Polygenicity of Cognitive Ability and Educational Attainment in Multiple Sclerosis

by

Gerald Trojillo

A Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

of the University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

at the

College of Pharmacy

University of Manitoba

Winnipeg, Manitoba

January 6, 2026

Table of Contents

this page is intentionally left blank

Abstract

Background: Lower educational attainment is associated with higher risk of multiple sclerosis (MS). In those with MS, cognition may be impaired. Cognitive ability and educational attainment are both heritable traits, with the latest large genome-wide association studies identifying >200 and >1,000 genomic loci associated with these traits, respectively but these studies did not investigate these traits in people with MS. Polygenic scores (PGS) consider the cumulative effects of multiple genes on a trait. Studies have not yet examined the association of the PGS for cognitive ability and educational attainment, specifically in persons with MS (PwMS).

Objectives: In PwMS, I aimed to determine whether (a) a cognitive ability PGS is positively associated with information processing speed and (b) an educational attainment PGS is positively associated with years of education and education level.

Methods: I used existing data from three cohorts: a study of psychiatric comorbidity in MS from Canada (IMID); an American 3-arm, multi-center, Phase III clinical trial on combination disease modifying therapy for MS (CombiRx); and the UK Biobank (UKB), a population-based study of 502,655 individuals. PwMS completed various of measures of information processing speed: Symbol Digit Modalities Test (IMID), Paced Auditory Serial Addition Test (CombiRx) or Digit Symbol Substitution Test (UKB) and reported their number of years of education (IMID, UKB) or highest education level achieved (all cohorts). I generated the PGS for cognitive ability and educational attainment based on the largest genome-wide association studies of both traits using SBayesR, and standardized to mean=0, standard deviation=1. I tested the association of the cognitive ability PGS with information processing speed and then the educational attainment PGS with education years or level in PwMS using regression. Analyses were completed by cohort and then pooled together using fixed-effect meta-analysis.

Results: I included 2092 PwMS of European genetic ancestry (213 IMID, 602 US, 1,277 UK). The average age and sex were similar across cohorts. Meta-analyses revealed in PwMS, higher cumulative cognitive ability PGS was associated with higher information processing speed (beta per standard deviation [SD]=0.59 standard error [SE]=0.23, P=0.01). Likewise, higher cumulative genetic score for educational attainment was associated with more years of education (beta per SD=0.56, SE=0.10, P<0.001) in PwMS (excluding US cohort) or higher education (beta per SD=0.39, SE=0.06, P<0.001). The variance explained by PGS were similar to non-MS cohorts (range cognitive ability PGS R^2: 0.035-0.094, educational attainment PGS R^2: 0.034-0.15).

Conclusions: Polygenicity for higher cognitive functioning and educational attainment were associated with information processing speed and education in PwMS, respectively. This study was able to show associations found in the MS population were consistent with similar studies done in general, non-MS populations.

List of Symbols and Abbreviations

DMT = Disease Modifying Therapy

DSST = Digit Symbol Substitution Test

EBV = Epstein-Barr Virus

EDSS = Expanded Disability Status Score

GWAS = Genome Wide Association Studies

HADS = Hospital Anxiety and Depression Scale

IMID = Immune-Mediated Inflammatory Disease

LD = Linkage Disequilibrium

MS = Multiple Sclerosis

OR = Odds Ratio

PASAT = Paced Auditory Serial Addition Test

PGS = Polygenic Scores

PwMS = People with Multiple Sclerosis

RRMS = Relapsing-relapsing Multiple Sclerosis

RRR = Recurrent Risk Ratio

SD = Standard Deviation

SE = Standard Error

SDMT = Symbol Digit Modalities Test

SNP = Single Nucleotide Polymorphisms

SPMS = Secondary Progressive Multiple Sclerosis

TOPMed = Trans-Omics for Precision Medicine

UKB = UK Biobank

95% CI = 95% Confidence Interval

Acknowledgements

I would like to start by thanking my supervisor, Dr. Kaarina Kowalec. In 2020, she took me in when I was a young and lost undergraduate student for summer research and showed me the ways of pursuing excellent genetics research. Then again in 2022 to join your lab once again, this time to start the journey on this very thesis. Kaarina, your unwavering support over the years has been so amazing and imperative to both my personal and career development as an MD/MSc student as I step closer to a career in medicine. From a fellow Sisler Spartan, I could not think of a better mentor to have!

To my advisory committee members, Dr. Britt Drögemöller and Dr. Ruth Ann Marrie. Both have provided valuable guidance throughout my Master's degree, from genetics, neurology, and data science. To Britt, your mentorship not only in my thesis but as well as during coursework has been invaluable to deepening my interests in human genomics, and something I'll hold dear in my career. To Ruth Ann, your work ethic and mentorship as a clinician scientist has been inspiring. This path to the MSc degree has been anything but linear and I want to thank my supervisor and committee members for your support.

I would like to thank various funding organizations throughout my research: the University of Manitoba Undergraduate Research Award, The University of Manitoba RFHS Graduate Studentship, and the CIHR-funded Advanced Degrees in Medicine Program (MD/MSc).

During grad school, I have come across some amazing people. From the College of Pharmacy graduate students, the Pharmacoepidemiology department, the amazing genetic counseling, genetic residents, and graduate students in my classes, the small but mighty Filipino graduate student community, to fellow endMS Summer School researchers. Last, to the people in

the Kowalec Lab past and present. Hayley, Urmi, Mira, Laila, Aziz, Mina, Ali, Gabrielle, Francisca, Rupinder, thank you!

Thank you to my classmates in medical school, who have been a joy to be around, and whose work ethic continues to inspire me every day. I'd also like to thank my friends for supporting me since day one! You know who you are!

To the Oklahoma City Thunder, my childhood team, for giving me pure inspiration during thesis writing by finally winning a championship the year I complete this degree. Thunder Up!

Last, to my family! I'll be the first Trojillo to complete a Master's degree, which in itself is a monumental accomplishment. But with full humility, all credit goes to my parents and brothers. From my older brother Kuya Julius, to my younger brother Joshua who are both setting the pace in life and who continue to motivate me to succeed. To my mom and dad, as immigrants starting from scratch in Canada, and who sacrificed a lot, paving the way for me and my brothers to thrive.

My time in grad school truly had its ups and downs, but I truly believe in the cliché "the journey is the reward", and I could not have done this without the help from everyone mentioned above.

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Multiple sclerosis

Multiple sclerosis (MS) is a demyelinating disease and characterized by demyelination and axonal injury in the central nervous system.[1] Damage to the myelin in nerves hinders the ability to send electrical impulses around the central nervous system, hindering its function with varying ranges of symptoms.[1] Demyelination results in gray and white matter lesions, with pathology studies suggesting that inflammation causes this effect.[2] Overall, persons with MS (PwMS) may experience chronic symptoms which can include mobility issues, vision problems, muscle weakness, fatigue, cognitive issues, and mental health issues such as anxiety and depression.[3,4]

The first episode of MS symptoms is termed clinically isolated syndrome when an individual presents to a clinic with MS symptoms, but do not meet the full criteria for a formal diagnosis.[5] A relapse is when a patient experiences new or worsening neurological deficits which can resolve or lead to worsening relapses over time.[3] MS is characterized by two primary types: relapsing-onset and progressive-onset.[3] Relapsing-onset MS can be further classified by two subtypes: relapsing-remitting MS and secondary progressive MS (SPMS), and the disease can be labelled as active or inactive.[3] Relapsing-remitting MS (RRMS) is the most common type of MS, comprising 85% of all MS cases.[3] PwMS experiences relapses which can consist of days to weeks of MS symptoms.[3] SPMS is characterized by progression of symptoms experienced in the RRMS stage (***Figure 1***), furthered intensified with irreversible symptoms and increased disability over time, showing no signs of improvement, with worsening baseline over time.[5] Prior to the introduction of disease modifying therapies, approximately 50% of people with RRMS progress to SPMS within 10 years, and approximately 95% within 25 years of diagnosis.[3] A third subtype of MS, primary progressive MS (PPMS) experience a gradual onset of MS symptoms from

diagnosis, and have symptoms identical to SPMS without exhibiting a gradual onset of MS symptoms and relapses found in patients with RRMS.[3,5]

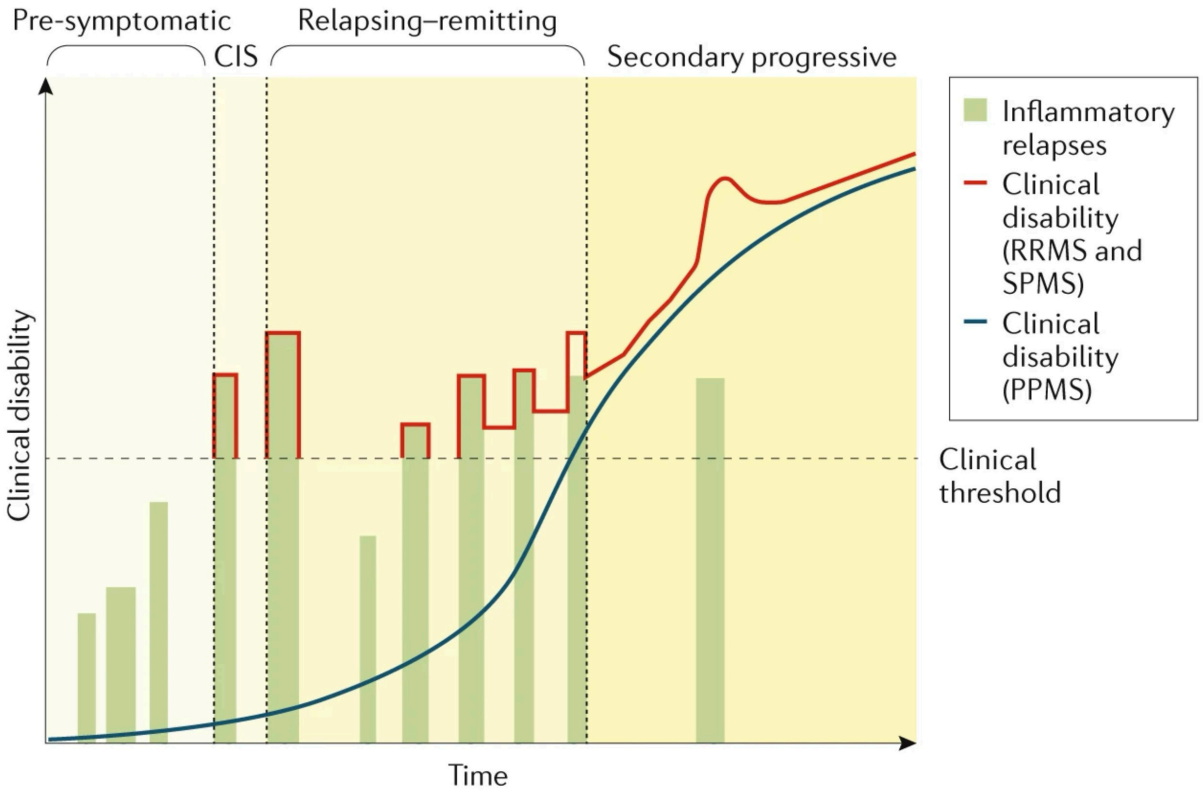

Figure 1. Schematic of the disease progression for relapsing-remitting multiple sclerosis (RRMS) and secondary progressive multiple sclerosis (SPMS).[3] Figure was reproduced with permission from CCC RightsLink.

Currently, there is no cure for MS, though disease-modifying treatments (DMTs) have been developed to reduce disease activity and disability worsening.[6] DMTs are given to those diagnosed with RRMS or clinically isolated syndrome to mitigate disease progression.[6] DMTs such as ocrelizumab and siponimod are used for treatment for PPMS and SPMS, respectively.[7] Examples of first line DMTs for RRMS include like injectible interferon beta and oral dimethyl fumarate, with second line agents such as natalizumab.[6] Third line agents include alemtuzumab and ocrelizumab.[8]

Comorbidities commonly arise in PwMS and can affect treatment and management for each patient.[9] Examples of physical comorbidities include hypertension, hyperlipidemia, and migraines, and psychiatric comorbidities include anxiety and depression.[4] When adjusting for age, sex, and socioeconomic status, these comorbidities in PwMS are associated with a 2-fold increased rate in hospitalizations (adjusted risk ratio = 2.21, 95%CI = 1.73-2.82) compared to PwMS without comorbidities, and a 4-fold increased rate in hospitalizations in PwMS with comorbidities compared the non-MS population without comorbidities (adjusted risk ratio = 3.85, 95% CI 3.40-4.35).[10] In terms of mortality, another study has observed an increase in mortality risk with physical and psychiatric comorbidities when comparing PwMS with comorbidities and PwMS without comorbidities.[10] Various chronic MS symptoms exhibited by PwMS, such as fatigue, movement disorders, and cognitive impairment, can have a long-term impact on a person's quality of life, including physical, professional, financial, and social well-being.[4] PwMS may be forced to remove themselves from work due to reduced mobility or cognitive impairment.[11] A study found that MS patients that became unemployed had worse outcomes in depressive symptoms, stemming from further disease progression that hinders their ability to engage in daily activities.[4]

1.2 Epidemiology of MS

The global prevalence of MS is 2.8 million people or 35.9 per 100,000 people, with an incidence rate of 2.1 per 100,000 people (***Figure 2***).[1] Reporting from 75 countries worldwide, 107,000 people are diagnosed with MS annually.[1] The average age of MS diagnosis is 32 years old.[1] This age of diagnosis in early adulthood differs from other neurological diseases such as dementia and Alzheimer's, where cases are found within the elderly population. For RRMS diagnosis, the average age of diagnosis was 34 years old (as of 2019), an increased age of onset compared to an average age of 23 years old in 1970.[12] MS affects all age groups, as 1.5% of people diagnosed with MS are children or adolescents.[3]

Canada has one of the highest prevalences of MS per capita worldwide, ranking fourth in MS rates with 250 MS cases per 100,000 people.[1] Canada ranks behind Germany, USA, and Denmark in prevalence of MS worldwide.[1] The prevalence of MS has increased in recent years in Canada. From 1991 to 2010, using administrative health data from British Columbia, the prevalence of MS in Canada was estimated to have increased by an average of 4.7% annually (***Figure 3***).[13] This increase in prevalence has been observed despite no significant change in the incidence of MS.[13] This study inferred that the increase in prevalence can be due to survivability rates increasing over the past decades, earlier diagnosis for MS, and immigration of prevalent MS cases.[13] A meta-analysis of 94 studies found a positive gradient between increasing latitude and prevalence of MS, observing an increase of 3.64 cases per 100,000 people per degree latitude.[14] Higher rates of MS are found in areas of temperate climates, closer to the North and South poles, such as North America, the United Kingdom, Northern Europe, Scandinavia, Australia, and New Zealand.[14] A latitude gradient is also observed within a country. An Australian study observed increased rates of MS in northern parts of the country in the states of Queensland (75 cases per 100,000 people)

compared to the southernmost (and furthest away from the Equator) state of Tasmania (139 cases per 100,000 people).[14]

Economically, MS has a large burden in North America. In Canada, individuals with MS face high annual healthcare costs; the total per capita health care costs in 2011 $16 800 CAD per patient, compared to a per capita cost of $2500 of individuals without MS.[15] Additionally, total MS care-related costs are projected to reach $2 billion CAD in 2031.[15]

Women are disproportionately more likely to be diagnosed with MS than men, making up approximately 75% of all MS cases worldwide.[16] In Canada, the sex ratio is 3.21:1 between females and males, increasing from a 1.4:1 sex ratio between females and males in 1977.[16]

Overall, the high prevalence in MS within Canada makes MS research a key topic of interest as investigating the disease can discover useful knowledge to mitigate disease severity amongst PwMS via identification of risk factors or discovering new treatments to lessen disease progression.

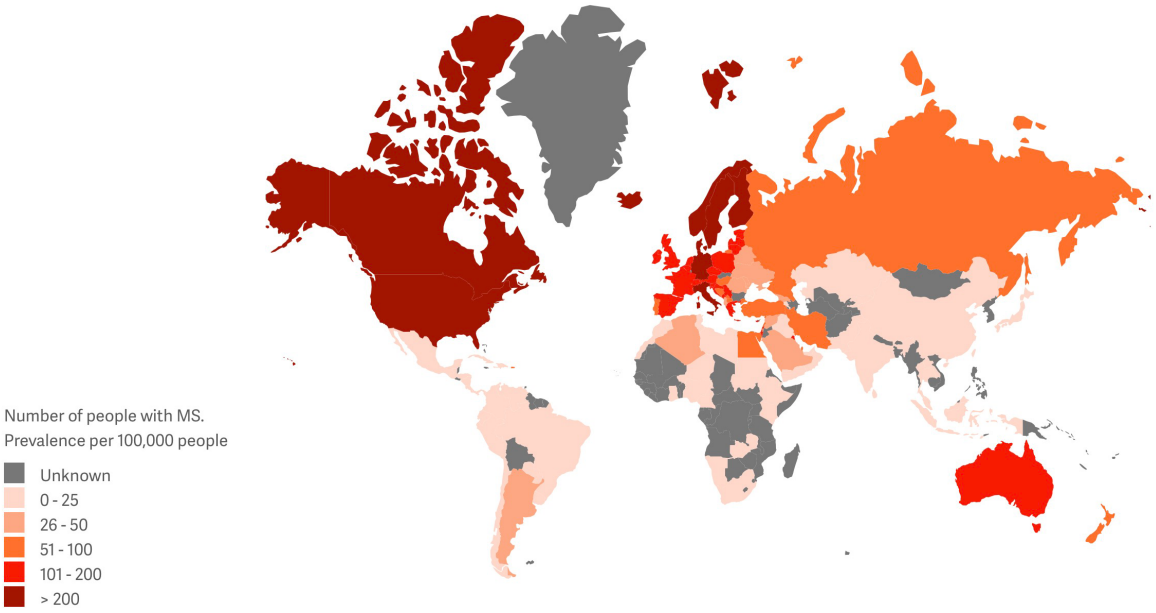

Figure 2. Map summarizing the prevalence of MS per 100,000 people.[17] Figure was reproduced with permission from CCC RightsLink.

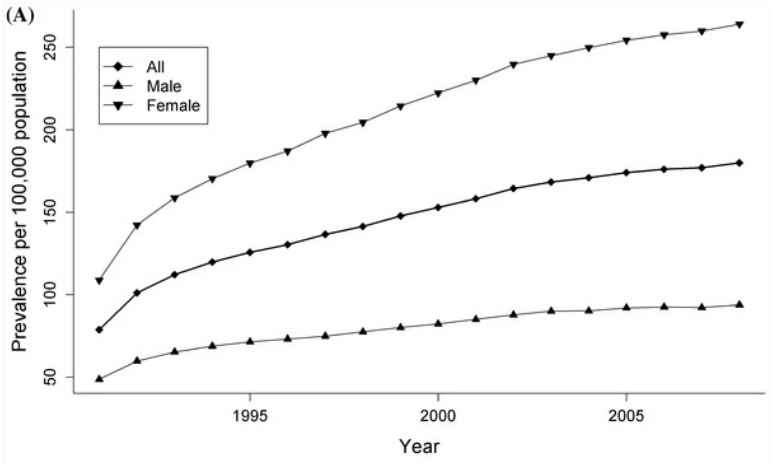

Figure 3. Age-standardized prevalence of multiple sclerosis from 1991-2008 in British Columbia, Canada.[13] Figure was reproduced from an open access article distributed under the terms of the Creative Commons CC BY license.

1.3 Genetic and environmental risk factors for MS

MS is a complex disease, triggered by genetic and environmental factors. The recurrence risk ratio (RRR) is a measure of familial aggregation of a disease, with a value greater than 1 indicating familial aggregation for a given disease. A meta-analysis study measuring the RRR for MS found that the sibling relative risk was $\lambda = 16.8$.[18] Heritability is the proportion of variance of a disease phenotype in a population that can be explained by genetic variance. A Swedish twin study including 74,757 twin pairs (with no MS) and 315 twins with MS (78 monozygotic, 237 dizygotic) computed the pedigree heritability of MS to be 0.64 (95% CI: 0.36-0.76).[19] Meaning 64% of the variability of MS can be attributed to genetic or shared environmental factors within the cohort tested. A study by Westerlind *et al.* from Sweden, using 28,396 individuals found a significant difference in the transmission of MS, with a 2.7-fold increased risk for MS from the mother to daughter ($p=2.2 \times 10^{-16}$), and a 1.65-fold increased risk of MS transmission from father to daughter ($p=0.014$).[19]

Genome-wide association studies (GWAS) investigate the association between common genetic variants and the occurrence of a disease or trait.[20] GWAS are detailed further in Section 1.6. The first GWAS for MS was completed in 2007 with a total of 12,360 participants and found 49 SNPs with an association for MS.[21] Further analysis with 1,540 parent-affected offspring trios found two variants that were located within the interleukin-2 and interleukin-7 receptors, with multiple single nucleotide polymorphisms (SNPs) in the *HLA-DR* locus.[21] Subsequent larger GWAS' of MS risk identified multiple variants within the *HLA-DRB1* loci.[22] One of the alleles, *HLA-RB1*1501*, is frequently detected in MS, though its effect in MS susceptibility varies by different studies.[22] Most recently, the largest GWAS of MS identified 233 genetic variants in 47,429 European genetic ancestry cases and 68,374 European genetic ancestry controls, and were

primarily outside the HLA region.[23] The current variance explained by common genetic variation for MS is 19.2%.[23] Currently, there are no large-scale GWAS of MS risk for populations of non-European ancestries. However, one Japanese study with 533 cases and 1,789 controls found 26 loci associated with MS, all found outside of the MHC region.[24]

Polygenic scores are an emerging field of study in MS research. Polygenic scores are used to summarise the cumulative genetic burden of a disease or trait by individual. From the latest GWAS of MS risk from the International Multiple Sclerosis Genetics Consortium,[23] a polygenic score (PGS) was calculated to assess the genetic risk of MS in an individual. PGS are further detailed in Section 1.7. The study had 47,429 cases and 68,374 controls from 15 data sets from Canada, the USA, Australia, and Europe.[23] As expected, the MS risk PGS was significantly higher in those with MS compared to the controls.[23] Additionally, relatives of PwMS have a higher PGS than controls, and siblings of PwMS had a higher PGS than the control group but a lower PGS than affected siblings.[20]

In addition to genetic risk factors for MS, there are many environmental risk factors for MS as well, including low vitamin D levels (OR = ~1.4) and low sun exposure (OR = ~2, ***Table 1)***.[25] These support the association between MS prevalence and geographic gradient, as regions further away from the equator experience lower levels of sunlight and UV exposure throughout the year compared to countries near the equator. Another environmental risk factor is smoking (OR = ~1.6).[25] A Swedish study found that among a cohort of >7500 individuals, 20.4% of all MS cases were attributable to first-hand or second-hand smoking.[26] Other environmental factors that are associated with an increased risk to MS include obesity and shift work.[25] Epstein-Barr virus (EBV) is another identifiable risk factor for MS (OR = ~3.6)[25], and in recent years has been an emerging field of MS research. An increased presence of antibodies for EBNA1, a protein that is associated

with EBV, is associated with MS status.[27] A study with >10,000,000 individuals (n=955 PwMS) in 2022 investigated the association between EBV and MS.[28] Results from the study found that there was a 32-fold increase in risk of MS for individuals in the military who became EBV-positive compared to individuals who were EBV-negative.[28]

Table 1. Summary of lifestyle and environmental risk factors for MS from the literature.[25]

Factor	Odds ratio for MS risk
Smoking	~1.6
Epstein-Barr virus serology	~3.6
Vitamin D <50 nmol/L	~1.4
Adolescent obesity/body mass index (BMI) >27	~2
Night work	~1.7
Low sun exposure	~2
Infectious mononucleosis	~2
Passive smoking	~1.3
Oral tobacco/nicotine	0.5
Alcohol	~0.6
Coffee	~0.7

Migration studies have shown that environment has a significant influence on MS risk (***Figure 4***); migrants who moved to Canada had an overall decreased risk of MS compared to people who are long-term residents in Canada.[29] Immigrant health data was investigated out of countries from Europe, South Asia, East Asia, Middle East, and Africa. and out of all immigrants, people from Iran (hazard ratio = ~3) and Lebanon (hazard ratio = ~1.5) had the highest risk for MS than immigrants originating from other countries.[29] The reason for higher MS risk is unclear but it is speculated that genetic factors such as *HLA-DRB1*01* are also associated with MS risk in the Middle Eastern population.[29] Furthermore, there is an associated increase in MS risk in people who immigrate to Canada at an earlier age or have a longer duration of stay in Canada.[29] Compared to those who immigrated at an age less or equal to 15 years old, older immigrants had a reduced risk for MS, with the age group between 46-65 with the lowest hazard ratio (0.31, 95% CI 0.24-0.81) compared to older age groups.[29] This suggests that the Canadian or a Western environment – including aspects like climate, lower vitamin D levels, pollutants, or lifestyle – have a significant influence on increasing the risk for MS.

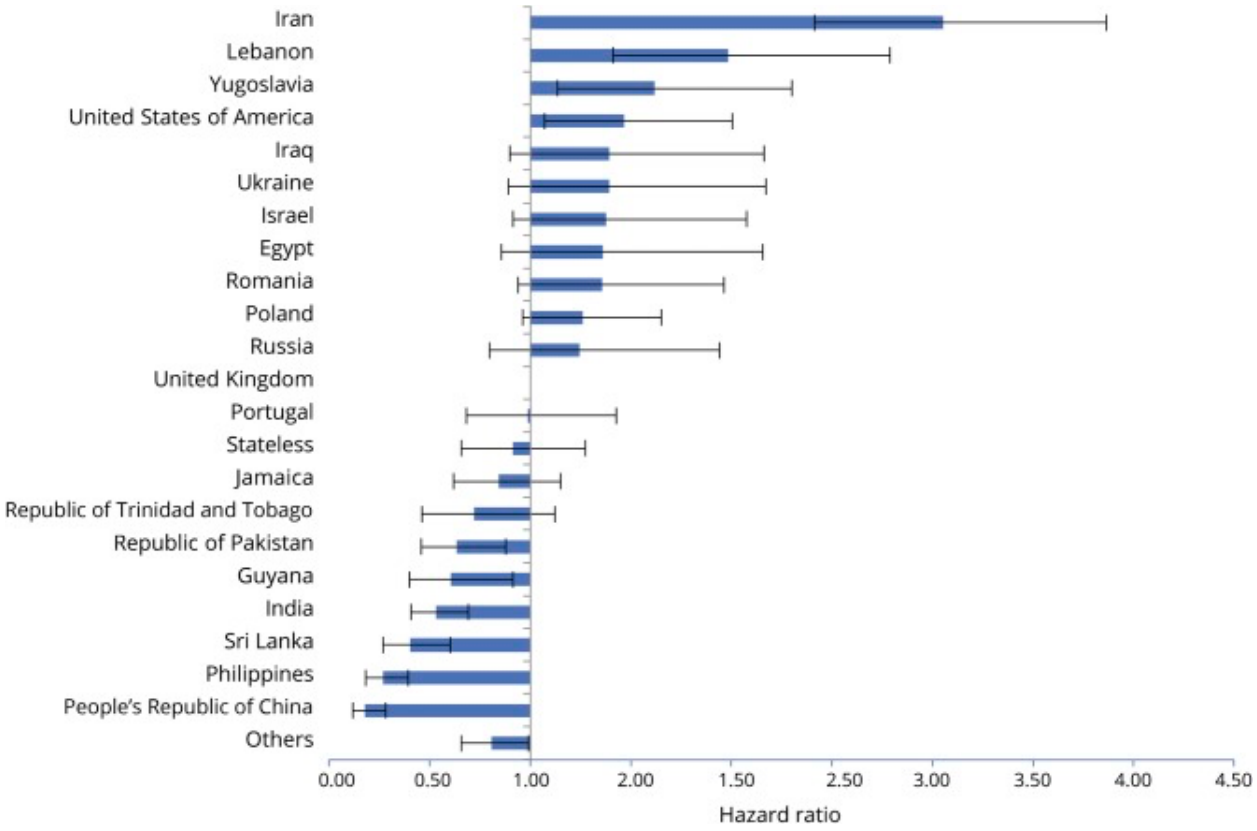

Figure 4. Risk of MS in Canadian immigrants by country of origin.[29] Figure was reproduced with permission from CCC RightsLink.

1.4 Cognitive ability in MS

Cognitive ability is defined as the brain's ability to take in, store, and retrieve information.[30] Domains of cognitive ability include memory, attention, processing speed, and decision making, among others.[31] Impairment in various different cognitive domains occurs in 40-65% of PwMS; cognitive impairment is associated with increased likelihood of affecting work performance and employment status.[31] Cognitive impairment in MS is often demonstrated as a reduction in any of the following: information processing speed, attention, impaired memory, and verbal fluency.[32] Information processing speed is a component of cognition that dictates how well and efficiently one can encode or process information such as visual or auditory information.[32] In general, grey matter and white matter pathologies contribute to cognitive impairment in PwMS, with white matter pathologies becoming more prevalent in later stages of progressive MS.[33] PwMS experiencing cognitive impairment face issues such as cognitive inefficiency and memory decline, causing those affected to take longer to complete mental tasks or hinder their ability to multitask.[31] Current treatment and management for PwMS experiencing lower cognitive ability include wellness-based symptom management such as dietary changes, aerobic exercise, avoiding multitasking, learning new material, story memory techniques, or undergoing cognitive rehabilitation, with the drug, fampridine, also demonstrating some benefit.[34–37]

Assessing cognitive ability can be done with various tests, including the Symbol Digit Modalities Test (SDMT). This test assesses information processing speed and is the most used test for processing speed in people with MS.[38] Another processing speed test is the Paced Auditory Serial Addition test (PASAT) which also assesses concentration and interference suppression.[39] Last, another information processing speed test is the Digit Symbol Substitution test (DSST) which specifically assesses processing speed by requiring participants to draw a

symbol that matches with its corresponding digit, which is similar to the SDMT, but with the question and response being inverted.[40]

Cognitive ability, like MS, is determined by both genetic and environmental factors.[41] The largest GWAS to date for cognitive ability identified >200 associated genomic loci, with a SNP heritability (h^2_{SNP}) of 0.19 (standard error = 0.01, ***Table 2***).[42] It included 269,867 European ancestry individuals across 14 cohorts from Europe, North America, and Australia.[42] The genetic loci identified were often located in brain-expressed genes and had been previously implicated with intelligence, nervous system development regulation, and central nervous system neuron differentiation.[42] For this GWAS, cognitive ability was operationalized using various cognitive tests scores from individuals, including verbal reasoning, mathematical reasoning, spatial visualization, and delayed recall to assess cognition, with each test highly correlated with one another.[42] Regarding the genetic correlation between MS and cognition, a study investigating the effects of MS-risk variants on brain structure and cognition found that the risk variant rs2283792 is associated with poorer memory.[43] Outside MS, the PGS for cognitive ability has been used in populations of people with first-time psychosis episodes, in which they found persons with a higher cognitive ability PGS were associated with a better course of symptoms and functionality from a psychotic episode compared to those with a lower cognitive ability PGS.[44] The replication of PGS in populations outside the original discovery populations is imperative to their more widespread use.

1.5 Educational attainment in MS

Educational attainment, defined as a person's highest level of education completed, is associated with a multitude of health outcomes. Higher educational attainment is associated with decreased risk of MS.[45] In PwMS, educational attainment is often used as a proxy measure for "cognitive reserve" which is thought to mitigate the impact of pathological brain changes and the risk of cognitive impairment.[46] PwMS experiencing cognitive impairment face issues such as cognitive inefficiency and memory decline.[46]

Educational attainment is also the result of genetic and environmental factors.[46] Educational attainment is a heritable trait ($h^2_{SNP} = 0.21$[47] , ***Table 2***); the most recent largest GWAS of educational attainment (measured using maximum number of years of education) included 3 million individuals and identified >3,000 associated genomic loci.[47] Higher educational attainment PGS was associated with lower relative risk to common diseases such as cardiovascular disease, type 2 diabetes, asthma, rheumatoid arthritis, and depression.[47] Results found that the polygenic index incremental R^2 value was 15.8, or that 15.8% of the educational attainment variance can be explained by the polygenic index.[47] Outside MS, the PGS for educational attainment has been used in populations of people with first time psychosis which found that similar to cognitive ability, higher educational attainment PGS were associated with a better course of symptoms, functionality, and working memory from a psychotic episode compared to those with a lower educational attainment PGS.[44] As mentioned in the section above, the replication of PGS in populations outside the original discovery populations is imperative to their more widespread use. For both cognitive ability and educational attainment, there is evidence of the two traits having a protective effect on cognition in a population of individuals with psychosis, and from a neurological disease perspective it is important to investigate whether such an association is seen in an MS population.

Table 2. Latest heritability values and GWAS results for cognitive ability and educational attainment.

	Cognitive Ability[42]	**Educational Attainment**[47]
Sample Size	269,827	3,037,499
Outcome	Various cognitive ability tests	Highest Years of Education
h^2_{SNP} value (SNP-based heritability)	0.19 (SE = 0.01)	0.21 (SE=0.007)
Number of genome-wide significant loci ($p<5 \times 10^{-8}$)	205	3,952

1.6 Genome-wide Association Studies

In a GWAS, genetic data are collected from individuals specific to a disease status like MS or a trait such as cognitive ability or educational attainment. The steps undertaken when conducting a GWAS are outlined in ***Figure 5.***[20] Samples from a control population can also be collected. Genetic data are generated using DNA extracted from saliva (or blood) and then genotyped using a GWAS microarray that captures ~700,000 to 1 million common genetic variants.[20] Linkage disequilibrium (LD) is the correlation of two alleles being linked to each other and the likelihood of two alleles being inherited together, and in a GWAS this concept explains what causal variants are derived from.[48] The resultant genotypes undergo quality control using software programs like PLINK that will allow for the filtering of variants through criteria such as identifying variants with missing genotypes, variants with a low minor allele frequency, or variants that fail the Hardy-Weinberg equilibrium test.[20] Base and target data are further analyzed to look for ambiguous, mismatched, and duplicate SNPs which are to be removed.[49] The cleaned genotype data then undergo imputation, the statistical inference of unobserved variants by leveraging known linkage disequilibrium, for up to 30-40 million genetic variants using a selected reference panel, such as the 1000 Genomes Project, the Haplotype Reference Consortium, or TOPMed. Further cleaning is applied to the imputed data to remove variants with low imputation quality or low allele frequency.

A GWAS tests the association between a set of genetic variants with its frequency in a trait. For GWAS, any genetic variants with a p-value smaller than 5×10^{-8} are generally assumed to be associated with the trait. This modified p-value threshold is much smaller than 0.05 because of the Bonferroni testing threshold that divides 0.05 by the number of common independent common genetic variants of which according to multiple studies is approximately one million genetic

variants.[20] The results of a GWAS can be visualized through a Manhattan plot, where the observed p-value (y-axis = $-\log_{10}(P)$) for every variant is plotted against the SNP's position within the genome.[20]

Current limitations for GWAS include the lack of diversity in the included participants which has led to a disproportionate number of genetic variants associated with populations predominantly of European ancestry. It was found that 79% of participants across all GWAS were of European ancestry, despite the European population comprising only 16% of the global population (***Figure 6***).[50] For example, the prediction accuracy of 17 anthropometric and blood panel traits for individuals of non-European ancestry in the UK Biobank was significantly lower than in those of European ancestry.[50] Traits included in this study included height, BMI, systolic blood pressure, and various blood panel measurements.[50] There are current efforts to increase diversity by increasing the recruitment of non-European participants into genetic studies, e.g. *H3Africa*, a consortium that leads efforts to increasing genetic studies for people of African ancestry.[35,51]

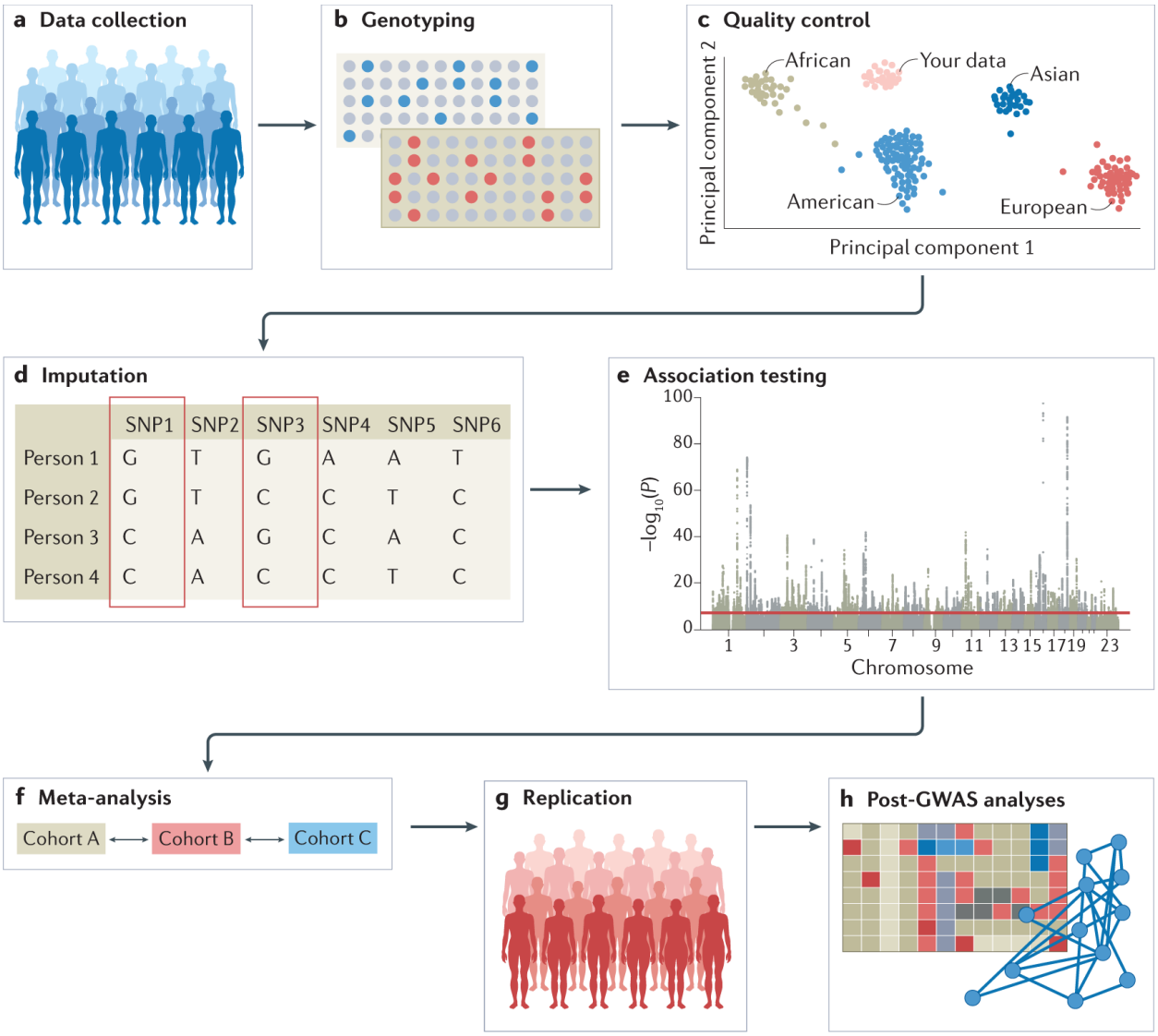

Figure 5. Overview of the steps for a genome-wide association study (GWAS).[20] Figure was reproduced with permission from CCC RightsLink.

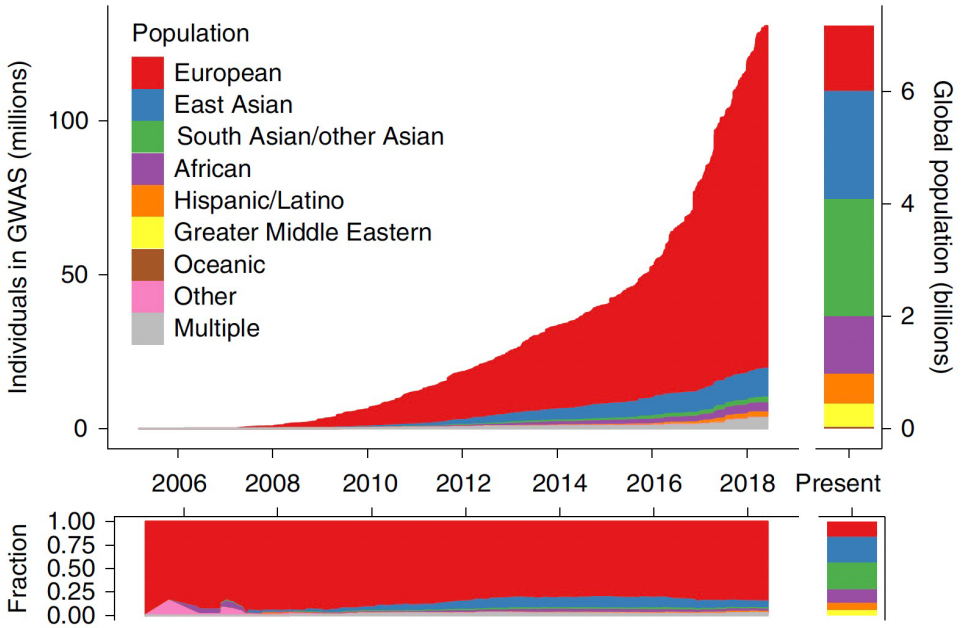

Figure 6. Genetic ancestry of participants in a GWAS over time.[50] Figure was reproduced with permission from CCC RightsLink.

1.7 Polygenic scores

Polygenic scores (PGS, also known as polygenic risk scores) of various traits consider polygenicity, the concept of multiple genes influencing a particular phenotype (for e.g., cognitive ability) to assess an individual's preponderance for a trait and are weighted by their effect sizes.[52] In general, the higher an individual's PGS score, the higher their relative risk for a disease or trait will be (***Figure 7***).[53] PGS can be used to identify those at higher risk of disease, e.g., higher MS risk PGS is associated with higher risk of MS. PGS can also be used to understand overlapping disease etiology, e.g., a study identified a role for increased hypertension polygenicity and lower cognitive ability scores, even in the absence of hypertension.[54]

PGS are generated from using 'training data' which are GWAS summary-level statistics and can be useful when investigating complex diseases like MS and traits like cognitive ability.[49] There are many methods to calculating a PGS. The basic premise of the PGS involves calculating a person's number of risk variants, weighted by the effect sizes (from the training data GWAS) in their sample, and then generating a score summing up the number of variants present. The original method was "p-value thresholding and clumping", which involved LD clumping to pool variants using LD.[55] The algorithm iteratively computes the correlation between other variants that were clumped together, and removes variants above a certain r^2 value.[56] This r^2 value is often set to a default value of 0.5.[56] The second phase to this method, p-value thresholding, involves removing variants that are smaller than a pre-selected p-value level of significance. This threshold can range from 1 (all variants) to $5x10^{-8}$. More contemporary methods for calculating PGS include *SBayesR,* which uses summary statistics from a GWAS and Bayesian multiple regression model to calculate a PGS.[57] Compared to the clumping and thresholding method, *SBayesR* outperformed in predictiveness for common traits such as BMI and height.[57]

Additionally, because PGS require a training dataset (existing GWAS), the majority of PGS developed are in European genetic ancestry individuals. As more diverse GWAS are conducted, more diverse methods (i.e., PRS-CSx) and PGS are becoming available.[58] Polygenic scores reporting guidelines are available and include reporting processes such as study type, design, study population, risk model development, risk model application, risk model evaluation, limitations, and data transparency.[59]

PGS are currently being developed for various diseases for potential use in a clinical setting, from a diagnostic and prognostic standpoint. There are many PGS developed for common multifactorial conditions such as depression, type 1 diabetes, and coronary artery disease.[60] Overall, when considering common diagnostic risk factors such as age, sex, and BMI, adding PGS for these conditions increased the area under the curve (AUC) by an average of approximately 10%.[60] However, replication in additional cohorts of individuals are needed for PGS to attain the necessary levels of clinical utility.

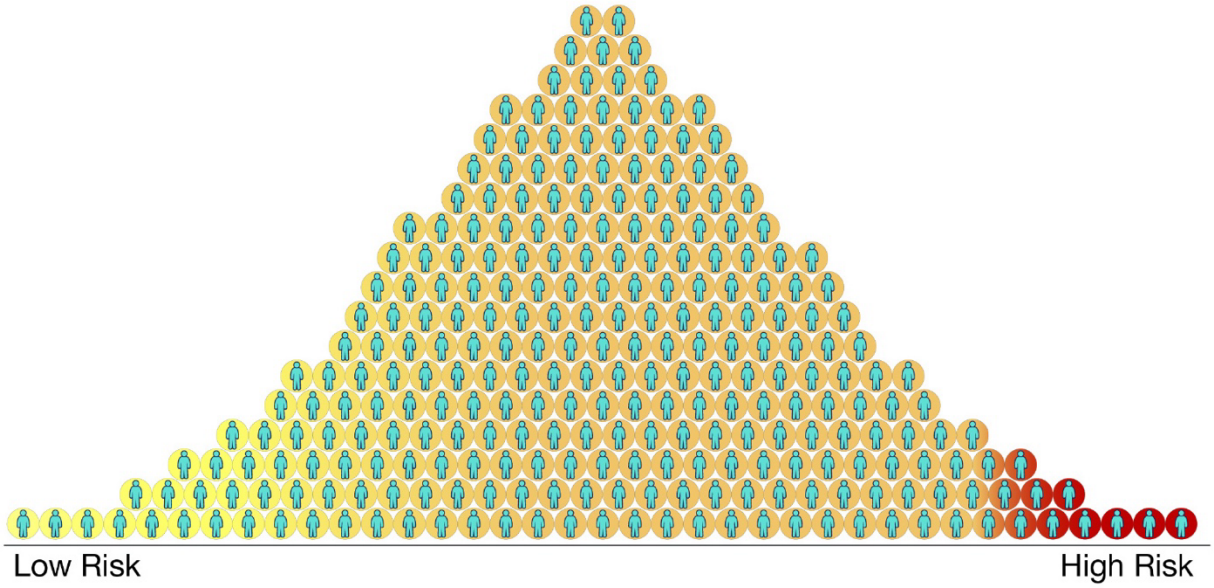

Figure 7. Visualisation of a polygenic score derived from variants discovered in a genome-wide association study, with the gradient representing an individual's polygenic score. The figure was reproduced under the Public Domain (Courtesy: National Human Genome Research Institute).

1.8 Knowledge Gap

Whether the PGS for cognitive ability or educational attainment are associated with education and cognition in PwMS is unknown. The aim for the study is to validate the use of cognitive ability PGS and educational attainment PGS specifically within PwMS. Investigating the PGS specifically within PwMS can give us insight to how this information can be used in MS either for research into the underlying MS disease processes or to improve care (as a potential clinical tool to assess the relative risk of a person's cognitive decline to potentially modify management plans) for PwMS that may be at higher risk for lower cognitive ability or cognitive reserve.[61]

The cognitive ability PGS and the educational attainment PGS utilize GWAS summary statistics from large, non-MS populations.[42,47] To ensure transferability from those studies to an MS population, I will examine the association of these two PGS with their respective traits specifically in PwMS. Additionally, educational attainment and cognitive ability are two phenotypes that can be assessed in many ways. For example, cognitive ability can be assessed using information processing speed, but this is only one domain of cognitive ability. Further, the number of education years is one measure of an individual's level of educational attainment and others could include highest level of education (instead of years). This project will intend to replicate in PwMS the association between the PGS for these two traits against information processing speed and educational attainment (years of education, education level), respectively.

1.9 Research objectives and hypotheses:

Objectives & Hypotheses

In my thesis, I will conduct analyses related to two objectives:

(1) Examine the association between cognitive ability PGS and information processing speed in PwMS using three cohorts from three countries.

Hypotheses: in PwMS, the cognitive ability PGS will be positively associated with increasing information processing speed.

(2) The second objective investigates if there is an association between educational attainment PGS and educational attainment in PwMS using 2 cohorts from 2 countries (years of education), or 3 cohorts from 3 countries (education level).

Hypothesis: in PwMS, the educational attainment PGS will be positively associated with years of education, and that educational attainment PGS will be positively associated with education level.

Chapter 2: Materials and Methods

Participants

I used data from three pre-existing studies or cohorts: (i) a cohort of MS from Manitoba, Canada ("IMID study");[61] (ii) a USA cohort of MS from the NIH-funded CombiRx study,[62] and (iii) a cohort of MS from the UK Biobank (UKB).[63]

Participants with MS from the IMID/Canadian cohort (N=247) were recruited from 2014-2016 as part of the previous CIHR-funded study focused on immune-mediated inflammatory diseases.[61] This study enrolled patients with the following diseases: inflammatory bowel disease, rheumatoid arthritis, and MS, as well as healthy controls and those with depression or anxiety and no immune diseases. For this thesis, I used only those with MS. All participants had a neurologist-confirmed diagnosis of MS.[61] For MS, the inclusion criteria for this study were that they were diagnosed with MS under the McDonald criteria.[61]

The CombiRx study is an NIH-funded, 3-arm, randomized, multi-center, Phase-III clinical trial of combination DMT for MS, which includes ~600 neurologist-diagnosed MS subjects recruited between 2005-2009 and followed for ~3 to 7 years.[62] The inclusion criteria for this study was that they: were aged between 18 and 60 years old, had a diagnosis of RRMS based on the McDonald criteria, had a EDSS score from 0 to 5.5 inclusive, had at least 2 MS exacerbations in the prior three years.[62]

The UK Biobank (UKB) enrolled and assessed 502,655 individuals aged 37-73 from the UK between 2006-2010.[63] The UKB does not include clinical diagnoses by neurologists; however, based on a previous study using the UKB, the authors indicated a greater variance explained in the MS liability by the MS risk PGS was found using a probable definition of MS,[64] which I used here. Probable PwMS were defined as those reporting ≥2 of the following: a self-reported lifetime MS diagnosis, an International Classification of Disease version-10 (ICD-10)

code for MS (ICD-10 code G35) recorded in hospital inpatient records, or self-reported lifetime MS medication (betaferon, interferon-beta, avonex, glatiramer, or copaxone). An additional definition was used for sensitivity analyses: "MS – possible" population were those who reported having ≥1 of the conditions.

Measures

Socio-demographics

For Canada, information collected at the baseline visit included: age, sex (female/male), BMI, annual household income (above $50,000 CAD, ≤$50,000 CAD, declined), smoking status smoking status (smoked ≥100 cigarettes in lifetime: yes/no), total number of years of education and highest level of education (above high school vs. high school and below).[61] Similar information was collected in the USA cohort at the baseline visit except income and years of education were not collected.[62] For UKB, similar information to Canada were captured. Specifically for the UKB, income level was divided into persons with income below or above £31,000 (approximate equivalent to $50,000 CAD) or declined.[63] Education data was collected as the person's highest education qualifications, and years of education were inferred by converting an individual's highest qualification with their corresponding International Standard Classification of Education (ISCED) level.[65] For the UKB cohort, an education level at a high school level and below was considered as a participant's highest level of education being Ordinary levels (O-levels), General Certificate of Secondary Education (GCSE), Certificate of Secondary Education (CSE), or equivalent, as noted in the participant's response in the demographic survey.[66]

Psychiatric morbidity

In the Canadian cohort, the presence/absence of lifetime anxiety or depression was captured using a validated self-report questionnaire. Lifetime anxiety or depression was also captured using the standard semi-structured interview (SCID) that can identify the presence of anxiety and mood disorders using the DSM-IV criteria.[61] In the USA cohort, anxiety or depression comorbidity were collected through the participants' medical history in addition to the participants' concomitant medication history, with each respective comorbidity being self-reported or matched with prior or current use of a medication specific to anxiety and/or depression.[62,67] In the UKB cohort, self-reported lifetime anxiety or depression were collected at the baseline visit.

Cognitive testing

Across all three cohorts, the cognitive domain of interest was processing speed. In the Canadian cohort, this was measured using the Symbol Digit Modalities Test (SDMT).[68,69] The SDMT was administered by trained research assistants at baseline visit. In the USA, the Paced Auditory Serial Addition Task (PASAT),[39] was administered during their baseline visit where they are examined by a clinician, with the PASAT being part of a larger assessment called the Multiple Sclerosis Functional Composite (MSFC).[70] In the UKB, participants completed the digit symbol substitution test (DSST).[40] For all tests, the raw scores were used in regression analyses. The SDMT and PASAT have been validated for use in MS for measuring processing speed.[69,71] The DSST was administered by the UK Biobank, and a study has shown high correlation with the SDMT.[72]

Genotyping

Canadian IMID samples were genotyped on the Illumina Global Screening Array (GSA) (version 2) at the Genome Québec Innovation Centre (McGill University). Genotypes were included if the call rate is >0.98 and the minor allele frequency is <0.01. Pseudo-autosomal region markers and markers with poor genotype clustering after manual inspection were removed. I excluded samples for having genotype missingness >0.02 (after first removing SNPs with a call rate <0.95), genotypic sex discordancy, autosomal heterozygosity >0.2 or were related (identity-by-descent >0.2). Variants were excluded with a call rate of <0.99 or deviating from Hardy-Weinberg equilibrium ($P<10^{-6}$). Samples and variants passing quality control were imputed using TOPMed as the reference panel on the Michigan Imputation Service.[73,74] Imputed data had the following filters applied: minor allele frequency (<0.01) and INFO score <0.7. USA samples were genotyped using the Illumina HumanOmni1-Quad chip array at the Translational Genomics Research Institute, Arizona, with genotype data available for self-reported white participants following the same quality control and imputation as the Canadian samples. UKB samples were genotyped using the Affymetrix UK Biobank Axiom Array, with genotype data available for central quality control and imputation. The reference panel used was the Haplotype Reference Consortium r1.1 and UK 10K Genomes.[75,76] I applied the same quality control to UK genotype data as that of the Canadian and USA samples. Data from all three cohorts were harmonized to ensure the risk alleles were coded to the same allele.

Genetic ancestry was identified using principal components analysis,[77] with patients excluded based on non-European ancestry when compared to 1000 Genomes Project (phase 3, v5) as the reference (N=2,493 unrelated individuals: 659 African, 347 Admixed, 504 East Asian, 503

Europeans, 480 South Asian).[78] Any samples that were further than 3 standard deviations (SD) from the 1000 Genomes Project European population on principal components 1 or 2 were removed. Exclusion of patients of non-European ancestry is due to the lack of sufficient transferability of European GWAS results to a non-European population.[50] Following the removal of non-European participants, the principal components were regenerated without the reference data and used as covariates in the analyses.

I computed the PGS for 2 complex traits to be used as exposures: cognitive ability and educational attainment. To derive these, summary statistics were obtained from previous GWAS'.[42,47] I acquired GWAS summary statistics without the UKB cohort included when computing the PGS in the UKB to avoid over-inflation. For the cognitive ability GWAS, the population included a total of 269,867 individuals that was used for the IMID and USA cohort.[42] For the UKB cohort, I excluded the UKB cohort from the GWAS (N = 74,214).[42] For the educational attainment GWAS, for the IMID and USA cohorts, I excluded the 23andMe population (N = 765,283). [47] For the UKB cohort, I excluded the 23andMe and UKB population from the GWAS (N = 324,162).[47] Data from all both studies were harmonized to ensure the risk alleles were coded to the same allele. Using SBayesR,[57] the PGS was calculated in the three test datasets as the sum of the variant dosages weighted by the effect from the discovery set across all variants under a Bayesian multiple regression model, applying shrinkage from mixture priors. The LD reference used was the UKBB with all other default parameters used. Scores were standardized to a mean=0, SD=1 for ease of interpretation.

Statistical analyses

Each cohort is described by their general health, socio-demographics and disease-specific characteristics using mean (SD) and N (%).

I tested the association of the cognitive ability PGS (exposure) with processing speed (outcome) and then the educational attainment PGS (exposure) with education years (outcome) using linear regression, along with adjustment for genetic ancestry principal components and other covariates that are common confounders (age, sex). For Objective 1 (cognitive ability), I additionally adjusted for education level. The USA cohort did not have years of education, but rather a categorized educational attainment, and thus, an additional (logistic) regression model was used to test the association between educational attainment PGS (exposure) and education level in all three cohorts (outcome: ≥ high school vs. < high school).

To determine the appropriate number of genetic ancestry principal components to be included as covariates in each analysis, scree plots were generated by sorting the eigenvalues for each principal component. Upon generation of the scree plots, the elbow of the scree plot designated the number of principal components to include in analyses for each respective cohort. This was performed in all three cohorts.

I computed the variance explained in the outcome by the respective PGS using either the Nagelkerke's pseudo-R^2 (using *DescTools* R package) for the logistic regression models or R^2 for linear regression models. A null model including only the genetic ancestry principal components (determined by scree plot) with the respective outcome was run, followed by running a full

model including the genetic ancestry principal components and the PGS. The R^2 difference between the null and full model for each respective regression model was calculated.

For the UKB cohort, I additionally conducted a sensitivity analysis to determine the effect of modifying the definition of PwMS to use ≥1 of self-reported lifetime MS diagnosis, MS drugs or an ICD-10 code for MS.

Findings across individual cohorts were pooled using fixed-effect inverse-variance weighted meta-analysis to account for the across study variation and quantify the variability in effect estimates due to heterogeneity (instead of chance) using the I^2 statistic (expressed as %).[79,80] All analyses were performed using *R* (v4.2.2) and RStudio (v2023.06.0+421) using the following packages: '*tidyverse'*, '*DescTools'*, '*utils'*, '*ggplot'*. A *p* value less than 0.05 was considered statistically significant.

Chapter 3: Results

4.1 Participant Characteristics

I included 2,092 total PwMS of European genetic ancestry (213 Canada, 602 USA, 1,277 UKB (***Table 3***). Across all cohorts, 1,542 were female (73.7%), with an average age of 50.1 years (SD: 8.7). Sociodemographics across all three cohorts were similar, including BMI, income, and smoking status (***Table 3***). For psychiatric comorbidities across all cohorts, 18.5% of PwMS self-reported anxiety, and 42.0% reported depression.

4.2 Cognitive testing

Overall, 988 (47.2%) people completed a cognitive test. In the IMID cohort, 213 PwMS had a mean raw SDMT score of 50.86 (SD = 12.3, ***Table 3***), with a standardized z-score of -0.664 (SD=1.10). From the 506 PwMS in CombiRx with a PASAT test, the mean raw score was 50.96 (SD = 9.5) and a standardized z-score of 0.00 (SD = 0.89, ***Table 3***). Among the 482 individuals with a DSST in the UKB cohort, the average raw DSST score was 17.6 (SD = 5.6), with an average standardized z-score of -0.32 (SD = 1.06, ***Table 3***).

4.3 Educational attainment

For Objective 2, 1,770 people reported either the number of years of education or their highest education level. Across all three cohorts, 66.1% of PwMS reported education above high school. The average years of education was 14.2 (SD = 2.65) in IMID and 15.1 (SD = 3.6) in UKB (***Table 3***). Participants in the USA cohort reported 69.8% completed education above high school, the highest percentage between the three cohorts (66.2% in IMID, 64.3% in UKB).

4.4 Regression analyses

I determined the following number of genetic ancestry principal components to use in further analysis based on the variance explained by: IMID: 3 (***Figure 8a.***), USA: 2 (***Figure 8b.***) and UKB: 3 (***Figure 8c.***). Using regression, I tested the association between the exposure (cognitive ability PGS) and the information processing speed in PwMS. In the Canada/IMID cohort, for every 1-standard deviation increase in cognitive ability PGS, the information processing speed significantly increased by 4.68 (SE: 1.79, $p = 9.4 \times 10^{-3}$, ***Table 4***). In the USA cohort, I found that for every 1-standard deviation increase in cognitive ability PGS, the information processing speed significantly increased by 2.71 (SE = 0.94, $p = 4.2 \times 10^{-3}$, ***Table 4***). In the UKB cohort, I found that for every 1-standard deviation increase in cognitive ability PGS, the information processing speed increased by 0.38 with no statistical significance (SE=0.24, p = 0.12, ***Table 4***). Pooling the three cohorts in a fixed effect meta-analysis, I found that for each 1-standard deviation increase in cognitive ability PGS, the information processing speed significantly increased by 0.59 (SE = 0.23, $p = 0.01$, I^2=78.2%, ***Table 4***). The variance explained by the cognitive ability PGS in the information processing speed ranged from 3.5-9.5% by cohort and was lowest in the USA cohort (R^2 =3.5%) and highest in the Canadian (R^2 =9.4%, ***Table 5***).

I tested the association between the educational attainment PGS and years of education for Canada and UKB cohort. I found that for every 1-standard deviation increase in educational attainment PGS, the number of years of education significantly increased by 0.86 (SE= 0.22, $p = 1.0 \times 10^{-4}$) in the IMID cohort and 0.49 (SE=0.11, $p = 1.2 \times 10^{-5}$) in the UKB (***Table 4***). Pooling the two cohorts in a fixed effect meta-analysis, for every 1-standard deviation increase in educational attainment PGS, the number of years of education significantly increased by 0.56 (SE = 0.10, $p < 0.001$, I^2 =55.8%). The variance explained by the educational attainment PGS in

the years of education was lower in the UKB cohort (R^2 =3.4%) compared to the Canada cohort (R^2 =15%, ***Table 5***).

I also used the outcome of categorized education level and performed a regression against the exposure which was the educational attainment PGS for each cohort. For every 1-standard deviation increase in educational attainment PGS, the level of education increased by 0.28 (IMID), 0.53 (USA) or 0.35 (UKB, ***Table 4***). Pooling the three cohorts in a fixed effect meta-analysis, for every 1-standard deviation increase in educational attainment PGS, the level of education increased significantly by 0.39 (SE = 0.06, p <0.001, I^2=0%). The variance explained by the educational attainment PGS in the education level ranged from 4.2-11% by cohort and was lowest in the UKB cohort (R^2 =4.2%) and highest in the Canada cohort (R^2 =11%, ***Table 5***).

4.5 Sensitivity analyses

In the UKB cohort, I included 2,098 individuals who were classified as 'MS - Possible' and 1,277 as 'MS – Probable'. In the 'MS – Possible' population, the number of people that completed the DSST was 482, whereas 289 in the 'MS Probable'. The basic characteristics (i.e., age, sex) were similar by definition (***Table 6***). In the regression (***Table 7***), each of the three outcomes (information processing speed, education years, or education level) were assessed with the respective PGS. I found that the standard errors for PGS were smaller for the 'MS Possible' cohort compared to the 'MS Probable' cohort, but there were no large changes in the effect estimates or direction of the effects (***Table 7***).

Table 3. Participant characteristics by cohort.

	Canada	**USA**	**UKB**	**Total**
N	213	602	1277	2,092
Female sex	174 (81.7)	434 (72.1)	934 (73.1)	1,542 (73.7)
Age, y	51.4 (12.6)	38.3 (9.5)	55.5 (7.5)	50.1 (8.7)
Body mass index (kg/m^2)	28.8 (7.4)	28.6 (6.6)	26.8 (5.1)	27.5 (5.8)
Ever-smoker	124 (58.2)	255 (42.4)	818 (64.1)	1,197 (57.2)
Income ($CAD)/£GBP				
<$50,000 / £31,000	68 (31.9)	N/A	628 (49.1)	696 (33.2)
≥$50,000 / £31,000	121 (56.8)	N/A	452 (35.4)	573 (27.2)
Declined	24 (11.3)	N/A	197 (15.4)	221 (10.6)
Lifetime depression	87 (40.8)	342 (56.8)	449 (35.2)	878 (42.0)
Lifetime anxiety	66 (31.0)	119 (19.8)	203 (15.9)	388 (18.5)
Years of education	14.2 (2.65)	N/A	15.1 (3.6)	15.0 (3.4)
Highest education level achieved				
High school or below	72 (33.8)	86 (14.3)	230 (18.0)	388 (18.5)
Above high school	141 (66.2)	420 (69.8)	821 (64.3)	1,382 (66.0)
Cognitive testing				
Information processing speed test performed	213 (100)	506 (84.1)	269 (21.1)	988 (47.2)
Raw information processing speed[#]	50.8 (12.3)	50.9 (9.5)	17.6 (5.6)	
Standardized information processing speed[#]	-0.664 (1.10)	0.00 (0.89)	-0.41 (1.05)	N/A
Polygenic scores				
Educational attainment	-0.06 (1.03)	0.03 (1.0)	0.00 (0.19)	N/A
Cognitive ability	0.00 (0.45)	0.00 (0.48)	0.00 ($2.2x10^{-8}$)	N/A
Categorical values presented as n (%), continuous as mean (SD).				

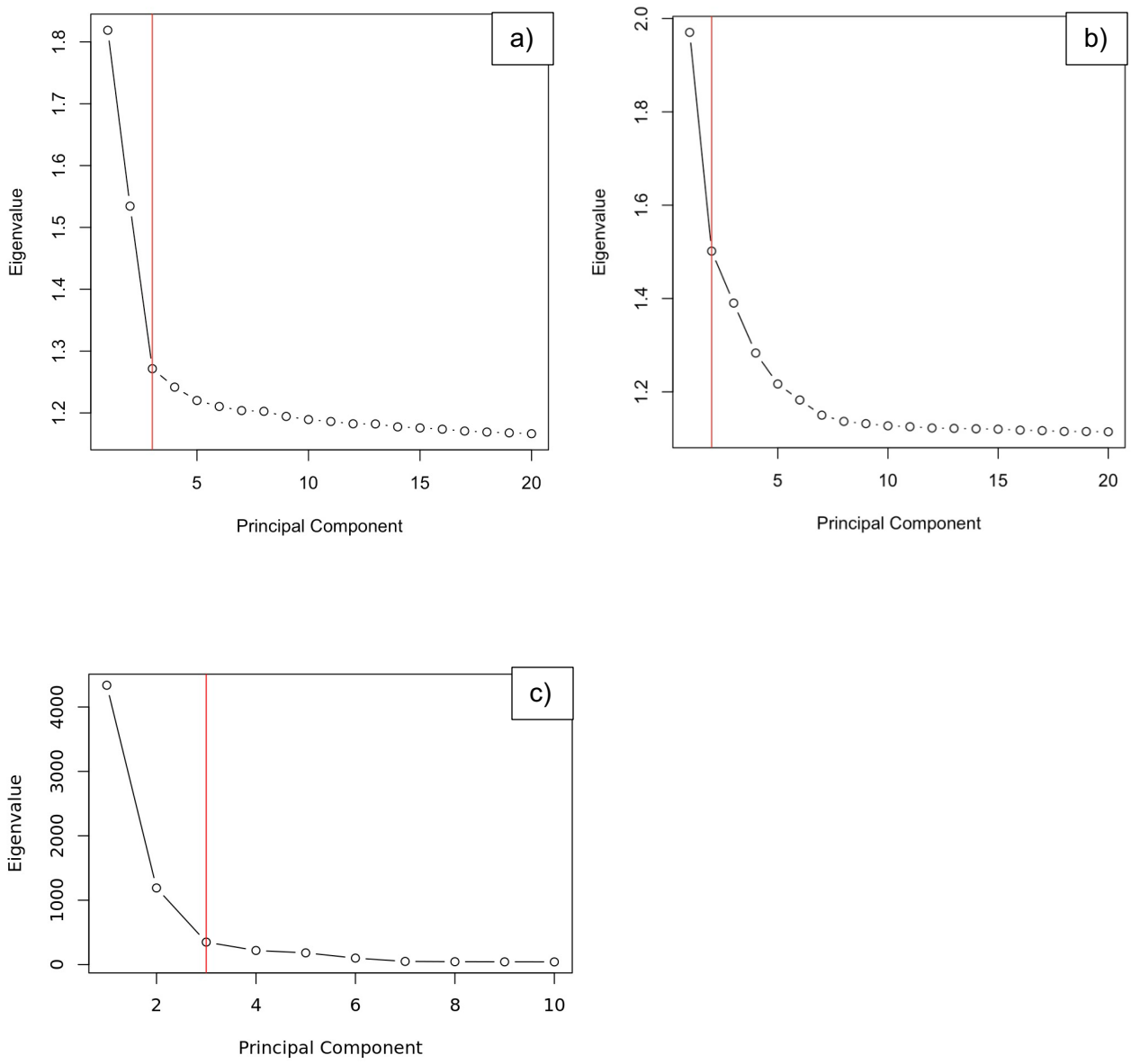

Figure 8. Scree Plot showing the proportion of variance explained by each principal components by cohort: (a.) IMID, (b.) USA, and (c.) UKB cohort. Each respective vertical line on each graph denotes the number of retained for each respective analysis.

Table 4: Regression results investigating the association between cognitive ability or educational attainment PGS and various outcomes in three cohorts of PwMS.

Exposure	**Outcome**	**Canada**			**USA**			**UKB**			**Pooled**		
		Beta	**SE**	***p***	**Beta**	**SE**	***p***	**Beta**	**SE**	***p***	**Beta**	**SE**	***p* (I^2)**
Cognitive Ability PGS	Information processing speed	4.68	1.79	9.4×10^{-3}	2.71	0.94	4.2×10^{-3}	0.56	0.34	0.1	0.59	0.23	0.01 (78.2%)
Educational attainment PGS	Years of education	0.86	0.22	1.0×10^{-4}	N/A			0.49	0.11	1.2×10^{-5}	0.56	0.10	<0.001 (55.8%)
Educational attainment PGS	Education level	0.28	0.19	0.14	0.53	0.13	4.7×10^{-5}	0.35	0.08	5×10^{-4}	0.39	0.06	<0.001 (0.0%)

PGS is standardized (to mean=0, standard deviation=1). Raw scores were derived from SDMT (Canada), PASAT (USA), and DSST (UKB). Covariates included in all models: age, sex, principal components and additionally for cognitive test score: education level.

Table 5. Variance explained (R^2) in the various outcomes by polygenic score for each cohort of people with MS.

Exposure	Outcome	Cohort			Published R^2
		Canada	USA	UKB	
Cognitive ability PGS	Information processing speed (continuous)	0.094	0.035	0.062	0.052 (ref: Savage et al., 2018, n = 269,827)
Educational Attainment PGS	Years of education (continuous)	0.150	N/A	0.034	0.158 (ref: Okbay et al., 2022, n = 3,037,499)
	Education level (categorical)	0.110	0.081	0.042	0.078 (ref: Okbay et al., 2022, n = 3,037,499)

For education level, Nagelkerke's pseudo-R^2 was used. For education years and processing speed, the R^2 was used.

Table 6. Demographic information for the UKB cohort, by the definition of MS.

	MS Possible (>1 condition)	**MS Probable (>2 conditions)**
N	2,098	1,277
Female sex	1,501 (71.5)	9,34 (73.1)
Age, y	55.4 (7.6)	55.5 (7.5)
Body mass index (kg/m^2)	27.0 (5.1)	26.8 (5.1)
Ever-smoker	1,329 (63.3)	818 (64.1)
Income ($CAD)/£GBP <$50,000 / £31,000	979 (46.7)	628 (49.1)
≥$50,000 / £31,000	801 (38.2)	452 (35.4)
Declined	318 (15.2)	197 (15.4)
Lifetime depression	754 (35.9)	449 (35.2)
Lifetime anxiety	383 (18.2)	203 (15.9)
Years of education	15.2 (3.6)	15.1 (3.6)
Highest education achieved ^		
High school or below	373 (17.8)	230 (18.0)
Above high school	1,359 (64.8)	821 (64.3)
Cognitive testing		
Information processing speed test performed	482 (22.9)	269 (21.1)
Raw information processing speed	17.9 (5.5)	17.6 (5.6)
Standardized information processing speed[#]	-0.32 (1.06)	-0.41 (1.05)
Polygenic scores		
Educational attainment	-7.5×10^{-3} (0.19)	-9.8×10^{-3} (0.19)
Cognitive ability	1.1×10^{-9} (2.2×10^{-8})	1.6×10^{-9} (2.3×10^{-8})

Categorical values presented as n (%), continuous as mean (SD). Conditions: (1) Self-reported lifetime MS diagnosis, (2) International Classification of Disease version-10 (ICD-10) code for MS (ICD-10 code: G35) recorded in hospital inpatient records, or (3) Self-reported lifetime MS medication (betaferon, interferon-beta, avonex, glatiramer, or copaxone).

Table 7. Sensitivity analyses varying the definition of MS to 'Possible MS' in the UK Biobank

Exposure	Outcome	UKB definition of MS	Canada			USA			UKB			Pooled		
			Beta	SE	*p*	Beta	SE	*p*	Beta	SE	*p*	Beta	SE	*p*
Cognitive Ability PGS	Information processing speed	Possible	4.68	1.79	$9.4x10^{-3}$	2.71	0.94	$4.2x10^{-3}$	0.38	0.24	0.12	0.59	0.23	0.01
		Probable	4.68	1.79	$9.4x10^{-3}$	2.71	0.94	$4.2x10^{-3}$	0.56	0.34	0.1	0.93	0.31	0.003
Educational attainment PGS	Years of education	Possible	0.86	0.22	$1.0x10^{-4}$	N/A	N/A	N/A	0.47	0.09	$9.8x10^{-8}$	0.53	0.08	<0.001
	Years of education	Probable	0.86	0.22	$1.0x10^{-4}$	N/A	N/A	N/A	0.49	0.11	$1.2x10^{-5}$	0.56	0.10	<0.001
	Education level	Possible	0.28	0.19	0.14	0.53	0.13	$4.7x10^{-5}$	0.31	0.06	$3.4x10^{-7}$	0.34	0.05	<0.001
	Education level	Probable	0.28	0.19	0.14	0.53	0.13	$4.7x10^{-5}$	0.35	0.08	$5x10^{-4}$	0.39	0.06	<0.001

Possible MS: n=482 for the cognitive test outcome, n=2,098 for the education outcomes. Probable MS: n=269 for the cognitive test outcome, n=1,277 for the education outcomes. PGS are standardized (to mean=0, standard deviation=1). SE=standard error. Covariates included in all models: age, sex, principal components and additionally for cognitive test score: education level.

Chapter 4: Discussion, Conclusions and Future Considerations

For this thesis, the overall goal was to validate the cognitive ability PGS and educational attainment PGS in PwMS, as these traits from a PGS standpoint were previously not studied in the MS population. Through PGS calculation and regression models, I was able to investigate the validity of cognitive ability PGS and educational attainment PGS against the participant's information processing speed and educational attainment, respectively. More specifically, objective 1, my hypothesis was that in PwMS, the cognitive ability PGS will be positively associated with increasing information processing speed. Based on the pooled results, the findings are consistent with my proposed hypothesis for Objective 1. For Objective 2, my hypothesis was that in PwMS, the educational attainment PGS will be positively associated with increasing years of education or education level. Based on the pooled results, my findings are consistent with my proposed hypothesis for Objective 2.

To the best of my knowledge, the thesis objectives were the first of its kind, both for cognitive ability and educational attainment specifically in PwMS. The findings in this study can provide a genetic basis for cognitive ability and educational attainment in PwMS. In addition to past findings found in the non-MS population, discovering the associations found in the study in an MS-specific population provides a rationale for future use of these PGS in an MS setting, in contrast to using a PGS derived from the general population. There is evidence that in a Chinese study that educational attainment is associated with a higher cognitive reserve, one's ability to withstand brain insults that can impact cognition, in PwMS.[81] For cognitive ability, these findings show evidence that we can compute a cognitive ability PGS in PwMS to assess in future studies which interventions to use to mitigate cognitive decline in PwMS during their disease

course. For educational attainment, this trait can be used as a proxy for cognitive reserve, and the educational attainment PGS can be used in PwMS to predict various cognitive outcomes during their disease course.

For the cognitive ability PGS, the variance explained in the information processing speed for the three cohorts were similar to that of published estimates (R^2=0.052).[42] This can mean that the models performed in all three MS cohorts similarly to the reference non-MS population. The original GWAS for cognitive ability used test results for various cognitive domains such as verbal reasoning, mathematical reasoning, and processing speed tests.[42] Out of the entire population included in the cognitive ability GWAS, only three cohorts used information processing speed tests, which consisted of 2.71% of the population.

In this study, each cohort used a different cognitive test to measure information processing speed: SDMT, PASAT, and the DSST, which have a differences in how they are administered, even though studies have shown high correlation between the SDMT and PASAT.[82] For the SDMT and the DSST, the two tests assess informational processing speed similarly, with the SDMT having participants draw out a corresponding matching symbol from a given number, and the DSST being the inverse by matching a number from a corresponding symbol. Between the DSST and the PASAT, a correlation study showed that both tests showed similar correlation scores against multiple domains of cognitive ability such as memory and attention.[71]

For educational attainment PGS, the variance explained for years of education in the IMID cohorts matched that of the published estimate. However, the UKB cohort was lower than the reference value, possibly due to a smaller sample size for training data for PGS calculation

for the UKB. The training data for the educational attainment GWAS included only 324,162 for generating the PGS in the UKB, whereas the sample size for the IMID cohort was 766,345 participants.[47] For the original education attainment GWAS, they used ISCED standards to measure years of education, which is consistent with the number of years of education calculated for the UKB cohort.

For the analyses on education level, there were differences in how education was measured. The Canada cohort recorded years of education, and the highest level of education attained. In the UKB cohort, the highest level of education was provided as one of six options (CSEs, O levels, A levels, NVQ, University, other professional). I used the ISCED to convert the UKB recorded highest education level to a participant's years of education. Despite the translation of highest education level to years of education, the UKB education data only includes up to a college or university degree as the highest educational level. However, despite not having the exact years of education achieved in the UKB, the beta estimates for the educational attainment PGS were similar between the Canada and UKB cohorts, albeit with higher cohort heterogeneity in the pooled results. In the USA cohort, only whether an individual finished their education above or below high school was available.

Compared to the Canada and USA cohort, the UKB had education data that follows the UK education system, which differs compared to Canada and the USA. For example, Canada and the USA considers secondary school to completed after grade 12 or at age 17/18. In the UK, the high school diploma equivalent is considered as compulsory education up until age 16 to earn a General Certificate of Secondary Education (GCSE), which is one less year of education than in Canada and the USA. This can influence the regression results, especially when analysing the

association between educational attainment PGS and education level, where the UKB cohort would have inherently less years of education than the IMID and USA cohort.

Considerations were made when determining the criteria to use to identify participants in the UKB with MS given that no MS diagnostic data is available. Two possible definitions were used based on a combination of self-reported diagnosis or MS medication or a hospital discharge code for MS. This classification was derived from a previous study that investigated autoimmune diseases in the UKB, where they indicated a greater variance explained in the MS liability by the MS risk PGS in those with a probable definition of MS,[64] compared to those meeting the possible definition of MS. However, as this study also included pharmaceutical drugs in their list of MS drug therapies that are not specific for MS (i.e., methotrexate and prednisolone), these were removed in my study, and I added a secondary definition of MS to increase the sample size. Adding the requirement of two criteria likely increased our specificity, making this population closer to the neurologist-diagnosed MS from the IMID and USA cohorts.

Relevance

The relevance of this thesis was to provide evidence for the application of PGS in MS. Previous studies for cognitive ability PGS and educational attainment PGS were completed in non-MS populations, and it is important that the PGS for these traits are found to be consistent in an MS population. This is due to the fact that compared to the general population, the MS population have a different clinical profile, especially when considering the differences in cognitive changes in the MS population over the disease course. This study provides insight on the two PGS in three cohorts of European genetic ancestry. It is important that these PGS be

validated in specific populations before they can be used as a clinical tool to influence management and prognosis or to understand disease etiological mechanisms. By investigating the association between the PGS and cognitive test scores, this can give insight for clinicians on the interaction between a person's genetics and their cognitive ability. The same can be said about educational attainment; by using years of education and education level as a proxy measure for cognitive reserve, and clinicians can use the information found in this study and consider a patient's level of education or education level and its impact on their cognitive function over their disease course, to similar to cognitive ability PGS be able to proactively introduce interventions to mitigate the effects of MS on cognitive decline.

Strengths

I employed a large sample size, with >2,000 individuals with MS and from multiple countries. Inclusion of the UKB cohort provided a significant increase in sample size, albeit without neurologist-defined MS cases, and fewer participants with information processing speed measured. Use of these cohorts to investigate the cognitive ability PGS and educational attainment PGS were applied to PwMS of European descent. Our conclusions are supported by the replication of our findings in each cohort, and consistent findings in each cohort. I also used the latest method for PGS computation, SBayesR, which has been shown to improve prediction R^2 compared to alternative methods of PGS computation such as p-value clumping and thresholding.[57]

Weaknesses

Although the thesis had many strengths, it was not without its limitations. First, the measure of education differed across all cohorts, with the USA cohort lacking the number of

years of education. Though I tested years of education with the IMID and UKB cohorts, I was limited to testing education level (above high school or high school and below) using all three cohorts. A limitation of the study is that the UKB population had a "healthy volunteer" bias, where people that enter the study are more likely to have healthier lifestyle choices and have lower all-cause symptoms.[83] This limitation was managed by having multiple filters to get as close to an MS diagnosis as possible, but this limitation can have an effect on educational attainment and cognitive ability. Additionally, the lack of a confirmed neurologist-diagnosis of MS for the patient population in the UKB is another limitation. However, I tried to counteract this by using multiple definitions of MS including a combination of whether the patient reported their condition, were diagnosed with MS in inpatient charting, or have used or are currently taking MS medications. In the analysis, I used age and sex as covariates for adjustment in the regression models. Further analysis can be done to explore other covariates to use in the regression model such as smoking status or income level (IMID and UKB only). Though there was a higher sample size by incorporating three cohorts, all three cohorts were restricted to those of European ancestry, meaning that these results can only be generalized to the European population, and not to people of non-European ancestry.

Future directions

One of the main limitations of the study is that our studied population was restricted to persons of exclusively European ancestry, which can limit the transference of our study in non-European populations. As genomic methods improve over time to improve predictiveness in non-European populations, this study can be replicated in cohorts that either contain non-European samples, or in cohorts that have a heterogeneous population, where trans-ethnic

polygenic scoring methods can be implemented. Furthermore, with improved methods in the future, we can potentially include persons of non-European ancestry from each cohort to find similarities or differences in the regression results.

In this study, I explored the genetic association of cognitive ability and educational attainment in PwMS. Further studies in this domain can be expanded to explore cognitive dysfunction. On a genetic basis, further studies can validate the PGS by observing if PwMS with cognitive dysfunction have a lower cognitive ability PGS and educational attainment PGS, respectively. This could be explored cross-sectionally or longitudinally, exploring an individual's cognitive function over time and see if there is an association with an increased cognitive decline with an increased cognitive ability PGS and educational attainment PGS, respectively.

This study can be moved further by reporting other values after the beta value and the standard error. For example, the findings in the study can be expressed as an odds ratio. Additionally, other parameters such as positive predictive value and negative predictive value can be calculated as a way to give additional ways to interpret the PGS for either trait, or to interpret any associations that were investigated in the study.

In this study, each cohort had a different cognitive test which tested cognitive ability in different ways, such as through visual or auditory means. Should a study be replicated with a different set of cohorts, it would be ideal to have cohorts that conducted the same cognitive test, or to have cohorts that completed multiple types of cognitive tests. Having multiple tests available can provide a more comprehensive view of someone's cognitive ability and provide the ability to compare the cognitive ability PGS against the raw cognitive test score for each respective test. Each cognitive test assesses a certain part of cognition, so having multiple tests can allow us to thoroughly investigate cognition in PwMS.

Conclusion

In conclusion, this study explored the intersection between genetics, education, cognition, and MS in cohorts from three different countries. I also investigated the genetic predisposition of two cognitive-related traits in PwMS using PGS to validate the associations seen in a non-MS population. Using a cognitive ability PGS and educational attainment PGS in PwMS, I was able to consider multiple facets of cognition. To the best of my knowledge, this is the first study done that explores these traits in the MS population.

Though PGS are not broadly used in clinics currently, this study did demonstrate, that in PwMS, both PGS were associated with their intended outcomes similarly to that from non-MS populations. Further studies in the MS population, especially in those of other ancestries, will assist with applying knowledge from this study, with studies to utilize these scores in connection with identifying those with poor cognitive outcomes in PwMS during their disease course. As precision medicine evolves into the mainstream and through increased used in clinic, this study lays the foundation for the use of PGS, originally developed in non-MS populations, to those with MS.

Bibliography

1. *Atlas of MS*. (The Multiple Sclerosis International Federation, 2020).
2. Lassmann, H. Multiple Sclerosis Pathology. Cold Spring Harb Perspect Med. *Mar* **1;8(3):a028936**, (2018).
3. Filippi, M., Bar-Or, A. & Piehl, F. Multiple sclerosis. *Nat Rev Primer* **4, 43**, (2018).
4. Genov, K. & Dimitrova, M. Correlation between cognitive abilities and social functioning in patients with multiple sclerosis. *Folia Med. (Plovdiv)* **64**, 33–40 (2022).
5. Lublin, F. D., Reingold, S. C. & Cohen, J. A. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* **83**, 278–286 (2014).
6. Review of Guidelines on Second-Line Therapy for Patients With Relapsing-Remitting Multiple Sclerosis: A 2024 Update: Rapid Review [Internet. in *Ottawa (ON): Canadian Agency for Drugs and Technologies in Health* (2024).
7. Hauser, S. L. & Cree, B. A. C. Treatment of Multiple Sclerosis: A Review. *Am. J. Med.* **133**, 1380-1390.e2 (2020).
8. Yamout, B. *et al.* Consensus recommendations for diagnosis and treatment of Multiple Sclerosis: 2023 revision of the MENACTRIMS guidelines. *Mult. Scler. Relat. Disord.* **83**, 105435 (2024).
9. Marrie, R. A. *et al. A Systematic Review of the Incidence and Prevalence of Comorbidity in Multiple Sclerosis: Overview. Multiple Sclerosis*. vol. 21 (Houndmills, Basingstoke, England, 2015).
10. Marrie, R. A. *et al.* Effect of comorbidity on mortality in multiple sclerosis. *Neurology* **85**, 240–247 (2015).

11. Jensen, A. R. The g factor: psychometrics and biology. *Novartis Found. Symp.* **233**, 37–121 (2000).

12. Romero-Pinel, L. *et al.* The age at onset of relapsing-remitting multiple sclerosis has increased over the last five decades. *Mult. Scler. Relat. Disord.* **68**, 104103 (2022).

13. Kingwell, E., Zhu, F. & Marrie, R. A. High incidence and increasing prevalence of multiple sclerosis in British Columbia, Canada: findings from over two decades (1991–2010. *J Neurol* **262**, 2352–2363 (2015).

14. Simpson, S., Jr *et al.* Latitude continues to be significantly associated with the prevalence of multiple sclerosis: an updated meta-analysis. *J. Neurol. Neurosurg. Psychiatry* **90**, 1193–1200 (2019).

15. Khakban, A. *et al.* Direct Health Care Costs Associated With Multiple Sclerosis: A Population-Based Cohort Study in British. *Neurology* **100**, 899–910 (2023).

16. Orton, S.-M. *et al.* Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol.* **5**, 932–936 (2006).

17. Walton, C. *et al.* Rising prevalence of multiple sclerosis worldwide. *Insights Atlas MS Third Ed. Mult. Scler.* **26**, 1816–1821 (2020).

18. O'Gorman, C., Lin, R., Stankovich, J. & Broadley, S. A. Modelling genetic susceptibility to multiple sclerosis with family data. *Neuroepidemiology* **40**, 1–12 (2013).

19. Westerlind, H. *et al.* Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. *Brain J. Neurol.* **137**, 770–778 (2014).

20. Uffelmann, E., Huang, Q. Q. & Munung, N. S. Genome-wide association studies. *Nat Rev Methods Primer* **1, 59**, (2021).

21. Consortium, I. M. S. G. *et al.* Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* **357**, 851–862 (2007).

22. Alcina, A. *et al.* Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. *PLoS One* **7**, (2012).

23. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **27;365(6460):eaav7188**, (2019).

24. Kira, J., Matsushita, T., Sato, S. & Yamamoto, K. A genome-wide association study (GWAS) in the Japanese population reveals novel genetic risk factors for multiple sclerosis and neuromyelitis optica. *J. Neurol. Sci.* **357**, 308–308 (2015).

25. Alfredsson, L. & Olsson, T. Lifestyle and environmental factors in multiple sclerosis. *Cold Spring Harb. Perspect. Biol.* **9**, 028944 (2019).

26. Hedström, A. K., Olsson, T. & Alfredsson, L. *Smoking is a major preventable risk factor for multiple sclerosis. Multiple sclerosis*. vol. 22 (Houndmills, Basingstoke, England, 2016).

27. Soldan, S. S. & Lieberman, P. M. Epstein–Barr virus and multiple sclerosis. *Nat Rev Microbiol* **21**, 51–64 (2023).

28. Bjornevik, K. *et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science. *N.Y* **375**, 296–301 (2022).

29. Rotstein, D. L. *et al.* MS risk in immigrants in the McDonald era: A population-based study in Ontario, Canada. *Neurology* **93**, 2203–2215 (2019).

30. Khera, T. & Rangasamy, V. Cognition and Pain: A Review. *Front Psychol* **12**, (2021).

31. Sumowski, J. F. *et al.* Cognition in multiple sclerosis. *Neurology* **90**, 278–288 (2018).

32. Ruano, L. *et al.* Age and disability drive cognitive impairment in multiple sclerosis across disease subtypes. *Mult Scler* **Aug;23(9):1258-1267**, (2017).

33. DeLuca, G. C., Yates, R. L., Beale, H. & Morrow, S. A. Cognitive impairment in multiple sclerosis: clinical, radiologic and pathologic insights. *Brain Pathol* **25**, 79–98 (2015).

34. Spain, R. Approach to Symptom Management in Multiple Sclerosis With a Focus on Wellness. vol. 28 1052–1082 (2022).

35. Feinstein, A. The CogEx trial – Cognitive rehabilitation and aerobic exercise for cognitive impairment in people with progressive multiple sclerosis: A randomised, blinded, sham-controlled trial. *Mult. Scler. J.* **29**, 1523–1525 (2023).

36. Chiaravalloti, N. D., Moore, N. B. & DeLuca, J. The efficacy of the modified Story Memory Technique in progressive MS. *Mult. Scler. J.* **26**, 354–362 (2020).

37. Papassotiropoulos, A. *et al.* The effect of fampridine on working memory: a randomized controlled trial based on a genome-guided repurposing approach. *Mol. Psychiatry* **30**, 2085–2094 (2025).

38. Costa, S. L., Genova, H. M., DeLuca, J. & Chiaravalloti, N. D. Information processing speed in multiple sclerosis: Past, present, and future. *Mult. Scler. J.* **23**, 772–789 (2017).

39. Diehr, M. C., Heaton, R. K., Miller, W. & Grant, I. The Paced Auditory Serial Addition Task (Pasat): Norms For Age, Education, And Ethnicity. *Assessment* **5**, 375–387 (1998).

40. Jaeger, J. Digit Symbol Substitution Test: The Case for Sensitivity Over Specificity in Neuropsychological Testing. *J Clin Psychopharmacol* **Oct;38(5):513-519**, (2018).

41. Whalley, L. J., Deary, I. J., Appleton, C. L. & Starr, J. M. Cognitive reserve and the neurobiology of cognitive aging. *Ageing Res Rev* **3**, 369–382 (2004).

42. Savage, J. E., Jansen, P. R. & Stringer, S. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat* **Genet50**, 912–919 (2018).

43. Ikram, M. A. *et al.* Genetic susceptibility to multiple sclerosis: Brain structure and cognitive function in the general population. *Mult. Scler. J.* **23**, 1697–1706 (2017).

44. Segura, A. G. *et al.* Link between cognitive polygenic risk scores and clinical progression after a first-psychotic episode. *Psychol. Med.* **53**, 4634–4647 (2023).

45. Bjørnevik, K. *et al.* Level of education and multiple sclerosis risk after adjustment for known risk factors: The EnvIMS study. *Mult. Scler. J.* **22**, 104–111 (2015).

46. Sumowski, J. F. & Leavitt, V. M. *Cognitive reserve in multiple sclerosis. Multiple sclerosis*. vol. 19 (Houndmills, Basingstoke, England, 2013).

47. Okbay, A., Wu, Y. & Wang, N. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet* **54**, 437–449 (2022).

48. Slatkin, M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* **9**, 477–485 (2008).

49. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).

50. Martin, A. R., Kanai, M. & Kamatani, Y. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584–591 (2019).

51. Mulder, N. J. *et al.* Development of Bioinformatics Infrastructure for Genomics Research. *New Age Genet. Genomics Res. Afr.* **12**, 91–98 (2017).

52. Wray, N. R. *et al.* From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry* **78**, 101–109 (2021).

53. National Human Genome Research Institute. Polygenic Risk Scores. genome.gov.

54. Rivier, C. A. *et al.* Polygenic Susceptibility to Hypertension and Cognitive Performance in Middle-aged Persons Without Stroke or Dementia. *Neurology* **101**, e512–e521 (2023).

55. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

56. Privé, F., Vilhjálmsson, B. J., Aschard, H. & Blum, M. G. B. Making the Most of Clumping and Thresholding for Polygenic Scores. *Am. J. Hum. Genet.* **105**, 1213–1221 (2019).

57. Lloyd-Jones, L. R., Zeng, J. & Sidorenko, J. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun* **10**, 5086 (2019).

58. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).

59. Wand, H., Lambert, S. A. & Tamburro, C. Improving reporting standards for polygenic scores in risk prediction studies. *Nature* **591**, 211–219 (2021).

60. Koch, S., Schmidtke, J., Krawczak, M. & Caliebe, A. Clinical utility of polygenic risk scores: a critical 2023 appraisal. *J. Community Genet.* **14**, 471–487 (2023).

61. Marrie, R. A., Graff, L. & Walker, J. Effects of psychiatric comorbidity in immune-mediated inflammatory disease: Protocol for a prospective study. *J Med Internet Res* **20**, (2018).

62. Lindsey, J. W., Scott, T. F. & Lynch, S. G. The CombiRx trial of combined therapy with interferon and glatiramer acetate in relapsing remitting MS: Design and baseline characteristics. *Mult Scler Relat Disord* **1**, 81–86 (2012).

63. Bycroft, C., Freeman, C. & Petkova, D. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

64. Glanville, K. P., Coleman, J. R. I., O'Reilly, P. F., Galloway, J. & Lewis, C. M. Investigating Pleiotropy Between Depression and Autoimmune Diseases Using the UK Biobank. *Biol. Psychiatry Glob. Open Sci.* **1**, 48–58 (2021).

65. International Standard Classification of Education (ISCED) 2011.

66. Amin, V., Fletcher, J. M., Sun, Z. & Lu, Q. Higher educational attainment is associated with longer telomeres in midlife: Evidence from sibling comparisons in the UK Biobank. *SSM - Popul. Health* **17**, 101018 (2022).

67. Salter, A., Kowalec, K., Fitzgerald, K. C., Cutter, G. & Marrie, R. A. Comorbidity is associated with disease activity in MS: Findings from the CombiRx trial. *Neurology* **4;95(5):e446-e456**, (2020).

68. Smith, A. *Symbol Digit Modalities Test*. (Western Psychological Services, 2002).

69. Benedict, R. H., DeLuca, J. & Phillips, G. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler J* **23**, 721–733 (2017).

70. Tiftikçioğlu, B. İ. Multiple sclerosis functional composite (Msfc): Scoring instructions. *Noro-Psikiyatri Arsivi* **55**, S46–S48 (2018).

71. Edgar, C. *et al.* Cognitive performance in relapsing remitting multiple sclerosis: a longitudinal study in daily practice using a brief computerized cognitive battery. *BMC Neurol* **11**, (2011).

72. Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PloS One* **15**, e0231627 (2020).

73. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

74. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–7 (2016).

75. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).

76. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).

77. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).

78. Auton, A. & Brooks, L. D. 1000 Genomes Project Consortium. *Nature* **526**, 68–74 (2015).

79. Biggerstaff, B. J. & Tweedie, R. L. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Stat Med* **16**, 753–768 (1997).

80. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat Med Stat Med* **21**, 1539–1558 (2002).

81. Zhong, T., Li, S., Liu, P., Wang, Y. & Chen, L. The impact of education and occupation on cognitive impairment: a cross-sectional study in China. *Front. Aging Neurosci.* **Volume 16-2024**, (2024).

82. Sonder, J. M., Burggraaff, J., Knol, D. L., Polman, C. H. & Uitdehaag, B. M. Comparing long-term results of PASAT and SDMT scores in relation to neuropsychological testing in multiple sclerosis. *Mult. Scler. J.* **20**, 481–488 (2014).

83. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).