Thesis

The cognitive reserve's effects modified by genetics. A research on Alzheimer's disease and psychological problems.

Name: Julian Brouwer

Student ID: 360551

Supervisor: Hans van Kippersluis

Place: Rotterdam

Date: 06/01/2020

Word count: 13.320

Abstract

Intro

Do education and occupation relate to the prevalence of Alzheimer's Disease (AD) / psychological problems (PP) and is there an interaction effect between the AD/Neuroticism polygenic score and education/occupation?

Societal- and healthcare costs for AD and PP currently account for profound budget spending. Total projected cumulative healthcare costs and unpaid caregiving for AD, in the USA between 2017-2030, account for $3.2 trillion respectively $4.5 trillion.
Total mental healthcare costs in the European Union, in 2010, accounted for $1,06 trillion. These costs are expected to double in 2030. Nearly 2 billion people suffer from PP every year.

Well known is that education and a cognitive (un)stimulating job are associated with one's risk profile for AD/PP. this mechanism is likely explained by the cognitive reserve hypothesis, a hypothetical reserve in the brain which is created by stimulating cognitive activities like education and one's occupation.

Data & Methods

In this study we have access to the so-called polygenic risk score (PGS). This is a weighted score of genes relating to AD/PP. For AD we use the PGS for AD, for PP we use the PGS of neuroticism (N). Neuroticism is the tendency in individuals to experience negative emotions/being emotionally instable which makes them easily nervous or upset. We examine the relationship of the cognitive reserve controlling for genetics. More interestingly we examine if there exists an interaction effect between the cognitive reserve and genetics. Does the cognitive reserve have a different effect for people with different PGSs? We will support the results with a transformed Grossman Model, where, controlling for genetics, you can invest in the cognitive reserve and individuals experience (natural) depreciation of the cognitive reserve stock. When the cognitive reserve falls below a certain threshold, the individual is classified as having AD or PP.

One of our main contributions is that we created a unique occupational index recoded from the 2000 United States Census, the International Standard Classification of Occupations (ISCO-08) and the Occupational Information Network (O*NET). Our study population consists of American white people who are 50 years and over.

We obtain interaction effects through the linear probability model (LPM) and control these models with logistic regressions and marginal effects.

Results

For AD we find no evidence of existing interaction effects. For PP we find moderate evidence suggesting there could exist an interaction between white collar and the PGS for N. We do find for AD that education relates to the development of the disease. More years of education lowers the probability of developing AD, occupational measures are found to be insignificant related. For PP we only found a significant relation for occupation, where being white collar lowers the probability of developing PP, white collar is partly explained by education due to the correlation between the two variables. Education, alone, appeared to have no effect.

Conclusion

We found that cognitive reserve measures relate to AD/PP even when controlling for genetics. For PP we even found suggestive evidence of a possible interaction between being white collar and the PGS of N. This interaction effect should be examined more in depth.

Table of Contents

Introduction

Alzheimer's Disease (AD) and psychological problems (PP) are known to be affected by both genetic factors, environmental factors, as well as individual behavior. Well established is the relationship between more education and, to a lesser extent, more cognitively stimulating occupations yielding better health outcomes for potentially developing AD and PP. The effect of education/occupation is usually explained by the cognitive reserve hypothesis, this states there is heterogeneity in the ability to cope with brain pathology. The cognitive reserve can be expanded by more education or a more cognitively stimulating occupation. This expansion of the cognitive reserve is sometimes found to alter the brain pathology of biomarkers in AD (Soldan, 2013) (Lo, 2013). (Vemuri, 2014) however, found no evidence for this. Likewise, heritability studies have well-established that genetic factors play a role in both AD and PP.

In this paper we will study the interaction between environmental and genetic factors on the diagnosis of having AD or PP. The environmental factors are expressed by the cognitive reserve (i.e. education and occupation). We will investigate this relationship within the framework of a transformed Grossman Model where you can invest in the cognitive reserve thereby counteracting the development of the disease / clinical symptoms.

The interaction effect between the factors determining the cognitive reserve and individual genetics has not been extensively examined. To our knowledge we are even the first to examine this interaction for PP. Could it be that, for a certain genetic profile, education/occupation has a far larger/smaller effect on developing AD/PP?

For AD we found two other studies to examine this interaction, (Garibotto, 2011) found no significant interaction between the APOE-e4 genotype (biomarker with the highest association for AD) and education/occupation. (Shadlen, 2005) found the opposite, the effect of education was different for carriers of the APOE-e4 variant, APOE-e4 has the highest association with AD. We use 17 other genetic markers besides APOE-e4, making the data more comprehensive for AD.

In this research we have access to the so-called polygenic score (PGS) of AD and neuroticism (N). The PGS is a weighted average of all genetic factors related to, in this study, the incidence of AD/N in humans. For AD it includes 17 other unique genetic markers related to AD besides the 'main' APOE-e4 variant. N includes 11 unique genetic markers. We use the PGS of N to estimate the dependent variable PP since we do not have the PGS of PP. This is a limitation, but of low impact since N has a strong relationship with developing psychological problems and is seen as the pre-stadium of developing a broad spectrum of mental disorders. N is best described as the tendency in individuals to experience negative emotions/being emotionally instable which makes them easily nervous or upset.

One of our main contributions is that we created a unique occupational index recoded from the 2000 United States Census, the International Standard Classification of Occupations (ISCO-08) and the Occupational Information Network (O*NET). With this unique dataset, we can examine the interaction between genetics and educational/occupational investments in the development of AD and PP more in depth. If there is an interaction found it would be interesting to examine whether people with different genetics would benefit from different treatments. In this way healthcare would become more tailor-made and thereby more efficient in terms of economic and health outcomes. For AD and PP this would eventually mean that there could exist a possibility to design custom treatments based on different genetic profiles.

Relevance

AD

5.5 million people in the United States are currently diagnosed with AD (1,68% of total population), a disease with high prevalence among people of 65 years and older. In the Netherlands, in 2015, 260.000 people were diagnosed with AD (1,54% of total population). These numbers are expected to increase profoundly due to the increase in life expectancy and the ageing of the baby boom generation, accounting for 13.8 million (USA) / 620.000 (The Netherlands) people living with AD in 2050 (He, Goodkind, & Kowal, 2015) (Alzheimer Nederland, 2019). In the Netherlands the current 65+ population consists of 3.3 million people and is expected to grow to 4.7 million (Centraal Bureau Statistiek, 2019).
The population of 65 years and older in the USA currently consists of 53 million people and is expected to grow to 88 million in 2050 (He, Goodkind, & Kowal, 2015). In the USA, in 2010, there were 455.000 new diagnoses of AD. In 2030 this will be 615.000 (35% higher) and is projected to be 960.000 in 2050 (110% increase from 2010) (Hebert, Beckett, Scherr, & Evans, 2001). Total projected cumulative healthcare costs / unpaid caregiving, in the USA between 2017-2030, account for $3.2 trillion respectively $4.5 trillion (Partnership to Fight Chronic Disease).

AD is a chronic progressive brain disease. It is a form of dementia, 70% of all dementia is AD. AD is mild progressive and slowly degrades one's memory and thinking, cognitive abilities diminish. In the latest stage of the disease it is nearly impossible for patients to carry out the simplest daily living activities and impair an individual's ability to carry out basic bodily functions such as walking and swallowing. AD is most seen among people at 60 years and older, although prevalence among people from the age of 40 onwards is seen.
Typically seen in the brain of AD patients are abnormal clumps (amyloid plaques) and tangled bundles of fibers (neurofibrillary tangles). A third symptom is the loss of contacts between neurons in the brain (J.Selkoe, 1991). At present time, these pathological symptoms can only be examined after death.

With AD being a more and more prevalent disease accompanied by the large increase in societal- and healthcare costs, it is of great importance to explore the risk factors of the incidence of AD and understanding through which mechanisms these effects work. In this way health policy can be made to benefit patients and society. For instance, it could be that, during the pre-clinical phase, treatment can be given to postpone the clinical symptoms of AD and thereby saving societal- and healthcare costs and preserve quality of life of the patient. By looking at the interaction between each PGS per individual and their education/occupation we can distill the unique effect of education and occupation per genetic profile. In this way health providers for instance could potentially classify people into different categories in which each category yields a different treatment.

Neuroticism & Psychological Problems

N is one of the Big Five Personality Traits, also known as the five-factor model (FFM) and the OCEAN model (Coetzer & Rothmann, 2003). N is not a medical condition but defines as a personality trait. It is best described as the tendency in individuals to experience negative emotions/being emotionally instable which makes them easily nervous or upset. It is not an absolute condition but rather a scale outcome. (Eysenck & Eysenck, 1975) Described the pathology of N as follows: characteristic individual differences in behavior are induced by determinants within the basic neural structures of the brain. It is based on activation thresholds in the visceral brain, this part of the brain is responsible

for the fight-or-flight response when facing danger. Those with high N scores have low activation thresholds in the visceral brain in the face of stressors, which makes them easily nervous and upset.

N is mainly perceived as a pre-stadium to develop other mental illnesses. It also is associated with worse physical health and higher comorbidity. N, opposed to other mental conditions, has been found to correlated with the greatest range of other mental health illnesses (Malouff, Thorsteinsson, & Schutte, 2005) thereby making it the one variable with the strongest explaining power for developing psychological problems. Although not qualifying as a medical condition accompanied by an official diagnose, psychologists and psychiatrists still value the N trait as an important indicator for mental health. Dr. Benjamin B. Lahey, of the University of Chicago's Departments of Health Studies and Psychiatry and Behavioral Neuroscience stated the following:

''*Although not widely appreciated, there is growing evidence that N is a psychological trait of profound public health significance. N is a robust correlate and predictor of many different mental and physical disorders, comorbidity among them, and the frequency of mental and general health service use*'' (Lahey, 2009).

(Kendler KS, 2006) found in a prospective study among 20.692 Swedish same sex twins a relation between N and major depression. Each 1 standard deviation difference in higher N scores was associated with a 31% greater risk for a first-ever diagnose of major depression in a 25-year period. A same positive association was found for schizophrenia. A birth cohort of 5.362 individuals showed that the probability of later in life meeting the criteria for schizophrenia was 93% higher for 16-year-olds with high N scores (Van Os J, 2001). Other scientific papers support the link between N and mental illnesses. These include for instance borderline, anxiety-, somatoform-, mood- and eating disorders (Malouff, Thorsteinsson, & Schutte, 2005). N even is associated with physical illnesses. This includes cardiovascular, asthma, atopic eczema (Smith TW, 2006).

Total costs for mental healthcare in the European Union, in 2010, accounted for $1,06 trillion. These costs are expected to double in 2030. A new Lancet commission report even said that mental illnesses are increasing in every country and will yield an economic cost of $16 trillion by 2030 globally. These costs mainly consist off lost productivity and early onset of the mental illness. Nearly 2 billion people suffer from mental illness every year (Patel, 2016).

For psychological problems it is of great interest as well to explore the relationship between the incidence of PP and the factors contributing to this. As with AD, depending on the PGS per individual and the interaction with education/occupation. Health providers could classify people into different categories yielding different, custom made, treatments.
In our dataset we use the PGS of neuroticism to predict if a person has ever been told by a doctor to have psychological problems (PP) like emotional-, nervous- or psychiatric problems. We unfortunately, as stated before, do not have data whether a person has been 'diagnosed' with N.

Objective and Research Question

We know from research that education and a cognitive (un)stimulating job are associated with one's risk profile for AD/PP. We don't know whether this represents a causal relationship and how this mechanism works. Likely is that this mechanism can be explained by the cognitive reserve hypothesis, a hypothetical reserve in the brain which is created by stimulating cognitive activities like education and one's occupation. This concept will be explained more in depth later on. Moreover, we don't know whether investments in the cognitive reserve can partially counteract genetic risk for AD/PP, and whether the effect of the cognitive reserve on AD/PP still exists controlling for genetic risk factors. In order to better understand how these mechanisms works, we explore the relationship between one's genetics and his or her investments in their cognitive reserve in the development of AD/PP. We explain this relation by the Grossman model, where one has an initial stock of health, does investments in their health and has depreciation of the health stock as a function of time. Could it be that a different PGS for AD/PP, also changes the effect in which education/occupation works? And if so, why would this be? And could this result be used for policy making and/or a better understanding of AD's/PP's risk factors?

Research question:

Do education and occupation relate to the prevalence of Alzheimer's Disease (AD) / psychological problems and is there an interaction effect between the AD/Neuroticism polygenic score and education/occupation?

Short Overview Thesis

First, we explain the cognitive reserve concept, the transformed Grossman Model and the Polygenic Risk Score. Then we will discuss our data variables and provide descriptive statistics. After that we will provide our results with the statistical models and output. In the following discussion we will critically evaluate the findings, our study design and match the results to theory. Ending with future research recommendations.

Theoretical Framework

Cognitive Reserve Theory

The cognitive reserve hypothesis posits that there is heterogeneity in the ability to cope with brain pathology, such as the plagues and tangles related to AD. The greater the cognitive reserve, the less risk for cognitive- or clinical impairment, since those with a higher reserve are able to work around their brain impairments (Stern Y. , 1994). A greater cognitive reserve can arise when the brain can use its networks more flexible and/or efficient, or with differences in brain anatomy such as a larger brain size or greater synapse counts. Education and occupation could be factors that positively affect the cognitive reserve.

The cognitive reserve is supported by several papers. For instance it was found that those with more educational attainment possess- greater brain weight, larger neurons and possess more finely branched neurons (Katzman, 1988). Also having greater brain size and having more education diminished the magnitude of cognitive decline following traumatic brain injury (Kesler, 2003). For the general population a larger head circumference is associated with both less memory related problems in older ages and with higher adult IQ (Gale, 2003).

Alzheimer's Disease and the Cognitive Reserve

The common finding is that limited educational attainment and a cognitive unstimulating occupation are related to greater risk of developing AD. This relationship is explained in two possible ways:

1. Bias in diagnostics tests ➔ Patients with higher education, and thus in general a more cognitively stimulating occupation, might be more difficult to diagnose since they can provide 'smarter' answers. A cognitive stimulating job is usually more accessible with a higher educational background.

However, one of the most important tests to test for AD, the Mini Mental state Exam (MMSE), is found by (Jorm, 1988) to be unbiased, this result is supported by (Jones, 2001).

2. Cognitive Reserve ➔ Patients with higher education build a 'cognitive reserve' which enables them to work around their brain impairments (Stern Y. , 1994). So, education has a direct effect on developing AD. The same is found for a cognitively stimulating job, although this effect is usually lower as found by (Andel, 2006).

More interestingly is adding a biological dimension to this cognitive reserve hypothesis. (Shadlen, 2005) found an interaction between years of education and APOE-e4 for AD. When a respondent had the APOE-e4 variant, it showed that education counteracted the associated cognitive decline, this effect of education was not found for the APOE-e2 and APOE-e3 variants. This result showed that education is likely to postpone the clinical cognitive impairment of AD and that education has a different effect conditional on the APOE status. Similar results for biomarkers were obtained by (Masliah, Terry, Alford, & DeTeresa, 1991), they found that there was a close relationship between neocortical synaptic density (synapses control the communication in the brain) and cognitive decline. (Arenaza-Urquijo, 2013) studied the effect of amyloid-β (Aβ42), the main composition of the AD related plagues in the brain, on the cognitive reserve. They did this by comparing the cortical thickness on different parts of the brain among subjects with different Aβ42 levels and cognitive reserve proxies. Despite the low number of observations, they found significant effects on different

parts of the brain with different pathology. Those with a higher cognitive reserve with the same level of amyloid-β had a different, better in terms of developing AD, brain pathology than those with a low cognitive reserve, meaning they could tolerate a more advanced neurodegenerative process. In other words, at any level of clinically severity, AD pathology (tangles and plaques), will be more advanced in individuals with more education opposed to less education (Roe, 2007).

Some studies also examined the deterioration after onset of the clinical symptoms of AD for people with a relative higher cognitive reserve. Although the cognitive reserve postpones the onset of clinical symptoms, it also means that after the onset of these symptoms, people with a higher cognitive reserve deteriorate faster. Meaning their cognitive impairments worsen faster (Andel, 2006).

Psychological problems and the Cognitive Reserve

For PP the same negative association with education and occupation is established. This association can be explained in three ways:

1. Those with poorer psychological health during their education period are more susceptible to drop out or achieve lower grades. Thereby limiting their access to job opportunities as well;

2. Certain characteristics, e.g. positive experiences of parenting and/or cognitive ability – may expose those to better health outcomes, to educational investment and success (Kjelsberg, 1999);

3. Through the cognitive reserve they can work around the 'brain impairments. For example, it improves a person's ability to act more rational, thereby suppressing impulsive/emotional behavior and being a calmer and emotionally stable person. In that way being more likely to thrive in education and/or their career path.

In this paper, we will focus on the cognitive reserve.
One could see PP as an impairment of the brain, which individuals can counteract by investing in the cognitive reserve. The pathway of the effect is described in point three above. While the first to link the cognitive reserve to mental disorders were (Barnett, Salmond, Jones, & Sahakian, 2006), we are the first to propose the hypothesis in which this specific pathway works.

The relation between the cognitive reserve and PP can be explained in three ways:

1. by affecting the risk for developing the mental disorder;
2. in the expression of symptoms within the mental disorders;
3. in patients' functional outcome, social and vocational outcomes, (Barnett, Salmond, Jones, & Sahakian, 2006).

Point two and three could be the result of how our proposed theory works. Being calmer and more rational prevents expression of symptoms and improves one social and vocational outcomes. E.g. those who are more rational could find it easier to incorporate the disorder in their daily living and thereby facing less social and vocational impairment.

There are multiple studies showing relationships between the cognitive reserve and mental disorders. Most of these studies use pre-morbid IQ as a measure for the cognitive reserve. Schizophrenia is known to have the strongest relation with the cognitive reserve as expressed by pre-morbid IQ (Koenen, 2009). A meta-analysis found a robust linear relationship, meaning that this effect is active at all points of the IQ spectrum (Golam, Khandakera, Barnettablan, White, & Jones,

2011). The risk on schizophrenia increased by 3,8% by every one-point decrease on the IQ spectrum. Similar significant relationships are found for depression, several anxiety disorders and bi-polar disorder (Koenen, 2009) (Grande, 2017).

Transformed Grossman Model

We explore the relationship between the cognitive reserve and genetic markers by the Grossman model. The Grossman model views health as a durable capital stock that yields an output of healthy time. A person inherits an initial stock of health that depreciates over time and can be increased with investments. Death occurs when stock health falls below a certain threshold (Grossman, 1973). We will adopt this model to explain having or not having AD/PP, as a function of (i) investments in human capital, like education and a cognitive stimulating occupation; (ii) depreciation of human capital that is partly genetically determined; and (iii) by the interaction between education/occupation and the genetic risk for AD/N. In this way, we are able to estimate the 'production function' for AD/PP, and we're able to capture the effect of education/occupation conditional on one's genetic propensity.

In line with the cognitive reserve hypothesis our model views the health outcome as dichotomous: having or not having AD/PP. We posit that the development of AD/PP can theoretically be linked to the Grossman model of health capital. AD/PP is then the binary outcome of one's cognitive reserve hitting a critical threshold. In line with the Grossman model, one's 'cognitive reserve' can be increased by investments in cognitive abilities (education, mentally stimulating job), and depreciates over time as one ages. The effect of cognitive investments will be different for different genetics, and the depreciation factor too may be influenced by one's genetic risk. The model is represented by the following equation (1).

$$\boldsymbol{H_t = I_t(G) - d_t(G)H_t} \quad \textbf{(1)}$$

where H_t is the cognitive reserve at time t, I_t measures investment in cognitive abilities at time *t*, the effectiveness of which is possibly influenced by one's genetic risk G, and d_t measures the depreciation rate over time, again potentially influenced by one's genetic risk G. AD/PP is then defined as $H_t < c$, where c is the critical threshold, which is unique for every individual, of the cognitive reserve below which an individual is classified as having AD/PP.

Empirically, we will model AD/PP risk as a function of educational attainment/occupation, genetic risk, and the interaction term between the PGS for AD/N and education/occupation, in order to estimate the theoretical production function specified in equation (1). In the model we capture the effect for education/occupation conditional on different genetics.

Diagnosis

First off, we ask ourselves when a Physician indeed classifies a person as having AD or PP.

Alzheimer's Disease

There is not a single test to diagnose AD. Physicians usually can clinically determine whether a patient has dementia, but the cause of this dementia can still be unknown. Physicians will conduct several neurological tests to rule other causes out, such as a stroke, Parkinson disease, brain tumors and other illnesses that impair cognitive/neurological functions. In addition, one of the most used assessments is given, the Mini-Mental State Exam (MMSE). This test checks everyday mental skills. The score shows the severity of dementia on a scale from 1 to 30, where less than 12 points is stated

as severe dementia. An average person with AD declines 2 to 4 points a year. Also, medical family history and behavioral changes are examined, and even brain imaging is used. (Alzheimer's Association, 2016)

Further research is currently emerging on the pre-clinical phase, we know that 15 years before the onset of the disease specific symptoms there are biological changes in de brain, for example in the cerebrospinal fluid, the protein 'Tau' changes. Such biological factors are called biomarkers. Research is still experimental and a correct study design to incorporate biomarkers in the diagnosis process are lacking today (Dubois, Hampelab, & Feldman, 2016). Unfortunately, we do not know how the diagnosis is made at individuals in our data.

Neuroticism & PP

Neuroticism is not an official clinical diagnose, it is described as a personality trait. N scores are measured with personality tests. Having psychological problems is the dependent variable in the model. It asks respondents the question: ''have you ever been told by a doctor to have psychological problems like emotional-, nervous- or psychiatric problems''? Since psychological problems is a broad definition of several mental illnesses, there is no general way of diagnosing patients. But most mental illnesses are diagnosed by psychiatrists supported by different questionnaires.

Genetic data

Our genetic data is from the Health and Retirement Study (HRS). The HRS combines two datasets to construct the PGS for N, the UK Biobank (UKB) and the Genetics of Personality Consortium (GPC). To construct the AD PGS, the HRS combined data from four international consortia: Alzheimer 's Disease Genetic Consortium (ADGC), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the European Alzheimer's Disease Initiative (EADI), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium. The diagnosis of AD is set by a medical professional.
For N, we only know what test is used for the UKB results (N=107.245) from the total observations (N= 170.911). The remaining data from the GPC (N=63.666) harmonized multiple unknown N tests. The UKB used the 12-item version of the Eysenck Personality Inventory N scale. This is a self-report instrument consisting of 57 yes or no questions. The questionnaire yields total scores for extraversion and neuroticism as well as a validity score. Eysenck defined 9 traits that are related to N: anxiety, depressed, guilt feelings, low self-esteem, tense, moody, hypochondriac, lack of autonomy and obsessive (Eysenck & Eysenck, 1975).

Research methods

Polygenic Risk Scores

The PGS is constructed from the human DNA. The human DNA consist of 46 chromosomes (23 pairs). Chromosomes all contain DNA made up of four nucleotides, abbreviated C (Cytosine), G (Guanine), A (Adenine), or T (Thymine), which line up in a particular sequence, in pairs, and make a long string. There are about 3.2 billion nucleotide pairs on all the human chromosomes: this is called the human genome.

99.9% of the human genome is identical, the parts of the human genome where there exists variation are called polymorphisms. The most common polymorphism is a single-nucleotide polymorphism (SNP). This is a variation on a certain nucleotide pair for only one of the nucleotides. For example, for a specific position on a chromosome where most pairs are A**C**, there is a group of people who have A**G**. This is a SNP, where C and G are called alleles for this position.

Each of these SNP's can have an effect of developing a certain disease/trait. Usually the effect of each SNP is small but combined they (can) have substantial impact. This combined result is called the Polygenic Score (PGS).

The genetic variants underlying certain diseases are identified from Genome-wide Association Studies (GWAS). These studies compare 2 groups of people. One group has the certain disease/trait whereas the other group does not. The genetic variants showing a significant and reproducible difference in frequency between groups are associated with the incidence of the disease/trait.

The PGS can be constructed in several ways, yielding different outcomes. For example, one can simply add up all the scores or weigh each SNP in how strongly they relate to the certain disease.

In our dataset: *''the PGS aggregates thousands to millions of individual loci* (fixed places on the chromosomes) *across the human genome and weights them by effect sizes derived from a GWAS as an estimate of the strength of their association to produce a single quantitative measure of genetic risk and to increase power in genetic analyses''* (Ware, Schmitz, Gard, & Faul, 2018).

The PGS is a standardized score, meaning one point on the PGS distribution is equal to exactly one standard deviation and the mean is 0.

The PGSs are obtained from different consortiums and combined by the Health Retirement Study (HRS). Our data has 12.090 observations from European ancestry (EA) and 3100 observations from African ancestry (AE).

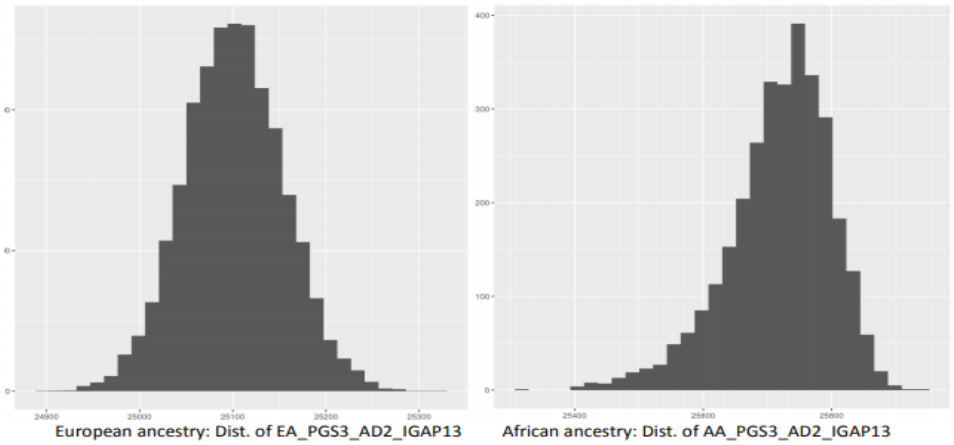
Alzheimer's Disease

There are four variants of the PGS for AD:

1. PGS for EA without the APOE genetic variants;
2. PGS for EA with the APOE genetic variants;
3. PGS for AE without the APOE genetic variants;
4. PGS for AE with the APOE genetic variants.

We will use the PGS for EA with the APOE genetic variants. EA has 4 times more (12.090 vs 3.100) observations and therefore is more normally distributed as we can see in figure 1.

Figure 1 – Distribution PGS of AD



There are three types of the apolipoprotein APOE gene, called alleles: APOε2, ε3 and ε4. The two single-nucleotide polymorphisms SNP's named: rs7412 and rs429358 make up which of the three alleles one has. The ε4 allele of APOE is the major genetic risk factor for AD (AD) (Kim, Basak, & Holtzman, 2009). Scientists yet have to confirm how this genetic variant works. The European ancestry PGSs contains 1,145,019 (1,145,021, adding the two APOE status variants) Single nucleotide polymorphisms (SNPs). Besides the APOE variants, the PGS of AD in this paper includes 19 other significant SNPs, this means that there are 19 genes in which there is significant genetic variation between people which affects the incidence of AD.

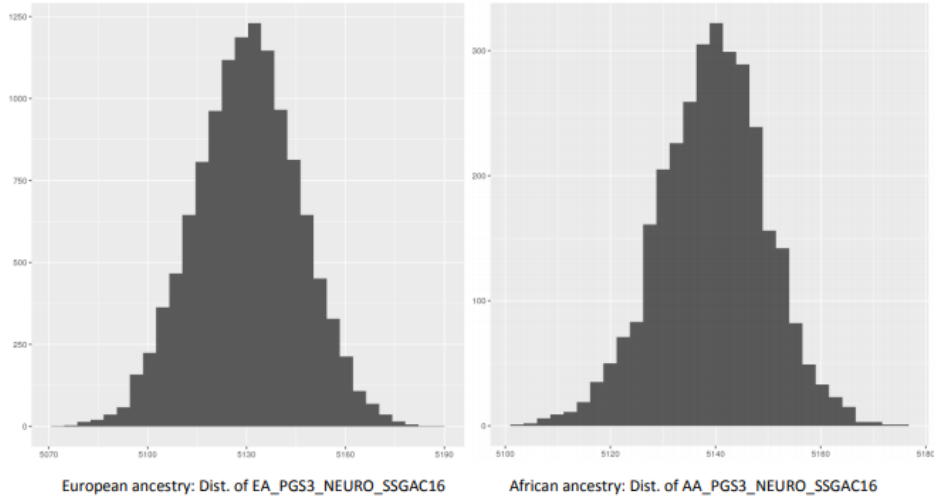
Neuroticism

For N there are two variants of the PGS:

- PGS for EA
- PGS for AA

We will use the PGS for EA, since it has 4 times more observations (12.090 vs 3.100) and therefore is more normally distributed as seen from the graphs in figure 2. The European ancestry PGSs contain 1,134,281 SNPs. the PGS of N in this paper includes 11 significant SNPs, this means that there are 11 genes in which there is significant genetic variation between people which affects the incidence of N.

Figure 2 – Distribution PGS of N



Data

The data is combined data from the RAND Health and Retirement Study (RAND HRS) dataset and the original HRS. The HRS is a longitudinal panel study that surveys a representative sample of approximately 20,000 people in The United States. This has been done approximately every two years since 1992. The study is done on subjects like health care, housing, assets, pensions, employment and disability. Since 2006 the HRS began collecting DNA through saliva samples. Yielding 3,100 non-Hispanic Black observations and 12,090 non-Hispanic White observations. All Respondents are randomly selected conditional on a minimum age of 50 during the data collections.

The original HRS is merged with the dataset from RAND HRS. The main variables are constructed as follows:

Dependent Variables

Alzheimer's Disease & Psychological Problems

The outcome variable indicates whether a doctor has ever told the respondent he/she has AD/PP. The RAND data asks the following question: 'Has a doctor ever told you, you had AD'? respectively 'has a doctor ever told you to have psychological problems like emotional, nervous, or psychiatric problems' These are dummy variables which takes the value 1 if 'yes' and 0 if 'no'. When the respondent indicated 'yes' in a previous wave, the question is skipped in the current and upcoming waves.

We do not know how the diagnose is made at individuals. We do know, as seen in the chapter 'Diagnosis', the most common methods to set a diagnosis.

Independent Variables

Education

Education indicates how many years of education is completed. We also include the squared term of education, since we expect the effect of education to be marginal diminishing from a certain threshold.

Polygenic Scores

PGS of AD and PGS of N have been standardized within ethnicity, to a standard normal curve (mean=0, standard deviation = 1).

Occupation

To see to what extent, one's job was/is cognitively stimulating we use the Industry code for job with longest tenure. This variable has 25 categories corresponding to the, issued by the US Government, 2000 census industry codes.

We then transform this variable according to the International Standard Classification of Occupations (ISCO-08) classification system. The ISCO-08 divides occupations into 4 different skill levels. The skill levels range from manual work to requiring an extensive body of theoretical and factual knowledge in a specialized field (International Labour Organization, 2012).

We assign each of the 25 industry codes from the 2000 census to one of the 4 skill levels of the ISCO-08. We do this by examining each industry code and corresponding job titles via the Integrated Public Use Microdata Series (IPUMS). Finally, we examine via the Occupational Information Network (O*NET), the activities/skills demanded by the included jobs. We then can match the industry code with a certain skill level from ISCO-08. We realize that each industry code has multiple occupations included, which require different skill levels. This is counteracted by changing the classification when in doubt and perform the same regressions. There were two categories where this occurred, and only changes were made one skill level up or down. We then checked our regressions models. Overall, we can say the classification is truthful.

To make this classification even more reliable we created a dummy variable for white-/ blue collar workers. The description of skill levels 1 and 2 match the more manual skilled workers, the descriptions of skill level 3 and 4 match the more theoretically skilled workers. When we were in doubt and changed the classifications, this was always within the same group (skill level 1&2 or 3&4). Thereby ending up with an even more reliable classification system.

Occupation

Occupation is divided into 25 categories according to the 2000 census.

Table 1 – Occupation categories according to 2000 census

Occupation sectors 2000 USA Census Categories	Freq.	Percent	Cum.
01 management	746	9,39	9,39
02 business operations specialists	149	1,88	11,3
03 financial specialists	142	1,79	13,1
04 computer + math	105	1,32	14,4
05 architecture + engineering	111	1,4	15,8
06 life/physical/social science	70	0,88	16,7
07 community + social services	147	1,85	18,5
08 legal	85	1,07	19,6
09 education/training/library	465	5,85	25,4
10 arts/design/entertainment	133	1,67	27,1
11 healthcare practice / tech	345	4,34	31,5
12 healthcare support	268	3,37	34,8
13 protective service	147	1,85	36,7
14 food prepping + serving	306	3,85	40,5
15 building/grounds/clean/maintenance	504	6,34	46,9
16 personal care + service	395	4,97	51,8
17 sales occupations	758	9,54	61,4
18 office + admin support	1.112	14	75,4
19 farm/fish/forestry	55	0,69	76,1
20 construction trades	421	5,3	81,4
21 extraction workers	5	0,06	81,4
22 install/maintenance/repair workers	263	3,31	84,7
23 production	684	8,61	93,4
24 transport/material moving	477	6	99,4
25 military specific	51	0,64	100
Total	7.944	100	

Transformation according to ISCO-08

Since each category has too few observations to yield significant results in the regressions, we made a new variable according to the ISCO-08 standard. This classifies occupations into 4 different skill levels. Summarized ISCO-08 describes their standard as follow:

Skill level 1

Occupations that typically involve the performance of simple and routine physical or manual tasks. They may require hand-held tools, such as shovels. They involve tasks such as cleaning, digging,

sorting, storing or assembling goods by hand, operating non-motorized vehicles and picking fruit and vegetables. Many occupations may require physical strength and/or endurance.

Skill level 2

Occupations that typically involve the performance of tasks such as operating machinery and electronic equipment, driving vehicles, maintenance and repair of electrical equipment/machinal equipment. For almost all occupations the ability to read information such as safety instructions, make written records and perform simple calculations is essential. Many occupations require a high level of manual dexterity. These occupations may include police officers, hairdressers, mechanics and building electricians.

Skill level 3

Occupations that typically involve the performance of technical and practical tasks that require an extensive body of factual, technical and procedural knowledge in a specialized field. Preparing detailed estimates of quantities and costs of materials and labour required for specific projects, coordinating, supervising, controlling and scheduling the activities of other workers. Occupations generally require high level of literacy and good interpersonal communication skills. The ability to understand complex written material, prepare factual reports and communicate verbally well in difficult circumstances. Occupations included are legal secretaries, sales representatives, computer support technicians, diagnostic medical radiographers and shop managers.

Skill level 4

Occupations that typically involve the performance of tasks that require complex problem-solving, decision-making and creativity based on an extensive body of theoretical and factual knowledge in a specialized field. Some tasks performed are analysis and research to extend the body of human knowledge in a particular field, diagnosis and treatment of disease, transferring knowledge to others and the design of structures or machinery and of processes for construction and production. Some occupations included are sales and marketing managers, civil engineers, secondary school teachers, medical practitioners, musicians, operating theatre nurses and computer systems analysts.

We constructed the new variable according to the ISCO-08 standard as follow:

We examined the occupation 2000 census via The Integrated Public Use Microdata Series (IPUMS), there we could see in greater detail what kind of (sub)sectors belonged to one of the 25 sectors from table 1.

For instance, for category 03 from table 1, financial specialists, the following jobs are considered:

1. Accountants and Auditors
2. Appraisers and Assessors of Real Estate
3. Budget Analysts
4. Credit Analysts
5. Financial Analysts
6. Personal Financial Advisors
7. Insurance Underwriters
8. Financial Examiners
9. Loan Counsellors and Officers
10. Tax Examiners, Collectors, and Revenue Agents

11. Tax Preparers

With these more detailed job categories we then matched them with the job descriptions given in The Occupational Information Network (O*NET). O*NET is sponsored by the U.S. Department of Labor and contains hundreds of occupational definitions. As an example, in the appendix figure 1A, the output of the description of an accountant is provided.

By combining each job description that belonged to one of the categories of the 2000 census as displayed in table 1, we could then classify each of the categories according to the ISCO-08 standard. Ending up with a new variable with 4 different categories according to the four different skill levels provided by ISCO-08 as seen in table 6 further on.

Transformation to White collar dummy

In order to obtain even more explaining power for the occupation variable, according to the ISCO-08 standard, we transformed the variable to white collar, a dummy variable. We did this to ensure each category had even more observations. Skill level 1 & 2 were grouped into white collar=0 (blue collar), Skill level 3 & 4 were grouped into white collar=1.

ISCO-08 clearly describes each skill level, whereas skill level 1 & 2 describes manual skilled labour. As described at skill level 2 which we consider to still be blue collar workers:
''Occupations that typically involve the performance of tasks such as operating machinery and electronic equipment, driving vehicles, maintenance and repair of electrical equipment/machinal equipment''.

And skill level 3 & 4 described jobs which require a more theoretical background. As described at skill level 3 which we consider as White collar workers:
''Occupations that typically involve the performance of technical and practical tasks that require an extensive body of factual, technical and procedural knowledge in a specialized field''.

Since each one of the 25 categories from de 2000 census contains multiple occupations, it is inevitable that there is some overlap for the different skill levels. Some jobs are better suited for skill level 2, whereas others belong into skill level 1. There were two instances where there was considerable doubt of our division. This was for management- and farm/fish/forestry occupations. This yielded three other possibilities to construct the occupation variable. We made these three other possible transformations and then checked the regression models. The most significant result/best explaining power was still obtained with our original ISCO-08 transformation without changes.

Control Variables

Male/female

This is a categorical variable that indicates if the respondent is male or female.

Age

Age is continuous variable that indicates one's age at the time of interview.

Results

Descriptive Statistics

Total population

For the analysis, the independent variables used are the same for the dependent variables PP and AD. Only the PGS for each corresponding outcome is different, e.g. the Alzheimer PGS is used for the dependent variable AD and the Neuroticism PGS is used for the dependent variable PP.

From the total study population, 20,96% (3.921) is ever told by a doctor they have emotional, nervous, or psychiatric problems (Table 3). Whereas for a doctor ever telling a respondent they have AD is 1,92% (360) (Table 4).

Table 2 – Descriptive statistics PGS, education and age

	Alzheimer PGS	Neuroticism PGS	Education	Age
Observations	12.090	12.090	37.368	18.747
Mean	0	0	12,05	67,9
S.D.	1	1	3,46	11,26
Min	(-)3,95	(-)3,71	0	50
1st quartile	(-)0,69	(-)0,67	11	59
Median	0	0,02	12	66
3rd quartile	0,68	0,68	14	76
Max	4,07	3,57	17	104

Age more explained

Alzheimer's Disease

10% of people 65 and older have AD. Prevalence is increasing with age. People between 65-74 have a prevalence of 3%, age 75-84 17%, age 85 and older 32%. From the population diagnosed with AD, 81% is older than 75 years of age. (Hebert, Beckett, Scherr, & Evans, 2001)

Neuroticism

The prevalence of N increases with age from adolescence through adulthood (Roberts & DelVecchio, 2000). Mean scores of N are highest in late adolescence and decline mildly through adulthood (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007).

Table 3 – Psychological Problems Diagnosis

Psychological Problems	% (Frequency)
Yes	20,96 (3.921)
No	79,04 (14.784)

Table 4 – Alzheimer Diagnosis

Alzheimer Diagnosis	% (Frequency)
Yes	1,92 (360)
No	98,08 (18.352)

Table 5 - Gender

Gender	% (Frequency)
Female	43,82 (21.066)
Male	56,18 (16.429)

Table 6 - ISCO-08 categories

ISCO-08 Job Categories		
Skill level	**1**	**2**
	food prepping + serving	Community + social services
	building/grounds/clean/maintenance	arts/design/entertainment
	farm/fish/forestry	healthcare support
	extraction workers	protective service
	transport/material moving	personal care + service
		sales occupations
		office + admin support
		construction trades
		install/maintenance/repair workers
		production
		military specific
Skill level	**3**	**4**
	business operations specialists	management
	education/training/library	financial specialists
		computer + math
		Architecture + engineering
		life/physical/social science
		legal
		healthcare practice /tech

Table 7 - Transformation of occupation according to the ISCO-08 standard

ISCO-08 Classification	Frequency	Percent	Cum.
1	1.347	16,96	16,96
2	4,379	55,12	72,08
3	614	7,73	79,81
4	1.604	20,19	100
Total	7.944	100	

Table 8 - transformation of ISCO-08 into White collar

white collar	Frequency	Percent	Cum.
0	5.726	72,08	72,08
1	2.218	27,92	100
Total	7.944	100	

Table 9 – Collinearity between variables regressed on PP

	PP	PGS Neuroticism	White collar	Education	Female	Age
PP	1,000					
PGS Neuroticism	0,054***	1,000				
White collar	-0,035***	-0,048***	1,000			
Education	-0,059***	-0,044***	0,418***	1,000		
Female	0,111***	-0,013	0,015	-0,019***	1,000	
Age	-0,045***	-0,001	0,053***	-0,087***	-0,002	1,000

****significant at 1% **significant at 5%*

Regression Models

Our regression model looks as follow:

$$Y = a + B_1(\text{EDU}) + B_2(\text{OCP}) + J(\text{PGS\_N}) / J(\text{PGS\_AD}) + r_1(\text{EDU} * \text{PGS\_N} / \text{PGS\_AD}) + r_2(\text{OCP} * \text{PGS\_N} / \text{PGS\_AD}) + k + e$$

Y=AD/PP, EDU= Education, OCP= Occupation, PGS_N= Corresponding PGS for N, PGS_AD= Corresponding PGS for AD k= Control Variables (Age, Female,)

This regression estimates how one's probability of being told by a healthcare professional to have AD/PP is affected by education, occupation, genetics and the interaction between genetics and education/occupation. Important to note is that each regression incorporates the PGS of AD when the dependent variable is AD or The PGS of N when the dependent variable is PP. This also holds for the interaction variables, education*PGS and occupation*PGS. The interaction is between education/occupation and the corresponding PGS for the outcome.

Linear Probability Model

We will estimate this equation with a linear probability model (LPM). The LPM has three main disadvantages:

1. Due to the linearity, the model can predict possibilities larger than 1 and smaller than 0. Which, obviously, is impossible.
2. A unit change in one of the independent variables has a constant effect.
3. The distribution of the disturbances ε_i is not normal which follows from deriving that $\varepsilon_i = 1 - x_i\beta$ when $y_i = 1$ and $-x_i\beta$ when $y_i = 0$. This means that the LPM has heteroskedasticity.

Although the LPM has many shortcomings, it still can be the preferred model due to the easy interpretable nature of the model. If the value of independent variables is about the average with little extreme values, it works well in general (Woolridge, 2009). For this paper in particular, the main advantage is that the interaction effects are directly interpretable. A logistic regression gives interaction effects that are not/very hard interpretable as described by (Ai, 2003).

As we saw from the transformed Grossman model, AD/PP is explained by investments in- and (natural) depreciation of the cognitive reserve. The cognitive reserve in this regression is explained by education and occupation. The depreciation and investments effects are different for every individual; therefore, we included the PGS and, most importantly, the interaction between this score and the cognitive reserve as expressed by r_1 and r_2. Finally, we will include the control variables age and female.

Binary Logistic Model

Due to the limitations of the LPM as stated before, a logit model is better suited for data without interactions variable(s). As well to check for the reliability of the LPM we can use a logit model, in particular by studying different population samples based on a certain variable. E.g. when you want to obtain the interaction between the continuous variable x_1 and binary variable x_2 for y. We could make two models; in the first model the population sample is conditional on x_2=0 respectively x_2=1 in the second model. We then can see the different effects for x_1 based on x_2 in each of the two models.

The logistic model is one of the most used models for regressions with a binary or categorical dependent variable because of the non-linear nature. The name is derived from the underlying distribution, a logistic distribution.

By using the cumulative distribution function (CDF) the values are ensured to be between zero and one. The probability of the dependent variable being one is:

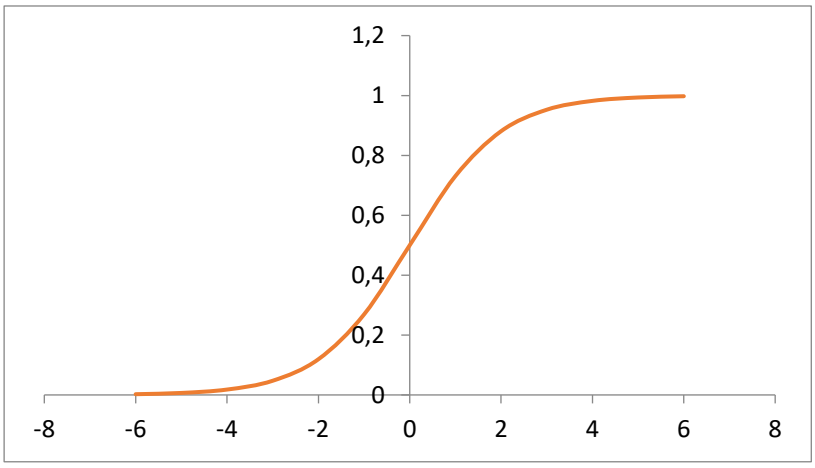
$$P(y_i = 1) = F(x_i\beta)$$

Where F denotes the CDF of the logistic density function:

$$F(x_i\beta) = \frac{e^{(x_i\beta)}}{1 + e^{(x_i\beta)}}, or \frac{1}{1 + e^{(-x_i\beta)}}$$

Graphically shown in figure 3

Figure 3 – CDF distribution



The distribution never exceeds zero or one on the y-axis. At the intersection on the y-axis the tangent line is the steepest at a value of y=0,5.

The logit model has one important assumption:

1. linear relationship between the log odds of the outcome and each predictor variable.

In other words; the continuous predictors of the model are linear with the log odds of the dependent variables. (Hosmer & Lemeshow, 2000) recommend using the Box-Tidwell approach to check for the linearity of the log odds. The Box-Tidwell approach provides a power transformation/log transformation upon the independent variables in order to make the model linear.

Besides that approach, we will use the Linktest as well for correct specification of the models. Linktest generates a logistic regression output of '_hat' and '_hatsq'. The output consists of two important variables as the predictors to rebuild the model:

1. _hat ➔ the linear predicted value
2. _hatsq ➔ linear predicted value squared

Since the variable '_hat' is the predicted value from the model, this variable should be significant in the model. This will be true, unless our model is mis-specified. Contrary to this, '_hatsq' should not have much explaining power, except by chance. If 'hatsq_' is significant then the linktest will be significant, this indicates that there are omitted variables or the link function is not correctly specified (STATA corp, 2019).

Alzheimer Disease Models

Our goal is to retrieve interaction effects for education and occupation with the PGS. Unfortunately, the independent variable, PGS of AD, regressed on the dependent variable of having AD in a logistic regression gave insignificant results. The same holds for the independent interaction variable PGS by education/occupation as seen in table 1A-4A in the appendix. This is most likely due to a low prevalence of AD in our dataset. 360 respondents stated to have been diagnosed with AD, this is 1,92% of all people to have answered this question.

To make the most out of the data, we can use the data to interpret the effects of age and education, which are significant related. White collar was found to be unrelated to AD, see table 5A in the appendix. We did a logistic regression and then checked conditional marginal effects for respondents with the following ages intervals: 55, 65, 75, 85, 95. We did the same for respondents with educational attainment of 1, 2, 3, 6, 9, 12, 15 and 17+ years of education. We did conditional marginal effects to account for the fact that age and education are not linear related to AD. As seen in figure 2A & 3A in the appendix.

With the Box Tidwell test, we tested for nonlinearities in the model. Both age and education were found to be non-linear. To correct for this, we included quadratic and exponential transformations in the model since the distribution of both is skewed to the left see figure 5A and 6A in the appendix. However, these transformations only made the model slightly better and more difficult to interpret. We therefore chose not to transform the variables.

From the logistic regression in table 10 we can interpret the sign of each variable. We can see that the more education obtained, the lower the probability on a diagnosis of Alzheimer. Keeping all other things fixed. When one's age increases, the probability on a diagnosis of AD, logically, increases. Keeping all other things fixed. Both effects are significant at a 1% level.

Table 10: Logistic Regression

Alzheimer Diagnosis	Coefficient	S.D.	P>Z	95% Confidence	Interval
Education	-0,089	0,015	0	-0,117	-0,06
Age	0,116	0,006	0	0,105	0,127
Constant	-11,585	0,499	0	-12,562	-10,607

To interpret the coefficient of age we perform conditional marginal effects to account for different effects for different age intervals. From table 11 we can derive that the probability of being diagnosed with AD is increasing with age. For instance; being aged 55 and becoming 56 increases the

probability of being diagnosed with AD by 0,022 percentage points keeping all other things fixed. This result is significant at a 1% level. The effect is marginally increasing in size regarding the higher age categories. Being aged 95 and becoming 96 increases the probability of being diagnosed with AD by 1,56 percentage points keeping all other things fixed. This result is significant at a 1% level

Table 11 – Conditional marginal effects

Age	Percentage Points	S.D	P>Z	95% Confidence	Interval
55	0,022	0,00002	0	0,0002	0,0003
65	0,069	0,00005	0	0,001	0,001
75	0,214	0,0001	0	0,002	0,002
85	0,627	0,0005	0	0,005	0,007
95	1,555	0,007	0	0,012	0,019

We performed conditional marginal effects for education as well. From table 12 we can see that education does lower one's probability as already indicated by table 9. However, we now can see that this effect has diminishing marginal returns, meaning that the effect of education on AD lessens in magnitude, keeping all other things fixed. Where having 9 years of education an additional year of education lowers the probability of being diagnosed with AD by 0,2 percentage points, having 17 or more years of education an additional year of education lowers this probability by 0,11 percentage points keeping all other things fixed. These results are significant at a 1% level.

Table 12 – Conditional marginal effects

Education *Years*	Percentage Points	S.D.	P>z	95% confidence	Interval
0	-0,378	0,001	0,00	-0,006	-0,002
1	-0,354	0,001	0,00	-0,005	-0,002
2	-0,33	0,001	0,00	-0,005	-0,002
3	-0,308	0,001	0,00	-0,005	-0,001
6	-0,249	0,001	0,00	-0,004	-0,001
9	-0,199	0,0004	0,00	-0,003	-0,001
12	-0,158	0,0003	0,00	-0,002	-0,001
15	-0,125	0,0002	0,00	-0,002	-0,001
17 & 17+	-0,106	0,00009	0,00	-0,001	-0,001

Psychological Problems Models

We start by investigating the relation between the PGS for Neuroticism and PP. We find a significant relation, in a logistic regression, between these variables as shown in Table 6A in the appendix. We then performed logistic regressions on several models with PP as dependent variable and each time one other relevant independent variable included. The variables included were female, age, education, white collar, occupation (ISCO-08). All variables, except occupation, had a significant relation with PP. We then made the following logistic model represented in table 13.

Table 13 – Logistic regression

Psychological Problems	Coefficient	S.D.	P>z	95% Confidence	interval
Female	0,701	0,088	0,00	0,529	0,874
Age	-0,027	0,006	0,00	-0,038	-0,016
White collar	-0,155	0,1	0,12	-0,351	0,04
PGS Neuroticism	0,173	0,043	0,00	0,089	0,257
Education	-0,046	0,021	0,02	-0,086	-0,006
Constant	0,702	0,453	0,12	-0,185	1,59

When interpreting the signs we can see that being female opposed to being male and having a higher PGS increases the probability (both significant at 1%), and being older (significant at 1%), or have more education (significant at 5%) lowers the probability to be told by a doctor you have psychological problems like emotional- ,nervous- or psychiatric problems. White collar is insignificant in the model. This is probably due to the collinearity between education and white collar as seen in table 9 in the descriptive statistics section.

When we remove education from the model in table 13, the coefficient of white collar changes from -0,155 to -.251 and becomes significant at a 1% level. Meaning the magnitude of the effect increases and white collar becomes significant. So, when education enters the model you can clearly see that education explains a large share of white collar's effect on the dependent variable.

To interpret the logistic regression, we estimate the logistic regression from table 13 with average marginal effects, we make some transformations to better specify the model. We use the 'Linktest' command in Stata tot test for this. At first our model was already correctly specified according to the Linktest. However, by looking at the data we found it to be appropriate to include education squared and age squared. This gave an even better model according to the Linktest. We then tried other transformations, but they didn't add extra explaining power to the model. See table 7A and 8A in the appendix for the Linktest output results.

After the correct specification of the model we calculate average marginal effects (for the whole study population) for all variables in the model and conditional marginal effects (for a subpopulation e.g. conditional on age=50) for age, education and the PGS for N since they are not linear related to PP. This relationship is graphically shown in figure 7A-9A in the appendix. The output results of the average marginal effects are shown in table 15 conditional marginal effects are shown in table 16 and 17.

Table 15 – Average marginal effects.

Psychological Problems	Marginal Probability in percentage points	S.D	Significant at
Female	11,837	0,014	1%
Age	-0,391	0,001	1%
White Collar	-3,514	0,016	5%
PGS Neuroticism	2,777	0,007	1%
Education	-0,476	0,004	Insignificant

Although the PGS does not relate in a perfect linear trend to PP, it does approach it well enough to interpret the average effects. To be comprehensive we will also perform conditional marginal effects for the PGS variable. Age does not approach a linear trend and so is not interesting to interpret, see figure 7A-9A in the appendix for a graphical representation. We will use conditional marginal effects for age due to the non-linear relationship.

It follows from table 15 that:

- being female, opposed to being male, increases the probability on average with 11,84 percentage points to being told by a doctor to have psychological problems, keeping all other things fixed. This effect is significant at a 1% level;
- Being white collar opposed to blue collar, decreases the probability of being told by a doctor to have psychological problems on average by 3,51 percentage point, keeping all other things fixed. This effect is significant at a 5% level;
- Having a one-point higher PGS, increases the probability of being told by a doctor to have psychological problems on average with 2,78 percentage points, keeping all other things fixed. This effect is significant at a 1% level. The PGS is a standardized score, meaning one point on the PGS distribution is equal to exactly one standard deviation.

Table 16 – Conditional marginal effects for age

Age *years*	Marginal Probability in percentage points	S.D	Significant at
50	-0,085	0,003	n/a
60	0,39	0,001	1%
70	-0,59	0,001	1%
80	-0,616	0,001	1%
90	-0,477	0,0004	1%

As seen in figure 6 in the appendix, the probability of a doctor telling you have psychological problems like emotional-, nervous- or psychiatric problems, is increasing with age, peaking around the age of 60 where after it diminishes. This is known from other literature. Unfortunately, we don't have data for individuals of 49 years and younger. From table 16 it shows that at (i) 60 (ii) 70 (iii) 80

(iv) 90 years of age, this probability (i) increases with 0,39-; (ii) decreases with 0,59-; (iii) decreases with 0,62-; (iv) decreases with 0,48 percentage points at an additional life year keeping all other things fixed. These effects are significant at a 1% level.

We also tested for conditional marginal effects for education. Although the average marginal effect for education in table 15 is insignificant, it still could be that certain specific years of education are significant. The results however were insignificant or there were too few observations per category to be valid. Please see table 9A and 10A in the appendix for these results.

The PGS for N always has a positive sign. This result is logical since the more the genetics are related to N, the higher you expect the probability of psychological problems to be. Besides the average marginal effects in table 15 it's interesting to see the effect in more detail. Table 17 displays the effects for different PGSs.

Table 17 - Conditional marginal effects for the PGS N

PGS N	Marginal Probability in percentage points	S.D	Significant at
-3,5	1,84267	0,003	1%
-2	2,30083	0,005	1%
0	2,78313	0,007	1%
2	3,24393	0,009	1%
3,5	3,62648	0,012	1%

For instance, having a PGS of 3,5, a one-point increase increases the probability of a doctor telling you have psychological problems by 3,63 percentage points, keeping all other things fixed. This effect is significant at a 1% level.

Interaction Effects

The aim of this paper is to explore whether interaction effects between the cognitive reserve and the PGSs exist. For PP we made several linear probability models with every time (i) a different independent variable, (ii) the PGS of N and (iii) the interaction between the independent variables from (i) and (ii). From these models we only derived a significant interaction between white collar and the PGS for N. There was no interaction between education and the PGS for N. These regression results for white collar and education are shown in the appendix, table 11A – 12A.

It is important to know whether there is a linear effect from the neuroticism PGS on the outcome, otherwise the coefficients' mean does not reflect the true effect. From Figure 9A in the appendix we can derive that the effect is approaching a linear trend, we therefore may interpret the coefficients' mean.

With this information we then make the LPM with interaction effects. We use the Ramsey RESET test to test for missing non-linearities, we find the model in table 18 to have the best specification. With a probability score of 0.1194 we reject the null hypothesis of the RESET test that there are missing non-linearities.

Table 18 – LPM

Neuroticism	Coefficient	S.D.	P>t	95% Confidence	Interval
Neuroticism PGS	0,038	0,009	0	0,020	0,055
White collar	-0,035	0,016	0,032	-0,067	-0,003
white collar#N PGS					
If white collar==1	-0,027	0,015	0,061	-0,056	0,001
Education	-0,076	0,025	0,002	-0,126	-0,027
Education^2	0,003	0,001	0,005	0,001	0,004
Female	0,117	0,014	0	0,089	0,144
Age	-0,05	0,001	0	-0,006	-0,003
Constant	0,999	0,179	0	0,648	1,351

In the model from table 18, with the other control variables included, the interaction effect has less explaining power. It has a P-value of 0,061, with 95% confidence we can say that the interval of the coefficient includes zero and thus has no effect.

To further explore this interaction, we construct two logit models, conditional on white collar being zero, respectively being one, see table 19 and 20. Hereby we can remove the interaction variable from the model, so we won't experience difficulties interpreting these interaction variables.
Only in the model conditional on white collar being 0, the PGS is significant. We can derive that conditional on being blue collar, if one's PGS rises with 1 then the probability of a doctor having told one to have psychological problems, on average, increases with 3,7% percentage points, keeping all other things fixed. This effect is significant at 1%. This finding is consistent with Figure 4, the effect of the PGS has a steeper line among blue-collar workers than among white-collar workers. The logit models even suggest that genetic risk only plays a significant role among blue collar workers. This is consistent with the cognitive reserve hypothesis, that when performing cognitively stimulating work (as in white collar work), the effect of genetic risk on psychological problems is attenuated.

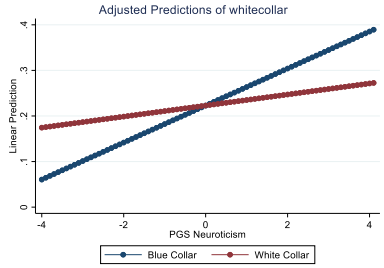
Table 19 – Marginal effects from logistic regression conditional on white collar==0

Psychological Problems	Percentage Points	S.D	P>z	95% Confidence	Interval
PGS Neuroticism	3,679	0,009	0	0,019	0,054
Education	-0,917	0,004	0,04	-0,018	-0,0004
Female	13,359	0,018	0	0,098	0,169
Age	-0,585	0,001	0	-0,008	-0,004

Table 20 – Marginal effects from logistic regression conditional on white collar==1

Psychological Problems	Percentage Points	S.D	P>z	95% Confidence	Interval
PGS Neuroticism	1,098	0,012	0,347	-0,012	0,034
Education	0,057	0,007	0,937	-0,014	0,015
Female	9,464	0,023	0,00	0,049	0,14
Age	-0,235	0,002	0,136	-0,005	0,0007

Figure 4 Probability on PP for white collar and blue collar conditional on the PGS of N



Discussion

The main aim of the paper was to examine if there exist interaction effects between genetics and the cognitive reserve. Our cognitive reserve was explained by education and occupation. Our study population consists of American white people who are 50 years and over. For AD we find no evidence of existing interaction effects. For PP we find moderate evidence suggesting there could exist an interaction between white collar and the PGS for N. We do find for AD that education relates to the development of the disease. More years of education lowers the probability of developing AD, occupational measures are found to be insignificant related. For PP we only found a significant relation for occupation, where education explains a part of this effect. The independent variable education, alone, had no effect. These results, for education for AD and occupation for PP, are supported by our transformed Grossman model; it is beneficial to invest in one's cognitive reserve in order to stay above the individual threshold of the cognitive reserve to prevent the development AD and PP. Although our cognitive reserve includes both education and occupation, these variables are never jointly, within the cognitive reserve concept, significant for PP or AD. Although as stated before, white collar's effect on PP is partly explained by education.

In line with our results for education, in multiple studies the cognitive theory is proposed and supported concerning AD. (Masliah, Terry, Alford, & DeTeresa, 1991) (Shadlen, 2005). Our result is even more important since our study could control for genetic risk, where most other studies could not. We can rule out that one's genetics could explain a part of the potential casual pathway of the effect for education for AD. Occupation was found to be unrelated for AD. For PP there is less literature concerning the link between the cognitive reserve and the development of mental illnesses. In general, in studies concerning education and occupation with respect to the cognitive reserve, occupation is mostly found to have the same negative effect as education but this effect is usually smaller (Andel, 2006). Whereas in our study controlling for genetic risk we only find a relationship for occupation. Education, even without controlling for genetic risk, as mentioned is insignificant for PP.
For the interaction effects we had expected there to be a possible interaction effect between genetics and the cognitive reserve for AD. This is shown by (Shadlen, 2005) where the APOE-e4 variant interacted with education. However due to the low observance of AD in the dataset we did not have enough explaining power to test for these interactions. Even the AD PGS was insignificantly related to AD. There still could exist an interaction effect, but our dataset was too limited.

For PP we were the first to propose that there could exist such an interaction effect, and we found suggestive evidence to accept this hypothesis. Where the LPM had insignificant results for the interaction term, in a logit model the PGS of N had a significant effect on PP for blue-collar workers, but not for white-collar workers. For every one-point increase in the PGS of N, conditional on being blue collar, the probability to have PP increases with 3,61 percentage points; keeping all other things fixed, significant at 1%. The other logistic model conditional on being white collar, the PGS of N was insignificant.

Note that the coefficient of the interaction term in the LPM is -0,027 (table 18), the coefficient of the PGS of N in the two different logistic models conditional on being blue collar and white collar is 0,037 respectively 0,011 (table 19 & 20). The difference in the two coefficient is almost equal to the interaction coefficient in the LPM model. Although not significant, the models do show a consistent result of a possible existing interaction. Our data just has not enough statistical power to produce potential significant results. This is why we would argue that there is suggestive evidence of an

interaction effect between genetics and being white- / blue collar on developing PP. This result should be explored further in future research. This could mean that those who are genetically more inclined to develop PP, can counteract this genetically driven development by enrolling into white collar occupations. In the other direction, our results are consistent with the interpretation that among those with a cognitively stimulating job, being a white collar worker, genetic risk for developing psychological problems is largely negligible. This effect of white collar occupations can be explained in 2 ways:

- Individuals are more cognitively stimulated in a white collar job and so build a larger cognitive reserve.
- Those who have a white collar job have more wealth. Which creates an environment of less stress and more access to healthcare.

Other results

One of our main contributions is the creation of the occupation variable, which initially exists of four categories according to the ISCO-08 division. We find no significant relation between this occupation variable for both PP and AD. This is probably due too low observations per category. When we transform this variable, again according to the ISCO-08 division, into the dummy variable white collar it does have explaining power, but only for PP. Where being white collar opposed to blue collar lowers the probability on developing PP. This effect is explained in multiple ways as mentioned earlier.

We found a quite large effect of being female compared to being male to develop PP. Where being female, opposed to male, increases this probability by 11 percentage points, keeping all other things fixed (significant at 1%). This result is somewhat surprising since general literature agrees that the prevalence of mental illness among men and women is equal, certain diseases are more common among men or women (Department of Mental Health and Substance Dependence, 2013). N for instance is on average more prevalent in women than men. (Eysenck & Eysenck, 1975) (Goodwin & Gotlib, 2004). Our dependent variable PP likely does not have a representative distribution on the mental disorder spectrum.
Gender differences are likely caused by the following pathways, differences in:

- Vulnerability (e.g. woman find it easier to talk about negative emotions);
- Stress exposure (e.g. women are far more likely to be victims of domestic violence or sexual assault);
- Biological factors (e.g. hormones, brain pathology etc.);
- Psychological factors (e.g. girls and boys are brought up different);
- Social factors (e.g. man possess greater role of power through occupations) (Mezulis & Harding, 2016).

For AD we did not find a different effect between being male and female. However, we know from literature that women experience a higher prevalence of AD. 66% of the Americans above 65 years with the disease are women (Hebert, Beckett, Scherr, & Evans, 2001). For people aged 71 and older 16% of women have AD or other forms of dementia, whereas for males this is 11% (Plassman, Langa, & Fisher, 2007)**.**

The most common argumentation for this discrepancy is that women have a higher life expectancy than men. Other arguments have been developed as well:

1. It is more common for men in middle ages to die from cardiovascular disease than women, so men who live after being 65 may have a healthier cardiovascular state and thus a lower risk of developing AD or other dementia since dementia is related to cardiovascular health (Chêne & Beiser, 2015);
2. Research found that the APOE-e4 genotype, the genotype with the highest relation with AD, may have a stronger relationship with AD in women than in men (Altmann, 2014). Although other research did not find such an association (Farrer, 1997);
3. Low education is found to be a risk factor of developing AD (Stern Y. A., 2012). Therefore, women born in the first half of the 20th century are at greater risk of developing AD (Rocca, 2014).

Limitations

Dataset

Due to the low observance of AD among the respondents even the AD PGS was not related to having AD. Furthermore, we do not know how the diagnosis of AD is set, and respondents could be in denial, too embarrassed or being too much cognitively impaired to acknowledge they had been diagnosed with AD. This could be reasons why the observance of AD was low. As with AD, we do not know how the diagnosis of PP is set and the same problem of wrong / dishonest responses could arise.

For both AD and PP diagnosis criteria may vary across countries, which makes the results less generalizable. Furthermore, the definition of PP is very broad, it essentially covers all diseases across the mental disorder spectrum. This means relationships found are very broad as well and don't relate to a certain mental disease but gives more of an indicative result. To test for the cognitive reserve hypothesis we use two measures, education and occupation. However, for AD only education is found to be significant and for PP only white collar is found to be significant. A part of the effect of white collar is explained by education. This is due to the correlation; one does first obtain enough education to be eligible for white collar jobs. This makes the cognitive reserve concept less strong in explaining power. We don't see this as a problem since both independently are important measures of the cognitive reserve. But it does diminish the strength of the cognitive reserve concept in this context.

However, this does not necessarily need to be true; The cognitive reserve is determined by education and occupation. Where education explains the first stage of life where education is obtained. After the education is completed one gets a certain occupation and this explains the second stage of life. So, education measures how one initially builds a cognitive reserve, where in the second stage one's occupation measures how well one does maintain this cognitive reserve or even improves it. So, the two variables measure the cognitive reserve in different stages of life.

Another concern is the age of all respondents which is 50 years and over. This is limited for PP since PP also develops in earlier stages of life. So, our study population for PP is not a representative sample and thereby it might not capture life circumstances closer to the diagnosis.

Finally, we explain having PP by the PGS of N. Although it's a good predictor since it is a pre-stadium of several psychological diseases. Neuroticism likely has a different relation for different mental disorders, in this sense PP is too broad defined.

Statistical design

As stated earlier, we try to retrieve interaction effects through the LPM. The LPM outcomes could be biased because of the nonlinear nature of having PP, since it's a binary variable. It was however still the preferred model due to the easy interpretable nature of the model. If the value of independent variables is about the average with little extreme values, it works well in general (Woolridge, 2009). With PP there could also exist reverse causality with education. Students could develop PP and thereby dropping out of college and thereby decreasing the attained years of education. This logically also has an effect on one's occupation. Since high education usually is a requirement for white collar jobs.

Furthermore (household) income can be described as a mediator. Individuals obtain a certain education and thereby they have access to (limited) occupations that yield a certain income. Where more income creates a more wealthy environment for the individual and likely more health supporting behavior and more funds to spend on healthcare (Dunn, 2000). This could affect the outcome variable.

(Parental) socio economic status (SES) is a confounder in this study. It is an important variable, especially one's childhood SES likely has a large impact. Higher SES gives more opportunity for education, there is more support from the parents and there is no need to work at young age since there are enough funds. For instance, prevalence of depressed mood or anxiety has been found to be 2,5 times higher among young people aged 10 to 15 years with low SES opposed to youths with high SES (Lemstra, 2008).

Another potential confounder is the PGS for each of the parents of the individual, since the PGS of the individual is the direct result from the parents' genetics. The parents PGS could also influence the dependent variables. Some choices in life could be genetically driven, e.g. the neighborhood a child is brought up etc. Likewise, family history of the disease is important as well to include. Family history is largely determined by genetics but also by environmental factors which could stimulate the disease to develop such as low SES.

Finally, we constructed the cognitive reserve with education and occupation. These are factors which indicate one's cognitive ability, however variables such as childhood IQ, pre-morbid IQ or current IQ measure the cognitive ability directly. With these variables we could construct a more powerful cognitive reserve and thereby possibly gaining more explaining power for the data.

Thus, our study suffers from the omitted variable bias. Other variables that much likely should be included are health variables such as alcohol intake, exercising and diet nutrients which for instance are related to depression (Berk, 2013). Social factors as place of birth, parental age, season of birth, neonatal vitamin d exposure, maternal health, access to healthcare are all related to mental health as found in a meta-study by (Agerbo, 2015). For AD there is consensus about the following risk factors smoking, hypertension, diabetes, obesity, lack of exercise, alcohol and head injuries (He, Goodkind, & Kowal, 2015).

We believe these limitations do have impact on our study results. However, most significant results we do obtain are comparable with other literature. However, we would probably obtain more sound results if these limitations where not an issue. The biggest impact could be when we were testing for interactions. This demands more explaining power from the data. It could be that interactions between genetics and the cognitive reserve do exist in our respondents for AD. These interaction effects have been found by (Shadlen, 2005) already, although there are studies which don't find these interactions such as (Garibotto, 2011). And finally, with less limitations we possibly would find more powerful results for interactions regarding PP. We already have suggestive evidence of an

interaction effect between white collar and genetics. More powerful data could potentially truly establish this relationship.

The Grossman Model

As with many economic models, our transformed Grossman model is a simplified model of reality. Our model assumes that when an individual's stock of the cognitive reserve falls below a certain threshold, this individual automatically has the disease. Our model does incorporate that this threshold is different for every individual, but even qualifying for a too low cognitive reserve does not automatically mean one will develop the disease.

Implications

Alzheimer Disease

We found suggestive evidence which supports the cognitive reserve hypothesis for AD. For AD this means the onset of clinical symptoms can be postponed to later stages in age. This could be done by investing in one's cognitive abilities during their lifetime, especially those exposed to greater risk of developing AD from genetic risk and low education. These results could be incorporated into health treatment to prevent the onset of symptoms. However, a causal pathway of the effect is needed first.

Investing in the cognitive reserve is supported by our transformed Grossman model, where individuals can invest in their cognitive reserve, thereby making sure it does not fall below their personal threshold (given by genetics and environmental factors) for the disease to develop.
To prevent the onset of AD they can invest in education to increase the cognitive reserve. It is according to our results not possible to modify the effect of the PGS through education, since there is no interaction effect. The depreciation in the model is positively related with age, this relation is exponential. This means that the depreciation is exponentially increasing in magnitude with age. Those who have a high cognitive reserve and thus express clinical symptoms in later stages deteriorate faster, but only when the first clinical symptoms start to arise, than those with a low cognitive reserve (Andel, 2006). This means that the depreciation rate of the cognitive reserve is lower for those with a high cognitive reserve, however when the first clinical symptoms start to express, the depreciation rate increases to a higher rate than those with a low cognitive reserve.

Psychological Problems

For PP the relationship with the cognitive reserve is harder to establish. Our hypothesis is that through the cognitive reserve one can work around their 'brain impairments'. For example, it improves a person's ability to act more rational, thereby suppressing impulsive/emotional behavior and being a calmer and emotionally stable person. In that way being more likely to thrive in education and/or their career path.

Where the cognitive reserve can be explained by three complementary mechanisms:

1. by affecting the risk for developing the mental disorder;
2. in the expression of symptoms within the mental disorder;
3. in patients' functional outcome (social and vocational outcomes) (Barnett, Salmond, Jones, & Sahakian, 2006).

We know occupation has an effect on developing PP, however through which mechanism this works is hard to establish. Although we do not know if the cognitive reserve theory is the correct pathway, it's interesting to put the results according to this theory into the Grossman model. Thereby showing the key differences between AD.
In the transformed Grossman model for PP, one can only invest in a white collar occupation to increase their cognitive reserve. If there exists an interaction between white collar and the PGS, for which we found suggestive evidence, then choosing a white collar job, if possible, makes the effect of the PGS largely negligible. It could be possible to modify the effect of the PGS through a white collar job. On a final note, the depreciation of the cognitive reserve increases with age to one's fifties/sixties where after this threshold the depreciation rates lowers again. Although this effect of age could be influenced by many factors and thus may not necessarily represent the true effect of age. It could be for instance that the elderly are mostly retired and therefore don't experience possible psychological problems at work, which makes the need for a diagnose and treatment less urgent. They can still function in daily life without the need of treatment.

Besides the cognitive reserve there are two other ways to describe the relationship with the development of PP.

1. Those with poorer psychological health during their education period are more susceptible to drop out or achieve lower grades. Thereby limiting their access to job opportunities as well;

2. Certain characteristics, e.g. positive experiences of parenting and/or cognitive ability – may expose those to better health outcomes, to educational investment and success (Kjelsberg, 1999).

Further research is needed to examine the link between the cognitive reserve and PP. If evidence is found for a causal relationship, then investing in the cognitive reserve could be incorporated in health treatment. The best-case scenario would be that the cognitive reserve could possibly reverse the biological cause of PP. It is however more likely that it could prevent the expression of symptoms.

Future Research Recommendations

Further research is needed to test for interaction effects between the cognitive reserve and genetics for AD. AD already has a strong proven relation between the disease and the cognitive reserve. One step further is to test for interaction effects between genetics and the cognitive reserve.
For PP there are multiple pathways in which variables like education and occupation relate to PP. So, for PP it is needed to examine this relationship more in depth. Could it be that the cognitive reserve could prevent the expressions of symptoms and, one step further, could the biological onset in the brain pathology be prevented? At last, since we found suggestive evidence of an existing interaction effect between genetics and occupations on PP, this is worthwhile to be explored further. This is especially interesting when differentiating the broad definition of PP into specific mental diseases such as depression, schizophrenia and more.

Bibliography

Agerbo, E. S. (2015). olygenic Risk Score, Parental Socioeconomic Status, Family History of Psychiatric Disorders, and the Risk for Schizophrenia. *JAMA Psychiatry*.

Ai, C. &. (2003). Interaction terms in logit and probit models. *Economics Letters*, 123 - 129.

Allen, J., Balfour, R., & Bell, R. (2014). Social determinants of mental health. *Internatonal Review of Psychiatry* , 392-407.

Altmann, A. T. (2014). Sex modifies theAPOE-related risk of developing Alzheimer disease. *Annals of Neurology*, 563 - 573.

Alzheimer Nederland. (2019, November 27). *Factsheet cijfers en feiten over dementie | Alzheimer Nederland.* Retrieved from Alzheimer Nederland: https://www.alzheimer-nederland.nl/factsheet-cijfers-en-feiten-over-dementie

Alzheimer's Association. (2016). 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia 12*, 459 - 509.

Andel, R. V. (2006). The effect of education and occupational complexity on rate of cognitive decline in Alzheimer's patients. *Journal of the International Neuropsychological Society, 12*, 310 - 320.

Arenaza-Urquijo, E. M.-L.-P.-F. (2013). Cognitive Reserve Proxies Relate to Gray Matter Loss in Cognitively Healthy Elderly with Abnormal Cerebrospinal Fluid Amyloid-β Levels. *Journal of Alzheimer's Disease 35*, 715 - 726.

Barnett, J. H., Salmond, C. H., Jones, P. B., & Sahakian, B. J. (2006, april 6). Cognitive reserve in neuropsychiatry. Cambridge University Press.

Berk, M. W. (2013). So depression is an inflammatory disease, but where does the inflammation come from? *BMC Medicine*.

Boyle, L. L. (2010). Trait Neuroticism, Depression, and Cognitive Function in Older Primary Care Patients. *The American Journal of Geriatric Psychiatry*, 305 - 312.

Centraal Bureau Statistiek. (2019, November 27). *Prognose bevolking; geslacht en leeftijd, 2019-2060.* Retrieved from Statline: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84346NED/table?ts=1574854436566

Chêne, G., & Beiser, A. (2015). Gender and incidence of dementia in the Framingham Heart Study from mid-adult life. *Alzheimer's & Dementia*, 310 - 320.

Coetzer, E. P., & Rothmann, S. (2003). The big five personality dimensions and job performance. *SA Journal of Industrial Psychology*, 29.

Department of Mental Health and Substance Dependence. (2013). *Gender Disparities in Mental Health.* Geneva: World Health Organization.

Dubois, B., Hampelab, H., & Feldman, H. H. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 292 - 323.

Dunn, J. R. (2000). Social inequality, population health, and housing: a study of two Vancouver neighborhoods. *Social Science & Medicine*, 563–587.

Eysenck, H. J., & Eysenck, S. B. (1975). *Manual of the Eysenck Personality Questionnaire (junior and Adult).* London: Hodder & Stoughton.

Farrer, L. A. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *JAMA: The Journal of the American Medical Association, 278*, 1349 - 1356.

Gale, C. R. (2003). Foetal and postnatal head growth and risk of cognitive decline in old age. *Brain*, 2273 - 2278.

Garibotto, V. B. (2011). Education and occupation provide reserve in both ApoE ε4 carrier and noncarrier patients with probable Alzheimer's disease. *Neurological Sciences*, 1037-1042.

Golam, M., Khandakera, J., Barnettablan, R., White, R., & Jones, B. (2011). A quantitative meta-analysis of population-based studies of premorbid intelligence and schizophrenia. *Schizophrenia Research*, 220-227.

Goodwin, R. D., & Gotlib, I. H. (2004). Gender differences in depression: the role of personality factors. *Psychiatry Research*, 135-142.

Grande, I. S.-M.-A. (2017). High cognitive reserve in bipolar disorders as a moderator of neurocognitive impairment. *Journal of Affective Disorders*, 621-627.

Grossman, M. (1973). *The Demand for Health: A Theoretical and Empirical Investigation.* California: NBER.

He, W., Goodkind, D., & Kowal, P. (2015). *An Aging World: 2015.* Washington DC: U.S. Government Publishing Office.

Hebert, L., Beckett, L., Scherr, P., & Evans, D. (2001). Annual incidence of Alzheimer disease in the United States projected to the years 2000 through 2050. *Alzheimer Dis Assoc Disord*, 73 - 169.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression.* New York: John Wiley & Sons, Inc.

International Labour Organization. (2012). *International Standard Classification of Occupations.* Geneva: ISCO-08.

J.Selkoe, D. (1991). The molecular pathology of Alzheimer's disease. *Center for Neurologic Diseases Harvard Medical School Brigham and Women's Hospital Boston*, 487 - 498.

Jelicic, M. B. (2003). Neuroticism Does Not Affect Cognitive Functioning in Later Life. *Experimental Aging Research*, 73 - 78.

Jones, R. N. (2001). Education Bias in the Mini-Mental State Examination. *International Psychogeriatrics 13*, 299 - 310.

Jorm, A. F. (1988). Educational level differences on the Mini-Mental State: the role of test bias. *Psychological Medicine 18*.

Kachman, D. J. (1987, June). *The Effects of Rational Emotive Education on the Rationality, Neuroticism and Defense Mechanisms of Adolescents.* Retrieved from Western Michigan University: https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=3232&context=dissertations

Katzman, R. T. (1988). Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques. *Annals of Neurology*, 138 - 144.

Kendler KS, G. M. (2006). Personality and Major Depression. *Archives of General Psychiatry*, 1113.

Kesler, S. R. (2003). Premorbid intellectual functioning, education, and brain size intraumatic brain injury: an investigation of the cognitive reserve hypothesis. *Applied Neuropsychology*, 153 - 162.

Kim, J., Basak, J. M., & Holtzman, D. M. (2009). The Role of Apolipoprotein E in Alzheimer's Disease. *Neuron*, 287 - 303.

Kjelsberg, E. (1999). A long-term follow-up study of adolescent psychiatric in-patients. Part IV. Predictors of a non-negative outcome. *Acta Psychiatrica Scandinavica*, 247-251.

Koenen, K. C. (2009). Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *American Journal of Psychiatry*, 50-57.

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, 241 - 256.

Lemstra, M. N. (2008). A systematic review of depressed mood and anxiety by SES in youth aged 10–15 years. *Canadian Journal of Public Health*, 125-129.

Lo, R. Y. (2013). Effect of Cognitive Reserve Markers on Alzheimer Pathologic Progression. *Alzheimer Disease & Associated Disorders*, 343-350.

M., W. J. (2009). *Introductory Econometrics: A Modern Approach.* Mason: South-Western Cengage Learning.

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, 101 - 114.

Masliah, E., Terry, R., Alford, M., & DeTeresa, R. (1991). Cortical and subcortical patterns of synaptophysinlike immunoreactivity in Alzheimer's disease. *The American Journal of Pathology*, 235 - 246.

Mezulis, A., & Harding, K. (2016). Gender Differences and Similarities in Mental Health. *Encyclopedia of Mental Health,*, 262-268.

Partnership to Fight Chronic Disease. (n.d.). *Alzheimer.* Retrieved from Fight Chronic Disease: https://www.fightchronicdisease.org/sites/default/files/PFCD_USA_Alzheimers_Factsheet_v13%20%281%29.pdf

Patel, V. S. (2016). Sustainable development and global mental health—a Lancet Commission. *The Lancet*, 1143-1145.

Perkins, A. M., & Corr, P. J. (2006). Cognitive ability as a buffer to neuroticism: Churchill's secret weapon? . *Personality and Individual Differences*, 39-51.

Plassman, B., Langa, K., & Fisher, G. (2007). Prevalence of Dementia in the United States: The Aging, Demographics, and Memory Study. *Neuroepidemiology*, 125-132.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 3-25.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on sychological Science*, 313-345.

Rocca, W. A. (2014). Sex and gender differences in the causes of dementia: A narrative review. . *Maturitas 79*, 196 - 201.

Roe, C. M. (2007). Education and Alzheimer disease without dementia: Support for the cognitive reserve hypothesis. *Neurology 68*, 223 - 228.

Shadlen, M. L. (2005). Education modifies the effect of apolipoprotein epsilon 4 on cognitive decline. *Neurobiology of Aging 26*, 17 - 24.

Smith TW, M. J. (2006). Review Personality and risk of physical illness. *Annual Review of Clinical Psychology*, 435 - 467.

Soldan, A. P. (2013). Relationship of cognitive reserve and cerebrospinal fluid biomarkers to the emergence of clinical symptoms in preclinical Alzheimer's disease. *Neurobiology of Aging*, 2827-2834.

STATA corp. (2019, oktober 10). *rlinktest.* Retrieved from Stata: https://www.stata.com/manuals13/rlinktest.pdf

Stern, Y. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA: The Journal of the American Medical Association, 271*, 1004 - 1010.

Stern, Y. A. (1995). Relationship between lifetime occupation and parietal flow. *Neurology 45*, 55 - 60.

Stern, Y. A. (2012). Cognitive reserve in ageing and Alzheimer's disease. . *The Lancet Neurology 11*.

Van Os J, J. P. (2001). Neuroticism as a risk factor for schizophrenia. *Psychol Med*, 1129 - 1134.

Vemuri, P. K. (2014). Cognitive reserve and Alzheimer's disease biomarkers are independent determinants of cognition. *Alzheimer's Disease Neuroimaging Initiative*.

Ware, E., Schmitz, L., Gard, A., & Faul, J. (2018). *HRS Polygenic Scores – Release 3.* Michigan: University of Michigan.

Woolridge, J. M. (2009). *Introductory Econometrics: A Modern Approach.* Mason: South-Western Cengage Learning.

Appendix

Figure 1A | Job description Accountant

Summary Report for:
13-2011.01 - Accountants

Updated 2019
Bright Outlook

Analyze financial information and prepare financial reports to determine or maintain record of assets, liabilities, profit and loss, tax liability, or other financial activities within an organization.

Sample of reported job titles: Accountant, Accounting Manager, Accounting Officer, Accounting Supervisor, Business Analyst, Certified Public Accountant (CPA), Cost Accountant, General Accountant, Project Accountant, Staff Accountant

View report: Summary | Details | Custom

Tasks | Technology Skills | Tools Used | Knowledge | Skills | Abilities | Work Activities | Detailed Work Activities | Work Context | Job Zone | Education | Credentials | Interests | Work Styles | Work Values | Related Occupations | Wages & Employment | Job Openings | Additional Information

Tasks

5 of 17 displayed

- Develop, maintain, and analyze budgets, preparing periodic reports that compare budgeted costs to actual costs.
- Prepare, examine, or analyze accounting records, financial statements, or other financial reports to assess accuracy, completeness, and conformance to reporting and procedural standards.
- Review accounts for discrepancies and reconcile differences.
- Prepare adjusting journal entries.
- Establish tables of accounts and assign entries to proper accounts.

back to top

Technology Skills

5 of 28 displayed Show 8 tools used

- **Accounting software** — Fund accounting software; Intuit QuickBooks; Sage 50 Accounting; Tax software
- **Data base user interface and query software** — Best Software CPAPayroll; Data entry software; Microsoft Access; Yardi
- **Enterprise resource planning ERP software** — Microsoft Dynamics; Oracle PeopleSoft; Sage ERP Accpac; SAP Business Objects
- **Financial analysis software** — Brentmark Estate Planning Quickview; Delphi Technology; Oracle E-Business Suite Financials; RSM McGladrey Auditor Assistant
- **Tax preparation software** — ATX Total Tax Office; BNA Income Tax Planning Solutions; Intuit Lacerte; Orrtax Software IntelliTax Classic

Knowledge

5 of 7 displayed

- **Economics and Accounting** — Knowledge of economic and accounting principles and practices, the financial markets, banking and the analysis and reporting of financial data.
- **Mathematics** — Knowledge of arithmetic, algebra, geometry, calculus, statistics, and their applications.
- **Clerical** — Knowledge of administrative and clerical procedures and systems such as word processing, managing files and records, stenography and transcription, designing forms, and other office procedures and terminology.
- **English Language** — Knowledge of the structure and content of the English language including the meaning and spelling of words, rules of composition, and grammar.
- **Law and Government** — Knowledge of laws, legal codes, court procedures, precedents, government regulations, executive orders, agency rules, and the democratic political process.

back to top

Skills

5 of 17 displayed

- **Active Listening** — Giving full attention to what other people are saying, taking time to understand the points being made, asking questions as appropriate, and not interrupting at inappropriate times.
- **Critical Thinking** — Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, conclusions or approaches to problems.
- **Mathematics** — Using mathematics to solve problems.
- **Reading Comprehension** — Understanding written sentences and paragraphs in work related documents.
- **Speaking** — Talking to others to convey information effectively.

back to top

Abilities

5 of 18 displayed

- **Mathematical Reasoning** — The ability to choose the right mathematical methods or formulas to solve a problem.
- **Written Comprehension** — The ability to read and understand information and ideas presented in writing.
- **Deductive Reasoning** — The ability to apply general rules to specific problems to produce answers that make sense.
- **Near Vision** — The ability to see details at close range (within a few feet of the observer).
- **Number Facility** — The ability to add, subtract, multiply, or divide quickly and correctly.

Work Activities

5 of 23 displayed

- **Interacting With Computers** — Using computers and computer systems (including hardware and software) to program, write software, set up functions, enter data, or process information.
- **Getting Information** — Observing, receiving, and otherwise obtaining information from all relevant sources.
- **Communicating with Supervisors, Peers, or Subordinates** — Providing information to supervisors, co-workers, and subordinates by telephone, in written form, e-mail, or in person.
- **Documenting/Recording Information** — Entering, transcribing, recording, storing, or maintaining information in written or electronic/magnetic form.
- **Establishing and Maintaining Interpersonal Relationships** — Developing constructive and cooperative working relationships with others, and maintaining them over time.

back to top

Detailed Work Activities

5 of 19 displayed

- Examine financial records or processes.
- Maintain data in information systems or databases.
- Analyze business or financial data.
- Examine financial records.
- Prepare financial documents.

back to top

Work Context

5 of 21 displayed

- **Electronic Mail** — 99% responded "Every day."
- **Importance of Being Exact or Accurate** — 92% responded "Extremely important."
- **Telephone** — 89% responded "Every day."
- **Indoors, Environmentally Controlled** — 91% responded "Every day."
- **Structured versus Unstructured Work** — 77% responded "A lot of freedom."

Job Zone

Title	Job Zone Four: Considerable Preparation Needed
Education	Most of these occupations require a four-year bachelor's degree, but some do not.
Related Experience	A considerable amount of work-related skill, knowledge, or experience is needed for these occupations. For example, an accountant must complete four years of college and work for several years in accounting to be considered qualified.
Job Training	Employees in these occupations usually need several years of work-related experience, on-the-job training, and/or vocational training.
Job Zone Examples	Many of these occupations involve coordinating, supervising, managing, or training others. Examples include accountants, sales managers, database administrators, graphic designers, chemists, art directors, and cost estimators.
SVP Range	(7.0 to < 8.0)

Figure 2A | Probability on AD by age

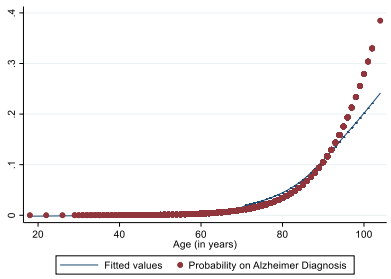


Figure 3A | Probability of AD by education

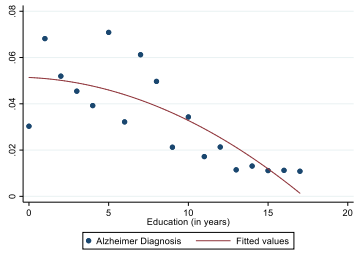


Figure 4A

Withdrawn

Figure 5A | Distribution of education

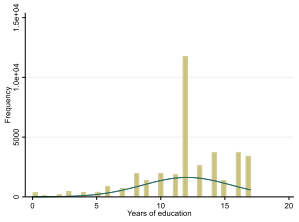


Figure 6A | Distribution of age

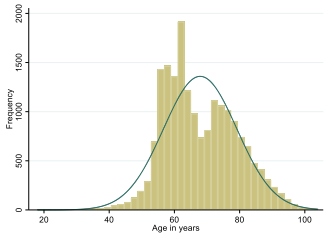


Figure 7A | Probability on PP by age and PGS of N

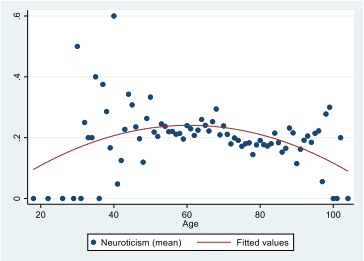


Figure 8A | Probability on PP by education and PGS of N

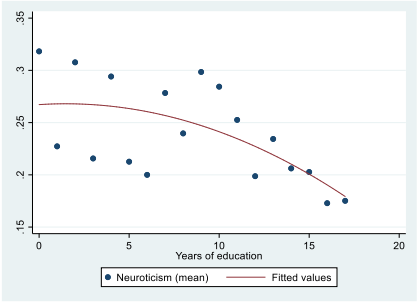


Figure 9A | Probability on PP by PGS of N

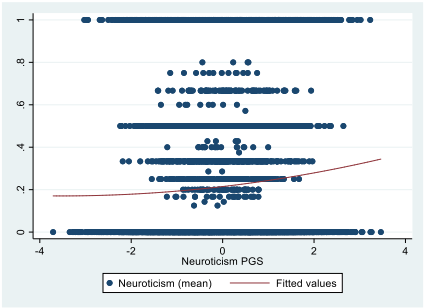


Table 1A | Logistic regression

Alzheimer	Coefficient	S.D.	z	P>z	95% Confidence	interval
PGS Alzheimer	0,1010394	0,0778682	1,3	0,194	-0,0515796	0,2536583
Constant	-3,992589	0,0781249	-51,11	0,00	-4,1457110	-3,839467

Table 2A | Logistic regression

Alzheimer	Coefficient	S.D.	z	P>z	95% Confidence	interval
Education*PGS AD	0,0001299	0,0000956	1,36	0,174	-0,0000575	0,0003172
Constant	0,0182595	0,001396	13,08	0,00	0,0155229	0,020996

Table 3A | Logistic regression

Alzheimer	Coefficient	S.D.	z	P>z	95% Confidence	interval
Occupation* PGS AD						
Skill level 1	0,0022186	0,0022723	0,98	0,329	-0,0022366	0,0066738
Skill level 2	0,0003598	0,0020915	0,17	0,863	-0,0037409	0,0044604
Skill level 3	-0,0028494	0,0029352	0,97	0,332	-0,0086043	0,0029056
Skill level 4	-0,0009679	0,0038626	0,25	0,802	-0,0085411	0,0066054
Constant	0,0049714	0,0012037	4,13	0,00	0,0026113	0,0073314

Table 4A | Logistic regression

Alzheimer	Coefficient	S.D.	z	P>z	95% Confidence	interval
White collar*PGS AD						
0	0,0006651	0,0017872	0,37	0,71	-0,0028389	0,0041692
1	-0,00153	0,0028434	0,54	0,591	-0,0071049	0,004045
Constant	0,0049731	0,0012038	4,13	0,00	0,0026128	0,0073334

Table 5A | Average marginal effects

Alzheimer	Percentage Points	S.D.	z	P>z	95% Confidence	Interval
whitecollar	-0,18111	0,001601	-1,13	0,258	-0,0049496	0,001327

Table 6A | Logistic regression

Psychological Problems	Coefficient	S.D.	Z	P>z	95% Confidence	interval
PGS N	0,1328685	0,025464	5,22	0	0,0829599	0,1827771
Constant	-1,2823	0,025289	-50,71	0	-1,331866	-1,232735

Table 7A | Linktest output – Original Model

PP	Coefficient	S.D.	Z	P>z	95% Confidence	Interval
_hat	1,375821	0,395648	3,48	0,001	0,6003658	2,151276
_hatsq	0,149381	0,152736	0,98	0,328	-0,1499753	0,448737
_cons	0,202122	0,2367	0,85	0,393	-0,2618014	0,666045

Table 8A | Linktest output – Transformed Model

PP	Coefficient	S.D.	Z	P>z	95% Confidence	Interval
_hat	1,160888	0,31394	3,7	0,00	0,5455776	1,776199
_hatsq	0,0636203	0,11848	0,54	0,591	-0,1685954	0,295836
_cons	0,0863579	0,197352	0,44	0,662	-0,3004441	0,47316

Table 9A - Conditional marginal effects for education

Education *years*	Marginal Probability in percentage points	Significant at
1	-6,581	1%
2	-7,127	1%
3	-7,253	1%
6	-5,63	1%
9	-3,173	1%
12	-1,306	1%
15	0,0928	n/a
18	1,5286	n/a

Table 10A | Education

Years of education	Freq.	Percent	Cum.
0.none	2	0.06	0.06
1	0	0	0
2	1	0.03	0.09
3	2	0.06	0.15
5	2	0.06	0.21
6	7	0.21	0.41
7	12	0.35	0.77
8	34	1.00	1.77
9	38	1.12	2.90
10	64	1.89	4.79
11	80	2.36	7.15
12	997	29.46	36.61
13	325	9.60	46.22
14	475	14.04	60.25
15	159	4.70	64.95
16	613	18.11	83.07
17 and 17+	573	16.93	100.00
Total	3,384	100.00	

Table 11A | LPM - interaction for education

Psychological Problems	Coefficient	S.D.	t	P>t	95% Confidence	interval
PGS for N	0,0531671	0,025155	2,11	0,035	0,0038574	0,102477
Education	-0,011331	0,001818	-6,23	0,00	-0,0148938	-0,00777
PGS for N*education	-0,0023813	0,001816	-1,31	0,19	-0,0059405	0,001178
Constant	0,3696558	0,025114	14,72	0,00	0,320427	0,418885

Table 12A | LPM - interaction for white collar

Psychological Problems	Coefficient	S.D.	t	P>t	95% Confidence	interval
PGS for N	0,0400833	0,009102	4,4	0,00	0,0222381	0,057929
white collar	-0,0459124	0,014474	-3,17	0,002	-0,07429	-0,01753
PGS for N*whitecollar	-0,0301302	0,014696	-2,05	0,04	-0,0589438	-0,00132
Constant	0,2397208	0,009152	26,19	0,00	0,2217774	0,257664