activité professionnelle) [43].

## Annexe 6. Calibration des modèles



### **Calibration des modèles : risque observé vs risque prédit.**

Le risque observé est estimé par la méthode de Kaplan Meier et le risque prédit est issu des modèles de Cox.

Modèle de base : modèle incluant l'âge, le sexe, le centre, le niveau d'éducation, la langue maternelle et APOE- $\epsilon 4$ .

## Annexe 7. Amélioration de la prédiction induite par l'ajout des GRSs à un modèle de base incluant les facteurs de risque cardio-vasculaire

*Tableau 1. Amélioration de la prédiction induite par l'ajout des GRSs à un modèle de base incluant les facteurs de risque cardio-vasculaire.*

	GRS-19		GRS-32	
	Valeur	IC 95 %	Valeur	IC 95 %
<b><math>\Delta</math>C-index</b>	0,00241	(-0,00167 ; 0,00648)	<b>0,00418</b>	<b>(0,000553 ; 0,00892)</b>
Modèle de base	0,769	(0,745 ; 0,792)	0,769	(0,745 ; 0,792)
Modèle de base + GRS	0,771	(0,748 ; 0,794)	0,773	(0,752 ; 0,794)
<b>cNRI</b>	<b>0,110</b>	<b>(0,0162 ; 0,235)</b>	<b>0,203</b>	<b>(0,0899 ; 0,319)</b>
Chez les MA	0,0543	(-0,0189 ; 0,144)	0,100	(0,0249 ; 0,180)
Chez les non MA	0,0561	(0,0113 ; 0,110)	0,103	(0,0442 ; 0,157)
<b>IDI</b>	0,00199	(-0,00046 ; 0,00811)	<b>0,00374</b>	<b>(0,000259 ; 0,0103)</b>

*Modèle de base : modèle incluant l'âge, le sexe, le centre, le niveau d'éducation, la langue maternelle, APOE- $\epsilon$ 4, le tabac, l'hypertension, l'hypercholestérolémie, le diabète et les ATCD cardio-vasculaires.*

**Tableau 2. Amélioration de la prédiction induite par l'ajout des GRSs par rapport à un modèle de base incluant les facteurs de risque cardio-vasculaire, mesurée séparément chez les porteurs et les non porteurs d'APOE-ε4.**

	GRS-19				GRS-32			
	Non porteurs d'APOE-ε4		Porteurs d'APOE-ε4		Non porteurs d'APOE-ε4		Porteurs d'APOE-ε4	
	Valeur	IC 95 %	Valeur	IC 95 %	Valeur	IC 95 %	Valeur	IC 95 %
<b>ΔC-index</b>	0,00181	(-0,00350 ; 0,00713)	0,00227	(-0,00461 ; 0,00915)	0,00410	(-0,00109 ; 0,00928)	0,00453	(-0,00483 ; 0,0139)
Modèle de base	0,778	(0,755 ; 0,801)	0,740	(0,696 ; 0,784)	0,778	(0,755 ; 0,801)	0,740	(0,696 ; 0,784)
Modèle de base + GRS	0,780	(0,757 ; 0,802)	0,742	(0,694 ; 0,791)	0,782	(0,757 ; 0,807)	0,745	(0,701 ; 0,788)
<b>cNRI</b>	0,102	(-0,0174 ; 0,247)	0,105	(-0,0832 ; 0,372)	<b>0,189</b>	<b>(0,0636 ; 0,319)</b>	<b>0,188</b>	<b>(0,0507 ; 0,401)</b>
Chez les MA	0,0486	(-0,0320 ; 0,156)	0,0116	(-0,0919 ; 0,207)	0,0966	(0,0147 ; 0,195)	0,0986	(-0,0854 ; 0,252)
Chez les non MA	0,0536	(-0,0124 ; 0,119)	0,0933	(-0,0603 ; 0,203)	0,0920	(0,0248 ; 0,159)	0,0895	(-0,0121 ; 0,213)
<b>IDI</b>	0,00294	(-0,000475 ; 0,0104)	0,000664	(-0,00157 ; 0,0119)	<b>0,00390</b>	<b>(0,000162 ; 0,0134)</b>	0,00344	(-0,000430 ; 0,0187)

*Modèle de base : modèle incluant l'âge, le sexe, le centre, le niveau d'éducation, la langue maternelle, le tabac, l'hypertension, l'hypercholestérolémie, le diabète et les ATCD cardio-vasculaires.*

## Annexe 8. Méthodes de pondération d'un GRS utilisant la dose allélique

*Méthodes de calcul du score de risque génétique en fonction de la pondération accordée aux SNPs.*

Méthode	Description	Pondération : $w_i$	Sélection des SNPs	Avantages	Inconvénients
<b>Comptage simple</b>	Somme du nombre d'allèles de chaque SNP	$w_i=1$	SNPs identifiés par la littérature (GWAS par exemple)	<ul style="list-style-type: none"> <li>• <b>Simplicité</b></li> <li>• Ne nécessite que l'identité des SNPs à inclure dans le GRS</li> </ul>	<ul style="list-style-type: none"> <li>• Fait l'assomption forte d'une <b>égale contribution des SNPs au GRS</b></li> </ul>
<b>Somme pondérée par l'OR</b>	Somme du nombre d'allèles de chaque SNP pondérée par sa taille d'effet estimée dans la littérature	$w_i=\beta_i=\log(OR_i)$ avec $OR_i$ issus de la littérature	Différents seuils de valeur de p peuvent être utilisés pour une sélection plus ou moins large	<ul style="list-style-type: none"> <li>• Relativement simple</li> <li>• Prend en compte <b>la taille d'effet des SNPs. Ces dernières peuvent être issues de large méta-analyses</b></li> <li>• Ne nécessite pas de population dédiée à l'estimation des <math>w_i</math></li> </ul>	<ul style="list-style-type: none"> <li>• Nécessite les ORs des SNPs</li> <li>• Les tailles d'effets des SNPs ne sont pas ajustées sur les autres SNPs ; <b>hypothèse d'additivité des tailles d'effet</b></li> </ul>
<b>Modèle de régression dédiée</b>	Construction d'un modèle de régression contenant les SNPs, dédié à l'estimation des coefficients	$w_i=\beta_i$ avec $\beta_i$ issus d'un modèle de régression dédié	SNPs identifiés par la littérature (GWAS par exemple)  Des méthodes de sélection de variables de type stepwise ou LASSO peuvent être employées pour affiner le modèle	<ul style="list-style-type: none"> <li>• Prend en compte la taille d'effet des SNPs</li> <li>• Les <b>tailles d'effets sont ajustées sur les autres SNPs</b></li> <li>• Ne nécessite pas d'information préalable</li> <li>• Des <b>méthodes de sélection</b> de SNPs permettent d'affiner le GRS</li> <li>• Des <b>interactions</b> entre SNPs peuvent éventuellement y être intégrées</li> </ul>	<ul style="list-style-type: none"> <li>• Nécessite une <b>population indépendante dédiée à la construction du modèle</b></li> <li>• Risque important de <b>surajustement</b>, notamment si la population utilisée est petite ou si les SNPs sont nombreux ; <b>risque de non reproductibilité des résultats</b></li> </ul>

## Annexe 9. Valeurs de référence pour l'interprétation des mesures d'amélioration de la prédiction

*Valeurs théoriques des mesures d'amélioration de la prédiction en fonction de la taille d'effet du nouveau prédicteur.*

	Taille d'effet FAIBLE	Taille d'effet MOYENNE	Taille d'effet ÉLEVÉ
<b>Apport d'un nouveau prédicteur de distribution Normale à un modèle de base</b>			
<b><math>\Delta</math>AUC/ <math>\Delta</math>C-index</b>	0,005	0,027	0,045
<b>IDI</b>	0,004	0,028	0,052
<b>cNRI</b>	0,160	0,395	0,622
<b>Apport d'un nouveau prédicteur binaire à un modèle de base</b>			
<b><math>\Delta</math>AUC/ <math>\Delta</math>C-index</b>	0,002	0,014	0,037
<b>IDI</b>	0,002	0,016	0,046
<b>cNRI</b>	0,104	0,296	0,523

*Données issues de simulations réalisées par Pencina et al [71], pour un nouveau prédicteur continue de distribution Normale, un nouveau prédicteur binaire, une incidence de la maladie de 10 % et un AUC/C-index de base de 0,75.*

## Annexe 10. Comparaison entre les participants inclus et exclus de l'analyse

### Comparaison des caractéristiques des participants inclus et exclus de l'analyse.

	Population à risque N = 9080	Non inclus dans l'analyse N = 3158		Inclus dans l'analyse N = 5922	P *
		valeur	valide		
<b>Age</b> (année)	74,2 ± 5,5	74,2 ± 5,7	3158	74,2 ± 5,5	0,8
<b>Femme</b>	5526 (60,9%)	1928 (61,1%)	3158	3598 (60,8%)	0,8
<b>Centre</b>			3158		< 0,001
Bordeaux	2027 (22,3%)	586 (18,6%)		1441 (24,3%)	
Dijon	4863 (53,6%)	996 (31,5%)		3867 (65,3%)	
Montpellier	2190 (24,1%)	1576 (49,9%)		614 (10,4%)	
<b>Niveau d'éducation</b>			3152		0,15
Bas	2989 (32,9%)	991 (31,4%)		1998 (33,7%)	
Moyen-bas	1460 (16,1%)	523 (16,6%)		937 (15,8%)	
Moyen-haut	2477 (27,3%)	885 (28,0%)		1592 (26,9%)	
Haut	2148 (23,7%)	753 (23,8%)		1395 (23,6%)	
<b>Langue maternelle : français</b>	8756 (96,4%)	2980 (94,4%)	3118	5776 (97,5%)	< 0,001
<b>Diabète</b>	874 (9,6%)	308 (11,7%)	2636	566 (9,6%)	0,002
<b>Hypertension</b>	7029 (77,4%)	2376 (75,2%)	3151	4653 (78,6%)	< 0,001
<b>Hypercholestérolémie</b>	4970 (54,7%)	1583 (57,2%)	2766	3387 (57,2%)	0,9
<b>Tabagisme actuel</b>	510 (5,6%)	185 (5,9%)	3154	325 (5,5%)	0,5
<b>ATCD de maladie CV</b>	856 (9,4%)	295 (9,3%)	3151	561 (9,5%)	0,9
<b>Porteur d'APOE-ε4</b>	1716 (20,2%)	506 (19,7%)	2571	1210 (20,4)	0,4
<b>GRS-19</b>	19,6 ± 2,6	19,5 ± 2,6	117	19,6 ± 2,6	0,7
<b>GRS-32</b>	25,8 ± 2,9	25,8 ± 2,9	117	25,8 ± 2,9	0,9
<b>MA incidentes</b>	497 (7,3%)	156 (6,0%)	2158	342 (6,2%)	0,6

Présente la moyenne ± écart-type des variables quantitatives et le nombre d'individus (%) des variables binaires ou catégorielles.

\*valeur de p issue de la comparaison entre inclus et exclus, estimée par un test de Student ou un test du Chi2.

**AUTEUR : Nom :** MAURY

**Prénom :** Fleur

**Date de Soutenance :** 24/03/2015

**Titre de la Thèse :** Évaluation de l'apport de deux scores de risque génétique pour la prédiction du risque de maladie d'Alzheimer dans l'étude des 3 Cités

**Thèse - Médecine - Lille 2015**

**Cadre de classement :** Santé publique

**Mots-clés :** Score de risque génétique, démence, Maladie d'Alzheimer, prédiction de risque, discrimination, épidémiologie.

**Contexte :** Pour lutter contre le fardeau croissant que représente la maladie d'Alzheimer (MA) pour la société, une stratégie consiste à agir tôt, en amont des manifestations cliniques, et à construire des outils simples et performants pour prédire le risque de développer la maladie. L'identification récente de plusieurs nouveaux facteurs de susceptibilité génétique, en plus du gène de l'apolipoprotéine E (APOE), représente une piste pour le développement de ces outils. L'objectif de ce travail était donc d'évaluer l'amélioration de la prédiction du risque de MA lors de l'ajout de ces facteurs de susceptibilité génétique à un modèle comprenant les principaux facteurs de risque de MA connus.

**Méthode :** L'analyse a été réalisée sur 5 922 participants âgés de 65 ans et plus de l'étude des 3 Cités. A partir des résultats de la dernière méta-analyse d'étude d'association pangénomique (GWAS), un score de risque génétique (GRS) composé des 19 loci associés à la MA (GRS-19), et un GRS comprenant également 13 loci supplémentaires (GRS-32) ont été calculés. L'association avec le risque de MA a été estimée par l'Hazard Ratio (HR) issu d'un modèle de Cox. L'amélioration de la prédiction induite par l'ajout d'un GRS à un modèle prédictif comprenant l'âge, le sexe, le niveau d'éducation, l'allèle APOE-ε4, le centre et la langue maternelle, a été évaluée à l'aide de la différence de C-index ( $\Delta$ C-index), du cNRI (continuous net reclassification improvement) et de l'IDI (integrated discrimination improvement).

**Résultats :** Après un suivi médian de 7,6 ans, 459 participants ont développé une MA. Les deux GRSs étaient significativement associés à la MA ( $HR_{GRS-19}$  [intervalle de confiance à 95%] = 1,05 [1,01 ; 1,09] et  $HR_{GRS-32}$  = 1,06 [1,03 ; 1,09], pour un allèle à risque supplémentaire). Cette association se traduisait par une amélioration faible et non significative de la prédiction du risque de développer une MA :  $\Delta$ C-index = 0,00263 [-0,00117 ; 0,00642], cNRI = 0,0941 [-0,00176 ; 0,216] et IDI = 0,00186 [-0,000274 ; 0,00686]. Les résultats semblaient néanmoins meilleurs avec le GRS-32.

**Conclusion :** Si les facteurs de susceptibilité génétique récemment identifiés par les GWAS ont ouvert de nouveaux axes de recherche pour l'exploration des mécanismes impliqués dans la MA, leur apport conjoint pour la prédiction de risque est encore insuffisant pour pouvoir les intégrer au sein d'une équation de risque.

**Composition du Jury :**

**Président :**

**Monsieur le Professeur Jean-Louis Salomez**

**Asseseurs :**

**Monsieur le Professeur Philippe Amouyel**

**Monsieur le Professeur Régis Beuscart**

**Madame le Docteur Florence Richard**