Exploring sub-phenotype variation in Alzheimer's disease

Ioanna Kyriaki Katzourou

2022

Cardiff University

Supervisors:

Valentina Escott-Price

Peter Holmans

Ganna Leonenko

Julie Williams

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

The main aim of this thesis is to explore the presence of distinct sub-phenotypes within Alzheimer's disease (AD) and to attempt to identify genetic and environmental factors that may influence the occurrence of these sub-phenotypes. Firstly, the rate of cognitive decline in AD was selected as a phenotypic domain of interest and examined. Secondly, the co-occurrence of behavioural and psychological symptoms in AD was investigated. A large phenotypic and genetic database including diverse phenotypic information on 4,163 individuals was used in this thesis, hereby termed the Cardiff Genetic Resource for AD (CAGRAD). A subset of this dataset that had been assessed longitudinally was used to derive a metric of cognitive decline in AD using mixed effects linear modelling. The metric derived was then used in various genetic studies, including a candidate gene study, a genome-wide association study, a pathway analysis and a polygenic risk score (PRS) analysis. The results of these analyses were replicated in an independent dataset. No genetic variants were found to be associated with the metric of cognitive decline, with the findings suggesting that the genetic risk for faster decline within AD is unrelated to the genetic background of AD. Subsequently, a principal component analysis of psychiatric symptoms of AD was performed to identify sub-phenotypes of co-occurring symptoms, which were then used in a PRS analysis, aiming to identify shared genetic architecture between them and common psychiatric conditions. Five psychiatric sub-phenotypes in AD were identified and were found to share genetic aetiology with psychiatric conditions. These results can help deconvolute the heterogeneity observed in AD symptomology and shed some light into aspects of AD that are often overlooked.

Acknowledgements

There are some people I am eternally grateful for, without whom this thesis might have never been written, or at the very least looked entirely different. First of all, my supervisors, Valentina and Peter, for the guidance and support, and their genuine enthusiasm for science. You are forces to be reckoned with. The Wellcome Trust, for supporting me throughout the completion of this PhD. Ganna, Beccy and the AD field team for their help in my endless data queries. Effie, for the trips and the coffees and the endless food and the software help and pretty much everything and also Valentina, for employing Effie and giving me the best work friend I could ever have asked for. Sam, for the sick tunes and the sharing of all that immense existential angst that I don't know how I could have finished this without. Chiara, for the hugs and the dancing and the love and for being a lovely human and a scientist to always look up to. Sophie, for sharing cats, frustration and a house during lockdown, I know it wasn't easy. Greg, for only requiring one glass of whiskey to proofread my manuscript. The Integrative Neuroscience programme directors, John, Frank, and Anthony, for believing in me and giving me this wonderful opportunity, and for making it so that I got to meet so many amazing people along the way to share our PhD highs and lows. Rod, for the snacks hidden in my freezer and discovered in hours of need. Zoe and Marios, for at least pretending to care about what this whole document is all about. My mother, father and brother, for giving me the space to do my own thing, even if they think it's not the best idea. My grandmother, for always thinking it is the best idea, and for making me who I am. My grandfather, for teaching me everything he knew, and for inspiring me to research the disease that killed him. Finally, Pakora, who is not a person, for being my partner in writing for the past months and purring along to my frustration.

Statement of others' work

Figures 1.1 and 1.4 were originally published by Alzheimer, A., reprinted with permission. Figure 1.3 was originally published by Sims, R., Hill, M. & Williams, J.,, reprinted with permission. Figure 2.5 was originally published at the ADNI website, reprinted with permission. The phenotypic data collection was performed by the AD field team at the MRC CNGG at Cardiff University. The genotyping of the DNA samples of the CAGRAD cohort was performed by research staff at the MRC CNGG at Cardiff University. The imputation of the genotypes of the CAGRAD cohort described in Chapter 2 was performed by Dr Ganna Leonenko. The genetic locations used for gene annotation in the pathway analysis described in Chapter 4.3.3 were provided by Dr Emily Baker. The summary statistics of the genome-wide association study published by Kunkle *et al.* without inclusion of GERAD that were used in the polygenic risk score analysis described in Chapter 4.3.5.2 were provided by Dr Ganna Leonenko.

Data used in the preparation of this thesis (Chapters 2, 3 and 4) were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at:
http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Authorship_List.pdf)

Abbreviations

Term	Definition
AD	Alzheimer's disease
BPSD	Behavioural and psychological symptoms of dementia
GWAS	Genome-wide association study
SNP	Single nucleotide polymorphism
PCA	Principal component analysis
PRS	Polygenic risk score
EOAD	Early-onset Alzheimer's disease
LOAD	Late-onset Alzheimer's disease
Aβ	Amyloid beta
MMSE	Mini Mental State Examination
AAO	Age at disease onset
NPI	Neuropsychiatric Inventory
SNP	Single nucleotide polymorphism
MDD	Major depressive disorder
BD	Bipolar disorder
NFT	Neurofibrillary tangle
β	Beta coefficient
ADNI	Alzheimer's disease neuroimaging initiative
CAGRAD	Cardiff genetic resource for Alzheimer's disease

Table of contents

Abstract ii
Acknowledgements iii
Statement of others' work iv
Abbreviations v
Table of contents vi
List of tables x
List of figures xii
Chapter 1 | Background 1
1.1. Introduction 1
1.2. Epidemiology 2
1.2.1. Prevalence and impact 2
1.2.2. Risk factors 3
1.2.3. Protective factors 5
1.3. Genetics 6
1.3.1. Autosomal dominant variants 7
1.3.2. APOE 7
1.3.3. Other variants 8
1.4. Pathology 9
1.5. Pathogenetic mechanism 11
1.6. Clinical presentation 12
1.6.1. Cognitive decline 12
1.6.2. BPSD 13
1.7. Diagnosis 15
1.8. Biomarkers 16
1.9. Treatment 19
1.10. Aims and objectives 21
Chapter 2 | Data description 23

2.1. Cardiff Genetic Resource for AD 23
2.1.1. Overview and aims 23
2.1.2 Study design and recruitment 23
2.1.3. Genetic data 25
2.1.4 Data harmonisation 25
2.1.5. Longitudinal data 27
2.1.6. Behavioural and psychiatric phenotypes 28
2.2. Alzheimer's Disease Neuroimaging Initiative 32
2.2.1. Overview and aims 32
2.2.2. Study design and recruitment 32
2.2.3. Genetic data 33
2.2.4. Data harmonisation 34
2.3. Cohort comparison 35
Chapter 3 | Modelling cognitive decline 37
3.1. Introduction 37
3.2. Aims and objectives 38
3.3. Methods 38
3.3.1. Mini Mental Score Examination 38
3.3.2. Computation of measures of cognitive decline 39
3.3.3. Mixed effect linear modelling 39
3.3.4. Cohort comparison 41
3.3.5. Replication 41
3.4. Results 42
3.4.1. Computation of measures of cognitive decline 42
3.4.2. Cohort comparison 46
3.4.3. Replication 47
3.5. Discussion 50
3.6. Conclusions 52
Chapter 4 | Genetics of cognitive decline 54
4.1. Introduction 54
4.2. Aims and objectives 55
4.3. Methods 55
4.3.1. Datasets 55
4.3.2. APOE 56

4.3.3. GWAS of decline 56
4.3.3.1. CAGRAD 56
4.3.3.2. ADNI 57
4.3.3.3 GWAS meta-analysis 57
4.3.4. Pathway analysis of decline 58
4.3.5. PRS analysis 59
4.3.5.1. PRS of decline 60
4.3.5.2. PRS of AD 61
4.4. Results 62
4.4.1. APOE 62
4.4.2. GWAS 64
4.4.2.1. CAGRAD 64
4.4.2.2. ADNI 66
4.4.2.3. Meta-analysis 69
4.4.3. Pathway analysis 71
4.4.4. PRS 75
4.4.4.1. PRS of decline 75
4.4.4.2. AD PRS 75
4.5. Discussion 76
4.6. Conclusions 79
Chapter 5 | Behavioural and psychological symptoms of dementia 80
5.1. Introduction 80
5.2. Aims and objectives 81
5.3. Methods 82
5.3.1. Dataset 82
5.3.2. Neuropsychiatric inventory 82
5.3.3. Cohort comparison 83
5.3.4. Principal component analysis 83
5.3.5. Contributing factors 85
5.3.6. PRS analysis 86
5.4. Results 87
5.4.1. Cohort comparison 87
5.4.2 Principal component analysis 88
5.4.3. Contributing factors 94

5.4.4. PRS analysis 98
5.5. Discussion 103
5.6. Conclusions 108
Chapter 6 | General Discussion 109
6.1. Overview 109
6.2. How can cognitive decline in AD be quantified, and does it have a genetic background? 109
6.3. What influences the appearance of behavioural sub-phenotypes in AD? 114
6.4. Considerations and future directions 121
6.5. Conclusion 122
Bibliography 123
Supplementary Material 157

List of tables

Table 2.1. Cross-validation. 26
Table 2.2. CAGRAD characteristics. 27
Table 2.3. CAGRAD missing values. 29
Table 2.4. NPI summary statistics.. 31
Table 2.5. ADNI characteristics. 35
Table 3.1. List of linear mixed effects models examined. 40
Table 3.2. Comparison of model fit....42
Table 3.3. The variables included in Model 5 as fixed effects and their effect on MMSE....43
Table 4.1. Number of SNPs....61
Table 4.2. Number of SNPs....62
Table 4.3. Association of APOE genotype with cognitive decline....64
Table 4.4. List of ten variants with the strongest association in the GWAS meta-analysis....70
Table 4.5. List of genes with the strongest association in the gene-based test. 71
Table 4.6. List pathways with the strongest association....72
Table 4.7. Association of AD risk pathways with cognitive decline. 74
Table 4.8. Association of PRS with cognitive decline....75
Table 4.9. Association of AD PRS with cognitive decline....75
Table 5.1. Training datasets used in the PRS analysis. 86
Table 5.2. Summary statistics of NPI for EOAD and LOAD. 88
Table 5.3. Component loadings for individuals with LOAD. 90
Table 5.4. Component loadings for individuals with EOAD. 91
Table 5.5. Component loadings for all individuals. 93
Table 5.6. Comparison between mild to moderate and moderate to severe AD. 94
Table 5.7. Comparison between early and late AD. 95
Table 5.8. Association of the five sub-phenotype scores with demographic factors. 96
Table 5.9. Association of the five sub-phenotype scores with APOE genotype.... 97
Table 5.10. Association of the PRS derived with the five sub-phenotype scores....101
Table 5.11. The association of all sub-phenotypes that were found to be associated with the same PRS.....102
Supplementary Table 1. Component loadings for LOAD when using a three-component structure....157
Supplementary Table 2. Component loadings for EOAD when using a three-component structure....158

Supplementary Table 3. Component loadings for LOAD when using a three-component structure.....159
Supplementary Table 4. Component loadings for EOAD when using a three-component structure.....159
Supplementary Table 5. Association of the PRS derived with the five sub-phenotype scores.....160

List of figures

Figure 1.1. Illustration by Alöis Alzheimer 2
Figure 1.2. Cognitive reserve hypothesis. 6
Figure 1.3. Genetic variants implicated in sporadic and familial AD. 8
Figure 1.4. Illustration by Alöis Alzheimer 9
Figure 2.1. Sites participating the first stage of CAGRAD. 24
Figure 2.2. Number of missing values per NPI domain for the two CAGRAD stages.....29
Figure 2.3. Missing data imputation..... 30
Figure 2.4. Missing data imputation..... 31
Figure 2.5. Sites participating in ADNI. 33
Figure 3.1. Density plot of measure of decline..... 44
Figure 3.2. Scatterplot of measure of decline..... 45
Figure 3.3. Graphical representation of the individual rate of decline.....46
Figure 3.4. Cognitive decline in individuals with AD and controls.....47
Figure 3.5. Cognitive decline in individuals with EOAD and LOAD. 47
Figure 3.6. Density plot of measure of decline..... 48
Figure 3.7. Cognitive decline in individuals with AD and controls.....48
Figure 3.8. Cognitive decline in individuals with EOAD and LOAD. 49
Figure 3.9. Cognitive decline in individuals that enrolled as AD and converted to AD and 49
Figure 4.1. Numbers of individuals..... 56
Figure 4.2. Cognitive decline by number of APOE ε4 alleles in CAGRAD. 63
Figure 4.3. Cognitive decline by number of APOE ε2 alleles in CAGRAD. 63
Figure 4.4. Cognitive decline by number of APOE ε4 alleles in ADNI..... 63
Figure 4.5. Cognitive decline by number of APOE ε2 alleles in ADNI. 64
Figure 4.6. Q-Q..... 65
Figure 4.7. Manhattan plot..... 66
Figure 4.8. Q-Q..... 67
Figure 4.9. Manhattan plot..... 67
Figure 4.10. The region in chromosome six containing the associated variants 68
Figure 4.11. Q-Q..... 69

Figure 4.12. Manhattan plot.....70
Figure 4.13. List of the pathways with the strongest association with cognitive decline. 73
Figure 5.1. Correlation plot.....89
Figure 5.2. Scree plot.....89
Figure 5.3. PCA.....90
Figure 5.4. Correlation plot.....92
Figure 5.5. PCA.....92
Figure 5.6. Distribution.....94
Figure 5.7. Comparison between mild to moderate and moderate to severe AD. 95
Figure 5.8. Comparison between early and late AD. 96
Figure 5.9. Comparison between individuals with one, two or zero APOE ε4 alleles.....97
Figure 5.10. Comparison between individuals with one, two or zero APOE ε2 alleles.....98
Figure 5.11. Association of AD PRS with the five BPSD sub-phenotypes.....99
Figure 5.12. Association of schizophrenia PRS with the five BPSD sub-phenotypes.....99
Figure 5.13. Association of bipolar disorder PRS with the five BPSD sub-phenotypes.....100
Figure 5.14. Association of depression PRS with the five BPSD sub-phenotypes.....100
Figure 5.15. Association of IQ PRS with the five BPSD sub-phenotypes.....101
Supplementary Figure 1. PCA when using a three-component structure. 157
Supplementary Figure 2. PCA when using a four-component structure. 158

Chapter 1 | Background

1.1. Introduction

The concept of senile dementia has existed for millennia, with mentions of age-related impaired memory and cognition even found in the works of Plato and Aristotle [1]. However, the first detailed accounts of what is today called Alzheimer's disease (AD) did not occur until the early 20th century. The clinical symptoms of AD were first described by Alöis Alzheimer in 1907 to describe a patient he was attending to at the state mental asylum in Frankfurt, named Auguste Deter [2]. Auguste was 50 years old and exhibited progressively deteriorating memory loss, along with psychosis, confusion and aggression, as documented in detail in Alzheimer's manuscript "Über eine eigenartige Erkrankung der Hirnrinde" [2]. In addition to providing what is the first in-depth description of the clinical manifestation of AD, Alöis also performed an autopsy on his patient's brain post-mortem, providing a detailed neuropathological description. He described what has come to be known as the landmark pathological findings of AD, neuritic plaques of accumulated amyloid beta (Aβ) and neurofibrillary tangles (NFTs) of hyperphosphorylated tau protein [3]. The term Alzheimer's disease was given later by the world-renowned psychiatrist Emil Kräpelin, Alöis's colleague and mentor. In the chapter on senile and presenile dementias of his psychiatry textbook "Psychiatrie: ein Lehrbuch für Studierende und Ärtzte" he described the clinical and pathological characteristics of what he called AD, and insinuated that old age is unlikely to be the solitary cause of these findings [4]. Indeed, it is now a known fact that AD is not a part of normal ageing, however the exact causes of it are still unclear.

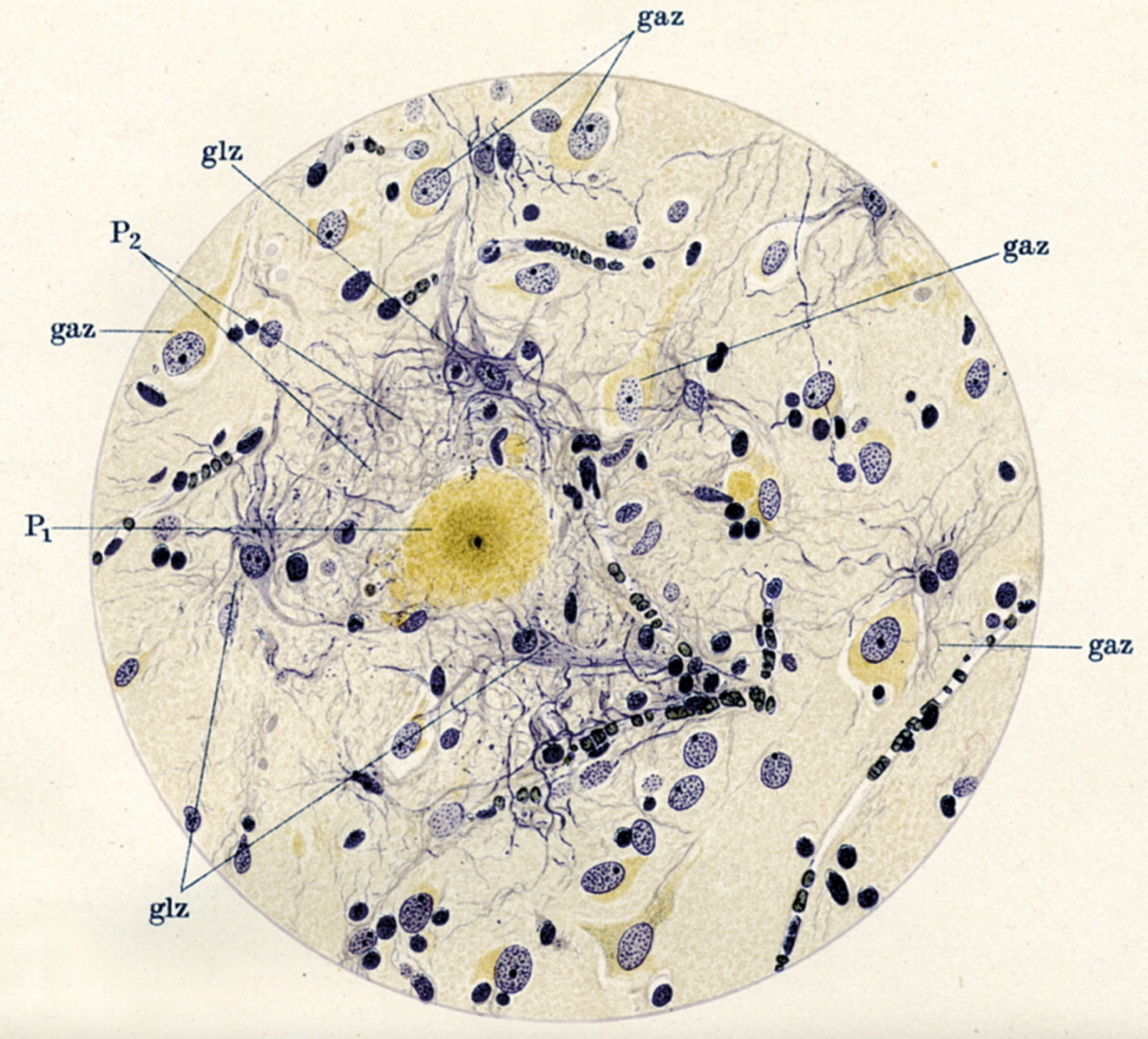


Figure 1.1. Illustration by Alöis Alzheimer describing the amyloid plaques and NFTs as seen in the post-mortem pathological examination of the brain of Auguste Deter, included in his 1911 manuscript [2]. P1 refers to the central part of the amyloid plaque, P2 to the periphery of the plaque , glz to glial cells and gaz to neurons. Image reprinted with permission.

1.2. Epidemiology

1.2.1. Prevalence and impact

AD is the most common form of dementia and the most prevalent neurodegenerative disease. It affects over 50 million individuals worldwide, and due to the global ageing of the population this figure is predicted to increase fourfold within the next 30 years [5]. Each year, an estimated 4.7 million new cases of AD are diagnosed, which amounts to one new case being diagnosed every 7 seconds [6]. The majority of the cases are in low-income countries, where the increase

in prevalence is also predicted to be the highest within the coming decades as the mean age of the population and life expectancy continue to increase [6]. As the proportion of elderly population rises, AD has the potential to become an intractable problem on a global scale. The incidence of AD increases exponentially with age until the age of 85, when it reaches an inflection point [7], [8]. Certain studies have shown that the incidence declines after a plateau is reached [9], while others indicate that the plateau remains until the end of life [10], [11].

The median survival after AD diagnosis is 6.2 years, ranging from 3 to 6, with many factors like age, sex, ethnicity, educational level, and comorbidities influencing survival time [12]–[14]. Notably, AD is the leading cause of death in England and Wales, accounting for 12.7% of all deaths registered [15], while it is also suggested that AD-related mortality as reported on death certificates is an underestimate of the actual number [16]. While deaths by other common non-communicable diseases like cardiovascular diseases and certain cancers are declining, deaths from AD continue to increase, partly due to the lack of effective treatments [17].

It is estimated that AD is responsible for over 10% of disability in adults over the age of sixty years, which is more than cardiovascular disease and stroke combined [18]. Over 40% of individuals with AD require institutionalisation, particularly in industrialised societies [5]. In low- to middle-income countries, informal caregiving by family members is often the only available care [19]. However, informal unpaid caregiving for individuals with AD is also common in high-income countries [17], [20]. This results in enormous societal costs associated with dementia care, which exceeded $1 trillion in 2019 worldwide [21]. In the UK, there is an estimated 850,000 people with AD, resulting in a total estimated societal cost of £26.3 billion per annum [22]. It is reasonable to assume that as the number of individuals living with AD continues to rise, the costs relating to AD care will continue to increase, causing a significant strain on nationalised healthcare systems worldwide. Therefore, the urgent need for effective interventions to prevent and treat AD is evident.

1.2.2. Risk factors

The strongest risk factor for AD is ageing, with the incidence dramatically increasing with age [23], from 3.9 cases for every 1000 individuals between the ages of 60 and 69, to 104.8 cases for ever 1000 individuals above the age of 90 [24]. The importance of age is such that it is used as a categorical marker, denoting the two main AD sub-categories, early-onset AD (EOAD) and

late-onset AD (LOAD). The arbitrary cut-off most commonly used is 65 years of age, considered to be the age after which the majority individuals will start exhibiting signs and symptoms [25]. The majority of AD cases are LOAD, with EOAD accounting for 5-10% of the total cases [25]. AD prevalence also differs by sex, with more women living with the disease than men, particularly in the ages of 80 years of age or above [26]. The incidence of AD has been found by some studies to be similar in men and women, suggesting that the higher prevalence in women stems from their higher longevity [27], while there are also studies reporting a higher incidence in women [28].

Many pre-existing health conditions have been linked to a higher risk of developing AD, most of them related to the cardiovascular system. Cardiovascular disease has been consistently linked with a higher risk of developing AD [29], [30], with peripheral arterial disease showing the strongest association [30], [31]. Cerebrovascular pathology usually coexists with AD-related pathology in post-mortem brain tissue of individuals with AD, however, it is considered that the vascular lesions do not contribute to the AD-related findings but exacerbate the dementia phenotype [32]. High serum cholesterol in mid-life also increases AD risk [33]. Hypertension, especially in mid-life, has been found to predispose to AD in some studies [34], however the results have been inconsistent, with studies with shorter follow-up periods not detecting any effect [35]. Type 2 diabetes mellitus, a disease strongly associated with cardiovascular disease, has also been associated with a higher risk of AD and neurodegeneration [36], especially if it is present at mid-life or has a long duration [37]. This could be due to a direct effect of hyperglycaemia and insulin resistance on brain pathology, due to the many vascular complications of diabetes, or even an indirect effect linked to the many cardiovascular comorbidities of diabetes mellitus [38]. Obesity has also been associated with a higher risk of developing AD later in life [39]–[41]. However, the association is moderate, and the biological pathways that could be implicated in this association are unclear at the moment. Obesity is the strongest risk factor for diabetes mellitus and cardiovascular disease, both significant predictors of AD risk. Therefore, it is possible that obesity has an indirect effect on AD risk through increasing the risk of diabetes and cardiovascular complications. A diet high in sugars and saturated fats, like the one commonly observed in industrialised societies, has been linked to cognitive impairment and neurodegeneration [42], still it is again unclear whether the effect of diet on AD risk is direct or indirect. Interestingly, low BMI in late life is also associated with a higher risk of developing AD, especially if the decline in BMI was rapid

[43]. However, that is considered to be due to the fact that weight loss precedes the onset of symptoms of AD and may be a prodromal sign rather than a risk factor [44]. Finally, an association has been observed between depression and development of AD later in life, though it is still undetermined whether history of depression is a risk factor, or depressive symptoms are indicative of preclinical AD [45]–[47].

1.2.3. Protective factors

It has been consistently observed that intellectually demanding activities protect from the development of AD [48]. This observation led to the cognitive reserve hypothesis, based on which sustained complex mental activity protects the central nervous system from injury and neurodegeneration, and has the potential to postpone the onset of dementia [49], [50]. It is hypothesised that individuals with a high cognitive reserve can tolerate more extensive AD pathology without exhibiting any dementia symptoms, due to their high premorbid cognitive function. However, once the a certain point is reached where the pathology overcomes the compensatory mechanisms of the cognitive reserve, symptoms of dementia commence to appear, and the course of disease is not affected further [50], [51]. Interestingly, it is suggested that part of the reason behind the increased AD prevalence in low- and middle-income countries compared to high-income ones is the lower levels of educational attainment observed there [19], which supports the cognitive reserve hypothesis. It should be noted, though, that cognitive reserve data are based on autobiographical observations, and derived by variables like educational attainment and level of complexity of profession that do not necessarily correspond to the mental activity of an individual.

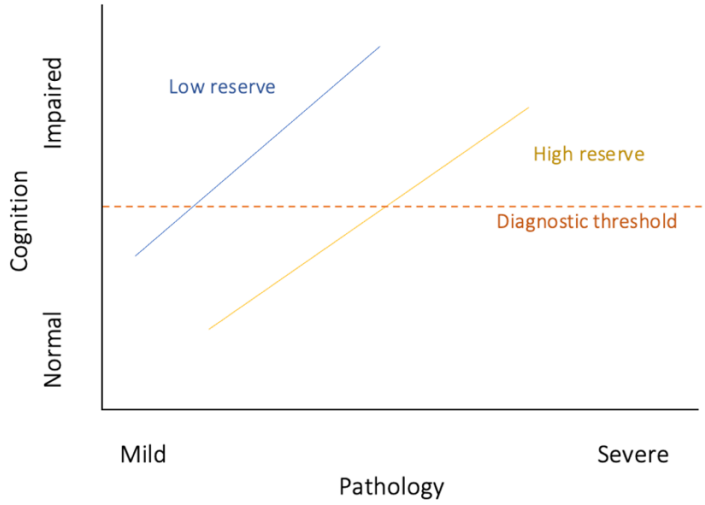


Figure 1.2. Cognitive reserve hypothesis. Individuals with a high premorbid cognitive reserve can tolerate more extensive neurodegenerative pathology before reaching the threshold of cognitive impairment required for a diagnosis of AD.

Increased physical activity in mid-life has been consistently found to lower the risk of developing AD, however the mechanisms by which that is achieved are unclear [52]. Physical activity is also associated with certain neuroanatomical findings that are associated with a lower AD risk, like higher grey matter and hippocampal volume [53], [54]. Moreover, physical activity is protective from many of the risk factors mentioned above, like diabetes mellitus, obesity and hypertension, so it is possible that the reduction in AD risk is achieved indirectly, through mitigating the effect of those risk factors. Lastly, the Mediterranean diet as well as diets with high seafood consumption have also been suggested as possible protective factors [55]–[57], though the evidence for that is still emerging, and it is unclear whether the diet alone affects dementia risk, or the protection is achieved through the associated cardiovascular benefits.

1.3. Genetics

Evidence from early studies suggests that AD is largely attributable to genetic factors, with an average estimated heritability of 75%, ranging from 58-79% for LOAD to over 90% for EOAD [58], [59]. In the past decades there has been tremendous progress in the domain of AD genetics, with numerous genetic loci having been implicated in the pathogenesis of AD.

1.3.1. Autosomal dominant variants

A small fraction of AD cases is attributable to autosomal dominant mutations that cause familial AD. So far, three genes have been found through linkage studies to cause this rare form of AD, *PSEN1*, *PSEN2* and *APP* [60], [61]. They encode presenilin and the amyloid precursor protein, respectively. Both proteins are involved in the cleavage of amyloid, with the mutations predisposing to increased Aβ generation and aggregation [60]. Mutations in these genes lead to a severe familial form of AD that is characterized by an early onset, often before 40 years of age, and a rapid cognitive and functional decline [25]. However, it should be noted that not all variants in these genes give rise to familial AD, with a number of them only slightly increasing the disease risk and others being unrelated to the development of AD [60], [62]. Mutations in *APP*, *PSEN1* and *PSEN2* have only been detected in a small fraction of familial AD, while there are clusters of familial AD for which the causative mutation is yet to be determined [60]. Even within the individuals with EOAD caused by one of the three known autosomal dominant mutations there is great variability in age of disease onset, which indicates that even when deleterious mutations are present there are additional factors that influence the onset and course of AD.

1.3.2. APOE

Early genetic studies implicated a region in chromosome 19 in the development of sporadic LOAD, which was later determined to be the *APOE* locus, the strongest genetic determinant of sporadic AD. It encodes the protein apolipoprotein E, the primary modulator of cholesterol in the brain with a complex role in lipid and protein homeostasis [63]. *APOE* has three alleles: ε2, ε3 and ε4. The *APOE* ε4 allele causes a three- to fourfold increase in the risk of developing AD, making it the strongest genetic risk factor for sporadic AD [64]. The effect of APOE ε4 on the risk of developing AD is dose-dependent, with an odds ratio (OR) of 14.7 for ε4 homozygotes compared an OR of 3.2 ε4 heterozygotes [65]. Conversely, the *APOE* ε2 allele leads to a twofold decrease in the risk of developing AD, making it the strongest genetic protective factor from AD [66]. *APOE* ε3 allele has minimal impact on AD risk. *APOE* also affects the presentation of AD, with ε4 being associated with an earlier age at disease onset [64]. It has been suggested that *APOE* influences the accumulation and clearance of Aβ and could be implicated in cerebral amyloid angiopathy and neuroinflammation, both processes associated with AD [66], [67].

However, the APOE ε4 allele does not completely account for the heritability of sporadic AD, explaining only 6% of AD heritability in isolation [68].

1.3.3. Other variants

Sporadic AD follows a pattern of complex heritability, its genetic risk being mediated by a multitude of common genetic variants that individually have a small effect on disease risk, but cumulatively infer a risk high enough to explain the heritable nature of AD. Genome-wide association studies (GWAS) have facilitated the discovery of various such variants. The first three loci with a genome-wide significant association (p-value < $5x10^{-8}$) with AD, *CLU*, *PICALM* and *CR1*, were discovered in 2009 [69], [70]. Since then, the number of known genome-wide significantly associated loci has risen to 75 [71], and the estimated heritability of AD as a result of common variants is 33% [68]. Interestingly, some of the variants, like *TREM2,* have a stronger association with EOAD [72]. The increase in AD risk inferred by those variants tends to be small, with effect sizes ranging from 0.1 to 0.13 [71], therefore they are not useful in isolation when trying to predict the likelihood of developing AD. Still, they can be useful in deconvoluting AD pathogenesis. Functional annotation of these loci has shown that the biological pathways that are implicated in the genetic risk for AD include immune response, lipid metabolism, endocytosis and vascular processes [73], [74]. It is hypothesised that some of the variants implicated in AD risk also affect Aβ and tau accumulation directly or indirectly, however the mechanisms through which this happens are unclear.

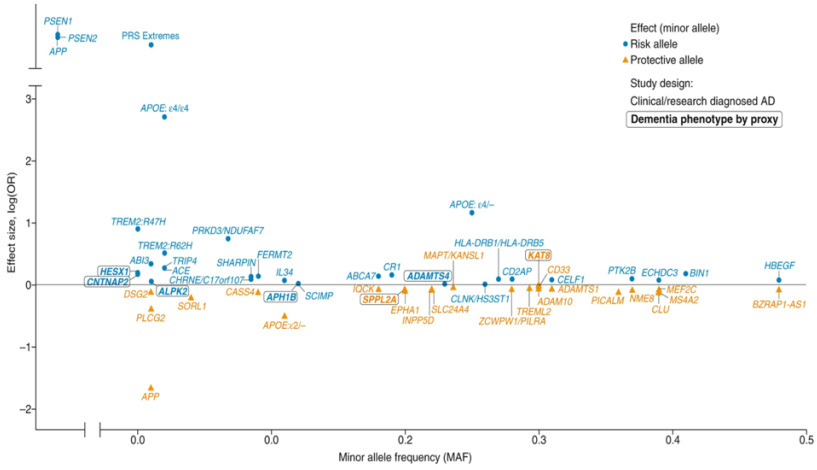


Figure 1.3. Genetic variants implicated in sporadic and familial AD. Protective variants are depicted in blue and risk variants are depicted in yellow. Image from Sims, R., Hill, M. & Williams, J., 2020, reprinted with permission.

1.4. Pathology

The main pathological landmarks of AD are microscopical, first described by Alöis Alzheimer in the early 20th century; extracellular amyloid plaques of accumulated Aβ and intracellular NFTs of hyperphosphorylated tau protein [2].

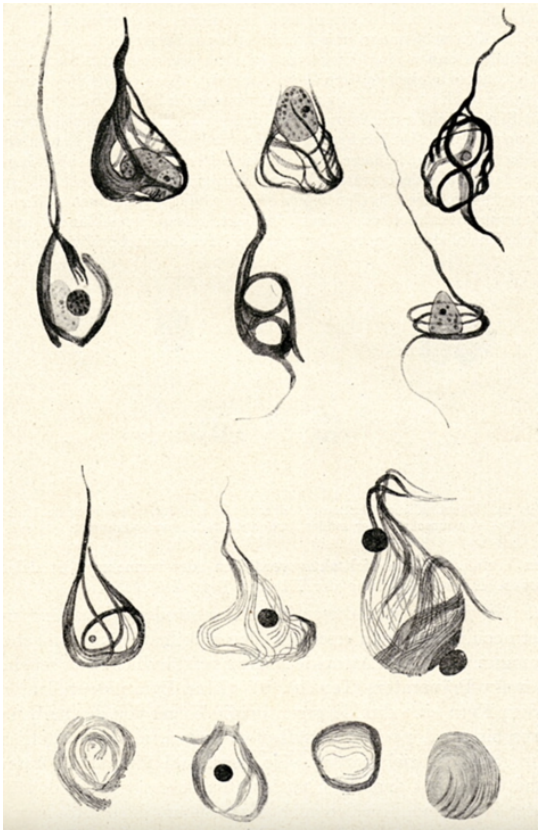


Figure 1.4. Illustration by Alöis Alzheimer depicting the amyloid plaques and NFTs as seen in the post-mortem pathological examination of the brain of Auguste Deter, included in his 1911 manuscript [2]. Image reprinted with permission.

Aβ is the result of the abnormal cleavage of the amyloid precursor protein (APP) by β and γ secretases, resulting in overproduction of $A\beta_{42}$ peptides, which fold into β pleated structures that subsequently accumulate into amyloid plaques [75]. Apart from forming amyloid plaques in the brain parenchyma, Aβ also forms deposits in cerebral blood vessels, with 85-95% of AD brain samples exhibiting some degree of amyloid angiopathy [76]. Tau is a microtubule associated protein present in axons. In AD, tau monomers first bind into oligomers, that subsequently aggregate and form NFTs. Tau in NFTs is hyperphosphorylated, which reduces its affinity to microtubules, however the contribution of the phosphorylation to the aggregation is unclear [76]. NFTs seem to have a stronger correlation with cognitive impairment than Aβ

[77]. Neuritic plaques are also present, consisting of tau-positive neurites and a central core of dense Aβ [75]. The pathological findings, including both neuritic plaques and NFTs tend to be exacerbated in individuals with EOAD compared to LOAD [78].

Neuronal loss is another pathological landmark of AD. It is correlated with the distribution of tau NFTs and is a good predictor of cognitive impairment [79]. However, synaptic loss appears to precede neuronal loss, as well as cognitive impairment, and is considered to be driven by amyloid and tau pathology [80]. Moreover, synaptic loss has the strongest association with cognitive impairment, higher than both tau accumulation and neuronal loss [81]. The loss of synapses could be caused by axonal dysfunction due to the tau protein abnormalities seen in AD, or by the accumulation of tau obstructing synaptic transmission. However, there is an enlargement of remaining synapses, indicating that compensating mechanisms are in place [76].

An inflammatory response is also visible in brain specimens of individuals with AD. Microglial activation, an innate immune response, is usually present surrounding neuritic plaques [82]. Reactive astrocytes are also present in the brain parenchyma of individuals with AD. They are also located around the amyloid plaques, though they are not as abundant as microglia [82]. The theory is that the astrocytes are attracted to the area by the cytokines released by the activated microglia, as they appear later in the course of the disease compared to the microglia [83]. It is possible that they have neurotoxic capabilities and they are linked to the neuronal loss seen in AD [84].

Detecting AD pathology requires a microscopic examination, achieved by biopsy or usually by post-mortem autopsy. By the time of death, there is usually extensive pathology and neuronal loss. Comorbidities are common, with over half AD cases also exhibiting non-AD pathological findings, with that number increasing with age [75]. The most common concurrent pathology is vascular pathology, found in over 40% of brain specimens of individuals with AD [85]. Additionally, there are some macroscopical pathological findings that are commonly found in individuals with AD, including moderate to severe cortical atrophy, enlargement of lateral ventricles and medial temporal atrophy affecting the hippocampus and amygdala [75]. The atrophy and ventricle enlargement results in a reduction of total brain volume and weight [75], [76]. However, none of these findings are specific to AD, and should only be used as an indicator and not a proof of AD. Amyloid deposition and tau pathology precede structural brain

abnormalities by decades, suggesting that macropathological findings are a result of the AD-specific micropathology [76].

1.5. Pathogenetic mechanism

Despite many breakthroughs in dementia research during the last decades, the causative mechanisms of AD remain largely unknown. The amyloid cascade hypothesis remains the leading view on AD pathogenesis. According to that, abnormal cleavage of APP resulting in higher concentration of the highly fibrillogenic $A\beta_{42}$ peptide causes Aβ deposition and leads to neuronal injury and death [86]. The amyloid hypothesis also proposes that tau pathology is initiated by Aβ-mediated neurotoxicity [86]. Early findings from genetic studies that discovered the three variants responsible for autosomal dominant AD were pivotal to the development of the amyloid cascade hypothesis [87]. However, the validity of the amyloid cascade hypothesis remains in question, as there is evidence of neurodegeneration prior to plaque pathology in early AD [88], [89]. Tau pathology has shown evidence of spreading between neurons, with potential direct cell to cell propagation of neurofibrillary tangles. It is suggested that this process precedes Aβ pathology. Moreover, there is evidence suggesting that Aβ accumulation is the result of neuronal damage, and not the cause of it [90].

The cholinergic hypothesis has also been suggested, according to which acetylcholine deficiency leads to the symptoms of AD [91]. This hypothesis is based on the post-mortem finding of low levels of acetylcholine producing enzymes in brains of AD patients, and further supported by the fact that acetylcholinesterase inhibitors improve cognition in AD [91]. However, the idea that the cognitive deficits that appear in AD are solely dependent on acetylcholine deficiency has been widely critiqued, among other reasons because the effect of acetylcholinesterase inhibitors in treating AD is moderate at best [92].

Neuroinflammation has been shown to be present in the brain tissue of individuals with AD and is considered to play an important role in the extensive neurodegeneration that takes place. It has been shown that activated microglia and astrocytes can promote Aβ accumulation and cause neuronal damage [93], [94]. Moreover, evidence from genetic studies suggests that there is an involvement of the innate immune system in AD pathophysiology [95]. It is, therefore, plausible that neuroinflammatory processes are implicated in AD, although at the

moment it is unknown how much of the contribution of neuroinflammation to AD burden is causative and how much is the result of pre-existing Aβ and tau pathology.

1.6. Clinical presentation

AD is a dementia, and as such the main symptom is cognitive decline, although the phenotype of AD can vary widely between individuals, suggesting the presence of various AD sub-phenotypes [96].

The clinical manifestation is similar between LOAD and the majority of EOAD cases. However, there is a subset of around 25% of individuals with EOAD that has a distinct clinical presentation, consisting of a number of non-memory cognitive deficits, including apraxia, dyscalculia, aphasia and visual and executive dysfunction, with relatively preserved episodic memory [97], [98]. Moreover, individuals with EOAD tend to experience a more aggressive form of the disease, with accelerated progression and shorter survival [97]. However, this is mostly documented in EOAD caused by autosomal dominant mutations, with the differences being less clear when examining sporadic EOAD [99].

1.6.1. Cognitive decline

An impairment of memory and cognition is the landmark symptom of AD. Episodic memory is the first aspect of cognition that becomes affected and its decline dominates the disease phenotype while being accompanied by other cognitive deficits, like impairment of executive function and language [100]. It is important to note that while cognitive decline is also a part of healthy ageing, and some of the changes seen in AD can overlap with the ones seen in cognitively healthy older adults, these changes are distinct, as cognitively healthy ageing individuals maintain their personality traits, behaviour, motivation, interests and executive functioning, which are commonly lost in AD [101], [102].

Mild cognitive impairment (MCI) is considered a prodromal phase of AD, during which the individual remains capable of living independently [103]. The loss of independence is one of the key clinical characteristics that differentiates MCI from AD [103]. Two types of MCI have been characterised, amnestic and non-amnestic, the former characterised by a gradual impairment of episodic memory while the latter mainly involving deficits in executive function, higher distractibility and impaired learning [104], [105]. However, only around 30% of the

individuals with MCI progress into developing AD [106], and the cognitive deficits seen in MCI tend to fluctuate over time and can be reversed [107]. The progression from MCI to AD is characterised by a loss of independence, as the individuals' executive function deteriorates further. Multiple cognitive and functional domains are affected in addition to episodic memory as the disease progresses, including language, visuospatial processing, cognitive flexibility and self-monitoring [100].

Cognitive function continues to decline for the duration of the disease, but the rate of decline has been found to vary widely between individuals. Over 30% of individuals with AD exhibit a rapid decline [108], [109]. Fast decliners have been shown to have a shorter survival and are more likely to be institutionalised [109], [110]. The reasons for the heterogeneity in the rate of decline observed in AD remain mostly unknown. A number factors have been suggested to affect the rate of decline, including age at onset [111], age [112], comorbidities [113], sex [114], education and occupational complexity [48]. However, the results are often conflicting, and it is yet unclear which of these factors directly influence the rate of cognitive decline in AD. The presence of a genetic predisposition to faster decline has also been suggested. The effects of the *APOE* genotype on the rate of decline have been widely researched, with studies finding conflicting results [115]–[119]. To date, there is no clear consensus on whether the number of *APOE* ε4 or ε2 alleles affects the rate of decline in AD. Other variants have also been examined [62], [120] and three GWAS have been attempted, two of them failing to uncover any significant association [121], [122] while the other detecting one variant of genome-wide significance [123], however in a very small dataset. Therefore, there is at present a lack of clarity regarding the possibility of a genetic determinant of the rate of decline and extensive further studies are required to confirm the presence of genetic risk factors for faster cognitive decline.

1.6.2. BPSD

In addition to the cognitive and functional decline, AD is often complicated by a wide array of non-cognitive symptoms, termed Behavioural and Psychological Symptoms of Dementia (BPSD). This term encompasses a variety of symptoms, including psychotic experiences, aggression, apathy, depression and eating and sleeping disturbances [124]. They are very common, affecting over 90% of individuals with AD [125], [126] and have been observed from

as early as Alöis Alzheimer's first account of the symptoms that Auguste Deter was experiencing [2]. Sometimes they precede the clinical diagnosis, with symptoms of anxiety, depression and apathy common during the preclinical stages of AD [127], [128].

BPSD have a very big impact on the individuals living with AD. They are associated with a faster cognitive and functional decline and a shorter survival [129]. Most importantly, they are linked to very high caregiver distress [130], [131]. Caregivers of individuals with AD find that BPSD are more distressing and difficult to manage than the cognitive symptoms [132], [133]. This is part of the reason why BPSD lead to earlier instutionalisation [134] and are considered to greatly exacerbate the cost of dementia care [135]. However, these symptoms are still a widely overlooked aspect of AD research, and not much is known about their pathogenesis.

A number of different hypotheses have been explored [136], [137]. One of the most prominent suggests that BPSD are a result of the neuroanatomical changes seen in the brains of individuals with AD. Extensive atrophy is often centred in brain areas responsible for impulse control, motivated behaviour and emotional expression and regulation, like the hippocampus, the frontal lobes, and the amygdala [124]. Loss of neurons and synapses in these brain areas could have detrimental effects on the individual's behaviour and emotions. Moreover, monoamine deficiency could also contribute to BPSD development. Monoamines like dopamine, norepinephrine and serotonin have been shown to be reduced in the brains of individuals with AD [138]. These neurotransmitters are implicated in the reward pathway and are crucial to mood regulation and motivation and are compromised in various psychiatric conditions including major depression, schizophrenia and anxiety [138]. However, the mechanism that leads to monoamine deficiency in AD is unclear. Another theory suggests that BPSD are a consequence of the extensive neuroinflammation that characterises AD, as neuroinflammation is also implicated in psychiatric conditions like major depression and schizophrenia [139], [140]. However, there is no substantial evidence for any of these hypotheses at the moment and the exact aetiology of BPSD remains elusive. It is possible that BPSD are a direct effect of the neurodegenerative processes that dominate AD. On the other hand, there is also the possibility that BPSD are the consequence of latent predisposition to psychopathology that this pervasive neurodegeneration brings to light. Extensive further research is required to deconvolute the pathogenesis of BPSD.

The presence of a genetic background to BPSD pathogenesis has been explored. Most genetic studies have focused on psychosis and have shown that it is possible that AD with psychosis is

a distinct AD sub-phenotype with a strong genetic component, as it tends to show familial aggregation and has an estimated heritability of 61% [141]. Recently, a GWAS of psychosis in AD revealed the first genome-wide significantly associated loci [142]. It has additionally been shown that psychosis in AD is associated with the genetic risk of developing schizophrenia [143], [144], suggesting that there could be a common pathway linking BPSD and neuropsychiatric disorders. However, evidence for a genetic predisposition to other BPSD is limited and largely unexplored.

1.7. Diagnosis

The diagnosis of AD is made clinically. However, the only definite way to diagnose AD is by a post-mortem pathological examination of the brain tissue, while a clinical diagnosis is restricted to probable AD. When all appropriate procedures are followed, a clinical diagnosis can have a good specificity when comparing individuals with AD with cognitively healthy individuals, however it is not as accurate when attempting to differentiate between different type of dementia [145]. The clinical diagnosis involves taking a detailed history from both the patient and another source of information, like a spouse, offspring or caregiver, and administering specialized neuropsychiatric assessment questionnaires, like the Mini Mental State Examination (MMSE) [146] or the Montreal Cognitive Assessment (MoCA) [147] for assessing cognition and the Neuropsychiatric Inventory (NPI) [148] for establishing the presence of BPSD. A thorough physical and neurological examination is performed in order to rule out somatic causes of the symptoms. Standardised diagnostic criteria, like the ones established by the Diagnostic and Statistical Manual of Mental Disorders (DSM) [149] or the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [150], are commonly employed. Moreover, neuroimaging is often also utilised to aid in the differential diagnosis of other brain disorders that could have a similar clinical presentation, which further increases the accuracy of the clinical diagnosis of AD [151]. The resulting accuracy of the clinical diagnosis of probable AD is relatively high, having a specificity of 81% and a sensitivity of 70%, when confirmed by post-mortem pathological examination [145]. On average, there is a distance of three years between the symptoms are first noticed by someone in the individual’s close circle

and an official diagnosis being made [152]. In individuals with EOAD there can be further delay caused by initial misdiagnoses [153].

A definite diagnosis of AD can be achieved post-mortem and is based upon specific neuropathological findings, largely similar to the ones described by Alöis Alzheimer in 1909 [3]. The neuropathological findings that need to be present to diagnose AD are amyloid plaques, particularly cored neuritic plaques, and NFTs of hyperphosphorylated tau protein filaments. The diagnosis is based on the density as well as the topography of the pathological findings [154], and includes a semiquantitative measurement of amyloid deposition, the Braak score for assessing NFTs [156] and the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) score [156] for assessing the presence of neuritic plaques [154], [157]. It is important to note here that the pathological findings can be apparent prior to the onset of any cognitive deficits [76] and can even be present in the absence of symptoms of dementia in elderly individuals [154]. The pathological findings suffice to reach a diagnosis of definite AD, regardless of the presence of cognitive or non-cognitive symptoms of dementia. For this reason, the prevalence of pathologically defined AD could up to three times higher than that of clinically diagnosed AD, indicating that AD-related pathology does not always lead to a dementia phenotype [158].

1.8. Biomarkers

As at the moment it is only possible to definitely diagnose AD post-mortem, there is a drive for developing accurate and valid biomarkers that will facilitate achieving a definite AD diagnosis in vivo. Particularly, the emphasis is on developing biomarkers that can aid to early detection of AD, before widespread neurodegeneration sets in, as that is considered the optimal timepoint for potential therapeutic interventions. Various types of biomarkers are being assessed, some with very promising results.

The biomarkers most widely utilized in AD research are neuroimaging-based. The use of structural magnetic resonance imaging (MRI) has been utilized to detect abnormalities in the brain structure of individuals with AD, the main finding being atrophy in the medial temporal lobe, a brain region associated with memory processing [159]. Using this finding, MRI has been proven effective in differentiating between individuals with AD and cognitively healthy controls with an accuracy of 85% [151], [159], however it is not useful for differential diagnosis between

AD and other dementias [160]. Functional neuroimaging has also been widely assessed for its use in AD diagnosis. Positron emission tomography (PET) fluorodeoxyglucose is used to assess differences in glucose metabolism as a proxy of cellular metabolic activity. It has been used to distinguish between individuals with AD and cognitively normal controls with very promising results, showing that AD is associated with reduced glucose metabolism in bilateral temporal parietal regions [161]. This method has been found to have a high accuracy even at earlier stages of AD and has received approval for diagnostic use in the USA [162]. However, as is the case with structural imaging, it is not very successful at differentiating between types of dementia [161]. PET with amyloid ligands is another promising functional neuroimaging technique that can achieve in vivo visualization of Aβ accumulation. A link between post-mortem amyloid pathology and amyloid binding in PET has been established, however it is still an emerging technique and further studies are required in order to assess its clinical validity [163].

Despite the promising findings, neuroimaging methods are costly and can only be performed in specialized centres. Therefore, they are not the type of biomarker that can be widely utilised as a diagnostic or especially a screening tool. On the other hand, fluid biomarkers offer the potential of widely accessible AD screening. An array of cerebrospinal fluid (CSF) biomarkers has been tested for AD prediction, including Aβ isoforms and total and hyperphosphorylated tau protein [164]. It has been shown that a combination of these three biomarkers can achieve a 95% sensitivity and 83% specificity in detecting early-stage AD within an MCI population [165]. Other potential CSF biomarkers include inflammatory markers and cytokines, as well as biomarkers associated with mitochondrial dysfunction and oxidative stress [166]. However, there is still some inconsistencies between CSF biomarker studies, and further analyses are required in order to arrive at a standardised method of using CSF biomarkers for AD prediction or diagnosis.

The development of plasma biomarkers could provide a great potential for dementia screening, as they only require a phlebotomy, a procedure that is minimally invasive, easy to implement and low in cost. Plasma Aβ is not an optimal biomarker, as it is not possible to determine if it is of cerebral or peripheral origin. However, measuring the ratio of $A\beta_{40}$ to $A\beta_{42}$ can predict amyloid PET burden and has a high accuracy of predicting AD, although it does not outperform CSF Aβ biomarkers [167]. Measuring tau-related biomarkers in plasma has also been assessed. Plasma phosphorylated tau and has equal predictive accuracy to CSF tau

biomarkers and has been shown to increase many years prior to disease onset [167], [168]. Total tau can also be measured, though its clinical validity is questionable [167]. Combining Aβ and tau plasma biomarkers could provide a method of detecting asymptomatic AD [167]. However, the domain of AD plasma biomarkers is still an area of active research, and standardised measuring protocols need to be developed before these biomarkers are ready to be used in mainstream clinical practice.

The main challenge for imaging and fluid biomarkers is the substantial overlap seen between AD and other forms of dementia, with 40% of AD patients having vascular pathology [169], 90% of individuals with Lewy body dementia having amyloid pathology [170] and almost all individuals with vascular dementia having concurrent AD pathology [171]. Therefore, there is a need for biomarkers that can adequately distinguish between AD and other type of dementia. Moreover, the optimal biomarker should be able to facilitate the identification of patients at the earliest stages of the disease, or even at-risk individuals prior to disease onset, in order to be useful as part of a screening protocol. Genetic biomarkers have the potential of achieving both of these aims, while having the added benefit of only requiring a non-invasive sampling procedure that can be done at the individual's place of residence. As described in Chapter 1.3, the past decade has seen dramatic developments in the domain of AD genetics, with 75 variants that increase the risk of developing sporadic AD having been discovered [71]. Those variants only infer a small effect on AD risk and are not individually useful for risk prediction. However, a combination of multiple risk variants of small effect could provide an indication of an individual's AD risk. Polygenic risk scores (PRS) can be of great value in this determining this. PRS combine all risk variants under a certain significance threshold that an individual carries and weigh them by their effect sizes, thus computing a score that can subsequently be used to predict the probability of that individual manifesting the disorder of interest [172]. Although PRS are not ready to be used in the clinic as of yet, they have the potential to identify individuals at high risk of developing a disorder many years prior to any symptoms or signs becoming apparent. Particularly, the use of PRS could be of great value in the domain of clinical trials, allowing for the selection of asymptomatic high-risk individuals for receiving interventions aimed at postponing or the altogether halting the onset of the disease. PRS have been proven to successfully predict AD [173], with the prediction accuracy increasing further when only assessing pathologically confirmed AD cases [174]. However, PRS have the disadvantage of generating a score that is dependent upon the dataset it was derived from, making it a difficult

method to generalise for use as a screening tool [175]. Moreover, the PRS methodology used in publications varies widely, with the prediction accuracy reported also varying, even when the studies are based on the same datasets [176]. Currently, the clinical utility of PRS for the purpose of AD prediction is limited as none of the approaches achieved a predictive accuracy over 75% [176], and intensive further studies are required in order to determine if PRS can be used for AD screening.

1.9. Treatment

Despite the rapid advances in the understanding of the biological underpinning of AD, no disease-modifying treatment for AD has been approved to date. There are several pharmacological agents aiming to cause cognitive enhancement in use, including donezepil, galantamine and rivastigmine, cholinesterase inhibitors, and memantine, a N-methyl-D-aspartate receptor antagonist, combination of which might show additive effects on cognition [177], [178]. However, those are only symptomatic treatments, aiming to mitigate cognitive decline, and their efficacy is limited to the early stages of AD [177]. Non-pharmacological interventions are also commonly employed, including physical activity and cognitive training. Although physical activity has been shown to delay the progression of MCI to AD, it has minimal effects on AD symptoms [104], [179]. Cognitive training does seem to have a moderate effect on cognition, although the effect is limited to the domain of cognition the training is focused on and do not translate to overall cognitive improvement. [104], [180].

A number of disease-modifying treatments have been trialled or are being currently trialled, the majority of them targeting Aβ. Active immunotherapy using Aβ fragments had very promising results in transgenic animal studies, resulting in reduction of amyloid pathology and a behavioural improvement [181]. However, the results were not as impressive in humans, with Aβ clearance being moderate and not combined with a clinical improvement, while also causing severe adverse effects in some cases [182]. Another medication category targets the cleavage of APP, aiming to increase the ratio of the $A\beta_{40}$ fragment to the toxic $A\beta_{42}$ fragment [183]. However, clinical trials of such agents have not been successful [184]. Direct disruption of the formation of Aβ aggregates is currently trialled [184], [185].

It is evident that there is a need for developing novel therapeutic agents that would be effective in significantly decreasing the severity of the symptoms of AD and decelerating its

course. A better understanding of the pathogenesis of AD could result in the characterisation of potential drug targets. Moreover, there is a growing consensus that interventions to delay or altogether halt the neurodegeneration seen in AD should be administered at a very early stage. At the point of diagnosis most individuals already exhibit extensive neuropathology [75], which is likely to be irreversible. It is plausible that targeting the known pathogenetic components of AD prior to them leading to neurodegeneration could mitigate their devastating effects. Identifying individuals with preclinical AD or MCI could be beneficial for clinical trials, as these individuals would be more likely to benefit from an intervention. However, the development of adequate screening tools against AD is necessary before this becomes a possibility.

The treatment options for BPSD are also limited, and there is no specific treatment clinically available. Antipsychotics, medications commonly used to treat schizophrenia and other psychotic disorders, are commonly used to treat certain BPSD, particularly psychosis, aggression, and agitation [124]. However, antipsychotic medication is linked to significant adverse side effects and an increase in mortality in individuals with AD [186], [187]. Moreover, it has been shown that antipsychotics have a moderate effect at best at controlling psychosis in AD, therefore they have a high risk to benefit ratio should not be prescribed [188]. Aggression, on the other hand, is often severe enough that the individual poses a danger to themselves or others and justifies the use of antipsychotics, as the benefits of a short term treatment can outweigh the risks [189]. Donezepil, galantamine and memantine show some effect in controlling BPSD, particularly delusions and aggression, and the effect seems not to be limited to the early stages of the disease [190]–[192]. Antidepressants are also commonly prescribed to combat symptoms of depression, agitation and apathy, however they are not effective and have been associated with adverse effects [193], [193], [194]. Overall, non-pharmacological interventions like behaviour management, music therapy and regular exercise seem to have a moderate effect on reducing BPSD without the various severe side effects associated with the pharmacological treatments [195]–[197]. Evidently, given the devastating impact that BPSD have on individuals with AD and their caregivers, better options for managing and preventing these debilitating symptoms are urgently needed. However, until more light is shed into the elusive pathogenesis of BPSD, the development of effective treatments will be facing insurmountable obstacles.

1.10. Aims and objectives

The main aim of this thesis is to utilise a richly phenotyped dataset in order to explore the wide phenotypic diversity observed in individuals with AD. Two main phenotypic domains were selected as areas of interest due to their acute impact on individuals with AD and their potential to provide vital insight into under-researched fields of this debilitating disease; the rate of cognitive decline and BPSD.

The rate of cognitive decline was selected as the first phenotypic domain of interest, as it is an aspect of utmost importance both to the individuals living with AD and the healthcare professionals involved in AD care. The ability to identify individuals at risk of a faster decline could facilitate the work of healthcare professionals in meticulously designing a personalised treatment plan that would benefit most each individual. Moreover, identifying predictors of a faster decline could also be of value in the domain of medicine discovery and clinical trials. Individuals that are at risk of having a more aggressive disease course with fast cognitive decline and deterioration are perhaps ideal candidates for clinical trials, as it is likely that they would exhibit the effects of a given intervention faster.

The second phenotypic facet of AD explored was BPSD. BPSD were selected because despite having a devastating impact on individuals with AD and their caregivers, they are the aspect of AD that is most commonly overlooked, resulting in limited knowledge regarding their pathogenesis and no specific treatment options. Gaining a better understanding in the epidemiology of BPSD, as well as the environmental and genetic factors that contribute to their occurrence is the necessary first step towards developing effective therapeutic interventions.

The main objectives of this thesis are to:

- Harmonise the phenotypic datasets of individuals with AD available at the MRC CNGG
- Identify a publicly available dataset suitable for replicating all proposed analyses
- Compute a measure of cognitive decline in AD
- Assess the effect of *APOE* genotype on cognitive decline in two datasets
- Perform a GWAS of cognitive decline followed by a pathway analysis in two datasets
- Derive a PRS of cognitive decline and determine its ability to predict the rate of decline in an independent dataset
- Perform a PRS analysis to assess the common genetic architecture between AD risk and cognitive decline in AD

- Define sub-phenotypes of BPSD in individuals with EOAD and LOAD
- Determine demographic factors that affect BPSD sub-phenotypes
- Asses the effect of *APOE* genotype on BPSD sub-phenotypes
- Perform PRS analyses to assess the common genetic architecture between BPSD sub-phenotypes and neuropsychiatric conditions

Chapter 2 | Data description

2.1. Cardiff Genetic Resource for AD

2.1.1. Overview and aims

The data used in this thesis constitute the Cardiff Resource for AD (CAGRAD) dataset. It contains considerable wealth of phenotypic information for 4,163 individuals with early- and late-onset AD (age at disease onset < 65 years or >= 65 years, respectively) and cognitively healthy elderly individuals. The collection of this dataset was funded by the Medical Research Council (MRC) through the Centre for Neuropsychiatric Genetics and Genomics (CNGG) at Cardiff University.

The main aim of the dataset is to create a vast resource of phenotypic information on individuals with AD that can be utilised by researchers investigating the complex phenotypic and genetic interactions that are at play in AD. This dataset has been a part of data analyses of several consortia, including the Genetic and Environmental Risk in Alzheimer's Disease (GERAD), the International Genomics of Alzheimer's Project (IGAP) and the European Alzheimer's DNA Biobank (EADB), that have resulted in numerous ground-breaking findings in the domain of AD genetics [69], [198], [199].

2.1.2 Study design and recruitment

The phenotypic information was collected in two stages over the last 20 years, starting in 2001. An emphasis was placed on recruiting individuals with early-onset AD (EOAD) during the second stage, starting in 2014, while the first stage included primarily individuals with late-onset AD (LOAD). For the first stage, participants were recruited in various sites in the United Kingdom and the Republic of Ireland by researchers from Cardiff University, the Institute of Psychiatry at King's College London, Cambridge University, Southampton University and Trinity

College Dublin. For the second stage, participants were recruited as part of AD Genetics Wales and England and AD Genetics Scotland.

Figure 2.1. Sites participating the first stage of CAGRAD.

The recruitment was done via various channels, including self-referral, memory clinics, GP surgeries, the NHS clinical research network, research registries and the Joint Dementia Research (JDR) platform. The study participants were assessed at specified research clinics or their places of residence, and they were accompanied by an informant, usually a family member, that could provide additional information. Individuals with AD and healthy controls were recruited. All AD participants were considered to have probable AD based on the DSM-IV diagnostic criteria [149] or definite AD based on the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) score [200] upon neuropathological examination. They were assessed for disease severity using multiple established neuropsychiatric instruments. The cognitively healthy individuals were matched to the individuals with AD in terms of age and sex and were screened for dementia using the Mini Mental State Examination (MMSE) [146] or Alzheimer's Disease Assessment Scale Cognitive subscale (ADAS-cog) [201] and were determined to be free of dementia. A vast amount of phenotypic information was collected, including numerous cognitive and neuropsychiatric assessments like the MMSE [146], the

ADAS-cog [201], the Clinical Dementia Rating (CDR) [202], the Activities of Daily Living (ADL) [203], the Neuropsychiatric Inventory (NPI) [148] and the Geriatric Depression Scale (GDS) [204]. Comprehensive medical and family history as well as detailed demographic information were also obtained. A blood or saliva specimen was collected during the first assessment from all participants in order to acquire a DNA sample for genotyping.

2.1.3. Genetic data

The DNA samples were genotyped in two stages. For the first stage, the genotyping was performed using the Illumina 610-quad, the Illumina HumanHap550 or the Illumina HumanHap300 array at the Sanger Institute in Cambridge or the MassARRAY and iPlexGOLD systems in Cardiff, as part of the Genetic and Environmental Risk for AD (GERAD) consortium [69]. For the second stage, genotyping was performed on the Illumina Infinium GSA array, and completed in three waves in Lille, Cardiff and Edinburgh, as part of the European Alzheimer's Disease DNA Biobank (EADB) dataset [198]. The genotype samples were imputed on the Michigan Imputation Server [205] using the Genome Reference Consortium Human Build 37 assembly (GRCh37) [206] of the Haplotype Reference Consortium (HRC) reference panel [207]. The imputation was performed by Dr Ganna Leonenko. After imputation 7,581,719 single nucleotide polymorphisms (SNPs) were present.

2.1.4 Data harmonisation

Before conducting any analyses, substantial work was required in order to harmonise these two stages and extract the information relevant for all subsequent analyses. Initially, the data from the first stage of the cohort was included in three databases, two in-depth phenotypic databases with extensive information on 3192 and 3661 individuals, and a working database with demographic and minimal clinical information on 2361 individuals. The working database included information from up to 10 consecutive assessments for each participant, while the other two databases only included information for four assessments. 2704 individuals were common between the two in-depth phenotypic databases and 1517 individuals were common between all three datasets. The information on these individuals was cross validated based on

seven variables, sex, date of birth (DOB), age at disease onset (AAO), and MMSE scores on four assessments. The number of identical values per variable are shown in Table 2.1.

Variable	Identical	Different	NA
Gender	1461	26	30
DOB	1424	53	40
Age of Onset	730	114	673
MMSE1	577	79	861
MMSE2	180	14	1323
MMSE3	68	9	1440
MMSE4	34	5	1478

Table 2.1. Number of identical values for the seven variables selected for cross-validation. (DOB: date of birth, MMSE: Mini Mental State Examination)

When there was a discrepancy between the datasets, the differing data points were first visually inspected for cases of obvious typographic errors, for example date of birth after the year 2000 or an MMSE score over 30. For all other discrepancies that could not be resolved by inspection, the original interview records were revisited to determine the correct value for each data point. After revisiting the interview records, it was established that one of the databases included the correct information for the majority of the differing data points and was used to resolve any discrepancies in all other variables examined in the subsequent analyses. After correcting all non-identical values, as well as values that were missing in one of the two databases, and removing duplicated entries, a merged dataset was created, containing 4163 individuals. Out of these, 2542 were AD cases, 1611 were healthy controls and 10 were healthy controls that converted to cases within the data collection timeframe.

The data from the second stage had been recently collected and consisted of a single database that was created by using software that automatically extracted the data from the hard copy interview records into a digital database. No major data manipulation was required to utilise the data of this cohort. After removing some duplicated values, the cohort consisted of 2679 individuals with AD. All data processing was conducted in R [208].

2.1.5. Longitudinal data

Out of the Cardiff Genetic Resource for AD dataset, 1054 individuals had been assessed at multiple timepoints, 616 AD cases and 438 healthy age-matched controls. Out of the AD cases, 540 had late-onset AD (LOAD), with onset of symptoms at 65 years of age and above, and 76 had early onset AD (EOAD). The demographic characteristics of this dataset are illustrated in Table 2.2.

	Mean	SD	Range
Controls			
Age at recruitment (years)	75.75	7.12	57-101
Age at last assessment (years)	80.09	7.41	62-103
Number of assessments	3.22	1.32	2-7
First MMSE	28.48	1.61	22-30
Last MMSE	27.92	2.32	11-30
Gender	Female (%)		Male (%)
	259 (59.13)		179 (40.87)
LOAD			
Age at recruitment (years)	81.89	6.10	67-94
Age at last assessment (years)	84.33	6.09	68-102
Number of assessments	3.13	1.14	2-8
First MMSE	16.82	8.52	0-30
Last MMSE	11.34	9.09	0-30
Gender	Female (%)		Male (%)
	377 (69.82)		163 (30.18)
EOAD			
Age at recruitment (years)	66.80	7.01	41-83
Age at last assessment (years)	69.85	7.18	44-84
Number of assessments	3.15	1.12	2-7
First MMSE	18.49	8.69	0-29
Last MMSE	12.96	10.30	0-30
Gender	Female (%)		Male (%)
	38 (50)		38 (50)

Table 2.2. Sample characteristics for the portion of CAGRAD that had available longitudinal data. (MMSE: Mini Mental State Examination)

For healthy controls, the mean age at recruitment was 75.75, mean age at last assessment was 80.09 and the mean number of assessments was 3.22. Mean MMSE score at first assessment was 28.48 and mean MMSE score at last assessment was 27.92. 59.13% of the individuals were female. For the individuals with LOAD, the mean age at recruitment was 81.89, mean age at last assessment was 84.33 and the mean number of assessments was 3.13. Mean MMSE score at first assessment was 16.82, mean MMSE score at last assessment was 11.34 and 69.82% of the individuals were female. For the individuals with EOAD, the mean age at recruitment was 66.80, mean age at last assessment was 69.85 and the mean number of assessments was 3.15. Mean MMSE score at first assessment was 18.49, mean MMSE score at last assessment was 12.96 and both genders were equally represented in the dataset.

2.1.6. Behavioural and psychiatric phenotypes

The Neuropsychiatric Inventory (NPI) has been used to assess the presence of behavioural and psychological symptoms of dementia (BPSD) in the CAGRAD cohort. The NPI is a neuropsychiatric assessment instrument commonly used to examine the presence of behavioural and psychiatric symptoms in individuals with dementia and is explained in detail in Chapter 5.3.1. In summary, it assesses 12 symptom domains: delusions, hallucinations, aggression, depression, anxiety, elation, apathy, disinhibition, irritability, aberrant motor behaviour (AMB), eating and sleeping disturbances, with each domain getting a score between 0 and 12 [148]. 1724 individuals from the first stage and 1831 from the second stage of the cohort had available NPI questionnaire scores, out of which 1334 and 1622, respectively, had no missing values for any of the NPI symptom domains. The missing values for the NPI domains are illustrated in Table 2.3 and Figure 2.2.

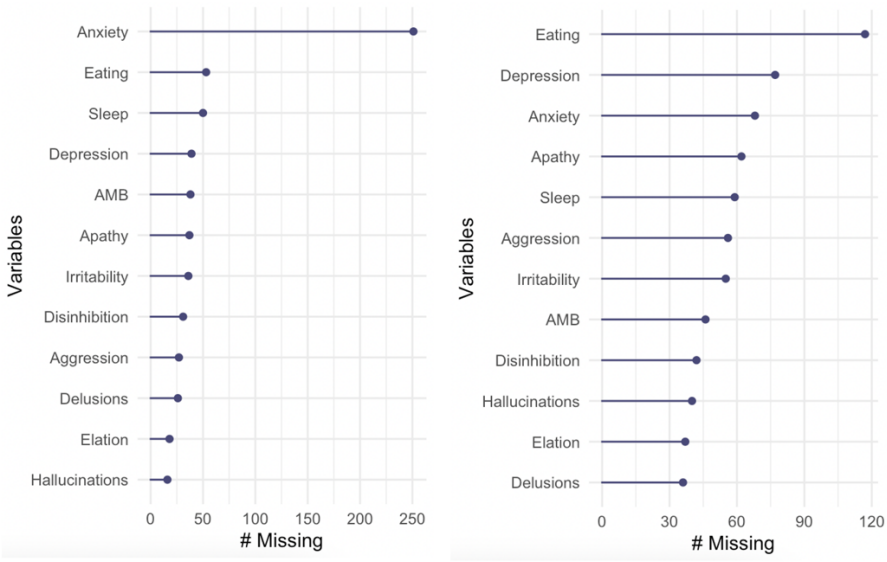


Figure 2.2. Number of missing values per NPI domain for the two CAGRAD stages, first stage on the left and second stage on the right. (AMB: aberrant motor behaviour)

	Stage 1		Stage 2	
	N missing	% Missing	N missing	% Missing
Delusions	26	1.32	36	1.97
Hallucinations	16	0.81	40	2.18
Aggression	29	1.47	56	3.06
Depression	43	2.18	77	4.21
Anxiety	252	12.76	68	3.71
Elation	18	0.91	37	2.02
Apathy	37	1.87	62	3.39
Disinhibition	33	1.67	42	2.29
Irritability	36	1.82	55	3.00
AMB	38	1.92	46	2.51
Sleep	50	2.53	59	3.22
Eating	57	2.88	117	6.39

Table 2.3. Number of missing values per NPI domain for the two CAGRAD stages. (AMB: aberrant motor behaviour)

A large number of participants from the first stage of the cohort had no entry for the domain of anxiety, as the data had been lost under unknown circumstances. As the proportion of missing data was 12.76%, it was decided that these data points would not be imputed, and the number of participants from the first stage of the cohort that have no information on anxiety would be removed from all subsequent analyses. For all other symptom domains, the missing values were imputed using the package mice() in R [209], which uses multiple imputations for each variable to replace multivariate missing data points. Fifty iterations were taken to impute each missing value using predictive mean matching (PMM) as an imputation method, and five imputed datasets were constructed for each variable. The missing data for each stage of the cohort was imputed separately. The distributions of the observed and imputed data for all the variables are illustrated in Figures 2.3 and 2.4. Based on these distributions, the imputed data points were considered to be an adequate approximation of the observed data points.

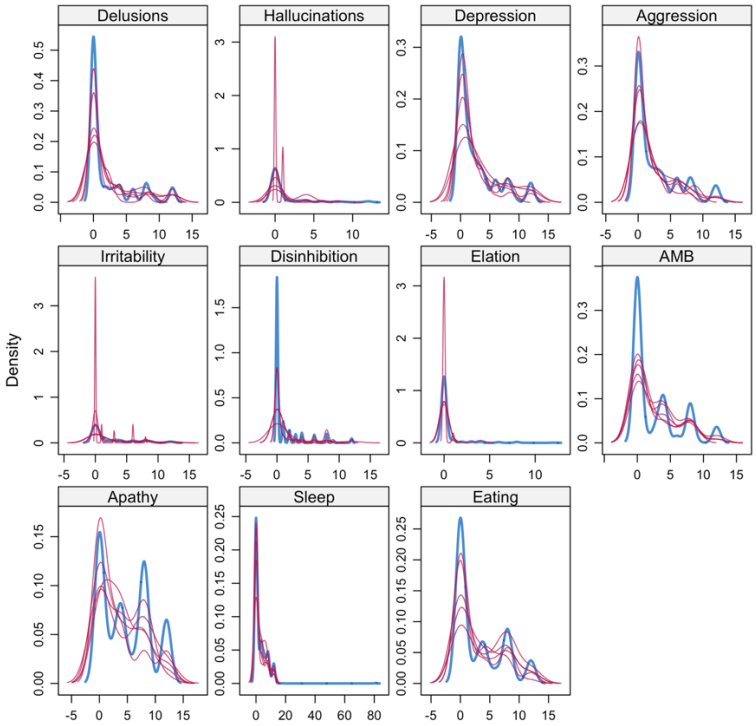


Figure 2.3. The density of the imputed data for each imputed dataset is showed in magenta while the density of the observed data is showed in blue. (AMB: aberrant motor behaviour)

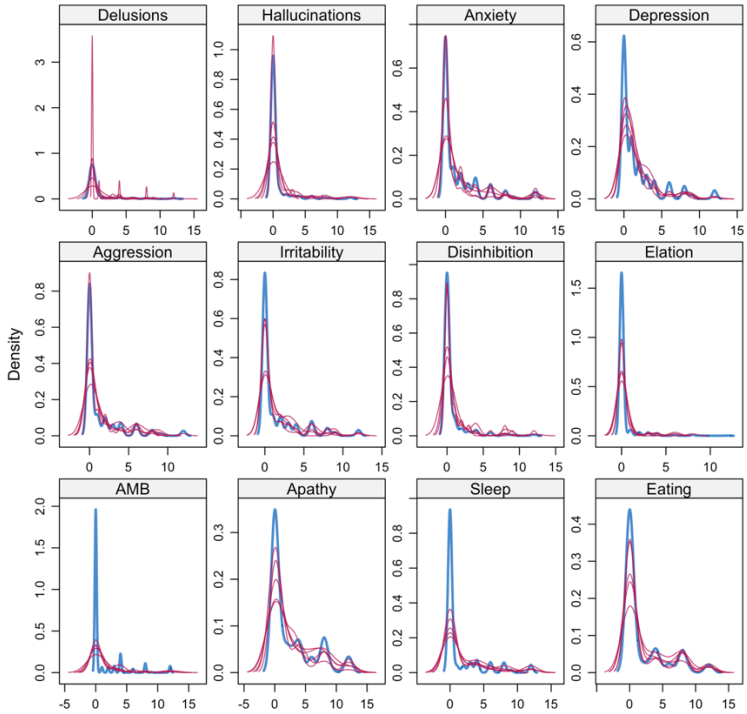


Figure 2.4. The density of the imputed data for each imputed dataset is showed in magenta while the density of the observed data is showed in blue. (AMB: aberrant motor behaviour)

NPI symptom domain	% present		Mean	
	1st stage	2nd stage	1st stage	2nd stage
Delusions	40.31	16.33	2.17	0.75
Hallucinations	22.10	14.64	1.03	0.54
Anxiety	42.87	43.36	2.14	1.63
Depression	55.86	52.81	2.50	1.77
Aggression	53.02	36.65	2.64	1.40
Irritability	44.08	36.81	2.31	1.52
Disinhibition	32.48	19.88	1.28	0.59
Elation	11.31	9.52	0.38	0.27
AMB	46.35	25.67	2.78	1.32
Apathy	69.32	48.94	4.85	2.62
Sleep	48.20	29.05	3.01	1.55
Eating	50.23	33.92	3.24	1.92

Table 2.4. Summary statistics of the 12 NPI domains for the two stages of CAGRAD. (AMB: aberrant motor behaviour)

The summary statistics of the 12 symptom domains after imputation are illustrated in Table 2.4. The most commonly occurring symptom was apathy (48.94 to 69.32%), followed by depression (52.81 to 55.86%), and the least prevalent was elation (9.52 to11.31%).

2.2. Alzheimer's Disease Neuroimaging Initiative

2.2.1. Overview and aims

The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset was used as a replication dataset for some of the analyses described in subsequent chapters. The ADNI dataset is one of the largest publicly available AD datasets in existence, the result of a multi-site longitudinal study conducted in multiple sites across North America that collected clinical, genetic, neuroimaging and biospecimen biomarkers. ADNI is a private and public partnership that has received funding from 20 pharmaceutical companies and two foundations through the Foundation for the National Institutes of Health [210].

The main aims of the ADNI study were three [210]:

- To derive methods of detecting AD at the earliest point possible, before dementia onset, and to monitor the progression of the disease using biomarkers.
- To facilitate advances in the diagnosis, prevention, and treatment of AD.
- To continue implementing the innovative open data access policy that ADNI supports.

ADNI is one of the most widely used datasets for AD and has resulted in the publication of numerous manuscripts looking at a wide range of phenotypes and biomarkers and has led to significant discoveries in the field of neuroimaging, genomics, and metabolomics [211]–[213].

2.2.2. Study design and recruitment

ADNI is a longitudinal study conducted across 57 sites in the Unites States of America and Canada, as shown in Figure 2.4 [210]. The main rationale behind the study was the identification of clinically relevant predictors of the early stages of AD, since interventions early in the disease course, stage when the degeneration is not as profound, are likely to have a greater impact than those that take place later in the disease course. ADNI aims to identify AD

markers that could be utilised for the purpose of diagnosing, monitoring progression and assessing the efficacy of novel treatments [213].

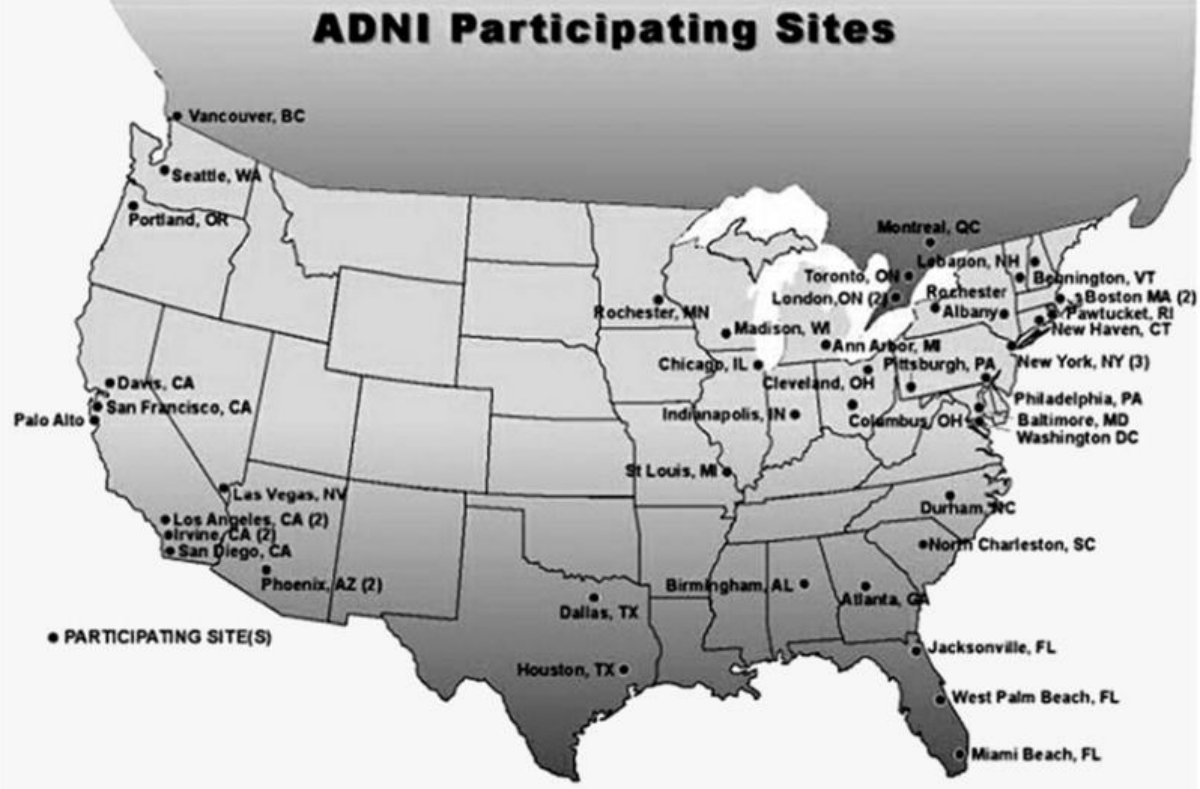


Figure 2.5. Sites participating in ADNI. Image reprinted with permission. [210]

For this purpose, the study recruited individuals with AD, individuals with early and late MCI and healthy elderly controls between the ages of 55 and 90 were recruited and monitored for several years. A wide range of information was collected from these individuals at the point of enrolment and at regular intervals throughout the duration of the study. The study participants had to undergo several tests and procedures, including multiple neuropsychological assessments, clinical evaluations, medication reviews, genotyping, lumbar puncture, and neuroimaging of various modalities [210]. The recruitment and data collection for ADNI was conducted in four stages: ADNI 1, ADNI GO, ADNI 2 and ADNI 3, resulting in a total of over 2,200 individuals recruited and monitored for over a decade [211], [214].

2.2.3. Genetic data

Genotypes were acquired for all participants upon enrolment using Illumina platforms. Illumina Human610-Quad BeadChip was used in ADNI 1 [215], including 620,901 variants, and Illumina

HumanOmniExpress BeadChip with 730,525 was used in ADNI 2, ADNI GO and ADNI 3 [211], [214]. APOE genotype was recorded for all participants upon study enrolment. The two SNPs (rs42935 and rs7412) that correspond to the APOE ε 2, 3 and 4 alleles were not included on the Illumina arrays and were therefore independently genotyped [215].

2.2.4. Data harmonisation

A specific subset of individuals from the ADNI dataset were used in this project. Only individuals with AD and healthy controls were of interest, therefore individuals with MCI that did not progress to AD at any point in the study were excluded. That resulted in 987 individuals to be included in the analysis, 577 individuals with AD and 410 healthy controls. Out of the individuals with AD, 518 had LOAD and 59 had EOAD. Age at disease onset is not known for the participants of ADNI. Therefore, for individuals that entered the study with a diagnosis of AD (N=336), disease duration was calculated as time elapsed from study enrolment. For individuals that developed AD while the study was ongoing (N=241), duration was defined as time elapsed since the first assessment in which they were classified as AD patients. The demographic characteristics of this dataset are illustrated in Table 2.5.

For healthy controls, the mean age at recruitment was 73.88, mean age at last assessment was 77.62 and the mean number of assessments was 3.36. Mean MMSE score at first assessment was 29.05 and mean MMSE score at last assessment was 28.86. 54.39% of the individuals were female. For the individuals with LOAD, the mean age at recruitment was 77.43, mean age at last assessment was 78.94 and the mean number of assessments was 3.47. Mean MMSE score at first assessment was 23.08, mean MMSE score at last assessment was 19.50 and 40.34% of the individuals were female. For the individuals with EOAD, the mean age at recruitment was 61.04, mean age at last assessment was 62.37 and the mean number of assessments was 2.12. Mean MMSE score at first assessment was 23.07, mean MMSE score at last assessment was 18.63 and 57.63% of the individuals were female.

	Mean	SD	Range
Controls			
Age at recruitment (years)	73.88	5.79	56.2-89.6
Age at last assessment (years)	77.62	6.56	60.28-97.89
Number of assessments	3.36	1.06	2-8
First MMSE	29.05	1.16	24-30
Last MMSE	28.86	1.41	22-30
Gender	Female (%)		Male (%)
	223 (54.39)		187 (45.61)
LOAD			
Age at recruitment (years)	77.43	5.99	65.08-94.45
Age at last assessment (years)	78.94	5.89	66-94.60
Number of assessments	3.47	1.11	2-9
First MMSE	23.08	3.14	2-30
Last MMSE	19.50	5.70	0-30
Gender	Female (%)		Male (%)
	209 (40.34)		309 (59.65)
EOAD			
Age at recruitment (years)	61.04	2.86	55.10-64.90
Age at last assessment (years)	62.37	3.05	55.60-67.99
Number of assessments	3.12	0.88	2-5
First MMSE	23.07	3.06	11-28
Last MMSE	18.63	6.03	2-27
Gender	Female (%)		Male (%)
	34 (57.63)		25 (42.37)

Table 2.5. Sample characteristics for the portion of ADNI that was used in this project. (MMSE: Mini Mental State Examination)

2.3. Cohort comparison

The subset of CAGRAD that had been longitudinally monitored and described in detail in Chapter 2.1.5 was used in a series of analyses that will be described in detail in Chapters 3 and

4, aiming to quantify the rate of cognitive decline in AD and explore its genetic background. The subset of ADNI described in Chapter 2.2.4 was used as a replication dataset for said analyses. A comparison of certain variables of interest between the individuals with AD included in the two datasets was first performed in order to assess the similarity of the individuals that were part of each cohort, and to examine the suitability of ADNI as a replication dataset. First, the distribution of age, age at disease onset and disease duration was compared between the two datasets using a t-test. The sex distribution was also compared using a chi squared test. All statistical analyses were conducted in R [208].

The age at recruitment was significantly higher in CAGRAD compared to ADNI (80.20 years and 76.39 years, respectively). Disease duration was also significantly higher in CAGRAD (mean 2.38 years and 1.57 years, respectively), while there was no significant difference in the age at disease onset between the cohorts (mean 74.81 years in CAGRAD and 75.46 years in ADNI). The higher age at recruitment and disease duration found in CAGRAD compared to ADNI could stem from the fact that ADNI was designed as a prospective study and enrolled more individuals with MCI than individuals that already had a diagnosis of AD [210]. As mentioned in Chapter 2.2.4, a large proportion (N=241) of the ADNI participants examined here were diagnosed with AD after recruitment. On the other hand, CAGRAD recruited individuals with AD often through healthcare services related to dementia, therefore it targeted mostly elderly individuals that had been already living with AD for some time prior to recruitment. Moreover, as explained in Chapter 2.2.4, no information regarding the age at disease onset was available for the individuals that enrolled in ADNI while after diagnosis. The disease duration for these individuals was approximated by the time elapsed between recruitment and each cognitive assessment, which is naturally going to be shorter than the actual disease duration. CAGRAD also had significantly more women participants than ADNI (67.37% and 43% respectively), still the higher proportion of women with AD seen in CAGRAD is representative of the observations on a population level that consistently show a higher AD prevalence in women compared to men [26]–[28]. Despite these differences, ADNI remained an adequate approximation of CAGRAD and the largest publicly available cohort of longitudinally assessed individuals with AD that was accessible at the time, therefore it was used as a replication dataset for the analyses in Chapters 3 and 4. However, given that the two cohorts differ in various domains, caution is advised when interpreting the results of the relevant chapters.

Chapter 3 | Modelling cognitive decline

3.1. Introduction

The severity of memory loss and the rate of deterioration are important factors to consider regarding AD, as individuals with a severe phenotype or a rapid decline are considerably more likely to require additional care resources. The ability to predict the rate of decline for AD would be beneficial to healthcare professionals, as it would allow for the construction of a more informed and personalised treatment plan. Moreover, it would provide valuable information on the prognosis to patients and their caregivers, and it would give the ability to make informed decisions regarding any care arrangements that need to be made. Most importantly, it could allow for identification of individuals that are expected to have a faster disease progression, and therefore could manifest the effects of any pharmacological interventions in a shorter timeframe, an attribute that could be particularly useful in the selection of participants for clinical trials.

Various methods of modelling and predicting disease progression and severity in dementia have been suggested. Machine learning algorithms have been previously employed to assess progression in dementia, using a wide variety of predictors, including neuroimaging data of different modalities [216], [217], amyloid positron emission tomography (PET) [216], and various neuropsychological instruments used to assess cognitive function [218]. Latent class models and mixed effects models have also previously been investigated [113], [219]. However, the majority of the studies on disease progression in dementia tend to focus on progression from MCI to AD, with few examining the cognitive decline after AD diagnosis. The main reason behind this is that there is a drive to detect dementia earlier, as none of the existing or trialled medications are effective in advanced stages of disease. However, the majority of AD diagnoses happen after the MCI stage, therefore it is important to develop prognostic tools for established AD.

In this chapter, I have explored methods that can be used to produce a metric of the rate of cognitive decline in AD that can be used in epidemiological and genetic analyses. I suggest the

use of mixed effects linear modelling for generating a quantitative measure of cognitive decline in longitudinal population-based datasets, as this type of modelling can tolerate the variability that can arise in population cohorts and the correlation that exists between measurements from the same individual [220].

3.2. Aims and objectives

The main aim of this chapter is to establish a method of quantifying and assessing the rate of cognitive decline in individuals with AD, that can subsequently be utilised in genetic analyses. The objectives of this chapter are:

- To use mixed effect linear modelling to generate a measure of cognitive decline in AD
- To validate the use of this measure in assessing cognitive decline
- To replicate this in an independent dataset

3.3. Methods

3.3.1. Mini Mental Score Examination

The Mini Mental Score Examination (MMSE) [146] was used to assess cognitive function in this analysis as it was the cognition metric most widely available in the dataset used. MMSE is the most commonly used instrument to assess cognitive function and screen individuals for dementia. MMSE is a standardised cognitive screening tool, widely utilised both in clinical and in research settings [221]. It is available in multiple languages and has been validated for use in many countries.

MMSE includes 11 questions focusing on memory, attention and language and takes around five to ten minutes to complete, making practical to use in elderly and demented individuals, while providing a thorough review of one's cognitive abilities [146]. However, it has the disadvantage of only focusing on the cognitive side of mental capacity, completely disregarding any psychological or behavioural abnormalities. The score of MMSE ranges from 0 to 30, with lower scores indicating more pronounced cognitive deficits. Scores below 25 indicate dementia, with scores between 20 and 25 being classed as mild dementia, between 10 and 19 as moderate and below 9 as severe [222]. Apart from its use in screening and diagnosis, MMSE

is also commonly used to monitor disease severity and progression in individuals diagnosed with dementia or MCI.

3.3.2. Computation of measures of cognitive decline

A subset of the CAGRAD dataset (N= 1054) that had available longitudinal data was used in this chapter, as described in Chapter 2.1.5. In order to assess cognitive decline, the change in cognition between different time points needed to be quantified. Using MMSE as a metric of cognitive function, a simple method of achieving this is dividing the difference between the MMSE score at the first and the last available assessment by the time that had elapsed between the two assessments. However, this method only takes into account the first and last assessment for each individual, and ignored the rest, therefore it was not deemed the ideal use of a dataset that provided more than two assessments for the majority of the participants.

3.3.3. Mixed effect linear modelling

Mixed effect linear modelling was subsequently selected as a method of assessing the rate of cognitive decline in this dataset as it would allow for inclusion of all available MMSE scores for each participant. Mixed effect linear models are an extension of linear modelling that allows for the inclusion both of random and of fixed effects that could influence a variable of interest. Fixed effects are parameters that vary within a large homogenous population, but do not vary within an individual, while random effects are parameters that can vary within individuals [223], however parameters can vary both within the population and within the individual and be considered as both fixed and random effects. There were multiple MMSE scores recorded for each participant in this dataset (N=1054), resulting in 2,899 observations when accounting for all assessments of each individual. The number of MMSE measurements recorded for each individual ranged from 2 to 8, with a mean of 3.17. Variability in these observations could arise either from within a group or between groups, a group being MMSE scores taken from the same participant, as it is likely that MMSE scores from the same participant will be correlated. A variety of individual-specific parameters can influence cognitive decline, like educational attainment [224] and level of physical fitness [225]. Mixed effect linear models facilitate accounting for that inter-individual variability without the data reduction that would arise if

the measurements were to be aggregated or examined separately. Moreover, mixed effect linear modelling has the added advantage of allowing for random effects that may influence the rate of decline and vary between individuals, and it is a type of modelling that has been found to tolerate well the variation in number of data points, as well as the variable time between data points that are commonly observed in population longitudinal datasets like the one used in this analysis [220]. A number of different mixed effects linear models were assessed prior to selecting the most preferable for this dataset. The models assessed are described in Table 3.1.

Model 1	MMSE ~ (1+ duration | ID)
Model 2	MMSE ~ age + (1 + duration | ID)
Model 3	MMSE ~ (1+ age | ID) + (1 + duration | ID)
Model 4	MMSE ~ age + (1 + duration | ID) + (1+ Age | ID)
Model 5	MMSE ~ age + sex + duration + (1 + duration | ID) + (1+ age | ID)
Model 6	MMSE ~ age^2 + age + sex + duration + (1 + duration | ID) + (1+ age | ID)

Table 3.1. List of linear mixed effects models examined. (1 + duration | ID) denotes that duration is considered a variable of random effect, therefore estimating a random intercept and a random slope for duration for each individual [226]. (MMSE: Mini Mental Score Examination, ID: individual)

MMSE score was the dependent variable in all the models assessed, as it was the metric of cognitive function selected for this analysis. Disease duration, defined as time elapsed between onset of AD symptoms and each cognitive assessment, was selected as the variable of interest, based on existing literature highlighting the fact that time elapsed since symptom onset affects cognitive decline more than age does in AD [114]. The random effect of disease duration on MMSE was used as a base model (Model 1, Table 3.1). Next, the inclusion of several additional independent variables was assessed. Age affects cognitive function therefore it was added in the model as a fixed effect (Model 2, Table 3.1) [102]. However, the influence of age on cognition can vary at random between individuals, so age at assessment was also added as a random effect, and subsequently both as a fixed and as a random effect (Models 3 and 4, Table 3.1). Sex and disease duration have been shown to affect cognitive decline therefore they were added as a fixed effect (Model 5, Table 3.1). Finally, age was added to the model as a quadratic term to account of the presence of a non-linear effect of age on cognitive function [102]. The

fit of each model was assessed using the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC), and by comparing the model fit over and above the previous model was tested using analysis of variance (ANOVA). The linear mixed effects models were generated using the package lme4() in R [226].

After assessing the different models, Model 5 from Table 3.1 was considered to be a satisfactory approximation of data. The random slopes for disease duration generated by the model were extracted for each individual and utilized as a measure of cognitive decline in all subsequent analyses. To establish the validity of the random slopes as a measure of cognitive decline, its correlation with the simpler decline measure described in Chapter 3.2.1 was assessed. All statistical analyses were conducted in R [208].

3.3.4. Cohort comparison

To assess the validity of the measures of cognitive decline generated, the derived rate of decline was compared between individuals with AD (N= 616) and healthy controls (N= 437) using linear regression, adjusting for age and sex. Afterwards, the controls were removed from the dataset, the model was constructed again using individuals with AD only and the random slopes were extracted and used as a new measure of cognitive decline. The rate of decline was compared between individuals with EOAD (N= 540) and LOAD (N= 76) using linear regression, adjusting for age and sex. All statistical analyses were conducted in R [208].

3.3.5. Replication

To assess the validity of the measure of cognitive decline generated, a replication of the modelling was attempted in an independent dataset. A subset of the ADNI [210] dataset (N= 987) was used for the replication, as described in Chapter 2.2.4. The mixed effects linear model that was selected in Chapter 3.3.2 was computed and the random slopes for disease duration were extracted to be used as a measure of cognitive decline. As above, the rate of decline was compared between individuals with AD (N= 577) and healthy controls (N= 410). The controls were then removed from the dataset, the model computed anew, and slopes derived and used as a measure of cognitive decline. The rate of decline was subsequently compared between individuals with EOAD (N= 59) and LOAD (N= 518). Additionally, for the ADNI dataset a

comparison between the individuals that enrolled in the study having a diagnosis of AD and the individuals that converted from MCI or healthy controls while enrolled in the study was performed, to establish whether there was a difference between the groups. Such a difference could have emerged as a result of the difference in the way disease duration was defined in these two groups. Out of the 577 individuals with AD in this dataset, were 336 recruited while having AD and 241 converted as the study was ongoing. The rate of decline in the two dataset was compared using logistic regression. All statistical analyses were conducted in R [208].

3.4. Results

3.4.1. Computation of measures of cognitive decline

Seven different methods of assessing cognitive decline were assessed. The first was a simple measure computed by dividing the difference between the first and last MMSE scores for each individual by the time that had elapsed between the recording of those scores. The six subsequent measures were generated using linear mixed effects modelling, with MMSE score being the dependent variable. After each amendment, the improvement of the model fit over and above the previous model was tested by ANOVA. The results are illustrated in Table 3.2.

	MODEL	AIC	BIC	ANOVA P-VALUE
1	MMSE ~ (1+ duration | ID)	18271	18301	
2	MMSE ~ age + (1 + duration | ID)	18122	18158	$< 2.2x10^{-16}$
3	MMSE ~ age + duration + (1 + duration | ID)	17614	17656	$< 2.2x10^{-16}$
4	MMSE ~ age + (1 + duration | ID) + (1+ age | ID)	17595	17654	$1.21x10^{-5}$
5	MMSE ~ age + sex + duration + (1 + duration | ID) + (1+ age | ID)	17578	17644	$1.26x10^{-5}$
6	MMSE ~ age^2 + age + sex + duration + (1 + duration | ID) + (1+ age | ID)	17578	17646	0.345

Table 3.2. Comparison of model fit for linear mixed effects models constructed. The p-value column indicates the improvement of the model fit when an additional predictor is added, tested using ANOVA. (BIC: Bayesian Information Criterion, AIC: Akaike Information Criterion, ANOVA: Analysis of Variance, MMSE: Mini Mental Score Examination, ID: individual)

Model 5, including age at assessment and disease duration as random and fixed effects and sex as fixed effect, was selected for assessing the rate of decline in this dataset, as the fit for this model was significantly better than for the previous ones, while the next model was not an improved fit. The decision was based after examining each model's BIC and AIC and the p-value of an ANOVA comparing the model after each amendment. The random effects of age at assessment and disease duration were included to account for individual-specific variation in cognitive decline. The fixed effects of sex, age at assessment and disease duration were significant predictors of cognitive decline (β= 0.143, p= 7.85×10^{-12}, β= -0.626, p= 0.0032, and β= -0.036, p= 0.025, respectively), as illustrated in Table 3.3.

	β	SE	p-value
duration	-0.036	0.05	0.025
sex	0.143	0.27	7.85×10^{-12}
age	-0.626	0.20	0.0032

Table 3.3. The variables included in Model 5 as fixed effects and their effect on MMSE. (β: beta coefficient, SE: standard error)

The direction of the effect indicates that younger individuals with shorter disease duration decline slower, while it also suggests that females decline faster than males. The addition of other variables like educational attainment or comorbid conditions in the model was decided against, as it could potentially lead to overfitting of the model and reduce the sample size due to a large proportion of missing data for such variables. The random slopes for disease duration derived from this model were extracted. Their distribution is shown in Figure 3.1, with the slopes for controls in yellow, the slopes for individuals with EOAD in red and the slopes for individuals with LOAD in blue. The slopes of the controls have a very narrow distribution with the mean being very close to zero, as expected by the minimal changes in MMSE observed in this group. For the individuals with EOAD and LOAD the distributions are wider and the means are lower, while the slopes also seem to follow a bimodal distribution.

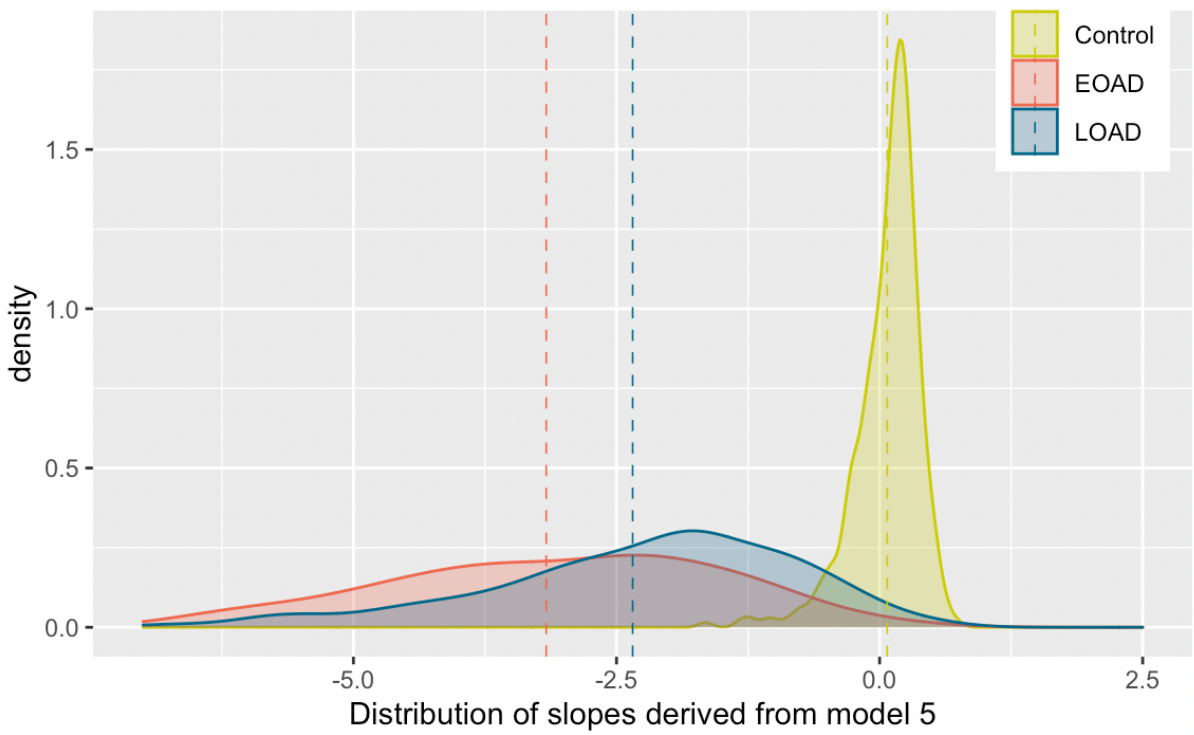


Figure 3.1. Density plot of the distribution of the random slopes derived by model 5. (EOAD: early-onset Alzheimer's disease, LOAD: late-onset Alzheimer's disease)

To assess whether the extracted slopes measure cognitive decline, their correlation with the simpler measure of decline was examined. The two measures were highly correlated (correlation coefficient = 0.34, p-value = 8.91×10^{-30}). The relationship is illustrated in Figure 3.2.

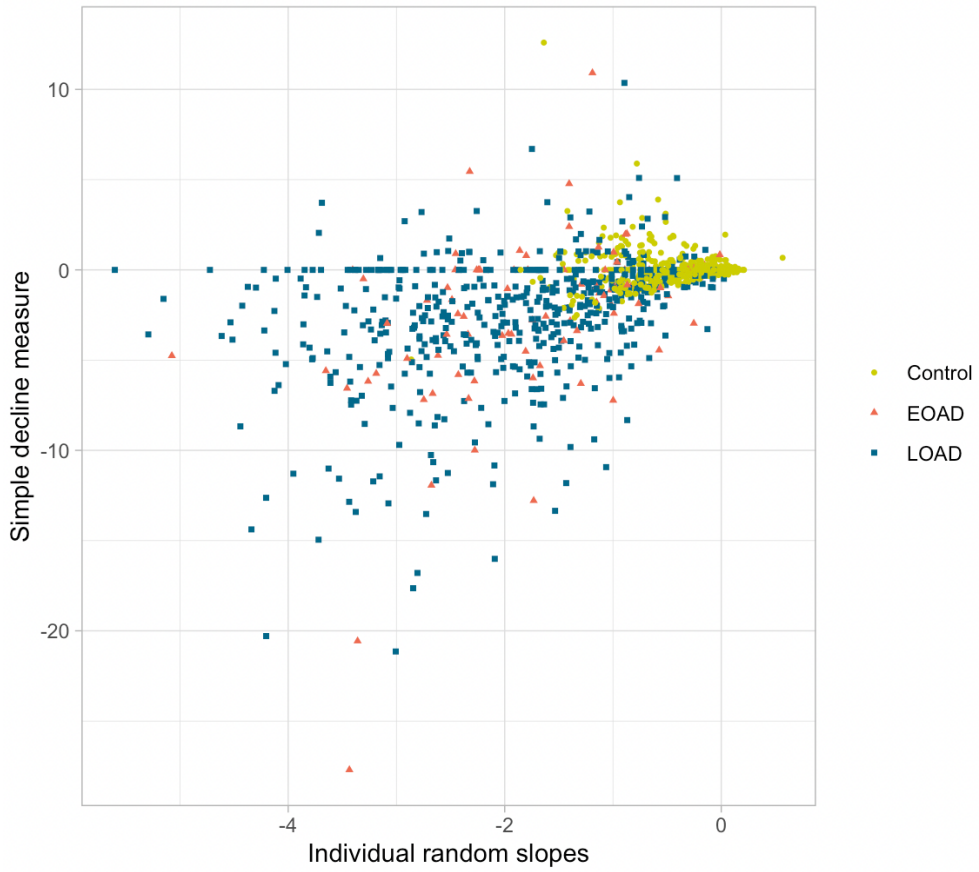


Figure 3.2. Scatterplot of the simple decline measure on the y axis and the individual random slopes derived from Model 5 om the x axis. Each point represents an individual, having a different colour depending on whether they are individuals with LOAD (blue), EOAD (red), or controls (yellow). The correlation of the two measures differed for the three groups of individuals, with the correlation coefficient being 0.29 for LOAD, 0.43 for EOAD and 0.18 for controls. (EOAD: early-onset Alzheimer's disease, LOAD: late-onset Alzheimer's disease)

Hence, the individual random slopes from model 5 are a valid measure of the rate of cognitive decline, while including more information than the simple measure. Therefore, the slopes were used to quantify cognitive decline in all subsequent analyses.

Figure 3.3 illustrates the rate of cognitive decline seen in this dataset. The dashed lines represent the random slopes and intercepts extracted for each individual from the mixed effect linear model, whereas the bold continuous lines represent the overall slope and intercept per group. Controls show only minimal cognitive decline associated with normal aging, whereas LOAD and EOAD individuals have a much steeper decline. The decline is more rapid for individuals with LOAD than with EOAD in this dataset, though this difference is not statistically significant (p-value = 0.307). The axes limits are extended above the maximum

values for MMSE and disease duration as that allows for a better visual representation of the different trajectories of the three groups.

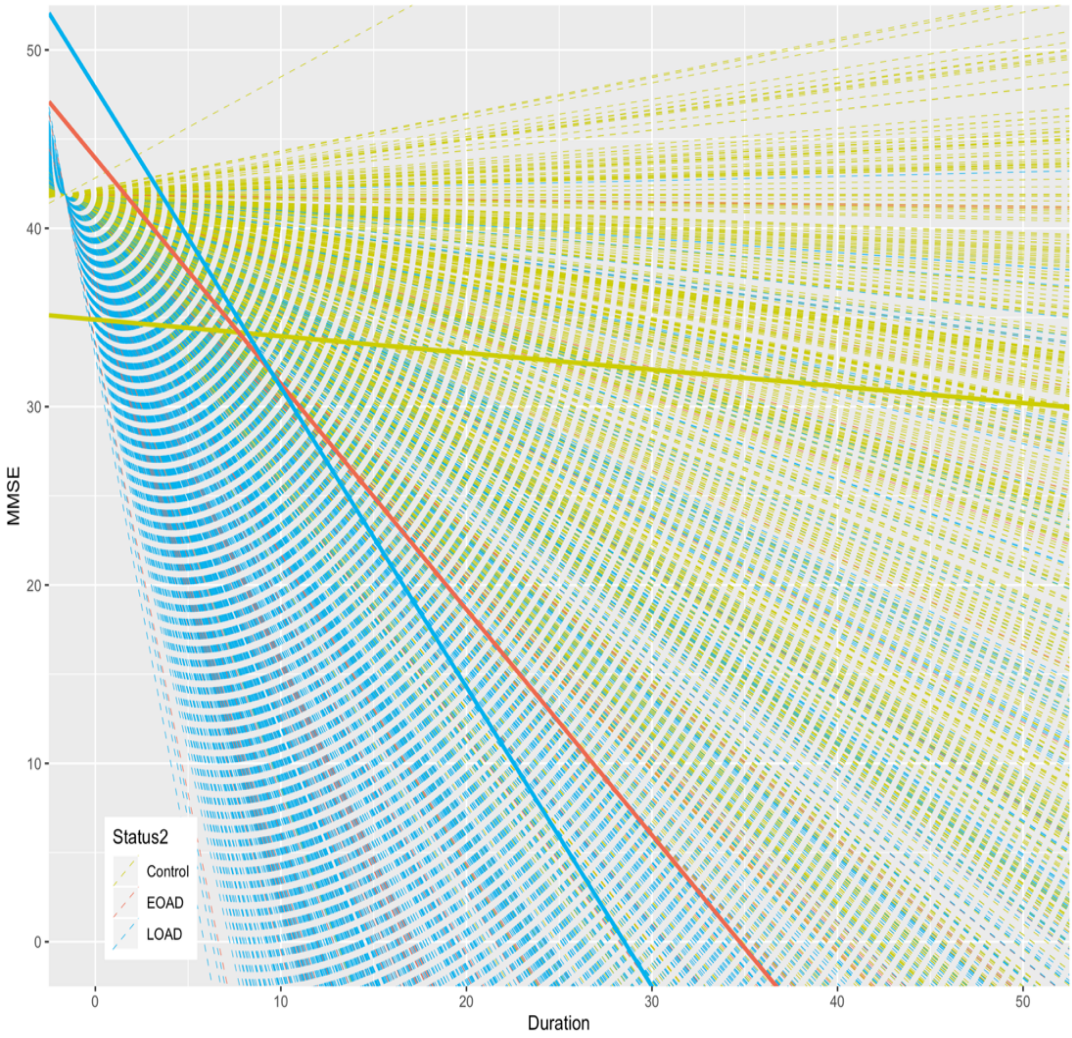


Figure 3.3. Graphical representation of the individual rate of decline. The dashed lines indicate the intercept and slope for each individual, while the bold continuous lines indicate the overall intercept and slope for each participant group. (EOAD: early-onset Alzheimer's disease, LOAD: late-onset Alzheimer's disease)

3.4.2. Cohort comparison

The rate of decline was compared between individuals with AD and controls using linear regression. The results are illustrated in Figure 3.4. As expected, individuals with AD declined significantly faster than controls (β= -2.93, p-value = 2.57×10^{-53}).

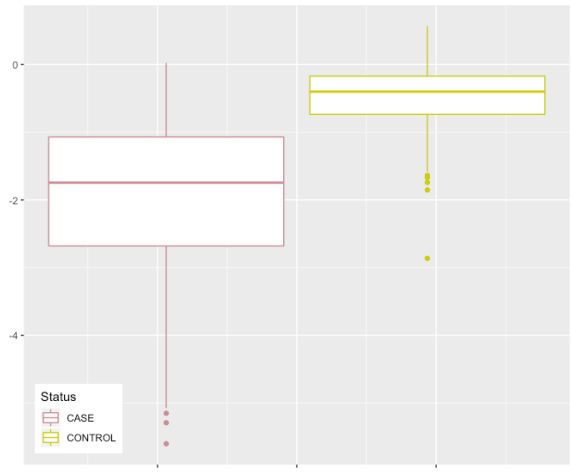


Figure 3.4. Cognitive decline in individuals with AD and controls.

Next, the controls were removed from the analysis, and rate of decline in individuals with LOAD and EOAD was compared. Interestingly, individuals with EOAD seem to decline slightly slower than individuals with LOAD, although the difference was not statistically significant (β= -0.158, p-value = 0.307). These results are illustrated in Figure 3.5.

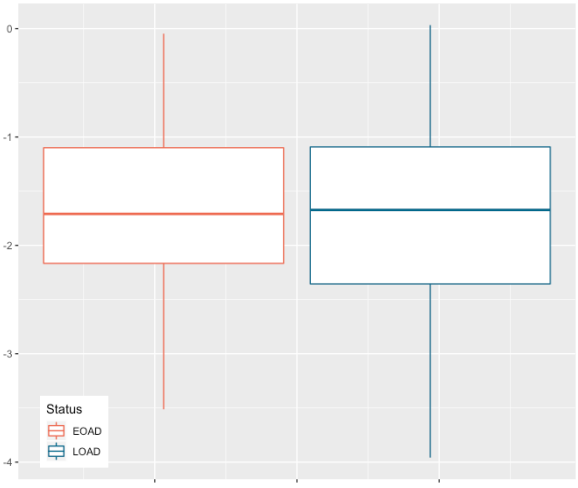


Figure 3.5. Cognitive decline in individuals with EOAD and LOAD. (EOAD: early-onset Alzheimer's disease, LOAD: late-onset Alzheimer's disease)

3.4.3. Replication

Again, a measure of decline was generated using the same mixed effects linear model (Figure 3.6) and compared between individuals with AD and controls. As above, cognitive decline was significantly faster in individuals with AD than in healthy controls (β = -4.30, p-value = 8.74×10^{-41}), as illustrated in Figure 3.7.

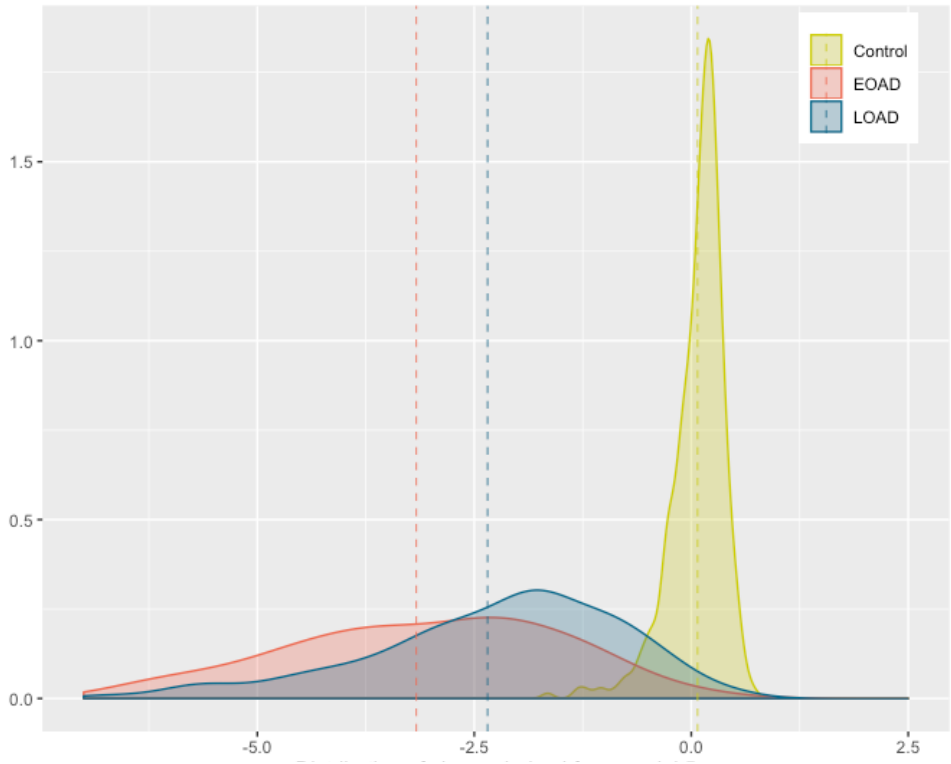


Figure 3.6. Density plot of the distribution of the random slopes derived in the replication dataset. (EOAD: early-onset Alzheimer's disease, LOAD: late-onset Alzheimer's disease)

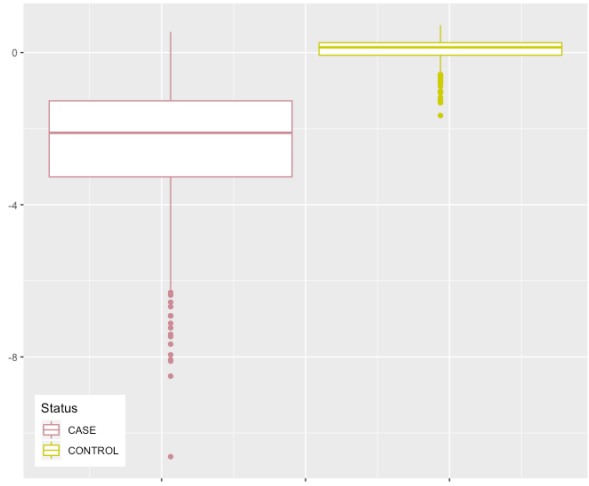


Figure 3.7. Cognitive decline in patients and controls in the replication dataset.

The controls were then removed from the dataset, and the measure of decline was computed again using AD cases only. In this dataset, cognitive decline was significantly more rapid in individuals with EOAD than individuals with LOAD, contrary to what was previously indicated

using the CAGRAD cohort (β = 0.154, p-value = 0.025). These results are illustrated in Figure 3.8.

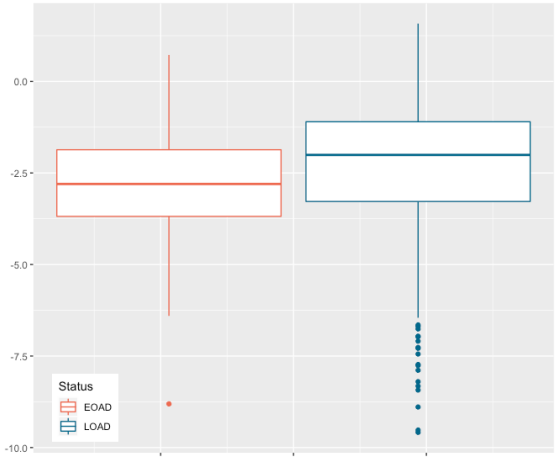


Figure 3.8. Cognitive decline in individuals with EOAD and LOAD in ADNI. (EOAD: early-onset Alzheimer's disease, LOAD: late-onset Alzheimer's disease)

Subsequently, the rate of decline was compared between the individuals that enrolled in the study while living with AD and the individuals that converted form MCI or healthy controls during the study, in order to examine if the different way of defining disease duration in these two groups could have an effect on the measure of decline generated. There was no significant difference between the two groups (β = 0.032, p-value = 0.531), indicating that any discrepancies in disease duration do not significantly affect the decline measure. The results are illustrated in Figure 3.9.

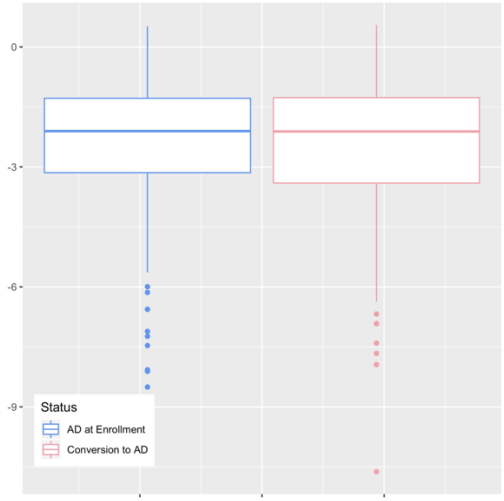


Figure 3.9. Cognitive decline in individuals that were diagnosed with AD at enrolment and converted to AD during the study in the replication dataset. (AD: Alzheimer's disease)

3.5. Discussion

The main aim of this chapter was to compute an adequate measure of assessing cognitive decline in individuals with AD using mixed effects linear modelling. Linear mixed effects models were selected as a method of assessing decline in this dataset as they allowed for including all available MMSE scores for each individual and account for inter-individual variability, and they can substantially tolerate the variance between datapoints commonly seen in population cohorts [220]. Perhaps mixed effects linear modelling was not the optimal choice for assessing decline, as it has been suggested that the cognitive and functional decline observed in AD might not follow a linear trajectory and thus non-linear models might be better equipped to represent that [116], [227]. However, the number of measurements available in CAGRAD (mean = 3.17, range from 2 to 8) would not allow for capturing such non-linearity, therefore mixed effects linear modelling was considered an adequate approximation of the rate of decline given the number of available data points. Also, mixed effects linear modelling is a method that has been used multiple times in literature to assess disease progression in AD [228], [229] as well as other neurodegenerative disorders like Huntington's disease [230] and is generally considered to adequately represent such phenotypes. MMSE score was utilized as a measure of cognitive function in this chapter. MMSE is a robust and well-established instrument for screening for dementia, commonly employed in clinical and research settings. Still, MMSE only assesses memory and cognition, while there are multiple other domains that deteriorate in AD. Alternative instruments for assessing dementia severity can create a more accurate picture of the individual's mental state, exploring areas such as executive function, emotional wellbeing, behaviour, and general daily functioning. For example, Clinical Dementia Rating (CDR) [202] and Activities of Daily Living (ADL) [203] are two instruments commonly used for monitoring dementia in a primary care setting and are increasingly being utilised in research [231]. Those instruments are perhaps preferable in assessing severity in AD, as AD is a complex disorder affecting multiple mental domains, however they make determining if a specific symptom is driving the deterioration less clear. As the aim of this chapter was to assess disease trajectory with regards to cognition, MMSE was deemed more suitable than any of the other available assessment tools. Moreover, it was the assessment most widely documented in the CAGRAD dataset, therefore using any other would have resulted in a smaller sample due to missing values.

Multiple models using MMSE as the dependent variable were assessed and the one with the best fit for this dataset out of them was selected. The model selected included age at assessment, gender and disease duration as fixed effects, and age at assessment and disease duration as random effects. Random slopes of disease duration were extracted from this model and used in further analyses. The slopes were highly correlated to a simpler measure of decline computed by assessing the difference in the MMSE scores of individuals within a given time, therefore they were considered a valid metric of cognitive decline. Initially, healthy controls as well as individuals with AD were included in the model in order to account for the cognitive decline that is caused by normal aging, over and above the decline caused by neurodegeneration, into the model. AD patients deteriorated significantly faster than controls, as was expected. However, it was later decided that as the difference in the rate of decline was considerable, it is possible that including controls into the model would positively bias the slopes derived, and result in a metric of decline that is not representative of the individuals with AD. Therefore, the controls were removed from the dataset and new slopes were derived from a model including only AD cases. Assessing individuals with EOAD and LOAD separately was also considered, but as the number of individuals with LOAD was small (N=76) there was a risk of model overfitting, and all AD cases were examined together.

The rate of decline in individuals with EOAD and LOAD was compared to examine if there is a difference in disease trajectory between the two groups. Interestingly, individuals with LOAD seem to decline slightly faster than individuals with EOAD in the CAGRAD cohort, although this result was not significant (p=0.307). Based on existing literature, there is a suggestion that patients with EOAD tend to deteriorate faster [232]–[235].However, there are studies showing no association of rate of decline with age at disease onset [236], and others showing that patients with an earlier onset decline slower [237], as shown in this chapter. A factor that could influence in this result is that average disease duration at recruitment was 6.32 years for LOAD individuals, compared to 8.74 years for EOAD. Therefore, if cognitive decline is not a linear process, it is possible that the two groups are on different phases of disease, which affect cognition differently, or even that the individuals in the EOAD group have already declined significantly at the point of recruitment, therefore they do not show much further decline as the study continues. However, since the mean MMSE at recruitment in the EOAD individuals is higher than it is in the LOAD individuals (18.48 and 16.86 respectively), it is unlikely that pronounced initial decline is the reason behind this finding. Moreover, another important

factor influencing this result is that age at symptom onset is often based on the patient's or caregiver's account and not on examination by a clinical or research professional. Therefore, the reliability of this variable is questionable. This can be problematic as the duration of the disease, defined as time from first manifestation of symptoms, is an important predictor of disease severity and progression in AD. However, most available datasets of individuals with AD, CAGRAD included, rely on self-reported age of onset, therefore that is the variable commonly used in literature. Finally, the sample size for the EOAD group was rather small (N=76), therefore any results drawn from it should be interpreted with caution.

To further examine the validity of the method of computing the metric of cognitive decline, a replication was attempted using the publicly available ADNI [210] dataset, as described in chapter 2.3. When comparing individuals with EOAD and LOAD in the ADNI cohort, individuals with EOAD showed a borderline significant accelerated decline compared to individuals with LOAD. However, as ADNI does not include information on age at disease onset, disease duration was calculated differently for this cohort than for the CAGRAD cohort, as described in chapter 2.2.4. This difference in the method of establishing the age at disease onset may account for some of the differences in results. However, the rate of decline in individuals within the ADNI cohort that were enrolled while having an AD diagnosis, and their age at disease onset is unknown and approximated, did not differ significantly from the ones that were diagnosed while the study was ongoing, and their age at disease onset is reliably recorded. Therefore, it is reasonable to assume that the minor differences in the definition of disease duration would not have a large effect on the measure of cognitive decline, and the difference between the two cohorts could be a true finding caused by different data trends present in each cohort.

3.6. Conclusions

Identifying the factors that determine how rapidly an individual with AD will deteriorate can be of great use in everyday clinical practice and clinical research. In this chapter I have attempted to compute a measure of cognitive decline in AD that could be used to explore the determinants of the rate of decline in AD. Mixed effects linear modelling of MMSE was selected as the preferred method of generating this measure. The validity of the measure was assessed by comparing it to a simpler measure of cognitive decline and by replicating the modelling in

an independent dataset. I have shown that the derived measure provides an adequate representation of the rate of cognitive decline in AD that I will subsequently be using in genetic analyses.

Chapter 4 | Genetics of cognitive decline

4.1. Introduction

The rate of cognitive decline in AD varies widely between individuals, with the factors that drive this heterogeneity remaining mostly unknown. A number of environmental risk factors for faster deterioration have been proposed, including low premorbid IQ [238], low educational attainment [239], sedentary life and stressors [240]. However, these factors alone cannot explain the vast differences in rate of decline seen between AD patients. It is possible that there are genetic factors that have an effect on the rate of decline.

The evidence for a genetic predisposition to faster decline in individuals with AD is inconclusive. Apolipoprotein E (*APOE*) ε4 allele is the strongest genetic risk factor for sporadic AD [186], and has therefore been hypothesised to be involved in the rate of disease progression, including cognitive decline. Numerous studies have examined the association of the *APOE* genotype with disease progression and cognitive decline in patients with AD. However, the results are conflicting, with some studies finding that the *APOE* ε4 allele is associated with faster progression [228], [241], [242] and other showing opposing results, with *APOE* ε4 alleles predisposing to a slower decline or having no effect at all [117], [243], [244].

Aside from *APOE*, the genetic background of disease progression in AD has also been explored at a genomic level. A number of genome-wide association studies (GWAS) of cognitive decline in AD have been reported, with two of them reporting variants of genome-wide significance [245], [246], however no variants of genome-wide or suggestive significance are common between the different studies. Polygenic risk scores (PRS) have also been utilised in order to explore the shared genetic architecture of cognitive decline in AD and AD risk, again bringing conflicting results [121], [228]. Euesden *et al.* failed to discover an association of AD PRS with cognitive decline in AD [121], while Del-Aguila *et al.* reported a nominal association between

AD PRS and AD progression [229], while both studies have utilised different approaches in measuring decline. Therefore, further exploration is required to deconvolute these findings, and shed some light into the genetic aetiology of cognitive decline in AD.

In this chapter, I have attempted to investigate the presence of a genetic background in cognitive decline in AD, using the measure of decline described in chapter 3.

4.2. Aims and objectives

The main aim of this chapter is to explore the genetic architecture of cognitive decline in AD and uncover genetic variants that predispose to more rapid decline using a variety of methods. The objectives of this chapter are:

- To assess the effect of the *APOE* genotype on cognitive decline in AD
- To perform a GWAS of cognitive decline
- To perform a genetic pathway analysis of cognitive decline
- To generate a PRS of decline and test its association with cognitive decline in an independent dataset
- To examine the association of a PRS of AD with cognitive decline

4.3. Methods

4.3.1. Datasets

A subset of individuals with AD from the CAGRAD dataset that had available longitudinal data (N= 616) was used in this chapter, as described in Chapter 2.1.5. Out of these individuals, 575 had available *APOE* genotypes, while 537 had available full genotype data. A subset of individuals with AD from ADNI [210] (N= 577), as described in Chapter 2.2.4, was used to replicate the analyses described in this chapter. Out of these individuals, 498 had available *APOE* genotypes and 373 had available full genotype data. The datasets are described in Figure 4.1. The measure of cognitive decline described in Chapter 3.3.3 was used in all the analyses in this chapter to quantify cognitive decline.

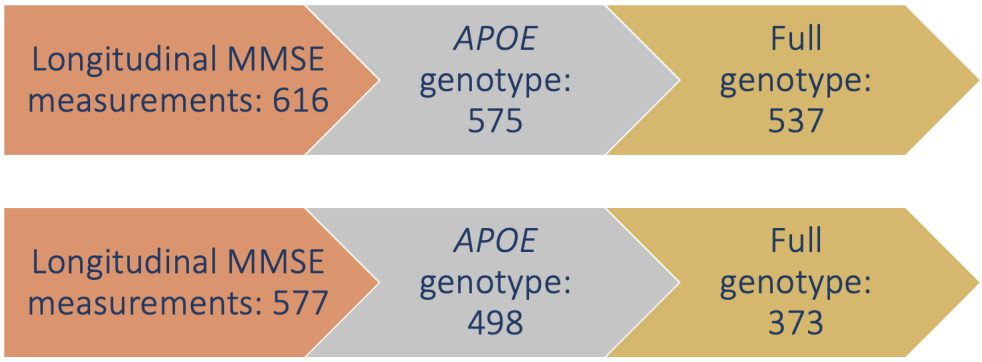


Figure 4.1. Number of individuals in CAGRAD (top) and ADNI (bottom) used in the analyses of this chapter. (MMSE: mini-mental state examination)

4.3.2. APOE

The association of the number of *APOE* ε4 and ε2 alleles with cognitive decline in AD was assessed using linear regression. The *APOE* genotype acquisition for CAGRAD is described in chapter 2.1.3. 575 individuals had available *APOE* genotype information and were included in the analysis. A replication of this analysis was attempted using ADNI dataset, as described in chapter 2.2.4. The *APOE* genotype acquisition for ADNI is described in chapter 2.2.3. 493 individuals had available *APOE* genotype information and were included in the analysis. The statistical analyses were performed in R [208].

4.3.3. GWAS of decline

4.3.3.1. CAGRAD

The genotyping of the CAGRAD cohort is described in chapter 2.1.3. 537 individuals had available genotype data and were included in this analysis. Quality control was performed before carrying onto the GWAS. Standard GWAS quality control protocols were observed. The dataset was screened for individuals with missing genotype rate of > 0.05, cryptic relatedness (identity by descent estimate ≥ 0.2) and individuals of non-European ancestry. After individual quality control, 529 individuals remained. To assess the population structure within the datasets, principal component analysis (PCA) was performed. No genetic outliers were identified. 7,581,719 single nucleotide polymorphisms (SNPs) were available for CAGRAD after

imputation. SNPs were screened for a missing genotype rate of > 0.05, a minor allele frequency (MAF) of < 0.01 and Hardy-Weinberg P < 10^{-6}. After SNP quality control, 5,984,522 SNPs remained. SNPs were tested for association with the rate of decline using linear regression, assuming an additive model, using the first four principal components identified in the PCA as covariates. The GWAS and quality control were performed in PLINK [247].

4.3.3.2. ADNI

A replication of this analysis was attempted using ADNI dataset, as described in chapter 2.3.4. The genotyping of ADNI is described in chapter 2.2.3. 373 individuals had available genotype data and were included in this analysis. Again, the same GWAS quality control protocols as above were observed. The dataset was screened for individuals with missing genotype rate of > 0.05, cryptic relatedness (identity by descent estimate ≥ 0.2) and individuals of non-European ancestry. After individual quality control, 373 individuals remained. To assess the population structure within the datasets, principal component analysis (PCA) was performed. No genetic outliers were identified. 8,718,573 single nucleotide polymorphisms (SNPs) were available for ADNI. SNPs were screened for a missing genotype rate of > 0.05, a MAF of < 0.05 and Hardy-Weinberg P < 10^{-6}. After SNP quality control, 6,137,396 SNPs remained. SNPs were tested for association with the rate of decline using linear regression, assuming an additive model, using the first ten principal components identified in the PCA as covariates. The GWAS and quality control were performed in PLINK [247].

4.3.3.3 GWAS meta-analysis

To increase the power of the rate of decline GWAS, a GWAS meta-analysis was performed using the METAL software [248]. METAL combines test statistics and p-values across multiple GWAS, taking direction of effect and sample size into account. It results in no less efficiency than analysing combined genotype data from all included studies, while overcoming the problems that can arise when combining genotype data, like differences in ethnicity or phenotype distribution, and is considerably less computationally demanding [248]. The sample size analysis scheme was used, which combines the variant p-values across studies while accounting for the individual study sample size and the direction of effect for each variant. Only variants present in both studies were included in the meta-analysis, and rare variants

were excluded. A threshold of MAF < 0.01 was used for CAGRAD and MAF < 0.05 for ADNI, as the size of the ADNI dataset was smaller, resulting in 4,643,781 SNPs.

4.3.4. Pathway analysis of decline

In order to identify biological pathways implicated in cognitive decline a pathway analysis was performed. Pathway analysis examines predefined gene sets associated with biological pathways for enrichment of genomic signal. This provides some insight into the biological mechanisms through which the genetic risk factors could lead to the development of the phenotype of interest, thus facilitating the interpretation of the observed results of genomic analyses. Pathway analysis can therefore be utilised for gaining some understanding of the molecular mechanisms behind a phenotype. An added benefit of pathway analysis is that it combines the signals from numerous related genetic variants, therefore allowing the discovery of associations in studies that lack the statistical power to detect single variant associations at the genome-wide significant level. Pathways are defined as sets of genes that are related to each other in terms of biological processes at the cellular or molecular level with specific start and end points. They can be generated by the researcher using specific gene networks that they are interested in or alternatively extracted from existing pathway databases. Pathway databases are manually curated by biologists based on existing literature and cross-referenced with other pathway databases and updated and corrected periodically.

The pathway analysis was performed using the MAGMA software [249]. Pathway analysis with MAGMA is performed in two consecutive steps. First, a gene-based test is performed to generate intermediate gene-based statistics, aggregating the effects of all the genetic variants within a gene, while accounting for the number of SNPs in each gene and the LD between the SNPs. The gene-based statistics that are generated are then used to determine the association of pathways with the phenotype of interest. Pathway analyses with MAGMA can be based on GWAS summary statistics or individual genotype data. In this analysis, the summary statistics of the GWAS meta-analysis described in chapter 4.3.2.1 were used. First, all the p -values from SNPs within a gene region, with an optional annotation window around the region, are combined and a single gene-level p-value is computed. If GWAS summary statistics are used as input data, alternative raw genotype data are required as a reference panel for LD structure. Here, a total of 18,514 genes were used for SNP annotation, using a window of 35kb before

and 10kb after the transcription region of each gene (as in Kunkle *et al.* [186]) , and raw genotype data from the European population of 1000 Genomes Project [250] were set as a reference for LD. The genetic locations used for annotation were provided by Dr Emily Baker. The gene-based statistics were subsequently used in a competitive test of gene set association. The competitive approach compares the association of the genes within a pathway to that of the genes outside the pathway with the phenotype of interest, as opposed to the self-contained approach that only examines the genes within a pathway and compares their association with the phenotype to the null hypothesis of no association. A competitive approach was selected as it is considered more reliable when exploring the biological background of a trait with high polygenicity [251]. Total of 10,271 gene sets from GeneOntology (GO) [252], Kyoto Encyclopaedia of Genes and Genomes (KEGG) [253], Pathway Interaction Database (PID) [254] and REACTOME [255] were used in the gene set analysis, as defined in Bellenguez *et al.*, 2020 [198]. The use of these gene sets was decided upon in order to allow for comparisons of the results of this pathway analysis with the pathways implicated in AD risk as described in Bellenguez *et al.* [198].

4.3.5. PRS analysis

The presence of a polygenic architecture in the rate of cognitive decline was assessed using polygenic risk score (PRS) analysis. PRS is an advantageous method of assessing the genetic aetiology of complex disorders that occur as a result of a combination of multiple genetic variants of small effect that are spread throughout the genome. PRS are commonly used to predict the likelihood of developing a phenotype using genetic data but can also be useful in determining if two different traits share some polygenic heritability. They allow for estimating how an individual's combination of genetic variants influences their risk of manifesting a phenotype of interest. GWAS provide an effect size for every variant examined, signifying the strength of the association of that variant with the phenotype of interest, usually either a beta coefficient or an odds ratio. In a PRS analysis, the number of risk variants present in the genome of an individual weighted by that effect size are added, resulting in a cumulative score:

$\hat{S} = \sum_{j=1}^{m} X_j \hat{\beta}_j$, where S is the PRS of the individual, m is the total number of independent SNPs included in the score, X is the number of risk alleles of SNP j in the genotype of the

individual and β is the effect size of SNP j. The association of this score with the phenotype of interest can then be assessed using regression analysis. The method followed to generate PRS was the one described by the International Schizophrenia Consortium [172]. It requires two independent datasets, a training dataset for which only GWAS summary statistics are required, and a validation dataset with individual genotype data. The effect sizes and p-values of each SNP are extracted from the training dataset and used to generate a PRS on the validation dataset.

A key consideration in PRS analyses is the number of variants that should be included. Clumping and thresholding are the two main methods utilised to determine the SNPs that should be included in the PRS. SNPs that are situated in close proximity are more likely to be in linkage disequilibrium (LD), therefore they do not offer individual predictive power. Clumping is a form of informed LD pruning, that removes SNPs in high LD while retaining the variant with the lowest p-value in the training dataset. Clumping will also retain multiple SNPs within a region if they have differing effect sizes. Thresholding refers to only including SNPs below a set significance threshold (e.g., p-value < 0.05). Since the optimal significance threshold for including SNPs in the PRS cannot be determined a priori, a range of inclusion thresholds are usually assessed and the association of the resulting PRS with the phenotype of interest is examined. There are alternative methods to clumping and thresholding, like LDpred, that has been shown to achieve higher prediction accuracy when used on large datasets [256]. However, a standard clumping and thresholding approach was utilised in this thesis as it is the most widely used variant selection method in PRS analyses and thus would provide results that could be easily comparable to existing literature on the utility of PRS in assessing cognitive decline in AD.

4.3.5.1. PRS of decline

First, a PRS of cognitive decline was generated and tested for association with cognitive decline in an independent dataset. The training dataset used here was CAGRAD, and ADNI was the validation dataset. The two datasets are completely independent. The PRS were generated using PLINK [247]. The common SNPs between the two datasets were identified and extracted. 4,643,781 SNPs were common between the datasets. LD clumping was performed in ADNI, using the individual SNP data from the CAGRAD GWAS. The clumping parameters used were

an r^2>0.1 and a physical distance threshold of 1MB. After clumping, 155,171 SNPs remained. Weighted PRS were computed for seven different significance thresholds (0.5, 0.3, 0.1, 0.05, 0.01, 0.001, 0.0001), using the number of risk alleles for each individual adjusted for the effect size of the allele as found in the training dataset. The number of SNPs included in the PRS for each threshold are shown in Table 4.1.

Inclusion threshold	N SNPs
0.5	116,399
0.3	85,754
0.1	39,316
0.05	23,040
0.01	6,160
0.001	790
0.0001	85

Table 4.1. Number of SNPs included for each p-value threshold assessed. (SNP: single nucleotide polymorphism)

The association between the PRS and the rate of cognitive decline was tested using linear regression, controlling for age, sex and the ten first principal components identified in ADNI, to account for population stratification. The regression analyses were performed in R [208].

4.3.5.2. PRS of AD

Next, PRS of AD were generated and their association with the rate of cognitive decline investigated, to examine the presence of a common genetic architecture between AD and the rate of cognitive decline in AD. The GWAS published by Kunkle *et al.* (including 35,274 individuals with AD and 59,163 healthy controls) [186] was used as a training dataset and CAGRAD as a validation dataset. A subset of CAGRAD was included in the training dataset, as part of GERAD [69]. Therefore, the summary statistics from this publication after removing the GERAD individuals were used in this analysis, provided by Dr Ganna Leonenko. PRS were generated using PLINK [247]. The common SNPs between the two datasets were identified and extracted. There were 5,066,123 SNPs in common. LD clumping was performed in CAGRAD, using the SNP association information from the AD GWAS. The clumping parameters used were an r^2>0.1 and a physical distance threshold of 1MB. After clumping, 136,283 SNPs remained. Weighted PRS were computed for seven different significance thresholds (0.5, 0.3, 0.1, 0.05,

0.01, 0.001, $5x10^{-8}$), using the number of risk alleles for each individual adjusted for the effect size of the allele as found in the training dataset. The number of SNPs included in the PRS for each threshold are shown in Table 4.2.

Inclusion threshold	N SNPs
0.5	100,863
0.3	73,771
0.1	34,007
0.05	19,973
0.01	5,438
0.001	796
$5x10^{-8}$	21

Table 4.2. Number of SNPs included for each p-value threshold assessed. (SNP: single nucleotide polymorphism, N: number)

The association between the PRS and the rate of cognitive decline was tested using linear regression, controlling for age, sex and the four first principal components identified in CAGRAD, to account for population stratification. The regression analyses were performed in R [208].

4.4. Results

4.4.1. APOE

The association of the number of *APOE* and alleles with cognitive decline was assessed using linear regression. There was no significant association between the rate of decline and either *APOE* ε4 or ε2 allele both in CAGRAD (β = 0.116, p-value= 0.938 and β = -0.003, p-value= 0.423 respectively) and in ADNI (β= -0.044, p-value= 0.689 and β= 0.633, p-value= 0.052, respectively). While not reaching statistical significance, there is a suggestion of a protective effect of APOE ε2 allele in ADNI. The results are illustrated in Figures 4.5- 4.5 and summarised in Table 4.3

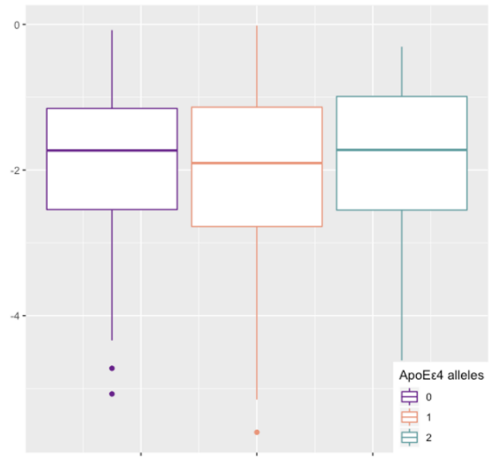


Figure 4.2. Cognitive decline by number of APOE ε4 alleles in CAGRAD.

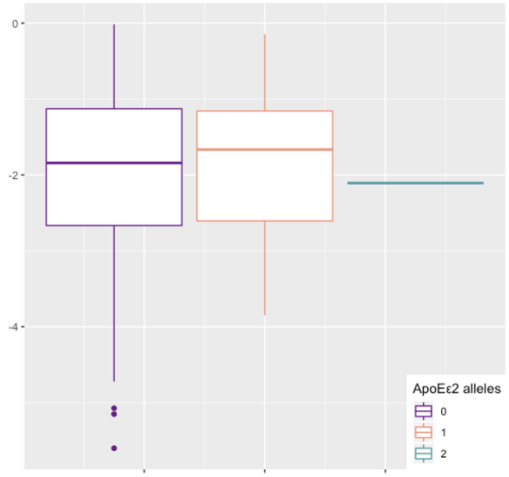


Figure 4.3. Cognitive decline by number of *APOE* ε2 alleles in CAGRAD.

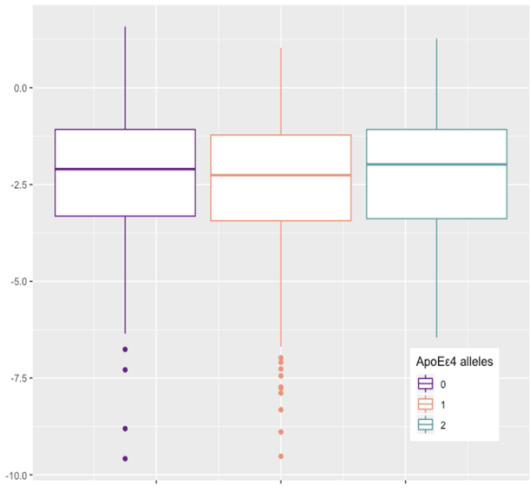


Figure 4.4. Cognitive decline by number of *APOE* ε4 alleles in ADNI.

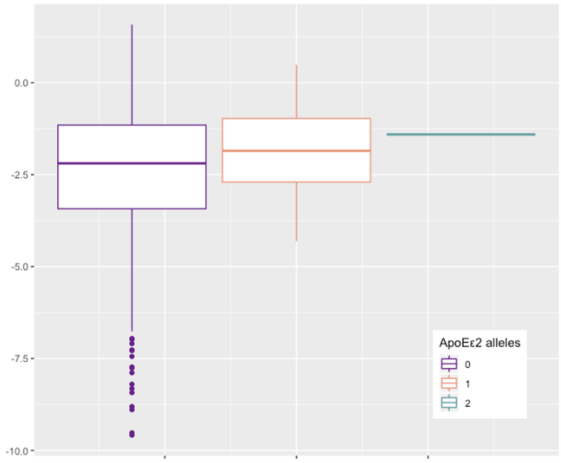


Figure 4.5. Cognitive decline by number of *APOE* ε2 alleles in ADNI.

.

Dataset	*APOE* ε2		*APOE* ε4	
	β	p-value	β	p-value
CAGRAD	0.116	0.938	-0.003	0.423
ADNI	0.633	0.052	-0.044	0.689

Table 4.3. Association of *APOE* genotype with cognitive decline for both cohorts.

4.4.2. GWAS

4.4.2.1. CAGRAD

A quantile-quantile (Q-Q) plot of the CAGRAD GWAS is shown in Figure 4.6. There is no obvious inflation in the dataset and the inflation factor is λ= 0.993.

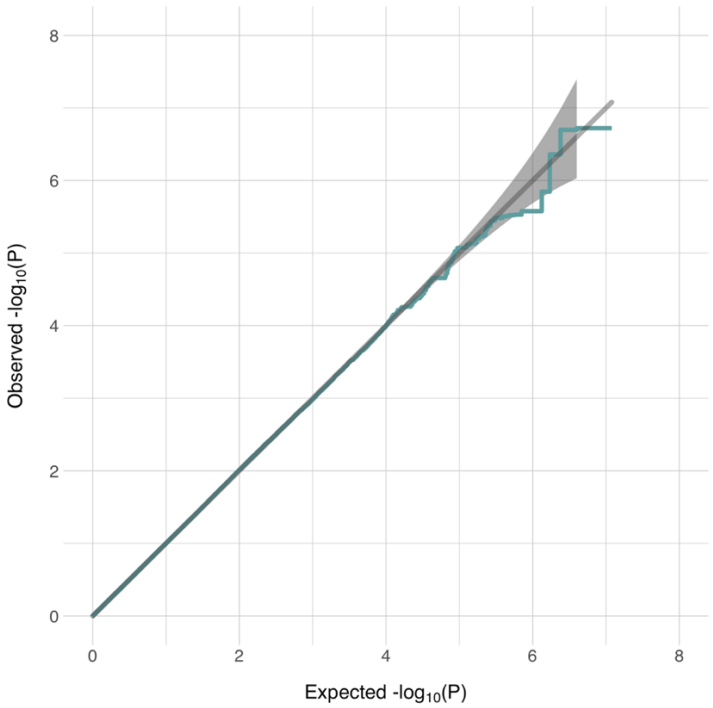


Figure 4.6. Q-Q plot of the expected versus the observed p-values of the CAGRAD GWAS. (Q-Q: quantile-quantile, GWAS: genome-wide association study)

No variants exceeded the threshold for genome-wide significance in this analysis (p-value < $5x10^{-8}$). However, three variants in chromosome 10 exceeded the suggestive significance threshold of p-value $<10^{-6}$ (rs17156614 with p-value = $1.90x10^{-7}$, rs78705533 with p-value = $2.01x10^{-7}$ and rs75302704 with p-value = $4.37x10^{-7}$). None of these variants are situated close to genes that have been previously implicated in AD risk. As expected from chapter 4.4.1, no variants within the *APOE* region (19:44905791 - 19:44909393) were associated with decline in this GWAS, with the lowest p-value in the region being 0.23. A Manhattan plot is shown in Figure 4.7.

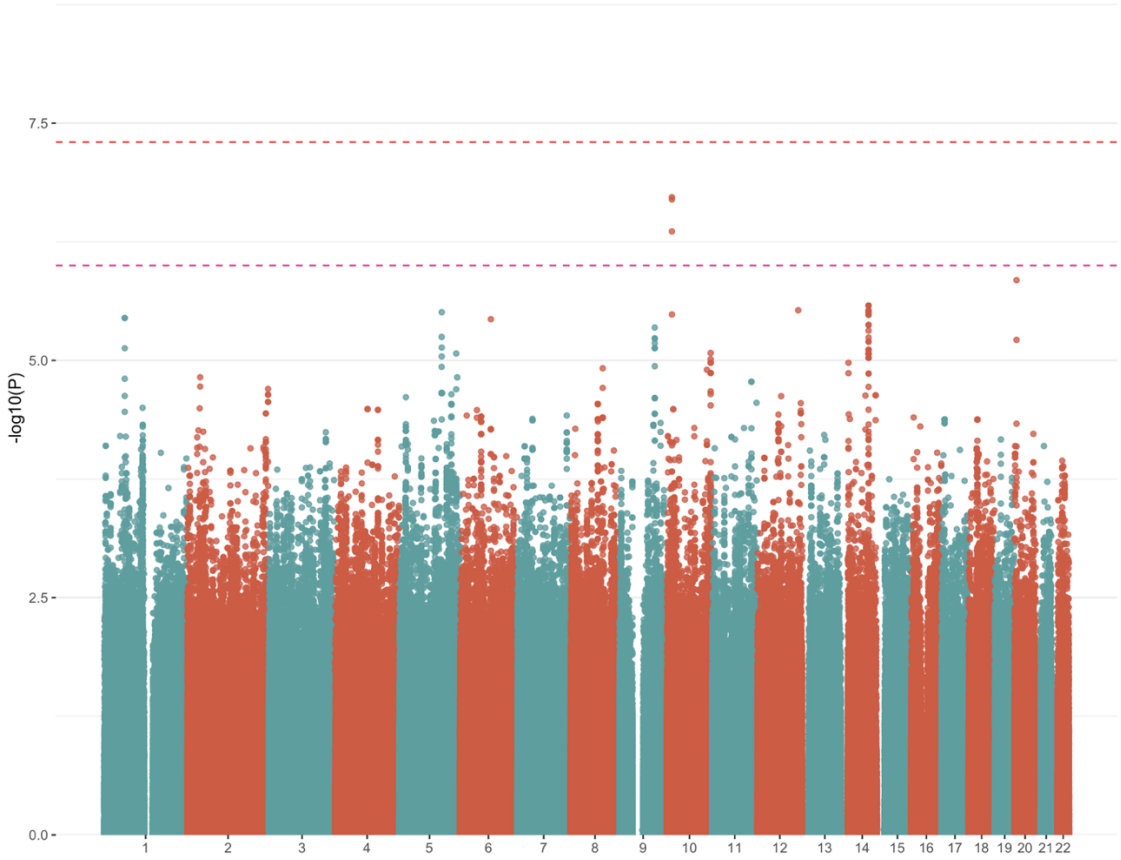


Figure 4.7. Manhattan plot of the CAGRAD GWAS. (GWAS: genome-wide association study)

4.4.2.2. ADNI

The Q-Q plot of the ADNI GWAS is illustrated in Figure 4.8. There seems to be some inflation for p-values lower than 10^{-6} and the genomic inflation factor is $\lambda = 1.001$. A Manhattan plot is shown in Figure 4.9.

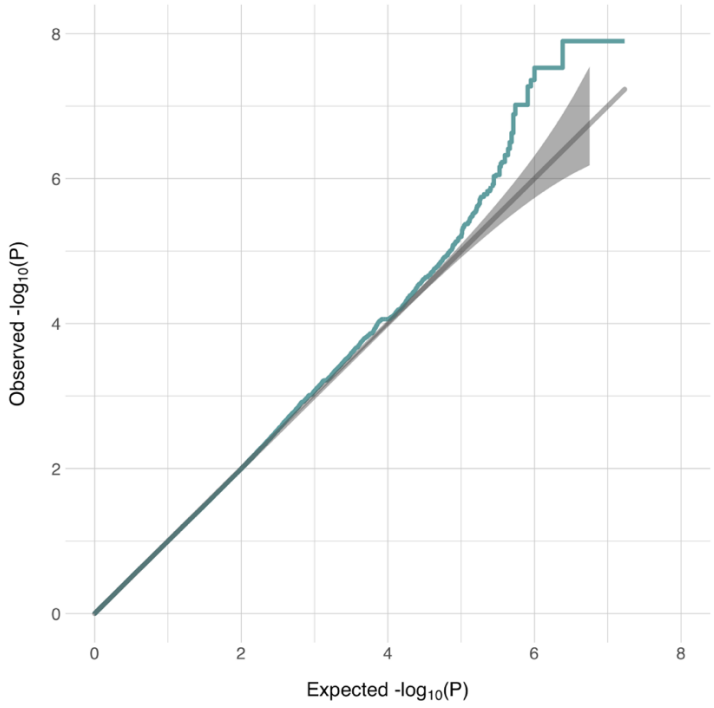


Figure 4.8. Q-Q plot of the expected versus the observed p-values of the ADNI GWAS. (Q-Q: quantile-quantile, GWAS: genome-wide association study)

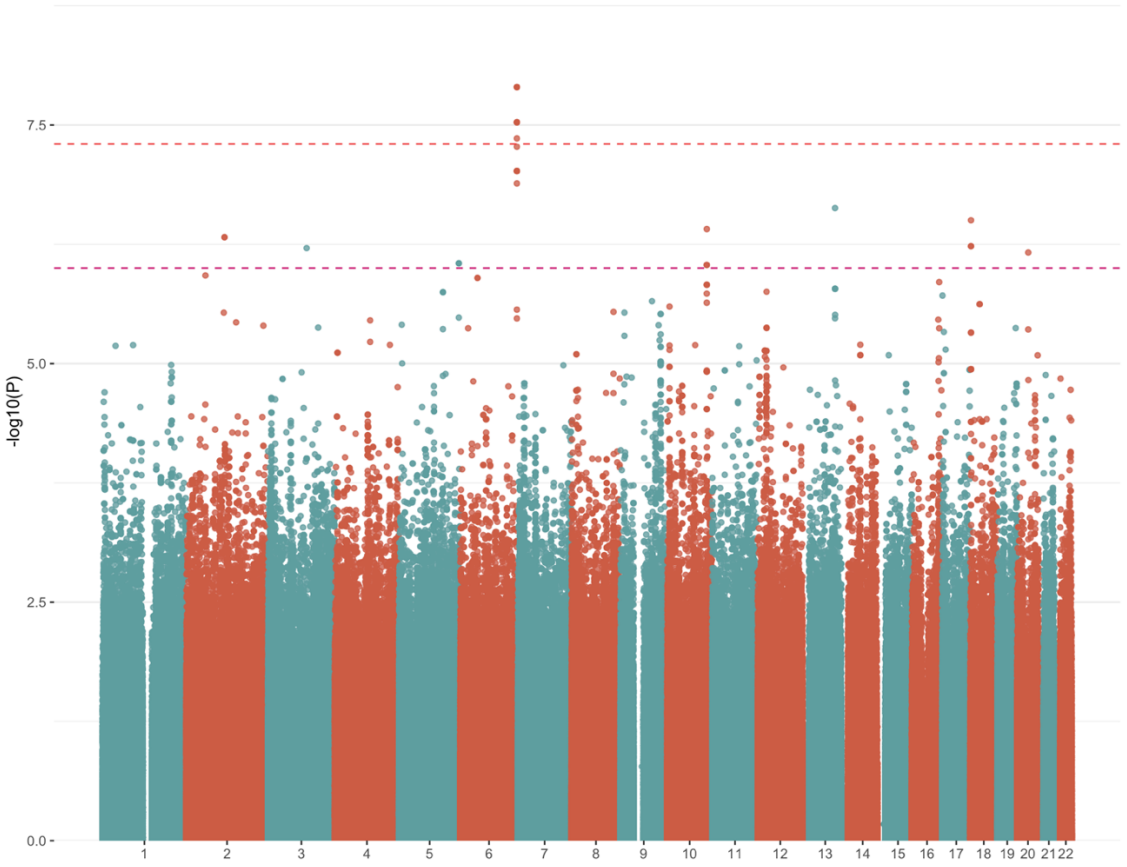


Figure 4.9. Manhattan plot of the ADNI GWAS. (GWAS: genome-wide association study)

Nine variants in chromosome six exceeded the threshold for genome-wide significance (rs80239946 with p-value = 1.27×10^{-8}, rs41266319 with p-value = 1.27×10^{-8}, rs73028356 with p-value = 1.27×10^{-8}, rs3807062 with p-value = 2.93×10^{-8} , rs73028322 with p-value = 2.93×10^{-8}, rs73028335 with p-value = 2.93×10^{-8}, rs58365143 with p-value = 2.93×10^{-8}, rs73028343 with p-value = 2.93×10^{-8} and rs749510 with p-value = 4.68×10^{-8}). None of the variants showed a suggestive association to cognitive decline in the CAGRAD GWAS, with the lowest p-value being 0.151. As expected from chapter 4.4.1, no variants within the APOE region were associated with decline, with the lowest p-value in the region being 0.34 in this GWAS. The region in chromosome six containing the nine significantly associated variants is illustrated in detail in Figure 4.10. The six variants are within the *FRMD1* gene region.

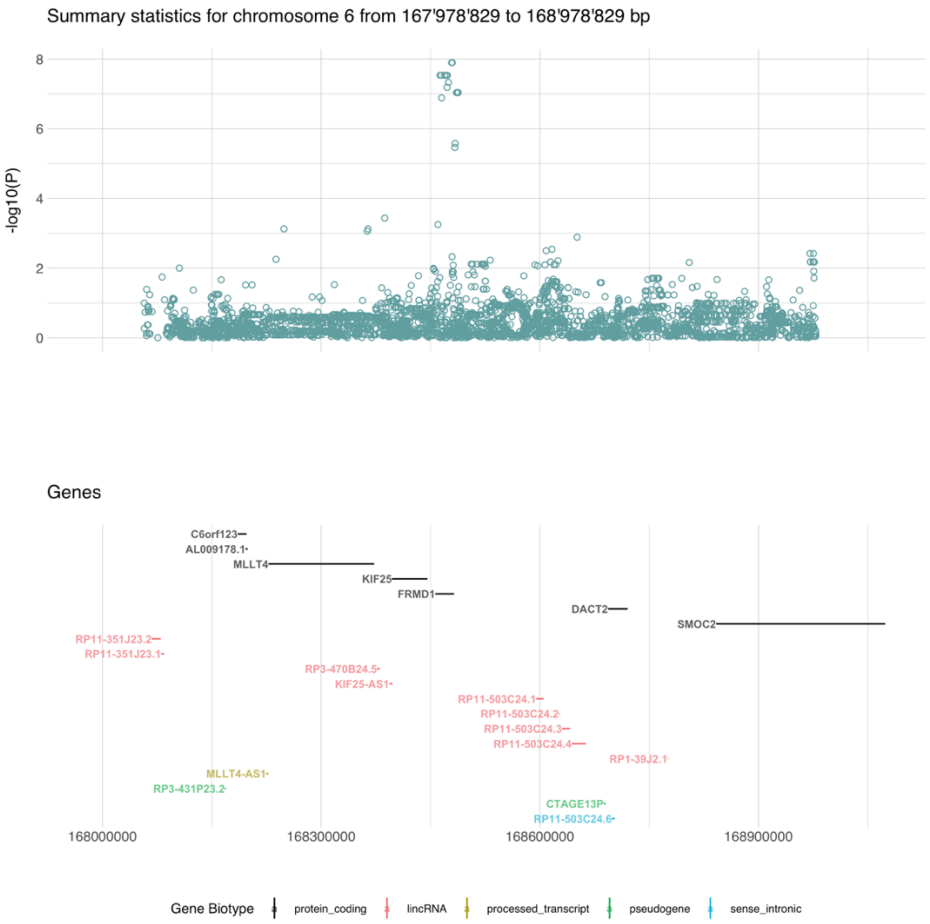


Figure 4.10. The region in chromosome six containing the nine variants that were significantly associated with cognitive decline in the ADNI GWAS. (GWAS: genome-wide association study)

4.4.2.3. Meta-analysis

A Q-Q plot of the GWAS meta-analysis (N= 902) is shown in Figure 4.11. There is no obvious inflation in the dataset and the genomic inflation factor is λ= 0.987.

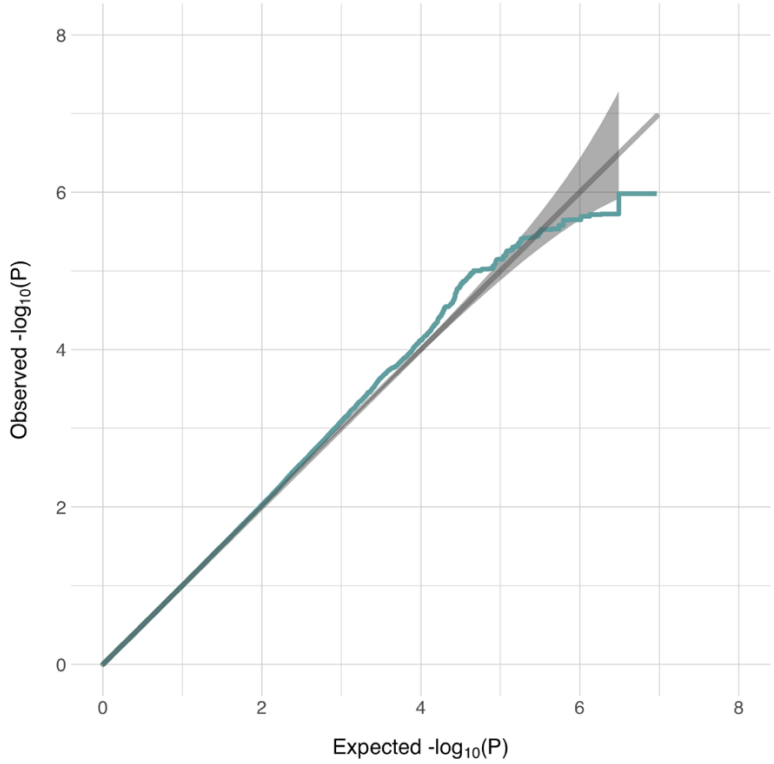


Figure 4.11. Q-Q plot of the expected versus the observed p-values of the GWAS meta-analysis. (Q-Q: quantile-quantile, GWAS: genome-wide association study)

No variant exceeded genome-wide significance in this meta-analysis, and no variant exceeded the suggestive significance threshold of p-value < 10^{-6} either. The most significant association was with in chromosome one (rs1286224 with p-value = 1.04×10^{-6}). A Manhattan plot is shown in Figure 4.12 and a list of the 10 independent SNPs with the strongest association to cognitive decline is shown in Table 4.4.

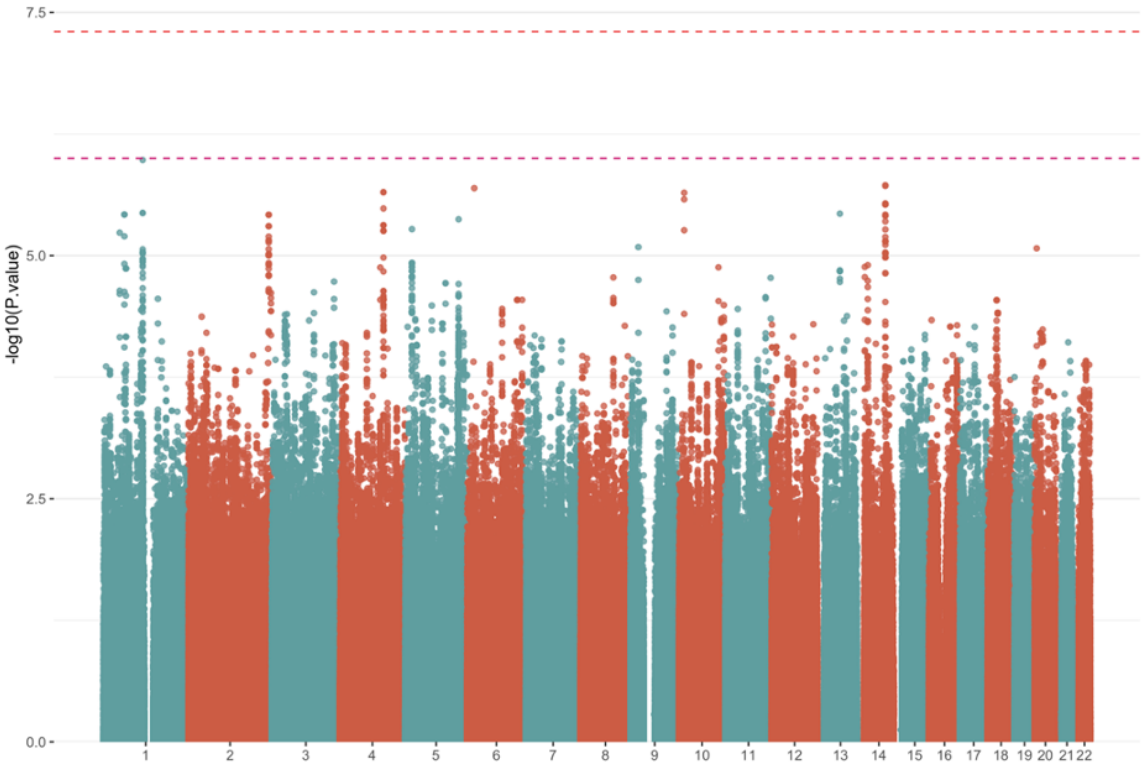


Figure 4.12. Manhattan plot of the GWAS meta-analysis. (GWAS: genome-wide association study

	Meta-analysis			CAGRAD		ADNI	
SNP	β	p	Direction	β	p	β	p
rs1286224	0.332	$1.04x10^{-6}$	--	-0.310	$3.16x10^{-5}$	-0.553	0.001
rs72697637	0.333	$1.89x10^{-6}$	--	-0.360	$3.08x10^{-6}$	-0.193	0.269
rs13204257	0.560	$2.02x10^{-6}$	++	0.532	$3.81x10^{-5}$	0.711	0.018
rs113548462	-0.463	$2.23x10^{-6}$	--	-0.449	$3.21x10^{-5}$	-0.542	0.027
rs78705533	-0.546	$2.26x10^{-6}$	+-	0.661	$2.01x10^{-7}$	-0.088	0.766
rs7323430	-0.382	$3.71x10^{-6}$	--	-0.364	$6.11x10^{-5}$	-0.378	0.228
rs1286830	-0.321	$3.79x10^{-6}$	--	-0.352	$3.58x10^{-6}$	-0.281	0.261
rs12694926	0.257	$3.81x10^{6}$	++	0.251	$3.62x10^{-5}$	0.099	0.402
rs6556570	0.270	$4.24x10^{6}$	++	0.393	$3.00x10^{-6}$	-0.166	0.389
rs16888141	0.449	$5.36x10^{-6}$	++	4.66	$2.44x10^{-5}$	0.298	0.373

Table 4.4. List of the ten variants with the strongest association in the GWAS meta-analysis, as discovered in the two independent GWAS. (GWAS: genome-wide association study, SNP: single nucleotide polymorphism, β: beta coefficient, p: p-value)

4.4.3. Pathway analysis

A gene-based test and pathway analysis were performed, using the GWAS meta-analysis results. No gene exceeded the commonly accepted significance threshold for gene-based testing (p-value < 2.5×10^{-6} [257]) in this analysis. The genes with the strongest association with rate of decline were *SGCD* and *KCNIP1* (p-values 1.46×10^{-5} and 7.35×10^{-5} respectively). The 10 genes that had the strongest association with cognitive decline at the gene-based test are illustrated in Table 4.5.

Gene name	Chromosome	N SNPs	P-value	Top SNP	Top SNP p-value
SGCD	5	1088	1.46×10^{-5}	rs6556570	2.24×10^{-6}
KCNIP1	5	823	7.35×10^{-5}	rs13177298	6.22×10^{-5}
SPATA32	17	12	1.02×10^{-4}	rs7210761	1.93×10^{-4}
CDCH	3	37	1.15×10^{-4}	rs6599013	9.29×10^{-5}
SHLD1	20	159	1.35×10^{-4}	rs1699231	8.50×10^{-5}
SAMD7	3	30	1.42×10^{-4}	rs6808802	8.30×10^{-5}
MAST4	5	933	2.01×10^{-4}	rs418015	2.34×10^{-4}
ZFAND1	8	34	2.01×10^{-4}	rs73695005	7.52×10^{-4}
MAS1	6	16	2.13×10^{-4}	rs2092583	1.80×10^{-4}
MPHOSPH6	16	143	3.00×10^{-4}	rs35839192	1.84×10^{-4}

Table 4.5. List of genes with the strongest association in the gene-based test. The SNP within each gene that had the strongest association with cognitive decline in the GWAS meta-analysis is also included. (N: number, SNP: single nucleotide polymorphism)

Subsequently, 10,271 gene sets from the GO, KEGG, PID and REACTOME databases, as described in Bellenguez *et al.* [198], were used to perform a pathway analysis. There is no commonly accepted significance threshold to correct the pathway analyses. After Bonferroni correction, the threshold for significance was p-value < 4.87×10^{-6}. No pathway exceeded the significance threshold in the pathway analysis. The 10 pathways with the strongest association with cognitive decline are illustrated in Table 4.6 and Figure 4.13.

Gene set name	N genes	p-value
GO:0016653 Oxidoreductase activity, acting on NAD(P)H, heme protein as acceptor	10	$5.31x10^{-6}$
R-HSA-8950505 Gene and protein expression by JAK-STAT signaling after Interleukin-12 stimulation	36	$7.7x10^{-5}$
GO:0070069 Cytochrome complex	26	$8.07x10^{-5}$
GO:2001021 Negative regulation of response to DNA damage stimulus	70	$1.01x10^{-4}$
R-HSA-389960 Formation of tubulin folding intermediates by CCT TriC	25	$1.34x10^{-4}$
GO:0031362 Anchored component of external side of plasma membrane	17	$1.72x10^{-4}$
GO:0046626 Regulation of insulin receptor signaling pathway	39	$2.55x10^{-4}$
GO:0048585 Negative regulation of response to stimulus	1242	$2.88x10^{-4}$
GO:0046209 Nitric oxide metabolic process	11	$2.94x10^{-4}$
R-HSA-447115 Interleukin-12 family signaling	54	$3.05x10^{-4}$

Table 4.6. List pathways with the strongest association with cognitive decline. (N genes: number of genes in pathway)

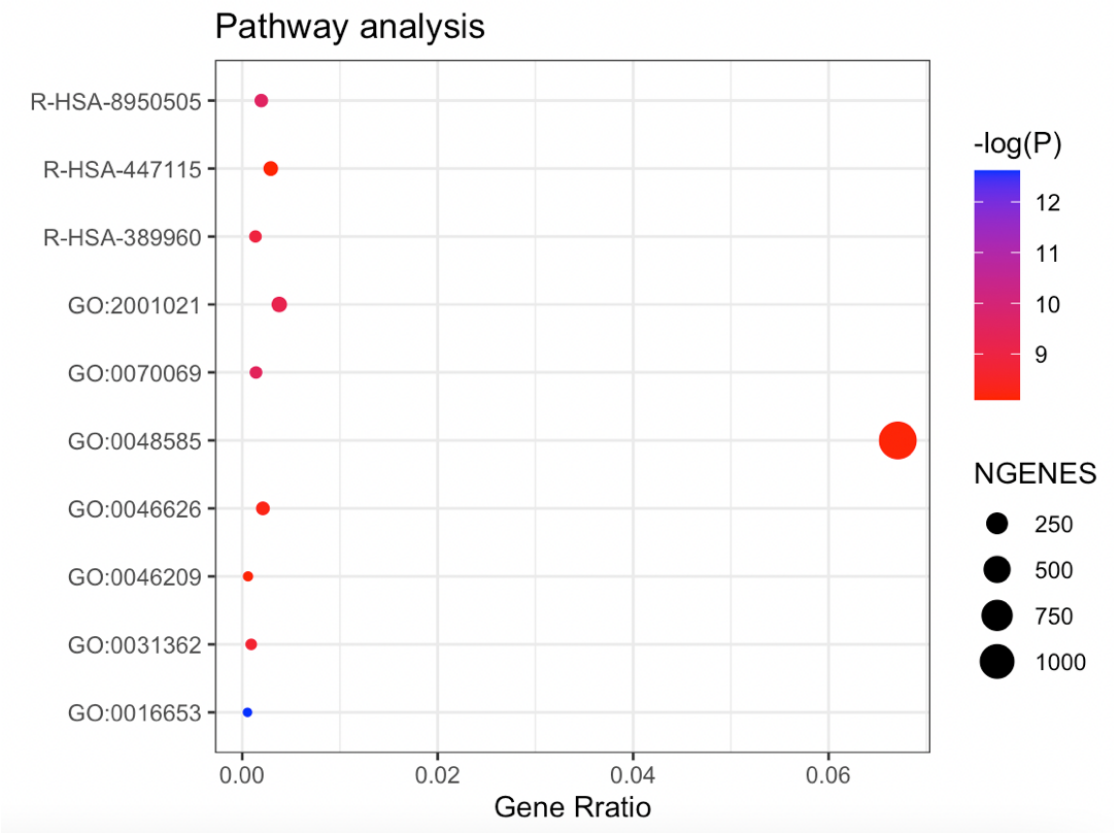


Figure 4.13. List of the ten pathways with the strongest association with cognitive decline. The size of the circle represents the number of genes included in the pathway while the colour of the circle represents the negative logarithm of the p-value of the association of cognitive decline with each of the pathways. (N genes: number of genes in pathway, Gene ratio: number of genes in pathway / total number of genes annotated)

In order to assess the presence of a shared biological pathways between AD risk and cognitive decline in AD, the association of the ten pathways with the strongest association with AD risk, as discovered by Bellenguez *et al.* [198], was assessed. None of the pathways showed a nominal association with cognitive decline, with the lowest p-value being 0.005. The results are illustrated in Table 4.7.

Gene set name	N genes	p-value	AD p-value
GO:1902992 Negative regulation of amyloid precursor protein catabolic process	14	0.084	9.49×10^{-17}
GO:1902430 Negative regulation of amyloid-beta formation	12	0.052	1.82×10^{-12}
GO:0048156 Tau protein binding	13	0.019	3.56×10^{-12}
R-HSA-5619055 Defective SLC24A4 causes hypomineralized amelogenesis imperfecta	1	0.145	8.28×10^{-11}
GO:1905245 Regulation of aspartic-type peptidase activity	11	0.249	9.83×10^{-10}
GO:1902991 Regulation of amyloid precursor protein catabolic process	32	0.058	1.57×10^{-9}
GO:1902003 Regulation of amyloid-beta formation	27	0.019	5.02×10^{-9}
GO:0032994 Protein-lipid complex	26	0.026	1.04×10^{-8}
GO:0033700 Phospholipid efflux	12	0.110	3.68×10^{-7}
GO:0034358 Plasma lipoprotein particle	23	0.005	4.09×10^{-7}

Table 4.7. Association of the ten pathways with the strongest association with AD risk discovered by Bellenguez *et al.* [198] with cognitive decline. (N genes: number of genes in pathway, AD p-value: p-value in the pathway analysis of AD risk published by Bellenguez *et al.*)

4.4.4. PRS

4.4.4.1. PRS of decline

A PRS of cognitive decline was constructed using CAGRAD as a training dataset and ADNI as a validation dataset. Six different p-value inclusion thresholds were assessed (0.5, 0.3, 0.1, 0.05, 0.01, 0.001, 0.0001). There was no association between the PRS derived and the rate of decline for any of the thresholds. The results are illustrated in Table 4.8.

Inclusion threshold	β	SE	p-value
0.5	0.003	0.101	0.976
0.3	-0.008	0.101	0.931
0.1	-0.051	0.101	0.612
0.05	-0.068	0.101	0.503
0.01	-0.010	0.101	0.925
0.001	0.157	0.101	0.122
0.0001	0.163	0.101	0.103

Table 4.8. Association of PRS with cognitive decline for the different inclusion thresholds assessed.

(β: beta coefficient, SE: standard error)

4.4.4.2. AD PRS

Seven p-value thresholds for inclusion in the PRS were assessed (0.5, 0.3, 0.1, 0.05, 0.01, 0.001, $5x10^{-8}$). There was no association between any of the PRS examined and the rate of cognitive decline in AD. The results are illustrated in Table 4.9.

Inclusion threshold	β	SE	p-value
0.5	-0.037	0.047	0.429
0.3	-0.014	0.047	0.757
0.1	-0.050	0.047	0.278
0.05	-0.056	0.047	0.226
0.01	-0.025	0.047	0.588
0.001	-0.031	0.047	0.502
$5x10^{-8}$	0.007	0.047	0.870

Table 4.9. Association of AD PRS with cognitive decline for the different inclusion thresholds assessed.

(β: beta coefficient, SE: standard error)

4.5. Discussion

The main aim of this chapter was to investigate the genetic background of cognitive decline in AD, using the measure of cognitive decline computed in Chapter 3. A number of different approaches were utilised in an attempt to uncover genetic variants that predispose to faster or slower decline.

Firstly, the association of the *APOE* genotype with cognitive decline was assessed. *APOE* is the strongest genetic predictor of AD [186], however its effect on cognitive decline is still under debate. Studies have shown that *APOE* ε4 alleles are correlated to greater cognitive decline in AD patients within a year [242], are significant covariates influencing disease progression in AD patients [241] and are associated with functional decline and disease severity [228]. However, other studies have found that the number of *APOE* ε4 has no effect on cognitive and functional impairment in AD [243], [244] and there are even studies finding that *APOE* ε4 alleles are associated with slower disease course [117]. In this chapter, *APOE* genotype was not found to affect the rate of decline in either of the two datasets examined. A metric of cognitive decline was extracted from a mixed effects linear model, and MMSE was used to assess cognition. Some of the studies finding association use other methods of calculating disease progression and cognitive decline [242], or use other instruments for assessment of dementia severity, like CDR [228] or ADAS-cog [241], [242]. Moreover, some studies include individuals with MCI as well as individuals with AD [228]. It is possible that the mechanisms that are implicated in progression from MCI to AD are not the same to the ones that cause the dramatic cognitive decline seen in AD patients. Therefore, study design is an important factor to take into account when comparing these results. Moreover, a study looking at neuroimaging progression biomarkers and found an association between the number of *APOE* ε4 alleles and the imaging progression markers examined [245]. However, the presence of neuroimaging findings typical of neurodegeneration is not necessarily correlated with the presence of a more severe clinical phenotype in individuals with AD, as data shows that cognitive deficits succeed neuroimaging findings by many years [258]. Therefore, combining cognitive assessments with imaging biomarkers could be beneficial for a more accurate estimation of the disease severity in AD, however that was not within the scope of this thesis, as CAGRAD does not include any imaging data.

Subsequently, the genetic predisposition to accelerated cognitive decline in AD was assessed at a genomic level, using multiple methods. Firstly, two independent GWAS of cognitive decline were performed in CAGRAD (N=529) and ADNI (N=373). No variant exceeded the threshold for genome-wide significance (p-value < $5x10^{-8}$) in CAGRAD, whereas nine variants in chromosome six reached genome-wide significance in ADNI (rs80239946 with p-value = $1.27x10^{-8}$, rs41266319 with p-value = $1.27x10^{-8}$, rs73028356 with p-value = $1.27x10^{-8}$, rs3807062 with p-value = $2.93x10^{-8}$, rs73028322 with p-value = $2.93x10^{-8}$, rs73028335 with p-value = $2.93x10^{-8}$, rs58365143 with p-value = $2.93x10^{-8}$, rs73028343 with p-value = $2.93x10^{-8}$ and rs749510 with p-value = $4.68x10^{-8}$). The variants are located within *FRMD1* gene, and seven out of the nine are located in intronic regions. The gene is protein-coding and the protein is mostly localised in the cell membranes and cytoskeleton and has been implicated in distal myopathy [259]. *FRMD1* has not been associated with AD or other dementias and is not highly expressed in brain tissue. None of the variants within *FRMD1* showed a nominal significance in the CAGRAD GWAS, with the lowest p-value within the region being 0.151. In order to further validate the findings of the two independent GWAS, a GWAS meta-analysis was performed. No significant association was found in the meta-analysis, and the significant associations from the ADNI GWAS failed to replicate, indicating that there is a high probability of them not being true findings. A small number of GWAS of cognitive decline have been published to date, with Sherva *et al.* reporting a genome-wide significantly associated variant in chromosome 11 near the *SPON1* gene (rs11023139, p-value= $7x10^{-11}$) [246] and Scelsi *et al.* reporting a genome-wide significantly associated variant in chromosome 4 near the *LCORL* gene (rs6850306, p-value= $1.03x10^{-8}$) [245]. None of the two variants had a nominal association in the GWAS meta-analysis (p-value= $5.39x10^{-4}$ and p-value= $4.16x10^{-4}$, respectively). Two more GWAS have been published that failed to uncover any significant association [121], [122]. None of the variants that were significant or of suggested significance in this analysis have been previously reported in any of the studies mentioned above as significant or suggestive. As in the *APOE* analysis, study design is an important factor to consider here, as the method used to assess the rate of cognitive decline in this analysis differs to the one used in all existing published GWAS, which could account for some of the differences in the results. Moreover, it is likely that the datasets used in this chapter were not large enough to uncover possible associations, as the samples usually required to uncover common variants of small effect tend to be in the tens of thousands.

Pathway analysis is advantageous in genetic studies with small sample sizes, like the one described in this chapter, as they allow for aggregation of the effect sizes of all genetic variants within the region of a gene, thus increasing the statistical power. For this reason, a pathway analysis was performed, using the results of the GWAS meta-analysis. No gene or gene set was significantly associated with the rate of decline in these analyses after correcting for multiple testing. However, there were some interesting suggestive findings. Many of the pathways with the strongest association with cognitive decline are involved in the immune response, implicating interleukin 12 (IL-12), a pro-inflammatory cytokine [260]. Inflammatory processes are widely suspected to be a driver of neurodegeneration in AD [261], and many inflammatory pathways have previously been implicated in AD pathogenesis [262]. However, plasma biomarkers associated with inflammation have been found not to affect cognitive decline and disease progression in AD [263], so it is possible that while inflammation is implicated in the pathogenesis of AD, it might have no effect on its clinical manifestation. None of the pathways that had the strongest association with cognitive decline in this analysis were associated with AD risk in the analysis published by Bellenguez *et al.,* [198]. Moreover, none of the top pathways in the Bellenguez *et al.,* publication was associated with rate of decline in the pathway analysis described in this chapter. This suggests that the genetic predisposition to cognitive decline in AD could be unrelated to the genetic risk for AD. However, as already mentioned, the pathway analysis described in this chapter is based on a GWAS performed on a small number of individuals (N= 902), therefore further analyses in larger datasets are necessary in order to confirm this hypothesis.

Lastly, the genetic architecture of cognitive decline in AD was assessed by two PRS analyses, one examining the presence of a polygenic component in cognitive decline and one exploring the shared genetic background between AD risk and severity. None of the PRS of decline generated were associated with decline in an independent dataset. However, the PRS were generated from the CAGRAD GWAS, which as mentioned above is based on a relatively small sample size, and therefore lacks the statistical power to uncover possible associations of small effect. While the size of the validation dataset in a PRS analysis tends not to have a tremendous effect on the power of the analysis, the size of the training dataset is of great importance, as it is where the effect sizes and p-values used for scoring are sourced from [264]. It is possible that replicating this analysis in larger datasets could result in different findings. PRS of AD derived from on the newest available GWAS comprising of clinically assessed individuals with

AD [186] also showed no association with the rate of decline. There are larger and more recent GWAS of AD, however they do not only include individuals with clinically assessed AD, but additionally include individuals with proxy AD, defined as having at least one first degree relative that has been diagnosed with AD [198]. While this method of phenotype ascertainment is generally regarded as valid, it was considered here that such a phenotype is less reliable than a clinical assessment and also less similar to the clinically assessed phenotypes present in the cohort that was used in this analysis as a validation dataset. Therefore, the publication by Kunkle *et al.*, [186] was selected as a training dataset was for the PRS analysis. The association of AD PRS with cognitive decline has previously been explored, with Del-Aguila *et al*. finding only a nominal association between a PRS derived from 21 high risk loci for AD and rate of decline [5], and Euesden *et al*. finding no association [121], [228]. Along with the findings of the PRS described in this chapter, these results could indicate that if there is a polygenic determinant of cognitive decline in individuals with AD, it is distinct from the polygenic risk for developing AD, which is also in accordance with the findings from the *APOE* and the pathway analysis. It is possible that the genetic variants that predispose to AD cause a cascade of events that lead to neurodegeneration but play no further role once the degeneration has been established. However, validating this hypothesis is outside the scope of this chapter.

4.6. Conclusions

The main aim of this chapter was to explore the genetic architecture of the rate of cognitive decline in AD. Four different approaches were utilised in order to examine different facets of that genetic architecture, namely candidate gene analysis, GWAS, pathway analysis and PRS analysis. No significant genetic determinants of cognitive decline were found in any of the analyses performed. It is possible that larger datasets are necessary in order to uncover genetic drivers of cognitive decline in AD. However, it is equally possible that there is no strong genetic predisposition to a faster rate of decline, and the rate of deterioration in AD patients is driven mainly by environmental factors, like physical activity levels, education and stress. Further analyses would be required to determine this.

Chapter 5 | Behavioural and psychological symptoms of dementia

5.1. Introduction

Despite cognitive decline being the landmark symptom of AD, the disorder is commonly complicated by the presence of additional non cognitive symptoms, termed behavioural and psychological symptoms of dementia (BPSD) [265]. BPSD is a term that encompasses various different disturbances, including among others psychotic symptoms, mood alterations, aggression, eating and sleeping problems, behavioural and sexual disinhibition, overactivity and apathy. BPSD are common, occurring in 50-90% of individuals with AD [266], [267], severely affect the quality of life of AD patients and their caregivers, and are associated with early institutionalisation, increased mortality and higher caregiver distress than cognitive decline in isolation [266], [268]–[273]. As BPSD are responsible for a substantial proportion of the social burden of dementia, they constitute a significant target for interventions. Despite their apparent significance, BPSD are still an understudied domain in dementia research, and the factors that drive the development of BPSD in AD patients have not yet been determined. Several demographic factors have been shown to influence BPSD. For example, sex seems to influence the type of BPSD that people develop, with men being more likely to exhibit aggression and sexual disinhibition, while women being more likely to experience psychotic symptoms and depression [274], [275]. However, the mechanism through which such factors can lead to the development of BPSD is unknown, and it is still unclear if there is a biological link between the neurodegeneration of AD and the occurrence of BPSD.

Although BPSD can be diagnosed individually, they tend to appear in association, with 50% of individuals with AD experiencing more than four BPSD simultaneously [276]. BPSD tend to cluster within domains of co-occurring symptoms, creating clinically distinct BPSD sub-phenotypes. In 2006, Hollingworth *et al.* reported the presence of four distinct BPSD sub-

phenotypes using a subset of the CAGRAD cohort [277], while others have reported different patterns of BPSD clustering [278]. There is no universally approved method of grouping BPSD, partially because of the heterogeneity in the groups of BPSD defined by different studies.

The aetiopathogenic mechanisms behind BPSD are largely unknown, however there is a strong indication of a genetic component [279]. *APOE*, the strongest genetic risk factor for developing AD [280], has been associated with certain behavioural symptoms and sub-phenotypes however findings being inconsistent [186]. With the exception of psychosis, the BPSD sub-phenotype that has been most thoroughly researched [281], the current knowledge regarding the genetic architecture of BPSD barely extends beyond candidate gene studies, and more research is required in order to identify predisposing genetic factors.

In this chapter, I have attempted to explore the presence of BPSD within a cohort of individuals with AD. I used a statistical method of defining BPSD sub-phenotypes similar to the one described by Hollingworth *et al.* [277] and examined the association of the sub-phenotypes I discovered with demographic and genetic risk factors, aiming to shed some light into the complex aetiology of these symptoms.

5.2. Aims and objectives

The main aim of this chapter is to explore the epidemiology and genetic background of BPSD in a sample of individuals with AD.

The objectives of this chapter are:

- To examine the distribution of BPSD in individuals with EOAD and LOAD
- To use principal component analysis to identify groups of co-occurring BPSD
- To examine the association of BPSD groups with demographic factors
- To investigate the presence of a shared genetic background between BPSD and common neuropsychiatric disorders and traits using PRS

5.3. Methods

5.3.1. Dataset

The dataset used in this chapter is described in detail in chapter 2.1.6. In summary, this dataset comprised of 3355 individuals with AD from the CAGRAD cohort. 1724 of them were collected during the first wave of data collection and 1831 during the second wave. Out of the individuals from the first wave, 171 had EOAD and 1506 had LOAD. Out of the individuals from the second wave, 1177 had EOAD and 642 had LOAD, while 53 individuals from the first wave and 9 from the second had no information on the age of disease onset. The genotyping of the dataset used is described in chapter 2.1.3.

5.3.2. Neuropsychiatric inventory

The neuropsychiatric inventory (NPI) [148] was used to assess the presence of BPSD in this dataset. The NPI is considered to be one of the most valid measures of BPSD. It is widely used in research and clinical settings and has been translated and validated for use in multiple languages [282]. The NPI uses the individual's primary caregiver as an informant, as the individuals themselves are not considered optimal informants as they might not be able to recall or describe the symptoms, and takes between 20 and 30 minutes to complete. Although originally designed specifically for AD, it is now commonly used for assessing BPSD in other forms of dementia. It assesses 12 behavioural domains: delusions, hallucinations, anxiety, depression, elation, aggression, disinhibition, irritability, aberrant motor behaviour (AMB), apathy, eating and sleeping disturbances. The NPI examines the frequency and severity of each symptom domain. The frequency of symptoms is rated from one (occasionally or less than once a week) to four (very frequently or more than once a day). The severity of symptoms is rated either one, two or three (mild, moderate, or severe respectively). The score of each symptom domain is obtained by multiplying the frequency by the severity rating, ranging from one to 12, however it is not possible for individuals to obtain scores of 5, 7, 10 or 11. If an individual does not exhibit a symptom at all, their score for that domain will be zero. The total NPI score is the sum of all domain scores [148].

The NPI questionnaire was part of the initial assessment during the CAGRAD phenotypic data collection, as described in Chapter 2.1.2. There were 3,355 individuals in the cohort with available data for at least one NPI domain. Missing NPI domain data for the individuals included in this analysis were imputed using the mice() package in R [209], as described in detail in chapter 2.1.6.

5.3.3. Cohort comparison

A comparison between individuals with EOAD (N= 1348) and LOAD (N= 2148) was first attempted, in order to assess the similarity of BPSD prevalence between the two groups of individuals. First, the percentage of individuals exhibiting each of the symptoms measured by the NPI was compared between the individuals with EOAD and LOAD using a chi squared test. Then, the distribution of the NPI domain scores were compared between individuals with EOAD and LOAD, only for the individuals that had a total score above zero for that symptom domain. As the NPI domain scores are not normally distributed a Mann-Whitney test was used. All statistical analyses were performed in R [208].

5.3.4. Principal component analysis

Despite offering rich phenotypic information, individual NPI domain scores offer limited clinical validity, while also analysing each domain individually reduces the statistical power of any analyses compared to analysing multiple domains together. Moreover, NPI domains are highly correlated, and assessing them individually fails to take that correlation into account. In order to increase the statistical power, reduce the number of statistical analyses performed and obtain clinically interpretable results, combining the scores of multiple NPI domains using a data reduction technique was considered appropriate. Two different types of data reduction techniques are commonly used to determine BPSD sub-phenotypes in individuals with AD, factor analysis and cluster analysis. In cluster analysis the individuals are grouped into non-overlapping clusters based on the symptoms that they exhibit, while in factor analysis the symptoms are grouped into factors, and an individual can exhibit multiple symptom groups in differing levels, while symptoms can belong in multiple factors [278]. Principal component analysis (PCA) has also been used to define BPSD groups. As with factor analysis, PCA identifies

variable correlation patterns in the data. However, PCA does not assume a latent factor structure underlying the observed data, but instead aims to examine the linear combination of the observed variables.

PCA was selected as a data reduction method in this analysis in order to approximate the approach used by Hollingworth *et al.* in a subset of the CAGRAD cohort [277]. PCA creates a number of variable components using a linear combination of the observed variables. The number of components is determined based on the proportion of variance explained by the components, the eigenvalue of each component and a visual examination of the scree plot [283]. A Pearson's correlation matrix is used as a standard in PCA, however when non normally distributed data are considered, non-parametric correlations can be used instead. PCA has the advantage of tolerating data that might include missing values, imprecise measurements and multicollinearity, and can be used for continuous, ordinal or categorical data [284], [285].

Since the values of 5, 7, 10 and 11 are unobtainable, the NPI domain scores are neither continuous nor normally distributed. They were therefore treated as ordinal variables. Polychoric correlations were used to estimate the correlation between the 12 NPI domain scores, as they are more appropriate for ordinal data, and tolerate deviation from normal distribution better than Pearson's correlations [286]. The polychoric correlation matrix was computed using the polycor() package in R [287]. The resulting correlation matrix was then utilised in a PCA analysis. The number of components were selected based on visual inspection of the scree plot, the Kaiser's eigenvalue > 1 criterion [283], the suggested number of components by the psych() R package [288] that was used to produce the scree plot, and the clinical meaningfulness of the resulting component structure. Based on these criteria, the use of three, four and five components was explored, and the resulting component structure was visually inspected. A non-orthogonal rotation of the components was selected, to account for intercorrelations between components, which were expected due to AD neurodegeneration being the common driving factor between them. Component loadings of > 0.5 were considered in the interpretation of the components.

The PCA was performed separately in individuals with EOAD and LOAD to assess if there is a different component structure between the two groups. The five-component structure was deemed the most clinically meaningful based on the combination of NPI symptom domains included in each of the components and thus was selected. Since the component structure was similar for both EOAD and LOAD individuals when using five components, the individuals

were subsequently combined and the PCA was performed anew. The component loadings for each individual were then extracted and used in all subsequent analyses. The principal component analysis was performed in R using the psych() package [288].

5.3.5. Contributing factors

It is still unclear whether certain BPSD are influenced by the disease stage of the individual, with certain symptoms having been reported to arise as the disease severity increases [127]. In order to assess the influence of disease severity on BPSD, the individuals were classed as having mild to moderate and moderate to severe AD (MMSE score ≥14 and <14, respectively), as defined by the NICE guidelines [289]. There were 2386 individuals with mild to moderate AD and 926 individuals with moderate to severe AD in the dataset. The scores derived from the PCA for each component were compared between the two groups using a Wilcoxon rank sum test. Following that, individuals were also separated based on disease duration, in order to assess its effect on BPSD. Disease duration was defined as time elapsed between the age at disease onset and the age at which the NPI was administered. The mean disease duration in this dataset (5.89 years) was used as a cut-off. There were 2077 individuals in the early stages and 1407 individuals in the late stages of AD in the dataset. The component scores were compared between the two groups using a Wilcoxon rank sum test.

Next, the influence of a number of demographic characteristics on BPSD was assessed, along with the strongest genetic risk factor for AD, *APOE*. The demographic factors examined were sex, age at disease onset, age at assessment, years in education, and disease duration. The association of the component scores with the demographic factors and the number of *APOE* ε4 and ε2 alleles was assessed using linear regression. As disease duration is linearly related to age and age at onset, the regression was performed sequentially, with disease duration being added at a first stage and all other demographic factors being added at a second stage, while the *APOE* ε2 and ε4 alleles were added last, at a third stage. Out of the 3355 individuals in the cohort, 2,845 had available information on APOE genotype and were included in the third stage. All statistical analyses were performed in R [208].

5.3.6. PRS analysis

The principles of PRS analysis are described in detail in chapter 4.3.4. In this chapter, PRS analysis was utilised to assess the presence of a shared polygenic architecture between common neuropsychiatric disorders and traits and the five BPSD domains established by the principal component analysis described in chapter 2.3.4. Five different traits were investigated: AD, bipolar disorder (BD), major depressive disorder (MDD), schizophrenia and IQ, using the latest available GWAS study for each phenotype as a training dataset. The training datasets used in this analysis are summarised in table 5.1. CAGRAD was used as a validation dataset, which is completely independent from all five training datasets. Out of the individuals with available BPSD data, 2,804 had available genotypes and were included in this analysis. There were 6,039,704 SNPs available in the validation dataset. The common SNPs between the validation dataset and the training datasets are illustrated in Table 5.1. LD clumping was performed in CAGRAD, using the individual SNP data from the training datasets. The clumping parameters used were an r^2>0.2, a p-value threshold of 1 and a physical distance threshold of 1MB. The numbers of SNPs available after clumping are illustrated in Table 5.1.

Trait	Dataset	Sample size	N SNPs	Common SNPs	SNPs after clumping
SCZ	PGC 2021	69,369 cases 236,642 controls	7,585,078	5,106,508	234,781
BD	PGC 2021 [290]	41,917 cases 371,549 controls	7,608,184	5,111,813	235,980
MDD	PGC 2019 [291]	246,363 cases 561,190 controls	7,237,821	5,000,603	231,746
AD	Kunkle *et al.,* 2019 [186]	32,070 cases 51,886 controls	11,469,978	5,111,357	235,753
IQ	Savage *et al.,* 2018 [292]	269,867	9,295,119	5,112,294	235,788

Table 5.1. Description of the training datasets used in the PRS analysis. (SCZ: schizophrenia, BD: bipolar disorder, MDD: major depressive disorder, AD: Alzheimer's disease, SNP: single nucleotide polymorphism, PGC: Psychiatric Genomics Consortium)

Weighted PRS were computed for four different significance thresholds (p-value < 0.5, 0.1, 0.05, 0.01), using the number of risk alleles for each individual adjusted for the effect size of the allele as found in the training datasets. The PRS were generated using PLINK [247]. The association between the PRS generated and the scores of the five BPSD sub-phenotypes was tested using linear regression, controlling for disease duration, age, sex and the ten first principal components identified in CAGRAD, to account for population stratification. When a PRS was associated with multiple sub-phenotypes, the association of it with all implicated sub-phenotypes was tested using multiple linear regression, controlling for disease duration, age, sex and population stratification. The regression analyses were performed in R [208].

5.4. Results

5.4.1. Cohort comparison

The NPI assess 12 symptom domains of BPSD, for each of which a score is calculated based on the symptom's frequency and severity. The summary statistics of the 12 symptom domains scores for individuals with EOAD and LOAD are illustrated in Table 5.2. The similarity of BPSD in individuals with EOAD and LOAD was assessed. First, the percentage of individuals exhibiting each symptom was compared between the two groups a contingency table analysis. Then, the distribution of each domain score was tested using a Mann-Whitney test. The results are shown in Table 5.2.

Symptom	% present		Mean			
	LOAD	EOAD	LOAD	EOAD	χ^2 p	Mann-Whitney p
Delusions	31.19	18.62	1.70	0.93	9.08×10^{-21}	$<2.2 \times 10^{-16}$
Hallucinations	18.85	16.24	0.83	0.62	0.06	0.03
Anxiety	40.55	47.03	1.87	1.83	1.89×10^{-4}	0.05
Depression	53.91	57.71	2.22	1.99	0.31	0.55
Aggression	47.35	39.76	2.16	1.71	1.32×10^{-5}	2.67×10^{-6}
Irritability	41.01	38.80	1.99	1.67	0.21	0.05
Disinhibition	27.93	22.11	1.00	0.73	1.46×10^{-4}	6.49×10^{-5}
Elation	10.00	11.42	0.32	0.34	0.20	0.18
AMB	39.62	30.05	2.27	1.62	1.15×10^{-8}	5.21×10^{-9}
Apathy	63.92	50.37	4.15	2.90	1.17×10^{-15}	$<2.2 \times 10^{-16}$
Sleep	41.81	31.90	2.21	1.76	4.63×10^{-8}	1.29×10^{-9}
Eating	45.25	35.83	2.84	2.08	5.26×10^{-9}	5.74×10^{-9}

Table 5.2. Summary statistics of the NPI domain scores for individuals with EOAD and LOAD and comparison between EOAD and LOAD. (AMB: aberrant motor behaviour, p = p-value, χ^2 = chi squared)

Delusions, aggression, disinhibition, AMB, apathy and eating and sleeping disturbances were significantly more prevalent in individuals with LOAD. Anxiety was significantly more prevalent in individuals with EOAD, while there was no significant difference for hallucinations, depression, irritability and elation between the two groups. Delusions, hallucinations, aggression, disinhibition, AMB, apathy and eating and sleeping disturbances had a significantly higher score in individuals with LOAD, while there was no significant difference between the groups for anxiety, depression, irritability and elation.

5.4.2 Principal component analysis

A PCA was performed in order to identify groups of co-occurring BPSD. First, polychoric correlation matrices were computed and then used in the PCA. The correlation plots of the NPI domain scores for the two CAGRAD waves are illustrated in Figure 5.1. the scree plots for individuals with EOAD and LOAD are shown in Figure 5.2.

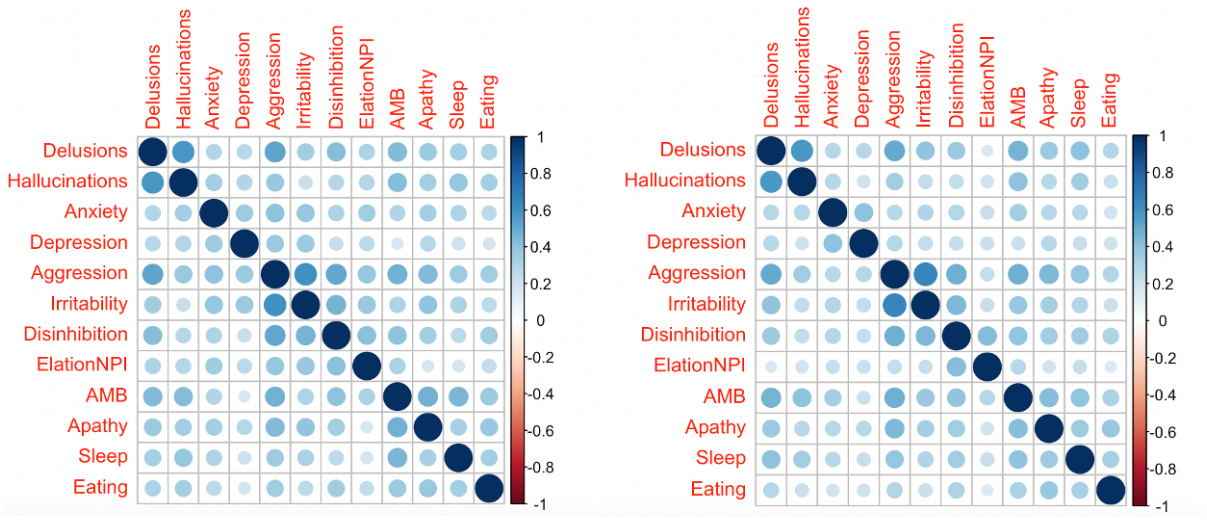


Figure 5.1. Correlation plot of NPI domain scores for individuals with EOAD (left) and LOAD (right). The colour of the circles denotes the direction of the correlation while the size of the circles represents the absolute value of the corresponding correlation coefficients. (AMB = aberrant motor behaviour)

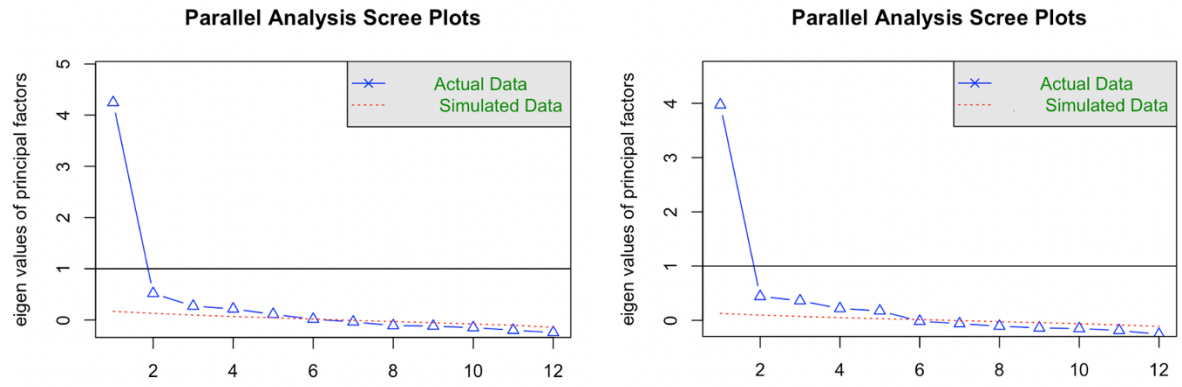


Figure 5.2. Scree plots of the eigenvalues scores for individuals with EOAD (left) and LOAD (right). They display the eigenvalues of the components in descending order. Components that appear before the curve levels off are considered valid to be retained [293].

Three, four and five components were explored. The five-component structure was considered to be the most clinically meaningful for both waves of the cohort and was selected. It is illustrated in Figure 5.3, and the component loadings are shown in Tables 5.3 and 5.4. A loading cut-off of 0.5 was selected for including symptoms into a component. The three- and four-component structures are illustrated in Supplementary Figures 1 and 2. and the component loadings for these structures are shown in Supplementary Tables 1 and 2.

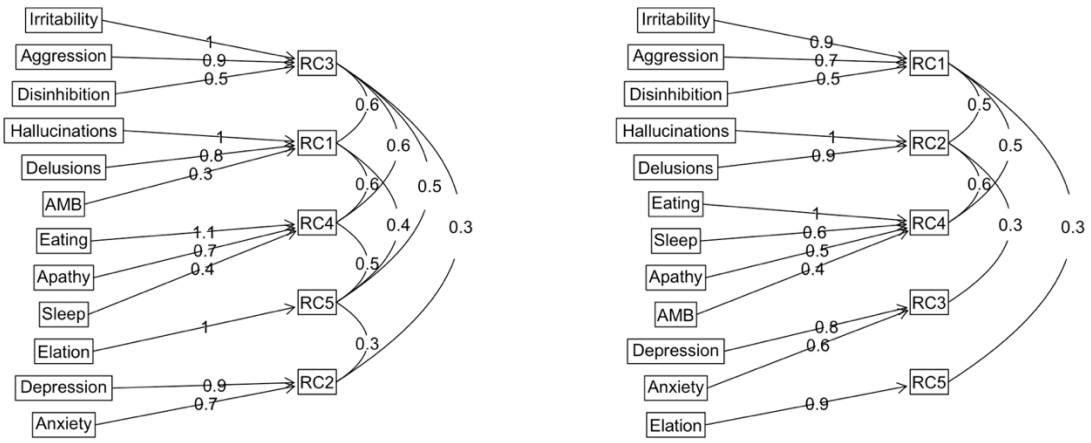


Figure 5.3. PCA of NPI domain scores for individuals with LOAD (left) and EOAD (right). The numbers on the arrows represent the loading of each NPI domain to a component while the numbers on the arcs represent the correlation between the components. (RC: rotated principal component, AMB = aberrant motor behaviour)

	RC1	RC2	RC3	RC4	RC5
Delusions	0.83	0.05	0.24	-0.13	-0.16
Hallucinations	1.02	0.00	-0.16	-0.18	0.01
Anxiety	0.10	0.66	-0.06	0.08	0.18
Depression	-0.02	0.90	0.04	-0.01	-0.05
Aggression	0.08	0.00	0.91	-0.03	-0.11
Irritability	-0.08	0.02	1.00	-0.15	-0.02
Disinhibition	-0.13	-0.05	0.49	0.09	0.46
Elation	-0.05	0.02	-0.10	-0.13	1.02
AMB	0.31	-0.05	0.17	0.20	0.20
Apathy	-0.07	0.08	0.26	0.68	-0.19
Sleep	0.38	-0.12	-0.08	0.44	0.15
Eating	-0.15	0.01	-0.19	1.06	-0.08
Eigenvalue	1.90	1.31	2.18	1.63	1.31
Proportional variance	0.16	0.11	0.18	0.14	0.11

Table 5.3. Component loadings for individuals with LOAD. The loadings for the NPI symptom domains that exceeded the threshold of 0.5 for each component are displayed in yellow. (RC: rotated principal component, AMB = aberrant motor behaviour)

	RC1	RC2	RC3	RC4	RC5
Delusions	0.21	0.91	-0.06	-0.23	-0.08
Hallucinations	-0.28	0.97	0.12	-0.01	0.04
Anxiety	0.11	0.21	0.56	0.03	0.12
Depression	0.19	-0.04	0.80	-0.03	0.09
Aggression	0.73	0.19	0.07	-0.07	0.04
Irritability	0.95	-0.19	0.22	-0.14	0.03
Disinhibition	0.52	0.05	-0.20	0.17	0.39
Elation	0.05	-0.03	0.16	0.09	0.87
AMB	0.12	0.40	-0.22	0.44	0.03
Apathy	0.47	-0.08	0.07	0.50	-0.37
Sleep	-0.15	0.15	0.26	0.56	-0.07
Eating	-0.15	-0.21	-0.04	0.99	0.18
Eigenvalue	2.19	1.98	1.30	1.72	1.15
Proportional variance	0.18	0.16	0.11	0.14	0.10

Table 5.4. Component loadings for individuals with EOAD. The loadings for the NPI symptom domains that exceeded the threshold of 0.5 for each component are displayed in yellow. (RC: rotated principal component, AMB = aberrant motor behaviour)

As the component structure was similar between individuals with EOAD and LOAD, the individuals were combined into one dataset and the PCA performed anew. The correlation of the NPI domain scores and the component structure for the combined dataset is illustrated in Figures 5.4 and 5.5 respectively. The component loadings are shown in Table 5.5.

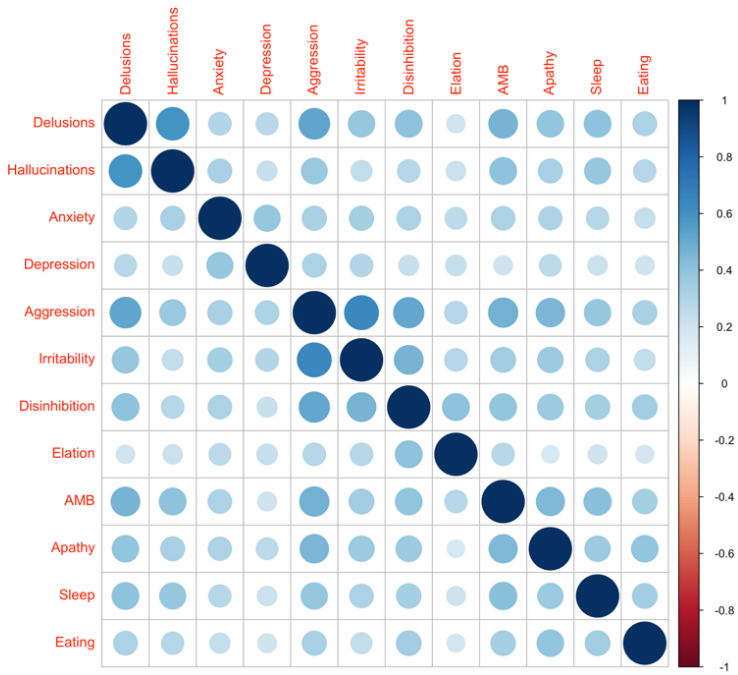


Figure 5.4. Correlation plot of NPI domain scores for all individuals. The colour of the circles denotes the direction of the correlation while the size of the circles represents the absolute value of the corresponding correlation coefficients. (AMB = aberrant motor behaviour)

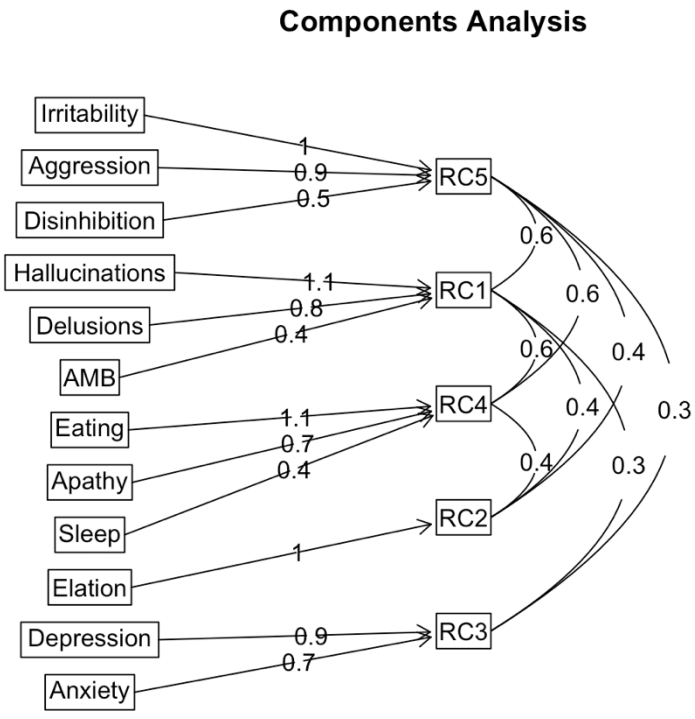


Figure 5.5. PCA of NPI domain scores for all individuals. (The numbers on the arrows represent the loading of each NPI domain to a component while the numbers on the arcs represent the correlation between the components. RC: rotated principal component, AMB = aberrant motor behaviour)

	RC1	RC2	RC3	RC4	RC5
Delusions	0.83	-0.12	-0.02	-0.14	0.25
Hallucinations	1.06	0.03	0.05	-0.16	-0.20
Anxiety	0.15	-0.16	0.66	0.06	-0.04
Depression	-0.04	0.02	0.85	0.01	0.10
Aggression	0.11	-0.06	0.01	-0.05	0.86
Irritability	-0.14	-0.02	0.08	-0.17	1.03
Disinhibition	-0.07	0.42	-0.11	0.16	0.49
Elation	-0.03	0.98	0.05	-0.06	-0.05
AMB	0.39	0.12	-0.11	0.31	0.11
Apathy	-0.06	-0.23	0.11	0.67	0.25
Sleep	0.39	0.01	0.02	0.44	-0.07
Eating	0.20	0.02	0.00	1.09	-0.22
Eigenvalue	1.04	1.22	1.28	1.75	2.17
Proportional variance	0.17	0.10	0.11	0.15	0.18

Table 5.5. Component loadings for all individuals. The loadings for the NPI symptom domains that exceeded the threshold of 0.5 for each component are displayed in yellow. (RC: rotated principal component, AMB = aberrant motor behaviour)

The five sub-phenotypes identified through the PCA were:

1. Psychosis, comprising of delusions and hallucinations
2. Elation, which constituted its own component
3. Affect, comprising of depression and anxiety
4. Behaviour, comprising of apathy and eating disturbances
5. Agitation, comprising of aggression, irritability, and disinhibition

AMB and sleeping disturbances did not have a loading above 0.5 for any of the components. The highest loading for AMB was on psychosis (0.39) while the highest loading for sleep was on behaviour (0.44). The distributions of the five sub-phenotype scores are illustrated in Figure 5.6.

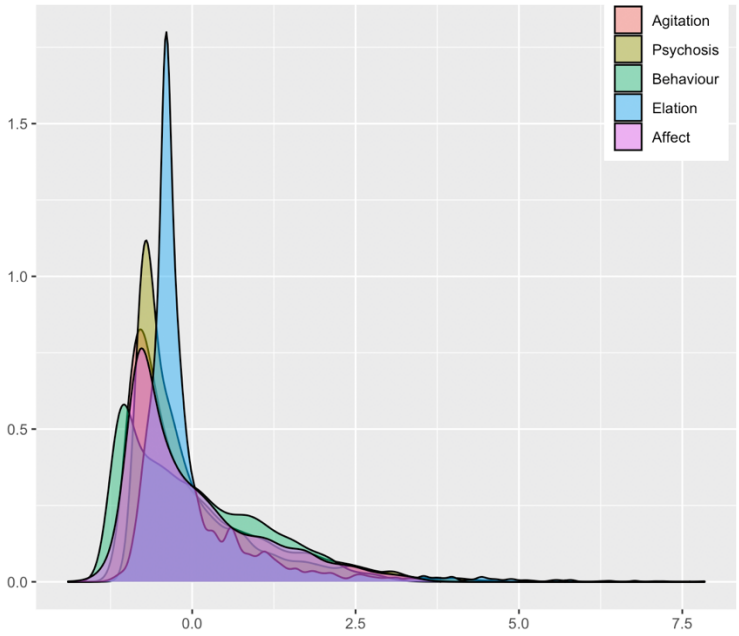


Figure 5.6. Distribution of the five sub-phenotype scores, showing a negative skew.

5.4.3. Contributing factors

Firstly, the scores of the five BPSD sub-phenotypes were compared between individuals with mild to moderate and moderate to severe AD using a Wilcoxon rank sum test, in order to assess the effect of disease severity on BPSD development. All sub-phenotypes had a significantly higher score in individuals with moderate to severe AD, as illustrated in Table 5.6 and Figure 5.7.

Sub-phenotype	Mean Severe AD	Mean Mild AD	Wilcoxon p-value
Psychosis	0.58	-0.28	$<2.2x10^{-16}$
Agitation	0.47	-0.21	$<2.2x10^{-16}$
Elation	0.19	-0.10	$6.16\ x10^{-8}$
Affect	0.13	-0.07	0.001
Behaviour	0.55	-0.25	$<2.2x10^{-16}$

Table 5.6. Comparison of the five sub-phenotype scores between individuals with mild to moderate and moderate to severe AD.

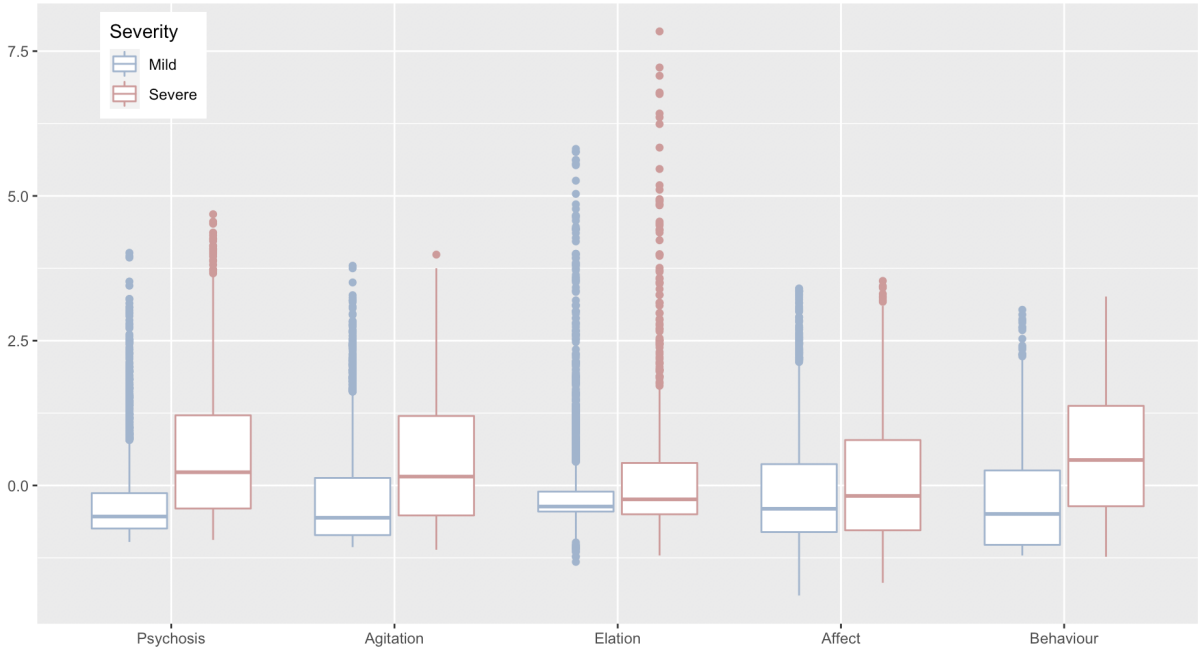


Figure 5.7. Comparison of the five sub-phenotype scores between individuals with mild to moderate and moderate to severe AD.

Then, the scores of the five BPSD sub-phenotypes were compared between individuals with early and late AD using a Wilcoxon rank sum test, in order to assess the effect of disease duration on BPSD development. Psychosis, agitation and behaviour sub-phenotypes had a significantly higher score in individuals with late AD, as illustrated in Table 5.7 and Figure 5.8.

Sub-phenotype	Mean Late AD	Mean Early AD	t-test p-value
Psychosis	0.19	-0.16	$<2.2\times10^{-16}$
Agitation	0.18	-0.13	$<2.2\times10^{-16}$
Elation	0.05	-0.04	0.009
Affect	0.06	-0.04	0.016
Behaviour	0.17	-0.13	2.74×10^{-16}

Table 5.7. Comparison of the five sub-phenotype scores between individuals with early and late AD.

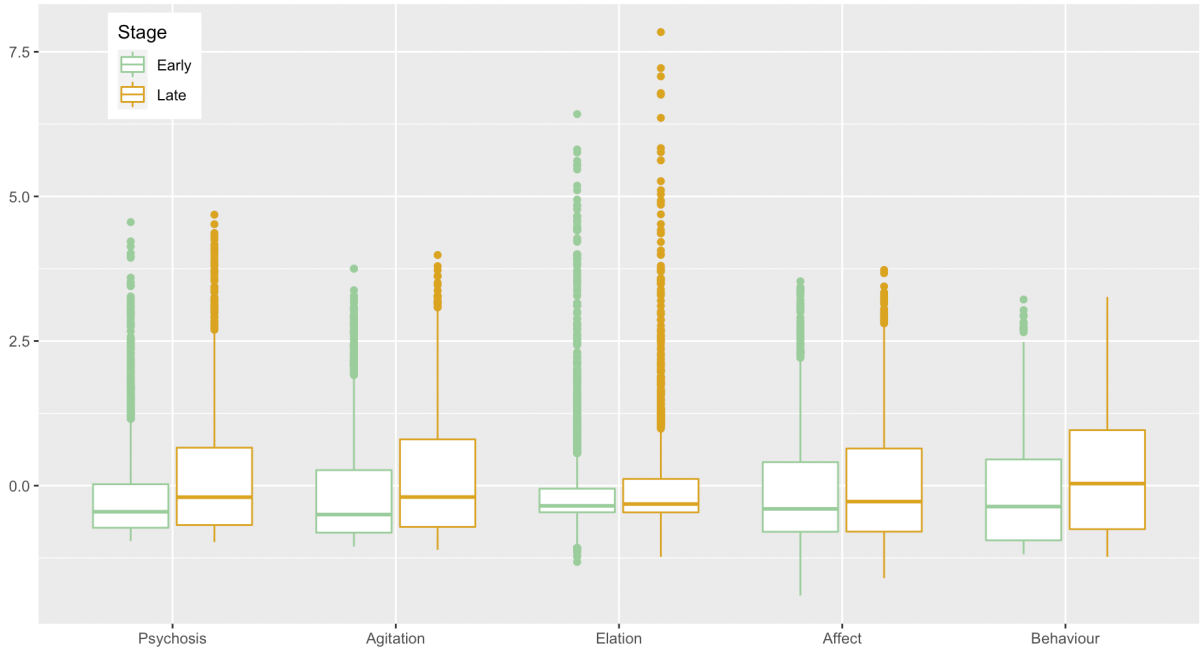


Figure 5.8. Comparison of the five sub-phenotype scores between individuals with early and late AD.

Then, a combination of factors that literature suggests could affect BPSD prevalence were tested for association with the five sub-phenotypes [267], [278], [279]. The factors examined were disease duration, age at assessment, age at disease onset, sex, MMSE score and educational attainment. The results are illustrated in Table 5.8.

	Psychosis		Agitation		Elation		Affect		Behaviour	
	β	p	β	p	β	p	β	p	β	p
Duration	0.01	$2.68x10^{-5}$	0.03	$1.42x10^{-7}$	0.01	0.956	0.01	0.381	0.02	$2.67x10^{-7}$
Sex	-0.02	0.560	0.15	$9.94x10^{-6}$	0.07	0.052	-0.09	0.011	0.01	0.696
AAO	0.03	$6.48x10^{-12}$	0.02	$9.20x10^{-7}$	0.01	0.079	0.01	0.204	0.03	$8.14x10^{-11}$
Age	-0.02	$2.53x10^{-5}$	-0.01	0.008	-0.01	0.047	-0.01	0.203	-0.01	0.003
MMSE	-0.04	$2.05x10^{-98}$	-0.03	$1.27 x10^{-52}$	-0.02	$6.30x10^{-12}$	-0.01	$9.21x10^{-6}$	-0.04	$6.78x10^{-79}$
EA	-0.02	$6.06x10^{-6}$	-0.03	$2.47x10^{-9}$	-0.01	0.097	0.01	0.589	-0.03	$1.66x10^{-7}$

Table 5.8. Association of the five sub-phenotype scores with demographic factors. (AAO = age at onset, EA = educational attainment, β = beta coefficient, p = p-value)

Multiple testing correction was necessary as a large number of tests were performed. The Bonferroni-corrected significance threshold was p-value < 0.01. Psychosis was significantly associated with disease duration (β= 0.01, p-value= $2.68x10^{-5}$), MMSE score (β= -0.04, p-value= $2.05x10^{-38}$), age (β= -0.0, p-value= $2.53x10^{-5}$), age at disease onset (β= 0.03, p-value= $6.48x10^{-12}$) and educational attainment (β= -0.02, p-value= $6.06x10^{-6}$). Agitation was significantly associated with disease duration (β= 0.03, p-value= $1.42x10^{-7}$), sex (β= 0.15, p-value= 9.94x10-

^{6}), age at disease onset (β= 0.01, p-value= $9.22x10^{-7}$), MMSE score (β= -0.03, p-value= $1.27x10^{-52}$) and educational attainment (β= -0.01, p-value= $2.47x10^{-9}$). Elation and affect were only significantly associated with MMSE score (β= -0.02, p-value= $6.30x10^{-12}$ and β= -0.01, p-value= $9.21x10^{-6}$). Behaviour was significantly associated with disease duration (β= 0.02, p-value= $2.67x10^{7}$), age at disease onset (β= 0.03, p-value= $8.14x10^{-11}$), MMSE score (β= -0.04, p-value= $6.78x10^{-79}$) and educational attainment (β= -0.03, p-value= $1.66x10^{-7}$).

Finally, the association of the sub-phenotype scores with the number of *APOE* ε2 and ε4 alleles was also assessed, while controlling for the factors examined above. The results are shown in Table 5.9 and Figures 5.9 and 5.10.

	ε4		ε2	
	β	p	β	p
Psychosis	0.035	0.083	0.064	0.252
Agitation	-0.015	0.501	0.062	0.306
Elation	0.151	$4.09x10^{-10}$	0.010	0.883
Affect	0.328	$8.92x10^{-37}$	0.044	0.529
Behaviour	-0.026	0.213	0.037	0.529

Table 5.9. Association of the five sub-phenotype scores with APOE genotype, while controlling for the examined demographic factors.

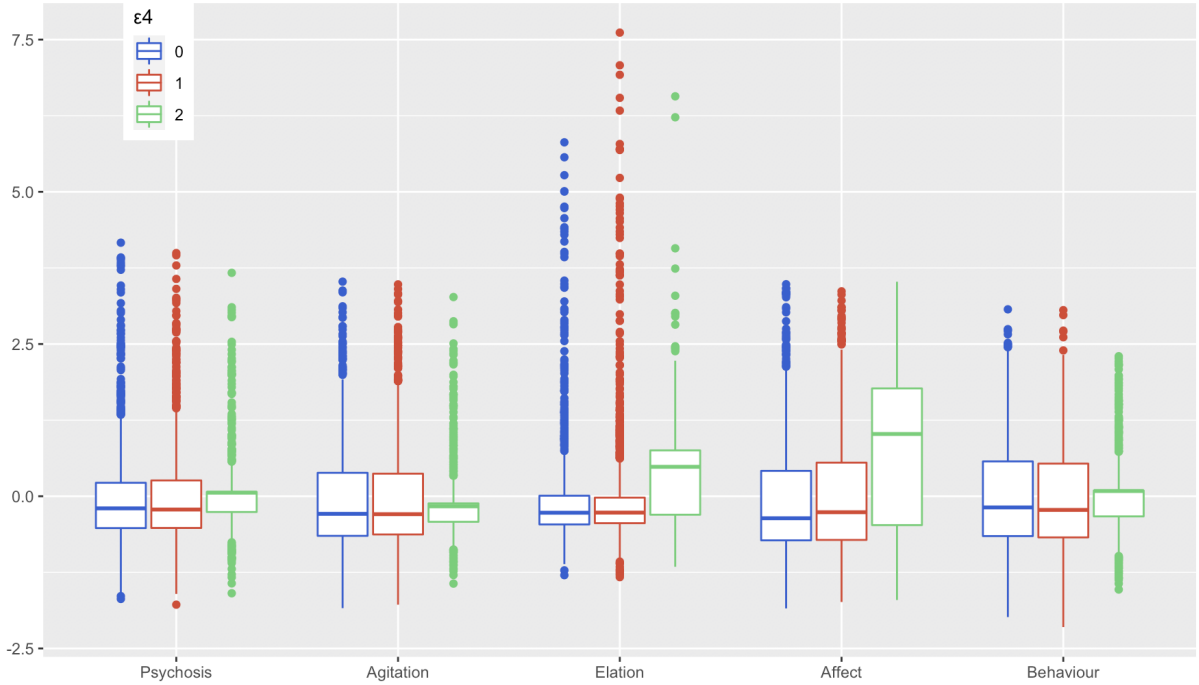


Figure 5.9. Comparison of the five sub-phenotype scores between individuals with one, two or zero *APOE* ε4 alleles, after controlling for sex, age, duration, age at onset and educational attainment.

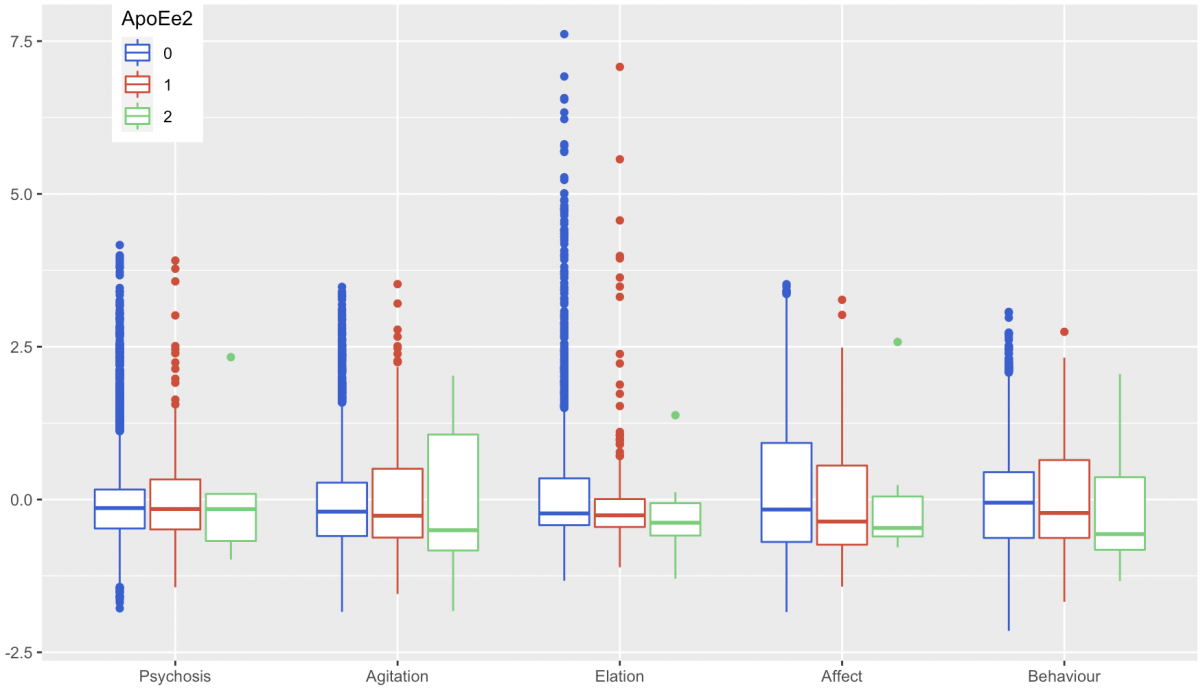


Figure 5.10. Comparison of the five sub-phenotype scores between individuals with one, two or zero *APOE* ε2 alleles, after controlling for sex, age, duration, age at onset and educational attainment.

The number of *APOE* ε4 alleles was significantly positively associated with elation and affect, while controlling for the demographic factors detailed above. The number of *APOE* ε2 alleles was not associated with any of the sub-phenotype scores.

5.4.4. PRS analysis

A PRS analysis was performed in order to assess the presence of a shared genetic architecture between the five BPSD sub-phenotypes and five common neuropsychiatric disorders and traits, AD, BPD, MDD, schizophrenia and IQ. Four inclusion thresholds were assessed for each of the traits examined, p-value <0.5, 0.1, 0.05 and 0.01, resulting in the generation of four PRS for each trait. The PRS that were generated were standardised and tested for association with each of the five sub-phenotype scores using linear regression, while controlling for age, sex, duration and population stratification. The results of the PRS analysis are summarised in Table 5.10 and Figures 5.11 to 5.15, while the full results are illustrated in Supplementary Table 3.

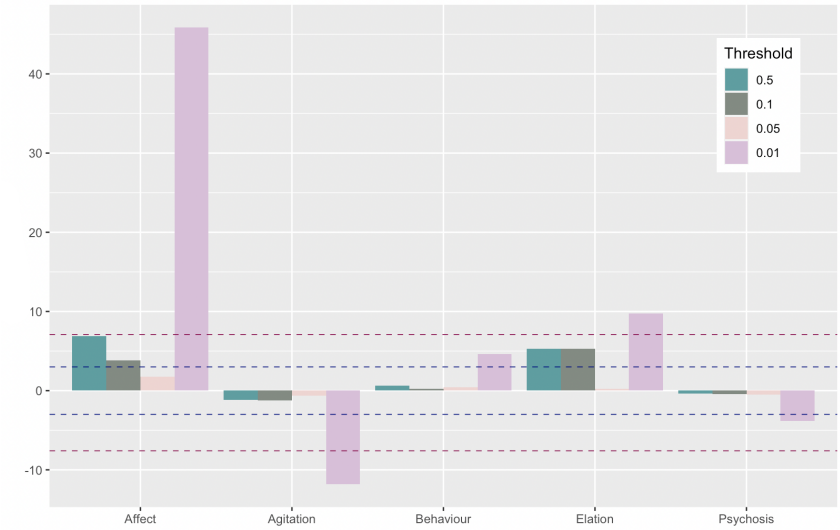


Figure 5.11. Association of AD PRS with the five BPSD sub-phenotypes. PRS derived using the four different inclusion thresholds are denoted by different colours, as explained in the legend. The negative logarithm of the p-value on the y-axis represents the strength of the association, while the direction of the y axis represents the direction of effect, with positive and negative associations represented as positive and negative values, respectively. The blue dashed line represents a standard p-value threshold of p-value < 0.05 white the magenta dashed line represents a Bonferroni-corrected p-value threshold of p-value < $5x10^{-4}$. (BPSD: behavioural and psychological symptoms of dementia)

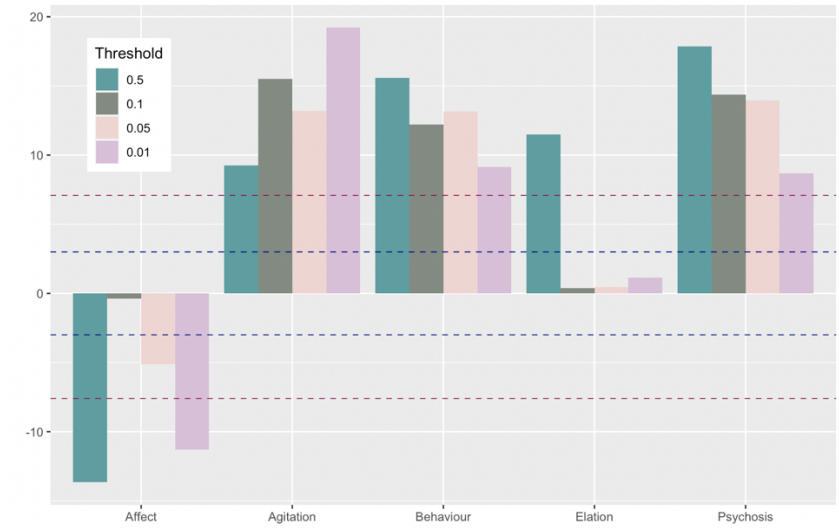


Figure 5.12. Association of schizophrenia PRS with the five BPSD sub-phenotypes. PRS derived using the four different inclusion thresholds are denoted by different colours, as explained in the legend. The negative logarithm of the p-value on the y-axis represents the strength of the association, while the direction of the y axis represents the direction of effect, with positive and negative associations represented as positive and negative values, respectively. The blue dashed line represents a standard p-value threshold of p-value < 0.05 white the magenta dashed line represents a Bonferroni-corrected p-value threshold of p-value < $5x10^{-4}$. (BPSD: behavioural and psychological symptoms of dementia)

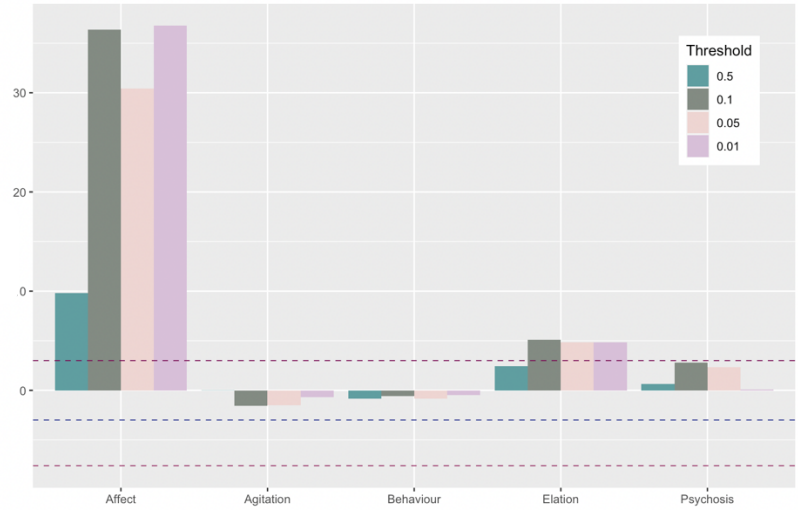


Figure 5.13. Association of bipolar disorder PRS with the five BPSD sub-phenotypes. PRS derived using the four different inclusion thresholds are denoted by different colours, as explained in the legend. The negative logarithm of the p-value on the y-axis represents the strength of the association, while the direction of the y axis represents the direction of effect, with positive and negative associations represented as positive and negative values, respectively. The blue dashed line represents a standard p-value threshold of p-value < 0.05 white the magenta dashed line represents a Bonferroni-corrected p-value threshold of p-value < $5x10^{-4}$. (BPSD: behavioural and psychological symptoms of dementia)

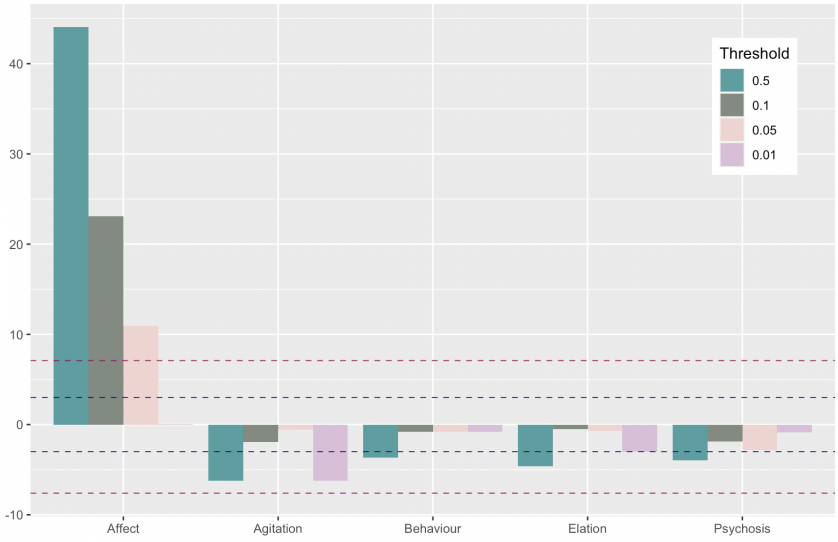


Figure 5.14. Association of major depressive disorder PRS with the five BPSD sub-phenotypes. PRS derived using the four different inclusion thresholds are denoted by different colours, as explained in the legend. The negative logarithm of the p-value on the y-axis represents the strength of the association, while the direction of the y axis represents the direction of effect, with positive and negative associations represented as positive and negative values, respectively. The blue dashed line represents a standard p-value threshold of p-value < 0.05 white the magenta dashed line represents a Bonferroni-corrected p-value threshold of p-value < 5x10-4. (BPSD: behavioural and psychological symptoms of dementia)

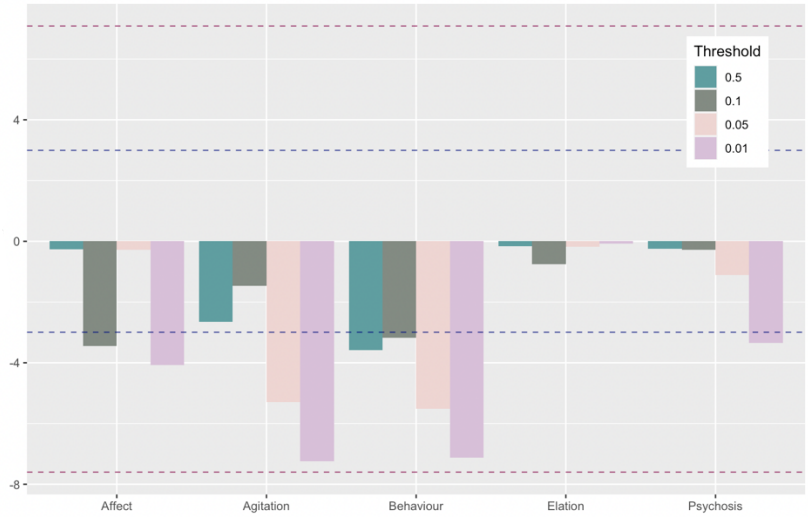


Figure 5.15. Association of IQ PRS with the five BPSD sub-phenotypes. PRS derived using the four different inclusion thresholds are denoted by different colours, as explained in the legend. The negative logarithm of the p-value on the y-axis represents the strength of the association, while the direction of the y axis represents the direction of effect, with positive and negative associations represented as positive and negative values, respectively. The blue dashed line represents a standard p-value threshold of p-value < 0.05 white the magenta dashed line represents a Bonferroni-corrected p-value threshold of p-value < 5x10-4. (BPSD: behavioural and psychological symptoms of dementia)

	Psychosis		Agitation		Elation		Affect		Behaviour	
PRS	β	p	β	p	β	p	β	p	β	p
AD P<0.01	-0.04	0.022	-0.08	$7.35x10^{-6}$	0.07	$5.70x10^{-5}$	0.18	$1.20x10^{-20}$	0.04	0.010
SZ P<0.5	0.10	$1.75x10^{-8}$	0.07	$9.63x10^{-5}$	0.09	$1.01x10^{-5}$	0.10	$1.18x10^{-6}$	0.10	$1.73x10^{-7}$
SZ P<0.01	0.06	$1.67x10^{-4}$	0.10	$4.48x10^{-9}$	-0.02	0.314	-0.15	$1.23x10^{-5}$	0.07	$1.06x10^{-4}$
BD P<0.01	-0.00	0.940	-0.01	0.513	0.05	0.008	0.16	$1.07x10^{-16}$	-0.01	0.632
MDD P<0.5	-0.04	0.019	-0.05	0.002	0.05	0.010	0.18	$7.29x10^{-20}$	-0.04	0.026

Table 5.10. Association of the PRS derived with the five sub-phenotype scores, while controlling for age, sex and the first ten principal components identified in the dataset. Only associations that reached statistical significance are shown here, while for each trait only the PRS derived using the inclusion threshold that achieved the strongest association with each phenotype is included. (AD = Alzheimer's disease, SZ = schizophrenia, BD bipolar disorder, MDD = major depressive disorder, β = beta coefficient, p = p-value)

As a large number of PRS analyses were performed here, the significance threshold was adjusted for multiple comparisons accordingly, using the Bonferroni multiple testing correction method. There were 100 tests performed (5 sub-phenotypes x 5 traits x 4 PRS inclusion thresholds = 100 tests), so the adjusted significance threshold was p-value < $5x10^{-4}$. After correcting for multiple testing, IQ PRS was not associated with any of the sub-phenotypes. AD PRS was significantly associated with agitation (β= -0.08, p-value = $7.35x10^{-6}$), elation (β= 0.07, p-value = $5.70x10^{-5}$), and affect (β= 0.18, p-value = $1.20x10^{-20}$). All the associations for AD PRS were observed when an inclusion threshold of p-value < 0.01 was employed for PRS generation. MDD and BD PRS were significantly associated with affect (lowest p-value = $7.29x10^{-20}$ and $1.07x10^{-16}$ respectively), with a positive direction of effect. Schizophrenia PRS was significantly associated with all five BPSD sub-phenotypes (lowest p-value = $1.75x10^{-8}$ for psychosis, $4.48x10^{-9}$ for agitation, $1.01x10^{-5}$ for elation, $1.18x10^{-6}$ for affect and $1.73x10^{-7}$ for behaviour), showing a positive association with all sub-phenotypes apart from affect, where the direction was negative.

AD and schizophrenia PRS were associated with more than one sub-phenotype, and the sub-phenotypes are highly intercorrelated. In order to assess if one of the sub-phenotypes was driving the association of the PRS with the other sub-phenotypes, a multiple linear regression was performed between the PRS that showed the strongest association for most of the sub-phenotypes (P<0.01 for AD and P<0.5 for schizophrenia) and all the associated sub-phenotypes, while controlling for age, disease duration, sex and population stratification. The results are illustrated in Table 5.11.

	Psychosis		Agitation		Elation		Affect		Behaviour	
PRS	β	p	β	p	β	p	β	p	β	p
AD P<0.01			-0.17	$1.57x10^{-14}$	0.00	$2.49x10^{-5}$	0.18	$2.83x10^{-21}$		
SZ P<0.5	0.07	0.004	0.07	$5.05x10^{-5}$	-0.05	0.136	-0.03	0.020	0.03	0.022

Table 5.11. The association of all sub-phenotypes that were found to be associated with the same PRS as per Table 5.11 was tested using multiple linear regression, while controlling for disease duration, age, sex and population stratification. (AD = Alzheimer's disease, SZ = schizophrenia, β = beta coefficient, p = p-value)

AD PRS remained significantly associated with elation, affect and agitation when all three sub-phenotypes were included in the linear regression. Schizophrenia PRS remained significantly associated with psychosis, behaviour, agitation and affect, while the association with elation failed to reach the significance threshold in this analysis. Interestingly, the strongest association was for the agitation sub-phenotype, which seems to be driving the associations of the other three sub-phenotypes that remained significant

5.5. Discussion

The aims of this chapter were to explore the prevalence of BPSD in CAGRAD, to use appropriate statistical methods of defining BPSD sub-phenotypes, and to examine the association of BPSD to certain demographic and genetic risk factors. The NPI is the psychiatric assessment that was most widely used in the CAGRAD cohort and was selected as a measure of BPSD for this chapter. It is an assessment tool for neuropsychiatric symptoms in individuals with dementia which provides information on 12 different BPSD domains [148]. First, the NPI scores were compared between individuals with LOAD and EOAD, in order to assess the generalisability of any analyses performed, and to decide if there were major differences in the pattern of BPSD burden between EOAD and LOAD. Most NPI domains were more common and had a higher score in individuals with LOAD. However, the distributions of the symptoms and the correlations between them were similar between the two groups of individuals.

Instead of investigating individual NPI symptom domains separately or looking at the overall NPI scores of the individuals in CAGRAD, a PCA was employed in order to identify clusters of co-occurring BPSD, that could constitute behavioural sub-phenotypes. The purpose of this was to drastically reduce the amount of datapoints examined, while also creating a more interpretable and comparable set of phenotypes from a clinical perspective. Investigating sub-syndromes of BPSD can increase the likelihood of uncovering significant associations, as clustering numerous datapoints together increases the statistical power of the analysis. Moreover, despite the rich phenotypic information that the 12-domain NPI provides, it is also evident that studying most of the domains in isolation has limited benefits from a clinical viewpoint, while using the total NPI score is an oversimplification of such a complex phenotype. Therefore, defining groups of correlated BPSD is not only reasonable from a statistical perspective, but can also have benefits in the clinical practice. As certain BPSD tend to co-

occur, it is reasonable to study them together, as they could share neurobiological pathogenetic mechanisms and respond to similar treatments.

Numerous examples of data reduction techniques that define BPSD sub-syndromes exist in literature, with studies having used latent class analysis, factor analysis, cluster analysis, as well as PCA, as in this chapter [278]. Interestingly, the sub-phenotype structure tends to be similar between studies, regardless of the method used to retrieve it, a finding that indicates that there are distinct BPSD syndromes. The number of sub-phenotypes reported varies, with three, four and five sub-phenotype structures having been described [278]. In 2006, Hollingworth *et al.* performed a similar analysis as the one here in a subset of just over 1,000 individuals from the CAGRAD cohort and reported the presence of four sub-phenotypes, namely psychosis, agitation, mood and behavioural dyscontrol [277]. In this analysis, a five sub-phenotype solution was selected as it provided the most interpretable results and was similar between individuals with EOAD and LOAD. The five BPSD sub-phenotypes identified were psychosis, agitation, elation, affect and behaviour. Psychosis included delusions and hallucinations, agitation included aggression, disinhibition and irritability, affect included anxiety and depression and behaviour included eating problems and apathy. Elation constituted a sub-phenotype in its own right, as it was not highly correlated with any of the other NPI domains, something that other studies have shown in the past [294]–[297]. Two symptoms, AMB and sleep, did not load strongly onto any of the components, which has also been found by others [295], [298]. The five BPSD sub-phenotypes discovered in this chapter are consistent with a large number of previously published studies that have used a variety of different methods of sub-phenotype definition [278]. This includes studies that have used alternative methods for assessing BPSD to the NPI, indicating that the sub-phenotype structure is not method- or NPI-specific.

In order to assess if the BPSD sub-phenotype scores differ depending on the stage of AD, the scores of the individuals with mild and severe as well as early and late AD were compared. All sub-phenotypes were found to have a significantly higher score in individuals with a more severe cognitive deficit. Previous studies have shown similar results, with Proitsi *et al.* finding an association between cognitive function and psychosis, agitation and behaviour, Hollingworth *et al.* finding an association between cognitive impairment and psychosis, behaviour and agitation, and Spalletta *et al.* finding an association between disease severity and psychosis and agitation [277], [299], [300]. Psychosis, agitation and behaviour also had

significantly higher scores in individuals with a longer disease duration, however no significant difference was found for elation and affect. Using a longitudinal design, Gonfrier *et al.* found similar results for affect, behaviour and agitation, however did not find a significant change in psychosis over time, while Garre-Olmo *et al.* found an increase in psychosis and behaviour and no change in affect [301], [302]. It is yet unclear how BPSD develop over time, and few longitudinal studies of BPSD have been published to date. As there were minimal longitudinal data on BPSD in CAGRAD, such a type of analysis was at the current time not feasible. However, it would shed some light into the natural history of BPSD and give a deeper insight into the way in which disease stage can affect them.

Next, a number of selected demographic characteristics were tested for association with the five the sub-phenotype scores. MMSE and disease duration were included in this analysis to account for disease stage. Sex, age, and age at disease onset were also included, as they have been previously implicated in BPSD severity [303]–[305]. Educational attainment was also included in this analysis, as it has been shown to influence the severity of cognitive impairment in AD [50]. Sex was associated with agitation, with males having higher scores for that sub-phenotype than females, as literature suggests [274], [299]. An association between the female sex and affect has previously been reported [274], [277], something that was not found in this analysis. MMSE score was associated with all sub-phenotypes, and disease duration with psychosis, agitation and behaviour, as expected. It is generally accepted that a more pronounced cognitive deficit is associated with a higher load of behavioural and psychological disturbances, and findings from human studies as well as animal models suggest that behavioural symptoms are correlated with a higher neuropathological burden [306], [307]. This could indicate that BPSD are a sign of a more aggressive neurodegenerative phenotype. Educational attainment was associated with psychosis, agitation, and behaviour, with individuals with more years in education having significantly lower scores for those sub-phenotypes. Education is a well-established protective factor for MCI and AD [50] and this result suggests that it may also protect from a more severe behavioural phenotype within AD, which also supports the idea that BPSD are connected to a more severe AD phenotype. Age was only associated with psychosis, with a negative direction of effect, indicating that younger individuals are more likely to experience psychotic symptoms. Age at disease onset was associated with psychosis, agitation and behaviour. Interestingly, the association was positive, indicating that individuals with a later onset have higher BPSD burden for these sub-

phenotypes. The effect of age of onset on BPSD is still unclear, with studies having found a positive [299], [308], negative [277], [309] or no association between them [299], [310]. However, these inconsistencies could have arisen from differences in cohort characteristics between the studies, differences in the data collection process, as well as differences in the methodology employed for defining the sub-phenotypes.

Finally, the association of the *APOE* genotype with the five sub-phenotype scores was explored. *APOE* ε4 allele is associated with a significantly higher risk of developing AD, as well as with an earlier onset of disease. It is considered to be involved in the biochemical mechanisms that lead to the development of neurodegenerative pathology in AD [311]. However, its implication in the development of BPSD remains unclear. Some studies have reported a relationship between *APOE* genotype and BPSD, with the number of ε4 alleles having been associated with increased risk of psychosis, aggression, depression, agitation and anxiety [312]–[316]. However, the results are very inconsistent, and a large number of studies have failed to detect any association between the number of *APOE* ε2 or ε4 alleles and the risk development of BPSD [277], [317]–[320]. In this chapter, the number of *APOE* ε4 alleles was found to be associated with the elation and affect sub-phenotypes, while the number of *APOE* ε2 alleles was not associated with any of the five sub-phenotype scores. This indicates that there might be a link between the neurodegeneration that *APOE* facilitates and the development of certain BPSD sub-phenotypes, however, given the inconsistent results mentioned above, further studies are required to validate this finding.

Finally, a PRS analysis was performed in order to explore the polygenic architecture of BPSD and uncover possible links between these symptoms and common neuropsychiatric disorders and traits, aiming to shed some light into the pathogenetic mechanisms of BPSD. Two main theories exist regarding the aetiology of most BPSD. These theories have been formulated around psychotic symptoms, but the same principle applies to all BPSD sub-phenotypes. The first theory is that BPSD are a direct result of the neurodegenerative processes that AD is characterised by. The other popular theory is that BPSD are a consequence of a latent psychopathology that is triggered by the profound neurodegeneration that takes place in AD. If the second theory is correct, it is possible that BPSD share some of the polygenic architecture that predisposes to psychiatric phenotypes. In this chapter, the polygenic architecture of BPSD was compared to that of three psychiatric conditions, BD, MDD and schizophrenia, to assess if genetic liability to these disorders would increase the risk of developing BPSD. Moreover, AD

and IQ, a trait that is strongly associated with AD [50], were included, in order to examine if genetic risk and protective factors for neurodegeneration would influence BPSD risk. Four inclusion p-value thresholds were examined, resulting in four sets of PRS for each trait. Psychosis, agitation and behaviour had a positive association with schizophrenia PRS, while agitation and elation were associated with AD and schizophrenia PRS. Affect was significantly associated with schizophrenia, AD, BD and MDD PRS, with the strongest association being for AD PRS (p-value = 1.20×10^{-20}). The strength of the association of the PRS of each trait and each BPSD sub-phenotype varied depending on the inclusion threshold used in PRS generation. Interestingly, schizophrenia PRS had a stronger association with agitation than with psychosis (lowest p-value = 4.48×10^{-9} and 1.75×10^{-8} respectively). None of the sub-phenotypes were associated with IQ PRS. There is limited evidence regarding the shared heritability of BPSD and other traits, with most published studies focusing solely on psychosis, and being inconsistent. DeMichele-Sweet *et al.* have previously found that a schizophrenia PRS was protective against psychosis in AD [281], while Creese *et al.* have found that a schizophrenia PRS increases the risk of developing psychosis in AD [321]. However, the method of establishing the phenotype of psychosis in both these studies differed compared to the one utilised here, with both studies having used a binary phenotype based on the presence of non-zero scores on the NPI domain of delusions and hallucinations. It is possible that the phenotype used here can better capture the varying degrees of psychosis burden seen in individuals with AD, however in-depth further analyses in larger datasets are required to further explore and validate this. Still, the findings of the PRS analyses described in this chapter suggest that BPSD might not be a direct result of some facet of AD pathology, but instead might be caused by latent predisposition to psychopathology that the extensive neurodegeneration that AD is characterised by brings forward. Similar analyses in individuals with Huntington's disease (HD), a rare neurodegenerative disorder that is cause by a single gene mutation, have reported similar results, with schizophrenia PRS being associated with psychosis and irritability in HD, while MDD PRS being associated with depression in HD [322]. This could indicate that this genetic predisposition is a wider characteristic of psychiatric symptoms in neurodegenerative disorders and not limited to AD. BPSD are most likely a complex trait, and studies focusing on common variation will only explain a small percentage of the variance of these symptoms. Epigenomic and functional genomic studies might be required to determine the complex interplay between genetic and environmental factors that leads to the development of these

symptoms. Moreover, combining clinical, genetic and neuroimaging data could aid in uncovering the ways in which risk factors associated with BPSD are linked to neuropathology. The analysis performed in this chapter uses one of the largest available datasets with information on BPSD, and to this day is the only one that examines the polygenic architecture of multiple BPSD sub-phenotypes. However, it should be noted that the results of this chapter are entirely based on the NPI, which is dependent on an informer other than the individual with AD. A level of bias could have been introduced by the caregivers distorting or exaggerating the patient's symptomology. However, the NPI is a validated tool of assessing BPSD widely used in research, and it is generally accepted that using caregiver information is universally considered a reliable method of assessing BPSD. Moreover, since the majority of the participants in CAGRAD do not have a definite pathologically confirmed AD diagnosis, it is plausible that some of the individuals might have another form of dementia, like frontotemporal dementia or Lewy body dementia, both of which are characterised by the presence of intense BPSD. This could inflate the BPSD prevalence in the dataset, however since AD is the most prevalent dementia in the population, it is reasonable to assume that if such misdiagnoses have occurred in CAGRAD, they will only affect a small number of individuals.

5.6. Conclusions

The aim of this chapter was to explore the complex epidemiology of BPSD and try and gain some insights into their aetiology. Five BPSD sub-phenotypes were defined; psychosis, agitation, elation, affect and behaviour. BPSD sub-phenotypes were found to be associated with multiple demographic factors such as age of disease onset, education, and sex, as well as the severity of cognitive impairment, the *APOE* ε4 allele, and PRS derived from four common neuropsychiatric conditions. However, the pathogenetic mechanisms behind these associations remains unclear. Further studies are required in order to uncover the biological underpinning of these complex phenotypes and lead the way into developing better treatment options for them.

Chapter 6 | General Discussion

6.1. Overview

Throughout this thesis I aimed to explore the variety of phenotypic presentations that can arise in individuals with AD using the CAGRAD cohort. CAGRAD is a large cohort of individuals (total N= 4,163) with AD that includes information on multiple demographic and phenotypic domains as well as genotypes (see Chapter 2). The rich phenotypic information available in this dataset allowed for a thorough investigation of the different facets of AD. After examining the variables available in the CAGRAD dataset and reviewing the literature on phenotypic variation in AD, two main areas of interest were established and explored throughout this thesis, the rate of cognitive decline and behavioural and psychological symptoms of dementia (BPSD). Using several different methods of statistical, epidemiological and genetic analyses, I have explored this dataset aiming to deconvolute the mechanisms that lead to the complex and variable manifestations that characterise AD.

6.2. How can cognitive decline in AD be quantified, and does it have a genetic background?

The rate of cognitive decline in individuals with AD was the first phenotypic domain of AD that was examined. The rate of cognitive decline in individuals with AD exhibits a wide variability, with some individuals declining rapidly after diagnosis and others surviving for up to a decade with relatively preserved cognitive function. The factors that drive this vast heterogeneity remain mostly unknown, and no validated methods of predicting the rate of decline are available.

Mixed effects linear modelling was utilised in order to compute a measure of cognitive decline in individuals with AD, using a subset of the CAGRAD cohort that had been longitudinally assessed (Chapter 3). MMSE score [146] was initially selected as the metric of cognition as it was the most widely documented cognitive assessment in the CAGRAD cohort. It is reasonable to assume that other assessment tools like MoCA [147] would be equally viable for this

analysis, however that was not examined here due to the lack of sufficient numbers of individuals with data on alternative cognitive assessments. Moreover, other assessment instruments like the CDR [202] and the ADL [203], that do not focus solely on cognition but also assess various functional domains are considered to be more effective at capturing the level of disease severity in AD and are therefore preferred by a number of researchers investigating the rate of decline in individuals with AD [228]. However, these instruments are designed to assess functional decline in addition to cognitive decline. In this thesis, only the cognitive component was of interest, as it is possible that the non-cognitive symptoms of AD might not share the exact same aetiology as cognitive ones. For this reason, the choice of MMSE was considered optimal for this analysis, and no other assessment instruments were examined.

Mixed effects linear modelling was selected for this purpose because it can tolerate the high variability often seen in longitudinal population cohorts, and it can also account for interindividual variation [220]. The dependent variable in the model was cognitive function as defined by the MMSE score [146] and the independent variable of interest was the random effect of disease duration. Additional independent variables were added to the model sequentially until the model constructed was considered to be an adequate representation of cognitive decline in this dataset, without overfitting. The variables included in the final model were sex, age, and disease duration. Perhaps there were other variables that are determinants of cognitive decline and could have improved the model. For example, it has been suggested that premorbid cognition as well as educational attainment might ameliorate the cognitive decline in AD due to increased cognitive reserve [50]. However, such variables were not available in sufficient numbers for the individuals that were included in this analysis and including them in the model would have significantly reduced the sample size. Therefore, it was decided that no further variables would be considered.

After the best model was selected, the random slopes for each individual were extracted and used as a measure of cognitive decline in all subsequent analyses. The rate of decline in individuals with EOAD and LOAD was first compared. In CAGRAD, there was no significant difference in the rate of cognitive decline between individuals with EOAD and LOAD, however there was a trend towards faster decline in individuals with LOAD. In an independent dataset that was used for replicating the results (ADNI [323]), individuals with EOAD declined significantly faster than individuals with LOAD. Age at disease onset was not known for a subset

of the individuals in ADNI included in the analyses and was approximated based on the age at recruitment. While there was no difference in the rate of decline between the individuals in ADNI for which age at disease onset was known and individuals for which it was not known, it is possible that the reliability of the variable and the validity of any results that include it might be affected. Many studies have investigated the difference in the rate of cognitive decline between individuals with EOAD and LOAD, with the majority concluding that EOAD is linked to a faster decline [232]–[234], though there are studies that have found the opposite effect [237] or no difference between EOAD and LOAD [236]. However, it is important to take into account that much of the literature on EOAD includes individuals with familial EOAD caused by one of the three recognised autosomal dominant mutations described in Chapter 1.3.1, which are considered to lead to a more malignant phenotype with rapid decline [25], while often the individuals included in the studies have not received genetic testing to determine if they carry one of the deleterious mutations.

Chapter 4 of this thesis utilised the quantitative measure of cognitive decline that was developed in Chapter 3 in analyses aimed to explore the presence of a genetic basis of cognitive decline. First, the effect of *APOE* genotype on cognitive decline was examined. As described in chapter 1.3.3, *APOE* is the strongest genetic determinant of AD, with ε4 allele being the strongest genetic risk factor for sporadic AD and ε2 allele being the strongest genetic protective factor against sporadic AD [66]. While the effect of *APOE* on disease risk is well established, it is not clear if and how much it affects the severity of the phenotype. No association was found between the rate of cognitive decline and the number of *APOE* ε4 or ε2 alleles in either of the datasets examined in this thesis. Previous studies on the effect of *APOE* genotype on severity and progression in AD have been controversial, with a number of publications suggesting that *APOE* ε4 predisposes to faster cognitive and functional decline and a higher disease severity [228], [241], [242], others finding no effect of *APOE* on decline [243], [244] while some even suggesting that *APOE* ε4 is associated with a slower disease progression [117]. It is therefore clear that there is no consensus regarding the effect that the *APOE* alleles might have on disease severity in AD.

Next, the presence of a complex genetic background to cognitive decline in AD was assessed. GWAS were performed on both CAGRAD and ADNI datasets independently and then these were meta-analysed, attempting to and uncover genetic variants of small effect that could predispose to a faster or slower cognitive decline. No variant exceeded the threshold for

genome-wide significance in CAGRAD, while nine variants on chromosome six were significant at a genome-wide level in ADNI. However, these associations failed to replicate in CAGRAD and consequently in the meta-analysis, therefore their validity is questionable. The GWAS meta-analysis did not reveal any associations that exceeded genome-wide significance. However, it should be taken into account that the sample size of the meta-analysis was 902 individuals, which is small for a GWAS. In order to detect common variants of small effect that are associated with a phenotype, GWAS usually require samples that are in the tens of thousands. Therefore, it is possible that increasing the sample size would have facilitated the detection of genome-wide significant associations. Four GWAS of cognitive decline in AD have previously been published, some failing to find any significant associations [121], [122], one study reporting a significant association on chromosome four [245] and one reporting a significant association on chromosome 11 [123]. However, the datasets used in these studies were also only comprising of a small number of individuals (N= 3,946 for the overall largest and N= 911 for the largest that discovered a significant association). Moreover, none of the significant results were independently confirmed and therefore should be interpreted with caution.

Finally, two polygenic risk score (PRS) analyses were performed in order to assess the presence of a polygenic component in cognitive decline in AD. First, a PRS of cognitive decline was computed based on the results of the CAGRAD GWAS and tested for association with the rate of decline in ADNI, without resulting in any significant associations. The statistical power of a PRS analysis is highly dependent on the size of the GWAS that is used as a training dataset, as this is where the effect sizes and p-values are derived from [264]. The training dataset in this PRS analysis was the GWAS performed on the CAGRAD cohort, with a sample size of 529 individuals. Therefore, it is likely that there was not enough statistical power in this GWAS to adequately represent the polygenic architecture of cognitive decline in AD, which would explain the lack of association of the PRS with the rate of decline in ADNI. It is also possible though that if a rapid cognitive decline is indeed genetically determined, this genetic risk is not polygenic. Attempting the PRS analysis in a larger dataset could help deconvolute the genetic basis of cognitive decline in AD. A second PRS was performed, aiming to determine the presence of a shared polygenic architecture between the risk of developing AD and cognitive decline after AD development. No association was observed between a PRS of AD derived from the largest and newest published AD GWAS [186]. The rationale behind this PRS analysis was to determine if the rate of cognitive decline in AD is an integral part of AD pathology or if it is

a distinct sub-phenotype with a separate set of risk factors and a unique genetic underpinning. The lack of association between AD PRS and the rate of cognitive decline in this analysis, along with the findings on the effects of *APOE* on decline as discussed above, suggest that the rate of cognitive decline in AD might be independent from the factors that predispose to AD development. The association of AD PRS with cognitive decline has also been explored by others, with Del-Aguila *et al*. only finding a nominal association between a PRS of AD and cognitive decline, and Euesden *et al*. finding no association at all [121], [228]. These results further reinforce the theory that a rapid cognitive decline in AD could constitute a distinct sub-phenotype and not a mere expression of more severe AD.

Based on the results from Chapter 4, it is not possible to reach a conclusion regarding the presence of a genetic background to faster or slower cognitive decline in AD. While some of the analysis performed gave rise to interesting suggestive findings, no significant associations were uncovered. As brought to attention multiple times throughout this chapter, the dataset that was used in all the genetic analyses described was relatively small, and thus lacks the power to detect significant associations when accounting for the multiple testing of SNPs in a GWAS. Recording longitudinal observations on a large number of individuals is a lengthy procedure that requires a considerable amount of resources and funding. For individuals with AD, it is an even more difficult endeavour, as the participants are often not ambulant, the presence of a caregiver in the assessments is often necessary and there is a high drop-out rate due to morbidity and mortality. This explains why most publications on the genetics of cognitive decline in AD are based on similarly small numbers of individuals. Unless a genetic variant has a very large effect on a phenotype, which is unlikely in the case of cognitive decline, the genotypes of a large number of individuals need to be examined to detect an association. Therefore, it is possible that increasing the number of individuals will uncover associations that now fail to exceed the predetermined genome-wide significance threshold. However, it could also be the case that the rate of cognitive decline in AD is not genetically determined. In order to reach a definite conclusion on this argument, larger studies on longitudinally monitored individuals with AD are necessary.

6.3. What influences the appearance of behavioural sub-phenotypes in AD?

Another domain of substantial phenotypic variability in AD are Behavioural and Psychological Symptoms of Dementia (BPSD). BPSD are hugely detrimental to individuals with AD and their caregivers, however they are often overlooked and are underrepresented in AD research. Their pathogenesis remains obscure and there are no effective therapeutic agents to combat them, as explained in Chapters 1.6.2 and 1.9. Chapter 5 of this thesis focused on BPSD and tried to explore their aetiology.

Initially, the distribution of different BPSD as defined by the 12-item NPI [148] was assessed in the CAGRAD cohort. Apathy was the most common BPSD domain in this cohort, appearing in almost 60% of the individuals. Apathy is commonly reported as the most prevalent BPSD in individuals with AD, and seems to worsen over the course of disease [324]. Depression was also very common, with a reported prevalence of 54.4%. Depression is common in the elderly and has often been found to precede the emergence of cognitive symptoms in AD [127], [325]. Depression in AD seems to be more aggressive than geriatric depression outside of dementia and does not respond well to antidepressant medication [193]. The least common symptom in the CAGRAD cohort was elation, that appeared in only 10% of the participants. This also concurs with the majority of published literature om BPSD, that confirm that elation is a rare symptom in individuals with AD [326].

Then, the differences in BPSD between individuals with EOAD and LOAD were examined to determine if the age at disease onset is a major determinant of the BPSD phenotype in AD. Anxiety was more common in individuals with EOAD, and there was no significant difference for hallucinations, depression, irritability and elation. All other BPSD were significantly more common in individuals with LOAD. A number of studies have looked at BPSD differences between individuals with EOAD and LOAD in the past. Some have found no significant effect of the age at disease onset on BPSD [299], [310], while other have reported certain differences [277], [309]. Interestingly, Gumus *et al.,* found that there were no significant differences in BPSD between individuals with EOAD and LOAD, with the exception of depression and anxiety that had significantly higher scores in individuals with EOAD [309], a result which is in partial agreement with what was found in Chapter 5.4.1. It should be highlighted that the individuals with EOAD assessed here all have sporadic EOAD, as carrying any of the known autosomal dominant mutations that cause familial AD was an exclusion criterion for CAGRAD. A large

proportion of the published literature on BPSD in EOAD is based on familial EOAD, while in many cases it is not specified whether the individuals carry any of the three known mutations mentioned in Chapter 1.3.1. As mentioned in Chapter 1.6, there is a subtype of familial EOAD that seems to constitute a distinct sub-phenotype of AD, with individuals exhibiting less pronounced cognitive decline and more non-cognitive symptoms [97]. Therefore, when assessing the phenotypic variation of EOAD, and particularly the presence of non-cognitive symptoms, this distinction is crucial.

Instead of examining each NPI symptom domain individually, it was considered preferable to use an appropriate statistical technique in order to detect groups of BPSD that are correlated. The main reasoning behind this was that symptoms that commonly co-occur could share an aetiological mechanism, therefore it is reasonable that they should be examined together when trying to determine their pathogenesis, while the grouping of symptoms also reduces the multiple testing burden. Principal component analysis (PCA) was selected for this purpose. First, individuals with EOAD and LOAD were examined separately, in order to determine if the two groups exhibit different patterns of behavioural sub-phenotypes. The component structure revealed by the PCA was similar for EOAD and LOAD individuals, therefore it was considered reasonable to combine them for a joint PCA and for all subsequent analyses. Five BPSD sub-phenotypes were discovered in this dataset through the PCA: 1) agitation, including aggression, disinhibition and irritability; 2) psychosis, including delusions, hallucinations and AMB; 3) affect, including depression and anxiety; 4) behaviour, including apathy, eating problems and sleeping disturbances; and 5) elation, that constituted a sub-phenotype on its own right. There was some intercorrelation between the sub-phenotypes, the largest observed between behaviour and psychosis (correlation coefficient = 0.63, p-value < 2.2×10^{-16}). The sub-phenotypes derived from the PCA when applying a five-component structure were sound from a clinical perspective, therefore no more component structures or data reduction techniques were examined.

Defining BPSD sub-phenotypes in individuals with AD is a subject that has been explored extensively in the past [278]. The methods used to define the sub-phenotypes varies between the different studies, with some using cluster analysis, others factor analysis and some using PCA, as done in Chapter 5.4.2 [278]. The number of sub-phenotypes discovered also varies between publications, with the majority reporting three, four or five sub-phenotypes [278]. In 2006, Hollingworth *et al.* [277] used a similar method to the one used here and a subset of the

CAGRAD cohort, and identified 4 sub-phenotypes instead of five. The structure of the sub-phenotypes was similar to the one found here, including psychosis, agitation, mood and behavioural dyscontrol [277]. In the publication by Hollingworth *et al*. a number of symptoms did not load highly to any of the sub-phenotypes. In this analysis, all symptoms had a loading higher than 0.4 for at least one of the sub-phenotypes. Moreover, here elation constituted its own sub-phenotype. Both these facts could explain the discrepancy in the number of sub-phenotypes, however the dataset used by Hollingworth *et al.* was only a small part of CAGRAD comprising of just over 1,000 individuals [277], and the differences could stem from the use of different datasets. The sub-phenotype structure discovered in Chapter 5.4.2 is similar to those that have generally been reported in the literature [278]. The majority of published studies have found delusions and hallucinations to form a psychosis sub-phenotype, anxiety and depression to form an affective sub-phenotype, aggression, disinhibition and irritability to form an agitation sub-phenotype and apathy, eating and sleeping problems and AMB to form a behavioural sub-phenotype [278]. Elation has often been found to form its own sub-phenotype [278]. The symptoms whose grouping differed the most between studies are apathy, sleeping problems and AMB [278]. It is important to note here that not all studies have used the 12 domain NPI that includes eating and sleeping problems, with some having used the older 10 domain version [278], which would affect the sub-phenotype structure. Moreover, not all studies on BPSD use the NPI as an assessment tool. Other instruments have also been utilised to assess the presence of BPSD, like the Behaviour Rating Scale for Dementia [327] or the Behavioural Pathology in Alzheimer's Disease Scale [328]. Interestingly, the sub-phenotype structure is similar regardless of the assessment instrument used [278], which indicates that the grouping of symptoms is an inherent characteristic and not specific to the symptom domains as defined by the NPI.

It could be argued that despite being data driven, the definition of these sub-phenotypes is still arbitrary, and that using the raw NPI data in all the genetic and epidemiological analyses performed would have been sounder from a clinical perspective, however there are some important arguments against this. Firstly, examining each NPI symptom domain individually would mean performing complex analyses on 12 separate variables, which would greatly increase the likelihood of type one error (false positive association) due to multiple testing, while correcting for the number of tests would significantly reduce the statistical power of the analyses resulting in false negative associations. Another point to consider here is that the

different NPI symptom domains are intercorrelated and analysing each of them in isolation would fail to account for this. Moreover, it is likely that symptoms that commonly co-occur might share a pathogenetic mechanism and might be influenced by the same variables, therefore they should be examined together when aetiology is concerned. Finally, and most importantly, examining individual NPI symptom domains has limited clinical utility for the majority of them. While symptoms like depression and apathy are common and severe enough to be considered a sub-phenotype alone, others like hallucinations and disinhibition are much rarer and provide limited clinically useful information when considered in isolation. Combining hallucinations with delusions into a psychosis sub-phenotype and disinhibition with irritability and aggression into an agitation sub-phenotype makes any associations that might be detected more interpretable.

After the sub-phenotypes had been defined, factors that could be linked to their appearance were examined. A number of different variables were tested for association with the sub-phenotypes, including disease severity, age at disease onset, age at interview, sex, MMSE score and educational attainment. In order to assess the differences in BPSD sub-phenotypes with disease stage, disease severity as defined by MMSE score [329] and disease duration were both explored. Individuals with severe AD (MMSE score < 14) had significantly higher scores in all five sub-phenotypes, while individuals with a longer duration had a significantly higher score in all sub-phenotypes apart from elation. Similar results have been previously reported [276], [277], [300]. For this reason, some researchers have opted to only study the presence of BPSD in individuals with severe AD, suggesting that the absence of BPSD at an early stage of AD is unreliable, as the individuals could still develop them at a later stage as the disease progresses [142], [277]. While there is merit in that viewpoint, it was decided against for this thesis as it would have led to a substantial decrease in the sample size and the statistical power of the various analyses performed. Moreover, there are no widely accepted timelines of BPSD development, therefore arbitrary thresholds would have to be drawn, which would not necessarily be identical for different BPSD sub-phenotypes. Therefore, it was decided to include all the individuals in the subsequent analyses, regardless of disease stage. Moreover, a small number of studies have examined BPSD longitudinally and have reported conflicting results [301], [302], [330] regarding their changes over the course of the disease. Ideally, monitoring the development of BPSD in the participants over time would help deconvolute the natural history of BPSD development. However, the NPI was only administered to the CAGRAD

participants at the point of recruitment and not repeated over the course of the data collection, therefore that was not possible.

The effect of a number of additional variables on BPSD sub-phenotypes was also explored. Age at disease onset was found to have a positive association with psychosis, agitation and behaviour. When individual NPI symptom domains were compared between individuals with EOAD and LOAD, it was also found that a young age at disease onset was associated with a lower BPSD burden for most symptoms, with the difference being smaller for elation and depression, while anxiety was interestingly more common in individuals with EOAD. Previous publications have found conflicting results regarding the effect of age at disease onset on BPSD prevalence, with some studies showing a higher BPSD burden with older onset [299], [308], others lower burden [277], [309] and others finding no effect [310]. However, as mentioned above, it is important to note that the individuals with EOAD here did not carry any of the causative mutations, as is often the case in literature. Sex was significantly associated with the agitation sub-phenotype, which was more common in men. That result is in line with previously published studies that have found men to me more likely to exhibit agitation and aggression [274]. The origins of this difference are unclear, and it is likely that there is a biological explanation, however it might also be related to stereotypical gender-related behaviours that suggest that women and men might express emotions of aggression and agitation in very different ways, with men being more likely to externalise it. A high educational attainment was associated with lower levels of psychosis, behaviour and agitation sub-phenotypes. This could be an indication that education protects from a more severe AD phenotype, as is suggested by the cognitive reserve hypothesis [50], described in Chapter 1.2.3, which supports that complex mental activity is protective against the detrimental effects of neurodegeneration. While these results are indicative of the type of demographic factors that can influence BPSD burden, it is unclear how exactly these factors exert their effects. It is possible that these factors are somehow implicated in the causative pathways that give rise to BPSD, while also some of these factors could influence the severity of the disease, which in turn leads to a more pronounced phenotype that includes a higher BPSD burden. Further studies with a more direct functional component, like studies using neuroimaging or translational animal studies, are required in order to shed some lights into how these factors could influence BPSD pathogenesis.

The presence of a genetic predisposition to BPSD was also examined. First, *APOE* genotype was explored as a determinant of BPSD severity. The association of the number of *APOE* ε2 and ε4

alleles with the five sub-phenotypes was first explored. *APOE* ε4 as found to be significantly associated with elation and affect, while *APOE* ε2 was not associated with any of the sub-phenotypes. While the possibility of *APOE* genotype having a direct effect on the biological pathways that give rise to these two sub-phenotypes cannot be excluded, it is more likely that this effect is based on the association of *APOE* ε4 with a more severe dementia phenotype, which is likely to involve non-cognitive as well as cognitive symptoms. Others have looked at the association of *APOE* with BPSD, with the number *APOE* ε4 alleles having been inconsistently associated with various BPSD [312], [314]–[316], [318], [320]. It is important to consider the method used to derive the quantitative phenotype used in any genetic studies of BPSD. In this study, component scores extracted from the PCA were used as metrics of BPSD severity for each of the five identified sub-phenotypes. Others have used the raw NPI score when assessing individual NPI domains [315], binary phenotypes [321] or scores derived from various models aimed to group BPSD into sub-phenotypes [277], [299]. Even when using binary phenotypes there is substantial heterogeneity, with some studies considering any presence of symptoms a sufficient phenotype [314] while others using arbitrary cut-offs of the relevant NPI domain scores [321], [331]. This heterogeneity in the definition of BPSD phenotypes is likely to be responsible for some of the inconsistencies in the findings between studies.

Lastly, the presence of a shared polygenic architecture between BPSD and three common psychiatric disorders, IQ and AD was assessed using PRS. IQ PRS was not associated with any of the five sub-phenotypes. IQ was included in the PRS analysis in order to assess if a genetic predisposition to intelligence would be associated with a lower BPSD burden. However, it is important to note that while a high IQ PRS is indicative of higher intelligence, it is not necessarily correlated with more complex mental activity throughout the life course, which is what has been found to mitigate the devastating effects of neurodegeneration [50]. Schizophrenia PRS was significantly associated with all five BPSD sub-phenotypes. What is particularly interesting is that the association to agitation was stronger than to psychosis (lowest p-value= 4.48×10^{-9}, β= 0.10 and 1.75×10^{-8}, β= 0.10 respectively), although it should be noted that the two sub-phenotypes were strongly correlated, which suggests that there is a common mechanism underlying them both. Moreover, it is important to consider that while schizophrenia is a psychotic disorder, with delusions and hallucinations being the prominent symptoms, it has a complex phenotype with a wide variety of accompanying symptoms, and it is unclear which parts of the phenotype of schizophrenia are most genetically determined.

MDD and BD PRS were significantly associated with the affect sub-phenotype (lowest p-value = $7.29x10^{-20}$ and $1.07x10^{-16}$ respectively). The polygenic background of BPSD has not been explored, with the exception of psychosis, the BPSD domain that has been most extensively researched. Psychosis in AD is considered to have a strong genetic component, with heritability estimates ranging between 30 and 60% [141], depending on the way in which psychosis as a phenotype has been defined. Studies exploring the association of schizophrenia PRS with psychosis in AD have previously found conflicting results, with DeMichele-Sweet *et al.* finding a negative association [281] and Creese *et al.* finding a positive association [321].

The results of the PRS analysis performed here suggest that there is a proportion of shared polygenic risk between BPSD and psychiatric conditions. This is a very exciting finding, as it points towards a specific direction regarding the aetiology of BPSD. Most theories on BPSD aetiology have focused on the detrimental results of neurodegeneration on brain morphology and function, suggesting that these changes are responsible for the diverse set of BPSD commonly seen in individuals with AD. However, if a genetic risk for psychiatric conditions is associated with BPSD, it could suggest that the neurodegeneration does not directly cause the BPSD. It may be that the individuals that develop debilitating BPSD have a genetic predisposition to psychopathology that had remained latent throughout their lives due to various protective factors. However, the extensive neurodegeneration of AD could have eroded the compensatory mechanisms that protected against the development of mental illness and brought this latent psychopathology to the surface. Similar associations of PRS derived from psychiatric disorders with psychiatric sub-phenotypes have been observed in Huntington's disease (HD) [322]. Ellis *et al.* found that schizophrenia PRS was associated with an increased risk of developing psychosis and irritability in HD, while MDD PRS was associated with an increased risk of depression in HD. These findings support the theory that BPSD could arise from a combination of risk factors independent of dementia risk, instead of being direct consequences of neurodegeneration. However, this is only an indication, and not a definite proof that this is indeed the origin of BPSD. Without gaining further insight into the pathogenetic mechanisms behind these symptoms it is difficult to reach an understanding regarding their origin.

6.4. Considerations and future directions

An important point to take into account is that the cohorts used in this thesis comprise solely of individuals of Caucasian descent. Therefore, it is likely that they are not representative of other ethnicities, especially as far as the genetic analyses are concerned, as it is widely known that different ethnic groups differ substantially in their genomic makeup. This issue is by no means limited to the cohorts used in this thesis. The vast majority of genetic studies that have been published to date only include Caucasian individuals of north European or north American descent and therefore fail to represent the majority of the human population on the planet. While it is a problem that researchers around the world are actively trying to combat, resolving it will require considerable time and resources [332]. At the moment, most GWAS studies published, and all GWAS used in the PRS analyses described in this thesis, are based on cohorts of Caucasian descent [333], therefore even in the event that the individuals in the cohorts used in this thesis were of a more diverse background, it would have still not been possible to perform the same analyses without a considerable reduction in statistical power. Therefore, any results that have been discussed throughout this thesis might be limited to a specific ethnicity and not necessarily generalisable, so it is important to be cautious about extrapolation of any conclusions made to more diverse populations.

As discussed in detail in Chapter 6.2, the results of Chapter 4 fail to provide a definite answer regarding the presence of a genetic background in cognitive decline in AD. At the point of designing the analyses described in Chapter 4, a number of additional sources of datasets that could be used to increase the sample size and/or attempt to replicate the results were investigated. No publicly available cohort of individuals with AD with longitudinal cognitive assessments and genotyping was identified at that time. Therefore, it was decided to limit the analyses to CAGRAD and ADNI and attempt a replication of the analyses in Chapter 4 using a larger dataset at a later date. However, that was not possible within this thesis due to time limitations. It is, however, a necessary step in confirming the presence or the absence of a genetic determinant of cognitive decline in AD.

While efforts were made to thoroughly explore the genetic epidemiology of BPSD in Chapter 5, there are some additional analyses that could greatly improve their understanding. To date, no GWAS of the full range of BPSD has been published, with existing publications focusing on individual symptoms, mostly psychosis [142], [331]. Therefore, conducting a GWAS for the five

sub-phenotypes would have been of great interest and could have offered valuable insight into the genetic predisposition to BPSD, however that was outside the scope of this thesis due to time limitations. Moreover, attempting to replicating the results of Chapter 5 in an independent dataset would help validate the findings. ADNI has available NPI data and would constitute an excellent replication dataset [323], even if the number of participants with AD in ADNI is not very high. At the point of designing and performing the analyses described in Chapter 5, the subset of ADNI data to which I had access to did not include any behavioural phenotypes, and time limitations did not allow for seeking an alternative dataset for validating the results of Chapter 5. For this reason, a replication of the findings was outside the scope of this thesis, however it would be essential in reinforcing any conclusions drawn from the results of Chapter 5.

6.5. Conclusion

The main aim of this thesis was to explore the vast variability of symptoms experienced by the individuals with AD, and to try and shed some light into the factors that influence this observed heterogeneity, while making use of a large and deeply phenotyped cohort of individuals with AD that was available at Cardiff University. The two main phenotypic areas that I decided to explore were the rate of cognitive decline and the various behavioural and psychological sub-phenotypes, two areas that I considered to be of paramount importance while also relatively underexplored. Several advanced statistical methods were employed in order to quantify the phenotypes of interest, and various analyses were designed in order to investigate their genetic epidemiology. This thesis offers some valuable insight into the genetic mechanism underlying areas of AD that hitherto have been obscure and provides the background upon which fascinating further studies into the genetic epidemiology of AD can be based.

Bibliography

[1] N. Papavramidou, 'The ancient history of dementia', *Neurol. Sci.*, vol. 39, no. 11, pp. 2011–2016, Nov. 2018, doi: 10.1007/s10072-018-3501-4.

[2] Alzheimer, Alöis, 'Über eine eigenartige Erkrankung der Hirnrinde', *Zentralbl Nervenh Psych*, vol. 18, pp. 177–179, 1911.

[3] Perusini, Gaetano, 'Über klinisch und histologisch eigenartige psychische Erkrankungen des späteren Lebensalters.', *Histol. Histopathol. Arb.*, no. 3, 1909.

[4] Kraepelin, Emil, 'Das senile und präsenile Irresein', in *Ein Lehrbuch für Studierende und Ärzte*, Leipzig: Verlag von Johann Ambrosius Barth, 1910, pp. 593–632.

[5] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, 'Forecasting the global burden of Alzheimer's disease', *Alzheimers Dement.*, vol. 3, no. 3, pp. 186–191, Jul. 2007, doi: 10.1016/j.jalz.2007.04.381.

[6] C. P. Ferri *et al.*, 'Global prevalence of dementia: a Delphi consensus study', *Lancet Lond. Engl.*, vol. 366, no. 9503, pp. 2112–2117, Dec. 2005, doi: 10.1016/S0140-6736(05)67889-0.

[7] L. Fratiglioni *et al.*, 'Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group', *Neurology*, vol. 54, no. 11 Suppl 5, pp. S10-15, 2000.

[8] W. A. Kukull *et al.*, 'Dementia and Alzheimer disease incidence: a prospective cohort study', *Arch. Neurol.*, vol. 59, no. 11, pp. 1737–1746, Nov. 2002, doi: 10.1001/archneur.59.11.1737.

[9] R. A. Miech, J. C. S. Breitner, P. P. Zandi, A. S. Khachaturian, J. C. Anthony, and L. Mayer, 'Incidence of AD may decline in the early 90s for men, later for women: The Cache County study', *Neurology*, vol. 58, no. 2, pp. 209–218, Jan. 2002, doi: 10.1212/wnl.58.2.209.

[10] F. Matthews, C. Brayne, and Medical Research Council Cognitive Function and Ageing Study Investigators, 'The incidence of dementia in England and Wales: findings from

the five identical sites of the MRC CFA Study', *PLoS Med.*, vol. 2, no. 8, p. e193, Aug. 2005, doi: 10.1371/journal.pmed.0020193.

[11] 'The incidence of dementia in Canada. The Canadian Study of Health and Aging Working Group', *Neurology*, vol. 55, no. 1, pp. 66–73, Jul. 2000.

[12] E. P. Helzner, N. Scarmeas, S. Cosentino, M. X. Tang, N. Schupf, and Y. Stern, 'Survival in Alzheimer disease: a multiethnic, population-based study of incident cases', *Neurology*, vol. 71, no. 19, pp. 1489–1495, Nov. 2008, doi: 10.1212/01.wnl.0000334278.11022.42.

[13] C. Helmer, P. Joly, L. Letenneur, D. Commenges, and J. F. Dartigues, 'Mortality with dementia: results from a French prospective community-based cohort', *Am. J. Epidemiol.*, vol. 154, no. 7, pp. 642–648, Oct. 2001, doi: 10.1093/aje/154.7.642.

[14] H. F. M. Rhodius-Meester *et al.*, 'Survival in memory clinic cohort is short, even in young-onset dementia', *J. Neurol. Neurosurg. Psychiatry*, vol. 90, no. 6, pp. 726–728, Jun. 2019, doi: 10.1136/jnnp-2018-318820.

[15] 'Deaths registered in England and Wales (series DR) - Office for National Statistics'. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredinenglandandwalesseriesdr/2017 (accessed Apr. 29, 2019).

[16] B. D. James, S. E. Leurgans, L. E. Hebert, P. A. Scherr, K. Yaffe, and D. A. Bennett, 'Contribution of Alzheimer disease to mortality in the United States', *Neurology*, vol. 82, no. 12, pp. 1045–1050, Mar. 2014, doi: 10.1212/WNL.0000000000000240.

[17] Alzheimer's Association, '2015 Alzheimer's disease facts and figures', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 11, no. 3, pp. 332–384, Mar. 2015, doi: 10.1016/j.jalz.2015.02.003.

[18] Mathers, C, Leonardi, M, 'Global burden of dementia in the year 2000'. World Health Organization (WHO), 2003. [Online]. Available: www.who.int/entity/healthinfo/statistics/bod_dementia.pdf.

[19] R. N. Kalaria *et al.*, 'Alzheimer's disease and vascular dementia in developing countries: prevalence, management, and risk factors', *Lancet Neurol.*, vol. 7, no. 9, pp. 812–826, Sep. 2008, doi: 10.1016/S1474-4422(08)70169-8.

[20] F. E. Matthews *et al.*, 'A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results

of the Cognitive Function and Ageing Study I and II', *Lancet Lond. Engl.*, vol. 382, no. 9902, pp. 1405–1412, Oct. 2013, doi: 10.1016/S0140-6736(13)61570-6.

[21] World Health Organization (WHO), 'The global dementia observatory reference guide'. 2019. [Online]. Available: https://apps.who.int/iris/handle/10665/272669

[22] M. Prince *et al.*, *Dementia UK: Second Edition - Overview*. Alzheimer's Society, 2014.

[23] Y. Hou *et al.*, 'Ageing as a risk factor for neurodegenerative disease', *Nat. Rev. Neurol.*, vol. 15, no. 10, pp. 565–581, Oct. 2019, doi: 10.1038/s41582-019-0244-7.

[24] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans, 'Alzheimer disease in the United States (2010-2050) estimated using the 2010 census', *Neurology*, vol. 80, no. 19, pp. 1778–1783, May 2013, doi: 10.1212/WNL.0b013e31828726f5.

[25] T. Ayodele, E. Rogaeva, J. T. Kurup, G. Beecham, and C. Reitz, 'Early-Onset Alzheimer's Disease: What Is Missing in Research?', *Curr. Neurol. Neurosci. Rep.*, vol. 21, no. 2, p. 4, 2021, doi: 10.1007/s11910-020-01090-y.

[26] R. Brookmeyer, S. Gray, and C. Kawas, 'Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset', *Am. J. Public Health*, vol. 88, no. 9, pp. 1337–1342, Sep. 1998, doi: 10.2105/ajph.88.9.1337.

[27] A. Ruitenberg, A. Ott, J. C. van Swieten, A. Hofman, and M. M. Breteler, 'Incidence of dementia: does gender make a difference?', *Neurobiol. Aging*, vol. 22, no. 4, pp. 575–580, Aug. 2001, doi: 10.1016/s0197-4580(01)00231-7.

[28] C. R. Beam, C. Kaneshiro, J. Y. Jang, C. A. Reynolds, N. L. Pedersen, and M. Gatz, 'Differences Between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease', *J. Alzheimers Dis. JAD*, vol. 64, no. 4, pp. 1077–1083, 2018, doi: 10.3233/JAD-180141.

[29] M. van Oijen, F. J. de Jong, J. C. M. Witteman, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler, 'Atherosclerosis and risk for dementia', *Ann. Neurol.*, vol. 61, no. 5, pp. 403–410, May 2007, doi: 10.1002/ana.21073.

[30] M. S. Beeri *et al.*, 'Coronary artery disease is associated with Alzheimer disease neuropathology in APOE4 carriers', *Neurology*, vol. 66, no. 9, pp. 1399–1404, May 2006, doi: 10.1212/01.wnl.0000210447.19748.0b.

[31] A. B. Newman *et al.*, 'Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort', *J. Am. Geriatr. Soc.*, vol. 53, no. 7, pp. 1101–1107, Jul. 2005, doi: 10.1111/j.1532-5415.2005.53360.x.

[32] M. M. Esiri, Z. Nagy, M. Z. Smith, L. Barnetson, and A. D. Smith, 'Cerebrovascular disease and threshold for dementia in the early stages of Alzheimer's disease', *Lancet Lond. Engl.*, vol. 354, no. 9182, pp. 919–920, Sep. 1999, doi: 10.1016/S0140-6736(99)02355-7.

[33] R. A. Whitmer, S. Sidney, J. Selby, S. C. Johnston, and K. Yaffe, 'Midlife cardiovascular risk factors and risk of dementia in late life', *Neurology*, vol. 64, no. 2, pp. 277–281, Jan. 2005, doi: 10.1212/01.WNL.0000149519.47454.F2.

[34] M. Kivipelto *et al.*, 'Midlife vascular risk factors and Alzheimer's disease in later life: longitudinal, population based study', *BMJ*, vol. 322, no. 7300, pp. 1447–1451, Jun. 2001, doi: 10.1136/bmj.322.7300.1447.

[35] C. Qiu, B. Winblad, and L. Fratiglioni, 'The age-dependent relation of blood pressure to cognitive function and dementia', *Lancet Neurol.*, vol. 4, no. 8, pp. 487–499, Aug. 2005, doi: 10.1016/S1474-4422(05)70141-1.

[36] F.-P. Lu, K.-P. Lin, and H.-K. Kuo, 'Diabetes and the risk of multi-system aging phenotypes: a systematic review and meta-analysis', *PloS One*, vol. 4, no. 1, p. e4144, 2009, doi: 10.1371/journal.pone.0004144.

[37] W. Xu, C. Qiu, M. Gatz, N. L. Pedersen, B. Johansson, and L. Fratiglioni, 'Mid- and late-life diabetes in relation to the risk of dementia: a population-based twin study', *Diabetes*, vol. 58, no. 1, pp. 71–77, Jan. 2009, doi: 10.2337/db08-0586.

[38] W. L. Xu, E. von Strauss, C. X. Qiu, B. Winblad, and L. Fratiglioni, 'Uncontrolled diabetes increases the risk of Alzheimer's disease: a population-based cohort study', *Diabetologia*, vol. 52, no. 6, pp. 1031–1039, Jun. 2009, doi: 10.1007/s00125-009-1323-x.

[39] A. L. Fitzpatrick *et al.*, 'Midlife and late-life obesity and the risk of dementia: cardiovascular health study', *Arch. Neurol.*, vol. 66, no. 3, pp. 336–342, Mar. 2009, doi: 10.1001/archneurol.2008.582.

[40] M. A. Beydoun, H. A. Beydoun, and Y. Wang, 'Obesity and central obesity as risk factors for incident dementia and its subtypes: a systematic review and meta-analysis', *Obes. Rev. Off. J. Int. Assoc. Study Obes.*, vol. 9, no. 3, pp. 204–218, May 2008, doi: 10.1111/j.1467-789X.2008.00473.x.

[41] M. Kivipelto *et al.*, 'Obesity and vascular risk factors at midlife and the risk of dementia and Alzheimer disease', *Arch. Neurol.*, vol. 62, no. 10, pp. 1556–1560, Oct. 2005, doi: 10.1001/archneur.62.10.1556.

[42] H. Francis and R. Stevenson, 'The longer-term impacts of Western diet on human cognition and the brain', *Appetite*, vol. 63, pp. 119–128, Apr. 2013, doi: 10.1016/j.appet.2012.12.018.

[43] F. Nourhashémi *et al.*, 'Body mass index and incidence of dementia: the PAQUID study', *Neurology*, vol. 60, no. 1, pp. 117–119, Jan. 2003, doi: 10.1212/01.wnl.0000038910.46217.aa.

[44] S. Floud *et al.*, 'Body mass index, diet, physical inactivity, and the incidence of dementia in 1 million UK women', *Neurology*, vol. 94, no. 2, pp. e123–e132, Jan. 2020, doi: 10.1212/WNL.0000000000008779.

[45] G. Dal Forno, M. T. Palermo, J. E. Donohue, H. Karagiozis, A. B. Zonderman, and C. H. Kawas, 'Depressive symptoms, sex, and risk for Alzheimer's disease', *Ann. Neurol.*, vol. 57, no. 3, pp. 381–387, Mar. 2005, doi: 10.1002/ana.20405.

[46] H. Amieva *et al.*, 'Prodromal Alzheimer's disease: successive emergence of the clinical symptoms', *Ann. Neurol.*, vol. 64, no. 5, pp. 492–498, Nov. 2008, doi: 10.1002/ana.21509.

[47] K. Andersen, A. Lolk, P. Kragh-Sørensen, N. E. Petersen, and A. Green, 'Depression and the risk of Alzheimer disease', *Epidemiol. Camb. Mass*, vol. 16, no. 2, pp. 233–238, Mar. 2005, doi: 10.1097/01.ede.0000152116.32580.24.

[48] R. Andel *et al.*, 'Complexity of work and risk of Alzheimer's disease: a population-based study of Swedish twins', *J. Gerontol. B. Psychol. Sci. Soc. Sci.*, vol. 60, no. 5, pp. P251-258, Sep. 2005, doi: 10.1093/geronb/60.5.p251.

[49] M. J. Valenzuela and P. Sachdev, 'Brain reserve and dementia: a systematic review', *Psychol. Med.*, vol. 36, no. 4, pp. 441–454, Apr. 2006, doi: 10.1017/S0033291705006264.

[50] Y. Stern, 'Cognitive reserve in ageing and Alzheimer's disease', *Lancet Neurol.*, vol. 11, no. 11, pp. 1006–1012, Nov. 2012, doi: 10.1016/S1474-4422(12)70191-6.

[51] M. E. Nelson, D. J. Jester, A. J. Petkus, and R. Andel, 'Cognitive Reserve, Alzheimer's Neuropathology, and Risk of Dementia: A Systematic Review and Meta-Analysis',

Neuropsychol. Rev., vol. 31, no. 2, pp. 233–250, Jun. 2021, doi: 10.1007/s11065-021-09478-4.

[52] M. Hamer and Y. Chida, 'Physical activity and risk of neurodegenerative disease: a systematic review of prospective evidence', *Psychol. Med.*, vol. 39, no. 1, pp. 3–11, Jan. 2009, doi: 10.1017/S0033291708003681.

[53] K. I. Erickson *et al.*, 'Exercise training increases size of hippocampus and improves memory', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 7, pp. 3017–3022, Feb. 2011, doi: 10.1073/pnas.1015950108.

[54] K. I. Erickson *et al.*, 'Physical activity predicts gray matter volume in late adulthood: the Cardiovascular Health Study', *Neurology*, vol. 75, no. 16, pp. 1415–1422, Oct. 2010, doi: 10.1212/WNL.0b013e3181f88359.

[55] R. J. Hardman, G. Kennedy, H. Macpherson, A. B. Scholey, and A. Pipingas, 'Adherence to a Mediterranean-Style Diet and Effects on Cognition in Adults: A Qualitative Evaluation and Systematic Review of Longitudinal and Prospective Trials', *Front. Nutr.*, vol. 3, p. 22, Jul. 2016, doi: 10.3389/fnut.2016.00022.

[56] M. C. Morris *et al.*, 'Association of Seafood Consumption, Brain Mercury Level, and APOE ε4 Status With Brain Neuropathology in Older Adults', *JAMA*, vol. 315, no. 5, pp. 489–497, Feb. 2016, doi: 10.1001/jama.2015.19451.

[57] N. Scarmeas, Y. Stern, R. Mayeux, and J. A. Luchsinger, 'Mediterranean diet, Alzheimer disease, and vascular mediation', *Arch. Neurol.*, vol. 63, no. 12, pp. 1709–1717, Dec. 2006, doi: 10.1001/archneur.63.12.noc60109.

[58] M. Gatz *et al.*, 'Role of genes and environments for explaining Alzheimer disease', *Arch. Gen. Psychiatry*, vol. 63, no. 2, pp. 168–174, Feb. 2006, doi: 10.1001/archpsyc.63.2.168.

[59] T. S. Wingo, J. J. Lah, A. I. Levey, and D. J. Cutler, 'Autosomal recessive causes likely in early-onset Alzheimer disease', *Arch. Neurol.*, vol. 69, no. 1, pp. 59–64, Jan. 2012, doi: 10.1001/archneurol.2011.221.

[60] M.-H. Dai, H. Zheng, L.-D. Zeng, and Y. Zhang, 'The genes associated with early-onset Alzheimer's disease', *Oncotarget*, vol. 9, no. 19, pp. 15132–15143, Mar. 2018, doi: 10.18632/oncotarget.23738.

[61] Y. Cai, S. S. A. An, and S. Kim, 'Mutations in presenilin 2 and its implications in Alzheimer's disease and other dementia-associated disorders', *Clin. Interv. Aging*, vol. 10, pp. 1163–1172, 2015, doi: 10.2147/CIA.S85808.

[62] O. Belbin, H. Beaumont, D. Warden, A. D. Smith, N. Kalsheker, and K. Morgan, 'PSEN1 polymorphisms alter the rate of cognitive decline in sporadic Alzheimer's disease patients', *Neurobiol. Aging*, vol. 30, no. 12, pp. 1992–1999, Dec. 2009, doi: 10.1016/j.neurobiolaging.2008.02.013.

[63] L. Puglielli, R. E. Tanzi, and D. M. Kovacs, 'Alzheimer's disease: the cholesterol connection', *Nat. Neurosci.*, vol. 6, no. 4, p. 345, Apr. 2003, doi: 10.1038/nn0403-345.

[64] E. H. Corder *et al.*, 'Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families', *Science*, vol. 261, no. 5123, pp. 921–923, Aug. 1993, doi: 10.1126/science.8346443.

[65] L. A. Farrer *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*, vol. 278, no. 16, pp. 1349–1356, Oct. 1997.

[66] C.-C. Liu, C.-C. Liu, T. Kanekiyo, H. Xu, and G. Bu, 'Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy', *Nat. Rev. Neurol.*, vol. 9, no. 2, pp. 106–118, Feb. 2013, doi: 10.1038/nrneurol.2012.263.

[67] J. M. Castellano *et al.*, 'Human apoE isoforms differentially regulate brain amyloid-β peptide clearance', *Sci. Transl. Med.*, vol. 3, no. 89, p. 89ra57, Jun. 2011, doi: 10.1126/scitranslmed.3002156.

[68] P. G. Ridge, S. Mukherjee, P. K. Crane, and J. S. K. Kauwe, 'Alzheimer's Disease: Analyzing the Missing Heritability', *PLoS ONE*, vol. 8, no. 11, p. e79771, Nov. 2013, doi: 10.1371/journal.pone.0079771.

[69] D. Harold *et al.*, 'Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease', *Nat. Genet.*, vol. 41, no. 10, pp. 1088–1093, Oct. 2009, doi: 10.1038/ng.440.

[70] J.-C. Lambert *et al.*, 'Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease', *Nat. Genet.*, vol. 41, no. 10, pp. 1094–1099, Oct. 2009, doi: 10.1038/ng.439.

[71] C. Bellenguez *et al.*, 'New insights on the genetic etiology of Alzheimer's and related dementia'. medRxiv, p. 2020.10.01.20200659, Dec. 14, 2020. doi: 10.1101/2020.10.01.20200659.

[72] C. Bellenguez *et al.*, 'Contribution to Alzheimer's disease risk of rare variants in TREM2, SORL1, and ABCA7 in 1779 cases and 1273 controls', *Neurobiol. Aging*, vol. 59, p. 220.e1-220.e9, Nov. 2017, doi: 10.1016/j.neurobiolaging.2017.07.001.

[73] A. C. Naj, G. D. Schellenberg, and for the A. D. G. Consortium (ADGC), 'Genomic variants, genes, and pathways of Alzheimer's disease: An overview', *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, vol. 174, no. 1, pp. 5–26, 2017, doi: 10.1002/ajmg.b.32499.

[74] C. Van Cauwenberghe, C. Van Broeckhoven, and K. Sleegers, 'The genetic landscape of Alzheimer disease: clinical implications and perspectives', *Genet. Med.*, vol. 18, no. 5, pp. 421–430, May 2016, doi: 10.1038/gim.2015.117.

[75] D. P. Perl, 'Neuropathology of Alzheimer's disease', *Mt. Sinai J. Med. N. Y.*, vol. 77, no. 1, pp. 32–42, Feb. 2010, doi: 10.1002/msj.20157.

[76] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman, 'Neuropathological alterations in Alzheimer disease', *Cold Spring Harb. Perspect. Med.*, vol. 1, no. 1, p. a006189, Sep. 2011, doi: 10.1101/cshperspect.a006189.

[77] L. M. Bierer *et al.*, 'Neocortical neurofibrillary tangles correlate with dementia severity in Alzheimer's disease', *Arch. Neurol.*, vol. 52, no. 1, pp. 81–88, Jan. 1995, doi: 10.1001/archneur.1995.00540250089017.

[78] G. A. Marshall, L. A. Fairbanks, S. Tekin, H. V. Vinters, and J. L. Cummings, 'Early-onset Alzheimer's disease is associated with greater pathologic burden', *J. Geriatr. Psychiatry Neurol.*, vol. 20, no. 1, pp. 29–33, Mar. 2007, doi: 10.1177/0891988706297086.

[79] T. Gómez-Isla *et al.*, 'Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease', *Ann. Neurol.*, vol. 41, no. 1, pp. 17–24, Jan. 1997, doi: 10.1002/ana.410410106.

[80] C. R. Overk and E. Masliah, 'Pathogenesis of synaptic degeneration in Alzheimer's disease and Lewy body disease', *Biochem. Pharmacol.*, vol. 88, no. 4, pp. 508–516, Apr. 2014, doi: 10.1016/j.bcp.2014.01.015.

[81] R. D. Terry *et al.*, 'Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment', *Ann. Neurol.*, vol. 30, no. 4, pp. 572–580, Oct. 1991, doi: 10.1002/ana.410300410.

[82] M. T. Heneka *et al.*, 'Neuroinflammation in Alzheimer's disease', *Lancet Neurol.*, vol. 14, no. 4, pp. 388–405, Apr. 2015, doi: 10.1016/S1474-4422(15)70016-5.

[83] S. A. Liddelow *et al.*, 'Neurotoxic reactive astrocytes are induced by activated microglia', *Nature*, vol. 541, no. 7638, pp. 481–487, Jan. 2017, doi: 10.1038/nature21029.

[84] A. K. Vehmas, C. H. Kawas, W. F. Stewart, and J. C. Troncoso, 'Immune reactive cells in senile plaques and cognitive decline in Alzheimer's disease', *Neurobiol. Aging*, vol. 24, no. 2, pp. 321–331, Apr. 2003, doi: 10.1016/s0197-4580(02)00090-8.

[85] R. N. Kalaria, 'Neuropathological diagnosis of vascular cognitive impairment and vascular dementia with implications for Alzheimer's disease', *Acta Neuropathol. (Berl.)*, vol. 131, no. 5, pp. 659–685, May 2016, doi: 10.1007/s00401-016-1571-z.

[86] G. Paroni, P. Bisceglia, and D. Seripa, 'Understanding the Amyloid Hypothesis in Alzheimer's Disease', *J. Alzheimers Dis. JAD*, vol. 68, no. 2, pp. 493–510, 2019, doi: 10.3233/JAD-180802.

[87] D. Scheuner *et al.*, 'Secreted amyloid β–protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease', *Nat. Med.*, vol. 2, no. 8, Art. no. 8, Aug. 1996, doi: 10.1038/nm0896-864.

[88] B. Schönheit, R. Zarski, and T. G. Ohm, 'Spatial and temporal relationships between plaques and tangles in Alzheimer-pathology', *Neurobiol. Aging*, vol. 25, no. 6, pp. 697–711, Jul. 2004, doi: 10.1016/j.neurobiolaging.2003.09.009.

[89] J. L. Price, P. B. Davis, J. C. Morris, and D. L. White, 'The distribution of tangles, plaques and related immunohistochemical markers in healthy aging and Alzheimer's disease', *Neurobiol. Aging*, vol. 12, no. 4, pp. 295–312, Aug. 1991, doi: 10.1016/0197-4580(91)90006-6.

[90] R. Ricciarelli and E. Fedele, 'The Amyloid Cascade Hypothesis in Alzheimer's Disease: It's Time to Change Our Mind', *Curr. Neuropharmacol.*, vol. 15, no. 6, pp. 926–935, 2017, doi: 10.2174/1570159X15666170116143743.

[91] H. Hampel *et al.*, 'The cholinergic system in the pathophysiology and treatment of Alzheimer's disease', *Brain J. Neurol.*, vol. 141, no. 7, pp. 1917–1933, Jul. 2018, doi: 10.1093/brain/awy132.

[92] A. Martorana, Z. Esposito, and G. Koch, 'Beyond the cholinergic hypothesis: do current drugs work in Alzheimer's disease?', *CNS Neurosci. Ther.*, vol. 16, no. 4, pp. 235–245, Aug. 2010, doi: 10.1111/j.1755-5949.2010.00175.x.

[93] M. Sastre, T. Klockgether, and M. T. Heneka, 'Contribution of inflammatory processes to Alzheimer's disease: molecular mechanisms', *Int. J. Dev. Neurosci. Off. J. Int. Soc. Dev. Neurosci.*, vol. 24, no. 2–3, pp. 167–176, May 2006, doi: 10.1016/j.ijdevneu.2005.11.014.

[94] R. G. Nagele, M. R. D'Andrea, H. Lee, V. Venkataraman, and H.-Y. Wang, 'Astrocytes accumulate A beta 42 and give rise to astrocytic amyloid plaques in Alzheimer disease brains', *Brain Res.*, vol. 971, no. 2, pp. 197–209, May 2003, doi: 10.1016/s0006-8993(03)02361-8.

[95] C. M. Karch and A. M. Goate, 'Alzheimer's disease risk genes and mechanisms of disease pathogenesis', *Biol. Psychiatry*, vol. 77, no. 1, pp. 43–51, Jan. 2015, doi: 10.1016/j.biopsych.2014.05.006.

[96] A. Atri, 'The Alzheimer's Disease Clinical Spectrum: Diagnosis and Management', *Med. Clin. North Am.*, vol. 103, no. 2, pp. 263–293, Mar. 2019, doi: 10.1016/j.mcna.2018.10.009.

[97] R. J. Bateman *et al.*, 'Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease', *Alzheimers Res. Ther.*, vol. 3, no. 1, p. 1, Jan. 2011, doi: 10.1186/alzrt59.

[98] N. S. Ryan and M. N. Rossor, 'Correlating familial Alzheimer's disease gene mutations with clinical phenotype', *Biomark. Med.*, vol. 4, no. 1, pp. 99–112, Feb. 2010, doi: 10.2217/bmm.09.92.

[99] A. Joshi, J. M. Ringman, A. S. Lee, K. O. Juarez, and M. F. Mendez, 'Comparison of clinical characteristics between familial and non-familial early onset Alzheimer's disease', *J. Neurol.*, vol. 259, no. 10, pp. 2182–2188, Oct. 2012, doi: 10.1007/s00415-012-6481-y.

[100] S. Baudic, G. D. Barba, M. C. Thibaudet, A. Smagghe, P. Remy, and L. Traykov, 'Executive function deficits in early Alzheimer's disease and their relations with episodic memory', *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.*, vol. 21, no. 1, pp. 15–21, Jan. 2006, doi: 10.1016/j.acn.2005.07.002.

[101] A.-M. Kirova, R. B. Bays, and S. Lagalwar, 'Working Memory and Executive Function Decline across Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease', *BioMed Res. Int.*, vol. 2015, p. 748212, 2015, doi: 10.1155/2015/748212.

[102] T. A. Salthouse, 'Trajectories of normal cognitive aging', *Psychol. Aging*, vol. 34, no. 1, pp. 17–24, Feb. 2019, doi: 10.1037/pag0000288.

[103] M. S. Albert *et al.*, 'The diagnosis of mild cognitive impairment 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. J. Alzheimers Assoc.*, vol. 7, no. 3, pp. 270–279, May 2011, doi: 10.1016/j.jalz.2011.03.008.

[104] F. T. Hane, M. Robinson, B. Y. Lee, O. Bai, Z. Leonenko, and M. S. Albert, 'Recent Progress in Alzheimer's Disease Research, Part 3: Diagnosis and Treatment', *J. Alzheimers Dis.*, vol. 57, no. 3, pp. 645–665, doi: 10.3233/JAD-160907.

[105] N. L. J. Saunders and M. J. Summers, 'Longitudinal deficits to attention, executive, and working memory in subtypes of mild cognitive impairment', *Neuropsychology*, vol. 25, no. 2, pp. 237–248, Mar. 2011, doi: 10.1037/a0021134.

[106] A. J. Mitchell and M. Shiri-Feshki, 'Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies', *Acta Psychiatr. Scand.*, vol. 119, no. 4, pp. 252–265, Apr. 2009, doi: 10.1111/j.1600-0447.2008.01326.x.

[107] F. Clément, S. Gauthier, and S. Belleville, 'Executive functions in mild cognitive impairment: emergence and breakdown of neural plasticity', *Cortex J. Devoted Study Nerv. Syst. Behav.*, vol. 49, no. 5, pp. 1268–1279, May 2013, doi: 10.1016/j.cortex.2012.06.004.

[108] M. E. Soto *et al.*, 'Rapid cognitive decline in Alzheimer's disease. Consensus paper', *J. Nutr. Health Aging*, vol. 12, no. 10, pp. 703–713, Dec. 2008, doi: 10.1007/bf03028618.

[109] L. Carcaillon, K. Pérès, J.-J. Péré, C. Helmer, J.-M. Orgogozo, and J.-F. Dartigues, 'Fast Cognitive Decline at the Time of Dementia Diagnosis: A Major Prognostic Factor for Survival in the Community', *Dement. Geriatr. Cogn. Disord.*, vol. 23, no. 6, pp. 439–445, 2007, doi: 10.1159/000102017.

[110] M. E. Soto *et al.*, 'Predictive value of rapid decline in mini mental state examination in clinical practice for prognosis in Alzheimer's disease', *Dement. Geriatr. Cogn. Disord.*, vol. 26, no. 2, pp. 109–116, 2008, doi: 10.1159/000144073.

[111] K. Stanley and Z. Walker, 'Do patients with young onset Alzheimer's disease deteriorate faster than those with late onset Alzheimer's disease? A review of the literature', *Int. Psychogeriatr.*, vol. 26, no. 12, pp. 1945–1953, Dec. 2014, doi: 10.1017/S1041610214001173.

[112] K. Stanley, T. Whitfield, K. Kuchenbaecker, O. Sanders, T. Stevens, and Z. Walker, 'Rate of Cognitive Decline in Alzheimer's Disease Stratified by Age', *J. Alzheimers Dis. JAD*, vol. 69, no. 4, pp. 1153–1160, 2019, doi: 10.3233/JAD-181047.

[113] M. L. Haaksma, A. Calderón-Larrañaga, M. G. M. Olde Rikkert, R. J. F. Melis, and J. S. Leoutsakos, 'Cognitive and functional progression in Alzheimer disease: A prediction model of latent classes', *Int. J. Geriatr. Psychiatry*, vol. 33, no. 8, pp. 1057–1064, Aug. 2018, doi: 10.1002/gps.4893.

[114] J. T. Tschanz *et al.*, 'Progression of cognitive, functional, and neuropsychiatric symptom domains in a population cohort with Alzheimer dementia: the Cache County Dementia Progression study', *Am. J. Geriatr. Psychiatry Off. J. Am. Assoc. Geriatr. Psychiatry*, vol. 19, no. 6, pp. 532–542, Jun. 2011, doi: 10.1097/JGP.0b013e3181faec23.

[115] R. F. Buckley *et al.*, 'Sex, Amyloid, and APOEε4 and risk of cognitive decline in preclinical Alzheimer's disease: findings from three well-characterized cohorts', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 14, no. 9, pp. 1193–1203, Sep. 2018, doi: 10.1016/j.jalz.2018.04.010.

[116] R. Camicioli and R. Kryscio, 'Nature throws curveballs: APOE and nonlinear decline in Alzheimer disease', *Neurology*, vol. 65, no. 12, pp. 1855–1856, Dec. 2005, doi: 10.1212/01.wnl.0000195338.67756.4b.

[117] B. D. Hoyt, P. J. Massman, C. Schatschneider, N. Cooke, and R. S. Doody, 'Individual Growth Curve Analysis of APOE ε4–Associated Cognitive Decline in Alzheimer Disease', *Arch. Neurol.*, vol. 62, no. 3, pp. 454–459, Mar. 2005, doi: 10.1001/archneur.62.3.454.

[118] M. Shinohara *et al.*, 'APOE2 eases cognitive decline during aging: clinical and preclinical evaluations', *Ann. Neurol.*, vol. 79, no. 5, pp. 758–774, May 2016, doi: 10.1002/ana.24628.

[119] A. E. van der Vlies, E. L. G. E. Koedam, Y. a. L. Pijnenburg, J. W. R. Twisk, P. Scheltens, and W. M. van der Flier, 'Most rapid cognitive decline in APOE epsilon4 negative Alzheimer's disease with early onset', *Psychol. Med.*, vol. 39, no. 11, pp. 1907–1911, Nov. 2009, doi: 10.1017/S0033291709005492.

[120] T.-J. Hsieh *et al.*, 'Association between Alzheimer's disease genes and trajectories of cognitive function decline in Han Chinese in Taiwan', *Aging*, vol. 13, no. 13, pp. 17237–17252, Jul. 2021, doi: 10.18632/aging.203204.

[121] J. Euesden, S. Gowrisankar, A. X. Qu, P. St. Jean, A. R. Hughes, and D. J. Pulford, 'Cognitive Decline in Alzheimer's Disease: Limited Clinical Utility for GWAS or Polygenic Risk Scores in a Clinical Trial Setting', *Genes*, vol. 11, no. 5, May 2020, doi: 10.3390/genes11050501.

[122] R. Sherva *et al.*, 'Genome-wide association study of rate of cognitive decline in Alzheimer's Disease patients identifies novel genes and pathways', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 16, no. 8, pp. 1134–1145, Aug. 2020, doi: 10.1002/alz.12106.

[123] R. Sherva *et al.*, 'Genome-wide association study of the rate of cognitive decline in Alzheimer's disease', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 10, no. 1, pp. 45–52, Jan. 2014, doi: 10.1016/j.jalz.2013.01.008.

[124] D. Wolinsky, K. Drake, and J. Bostwick, 'Diagnosis and Management of Neuropsychiatric Symptoms in Alzheimer's Disease', *Curr. Psychiatry Rep.*, vol. 20, no. 12, p. 117, Oct. 2018, doi: 10.1007/s11920-018-0978-8.

[125] M. Steinberg *et al.*, 'Point and 5-year period prevalence of neuropsychiatric symptoms in dementia: the Cache County Study', *Int. J. Geriatr. Psychiatry*, vol. 23, no. 2, pp. 170–177, Feb. 2008, doi: 10.1002/gps.1858.

[126] Q.-F. Zhao *et al.*, 'The prevalence of neuropsychiatric symptoms in Alzheimer's disease: Systematic review and meta-analysis', *J. Affect. Disord.*, vol. 190, pp. 264–271, Jan. 2016, doi: 10.1016/j.jad.2015.09.069.

[127] B. C. Jost and G. T. Grossberg, 'The Evolution of Psychiatric Symptoms in Alzheimer's Disease: A Natural History Study', *J. Am. Geriatr. Soc.*, vol. 44, no. 9, pp. 1078–1081, 1996, doi: 10.1111/j.1532-5415.1996.tb02942.x.

[128] B. C. Jost and G. T. Grossberg, 'The Natural History of Alzheimer's Disease: A Brain Bank Study', *J. Am. Geriatr. Soc.*, vol. 43, no. 11, pp. 1248–1255, 1995, doi: 10.1111/j.1532-5415.1995.tb07401.x.

[129] M. E. Peters *et al.*, 'Neuropsychiatric symptoms as predictors of progression to severe Alzheimer's dementia and death: the Cache County Dementia Progression Study', *Am. J. Psychiatry*, vol. 172, no. 5, pp. 460–465, May 2015, doi: 10.1176/appi.ajp.2014.14040480.

[130] T. M. Terum, J. R. Andersen, A. Rongve, D. Aarsland, E. J. Svendsboe, and I. Testad, 'The relationship of specific items on the Neuropsychiatric Inventory to caregiver burden in dementia: a systematic review', *Int. J. Geriatr. Psychiatry*, vol. 32, no. 7, pp. 703–717, Jul. 2017, doi: 10.1002/gps.4704.

[131] Y.-A. Chen *et al.*, 'Association Between Caregivers' Burden and Neuropsychiatric Symptoms in Female Patients with Alzheimer's Disease with Varying Dementia Severity', *J. Multidiscip. Healthc.*, vol. 14, pp. 929–940, 2021, doi: 10.2147/JMDH.S298196.

[132] R. F. Allegri *et al.*, 'Neuropsychiatric symptoms as a predictor of caregiver burden in Alzheimer's disease', *Neuropsychiatr. Dis. Treat.*, vol. 2, no. 1, pp. 105–110, Mar. 2006.

[133] B. Kim, J. I. Kim, H. R. Na, K. S. Lee, K.-H. Chae, and S. Kim, 'Factors influencing caregiver burden by dementia severity based on an online database from Seoul dementia management project in Korea', *BMC Geriatr.*, vol. 21, no. 1, p. 649, Nov. 2021, doi: 10.1186/s12877-021-02613-z.

[134] C. Steele, B. Rovner, G. A. Chase, and M. Folstein, 'Psychiatric symptoms and nursing home placement of patients with Alzheimer's disease', *Am. J. Psychiatry*, vol. 147, no. 8, pp. 1049–1051, Aug. 1990, doi: 10.1176/ajp.147.8.1049.

[135] D. L. Murman, Q. Chen, M. C. Powell, S. B. Kuo, C. J. Bradley, and C. C. Colenda, 'The incremental direct costs associated with behavioral symptoms in AD', *Neurology*, vol. 59, no. 11, pp. 1721–1729, Dec. 2002, doi: 10.1212/01.wnl.0000036904.73393.e4.

[136] Y. E. Geda *et al.*, 'Neuropsychiatric symptoms in Alzheimer's disease: Past progress and anticipation of the future', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 9, no. 5, pp. 602–608, Sep. 2013, doi: 10.1016/j.jalz.2012.12.001.

[137] K. L. Lanctôt *et al.*, 'Neuropsychiatric signs and symptoms of Alzheimer's disease: New treatment paradigms', *Alzheimers Dement. Transl. Res. Clin. Interv.*, vol. 3, no. 3, pp. 440–449, Aug. 2017, doi: 10.1016/j.trci.2017.07.001.

[138] G. Šimić *et al.*, 'Monoaminergic neuropathology in Alzheimer's disease', *Prog. Neurobiol.*, vol. 151, pp. 101–138, Apr. 2017, doi: 10.1016/j.pneurobio.2016.04.001.

[139] H. Barron, S. Hafizi, A. C. Andreazza, and R. Mizrahi, 'Neuroinflammation and Oxidative Stress in Psychosis and Psychosis Risk', *Int. J. Mol. Sci.*, vol. 18, no. 3, p. 651, Mar. 2017, doi: 10.3390/ijms18030651.

[140] B. T. Baune *et al.*, 'Inflammatory biomarkers predict depressive, but not anxiety symptoms during aging: the prospective Sydney Memory and Aging Study', *Psychoneuroendocrinology*, vol. 37, no. 9, pp. 1521–1530, Sep. 2012, doi: 10.1016/j.psyneuen.2012.02.006.

[141] S.-A. Bacanu, B. Devlin, K. V. Chowdari, S. T. DeKosky, V. L. Nimgaonkar, and R. A. Sweet, 'Heritability of psychosis in Alzheimer disease', *Am. J. Geriatr. Psychiatry Off. J. Am. Assoc. Geriatr. Psychiatry*, vol. 13, no. 7, pp. 624–627, Jul. 2005, doi: 10.1176/appi.ajgp.13.7.624.

[142] M. A. A. DeMichele-Sweet *et al.*, 'Genome-Wide Association Identifies the First Risk Loci for Psychosis in Alzheimer Disease', *Mol. Psychiatry*, vol. 26, no. 10, pp. 5797–5811, Oct. 2021, doi: 10.1038/s41380-021-01152-8.

[143] M. a. A. DeMichele-Sweet *et al.*, 'Genetic risk for schizophrenia and psychosis in Alzheimer disease', *Mol. Psychiatry*, vol. 23, no. 4, pp. 963–972, Apr. 2018, doi: 10.1038/mp.2017.81.

[144] B. Creese *et al.*, 'Examining the association between genetic liability for schizophrenia and psychotic symptoms in Alzheimer's disease', *Transl. Psychiatry*, vol. 9, no. 1, p. 273, Oct. 2019, doi: 10.1038/s41398-019-0592-5.

[145] D. S. Knopman *et al.*, 'Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology', *Neurology*, vol. 56, no. 9, pp. 1143–1153, May 2001, doi: 10.1212/wnl.56.9.1143.

[146] M. F. Folstein, S. E. Folstein, and P. R. McHugh, '"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician', *J. Psychiatr. Res.*, vol. 12, no. 3, pp. 189–198, Nov. 1975.

[147] Z. S. Nasreddine *et al.*, 'The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment', *J. Am. Geriatr. Soc.*, vol. 53, no. 4, pp. 695–699, Apr. 2005, doi: 10.1111/j.1532-5415.2005.53221.x.

[148] J. L. Cummings, M. Mega, K. Gray, S. Rosenberg-Thompson, D. A. Carusi, and J. Gornbein, 'The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia', *Neurology*, vol. 44, no. 12, pp. 2308–2314, Dec. 1994, doi: 10.1212/wnl.44.12.2308.

[149] *Diagnostic and statistical manual of mental disorders : DSM-IV*. Fourth edition. Washington, DC : American Psychiatric Association, [1994] ©1994, 1994. [Online]. Available: https://search.library.wisc.edu/catalog/999733358502121

[150] G. M. McKhann *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. J. Alzheimers Assoc.*, vol. 7, no. 3, pp. 263–269, May 2011, doi: 10.1016/j.jalz.2011.03.005.

[151] G. Waldemar *et al.*, 'Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline', *Eur. J. Neurol.*, vol. 14, no. 1, pp. e1-26, Jan. 2007, doi: 10.1111/j.1468-1331.2006.01605.x.

[152] D. van Vliet *et al.*, 'Time to diagnosis in young-onset dementia as compared with late-onset dementia', *Psychol. Med.*, vol. 43, no. 2, pp. 423–432, Feb. 2013, doi: 10.1017/S0033291712001122.

[153] B. Draper *et al.*, 'Time to diagnosis in young-onset dementia and its determinants: the INSPIRED study', *Int. J. Geriatr. Psychiatry*, vol. 31, no. 11, pp. 1217–1224, Nov. 2016, doi: 10.1002/gps.4430.

[154] B. T. Hyman *et al.*, 'National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 8, no. 1, pp. 1–13, Jan. 2012, doi: 10.1016/j.jalz.2011.10.007.

[155] H. Braak and E. Braak, 'Neuropathological stageing of Alzheimer-related changes', *Acta Neuropathol. (Berl.)*, vol. 82, no. 4, pp. 239–259, 1991, doi: 10.1007/BF00308809.

[156] S. S. Mirra *et al.*, 'The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease', *Neurology*, vol. 41, no. 4, pp. 479–486, Apr. 1991, doi: 10.1212/wnl.41.4.479.

[157] I. Alafuzoff *et al.*, 'Staging of neurofibrillary pathology in Alzheimer's disease: a study of the BrainNet Europe Consortium', *Brain Pathol. Zurich Switz.*, vol. 18, no. 4, pp. 484–496, Oct. 2008, doi: 10.1111/j.1750-3639.2008.00147.x.

[158] D. W. Dickson *et al.*, 'Identification of normal and pathological aging in prospectively studied nondemented elderly humans', *Neurobiol. Aging*, vol. 13, no. 1, pp. 179–189, Feb. 1992, doi: 10.1016/0197-4580(92)90027-u.

[159] L. O. Wahlund, P. Julin, S. E. Johansson, and P. Scheltens, 'Visual rating and volumetry of the medial temporal lobe on magnetic resonance imaging in dementia: a comparative study', *J. Neurol. Neurosurg. Psychiatry*, vol. 69, no. 5, pp. 630–635, Nov. 2000, doi: 10.1136/jnnp.69.5.630.

[160] P. Scheltens, N. Fox, F. Barkhof, and C. De Carli, 'Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion', *Lancet Neurol.*, vol. 1, no. 1, pp. 13–21, May 2002, doi: 10.1016/s1474-4422(02)00002-9.

[161] M. B. Patwardhan, D. C. McCrory, D. B. Matchar, G. P. Samsa, and O. T. Rutschmann, 'Alzheimer disease: operating characteristics of PET--a meta-analysis', *Radiology*, vol. 231, no. 1, pp. 73–80, Apr. 2004, doi: 10.1148/radiol.2311021620.

[162] K. A. Johnson *et al.*, 'Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 9, no. 1, p. e-1-16, Jan. 2013, doi: 10.1016/j.jalz.2013.01.002.

[163] A. Nordberg, 'PET imaging of amyloid in Alzheimer's disease', *Lancet Neurol.*, vol. 3, no. 9, pp. 519–527, Sep. 2004, doi: 10.1016/S1474-4422(04)00853-1.

[164] V. Welge *et al.*, 'Combined CSF tau, p-tau181 and amyloid-beta 38/40/42 for diagnosing Alzheimer's disease', *J. Neural Transm. Vienna Austria 1996*, vol. 116, no. 2, pp. 203–212, Feb. 2009, doi: 10.1007/s00702-008-0177-6.

[165] O. Hansson, H. Zetterberg, P. Buchhave, E. Londos, K. Blennow, and L. Minthon, 'Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study', *Lancet Neurol.*, vol. 5, no. 3, pp. 228–234, Mar. 2006, doi: 10.1016/S1474-4422(06)70355-6.

[166] N. Mattsson, K. Blennow, and H. Zetterberg, 'CSF Biomarkers', *Ann. N. Y. Acad. Sci.*, vol. 1180, no. 1, pp. 28–35, 2009, doi: 10.1111/j.1749-6632.2009.04944.x.

[167] A. Leuzy, N. C. Cullen, N. Mattsson-Carlgren, and O. Hansson, 'Current advances in plasma and cerebrospinal fluid biomarkers in Alzheimer's disease', *Curr. Opin. Neurol.*, vol. 34, no. 2, pp. 266–274, Apr. 2021, doi: 10.1097/WCO.0000000000000904.

[168] S. Palmqvist *et al.*, 'Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders', *JAMA*, vol. 324, no. 8, pp. 772–781, Aug. 2020, doi: 10.1001/jama.2020.12134.

[169] A. Heyman *et al.*, 'Cerebral infarcts in patients with autopsy-proven Alzheimer's disease: CERAD, part XVIII. Consortium to Establish a Registry for Alzheimer's Disease', *Neurology*, vol. 51, no. 1, pp. 159–162, Jul. 1998, doi: 10.1212/wnl.51.1.159.

[170] M. A. DeTure and D. W. Dickson, 'The neuropathological diagnosis of Alzheimer's disease', *Mol. Neurodegener.*, vol. 14, no. 1, p. 32, Aug. 2019, doi: 10.1186/s13024-019-0333-5.

[171] H. Lewis *et al.*, 'Quantification of Alzheimer pathology in ageing and dementia: age-related accumulation of amyloid-beta(42) peptide in vascular dementia', *Neuropathol. Appl. Neurobiol.*, vol. 32, no. 2, pp. 103–118, Apr. 2006, doi: 10.1111/j.1365-2990.2006.00696.x.

[172] The International Schizophrenia Consortium, 'Common polygenic variation contributes to risk of schizophrenia and bipolar disorder', *Nature*, vol. 460, no. 7256, pp. 748–752, Aug. 2009, doi: 10.1038/nature08185.

[173] V. Escott-Price *et al.*, 'Common polygenic variation enhances risk prediction for Alzheimer's disease', *Brain J. Neurol.*, vol. 138, no. Pt 12, pp. 3673–3684, Dec. 2015, doi: 10.1093/brain/awv268.

[174] V. Escott-Price, A. J. Myers, M. Huentelman, and J. Hardy, 'Polygenic Risk Score Analysis of Pathologically Confirmed Alzheimer's Disease', *Ann. Neurol.*, vol. 82, no. 2, pp. 311–314, Aug. 2017, doi: 10.1002/ana.24999.

[175] E. Baker and V. Escott-Price, 'Polygenic Risk Scores in Alzheimer's Disease: Current Applications and Future Directions', *Front. Digit. Health*, vol. 2, p. 14, 2020, doi: 10.3389/fdgth.2020.00014.

[176] J. R. Harrison, S. Mistry, N. Muskett, and V. Escott-Price, 'From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review', *J. Alzheimers Dis.*, vol. 74, no. 4, pp. 1271–1283, doi: 10.3233/JAD-191233.

[177] J. L. Cummings, G. Tong, and C. Ballard, 'Treatment Combinations for Alzheimer's Disease: Current and Future Pharmacotherapy Options', *J. Alzheimers Dis.*, vol. 67, no. 3, pp. 779–794, doi: 10.3233/JAD-180766.

[178] J. L. Cummings, T. Morstorf, and K. Zhong, 'Alzheimer's disease drug-development pipeline: few candidates, frequent failures', *Alzheimers Res. Ther.*, vol. 6, no. 4, p. 37, 2014, doi: 10.1186/alzrt269.

[179] L. S. Nagamatsu, T. C. Handy, C. L. Hsu, M. Voss, and T. Liu-Ambrose, 'Resistance training promotes cognitive and functional brain plasticity in seniors with probable mild cognitive impairment', *Arch. Intern. Med.*, vol. 172, no. 8, pp. 666–668, Apr. 2012, doi: 10.1001/archinternmed.2012.379.

[180] G. W. Rebok *et al.*, 'Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults', *J. Am. Geriatr. Soc.*, vol. 62, no. 1, pp. 16–24, Jan. 2014, doi: 10.1111/jgs.12607.

[181] D. Schenk *et al.*, 'Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse', *Nature*, vol. 400, no. 6740, pp. 173–177, Jul. 1999, doi: 10.1038/22124.

[182] C. Holmes *et al.*, 'Long-term effects of Abeta42 immunisation in Alzheimer's disease: follow-up of a randomised, placebo-controlled phase I trial', *Lancet Lond. Engl.*, vol. 372, no. 9634, pp. 216–223, Jul. 2008, doi: 10.1016/S0140-6736(08)61075-2.

[183] T. E. Golde, E. H. Koo, K. M. Felsenstein, B. A. Osborne, and L. Miele, 'γ-Secretase inhibitors and modulators', *Biochim. Biophys. Acta*, vol. 1828, no. 12, pp. 2898–2907, Dec. 2013, doi: 10.1016/j.bbamem.2013.06.005.

[184] T. E. Golde, 'Disease-Modifying Therapies for Alzheimer's Disease: More Questions than Answers', *Neurotherapeutics*, pp. 1–19, Feb. 2022, doi: 10.1007/s13311-022-01201-2.

[185] J. A. Hey *et al.*, 'Clinical Pharmacokinetics and Safety of ALZ-801, a Novel Prodrug of Tramiprosate in Development for the Treatment of Alzheimer's Disease', *Clin. Pharmacokinet.*, vol. 57, no. 3, pp. 315–333, Mar. 2018, doi: 10.1007/s40262-017-0608-3.

[186] B. W. Kunkle *et al.*, 'Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing', *Nat. Genet.*, vol. 51, no. 3, p. 414, Mar. 2019, doi: 10.1038/s41588-019-0358-2.

[187] T. Gerhard *et al.*, 'Comparative mortality risks of antipsychotic medications in community-dwelling older adults', *Br. J. Psychiatry J. Ment. Sci.*, vol. 205, no. 1, pp. 44–51, Jul. 2014, doi: 10.1192/bjp.bp.112.122499.

[188] L. S. Schneider, K. Dagerman, and P. S. Insel, 'Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials',

Am. J. Geriatr. Psychiatry Off. J. Am. Assoc. Geriatr. Psychiatry, vol. 14, no. 3, pp. 191–210, Mar. 2006, doi: 10.1097/01.JGP.0000200589.01396.6d.

[189] L. S. Schneider *et al.*, 'Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease', *N. Engl. J. Med.*, vol. 355, no. 15, pp. 1525–1538, Oct. 2006, doi: 10.1056/NEJMoa061240.

[190] S. G. Di Santo, F. Prinelli, F. Adorni, C. Caltagirone, and M. Musicco, 'A meta-analysis of the efficacy of donepezil, rivastigmine, galantamine, and memantine in relation to severity of Alzheimer's disease', *J. Alzheimers Dis. JAD*, vol. 35, no. 2, pp. 349–361, 2013, doi: 10.3233/JAD-122140.

[191] C.-C. Tan *et al.*, 'Efficacy and safety of donepezil, galantamine, rivastigmine, and memantine for the treatment of Alzheimer's disease: a systematic review and meta-analysis', *J. Alzheimers Dis. JAD*, vol. 41, no. 2, pp. 615–631, 2014, doi: 10.3233/JAD-132690.

[192] S. Gauthier, H. Loft, and J. Cummings, 'Improvement in behavioural symptoms in patients with moderate to severe Alzheimer's disease by memantine: a pooled data analysis', *Int. J. Geriatr. Psychiatry*, vol. 23, no. 5, pp. 537–545, May 2008, doi: 10.1002/gps.1949.

[193] A. K. Leonpacher *et al.*, 'Effects of Citalopram on Neuropsychiatric Symptoms in Alzheimer's Dementia: Evidence From the CitAD Study', *Am. J. Psychiatry*, vol. 173, no. 5, pp. 473–480, May 2016, doi: 10.1176/appi.ajp.2016.15020248.

[194] A. P. Porsteinsson *et al.*, 'Effect of citalopram on agitation in Alzheimer disease: the CitAD randomized clinical trial', *JAMA*, vol. 311, no. 7, pp. 682–691, Feb. 2014, doi: 10.1001/jama.2014.93.

[195] A. Kurz *et al.*, 'CORDIAL: cognitive rehabilitation and cognitive-behavioral treatment for early dementia in Alzheimer disease: a multicenter, randomized, controlled trial', *Alzheimer Dis. Assoc. Disord.*, vol. 26, no. 3, pp. 246–253, Sep. 2012, doi: 10.1097/WAD.0b013e318231e46e.

[196] S. Guétin *et al.*, 'Effect of music therapy on anxiety and depression in patients with Alzheimer's type dementia: randomised, controlled study', *Dement. Geriatr. Cogn. Disord.*, vol. 28, no. 1, pp. 36–46, 2009, doi: 10.1159/000229024.

[197] L. Teri *et al.*, 'Exercise plus behavioral management in patients with Alzheimer disease: a randomized controlled trial', *JAMA*, vol. 290, no. 15, pp. 2015–2022, Oct. 2003, doi: 10.1001/jama.290.15.2015.

[198] C. Bellenguez *et al.*, 'New insights on the genetic etiology of Alzheimer's and related dementia', *medRxiv*, p. 2020.10.01.20200659, Dec. 2020, doi: 10.1101/2020.10.01.20200659.

[199] J.-C. Lambert *et al.*, 'Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease', *Nat. Genet.*, vol. 45, no. 12, pp. 1452–1458, Dec. 2013, doi: 10.1038/ng.2802.

[200] H. C. Rossetti, C. Munro Cullum, L. S. Hynan, and L. H. Lacritz, 'The CERAD Neuropsychologic Battery Total Score and the progression of Alzheimer disease', *Alzheimer Dis. Assoc. Disord.*, vol. 24, no. 2, pp. 138–142, Jun. 2010, doi: 10.1097/WAD.0b013e3181b76415.

[201] W. G. Rosen, R. C. Mohs, and K. L. Davis, 'A new rating scale for Alzheimer's disease', *Am. J. Psychiatry*, vol. 141, no. 11, pp. 1356–1364, Nov. 1984, doi: 10.1176/ajp.141.11.1356.

[202] 'The Clinical Dementia Rating (CDR) | Neurology'. https://n.neurology.org/content/43/11/2412.2 (accessed May 06, 2020).

[203] 'Barthel Activities of Daily Living (ADL) Index', *Occas. Pap. R. Coll. Gen. Pract.*, no. 59, p. 24, Apr. 1993.

[204] F. Scogin, N. Rohen, and E. Bailey, 'Geriatric Depression Scale', in *Handbook of psychological assessment in primary care settings*, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2000, pp. 491–508.

[205] S. Das *et al.*, 'Next-generation genotype imputation service and methods', *Nat. Genet.*, vol. 48, no. 10, pp. 1284–1287, Oct. 2016, doi: 10.1038/ng.3656.

[206] 'GRCh37 - hg19 - Genome - Assembly - NCBI'. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/ (accessed Mar. 21, 2022).

[207] S. McCarthy *et al.*, 'A reference panel of 64,976 haplotypes for genotype imputation', *Nat. Genet.*, vol. 48, no. 10, Art. no. 10, Oct. 2016, doi: 10.1038/ng.3643.

[208] R. Development Core Team, 'R: A Language and Environment for Statistical Computing', *R Found Stat Comput*, vol. 1, Jan. 2011.

[209] S. van Buuren and K. Groothuis-Oudshoorn, 'mice: Multivariate Imputation by Chained Equations in R', *J. Stat. Softw.*, vol. 45, pp. 1–67, Dec. 2011, doi: 10.18637/jss.v045.i03.

[210] S. G. Mueller *et al.*, 'The Alzheimer's Disease Neuroimaging Initiative', *Neuroimaging Clin. N. Am.*, vol. 15, no. 4, pp. 869–xii, Nov. 2005, doi: 10.1016/j.nic.2005.09.008.

[211] M. W. Weiner *et al.*, 'The Alzheimer's Disease Neuroimaging Initiative: Progress report and future plans', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 6, no. 3, pp. 202–11.e7, May 2010, doi: 10.1016/j.jalz.2010.03.007.

[212] 'Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014 - PubMed'. https://pubmed.ncbi.nlm.nih.gov/26194320/ (accessed Nov. 16, 2021).

[213] M. W. Weiner *et al.*, 'The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 9, no. 5, pp. e111–e194, Sep. 2013, doi: 10.1016/j.jalz.2013.05.1769.

[214] M. W. Weiner *et al.*, 'The Alzheimer's Disease Neuroimaging Initiative 3: continued innovation for clinical trial improvement', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 13, no. 5, pp. 561–571, May 2017, doi: 10.1016/j.jalz.2016.10.006.

[215] A. J. Saykin *et al.*, 'Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 6, no. 3, pp. 265–273, May 2010, doi: 10.1016/j.jalz.2010.03.013.

[216] N. Franzmeier *et al.*, 'Predicting sporadic Alzheimer's disease progression via inherited Alzheimer's disease-informed machine-learning', *Alzheimers Dement.*, vol. n/a, no. n/a, Feb. 2020, doi: 10.1002/alz.12032.

[217] M. Wang, D. Zhang, D. Shen, and M. Liu, 'Multi-task exclusive relationship learning for alzheimer's disease progression prediction with longitudinal data', *Med. Image Anal.*, vol. 53, pp. 111–122, Apr. 2019, doi: 10.1016/j.media.2019.01.007.

[218] N. Bhagwat, J. D. Viviano, A. N. Voineskos, and M. M. Chakravarty, 'Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data', *PLoS Comput. Biol.*, vol. 14, no. 9, Sep. 2018, doi: 10.1371/journal.pcbi.1006376.

[219] C. Wattmo and Å. K. Wallin, 'Early- versus late-onset Alzheimer's disease in clinical practice: cognitive and global outcomes over 3 years', *Alzheimers Res. Ther.*, vol. 9, no. 1, p. 70, Aug. 2017, doi: 10.1186/s13195-017-0294-2.

[220] N. M. Laird and J. H. Ware, 'Random-Effects Models for Longitudinal Data', *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982, doi: 10.2307/2529876.

[221] K. K. F. Tsoi, J. Y. C. Chan, H. W. Hirai, S. Y. S. Wong, and T. C. Y. Kwok, 'Cognitive Tests to Detect Dementia: A Systematic Review and Meta-analysis', *JAMA Intern. Med.*, vol. 175, no. 9, pp. 1450–1458, Sep. 2015, doi: 10.1001/jamainternmed.2015.2152.

[222] J. S. Benoit, W. Chan, L. Piller, and R. Doody, 'Longitudinal Sensitivity of Alzheimer's Disease Severity Staging', *Am. J. Alzheimers Dis. Other Demen.*, vol. 35, p. 1533317520918719, Dec. 2020, doi: 10.1177/1533317520918719.

[223] 'Applied Regression Analysis and Generalized Linear Models', *SAGE Publications Inc*, Nov. 17, 2021. https://us.sagepub.com/en-us/nam/applied-regression-analysis-and-generalized-linear-models/book237254 (accessed Nov. 23, 2021).

[224] S. A. P. Clouston *et al.*, 'Education and Cognitive Decline: An Integrative Analysis of Global Longitudinal Studies of Cognitive Aging', *J. Gerontol. B. Psychol. Sci. Soc. Sci.*, vol. 75, no. 7, pp. e151–e160, Aug. 2020, doi: 10.1093/geronb/gbz053.

[225] J. N. Barnes, 'Exercise, cognitive function, and aging', *Adv. Physiol. Educ.*, vol. 39, no. 2, pp. 55–62, Jun. 2015, doi: 10.1152/advan.00101.2014.

[226] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-Effects Models Using lme4', *J. Stat. Softw.*, vol. 67, no. 1, Art. no. 1, Oct. 2015, doi: 10.18637/jss.v067.i01.

[227] C. a. R. Martins, A. Oulhaj, C. A. de Jager, and J. H. Williams, 'APOE alleles predict the rate of cognitive decline in Alzheimer disease: a nonlinear model', *Neurology*, vol. 65, no. 12, pp. 1888–1893, Dec. 2005, doi: 10.1212/01.wnl.0000188871.74093.12.

[228] J. L. Del-Aguila *et al.*, 'Assessment of the genetic architecture of Alzheimer's Disease risk in rate of memory decline', *J. Alzheimers Dis. JAD*, vol. 62, no. 2, pp. 745–756, 2018, doi: 10.3233/JAD-170834.

[229] X. R. Chen, Y. Shao, M. J. Sadowski, and Alzheimer's Disease Neuroimaging Initiative, 'Segmented Linear Mixed Model Analysis Reveals Association of the APOEε4 Allele with Faster Rate of Alzheimer's Disease Dementia Progression', *J. Alzheimers Dis. JAD*, vol. 82, no. 3, pp. 921–937, 2021, doi: 10.3233/JAD-210434.

[230] D. J. H. Moss *et al.*, 'Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study', *Lancet Neurol.*, vol. 16, no. 9, pp. 701–711, Sep. 2017, doi: 10.1016/S1474-4422(17)30161-8.

[231] K. Juva, R. Sulkava, T. Erkinjuntti, R. Ylikoski, J. Valvanne, and R. Tilvis, 'Staging the severity of dementia: comparison of clinical (CDR, DSM-III-R), functional (ADL, IADL) and cognitive (MMSE) scales', *Acta Neurol. Scand.*, vol. 90, no. 4, pp. 293–298, 1994, doi: 10.1111/j.1600-0404.1994.tb02724.x.

[232] J. Barnes, J. W. Bartlett, D. A. Wolk, W. M. van der Flier, and C. Frost, 'Disease course varies according to age and symptom length in Alzheimer's disease', *J. Alzheimers Dis. JAD*, vol. 64, no. 2, pp. 631–642, 2018, doi: 10.3233/JAD-170841.

[233] D. Peng *et al.*, 'Demographic and clinical characteristics related to cognitive decline in Alzheimer disease in China', *Medicine (Baltimore)*, vol. 95, no. 26, Jul. 2016, doi: 10.1097/MD.0000000000003727.

[234] M. Canevelli *et al.*, 'Predicting the Rate of Cognitive Decline in Alzheimer Disease: Data From the ICTUS Study', *Alzheimer Dis. Assoc. Disord.*, vol. 30, no. 3, pp. 237–242, Sep. 2016, doi: 10.1097/WAD.0000000000000124.

[235] L. Nelson and N. Tabet, 'Slowing the progression of Alzheimer's disease; what works?', *Ageing Res. Rev.*, vol. 23, pp. 193–209, Sep. 2015, doi: 10.1016/j.arr.2015.07.002.

[236] R. Katzman *et al.*, 'Comparison of rate of annual change of mental status score in four independent studies of patients with Alzheimer's disease', *Ann. Neurol.*, vol. 24, no. 3, pp. 384–389, Sep. 1988, doi: 10.1002/ana.410240306.

[237] R. S. Eldholm *et al.*, 'Progression of Alzheimer's Disease: A Longitudinal Study in Norwegian Memory Clinics', *J. Alzheimers Dis. JAD*, vol. 61, no. 3, pp. 1221–1232, 2018, doi: 10.3233/JAD-170436.

[238] V. N. Pavlik, R. S. Doody, P. J. Massman, and W. Chan, 'Influence of premorbid IQ and education on progression of Alzheimer's disease', *Dement. Geriatr. Cogn. Disord.*, vol. 22, no. 4, pp. 367–377, 2006, doi: 10.1159/000095640.

[239] M. G. Jansen *et al.*, 'Positive Effects of Education on Cognitive Functioning Depend on Clinical Status and Neuropathological Severity', *Front. Hum. Neurosci.*, vol. 15, p. 498, 2021, doi: 10.3389/fnhum.2021.723728.

[240] G. A. Edwards III, N. Gamez, G. Escobedo Jr., O. Calderon, and I. Moreno-Gonzalez, 'Modifiable Risk Factors for Alzheimer's Disease', *Front. Aging Neurosci.*, vol. 11, p. 146, 2019, doi: 10.3389/fnagi.2019.00146.

[241] K. Ito *et al.*, 'Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database', *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 7, no. 2, pp. 151–160, Mar. 2011, doi: 10.1016/j.jalz.2010.03.018.

[242] N. Hirono, M. Hashimoto, M. Yasuda, H. Kazui, and E. Mori, 'Accelerated Memory Decline in Alzheimer's Disease With Apolipoprotein ε4 Allele', *J. Neuropsychiatry Clin. Neurosci.*, vol. 15, no. 3, pp. 354–358, Aug. 2003, doi: 10.1176/jnp.15.3.354.

[243] C. L. Allan and K. P. Ebmeier, 'The influence of ApoE4 on clinical progression of dementia: a meta-analysis', *Int. J. Geriatr. Psychiatry*, vol. 26, no. 5, pp. 520–526, May 2011, doi: 10.1002/gps.2559.

[244] T. Kleiman *et al.*, 'Apolipoprotein E ε4 Allele Is Unrelated to Cognitive or Functional Decline in Alzheimer's Disease: Retrospective and Prospective Analysis', *Dement. Geriatr. Cogn. Disord.*, vol. 22, no. 1, pp. 73–82, 2006, doi: 10.1159/000093316.

[245] M. A. Scelsi *et al.*, 'Genetic study of multimodal imaging Alzheimer's disease progression score implicates novel loci', *Brain J. Neurol.*, vol. 141, no. 7, pp. 2167–2180, Jul. 2018, doi: 10.1093/brain/awy141.

[246] H.-R. Kim, T. Lee, J. K. Choi, Y. Jeong, and for the A. D. N. Initiative, 'Genetic variants beyond amyloid and tau associated with cognitive decline: A cohort study', *Neurology*, vol. 95, no. 17, pp. e2366–e2377, Oct. 2020, doi: 10.1212/WNL.0000000000010724.

[247] S. Purcell *et al.*, 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *Am. J. Hum. Genet.*, vol. 81, no. 3, pp. 559–575, Sep. 2007, doi: 10.1086/519795.

[248] C. J. Willer, Y. Li, and G. R. Abecasis, 'METAL: fast and efficient meta-analysis of genomewide association scans', *Bioinforma. Oxf. Engl.*, vol. 26, no. 17, pp. 2190–2191, Sep. 2010, doi: 10.1093/bioinformatics/btq340.

[249] C. A. de Leeuw, J. M. Mooij, T. Heskes, and D. Posthuma, 'MAGMA: Generalized Gene-Set Analysis of GWAS Data', *PLoS Comput. Biol.*, vol. 11, no. 4, Apr. 2015, doi: 10.1371/journal.pcbi.1004219.

[250] 1000 Genomes Project Consortium *et al.*, 'A map of human genome variation from population-scale sequencing', *Nature*, vol. 467, no. 7319, pp. 1061–1073, Oct. 2010, doi: 10.1038/nature09534.

[251] C. A. de Leeuw, B. M. Neale, T. Heskes, and D. Posthuma, 'The statistical properties of gene-set analysis', *Nat. Rev. Genet.*, vol. 17, no. 6, Art. no. 6, Jun. 2016, doi: 10.1038/nrg.2016.29.

[252] M. Ashburner *et al.*, 'Gene ontology: tool for the unification of biology. The Gene Ontology Consortium', *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000, doi: 10.1038/75556.

[253] 'KEGG: kyoto encyclopedia of genes and genomes - PubMed'. https://pubmed.ncbi.nlm.nih.gov/10592173/ (accessed Mar. 25, 2022).

[254] 'PID: the Pathway Interaction Database - PubMed'. https://pubmed.ncbi.nlm.nih.gov/18832364/ (accessed Mar. 25, 2022).

[255] A. Fabregat *et al.*, 'The Reactome pathway Knowledgebase', *Nucleic Acids Res.*, vol. 44, no. D1, pp. D481-487, Jan. 2016, doi: 10.1093/nar/gkv1351.

[256] B. J. Vilhjálmsson *et al.*, 'Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores', *Am. J. Hum. Genet.*, vol. 97, no. 4, pp. 576–592, Oct. 2015, doi: 10.1016/j.ajhg.2015.09.001.

[257] A. Kiezun *et al.*, 'Exome sequencing and the genetic basis of complex traits', *Nat. Genet.*, vol. 44, no. 6, pp. 623–630, May 2012, doi: 10.1038/ng.2303.

[258] S. A. Mofrad, A. J. Lundervold, A. Vik, and A. S. Lundervold, 'Cognitive and MRI trajectories for prediction of Alzheimer's disease', *Sci. Rep.*, vol. 11, no. 1, p. 2122, Jan. 2021, doi: 10.1038/s41598-020-78095-7.

[259] 'GRCh37.p13 - hg19 - Genome - Assembly - NCBI'. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.25/ (accessed Mar. 26, 2022).

[260] G. Trinchieri, 'Interleukin-12 and the regulation of innate resistance and adaptive immunity', *Nat. Rev. Immunol.*, vol. 3, no. 2, pp. 133–146, Feb. 2003, doi: 10.1038/nri1001.

[261] J. W. Kinney, S. M. Bemiller, A. S. Murtishaw, A. M. Leisgang, A. M. Salazar, and B. T. Lamb, 'Inflammation as a central mechanism in Alzheimer's disease', *Alzheimers Dement. Transl. Res. Clin. Interv.*, vol. 4, pp. 575–590, Sep. 2018, doi: 10.1016/j.trci.2018.06.014.

[262] C. M. Kloske *et al.*, 'Inflammatory Pathways Are Impaired in Alzheimer Disease and Differentially Associated With Apolipoprotein E Status', *J. Neuropathol. Exp. Neurol.*, vol. 80, no. 10, pp. 922–932, Oct. 2021, doi: 10.1093/jnen/nlab085.

[263] J. Hazen *et al.*, 'The Association Between Circulating Inflammatory Markers and the Progression of Alzheimer Disease in Norwegian Memory Clinic Patients With Mild Cognitive Impairment or Dementia', *Alzheimer Dis. Assoc. Disord.*, vol. 34, no. 1, pp. 47–53, Mar. 2020, doi: 10.1097/WAD.0000000000000342.

[264] F. Dudbridge, 'Power and Predictive Accuracy of Polygenic Risk Scores', *PLOS Genet.*, vol. 9, no. 3, p. e1003348, Mar. 2013, doi: 10.1371/journal.pgen.1003348.

[265] S. I. Finkel, J. Costa e Silva, G. Cohen, S. Miller, and N. Sartorius, 'Behavioral and psychological signs and symptoms of dementia: a consensus statement on current knowledge and implications for research and treatment', *Int. Psychogeriatr.*, vol. 8 Suppl 3, pp. 497–500, 1996, doi: 10.1017/s1041610297003943.

[266] E. C. Hersch and S. Falzgraf, 'Management of the behavioral and psychological symptoms of dementia', *Clin. Interv. Aging*, vol. 2, no. 4, pp. 611–621, Jan. 2008, doi: 10.2147/CIA.S1698.

[267] G. M. Savva *et al.*, 'Prevalence, correlates and course of behavioural and psychological symptoms of dementia in the population', *Br. J. Psychiatry J. Ment. Sci.*, vol. 194, no. 3, pp. 212–219, Mar. 2009, doi: 10.1192/bjp.bp.108.049619.

[268] C. G. Lyketsos *et al.*, 'Major and minor depression in Alzheimer's disease: prevalence and impact', *J. Neuropsychiatry Clin. Neurosci.*, vol. 9, no. 4, pp. 556–561, 1997, doi: 10.1176/jnp.9.4.556.

[269] T. González-Salvador *et al.*, 'Quality of life in dementia patients in long-term care', *Int. J. Geriatr. Psychiatry*, vol. 15, no. 2, pp. 181–189, Feb. 2000, doi: 10.1002/(sici)1099-1166(200002)15:2<181::aid-gps96>3.0.co;2-i.

[270] C. T. Chen *et al.*, 'Neuropsychiatric symptoms in Alzheimer's disease: associations with caregiver burden and treatment outcomes', *QJM Int. J. Med.*, vol. 110, no. 9, pp. 565–570, Sep. 2017, doi: 10.1093/qjmed/hcx077.

[271] S. Toot, T. Swinson, M. Devine, D. Challis, and M. Orrell, 'Causes of nursing home placement for older people with dementia: a systematic review and meta-analysis', *Int. Psychogeriatr.*, vol. 29, no. 2, pp. 195–208, Feb. 2017, doi: 10.1017/S1041610216001654.

[272] J. Wancata, J. Windhaber, M. Krautgartner, and R. Alexandrowicz, 'The consequences of non-cognitive symptoms of dementia in medical hospital departments', *Int. J. Psychiatry Med.*, vol. 33, no. 3, pp. 257–271, 2003, doi: 10.2190/ABXK-FMWG-98YP-D1CU.

[273] K. Ornstein and J. E. Gaugler, 'The problem with "problem behaviors": a systematic review of the association between individual patient behavioral and psychological symptoms and caregiver depression and burden within the dementia patient–caregiver dyad', *Int. Psychogeriatr.*, vol. 24, no. 10, pp. 1536–1552, Oct. 2012, doi: 10.1017/S1041610212000737.

[274] T. Kitamura, M. Kitamura, S. Hino, N. Tanaka, and K. Kurata, 'Gender differences in clinical manifestations and outcomes among hospitalized patients with behavioral and psychological symptoms of dementia', *J. Clin. Psychiatry*, vol. 73, no. 12, pp. 1548–1554, Dec. 2012, doi: 10.4088/JCP.11m07614.

[275] H. Lövheim, P.-O. Sandman, S. Karlsson, and Y. Gustafson, 'Sex differences in the prevalence of behavioral and psychological symptoms of dementia', *Int. Psychogeriatr.*, vol. 21, no. 3, pp. 469–475, Jun. 2009, doi: 10.1017/S1041610209008497.

[276] G. B. Frisoni *et al.*, 'Behavioral syndromes in Alzheimer's disease: description and correlates', *Dement. Geriatr. Cogn. Disord.*, vol. 10, no. 2, pp. 130–138, Apr. 1999, doi: 10.1159/000017113.

[277] P. Hollingworth *et al.*, 'Four Components Describe Behavioral Symptoms in 1,120 Individuals with Late-Onset Alzheimer's Disease', *J. Am. Geriatr. Soc.*, vol. 54, no. 9, pp. 1348–1354, 2006, doi: 10.1111/j.1532-5415.2006.00854.x.

[278] R. M. van der Linde, T. Dening, F. E. Matthews, and C. Brayne, 'Grouping of behavioural and psychological symptoms of dementia', *Int. J. Geriatr. Psychiatry*, vol. 29, no. 6, pp. 562–568, Jun. 2014, doi: 10.1002/gps.4037.

[279] M. Flirski, T. Sobow, and I. Kloszewska, 'Behavioural genetics of Alzheimer's disease: a comprehensive review', *Arch. Med. Sci. AMS*, vol. 7, no. 2, pp. 195–210, Apr. 2011, doi: 10.5114/aoms.2011.22068.

[280] V. I. Zannis *et al.*, 'Proposed nomenclature of apoE isoproteins, apoE genotypes, and phenotypes', *J. Lipid Res.*, vol. 23, no. 6, pp. 911–914, Aug. 1982.

[281] M. a. A. DeMichele-Sweet *et al.*, 'Genetic risk for schizophrenia and psychosis in Alzheimer disease', *Mol. Psychiatry*, vol. 23, no. 4, pp. 963–972, Apr. 2018, doi: 10.1038/mp.2017.81.

[282] C. K. Y. Lai, 'The merits and problems of Neuropsychiatric Inventory as an assessment tool in people with dementia and other neurological disorders', *Clin. Interv. Aging*, vol. 9, pp. 1051–1061, 2014, doi: 10.2147/CIA.S63504.

[283] I. T. Jolliffe, Ed., 'Choosing a Subset of Principal Components or Variables', in *Principal Component Analysis*, New York, NY: Springer, 2002, pp. 111–149. doi: 10.1007/0-387-22440-8_6.

[284] J. E. Jackson, *A User's Guide to Principal Components*. John Wiley & Sons, 2005.

[285] I. T. Jolliffe and J. Cadima, 'Principal component analysis: a review and recent developments', *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.

[286] F. Drasgow, 'Polychoric and polyserial correlations', *Encycl. Stat.*, vol. 7, pp. 68–74, Jan. 1986.

[287] Fox, John, 'polycor: Polychoric and Polyserial Correlations. R package version 0.8-0. https://CRAN.R-project.org/package=polycor'. 2021. [Online]. Available: https://CRAN.R-project.org/package=polycor

[288] Revelle, William, *psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.1.9*. 2017. Accessed: Jan. 10, 2022. [Online]. Available: https://CRAN.R-project.org/package=psych

[289] *National Institute for Health and Care Excellence: Guidelines*. National Institute for Health and Care Excellence (UK), 2003.

[290] M. N *et al.*, 'Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology', *Nat. Genet.*, vol. 53, no. 6, Jun. 2021, doi: 10.1038/s41588-021-00857-4.

[291] D. M. Howard *et al.*, 'Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions', *Nat. Neurosci.*, vol. 22, no. 3, pp. 343–352, Mar. 2019, doi: 10.1038/s41593-018-0326-7.

[292] J. E. Savage *et al.*, 'Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence', *Nat. Genet.*, vol. 50, no. 7, pp. 912–919, 2018, doi: 10.1038/s41588-018-0152-6.

[293] 'The Scree Test For The Number Of Factors: Multivariate Behavioral Research: Vol 1, No 2'. https://www.tandfonline.com/doi/abs/10.1207/s15327906mbr0102_10 (accessed Mar. 23, 2022).

[294] S.-T. Cheng, T. Kwok, and L. C. W. Lam, 'Neuropsychiatric symptom clusters of Alzheimer's disease in Hong Kong Chinese: prevalence and confirmatory factor analysis of the Neuropsychiatric Inventory', *Int. Psychogeriatr.*, vol. 24, no. 9, pp. 1465–1473, Sep. 2012, doi: 10.1017/S1041610212000609.

[295] S. Germain *et al.*, 'Does Cognitive Impairment Influence Burden in Caregivers of Patients with Alzheimer's Disease?', *J. Alzheimers Dis. JAD*, vol. 17, pp. 105–14, May 2009, doi: 10.3233/JAD-2009-1016.

[296] M. Poletti, A. Nuti, G. Cipriani, and U. Bonuccelli, 'Behavioral and Psychological Symptoms of Dementia: Factor Analysis and Relationship with Cognitive Impairment', *Eur. Neurol.*, vol. 69, no. 2, pp. 76–82, 2013, doi: 10.1159/000341956.

[297] S. U. Zuidema, J. F. M. de Jonghe, F. R. J. Verhey, and R. T. C. M. Koopmans, 'Neuropsychiatric Symptoms in Nursing Home Patients: Factor Structure Invariance of the Dutch Nursing Home Version of the Neuropsychiatric Inventory in Different Stages of Dementia', *Dement. Geriatr. Cogn. Disord.*, vol. 24, no. 3, pp. 169–176, 2007, doi: 10.1159/000105603.

[298] P. Aalten *et al.*, 'Neuropsychiatric Syndromes in Dementia', *Dement. Geriatr. Cogn. Disord.*, vol. 24, no. 6, pp. 457–463, 2007, doi: 10.1159/000110738.

[299] P. Proitsi *et al.*, 'A Multiple Indicators Multiple Causes (MIMIC) model of Behavioural and Psychological Symptoms in Dementia (BPSD)', *Neurobiol. Aging*, vol. 32, no. 3, pp. 434–442, Mar. 2011, doi: 10.1016/j.neurobiolaging.2009.03.005.

[300] G. Spalletta *et al.*, 'Neuropsychiatric symptoms and syndromes in a large cohort of newly diagnosed, untreated patients with Alzheimer disease', *Am. J. Geriatr. Psychiatry Off. J. Am. Assoc. Geriatr. Psychiatry*, vol. 18, no. 11, pp. 1026–1035, Nov. 2010, doi: 10.1097/JGP.0b013e3181d6b68d.

[301] S. Gonfrier, S. Andrieu, D. Renaud, B. Vellas, and P. H. Robert, 'Course of neuropsychiatric symptoms during a 4-year follow up in the REAL-FR cohort', *J. Nutr. Health Aging*, vol. 16, no. 2, pp. 134–137, Feb. 2012, doi: 10.1007/s12603-011-0147-9.

[302] J. Garre-Olmo, S. López-Pousa, J. Vilalta-Franch, M. de Gracia Blanco, and A. B. Vilarrasa, 'Grouping and Trajectories of Neuropsychiatric Symptoms in Patients with Alzheimer's Disease. Part II: Two-Year Patient Trajectories', *J. Alzheimers Dis.*, vol. 22, no. 4, pp. 1169–1180, Jan. 2010, doi: 10.3233/JAD-2010-101215.

[303] K.-H. Chang, C.-C. Huang, C.-M. Chen, H.-C. Wu, and H.-C. Kuo, 'Differences in Clinical Presentation of Behavioral and Psychological Symptoms of Dementia in Alzheimer's Disease According to Sex and Education Level', *J. Alzheimers Dis. JAD*, vol. 78, no. 2, pp. 711–719, 2020, doi: 10.3233/JAD-200507.

[304] W. J. Deardorff and G. T. Grossberg, 'Behavioral and psychological symptoms in Alzheimer's dementia and vascular dementia', *Handb. Clin. Neurol.*, vol. 165, pp. 5–32, 2019, doi: 10.1016/B978-0-444-64012-3.00002-2.

[305] A. Kolanowski *et al.*, 'Determinants of behavioral and psychological symptoms of dementia: A scoping review of the evidence', *Nurs. Outlook*, vol. 65, no. 5, pp. 515–529, Oct. 2017, doi: 10.1016/j.outlook.2017.06.006.

[306] F. Kosel, J. M. S. Pelley, and T. B. Franklin, 'Behavioural and psychological symptoms of dementia in mouse models of Alzheimer's disease-related pathology', *Neurosci. Biobehav. Rev.*, vol. 112, pp. 634–647, May 2020, doi: 10.1016/j.neubiorev.2020.02.012.

[307] N. B. Farber *et al.*, 'Increased neocortical neurofibrillary tangle density in subjects with Alzheimer disease and psychosis', *Arch. Gen. Psychiatry*, vol. 57, no. 12, pp. 1165–1173, Dec. 2000, doi: 10.1001/archpsyc.57.12.1165.

[308] D. van Vliet *et al.*, 'Prevalence of neuropsychiatric symptoms in young-onset compared to late-onset Alzheimer's disease - part 1: findings of the two-year longitudinal NeedYD-study', *Dement. Geriatr. Cogn. Disord.*, vol. 34, no. 5–6, pp. 319–327, 2012, doi: 10.1159/000342824.

[309] M. Gumus, N. Multani, M. L. Mack, and M. C. Tartaglia, 'Progression of neuropsychiatric symptoms in young-onset versus late-onset Alzheimer's disease', *GeroScience*, vol. 43, no. 1, pp. 213–223, Jan. 2021, doi: 10.1007/s11357-020-00304-y.

[310] M. do C. Ferreira, M. J. Abreu, C. Machado, B. Santos, Á. Machado, and A. S. Costa, 'Neuropsychiatric Profile in Early Versus Late Onset Alzheimer's Disease', *Am. J. Alzheimers Dis. Other Demen.*, vol. 33, no. 2, pp. 93–99, Mar. 2018, doi: 10.1177/1533317517744061.

[311] J. Kim, J. M. Basak, and D. M. Holtzman, 'The role of apolipoprotein E in Alzheimer's disease', *Neuron*, vol. 63, no. 3, pp. 287–303, Aug. 2009, doi: 10.1016/j.neuron.2009.06.026.

[312] D. L. Woods, B. Bushnell, H. Kim, D. Geschwind, and J. Cummings, 'Apolipoprotein epsilon4 status is associated with behavioral symptoms in nursing home residents with dementia', *Int. Psychogeriatr.*, vol. 21, no. 4, pp. 722–728, Aug. 2009, doi: 10.1017/S1041610209009235.

[313] K. F. Zdanys *et al.*, 'Apolipoprotein E epsilon4 allele increases risk for psychotic symptoms in Alzheimer's disease', *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.*, vol. 32, no. 1, pp. 171–179, Jan. 2007, doi: 10.1038/sj.npp.1301148.

[314] D. Craig, D. Hart, K. McCool, S. McIlroy, and A. Passmore, 'Apolipoprotein E e4 allele influences aggressive behaviour in Alzheimer's disease', *J. Neurol. Neurosurg. Psychiatry*, vol. 75, no. 9, pp. 1327–1330, Sep. 2004, doi: 10.1136/jnnp.2003.032276.

[315] J. Robertson, J. Curley, J. Kaye, J. Quinn, T. Pfankuch, and J. Raber, 'apoE isoforms and measures of anxiety in probable AD patients and Apoe-/- mice', *Neurobiol. Aging*, vol. 26, no. 5, pp. 637–643, May 2005, doi: 10.1016/j.neurobiolaging.2004.06.003.

[316] G. Spalletta, S. Bernardini, L. Bellincampi, G. Federici, A. Trequattrini, and C. Caltagirone, 'Delusion symptoms are associated with ApoE epsilon4 allelic variant at the early stage of Alzheimer's disease with late onset', *Eur. J. Neurol.*, vol. 13, no. 2, pp. 176–182, Feb. 2006, doi: 10.1111/j.1468-1331.2006.01165.x.

[317] R. A. Sweet *et al.*, 'Apolipoprotein E and alpha-1-antichymotrypsin genotypes do not predict time to psychosis in Alzheimer's disease', *J. Geriatr. Psychiatry Neurol.*, vol. 15, no. 1, pp. 24–30, 2002, doi: 10.1177/089198870201500106.

[318] B. Borroni *et al.*, 'Role of BDNF Val66Met functional polymorphism in Alzheimer's disease-related depression', *Neurobiol. Aging*, vol. 30, no. 9, pp. 1406–1412, Sep. 2009, doi: 10.1016/j.neurobiolaging.2007.11.023.

[319] E. Grünblatt *et al.*, 'Genetic risk factors and markers for Alzheimer's disease and/or depression in the VITA study', *J. Psychiatr. Res.*, vol. 43, no. 3, pp. 298–308, Jan. 2009, doi: 10.1016/j.jpsychires.2008.05.008.

[320] A. L. Pritchard *et al.*, 'The effect of the apolipoprotein E gene polymorphisms and haplotypes on behavioural and psychological symptoms in probable Alzheimer's disease', *J. Neurol. Neurosurg. Psychiatry*, vol. 78, no. 2, pp. 123–126, Feb. 2007, doi: 10.1136/jnnp.2006.092122.

[321] B. Creese *et al.*, 'Examining the association between genetic liability for schizophrenia and psychotic symptoms in Alzheimer's disease', *Transl. Psychiatry*, vol. 9, no. 1, p. 273, Oct. 2019, doi: 10.1038/s41398-019-0592-5.

[322] N. Ellis *et al.*, 'Genetic Risk Underlying Psychiatric and Cognitive Symptoms in Huntington's Disease', *Biol. Psychiatry*, vol. 87, no. 9, pp. 857–865, May 2020, doi: 10.1016/j.biopsych.2019.12.010.

[323] S. G. Mueller *et al.*, 'The Alzheimer's Disease Neuroimaging Initiative', *Neuroimaging Clin. N. Am.*, vol. 15, no. 4, pp. 869–xii, Nov. 2005, doi: 10.1016/j.nic.2005.09.008.

[324] A. L. Teixeira, M. M. Gonzales, L. C. de Souza, and S. L. Weisenbach, 'Revisiting Apathy in Alzheimer's Disease: From Conceptualization to Therapeutic Approaches', *Behav. Neurol.*, vol. 2021, p. 6319826, Aug. 2021, doi: 10.1155/2021/6319826.

[325] K. Javaherian *et al.*, 'Examining the Complicated Relationship Between Depressive Symptoms and Cognitive Impairment in Preclinical Alzheimer Disease.', *Alzheimer Dis. Assoc. Disord.*, vol. 33, no. 1, pp. 15–20, 2019, doi: 10.1097/WAD.0000000000000284.

[326] C.-Y. Kwon and B. Lee, 'Prevalence of Behavioral and Psychological Symptoms of Dementia in Community-Dwelling Dementia Patients: A Systematic Review', *Front. Psychiatry*, vol. 12, p. 741059, Oct. 2021, doi: 10.3389/fpsyt.2021.741059.

[327] B. Reisberg, J. Borenstein, S. P. Salob, S. H. Ferris, E. Franssen, and A. Georgotas, 'Behavioral symptoms in Alzheimer's disease: phenomenology and treatment', *J. Clin. Psychiatry*, vol. 48 Suppl, pp. 9–15, May 1987.

[328] J. E. Overall and D. R. Gorham, 'The Brief Psychiatric Rating Scale', *Psychol. Rep.*, vol. 10, no. 3, pp. 799–812, Jun. 1962, doi: 10.2466/pr0.1962.10.3.799.

[329] 'Overview | Dementia: assessment, management and support for people living with dementia and their carers | Guidance | NICE'. https://www.nice.org.uk/guidance/ng97 (accessed Jan. 13, 2022).

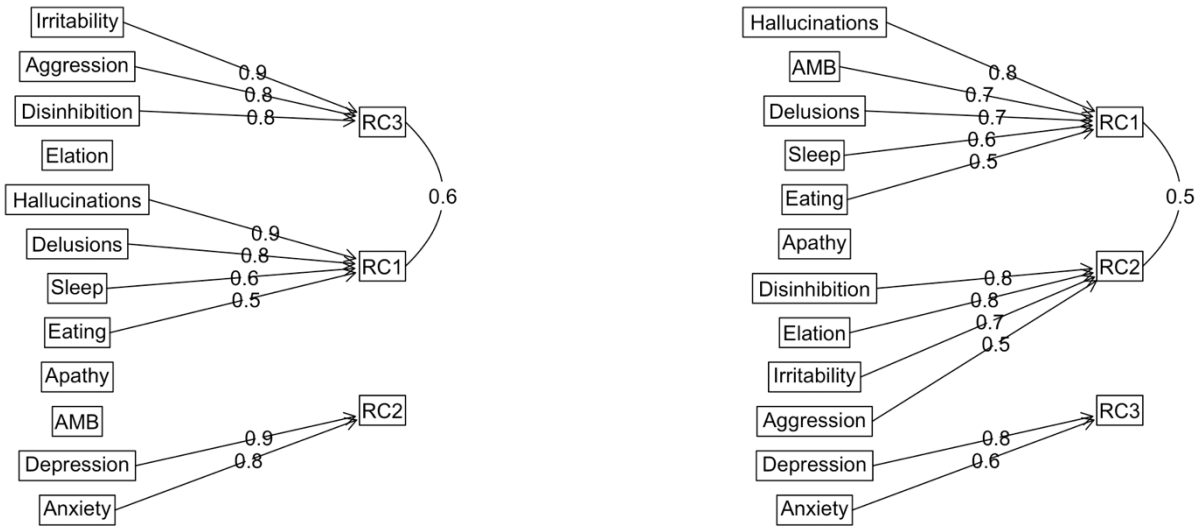
[330] J. Vilalta-Franch *et al.*, 'Syndromic association of behavioral and psychological symptoms of dementia in Alzheimer disease and patient classification', *Am. J. Geriatr. Psychiatry Off. J. Am. Assoc. Geriatr. Psychiatry*, vol. 18, no. 5, pp. 421–432, May 2010, doi: 10.1097/JGP.0b013e3181c6532f.

[331] P. Hollingworth *et al.*, 'Genome-wide association study of Alzheimer's disease with psychotic symptoms', *Mol. Psychiatry*, vol. 17, no. 12, pp. 1316–1327, Dec. 2012, doi: 10.1038/mp.2011.125.

[332] L. A. Hindorff *et al.*, 'Prioritizing diversity in human genomics research', *Nat. Rev. Genet.*, vol. 19, no. 3, pp. 175–185, Mar. 2018, doi: 10.1038/nrg.2017.89.

[333] J. MacArthur *et al.*, 'The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)', *Nucleic Acids Res.*, vol. 45, no. D1, pp. D896–D901, Jan. 2017, doi: 10.1093/nar/gkw1133.

Supplementary Material



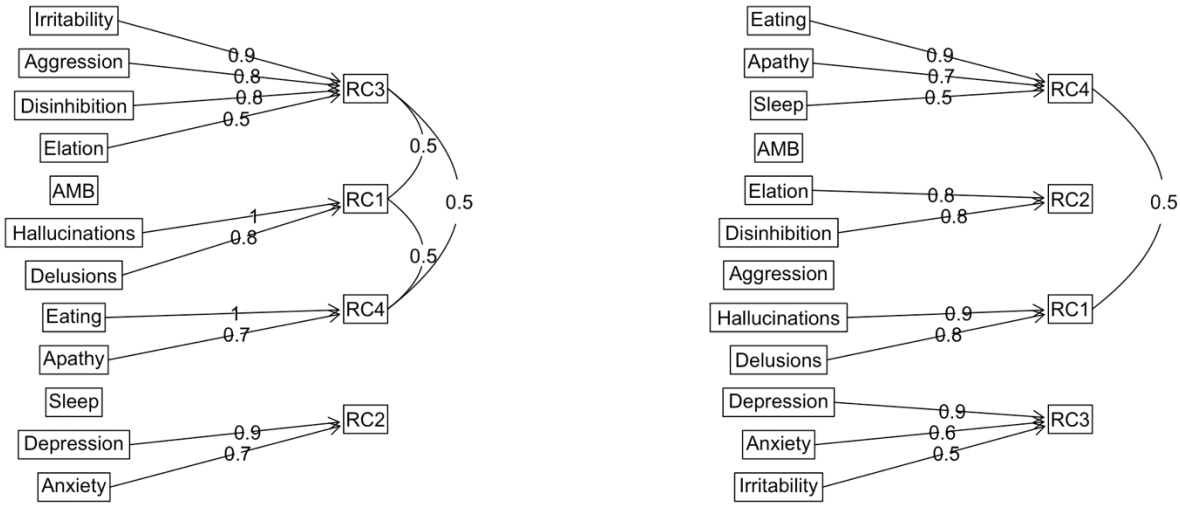
Supplementary Figure 1. PCA of NPI domain scores for individuals with LOAD (left) and EOAD (right) when using a three-component structure. (RC: rotated principal component, AMB = aberrant motor behaviour)

LOAD	RC1	RC2	RC3
Delusions	0.75	-0.02	0.07
Hallucinations	0.86	0.07	-0.20
Anxiety	0.14	0.75	-0.07
Depression	0.04	0.87	-0.18
Aggression	0.20	-0.16	0.78
Irritability	-0.02	-0.14	0.89
Disinhibition	-0.07	0.09	0.76
Elation	-0.31	0.45	0.48
AMB	0.41	0.05	0.32
Apathy	0.43	0.03	0.26
Sleep	0.02	0.02	0.11
Eating	0.51	0.10	0.01
Eigenvalue	2.55	1.56	2.46
Proportional variance	0.21	0.13	0.21

Supplementary Table 1. Component loadings for LOAD individuals when using a three-component structure. (RC: rotated principal component, AMB = aberrant motor behaviour)

EOAD	RC1	RC2	RC3
Delusions	0.68	-0.03	0.17
Hallucinations	0.84	-0.28	0.18
Anxiety	0.25	0.05	0.57
Depression	0.01	0.04	0.78
Aggression	0.18	0.52	0.28
Irritability	-0.18	0.65	0.44
Disinhibition	0.10	0.80	-0.11
Elation	-0.18	0.77	0.08
AMB	0.69	0.24	-0.20
Apathy	0.47	0.11	0.10
Sleep	0.61	-0.09	0.12
Eating	0.52	0.32	-0.31
Eigenvalue	2.75	2.32	1.61
Proportional variance	0.23	0.19	0.13

Supplementary Table 2. Component loadings for EOAD individuals when using a three-component structure. (RC: rotated principal component, AMB = aberrant motor behaviour)



Supplementary Figure 2. PCA of NPI domain scores for individuals with LOAD (left) and EOAD (right) when using a four-component structure. (RC: rotated principal component, AMB = aberrant motor behaviour)

LOAD	RC1	RC2	RC3	RC4
Delusions	0.84	-0.01	0.13	-0.11
Hallucinations	0.98	0.08	-0.13	-0.13
Anxiety	0.11	0.74	-0.07	0.09
Depression	0.04	0.85	-0.18	0.06
Aggression	0.14	-0.15	0.81	0.03
Irritability	-0.01	-0.13	0.92	-0.08
Disinhibition	-0.15	0.09	0.77	0.06
Elation	-0.15	0.45	0.51	-0.23
AMB	0.28	0.05	0.33	0.16
Apathy	-0.05	0.01	0.19	0.66
Sleep	0.33	0.01	0.10	0.37
Eating	0.17	0.06	-0.11	0.97
Eigenvalue	1.88	1.53	2.50	1.55
Proportional variance	0.16	0.13	0.21	0.13

Supplementary Table 3. Component loadings for LOAD individuals when using a four-component structure. (RC: rotated principal component, AMB = aberrant motor behaviour)

EOAD	RC1	RC2	RC3	RC4
Delusions	0.81	0.14	0.03	-0.10
Hallucinations	0.94	-0.09	0.03	-0.06
Anxiety	0.24	0.01	0.60	-0.02
Depression	0.01	-0.07	0.88	-0.09
Aggression	0.12	0.48	0.29	0.11
Irritability	-0.25	0.52	0.53	0.07
Disinhibition	0.01	0.77	-0.14	0.21
Elation	0.04	0.84	0.00	-0.20
AMB	0.36	0.21	-0.23	0.50
Apathy	-0.12	-0.08	0.25	0.72
Sleep	0.38	-0.23	0.20	0.53
Eating	-0.15	0.08	-0.19	0.88
Eigenvalue	1.86	2.04	1.75	2.91
Proportional variance	0.15	0.17	0.15	0.16

Supplementary Table 4. Component loadings for EOAD individuals when using a four-component structure. (RC: rotated principal component, AMB = aberrant motor behaviour)

	Psychosis		Agitation		Elation		Affect		Behaviour	
PRS	β	p	β	p	β	p	β	p	β	p
AD P<0.5	-0.01	0.706	-0.02	0.317	0.06	0.005	0.07	0.001	0.01	0.525
AD P<0.1	-0.01	0.650	-0.02	0.300	0.04	0.005	0.05	0.022	0.01	0.781
AD P<0.05	-0.01	0.592	-0.01	0.531	0.01	0.790	0.03	0.177	0.01	0.648
AD P<0.01	-0.04	0.022	-0.08	$7.35x10^{-6}$	0.07	$5.70x10^{-5}$	0.18	$1.20x10^{-20}$	0.04	0.010
SZ P<0.5	0.10	$1.75x10^{-8}$	0.07	$9.63x10^{-5}$	0.09	$1.01x10^{-5}$	0.10	$1.18x10^{-6}$	0.10	$1.73x10^{-7}$
SZ P<0.1	0.09	$5.71x10^{-7}$	0.09	$1.83x10^{-7}$	0.01	0.670	-0.01	0.688	0.08	$5.07x10^{-6}$
SZ P<0.05	0.08	$8.66x10^{-7}$	0.09	$1.88x10^{-6}$	-0.01	0.624	-0.06	0.006	0.07	$1.96x10^{-6}$
SZ P<0.01	0.06	$1.67x10^{-4}$	0.10	$4.48x10^{-9}$	-0.02	0.314	-0.15	$1.23x10^{-5}$	0.07	$1.06x10^{-4}$
BD P<0.5	-0.01	0.517	0.00	0.976	0.03	0.089	0.12	$5.58x10^{-5}$	-0.01	0.433
BD P<0.1	-0.03	0.060	-0.02	0.210	0.05	0.006	0.16	$1.58x10^{-16}$	-0.01	0.572
BD P<0.05	-0.03	0.097	-0.02	0.218	0.05	0.008	0.15	$6.48x10^{-15}$	-0.01	0.435
BD P<0.01	-0.00	0.940	-0.01	0.513	0.05	0.008	0.16	$1.07x10^{-16}$	-0.01	0.632
MDD P<0.5	-0.04	0.019	-0.05	0.002	0.05	0.010	0.18	$7.29x10^{-20}$	-0.04	0.026
MDD P<0.1	-0.02	0.151	-0.03	0.142	0.01	0.611	0.13	$9.53x10^{-11}$	-0.01	0.445
MDD P<0.05	-0.03	0.061	-0.01	0.562	-0.01	0.492	0.09	$1.74x10^{-5}$	-0.01	0.443
MDD	-0.01	0.423	-0.06	0.002	-0.01	0.050	0.00	0.953	-0.01	0.448

P<0.01										
IQ P<0.5	-0.01	0.478	-0.05	0.070	-0.04	0.849	-0.01	0.774	-0.04	0.028
IQ P<0.1	-0.01	0.756	-0.02	0.230	-0.01	0.470	-0.04	0.032	-0.03	0.042
IQ P<0.05	-0.01	0.328	-0.05	0.005	0.00	0.834	-0.01	0.758	-0.05	0.004
IQ P<0.01	-0.04	0.035	-0.07	$7.17x10^{-4}$	0.00	0.922	-0.05	0.017	-0.06	$8.12x10^{-4}$

Supplementary Table 5. Association of the PRS derived with the five sub-phenotype scores, while controlling for age, sex and the first ten principal components identified in the dataset. (AD = Alzheimer's disease, SZ = schizophrenia, BD bipolar disorder, MDD = major depressive disorder, β = beta coefficient, p = p-value)