



SOCIAL DETERMINANTS OF HEALTH IN OLDER AGE:

GENE-ENVIRONMENT INTERACTIONS

ACROSS THE LIFE COURSE

Yan Liu

Nuffield College

University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Sociology

June 2026

Word Count: 53,260

Abstract

This thesis investigates how social and environmental factors shape health outcomes in later life and how these effects vary according to genetic risks, conceptualized through gene–environment interactions ($G \times E$). Health in older age reflects a dynamic life-course process influenced by biological, social, and environmental exposures, with growing evidence that most conditions arise from the interplay of genetic risks and contextual environments. Leveraging advances in genome-wide association studies and polygenic scoring, this project applies quasi-experimental designs to examine how socioeconomic status and life stressors interact with genetic liability to influence cognitive function and mental health in older adults in the United Kingdom.

The three essays use data from the English Longitudinal Study of Ageing (ELSA). The first applies group-based trajectory modelling, identifying distinct patterns of memory and depressive symptoms. Results indicate cognitive stability, modest reductions in depressive symptoms, and strong socioeconomic gradients, with disadvantages concentrated among older, female, and socioeconomically deprived groups.

The second employs a regression discontinuity design exploiting the 1947 UK school reform to estimate the causal effect of education on late-life cognition and its interaction with genetic risks. Findings suggest that while additional schooling benefits cognition, it may amplify the detrimental influence of the APOE4 allele.

The last essay examines mental health consequences of the COVID-19 pandemic, showing that men with higher polygenic scores for major depressive disorder experienced greater increases in depressive symptoms, underscoring differential vulnerability to environmental shocks.

Taken together, the thesis provides novel evidence of the heterogeneity of cognitive and mental health trajectories in later life and substantive understanding of gene–environment interplay across the life course. By integrating polygenic scores with quasi-experimental designs, the analyses highlight how social exposures—such as education and the COVID-19 pandemic—interact with genetic risks to shape ageing outcomes.

Acknowledgements

This dissertation represents not only the culmination of several years of study, but also the support, encouragement, and guidance of many people to whom I owe a great debt of gratitude.

First and foremost, I would like to express my deepest appreciation to my supervisor, Jennifer Dowd, for her invaluable guidance, patience, and insight throughout every stage of this research. Her rigorous scholarship and unwavering support have been a constant source of inspiration, especially during times of uncertainty and transition, when her advice and trust were instrumental in helping me stay focused and move forward.

I am also grateful to my Transfer and Confirmation examiners, Charles Rahal, Evelina Akimova, and Xuejie Ding, for their thoughtful feedback and constructive suggestions, which greatly enriched this work. I am deeply thankful to the Leverhulme Centre for Demographic Science Graduate Scholarship, without which this research would not have been possible. I am also indebted to Eleni Kechagia-Ovseiko at Nuffield College and Christiaan Monden at the Department of Sociology, who offered kindness and practical help when I felt overwhelmed.

I extend my thanks to colleagues and friends in Oxford for their advice and support, whether in refining ideas, sharing resources, or offering encouragement during challenging times. On a more personal note, I am profoundly thankful to my family for their unconditional love and support. To my parents, thank you for instilling in me the values of perseverance and curiosity.

Finally, I dedicate this thesis to all those who believed in me when I doubted myself. This achievement is as much theirs as it is mine.

Table of Contents

CHAPTER 1	- 1 -
INTRODUCTION AND THESIS OUTLINE	- 1 -
1.1 GENERAL INTRODUCTION	- 1 -
1.2 THESIS OUTLINE	- 3 -
1.3 LITERATURE REVIEW	- 7 -
<i>1.3.1 Rationales for the Study of Gene-Environment Interactions</i>	<i>- 7 -</i>
<i>1.3.2 Examining $G \times E$ by Leveraging Polygenic Scores: Opportunities and Cautions</i>	<i>- 14 -</i>
<i>1.3.3 Gene-environment Correlation and Quasi-Experimental Approaches</i>	<i>- 22 -</i>
<i>1.3.4 $G \times E$ in Age-Related Cognitive Function Decline</i>	<i>- 28 -</i>
<i>1.3.5 $G \times E$ in Mental Health</i>	<i>- 37 -</i>
<i>1.3.6 Sex Differences in $G \times E$: Biological and Social Perspectives</i>	<i>- 42 -</i>
1.4 THEORETICAL FRAMEWORK	- 46 -
1.5 THESIS AIMS AND HYPOTHESES	- 51 -
1.6 DATA AND GENERAL METHODOLOGY	- 53 -
<i>1.6.1 Benefits of Studying the Context of the United Kingdom</i>	<i>- 53 -</i>
<i>1.6.2 The English Longitudinal Study of Ageing (ELSA)</i>	<i>- 56 -</i>
<i>1.6.3 Measurement of Outcome Variables</i>	<i>- 59 -</i>
<i>1.6.4 Genetic Biomarkers: Construction of PGSs and PCs in ELSA</i>	<i>- 66 -</i>
CHAPTER 2	- 72 -
TRAJECTORIES OF COGNITION AND MENTAL HEALTH IN OLDER AGE	- 72 -
2.1 INTRODUCTION	- 72 -
2.2 METHODS	- 76 -
<i>Study Design and Participants</i>	<i>- 76 -</i>
<i>Measures</i>	<i>- 77 -</i>
2.3 STATISTICAL ANALYSIS	- 81 -
2.4 RESULTS	- 84 -
2.5 DISCUSSION	- 102 -
2.6 SUPPLEMENTS	- 108 -
CHAPTER 3	- 114 -
THE IMPACT OF COMPULSORY SCHOOLING ON COGNITION AND ITS INTERACTION WITH GENETIC SCORES	- 114 -
3.1 INTRODUCTION	- 114 -
3.2 DATA	- 119 -
<i>Measures</i>	<i>- 119 -</i>
3.3 ANALYTICAL STRATEGY	- 123 -
3.4 RESULTS	- 134 -
3.5 DISCUSSION	- 147 -
3.6 SUPPLEMENTS	- 154 -
CHAPTER 4	- 163 -
GENE-ENVIRONMENT INTERACTION DURING A PANDEMIC: HOW POLYGENIC SCORES SHAPE MENTAL HEALTH RESPONSES TO COVID-19	- 163 -
4.1 INTRODUCTION	- 163 -
4.2 METHODS	- 171 -
<i>Data and Participants</i>	<i>- 171 -</i>
<i>Measures</i>	<i>- 171 -</i>
4.3 STATISTICAL ANALYSIS	- 175 -
<i>Identification Strategy: The Two-Way Fixed Effects (TWFE) Framework</i>	<i>- 175 -</i>
4.4 RESULTS	- 182 -
4.5 DISCUSSION	- 195 -
4.6 SUPPLEMENT	- 200 -

CHAPTER 5	- 216 -
CONCLUSION.....	- 216 -
SUMMARY OF FINDINGS	- 216 -
THEORETICAL CONTRIBUTIONS	- 219 -
STRENGTHS AND LIMITATIONS	- 223 -
POLICY IMPLICATIONS	- 227 -
GENERALISABILITY BEYOND THE UNITED KINGDOM	- 229 -
DIRECTIONS FOR FUTURE RESEARCH	- 230 -
CONCLUDING REMARKS.....	- 232 -
REFERENCES.....	- 234 -

List of Tables

- Table 1.1** Summary of ELSA Waves Used Across Empirical Chapters
- Table 2.1** Summary of longitudinal data on the scores of memory and depressive symptoms
- Table 2.2** Baseline sample characteristics (%), by memory trajectories and overall
- Table 2.3** Baseline sample characteristics (%), by depressive symptoms trajectories and overall
- Table 2.4** Average marginal effects for a multinomial logistic regression model for memory trajectories
- Table 2.5** Average marginal effects for a multinomial logistic regression model for DS trajectories
- Table 2.6** Trajectories of memory and the effects of time-varying predictors on trajectory shapes, stratified by trajectory group
- Table 2.7** Trajectories of depressive symptoms and the effects of time-varying predictors on trajectory shapes, stratified by trajectory group
- Table 3.1** Distribution of age of leaving full-time education by sex and date of birth (individuals born five years on either side of the cut-off date of April 1, 1933) (N, %)
- Table 3.2** Cognition scores and comparison by sex and age of leaving full-time education (for individuals born 5 years before and after the 1933 cut-off)
- Table 3.3** Correlations between Study Variables
- Table 3.4** Descriptive statistics of variables of the included and non-included samples
- Table 3.5** Memory and executive function scores by sex and age of leaving full-time education (for individuals born 5 years before and after the 1933 cut-off)
- Table 3.6** First stage results: effect of the reform on years of education
- Table 4.1** Measures of mental disorders
- Table 4.2** The summary statistics for PGS for DS, MDD, AD and Loneliness
- Table 4.3** Correlations between Study Variables
- Table 4.4** Statistical Analysis for Different Outcomes
- Table 4.5** Comparison of the sample included vs. non-included from the ELSA COVID-19 longitudinal study
- Table 4.6** Descriptive changes in mental health before and during COVID-19 by sex

List of Figures

Figure 1.2 Different lifespan trajectories of fluid intelligence and crystallized intelligence

Figure 1.1 Situating the Thesis Within the Life Course Cube Framework

Figure 2.1 Memory trajectories

Figure 2.2 Depressive symptoms trajectories

Figure 3.1 Fraction leaving school by age 14 by month of birth for males and females

Figure 3.2 Two-stage least squares effect of an additional year of schooling at age 15 on cognition controlling for polygenic scores but excluding gene–environment interactions

Figure 3.3 Two-stage least squares effect of an additional year of schooling at age 15 on memory and executive function including interactions with individual PGS vs three PGSs

Figure 3.4 Effect of an additional year of schooling at age 15 on executive function by APOE4 status for females with 95% CIs

Figure 4.1 Mental Health Trajectories before and after the COVID-19 Pandemic

Figure 4.2 Post $\times$ PGS Interaction Effects on Mental Health Outcomes (Males)

Figure 4.3 Post $\times$ PGS Interaction Effects on Mental Health Outcomes (Females)

Figure 4.4 Marginal effects of PGS MD $\times$ Post COVID-19 with 95% CIs

Abbreviation

- **AD** – Alzheimer’s disease
- **AD-PGS** – Alzheimer’s disease polygenic score
- **AIC** – Akaike Information Criterion
- **AME** – Average marginal effect
- **APOE** – Apolipoprotein E
- **APOE4** – Apolipoprotein E ϵ 4 allele
- **BIC** – Bayesian Information Criterion
- **CI** – Confidence interval
- **CSL** – Compulsory schooling law
- **DS** – Depressive symptoms
- **ELSA** – English Longitudinal Study of Ageing
- **GBTM** – Group-based trajectory modelling
- **G×E** – Gene–environment interaction
- **GWAS** – Genome-wide association study
- **HRS** – Health and Retirement Study
- **MCI** – Mild cognitive impairment
- **MDD** – Major depressive disorder
- **OR** – Odds ratio
- **PC** – Principal component
- **PGS** – Polygenic score
- **PRS** – Polygenic risk score
- **SD** – Standard deviation
- **SES** – Socioeconomic status
- **SRH** – Self-rated health

Glossary of Terms

Assortative mating: When couples who produce offspring select one another based on particular phenotypes.

Genome-wide association study (GWAS): A study that estimates associations between genetic variants (usually SNPs) and traits within one dataset, typically using individual-level genotype and phenotype data. A genome-wide significance threshold of $p < 5 \times 10^{-8}$ is commonly applied.

Genome-wide association meta-analysis (GWAMA): A meta-analysis that combines GWAS results (summary statistics) from multiple cohorts to increase sample size and statistical power.

Imputation: A statistical method that infers unobserved genotypes by using a reference panel from complete genome sequence data.

Indirect genetic effect: The effect of a genetic variant in one individual on the trait of another individual through environmental pathways (e.g., parental genotype influencing child outcomes).

Linkage disequilibrium (LD): The correlation between alleles at different loci (i.e., the nonrandom association of alleles at different loci).

Principal component (PC): A linear combination of SNP genotypes that captures major axes of genetic variation in a sample, where each SNP contributes with a 'loading'.

Rubin's rules: A statistical procedure for combining results from multiple imputed datasets, accounting for both within- and between-imputation variability.

Single-nucleotide polymorphism (SNP): A variation in a single base pair (A, T, G, or C) at a specific location in the genome.

Chapter 1

Introduction and Thesis Outline

1.1 General Introduction

This thesis aims to examine the impact of social and environmental factors on health outcomes of older people and how the impact might vary by genetic risks, conceptualized through gene-environment interactions ($G \times E$). Health in older age is the result of a dynamic process throughout the life course based on a combination of genetic, biological, social and environmental factors. It has been increasingly accepted that the aetiology of most common diseases involves not only discrete genetic and environmental factors but also interactions between genetic and environmental exposures. Gene-environment interaction ($G \times E$) processes impacting health likely play out over the entire life course. The availability of genetic data in large population surveys has resulted in an explosion of research on gene-environment interactions. Understanding the effects of critical social and environmental exposures on later life health outcomes and how these vary by underlying genetic susceptibility can help inform targeted interventions going forward.

For the “gene” examined in the gene-environment interaction, complex traits such as cognitive function and depression are highly polygenic, which means that they are influenced by thousands upon thousands of genetic loci scattered throughout the genome, each with a tiny effect (Chabris et al., 2015). Candidate genetic variant studies were operating under a naive assumption that a few genes with large effects are responsible for the heritability of most traits, and they were therefore conducted with sample sizes that were underpowered (Chabris et al., 2015). Hence instead of candidate

genes, I will use polygenic scores in my project, which is an estimate of an individual's genetic liability to a trait or disease, calculated according to their genotype profile and relevant genome-wide association study data (Barr et al., 2022; Choi et al., 2020; Dudbridge, 2013).

In most cases, the “environment” in $G \times E$ is not randomly assigned. This means it may be confounded by other non-genetic variables and could be correlated with the genetic variant in question, known as gene-environment correlation (r_{GE}). I will incorporate quasi-experimental designs into the $G \times E$ framework to leverage exogenous environmental variation to reduce the influence of gene-environment correlation and strengthen causal inference.

The overall aim of my thesis is to advance our understanding of the social and environmental determinants of health in later life across the life course, with a specific focus on gene–environment interactions. Specifically, it examines:

1. Trajectories of two key late-life health outcomes—cognitive function and mental health—and their predictors.
2. Education as a lifelong socioeconomic influence on cognition, both directly and indirectly, with a particular focus on underlying genetic risk and improved causal inference in gene–environment ($G \times E$) research.
3. The COVID-19 pandemic as a stressful life event affecting mental health, with a particular focus on underlying genetic risk and improved causal inference in gene–environment ($G \times E$) research.

1.2 Thesis Outline

This thesis is paper based.

Chapter 1 provides a comprehensive foundation for this thesis. Firstly, the review of existing literature explores the progress of genetic discovery and the aetiology of the core phenotypes. I describe the transition from candidate gene studies to Genome-Wide Association Studies (GWAS), the emergence of Polygenic Scores (PGS), and the persistent methodological challenges in identifying robust gene-environment interactions (G×E). This is followed by a detailed examination of the epidemiology and aetiology of age-related cognitive decline and mental health, which establishes the phenotypic focus of this research. Secondly, I construct a life-course framework of health in late life, which identifies critical knowledge gaps in how genetic risks manifest over time. Based on this synthesis, I define the specific aims and objectives of this DPhil thesis, highlighting the original contributions intended to address these gaps in the fields of social genomics and gerontology. Lastly, the empirical context and core methodology of the research are presented. I discuss the motivation for focusing on the United Kingdom context, followed by a detailed description of the English Longitudinal Study of Ageing (ELSA) and its genetic and Covid database. This section establishes the rigor of the subsequent analyses by detailing the construction protocols for the genetic predictors and the validation of the primary outcome measures.

Chapter 2 is the first empirical study to apply group-based trajectory modelling to examine distinct longitudinal patterns of cognitive function and depressive symptoms among older adults in England, using six waves of data from the English Longitudinal Study of Ageing (ELSA). The findings indicate memory stability and a modest reduction in depressive symptoms at the population level. Socioeconomic and

demographic factors were strongly associated with trajectory group membership: individuals with higher education and wealth were substantially more likely to sustain high memory function and low depressive symptoms, whereas older age, female gender, and early-life disadvantage were predictive of less favourable trajectories. Analyses with time-varying predictors showed that poor self-rated health, disability, and multimorbidity were linked to lower memory and higher depressive symptoms, especially in lower-functioning groups. Collectively, these findings highlight both the stability and variability of cognitive and mental health trajectories in later life, while underscoring the cumulative impact of socioeconomic advantage and disadvantage across the life course. This chapter therefore provides an essential foundation for the subsequent analyses of gene–environment interactions in these two health outcomes.

Chapter 3 is the second empirical paper that explores the impact of education on cognitive function and how this impact interacts with genetic risks. Building off of work by Banks and Mazzonna (Banks & Mazzonna, 2012), I exploit the reform of compulsory schooling in England in 1947, which raised the minimum school-leaving age from 14 to 15 for individuals born after a specific cut-off in March 1933 and has been shown to have strong impacts on educational attainment. I use a fuzzy regression discontinuity design and compare individuals born before and after the cut-off to identify the causal effect of the policy-induced one additional year of education change on later life cognition and how this interacted with genetic risk. Data come from the English Longitudinal Study of Ageing (ELSA), the ELSA Special License data and the ELSA Genetic data. The findings show that education affects cognitive function and may accelerate the detrimental effects of the APOE4 allele on cognition. However, larger sample sizes are required to robustly validate these findings. This paper was examined during the Transfer of Status.

Chapter 4 is the last empirical paper that explores the impact of exposure to the COVID-19 pandemic on mental health and how this interacts with genetic risk. I use two-way fixed effects model regression to estimate the causal effects of the COVID-19 pandemic on mental health outcomes among older people using the panel data from the ELSA before the COVID-19 pandemic (Wave 9 in 2018- 2019) and on 2 occasions in 2020 (June/July and November/December). I use polygenic scores (PGS) for mental-health-related phenotypes in the ELSA Genetic data to examine whether and how the effect is moderated by genetic risks. I find that men with higher PGSs show more depressive symptoms than those with lower PGSs for MDD. However, larger sample sizes are required to robustly validate these findings. This paper was examined during the Confirmation of Status and was orally presented at the Population Association of America Annual Meeting in 2023.

Chapter 5 brings these strands together to argue that late-life health is a product of the continuous, dynamic interplay between genetic risks, social determinant, and historical events. By integrating longitudinal modelling with quasi-experimental causal identification, this thesis moves beyond the static "nature versus nurture" debate toward a more nuanced, biosocial understanding of the ageing process. A central contribution of this synthesis is the demonstration that gene-environment interactions are not monolithic, providing evidence that different genetic markers follow distinct theoretical logics: while the COVID-19 analysis reveals a classic diathesis-stress pattern where acute shocks activate latent vulnerabilities (PGS MDD), the findings for female APOE4 carriers support the differential susceptibility framework, as the 1947 reform led to a substantial 2.5-unit (1.77 SD) reduction in executive function composite score, identifying APOE4 as a "plasticity factor" rather than a simple risk gene. Critically, this chapter directly confronts the challenge of modest statistical power and small effect

sizes ($f^2 < 0.02$) inherent in sociogenomics, arguing that the significance of these findings lies in their causal robustness and subgroup relevance rather than population-level variance explained. Even a statistically small R^2 can represent a major life-course shift for specific high-risk groups when identified through natural experiments that minimize selection bias. Substantively, the results underscore that while education is a fundamental cause of health, its benefits are filtered through biological heterogeneity, and policy interventions must therefore adopt a biosocial perspective that addresses both structural disadvantage and biological diversity to promote more equitable ageing.

1.3 Literature Review

1.3.1 Rationales for the Study of Gene-Environment Interactions

The discourse on human development has historically been dominated by the "nature versus nurture" dichotomy, a term coined by Galton (1875) to describe the competing influences of innate heredity and external experience. For over a century, this debate sought to partition phenotypic variance into mutually exclusive biological or environmental components, a statistical reductionism that often overlooked the dynamic synergy between the two (Ridley, 2003; Tabery, 2023). However, the contemporary understanding of human complexity has moved beyond this artificial separation. It is now increasingly recognized that the interaction between genetic and environmental variation is not merely an analytical nuance, but a fundamental mechanism underlying the aetiology of diverse traits and diseases, ranging from mental disorders to Alzheimer's disease and diabetes (Hunter, 2005; Caspi & Moffitt, 2018).

By transcending the "nature versus nurture" framework, the investigation of gene–environment interactions ($G \times E$) offers several transformative contributions to the biological and social sciences. First, it offers insights into the aetiology of diseases by demonstrating how risk develops through the interplay of biological predispositions and contextual exposures (Hunter, 2005; Le Marchand & Wilkens, 2008). Second, it improves treatment and prognostic models by integrating genotype information, moving research closer to personalized medicine (Thomas, 2010). Finally, it identifies genetic subgroups with heightened exposure-specific risks, thereby informing targeted prevention strategies (McAllister et al., 2017).

From Heritability Estimates to Molecular Realities

In recent years, advances in genetic research have introduced new study designs, analytical strategies, and statistical methods for exploring gene-environment interactions. One of the foundational concepts in this area is heritability, defined as the proportion of variation in a trait within a population that can be attributed to genetic differences (Visscher et al., 2008). Historically, twin studies served as the primary method for estimating heritability, comparing monozygotic (identical) and dizygotic (non-identical) twins to distinguish between genetic and environmental contributions (Galton, 1875). For example, a Swedish study of 15,493 twins reported that lifetime major depression was moderately heritable, with estimates of around 37% (Gatz et al., 2006; Kendler et al., 2006). Despite their influence, twin studies face limitations. They rely on assumptions such as non-random mating and the equal environment hypothesis, and they cannot identify specific genes or exposures. Moreover, they are vulnerable to gene-environment correlations that can obscure causal pathways (Røysamb & Tambs, 2016).

Since the completion of the Human Genome Project in 2003, considerable effort has been devoted to alternative estimates of heritability based on molecular genetic data. Early candidate gene studies tested prior hypothesis-driven candidate genes based on the loci's supposed biological functions, such as the widely cited (but heavily debated) association between the 5-HTTLPR serotonin transporter gene and sensitivity to stressful life events (Caspi et al., 2002). However, candidate gene studies have prior hypothesis-driven candidate genes, low statistical power, unsuccessful replications, and publication bias toward positive findings, which are “wrong or misleading” (Hewitt, 2012). Very few exceptional cases are for certain traits and diseases for which effect sizes are large and associations are replicated, for instance, the discovery of a major risk allele for late-onset Alzheimer's disease (AD) - apolipoprotein E (APOE) gene. The

APOE gene explains 13% of the heritability in Alzheimer's disease and the presence of each risk allele translates into a fivefold increase in the risk of developing AD (Battista et al., 2016). Yet it is also noteworthy that, except in these instances such as the APOE4 gene, most reported candidate gene associations have not replicated in GWA studies (Hewitt, 2012). A broad array of traits (e.g., cognitive function, mental health, height, body mass index, educational attainment and risk-taking) are complex and highly polygenic. Population variation in these complex traits is attributable to many genetic variants throughout the genome, each individually exhibiting a relatively small effect.

The Rise and Limits of Polygenic Scores

The limitations of candidate gene studies spurred the development of genome-wide association studies (GWAS), enabled by the International HapMap Project in 2005. A genome-wide association study is an approach that involves rapidly scanning markers across the complete sets of deoxyribonucleic acid (DNA) or genomes of many people to find genetic variations, namely single-nucleotide polymorphisms (SNPs), associated with a particular disease/trait (Uffelmann et al., 2021). This approach has become the most powerful tool for gene discovery and has also enabled the development of polygenic scores (PGSs), indices that aggregate the effects of numerous genetic variants across the genome on specific traits or diseases (So & Sham, 2017). A PGS is typically constructed by summing trait-associated alleles, weighted by their corresponding effect sizes estimated from a reference genome-wide association meta-analysis (GWAMA) (Choi et al. 2020). The resulting scores are continuous measures that capture an individual's genetic predisposition to a wide range of outcomes, from educational attainment to mental health conditions (Choi et al., 2018; Dudbridge, 2013; Purcell et al., 2009).

Polygenic scores offer great promise to incorporate genetics into social science to better understand how genetic effects are contingent upon social environments (Mills & Rahal, 2019). For example, Barcellos et.al (Barcellos et al., 2018a) investigated whether the educational policy affected health differently for people with different genetic backgrounds using the UK Biobank data. They used PGSs for body mass index (BMI) and educational attainment and found that education reduced the gap in unhealthy body size between those in the top and bottom terciles of the PGS for BMI from 20 to 6 percentage points. In the long run, PGSs might be used to inform therapeutic intervention, disease screening and life planning (Torkamani et al., 2018). For example, if healthy lifestyle choices were preferentially targeted to and employed by women in the top decile of genetic risk of breast cancer, an estimated percentage of 20% of all preventable breast cancer cases would be avoided (Maas et al., 2016).

While polygenic scores have been shown to predict complex phenotypes, they are far from perfect as measures of genetic influence. On one hand, PGSs do not account for all of the genetic effects. They account for common genetic variants but not rare variants. However, the growing accessibility of whole-exome and whole-genome sequencing raises the strategy of including both common and rare variants in these genetic profiles (Biddinger et al., 2022; Bishop et al., 2021). On the other hand, PGSs are not a clean way to separate genetics from other effects that contribute to variation among people's outcomes (Harden & Koellinger, 2020). GWAS results are not entirely immune to unobserved (e.g., demographic, social or environmental) confounds, such as assortative mating, population stratification, socioeconomic status, and indirect genetic effects of relatives (Howe et al., 2019; Kong et al., 2018; Selzam et al., 2019; Shen & Feldman, 2020; Trejo & Domingue, 2018) For example, a recent study of genetic data from parents and offspring (Kong et al., 2018) showed that the parental genotype that children

do not inherit is still associated with children's educational attainment, suggesting that the genetic nurture effect is an indirect link between parental genotypes and children's characteristics and GWAS results for educational attainment partly represent these intergenerational pathways. Thus, despite the optimal strategy for gene discovery and polygenic prediction remains to maximize sample sizes using unrelated individuals, validating PGSs within families is vitally important for sifting out causation from correlation, especially for phenotypes sensitive to demographic and indirect genetic effects, such as educational attainment (Howe et al., 2022).

Addressing the "Missing Heritability" via $G \times E$

A further challenge is the problem of missing heritability. Although PGSs may be inflated due to shared environmental effects within the family, heritability estimates derived from GWASs explained only a fraction of the total twin and family heritability (Manolio et al., 2009). Substantial missing heritability remains in most complex diseases. For example, heritability estimates from GWAS amount to about 23 % of the variance in liability in schizophrenia (SZ) and about 11 % for major depressive disorder (MDD). This is far lower than the findings based on family studies that heritability for SZ accounts for 60–80 % of disease risk and more than 40% for MDD (Avramopoulos, 2018; Flint & Kendler, 2014; van Calker & Serchov, 2021). The missing heritability may be attributed to a combination of causes, including a large number of common variants of extremely small effect sizes which elude detection, rare variants to which GWAS are insensitive, epigenetic mechanisms, such as DNA methylation and histone modifications, and noncoding RNAs (ncRNA) (Génin, 2020) (Maroille & Tarailo-Graovac, 2019; Schuebel et al., 2016). Yet a further possibility is that GWAS are based on combining data from multiple periods, countries, and birth cohorts. Thus, this will

introduce heterogeneity in genetic effects, as a function of environmental factors to which some but not all individuals are exposed, implying a gene-environment interaction ($G \times E$) (Conley, 2016b; Tropf et al., 2017). A recent study using UK Biobank data estimated $G \times E$ heritability for BMI, finding that smoking, alcohol intake, and physical activity significantly moderated genetic effects, thereby suggesting that $G \times E$ may account for at least part of the missing heritability (Jung et al., 2023).

Theoretical Models of GxE

While empirical advances have revealed the scale of missing heritability and the role of environmental moderation, they cannot fully explain the mechanisms underlying $G \times E$. To bridge this gap, researchers have developed theoretical models that provide frameworks for interpreting how genetic risks interact with different contexts. These models help translate statistical patterns into clearer insights about health and disease. To better interpret these complexities, researchers have proposed three influential theoretical models of $G \times E$: the diathesis–stress (social trigger) model, the vantage sensitivity model, and the differential susceptibility model (Boardman et al., 2012; Mills & Tropf, 2020). The diathesis-stress model poses that the risky genotype will have a more deleterious effect on a trait when triggered by an environmental stressor. The emphasis of this model is on the toxicity of stressful environments. For instance, a study supported this model by using the PGS for major depressive disorder and stressful life events effects on depressive symptoms. In a sample of 4919 participants throughout Scotland, it showed that a higher PGS for major depressive disorders was associated with an increased risk of depressive symptoms only after a series of stressful life events (Arnau-Soler et al., 2019). Whereas the diathesis-stress model focuses almost exclusively on negative environmental exposures, the vantage sensitivity model suggests

that sensitive individuals derive the greatest benefits from a trait in nurturing, supportive, and stress-free environments. The differential-susceptibility model poses that plasticity varies by individual with some more genetically susceptible to both positive and negative environments, while others remain resilient across all situations. Examining and validating all these theoretical models with updated data is an important contribution to this larger body of $G \times E$ work. Testing these models with updated data is critical to advance $G \times E$ research and clarify how genes and environments jointly shape human outcomes.

Equity and the Ethics of Environmental Genomes

The ethical, legal, and social dimensions of genomic research must be understood not only as a set of risks to be managed but as active principles that shape how science is conducted and applied. Early debates highlighted foundational issues such as privacy, informed consent, and the potential for genetic discrimination (Clayton, 2003), yet these concerns alone are no longer adequate. As genomic technologies expand into population studies and clinical care, the pressing challenge lies in creating structures that address equity, trust, and accountability in practice. Large-scale genome-wide association studies have transformed biomedical science, but they remain disproportionately focused on populations of European ancestry, raising serious concerns about both scientific validity and social justice (Mills & Rahal, 2019). Existing governance has not fully kept pace with the complexity of data sharing and diverse population needs, a gap that becomes particularly significant for research on gene–environment interactions, where biological data are inseparable from lived social contexts (Dolan, 2022). Recent scholarship has emphasized that responsible genomics requires more than mitigating harm: it demands translational frameworks that integrate justice into every stage of

research, from design and recruitment to communication and application (Halley et al., 2024; Dolan et al., 2024). This perspective underscores the necessity of treating justice as a guiding principle, ensuring that genomic and $G \times E$ research contributes to reducing—rather than reinforcing—health disparities.

In sum, research on gene–environment interactions has progressed from early heritability estimates and candidate gene studies to genome-wide approaches and polygenic scores, revealing both the potential and limitations of genetic prediction. While GWAS and PGSs have expanded the scale of discovery and opened new avenues for integrating genetics into social science and medicine, they also highlight persistent challenges such as missing heritability, confounding by social structures, and unequal representation across populations. Theoretical models—including diathesis–stress, vantage sensitivity, and differential susceptibility—help interpret these complex patterns by framing how genetic risks are amplified or moderated in different environments. At the same time, ethical and social considerations underscore that genomics cannot be treated purely as a technical enterprise: equity, accountability, and justice must be embedded in research design and application to ensure that $G \times E$ findings reduce rather than reinforce health disparities. Together, these rationales justify a sustained focus on $G \times E$ research as an essential pathway to understanding human health and guiding more just and effective interventions.

1.3.2 Examining $G \times E$ by Leveraging Polygenic Scores: Opportunities and Cautions

In the study of gene–environment interaction ($G \times E$), several conceptual and methodological challenges need to be acknowledged. Unlike monogenic diseases, where single genetic variants can strongly predict outcomes, complex behavioral and health

traits—such as educational attainment or mental health—are highly polygenic, with many genetic variants each exerting small effects (Boyle et al., 2017; Mills & Tropf, 2020). Hence, early $G \times E$ research using candidate genetic variant studies often failed due to overly simplistic assumptions that a handful of variants could account for substantial phenotypic variance (Border et al., 2019).

Modern $G \times E$ research overcomes the limitations of candidate gene approaches by utilising Polygenic Scores (PGS), which aggregate the additive genetic influences of thousands to millions of variants across the entire genome for a specific phenotype (So & Sham, 2017). Rather than focusing on a single locus, a PGS captures the infinitesimal genetic architecture of complex traits by weighing individual genetic variants—typically single-nucleotide polymorphisms (SNPs)—according to their marginal effect sizes derived from independent, large-scale genome-wide association study (GWAS) discovery samples (Choi & O’ Reilly, 2020). By condensing this vast genomic information into a single, continuous measure of genetic liability, PGSs notably enhance statistical power and offer a more robust proxy for biological predisposition, thereby enabling a more nuanced investigation of how cumulative genetic risk interacts with environmental exposures (Trejo et al., 2018). The precise technical construction and quality control protocols for the scores are detailed in Section 1.6.4.

One of the most widely recognized strengths of polygenic scores (PGSs) lies in their ability to aggregate the effects of thousands of common genetic variants into a single index, thereby greatly enhancing predictive accuracy relative to single-variant approaches. This aggregation enables researchers to stratify individuals according to their genetic predisposition to a wide range of outcomes, from body mass index and cardiovascular disease to educational attainment and depression (Lewis & Vassos, 2020;

Torkamani et al., 2018). Importantly, PGSs can identify individuals at the tails of the risk distribution who may be especially vulnerable to adverse health outcomes. For example, women in the top decile of a breast cancer PGS can benefit disproportionately from targeted lifestyle interventions and screening programs, which could potentially prevent up to 20% of cases (Maas et al., 2016). Such applications highlight the potential of PGSs to improve precision medicine by informing early intervention, personalized prevention strategies, and optimized clinical decision-making.

Beyond prediction, PGSs have transformed the methodological landscape of human genetics by enabling researchers to quantify genetic predisposition in large-scale population cohorts. They provide a continuous measure of genetic liability that can be incorporated into epidemiological models, thereby improving causal inference in the study of complex traits (So & Sham, 2017). By serving as proxies for aggregate genetic risk, PGSs allow researchers to test hypotheses about how genetic influences interact with environmental factors across the life course. For instance, Barcellos et al. (2018) demonstrated that educational reforms in the United Kingdom reduced the adverse health gap between individuals with high and low genetic risk for obesity, illustrating how social policies can moderate genetic vulnerabilities. Similarly, in sociology and economics, PGSs have been used to examine how inequality, education, and labor market conditions interact with genetic predispositions, providing a richer account of social outcomes (Mills & Rahal, 2019; Harden & Koellinger, 2020). In this sense, PGSs help move scholarship beyond simplistic “nature versus nurture” debates toward an integrated framework where biology and environment are studied in tandem.

However, the application of polygenic scores in GxE research requires navigating a series of substantial conceptual and methodological hurdles:

Interpretive and Explanatory Constraints

A foundational principle in sociogenomic research is that polygenic scores (PGSs) capture relative—not absolute—genetic predisposition (Lewis & Vassos, 2020; Torkamani et al., 2018). Derived from genome-wide association studies (GWAS), a PGS reflects how an individual’s genetic liability compares to others within the same discovery population. Thus, a person with a high PGS for a trait such as educational attainment or depression is at greater genetic risk relative to peers, but this does not translate into a fixed or deterministic probability of developing the trait. Without careful contextualization, PGS findings may be misinterpreted as fixed biological risks, reinforcing deterministic narratives and potentially contributing to stigma or misinformed policy applications (Harden, 2021).

At the same time, their explanatory power remains modest for most complex traits. Even for highly heritable outcomes such as educational attainment or schizophrenia, current PGSs typically account for only a small proportion of total phenotypic variance (Duncan et al., 2019). They capture primarily the additive effects of common variants, while neglecting rare variants, structural variation, epigenetic regulation, and gene–gene interactions (Boyle et al., 2017). As a result, much of the “missing heritability” remains unresolved, leaving PGSs as incomplete proxies of genetic influence.

Portability and Population Stratification

Beyond interpretive issues, the scientific validity of G×E findings is further challenged by portability problems. Because most GWAS have been conducted in European-ancestry samples, predictive accuracy declines sharply in non-European populations due to allele frequency differences, linkage disequilibrium structures, and heterogeneous environments (Mills & Rahal, 2019; Martin et al., 2019). Even within ancestry groups,

predictive performance can vary with socioeconomic stratification or institutional contexts, undermining claims of universal validity (Mostafavi et al., 2020). Thus, while PGSs can powerfully rank-order risk within discovery populations, their utility across diverse groups remains constrained until multi-ancestry GWAS become standard (Carlson et al., 2022). This dual reality—conceptual promise coupled with empirical constraint—underscores the importance of interpreting PGSs as probabilistic and context-dependent tools rather than deterministic indicators of outcome risk.

To mitigate these biases, researchers must rigorously account for population stratification—systematic differences in allele frequencies that correlate with geography or culture rather than the trait itself. If these ancestral lineages also differ in their social environments, dietary habits, or geographical exposures, failing to control for population structure can result in spurious $G \times E$ associations, where genetic noise is misinterpreted as a biological moderated effect. The inclusion of Principal Components (PCs) derived from genome-wide genotype data as covariates is the methodological gold standard for addressing this confounding.

Mathematically, Principal Component Analysis (PCA) is a dimension-reduction technique that transforms a high-dimensional matrix of hundreds of thousands of single-nucleotide polymorphisms (SNPs) into a smaller set of uncorrelated variables (components). This is typically achieved through an eigenvalue decomposition of the genetic relationship matrix, where each successive principal component captures a decreasing proportion of the total genetic variance across the sample (Price et al., 2006). Biologically, the top PCs function as essential proxies for an individual's genetic ancestry; for instance, the first few PCs typically reflect geographical clines, such as the

north-to-south genetic gradients observed across Europe (Novembre et al., 2008). In a regression framework, these components act as "anchors" for ancestral background.

The determination of the optimal number of Principal Components (PCs) to include as covariates involves a strategic trade-off between insufficient correction and statistical over-correction. Theoretically, the number of significant PCs can be identified through a Scree Plot or the Tracy-Widom test, which identifies the "elbow" point where the marginal variance explained by additional components drops to a level indistinguishable from random noise (Patterson et al., 2006). In highly diverse or admixed populations, such as those in the United States or Latin America, extending the adjustment to 20 or even 40 PCs is often necessary to account for fine-scale admixture and subtle founder effects that remain undetected by the primary axes of variation (Patterson et al., 2006). However, in the context of G×E research, an indiscriminate increase in PCs may introduce over-correction bias; because PCs can correlate with geographically-patterned environmental variables (e.g., regional deprivation or climate), excessive adjustment risks absorbing meaningful environmental variance, thereby attenuating the very interaction effects under investigation (Conley et al., 2017).

For research utilizing large-scale British cohorts, such as the UK Biobank (UKB) or the English Longitudinal Study of Ageing (ELSA), the inclusion of the top 10 Principal Components has emerged as a robust and empirically validated standard. These cohorts are characterized by a relatively high degree of genetic homogeneity within the White British sub-sample, where the first few PCs typically capture the primary north-to-south and east-to-west geographic clines (Novembre et al., 2008). Systematic sensitivity analyses within these biobanks have consistently demonstrated that while higher-order PCs (e.g., 11–40) may capture extremely localized clusters, they explain a negligible

proportion of additional genetic variance and rarely alter the point estimates for complex trait associations (Bycroft et al., 2018; Privé et al., 2020). Consequently, following the prevailing convention in sociogenomic literature, this study adopts the top 10 PCs as sufficient controls to maintain the balance between neutralizing ancestral stratification and preserving the environmental variation necessary for robust $G \times E$ inference. The precise technical construction for the PCs is detailed in Section 1.6.4.

Collider Bias and Systematic Biases

Even when ancestral stratification is properly managed through Principal Components, $G \times E$ research remains vulnerable to more localized forms of confounding. Collider bias introduces a further challenge. Because genetic and environmental factors often covary, including heritable covariates such as education, income, or health behaviors in models risks introducing spurious associations by conditioning on shared causes. This problem inflates or obscures $G \times E$ estimates, producing results that may reflect statistical artifacts rather than genuine biological interactions (Akimova et al., 2022). This is closely linked to "genetic nurture," where parental genotypes influence offspring outcomes indirectly by shaping the household environment (Kong et al., 2018). Without within-family designs to isolate direct genetic effects, $G \times E$ estimates may capture intergenerational environmental correlations rather than genuine genetic sensitivity.

Methodological issues further complicate the use of polygenic scores (PGSs) in $G \times E$ research. Standard GWAS estimate SNP effects on conditional means, privileging variants with consistent effects across contexts ("dimmer-type") while excluding context-sensitive loci whose effects reverse under different environments ("lens-type") (Domingue et al., 2020). This creates a systematic design bias that favors stability over variability, meaning that PGSs are more likely to capture main genetic effects but

systematically underrepresent environment-dependent effects. As a result, many $G \times E$ studies may fail to detect meaningful interactions not because they do not exist, but because the genetic instruments were not designed to capture them. Recent methodological innovations aim to address these blind spots. One proposal is to distinguish between mean polygenic scores (mPGS), which summarize an individual's average expected outcome, and variance polygenic scores (vPGS), which capture inter-individual differences in environmental sensitivity or plasticity (Johnson et al., 2021). This distinction is conceptually promising, as it aligns more directly with sociogenomic theories of differential susceptibility. However, empirical work remains limited, and questions persist about their stability, predictive validity, and interpretability across different populations and exposures. Without robust validation in large and diverse datasets, these tools risk being more appealing in theory than in practice.

In synthesis, while technical advances—including machine learning (LASSO, Bayesian shrinkage) and trans-ethnic GWAS—promise to improve prediction and portability (Maier et al., 2018; Wojcik et al., 2019), they also intensify concerns regarding transparency and equity. Taken together, PGSs are both an indispensable advance and a deeply imperfect instrument. They move the field beyond candidate gene models and enable large-scale, quantitative studies of genetic influence, yet their context dependence, methodological biases, and modest predictive validity demand critical caution. Rather than treating PGS as fixed genetic markers, researchers must view them as probabilistic tools whose meaning depends on discovery samples, environmental contexts, and analytic strategies. This recognition underscores the need to complement PGS-based analyses with quasi-experimental designs and other causal inference strategies, which can more robustly disentangle genetic predisposition from environmental influence.

1.3.3 Gene-environment Correlation and Quasi-Experimental Approaches

A persistent challenge in $G \times E$ research is the conceptual issue of modelling interactions between variables that are themselves correlated. It is unknown whether a specific genetic variant or a variant associated with the specific genetic variant may be the true cause of a phenotype. Added to this inferential issue is that, for most ‘E’ examined in $G \times E$ research, the environment is not randomly assigned, may be confounded with other non-genetic variables, and could be correlated with the genetic variant in question, which is another form of gene–environment interplay, namely gene–environment correlation (rGE). There are different rGE processes including evocative, active and passive.

Evocative rGE is when a heritable trait evokes some responses from the environment. For example, if someone is naturally prone to being introverted or shy, they may appear unapproachable or aloof, which will in turn reinforce that trait through reactions to them. Active rGE is where such a heritable trait is associated with what kind of environments individuals actively select or create for themselves (Horwitz et al., 2011; McGue et al., 2005). Passive rGE is the association between the genotype a child inherits from parents and the environment where she or he is raised (Horwitz et al., 2011; Kendler, 1996). For instance, assortative mating among highly educated parents not only transmits cognitive ability genetically but also constructs learning-intensive environments, producing “genetic nurture” effects (Kong et al., 2018), suggesting that the genetic nurture effect is an indirect link between parental genotypes and children's characteristics.

Because environments are not randomly assigned and often overlap with genetic predispositions, distinguishing genuine $G \times E$ effects from rGE requires research designs that approximate randomization. One promising strategy is the use of natural

experiments and quasi-experimental approaches such as regression discontinuity, instrumental variables, and difference-in-differences (Fletcher, 2017; Fletcher & Conley, 2013). A few studies have recently contributed to incorporating quasi-experimental designs that leverage exogenous environmental variation into a $G \times E$ framework to reduce the influence of rGE and strengthen the ability to demonstrate causal inferences.

Classic examples include Fletcher (2014), who used a regression discontinuity design exploiting whether Add Health respondents were interviewed before or after the 9/11 terrorist attacks, treating the event as an exogenous stress shock and found that 9/11 increased reports of sadness overall but did not confirm the hypothesized interaction with the 5-HTT polymorphism while the variation in DRD4 buffered the depressive impact of stress. Similarly, Guo et al. (2010) leveraged legal thresholds for alcohol and smoking and found a protective effect in the dopamine transporter gene (DAT1) against risky behaviours. Lee et al. (2013) examined the effects of the Great Recession on maternal harsh parenting and found that mothers' responses to adverse economic conditions were moderated by the Taq1A polymorphism of one dopamine gene (DRD2). These studies illustrate the power of leveraging exogenous variation to clarify whether observed gene–environment interplay is genuinely causal.

Recent research has increasingly applied polygenic scores (PGSs) in combination with quasi-experimental designs, extending the scope of $G \times E$ research. Domingue and colleagues (Domingue, et al., 2017) conducted a polygenic score (PGS) study to test whether genetic scores to better subjective well-being can buffer against the risk of the development of depression following a stressful life event: spousal death. Utilizing an RDD, Domingue found that having a higher PGS for subjective well-being, which may reflect the opposite end of depression, buffered against increased depressive symptoms

following spousal death. To estimate the causal effect of genes and education on health, Barcellos, et al (Barcellos et al., 2018b) exploited the raising of the minimum school leaving age in 1972 from 15 to 16 in the UK, which induced sharp across-cohort differences in educational attainment. Using data from the UK Biobank, they combined an RDD with polygenic scores (PGSs) and found that the additional year of schooling benefited those with a higher genetic risk of obesity the most and higher educational attainment can reduce the gap in unhealthy body size between those in the top and bottom terciles of genetic risk of obesity from 20 to 6 percentage points.

Fletcher (Fletcher & Venkataramani, 2018) also provided some of the first evidence on causal gene-environment interactions using polygenic scores for cognition from the Health and Retirement Study (HRS) data. They leveraged a health policy shock – the introduction of sulfonamide antibiotics in 1937 – to achieve exogenous variation in the broader disease environment in an IV model and found that the effects of early-life exposure to pneumonia on cognitive outcomes were largest for individuals with higher genetic endowments, and null for those with lower endowments. Moreover, they conducted a placebo check to rule out the potential correlation between the cognitive PGS and other genetic and socioeconomic measures that may also modify the effects of early exposure to infant health shocks.

Using the HRS data, Schmitz and Conley (Schmitz & Conley, 2016) examined whether the impact of late-career unemployment on body mass index (BMI) can be moderated by genotype. They used job loss due to a business closure as an exogenous environmental shock to avoid confounding. They incorporated the PGS for BMI into a DID design, where they compared the body size of workers before and after the layoff with those who had not experienced job loss in later life. Results indicated that the $G \times E$ effect is

significant for genetically-at-risk individuals who were not overweight prior to job loss. For each standard deviation increase in the BMI PGS, the BMI of all individuals aged 50-61 prior to job loss increased by an average of 0.458 kg/m² after a business closure and by 0.953 kg/m² for men.

Building on these foundations, several recent studies highlight the growing sophistication of quasi-experimental $G \times E$ research. Pereira, Rietveld, and van Kippersluis (2025) instrumented the maternal smoking decision with a genetic variant (rs1051730) located in the nicotine receptor gene CHRNA3 and deal with the underreporting of maternal smoking by using a biomarker of nicotine collected during pregnancy, finding strong main negative effect of child's PGS for birth weight and maternal smoking on the child's birth weight but no evidence of meaningful interactions between PGS and maternal smoking.

Von Hinke and Sørensen (2023) examined the London smog of 1952 as early-life environmental shocks and leveraged a difference-in-differences (DiD) framework, sharpening geographic and temporal variation (a brief, severe smog event) for a convincing causal design. This study showed that individuals with higher polygenic scores for education showed larger smog-related declines in educational attainment; similarly, those with higher genetic risk for respiratory issues were more vulnerable to pollution-related respiratory disease, providing nuanced insight into the moderation effect of genetic risks for education and respiratory health on the long-term outcomes of pollution exposure.

Cheesman et al. (2022) used population-wide register data from Norway combined with polygenic scores (PGS) from the Norwegian Mother, Father and Child Cohort Study (MoBa) and incorporated multi-level modeling (family, school, and neighborhood fixed

effects), making this effectively a quasi-experimental design—since siblings, classmates, and neighbors provide natural counterfactuals for isolating $G \times E$. They found that high-quality schools can amplify genetic potential for educational achievement, while disadvantaged neighborhoods dampen it, which supports the differential susceptibility model - environments can either magnify or suppress genetic influences, rather than effects being uniform across contexts.

Biroli et al. (2024) synthesized methodological lessons from economics, emphasizing how econometric designs such as regression discontinuity and instrumental variables can sharpen causal claims in $G \times E$ research. Using data from ALSPAC, the study exploits exogenous variation in school-starting age (“being old-for-grade”) to test how it moderates the impact of polygenic indices for educational attainment. They show that while family environments may amplify inequality tied to genetic endowments, schools can play a compensatory role by reducing these disparities. Together, these studies underscore a methodological shift toward exploiting exogenous variation, moving beyond correlational evidence to more credible causal inference.

This recognition provides the central motivation for my study: by combining PGSs with quasi-experimental designs, I aim to move beyond descriptive associations and provide stronger evidence of causal $G \times E$ mechanisms. In doing so, my thesis contributes both methodologically—by demonstrating the value of leveraging exogenous variation—and substantively, by showing how genetics and environments interact in ways that can inform health, education, and social policy. In my thesis, I apply polygenic scores and quasi-experimental designs to improve causal inference for $G \times E$ research. However, there remain several persistent challenges in the study of $G \times E$ and the use of PGSs for the detection of $G \times E$. One problem is that the theoretical models underlying $G \times E$ are

not formalized, with an inability to articulate plausible biological pathways for $G \times E$, leading to a black box explanation. Another problem is the validation of PGSs because PGSs capture a broad range of information and therefore the associations between an individual's PGS and their downstream outcomes cannot be readily interpreted as the causal effect of genes due to confounding such as social genetic effect. A third issue is the inconsistent measurement of exposure to the environment or trait. When including measurements that represent exogenous environmental shocks, researchers need to carefully consider their actual exogeneity (Conley, 2017). Finally, adequate statistical power and large samples are required to detect effects in $G \times E$ models.

This thesis focuses on gene-environment interactions in late-life health outcomes, with a focus on including age-related cognitive decline and mental health. Generally, an individual's cognitive ability develops from birth, stabilises in adulthood and declines with increasing age. Mental health, such as depression and anxiety, affect a large percentage of the adult population, often with profound consequences for well-being. Understanding the mechanisms for age-related cognitive decline and mental health is crucial not only for obtaining a better understanding of these two health outcomes, maximising cognitive function and mental health or delaying the onset of cognitive decline or mental disorders but also for improving the quality of life of individuals and may benefit the wider society. Therefore, the following two sections provide a necessary overview of age-related cognitive decline and mental health, outlining their epidemiological patterns, key risk factors, evidence from quasi-experimental studies, and current approaches to detecting $G \times E$.

1.3.4 G×E in Age-Related Cognitive Function Decline

Maintaining cognitive function is one of the greatest challenges of ageing. A useful distinction between two categories of cognitive function—fluid ability and crystallized ability—was first proposed by Cattell (1963) and elaborated by Horn and Cattell (1967). Crystallized abilities, such as vocabulary, general knowledge, and numerical skills, tend to remain relatively stable or even improve with age, reflecting accumulated experience and learning (Craik & Bialystok, 2006). By contrast, fluid abilities—including memory, executive functioning, reasoning, and processing speed—decline steadily from midlife onward, as they rely on efficiency in processing novel information and cognitive flexibility (Craik & Bialystok, 2006).

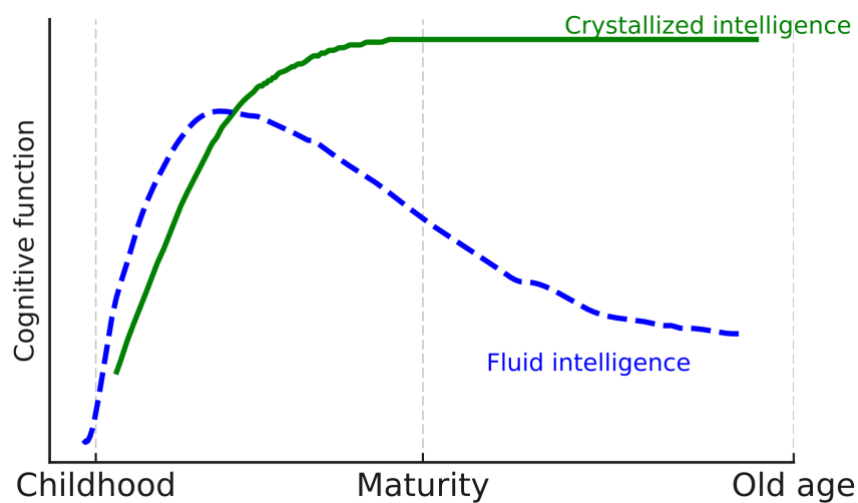


Figure 1.2 Different lifespan trajectories of fluid intelligence and crystallized intelligence

Notes: Based on Horn and Cattell (1967). Fluid intelligence (e.g., processing speed, working memory) peaks in early adulthood and declines with age. Crystallised intelligence (e.g., vocabulary, accumulated knowledge) remains relatively stable or continues to increase into later life. Both domains are examined in this thesis via memory and executive function outcomes.

Age-related decline in cognitive function - or normal (non-pathological, normative, usual) cognitive ageing - has been attributed to normative developmental processes and distinguished from the precipitous cognitive decline attributed to pathologic processes such as those underlying Alzheimer's disease (AD) and other dementias (Deary et al., 2009). Yet, evidence shows marked heterogeneity in late-life trajectories: some older adults experience sharp declines, whereas others maintain high cognitive performance well into advanced age—even in the presence of brain pathology (Boyle et al., 2021; Lindenberger, 2014; Lyons et al., 2017). This variability has fueled a large body of work on the determinants of cognitive ageing, with a view toward identifying factors that protect against decline (Harada et al., 2013).

One of the most influential explanatory frameworks is cognitive reserve (CR), which refers to the capacity of the brain to compensate for age-related changes or pathology through accumulated resources such as education, occupational complexity, and cognitively stimulating activities (Stern, 2002; Stern et al., 2023). While CR is not the central focus of this thesis, acknowledging it is important for contextualizing how social and environmental exposures, particularly education, shape resilience to decline. CR is often conceptualized alongside complementary mechanisms of brain maintenance (resisting pathology) and compensation (functional reorganization to sustain performance) (Stern et al., 2020). Together, these frameworks emphasize the role of lifelong experiences in shaping individual variability in late-life outcomes.

A substantial literature identifies genetic, biological, social, and environmental factors as contributors to this heterogeneity. For example, sex differences are well documented: women tend to perform better in episodic memory tasks, while men often score higher in visuospatial domains (De Frias et al., 2006; Pauls et al., 2013). Genetic risks are also

salient, most notably the APOE 4 allele, which predicts both steeper cognitive decline and increased conversion from cognitive decline to dementia (Hong et al., 2015). At the same time, APOE4 carriers are more likely to report cognitive decline (Krell-Roesch et al., 2015), highlighting the complex interplay between genetic predispositions and self-perceived decline.

Environmental and lifestyle factors act as powerful modifiers. Higher education and sustained engagement in cognitively stimulating activities bolster reserve and delay decline (Stern, 2012). Social engagement and meaningful roles support cognitive maintenance through enriched stimulation and stress buffering (Seeman et al., 2011). By contrast, chronic health conditions such as hypertension, diabetes, and cardiovascular disease are consistently linked to accelerated decline (Livingston et al., 2017). Unhealthy behaviors such as smoking and inactivity (Yaffe et al., 2014) further undermine resilience.

Critically, as Dixon and Lachman (2019) argue, cognitive ageing cannot be understood through single risk or protective factors in isolation. Instead, it is best conceptualized as a multidomain, interactive process, where education, health behaviors, social integration, and psychological well-being accumulate and interact across the life course. Interventions targeting multiple domains simultaneously—rather than narrow single exposures—are therefore most promising for promoting healthy cognitive ageing and mitigating decline.

Educational Attainment, Schooling and Cognition

The protective effect of educational attainment on cognitive function in older age has been well-established in the literature (Arvanitakis et al., 2004; McArdle et al., 2002; Zahodne et al., 2011). Multiple mechanisms may be at play. “Cognitive reserve” is

perhaps the most influential theory of the mechanisms, which proposes that individuals with higher levels of education and intellectually stimulating activities build up a reserve of cognitive resources in their brains. This reserve can help compensate for age-related changes and brain pathology, delaying the onset of cognitive decline (Stern, 2012; Stern et al., 2019; Stern, 2002). Furthermore, education may influence individuals' lifestyle choices, such as diet, physical activity, smoking and alcohol consumption, and can increase awareness of various risk factors for cognitive declines, such as hypertension, diabetes and depression (Clare et al., 2017; Lovden et al., 2020).

Many studies have examined the association between educational attainment and cognitive function. However, there are generally two kinds of threats to causal inference in these observational studies: (a) reverse causation, whereby an observed correlation between schooling and cognition can be explained by poor cognitive function causing lower educational attainment, rather than the opposite direction; and (b) confounding of the association between educational attainment and cognition by unobserved genetic endowments or other early life environmental conditions. The reverse causation and omitted variables that are correlated with education and cognition will lead to the problem of endogeneity (Anstey & Christensen, 2000; Card, 2001).

To address these methodological challenges, recent studies have explored the causality between educational attainment and cognitive function using quasi-experimental methods by exploiting exogenous policy shocks that lead to differences in educational attainment (Crespo et al., 2014; Mazzonna, 2014; Nguyen et al., 2016). One leading example of these policy changes is compulsory schooling laws (CSLs), which establish a minimum school-leaving age and require school-age children to attend school for a certain period.

Glymour et al. investigated state compulsory schooling laws (CSLs) in the US during 1907-1961 and used an instrumental variable (IV) approach to estimate the impact of education on cognitive outcomes at older ages (Glymour, 2008). The study found significant effects of education on memory but not mental status. Each additional year of schooling improved memory scores by 0.18 standard deviations.

Banks and Mazzonna (Banks & Mazzonna, 2012) exploited the reform of compulsory schooling in England in 1947, which raised the minimum school-leaving age from 14 to 15 for individuals born after a specific cut-off in March 1933, and used a fuzzy regression discontinuity design to compare individuals born before and after the cut-off. The study found that education improved old-age memory for both males and females as well as executive functioning for males. In the smallest sample (those who left education < 16), the effect of an additional year of education on memory scores was approximately 0.5 of a standard deviation for males and around 0.4 for females. The effect on executive functioning scores for males ranged between 0.37 and 0.63 of a standard deviation.

Schneeweis (2014) et al. (Schneeweis et al., 2014) conducted an IV analysis and exploited the compulsory schooling reforms in six European countries: Austria, the Czech Republic, Denmark, France, Germany, and Italy. The study extended the set of cognitive outcome variables including immediate and delayed memory, verbal fluency, numeracy and orientation-to-date. They found a positive causal effect of education on memory performance. One year of education increased the delayed memory score in older age by about 10% of a standard deviation. They did not find consistent effects on verbal fluency, numeracy, or orientation-to-date. Furthermore, the study provided

evidence for a protective effect of longer schooling on cognitive decline, which is captured by the change in cognitive assessment scores within two years.

Importantly, recent reviews emphasize that results across CSL-based quasi-experiments are not always consistent, reflecting both heterogeneity in study designs and the complexity of mechanisms linking education to health and cognition. Glymour and Manly (2018) argue that education influences multiple domains—including economic resources, occupational complexity, and health behaviors—such that effects may vary across populations and historical contexts. More recently, Schwartz and Glymour (2023) highlight the need to bridge life-course epidemiology with causal inference, suggesting that education's cognitive benefits cannot be understood without situating quasi-experimental evidence within broader social and developmental trajectories. Taken together, the accumulated evidence supports a robust but context-dependent causal effect of education on cognitive ageing, reinforcing the need to integrate both methodological rigor and life-course perspectives in future research.

There is substantial heterogeneity across these study designs, including the type of CSLs variation, the birth cohort included, the analytic method and the measurement of cognitive function. Moreover, it is also possible that educational attainment and cognition are nonlinearly related. For example, finishing a degree may be associated with health outcomes beyond the continuous effect of time in school, or additional education may have diminishing marginal effects on cognition at higher levels of education. Notwithstanding different study designs and some uncertainty about the exact magnitude of the association between education and cognitive function, all these studies took advantage of CSLs as exogenous policy shocks to education to identify the causal effect of educational attainment on cognitive outcomes. We can safely conclude that this

association constitutes a highly consistent and replicable finding. Together, these findings underscore that while the protective role of education on cognitive ageing is robust, it is also context-dependent, requiring methodological rigor and attention to social heterogeneity in future research. (Galama et al., 2018; Hamad et al., 2018; Lovden et al., 2020).

Genetic Factors of Cognition and Gene-Environment Interactions

Researchers have found consistent replicable results of the effects of the APOE gene of chromosome 19 in Alzheimer's disease (AD) on cognitive decline among non-demented older adults. The presence of the APOE ϵ 4 allele is significantly related to an increased risk of developing AD (Battista et al., 2016), and a greater decline in memory (Andrews et al., 2015) and speed performance (Hofer et al., 2002). Since cognitive function is a polygenic complex trait, more recent studies have used polygenic scores (PGSs) derived from GWAS to predict cognitive ability in later life. However, there are inconsistent findings regarding significant associations between PGS and cognitive decline among older adults without dementia (Harris et al., 2014; Marden et al., 2016; Mormino et al., 2016). Previous studies have shown that a PGS for cognitive function (based on a GWAS of 48462 people) predicts 1.27% of the variance in cognitive function in an independent cohort (Davies et al., 2015). One can also examine PGSs for correlated traits, for example, a higher PGS for schizophrenia is correlated with greater life-course cognitive decline (McIntosh et al., 2013). Some studies reported PGSs for AD were not associated with cognitive decline (Riaz et al., 2021), while others reported a significant effect of PGSs for AD on cognitive decline (Chaudhury et al., 2019; Marden et al., 2016). This highlights the need for further research.

Gene-environment interactions for cognition have been reported in several studies (Harden et al., 2007; Rhemtulla & Tucker-Drob, 2012; Rowe et al., 1999; Scarr-Salapatek, 1971; Tucker-Drob et al., 2011) among which the APOE gene is one of the crucial genetic risk factors for cognitive decline (Davies et al., 2014). Several studies suggested that education can moderate the relationship between APOE4 and cognitive decline (McArdle & Prescott, 2010; Seeman et al., 2005) or dementia (Arenaza-Urquijo et al., 2015). For instance, McArdle and Prescott (2010), using U.S. HRS data, found steeper memory decline for APOE4 carriers but only among individuals with eight years of schooling or less. Similarly, Wang et al. (2012) showed that higher early-life education buffered the adverse effects of APOE4 on dementia risk in Northern Europe, while Chang et al. (2021) reported that the negative relationship between APOE4 and executive function was alleviated by greater educational attainment. These findings are consistent with the cognitive reserve hypothesis (Stern, 2009), suggesting that intellectually enriching environments help individuals better withstand neuropathological burden. Yet, the moderating role of education may vary by sex, given documented gender differences in both cognition and schooling (Ishioka et al., 2016). However, the evidence on the interaction between APOE4, education, and sex is still scarce.

Building upon earlier findings, contemporary research provides more refined evidence on the interplay of genes and environments in shaping cognitive trajectories. Hsu et al. (2025) investigated whether education modifies genetic influences on late-life cognitive ability. Using polygenic scores (PGSs) and longitudinal data, they found that higher educational attainment attenuates genetic risks, particularly for fluid cognition, supporting the hypothesis that education can buffer inherited vulnerabilities. Najar et al. (2023) followed a population-representative cohort for 16 years to test whether PGSs for

Alzheimer's disease (AD) predict cognitive decline. They found that individuals with higher AD-PGSs showed significantly faster decline, but effects were moderated by baseline cognition and lifestyle factors, emphasizing the joint influence of genetic liability and long-term behavioral environments. Liu et al. (2023) conducted a longitudinal U.S. study examining PGSs, baseline cognition, lifestyle, and diabetes. Results indicated that high AD-PGS and poor baseline cognition predicted steep decline, but protective lifestyle behaviors (e.g., physical activity, non-smoking) mitigated these risks. This suggests that modifiable behaviors can offset genetic vulnerabilities. Langella et al. (2023) focused on autosomal dominant AD, examining APOE genotype and education. They found that high educational attainment was associated with better cognitive outcomes even among genetically high-risk individuals, reinforcing the idea that education provides resilience against deterministic genetic pathways.

Taken together, this literature shows that cognitive ageing is shaped by both genetic endowment and modifiable environments, with education and lifestyle emerging as especially powerful moderators of genetic risk. The evidence underscores the importance of moving beyond purely correlational studies to quasi-experimental and life-course frameworks that capture how genes and environments interact dynamically across time. For the purposes of this thesis, these insights motivate the integration of polygenic scores with quasi-experimental designs, aiming to improve causal inference on how educational attainment shapes cognitive trajectories in genetically diverse populations.

1.3.5 G×E in Mental Health

Mental health is defined by the World Health Organisation (2022c) as a state of mental well-being that enables individuals to cope with life stresses, realize their abilities, and contribute meaningfully to their communities. It possesses both intrinsic and instrumental value, underpinning our collective capacity to build relationships and navigate the world. While mental health conditions formally include a range of disorders—most notably depressive and anxiety disorders, which are characterized by persistent distress, excessive worry, and impaired functioning (World Health Organisation, 2022b)—modern research, particularly in the context of healthy ageing, increasingly views mental health as a broader continuum. This spectrum encompasses not only the presence of clinical symptoms but also the subjective social and evaluative states that define an individual's psychological resilience. Within this framework, loneliness serves as a critical social dimension of mental health, acting as a primary stressor that reflects the subjective distress of social isolation (Beutel et al., 2017). Furthermore, subjective well-being and quality of life (QoL) provide a holistic assessment of an individual's overall functioning and life satisfaction (Diener et al., 2018). In this thesis, these four indicators—depression, anxiety, loneliness, and QoL—are examined in tandem to capture the comprehensive psychological toll of exogenous environmental shocks, moving beyond a purely clinical focus to a multidimensional evaluation of mental health trajectories.

Multiple individual, genetic, social and environmental determinants may combine to protect or undermine our mental health and shift our position on the mental health continuum throughout the life course (World Health Organisation, 2022b). Genetics and individual factors such as emotional skills and substance use can make people more

vulnerable to mental health problems. Exposure to unfavourable social, economic, geopolitical and environmental circumstances – including poverty, violence, pandemic, inequality and environmental deprivation – also increases the risk of experiencing mental disorders. Older people with mental disorders have poorer functioning compared to those with chronic medical conditions such as lung disease, hypertension or diabetes. Mental disorders also increase the perception of poor health, the utilization of health care services and costs. Moreover, mental disorders can also have a large impact on families, caregivers and communities. Therefore, it is important to strengthen and promote mental health in older adults and to integrate effective strategies into policies and plans for healthy, active ageing, and for preventing mental illness.

Mental health is multifactorial in aetiology, influenced by both genetic and environmental factors (Plomin et al., 1994) and behavioural genetics research also suggests accounting for the interaction effect between genetic and environmental factors rather than independent main effects on psychiatric disorders (Assary et al., 2018; Uher, 2014). Caspi et al. were the first to examine the interaction between a specific candidate gene and adverse environmental influences on psychiatric disorders (Caspi et al., 2002) (Caspi et al., 2003). They found that the association of life stress with depression was moderated by a polymorphism in the 5-HTT gene, and individuals with one or two copies of the short allele of the 5-HTT promoter polymorphism exhibited a higher risk of depression and suicidality in relation to stressful life events than individuals homozygous for the long allele (Caspi et al., 2003). However, these early hypothesis-driven studies adopted a bottom-up approach which requires a strong biological hypothesis to select appropriate candidate genes, but knowledge regarding specific biological mechanisms underlying these mental disorders is limited. Therefore, the risk of selecting inappropriate candidate genes is high (Bosker et al., 2011) (Collins et al.,

2012). In addition, candidate gene studies have low statistical power, unsuccessful replications, and publication bias toward positive findings (Karg et al., 2011; Munafo et al., 2009; Risch et al., 2009; Uher & McGuffin, 2010), which are “wrong or misleading” (Hewitt, 2012).

In contrast to candidate gene studies, genome-wide association studies (GWASs) can take a hypothesis-free approach that does not require any a priori assumptions regarding the role of specific genes in a disorder (Kwon & Goate, 2000). More broadly, results from GWASs have profoundly reshaped our understanding of the genetic architecture of mental disorders in two important ways. First, previous estimates of the effect size associated with single variants were far too optimistic, and individual common variants are associated with small effects [e.g., odds ratio (OR) < 1.1] (Caspi et al., 2003) (Caspi et al., 2002) (Egan et al., 2001) (Thomasson et al., 1991). Second, mental disorders are a complex and highly polygenic trait, which are influenced by multiple genetic variants of small effects throughout the genome rather than a few selected candidate genes of large effects (Donnelly, 2008). Many risk loci have emerged from genome-wide association studies (GWAS) of, among others, major depression, anxiety, schizophrenia, bipolar disorder, and attention-deficit/hyperactivity disorder from the Psychiatric Genomics Consortium (PGC) and other efforts (Sullivan et al., 2018). The additive effects of independent and commonly occurring variants, when weighted by their association with a disorder or related construct in an adequately powered GWAS, reliably predict variance in the same and related phenotypes. These additive scores, which we refer to as polygenic risk scores (PGSs), along with other polygenic approaches (e.g., pathway-based analyses, machine learning, linkage disequilibrium (LD) score regression) hold tremendous potential to contribute to our aetiologic understanding of psychology and

the quantification of individual risk, both of which may ultimately advance the characterization and treatment of mental disorder (Bogdan et al., 2018).

Recent GWAS-derived heritability ($SNP-h^2$) estimates show that common genetic variants account for a modest but significant portion of liability for psychiatric traits. For schizophrenia, $SNP-h^2$ is estimated at around 25%, compared with total heritability of ~60–80% from twin studies (Owen et al., 2023). Major depressive disorder (MDD) shows lower $SNP-h^2$ —about 8.4% on the liability scale (Adam et al., 2025) — underscoring a substantial 'missing heritability'. Notably, these broader mental health constructs possess a distinct polygenic architecture similar to clinical disorders; for instance, the $SNP-h^2$ for subjective well-being is estimated at 5–10% (Okbay et al., 2016), while loneliness exhibits a comparable genetic influence of approximately 4.2% (Abdellaoui et al., 2019). Broader latent traits further illustrate this gradient, with GWAS-based heritability estimated at 8.1% for depression, 29.1% for schizophrenia, and 27.4% for bipolar disorder (Dattani et al., 2023).

Across both clinical and non-clinical outcomes, GWAS heritability estimates are consistently lower than those derived from twin studies. This discrepancy is in part attributed to many common variants of extremely small effect sizes that elude detection, rare variants to which GWAS are insensitive, and rare structural variations. Yet a further possibility is that GWAS, which are based on combining data from multiple periods, countries, and birth cohorts, will introduce heterogeneity in genetic effects as a function of environmental factors to which some but not all individuals are exposed, implying a gene-environment interaction (GxE) (Conley, 2016a; Tropf et al., 2017). Therefore, it should ultimately be possible to account for additional missing heritability across the mental health spectrum by examining how environmental exposures modify genetic risk

(GxE), justifying the use of polygenic scores (PGS) for both symptomatic and evaluative mental health outcomes to capture the full range of genetic sensitivity to environmental change.

Using polygenic scores in the $G \times E$ framework, many studies have examined gene-environment interactions and depression. Colodro-Conde et al. (Colodro-Conde et al., 2018) found that the PGS for major depressive disorders (MDD) significantly interacted with stressful life events (SLEs) to increase the risk of depression. Iob and Steptoe (Iob et al., 2021) similarly found that the associations between adverse childhood experiences (ACEs) and depressive symptoms were larger among older adults with a higher PGS for MDD (OR 1.13, 95% CI 1.04–1.23), in line with the diathesis–stress model (Monroe & Simons, 1991; Robins & Block, 1989). In contrast, one study showed that individuals exposed to severe childhood trauma were at the greatest risk if they also had a *low* PGS for depression (Mullins, 2016). Others reported independent but not interactive associations of ACEs and PGSs with depression (Musliner et al., 2015; Peyrot et al., 2018). A few $G \times E$ studies employed polygenic scores to examine other mental disorders. One study derived a polygenic score of environmental sensitivity from the psychiatric genetics consortium in 1,026 monozygotic twin pairs and found that parenting had little effect on emotional problems for those with a low PGS. However, in those with a high PGS, negative parenting was associated with increased emotional problems, while positive parenting was associated with decreased emotional symptom scores (Keers et al., 2016). This finding implies that individuals with the greatest environmental sensitivity may be more likely to develop mental disorders in adverse environments but also benefit more from favourable environments, supporting the differential-susceptibility model (Belsky & Pluess, 2009). Another study explored both depression and well-being outcomes and reported significant independent effects of the

PGS for depression and peer victimisation in predicting the risk of depressive symptoms and well-being in early adulthood, but no clear interaction effects (Armitage et al., 2022).

Using quasi-experimental designs, Domingue et al. (Domingue, 2017) tested whether higher subjective well-being polygenic scores buffered survivors' risk of developing depressive symptoms following the death of their spouse from 8,588 non-Hispanic white adults in the Health and Retirement Study (HRS). This study found that individuals who carry more well-being-associated alleles tended to develop fewer additional depressive symptoms following the death of a spouse. While spousal death is not exogenous, the genetic buffering was not affected by any gene-environment correlation between the genetics of subjective well-being and exposure to the death of a spouse; HRS respondents' subjective well-being polygenic scores were unrelated to their risk of losing a spouse during follow-up. Fletcher (Fletcher, 2014) exploited a quasi-experimental variation in the environmental exposure (whether a respondent was interviewed before or after the 9/11 attacks) using Wave 3 of the National Longitudinal Study of Adolescent Health (Add Health). This study showed that the effects of this external stressor were moderated by two candidate genes. These research designs can be adapted to other causal inferences in gene-environment interaction studies for other specific life stressors, such as the COVID-19 pandemic.

1.3.6 Sex Differences in G×E: Biological and Social Perspectives

The systematic exploration of sex differences constitutes a core contribution of this thesis, moving beyond treating sex as a mere covariate to recognizing it as a fundamental moderator of both cognitive and mental health dynamics. The rationale for

conducting sex-stratified analyses is rooted in the complex interplay between biological dimorphism and divergent social-structural experiences across the life course.

Biological Foundations: Chromosomes, Hormones, and Genetic Architecture

From a biological perspective, sex differences are not merely phenotypic but are encoded at the most fundamental levels of cellular and systemic organization. This biological dimorphism provides the primary justification for stratifying G×E analyses, as the "G" (genetic risk) and the "E" (environmental trigger) often interact within entirely different physiological contexts for men and women.

First, sex differences are rooted in sex chromosomes (X and Y). Beyond the presence of Y-linked genes in males, the biological reality for females involves X-chromosome inactivation (XCI). However, research indicates that approximately 15% of genes on the human X chromosome escape inactivation, leading to higher expression doses of certain genes in females compared to males (Khramtsova et al., 2019). This "dosage effect" can fundamentally alter the baseline genomic variance and the subsequent polygenic risk for neurodevelopmental and neurodegenerative traits.

Second, sex hormones—primarily estrogens and androgens—act as powerful epigenetic and transcriptional regulators. Estrogens, for instance, have neuroprotective properties and can directly modulate the transcriptional activity of genes involved in the stress response. A key mechanism is the hormonal regulation of the human corticotropin-releasing hormone (CRH) gene, a central driver of the Hypothalamic-Pituitary-Adrenal (HPA) axis (Vamvakopoulos & Chrousos, 1994). Estrogen response elements (EREs) located in the promoter regions of various stress-related genes mean that the same environmental stressor (E) may trigger a significantly different transcriptional response (G) depending on the hormonal milieu.

Finally, large-scale genomic analyses have identified that the genetic architecture of complex traits—including cognitive function and psychiatric outcomes—exhibits significant sexual dimorphism. This "sex-specific genetic architecture" implies that the SNP-based heritability and the specific loci contributing to a Polygenic Score (PGS) may vary by sex (Khramtsova et al., 2019). Consequently, a pooled G×E analysis risks "cancelling out" effects that are potent in one sex but absent in the other, necessitating a stratified approach to capture the true magnitude of genetic sensitivity to environments.

Social and Policy Perspectives

Beyond innate biology, the social environment is profoundly gendered, leading to distinct exposure patterns and life-course trajectories. Men and women are often subject to different societal expectations, which shape their "environmental landscape" long before an exogenous shock occurs.

These gendered experiences are often reinforced by macro-level policies and structural shifts. For example, historical changes in educational policies, such as reforms in compulsory schooling laws, may yield differential cognitive returns for men and women due to the gendered nature of labor market opportunities and social roles prevalent at the time of the reform. Similarly, broad environmental shocks—ranging from economic recessions to global health crises—do not interact with a monolithic population. Instead, they hit individuals whose resilience is already pre-conditioned by a lifetime of gendered socialisation and resource access (Kuehner, 2003).

Empirical Necessity of Stratified G×E Analysis

The necessity of this stratified approach is underscored by foundational research in Polygenic Score (PGS) dynamics. Empirical evidence suggests that the interaction

between genetic risks and environmental factors (such as socioeconomic deprivation or life events) frequently exhibits sex-specific patterns (Assary et al., 2020; Silberg et al., 2001). These findings reflect what is increasingly recognized as a "gene-environment" causation process that is inherently contingent upon biological sex (Uher & Zwickler, 2017). By adopting a sex-stratified analytical strategy, this thesis aims to uncover the nuanced ways in which biological sex moderates the impact of both individual-level stressors and broad environmental shifts on health trajectories in later life.

1.4 Theoretical Framework

The theoretical framework of this thesis integrates the life-course perspective with theories of the social determinants of health to examine the dynamic interplay between genetic predispositions and environmental exposures. This approach emphasizes that health in later life is not a static outcome but a cumulative product of biological, behavioural, and psychosocial pathways operating across key developmental periods—from childhood and adolescence into old age (Ben-Shlomo & Kuh, 2002; Whalley et al., 2006).

The Life-Course Paradigm and the Life Course Cube

The theoretical framework of this study is grounded in the life-course paradigm articulated by Glen H. Elder Jr. (1998) and further operationalized through the “life course cube” proposed by Bernardi, Huinink, and Settersten (2019). Elder’s paradigm provides four foundational principles: *human agency*, *historical context*, *the timing of experiences*, and *linked lives*. These principles highlight that individuals shape their trajectories within the constraints of historical eras and settings, such as policy reforms or global crises (Elder et al., 2003).

To translate these principles into an analytical structure, the Life Course Cube offers a three-dimensional heuristic for analyzing lives across intersecting axes: Time (past experiences and future trajectories), Life Domains (e.g., education, health), and Levels of Analysis (ranging from micro-biological predispositions like genetics to macro-structural policies). Time axis encompasses past experiences, present conditions, and future trajectories, recognizing both cumulative and time-specific effects. In the context of this thesis, it allows for the tracking of health from early-life exposures to late-life

outcomes. Life domains capture various spheres of human activity—such as family, education, employment, and health—through which resources and risks are organized. By focusing on the education and health domains, this thesis isolates specific environments that may interact with biological traits. Levels of analysis range from micro-biological predispositions (e.g., genetics) to individual actions, and up to supra-individual structures such as institutions and macro-policies. This multidimensional model is particularly suited for studying gene–environment interplay ($G \times E$) as it facilitates the integration of biological data within complex social and historical contexts. The "G" (Genetic risk) is situated at the micro-biological level, while the "E" (Environment) is sourced from supra-individual levels (such as the 1947 education reform or the COVID-19 pandemic) and life domains (education and health). The interaction between them is then traced along the "Time" axis, revealing how the expression of genetic liability is conditionally moderated by an individual's location within the cube.

Core Life-Course Mechanisms: Integrating Social Determinants and Biological Outcomes

Within this paradigm, three specific mechanisms bridge the gap between structural social determinants and individual biological outcomes, explaining how "place and circumstance" get under the skin over time:

Cumulative Advantage/Disadvantage (CAD) and Social Stratification: This model suggests that health inequalities in later life are the product of systemic social stratification. According to the CAD model, early-life socioeconomic determinants—such as parental wealth and neighborhood environment—are not static; they are magnified over time through recursive processes (Dannefer, 2003). From a SDoH

perspective, this represents the "accumulation of risk," where sustained exposure to favorable or unfavorable social conditions creates divergent health trajectories. By the time individuals reach old age, these longitudinal social gradients manifest as significant biological heterogeneities in cognitive and mental health.

Sensitive Periods and Structural Interventions: This mechanism identifies specific ontogenetic windows, such as adolescence, where the individual exhibits heightened biological plasticity to shifts in social determinants. Structural factors, such as educational policies or schooling reforms, act as powerful social determinants during these periods (Kuh et al., 2003). When a macro-level intervention (e.g., the 1947 reform) alters the educational landscape, it effectively modifies the "social dosage" an individual receives during a sensitive window. This interaction determines the development of "cognitive reserve," illustrating how timely social investments can pivot a life trajectory toward resilience.

Diathesis-Stress and Acute Environmental Shocks: While CAD focuses on chronic accumulation, the Diathesis-Stress model addresses how acute social determinants—such as global pandemics or economic crises—interact with underlying vulnerability. This model posits that genetic liability (diathesis) moderates the impact of environmental stressors, where individuals with fewer social resources and higher genetic risk suffer disproportionately from acute shocks (Caspi et al., 2002). Within the SDoH framework, this explains the "differential vulnerability" of populations: an exogenous social crisis (like COVID-19) does not strike a level playing field but instead activates latent biological risks more severely in socially marginalized groups.

Situating the Empirical Chapters within the Framework

Each empirical study in this thesis occupies a distinct "slice" of the Life Course Cube, moving from established social patterns to causal G×E interactions:

Chapter 2 (The Baseline): Situated within the “time” axis and “health” domain, this study identifies longitudinal trajectories of cognitive and mental health. By applying the CAD model, it establishes a phenotypic baseline, demonstrating how socioeconomic factors shape health inequalities over time without yet considering genetic factors.

Chapter 3 (Sensitive Period Intervention): Positioned at the intersection of the "education" domain and "biological" level, this study utilizes the 1947 schooling reform as a natural experiment. It exemplifies Elder’s principle of "lives embedded in historical time" and tests the Sensitive Periods hypothesis, examining how structural changes in adolescence interact with genetic risks (e.g., APOE 4).

Chapter 4 (Acute Environmental Shock): Located as a historical rupture in the "time" axis, this study examines the COVID-19 pandemic. It applies the Diathesis-Stress model to evaluate how an acute, supra-individual shock interacts with polygenic scores (PGS) to alter mental health stability.

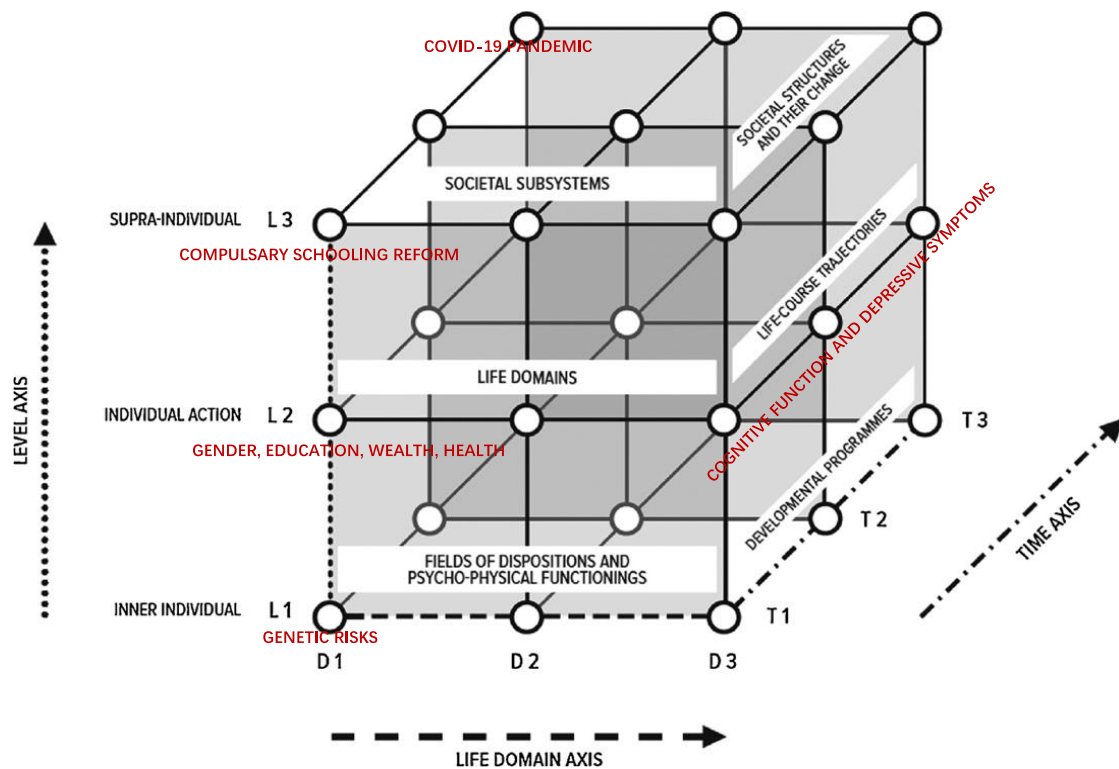


Figure 1.2 Situating the Thesis Within the Life Course Cube Framework

Note: this figure presents this integrated theoretical framework and illustrates how these studies are situated within the life course cube, integrating the life-course perspective with genetic risk to form the conceptual basis for the hypotheses guiding this thesis

By integrating these frameworks, this thesis addresses several critical gaps in the existing literature. First, it moves beyond cross-sectional or static $G \times E$ research by situating genetic risks within a longitudinal, life-course context. Second, it addresses the endogeneity of environmental exposures by leveraging quasi-experimental designs within the life-course cube. The primary contribution of this thesis lies in its holistic mapping of ageing. It demonstrates that gene-environment interaction is not a fixed biological destiny but is continually moderated by the timing, the duration, and the intensity of social environments. Ultimately, this integration provides a robust evidence base for identifying vulnerable subgroups and informing targeted interventions in an ageing society.

1.5 Thesis Aims and Hypotheses

To summarise, it is important to understand risk factors for cognitive ageing and mental health among older adults for healthy and active ageing in late life. Central to this thesis is the exploration of Gene-Environment Interaction ($G \times E$), positing that health outcomes in later life are not merely the sum of nature and nurture, but the result of their dynamic and reciprocal interplay across the life course. Complex traits like cognitive ageing and mental health are multifactorial and polygenic. Risk factors range from genetic polymorphisms such as APOE genotype and polygenic scores for specific traits to potentially more modifiable social and environmental factors throughout the life course such as educational attainment in early life, and stressful life events in late life. More recently, the possibility of interactions between genetic characteristics and other, potentially modifiable, characteristics of the individual or environment has gained attention and will offer opportunities for a more comprehensive understanding of the complex process of cognitive ageing and mental health. The results from applying GWAS data and quasi-experimental designs can be used to improve causal inference for $G \times E$ research.

The overall aim of this project is to advance understanding of the social and environmental determinants of health in later life, with a particular focus on gene–environment interactions. Specifically, it examines:

1. The trajectories of two key late-life health outcomes—cognitive function and mental health—and their predictors, to establish a phenotypic baseline and to demonstrate how socioeconomic predictors shape health inequalities over time without yet considering genetic factors.

2. Education as a lifelong socioeconomic influence, both directly and indirectly, on cognitive function, with attention to gene–environment interplay.
3. The COVID-19 pandemic as a stressful life event affecting mental health and its interaction with genetic risk.

Building on the theoretical framework and prior literature, this thesis employs quasi-experimental designs in combination with gene–environment (G×E) analysis to test the following hypotheses:

- I. Memory function is expected to decline, while depressive symptoms will increase among older adults. Considerable heterogeneity will exist in these trajectories. The combined influence of sociodemographic disadvantages (e.g., older age, female sex, early-life adversity) and adverse health or life circumstances (e.g., poor self-rated health, lack of paid work, disability, multimorbidity) is anticipated to accelerate memory decline and exacerbate depressive symptoms.
- II. Increased years of education has a positive effect on later life cognitive function, and this effect varies by polygenic risk for low cognitive function or dementia. Specifically, higher genetic risks (e.g. APOE4 carrier/ lower PGSs for general cognition/ higher PGSs for Alzheimer’s disease) may dampen the positive effect of increased years of education on late-life cognition.
- III. The COVID-19 pandemic, as a stressful life event in late life, has a negative impact on mental health among older people, and this impact will be moderated by genetic risks for mental disorders. In particular, the negative effect of the pandemic may be more pronounced among individuals with higher genetic risks (e.g. higher PGSs for depressive symptoms/ major depressive disorders /anxiety disorders /loneliness).

1.6 Data and General Methodology

1.6.1 Benefits of Studying the Context of the United Kingdom

The increased longevity of the population is one of the greatest achievements of the modern age but is also the source of some concern from the perspectives of health, economic and social policy. The UK population is now living longer than ever before. In mid-2043, there are projected to be many more people at older ages. This partly reflects the 1960s baby boomers now being aged around 80 years but also general increases in life expectancy. In mid-2018, there were 1.6 million people aged 85 years and over; by mid-2043, this is projected to nearly double to 3.0 million (Nash, 2018). If older people are experiencing these extra years in good health and live in a supportive environment, their ability to do the things they value will have few limits. However, if these added years are dominated by rapid declines in health and mental capacity, the implications for older people and society as a whole are much more negative.

Unfortunately, there are increasing numbers of older adults living with low levels of life satisfaction and well-being. The prevalence of depressive symptoms in the UK ranged from 11.3% for mild, to 3.3% for severe symptoms, and the prevalence of probable depressive disorder was 7.5% (Kuan et al., 2023). Based on primary care records, the prevalence of documented mild cognitive impairment was approximately 0.22% between 2019 and 2022 in the UK, although this figure likely reflects substantial underdiagnosis in routine practice (NICE, 2022). Therefore, ensuring the best possible health in older age is crucial if we are to achieve sustainable development in the UK. There are some advantages when studying the UK context. The policy changes in the UK context enable us to use natural experiments, such as compulsory schooling laws and clear lockdowns and restrictions in response to the COVID-19 pandemic, to

examine the effect of biological exposures or social interventions. Moreover, there are high-quality long-term data in the UK context.

Compulsory Schooling Laws in Britain

The laws governing the length of compulsory education in Britain are national, which is a contrast to US state-regulated compulsory school laws. The British compulsory schooling laws specify the maximum age by which children must start school and the minimum age at which children can leave school. The maximum age by which children must start school is currently five. The 1944 Education Act raised the minimum school leaving age from 14 to 15 and a Ministerial order later in 1944 specified that it would be raised on April 1, 1947. The 1972 Act raised the age to 16 on September 1, 1972. The 1947 change meant that students could not leave school until partway through grade nine. Whilst children born before 1 April 1933 could leave school when they turned 14, the 1947 reform stated that those born after that date could not leave until they turned 15. This law change affects a large proportion of the population (Oreopoulos, 2006). Using the ELSA data described in more detail below in Section 1.6.2, Figure 3.1 in Section 3.3 of Chapter 3 shows the fraction of males and females who left full-time education by the age of 14. It is evident that before the reform about 60% of the population left the school at the age of 14, whilst after its implementation this proportion dramatically decreased to below 10%.

UK Government Coronavirus Lockdowns and Restrictions

In the wake of the World Health Organization (WHO) declaring COVID-19 a public health emergency of international concern, hundreds of laws have been made in the UK in response to the pandemic. (1) First national lockdown (March to June 2020): England was in national lockdown between late March and June 2020. All “non-essential” high

street businesses were closed, and people were ordered to stay at home. Starting in May 2020, the laws were slowly relaxed. (2) Minimal lockdown restrictions (July to September 2020): Most lockdown restrictions were lifted on 4 July 2020. (Lai et al., 2017) Reimposing restrictions (September to October 2020): On 14 September, the “rule of six” – indoor and outdoor social gatherings above six were banned in England. On 14 October, the Government rationalised local restrictions by introducing a “three-tier system”. At first, most of the country was placed in the least restrictive tier one, which had similar restrictions to the previous national rules. As time went on, more of the country was placed in the higher two tiers. (4) Second national lockdown (November 2020): On 5 November, the second national lockdown came into force in England. Non-essential high street businesses were closed, and people were prohibited from meeting those not in their “support bubble” inside. Reintroducing a tiered system (December 2020): On 2 December, the tiered system was reintroduced with modifications. On 19 December, the Prime Minister announced that Tier 4 would be introduced, which were like those imposed during the second national lockdown. Christmas mixing rules were tightened. (6) Third national lockdown (January to March 2021): National restrictions were reintroduced for a third time on 6 January, which were more like those in the first lockdown. People were once again told to stay at home but could still form support bubbles. (7) Exiting from lockdown: On 8 March 2021, England began a phased exit from lockdown, known as the roadmap out of lockdown. England moved through the roadmap as planned but step four was delayed by four weeks to allow more people to receive their first dose of a coronavirus vaccine.

1.6.2 The English Longitudinal Study of Ageing (ELSA)

This thesis used data from the English Longitudinal Study of Ageing (ELSA), which was designed as a sister study to the Health and Retirement Study (HRS) in the USA and was set up to document the experience of ageing in England in the 21st century. It provides high-quality longitudinal data that can be used to investigate changes in economic circumstances, social status, physical and mental health, cognitive function, social relationships and biology, as people prepare for and move into retirement and old age.

ELSA was selected as the primary data source for this research over other large-scale British datasets, such as the UK Biobank (UKB) or Understanding Society (UKHLS), for several strategic reasons. First, while the UK Biobank offers a superior genetic sample size ($n \approx 500,000$), its recruitment (2006–2010) and follow-up primarily rely on administrative health linkages rather than the frequent, face-to-face longitudinal phenotypic measurements required for the dynamic trajectory analysis in Chapter 2. Second, unlike the UKHLS (Understanding Society), which covers a broad age range from 16 to 100+ years to study general household dynamics, ELSA is specifically tailored to the 50+ population. This focus ensures that the cognitive and mental health instruments used in ELSA are age-validated and sensitive enough to detect subtle declines in later life. Third, ELSA's rigorous harmonization with the HRS allows for the application of validated polygenic scoring (PGS) methodologies, ensuring the international comparability of my G×E findings.

The ELSA is a panel study of a representative cohort of men and women living in England aged 50 and over living in private households in England. The original sample was drawn from the Health Survey for England (Schikowski et al., 2015) in 1998, 1999

and 2001 which includes an initial sample of about 12,000 people ranging in age from 50 to 100 years. The study commenced in 2002, and the sample has been followed up every 2 years. To address the specific research questions of this thesis, different combinations of ELSA waves were utilized across the empirical chapters, as summarized in Table 1.1. More than 18,000 respondents have been interviewed so far and the sample has been refreshed four times now, at Wave 3, 4, 6, 7 and 9. Data are collected using computer-assisted personal interviews and self-completion questionnaires, with additional nurse visits for the assessment of biomarkers every 4 years. The original sample consisted of 11 391 members ranging in age from 50 to 100 years. ELSA is harmonized with ageing studies in other countries to facilitate international comparisons and is linked to financial and health registry data (Steptoe, 2013).

Table 1.1 Summary of ELSA Waves Used Across Empirical Chapters

Empirical Study	ELSA Waves Used	Data Collection Period	Analytical Focus
Chapter 2	Waves 4 – 9	2008/09 – 2018/19	Longitudinal trajectories of cognitive and mental health over a 10-year window.
Chapter 3	Waves 1 – 9	2002/03 – 2018/19	Quasi-experimental G×E analysis of the 1947 schooling reform (Executive function restricted to W1–5).
Chapter 4	Wave 9 & COVID-19 Waves 1–2	2018/19 & 2020	Impact of the pandemic on mental health stability moderated by genetic risk (PGS).

For the genetic analysis, the ELSA provided the genetic data of a wide range of age-related traits and outcomes in association with the genetic factors in the sample of older people born in England. The genome-wide genotyping was performed in 2013-2014 at the University College London. This involved genotyping of 7,597 ELSA participants of European ancestry using the Illumina HumanOmni2.5 BeadChips (HumanOmni2.5-4v1, HumanOmni2.5-8v1.3), which measures ~2.5 million markers that capture the genomic variation down to 2.5% minor allele frequency. Genotyping was performed in two batches. Allele frequencies were compared between the batches after filtering for 5% of missingness. The correlation was calculated between the batches for a number of chromosomes and exceeded 99%. After post-genotyping quality assurance, such as excluding ethnic outliers (self-reported) and duplicates, the GWAS data was available for a total of 7,412 ELSA participants and 2,230,767 SNPs. Using methods employed in the Health and Retirement Study, the ELSA carried out robust quality control of the genotyped genetic data. In total, a final sample of 7,183 participants (96.9% of 7,412 original cohort) of European ancestry and 1,372,240 (61.5% of 2,230,767) variants remained after quality control. In ELSA, the methods employed for creating polygenic scores (PGSs) are those outlined by the Health and Retirement Study (HRS), in order to harmonise the research across age-related longitudinal studies by adopting a consistent methodology for creating PGSs. PGSs for each phenotype are based on a single, replicated GWAS. PGS data and detailed report describing the methods employed and a full list of PGSs available in ELSA can be found on the ELSA website (ELSA Project, 2023). To provide further details on the construction steps used to generate the PGSs and Principal Components (PCs), Section 1.6.4 presents these details.

Moreover, the ELSA responded agilely to COVID-19 and enabled research into the social and health impacts of the pandemic. During the COVID-19 outbreak, ELSA

conducted two additional waves of collection to measure the socioeconomic effects and psychological impact of the pandemic on the older population of England. All participants for the ELSA COVID-19 Substudy were selected from the existing ELSA sample. The first wave of the COVID-19 Substudy, which took place in June-July 2020, achieved 7040 completed interviews; the second wave, which took place in November-December 2020, achieved 6794 completed interviews. These pandemic waves are unique as they can be benchmarked against nearly two decades of pre-pandemic baseline data (Waves 1-9), providing a longitudinal depth that UK Biobank and other cross-sectional cohorts cannot replicate for this age group. In the COVID-19 substudies, mental health was assessed using the Center for Epidemiologic Studies Depression Scale, the Generalised Anxiety Disorder 7 assessment, and well-being measures. The COVID-19 questions concerning the experience related to COVID-19, health issues and access to health care, mental health (e.g. depression and anxiety), health behaviours (e.g. smoking, drinking alcohol, physical activity, eating behaviours, sedentary behaviours, and sleep), demographics, financial circumstances, work, social connections and isolation.

1.6.3 Measurement of Outcome Variables

This section provides a comprehensive exposition of the primary outcomes throughout this thesis. Given that several core outcomes—such as memory performance and depressive symptoms—are analysed across multiple empirical chapters (Chapters 2, 3, and 4), they are detailed here in their entirety to ensure methodological transparency and to establish a unified reference point.

By consolidating these definitions within this foundational chapter, the thesis aims to maximize structural efficiency and avoid redundant descriptions in subsequent empirical

sections. Consequently, the following chapters will provide only concise summaries of these variables, cross-referencing this section for specific details on scaling, clinical thresholds, and longitudinal standardisation protocols. This approach not only ensures that the same measurement constructs are applied consistently across different research questions but also allows the empirical chapters to focus more effectively on their respective analytical models and results.

Cognitive Function: Memory and Executive Function

The cognitive assessment module in ELSA was designed to capture multiple dimensions of cognitive health. This study focuses on two primary domains that are most sensitive to age-related decline and genetic influence.

Memory Performance (Waves 1–9)

Memory is a cornerstone of cognitive aging research and serves as a sensitive early biomarker for neurodegenerative pathologies. In ELSA, memory was assessed using a Word-list Learning Test. During the interview, a computer-generated list of 10 common nouns was read to the participant. The assessment comprised two phases: 1) Immediate Recall: The participant was asked to recall as many words as possible immediately after the presentation. 2) Delayed Recall: After a latency period of approximately five minutes—during which the respondent completed other non-cognitive survey items—the participant was again asked to recall the original list.

An aggregate Composite Memory Score (range 0–20) was constructed by summing the scores from both phases. This approach stabilizes the measurement by accounting for both short-term encoding and long-term retrieval capabilities. To facilitate longitudinal

modeling and G×E interactions, these raw scores were Z-score standardized within each wave to a mean of 0 and a standard deviation of 1.

Executive Function (Waves 1–5)

Executive function, reflecting the ability to organize thoughts and inhibit irrelevant information, was operationalized through two distinct tasks: 1) Semantic Verbal Fluency: Participants were tasked with naming as many different animals as possible within a one-minute interval. This measures the fluidity of access to the semantic network and frontal lobe integrity. 2) Letter Cancellation: A pencil-and-paper task requiring participants to scan a grid of random letters and cross out specific targets (P and W) under time pressure (one minute). This evaluates visual search speed, sustained attention, and processing accuracy.

The Executive Function Index was derived by aggregating the standardized scores of both tasks. Due to changes in survey administration, this composite is primarily utilized in analyses spanning Waves 1 through 5 to maintain instrumental fidelity (Banks et al., 2006; Steptoe et al., 2013).

Mental Health and Psychological Wellbeing

While mental health is formally defined by clinical disorders—most notably depressive and anxiety disorders characterized by persistent distress and impaired functioning (World Health Organization, 2022)—modern gerontological research increasingly views psychological health as a broad continuum. This spectrum extends beyond the binary presence or absence of clinical symptoms to encompass the subjective social and evaluative states that underpin an individual’s psychological resilience and overall flourishing. Within this multidimensional framework, this thesis examines four distinct

yet interrelated indicators to capture the comprehensive psychological toll of exogenous environmental shocks. Depression and anxiety represent the traditional clinical axes of mental distress. Loneliness is integrated as a critical social dimension of mental health, serving as a primary stressor that reflects the subjective distress resulting from social isolation (Beutel et al., 2017). Finally, Quality of Life (QoL) and subjective well-being offer a holistic assessment of an individual's evaluative functioning and life satisfaction (Diener et al., 2018). By analyzing these four measures in tandem, this study moves beyond a narrow clinical focus to provide a robust evaluation of mental health trajectories in the aging population.

Depressive Symptoms (CES-D 8)

Depressive symptoms were assessed using the 8-item version of the Center for Epidemiological Studies Depression Scale (CES-D 8). Originally developed as a 20-item self-report tool (Radloff, 1977), the CES-D has undergone various structural modifications to facilitate large-scale longitudinal research. In the ELSA, the CES-D 8 has served as the standardized and consistent measure of depressive symptoms throughout its history. The decision to utilize the 8-item version, rather than the 10-item or original 20-item versions, is informed by a rigorous methodological rationale centered on longitudinal integrity and cross-national harmonization:

1) Longitudinal Consistency and Data Integrity: The primary justification for adopting the CES-D 8 is the necessity of maintaining instrumental fidelity across the study's timeline. A meticulous inspection of ELSA survey instruments confirms that the CES-D 8 items represent the consistent core administered across Wave 1 through Wave 9, as well as the specific COVID-19 Substudy waves (Steptoe et al., 2013). Crucially, to minimize respondent burden during the pandemic, the COVID-19 Substudy exclusively

administered these eight core items. Therefore, adopting the CES-D 8 is the only viable strategy to ensure strict longitudinal measurement invariance and avoid "measurement artefacts"—spurious changes in scores resulting from shifts in scale dimensionality rather than true phenotypic variations (Zaninotto et al., 2022).

2) Cross-National Harmonization and the "HRS" Standard: The CES-D 8 serves as the global "gold standard" for the international network of ageing studies. ELSA was designed to be directly comparable with the U.S. Health and Retirement Study (HRS) and the Survey of Health, Ageing and Retirement in Europe (SHARE). Methodological documentation indicates that the transition to the 8-item version was a strategic refinement to reduce cognitive burden and survey fatigue among older respondents (Steffick, 2000). Research confirms that the CES-D 8 maintains an exceptionally high correlation (>0.90) with the full 20-item scale (González et al., 2017), ensuring that findings are comparable with the broader body of genomic and epidemiological evidence.

This thesis adopted a cut-off score of 4 or higher on the CES-D 8 scale. Drawing on foundational HRS documentation (Steffick, 2000), this specific threshold has been validated to reflect the conventional 16-point score used in the original 20-item scale (Radloff, 1977). This equivalence ensures the measure captures clinically significant depressive symptoms with precision, aligning with recent authoritative ELSA pandemic research (Di Gessa & Zaninotto, 2023; Zaninotto et al., 2022).

Loneliness

Loneliness was assessed using a composite measure consisting of the 3-item Revised University of California (UCLA) Loneliness Scale and an additional direct item asking participants how often they feel lonely (Hughes et al., 2004). Each component was rated

on a 3-point Likert scale (1 = “hardly ever/never”; 2 = “some of the time”; 3 = “often”). The individual scores were summed to produce a total index ranging from 4 to 12, with higher scores indicating greater perceived loneliness and social-emotional distress.

Quality of Life (QoL)

Subjective Quality of Life was evaluated using the CASP-12 scale (Control, Autonomy, Self-realisation, and Pleasure), a 12-item self-completion questionnaire specifically developed to assess wellbeing in older populations (Wiggins et al., 2008). The scale measures three distinct domains: control and autonomy, pleasure, and self-realisation. Each item is scored on a 4-point scale; the items were summed to create an index (range: 12–48). In the coding used for this study, higher scores reflect poorer quality of life, allowing for a consistent directional interpretation of "psychological distress" across all mental health indicators.

Anxiety (GAD-7)

Anxiety was assessed via the 7-item Generalized Anxiety Disorder scale (GAD-7) (Spitzer et al., 2006). This instrument evaluates the frequency of various anxiety symptoms (e.g., feeling nervous, anxious, or on edge) over the preceding two weeks. Each item is scored on a 4-point scale (0 = “Not at all” to 3 = “Nearly every day”), with total scores ranging from 0 to 21. Higher values indicate a greater burden of anxiety symptoms. It is important to note that the GAD-7 was introduced specifically during the ELSA COVID-19 Substudy waves; consequently, it is utilized in this thesis to capture acute psychological responses to the pandemic, acknowledging the absence of pre-pandemic baseline data for this specific instrument.

The strategic selection of outcome variables across this thesis is dictated by the specific analytical objectives and data constraints of each research stage. In Chapter 2, which establishes a ten-year longitudinal trajectory, the analysis is strictly concentrated on Memory and Depressive Symptoms (CES-D 8). These two indicators were selected as the primary representatives of cognitive and mental health due to their high instrumental fidelity and uninterrupted availability from Wave 4 through Wave 9. Prioritizing these gold-standard measures ensures maximum statistical power and representative population trends, avoiding the significant sample attrition and selection bias associated with auxiliary psychosocial modules (e.g., loneliness and quality of life), which are primarily collected via self-completion questionnaires and often suffer from higher rates of non-random missingness.

In contrast, Chapters 3 and 4 shift to a detailed investigation of gene-environment interactions following specific exogenous shocks. To capture the full spectrum of cognitive and mental health, this framework is expanded to include a wider range of indicators including Executive Function, Loneliness, Quality of Life, and Anxiety (GAD-7). While some of these measures have restricted availability—such as Executive Function (Waves 1–5) and Anxiety (introduced only during the COVID-19 pandemic)—their inclusion in these specific chapters provides the necessary phenotypic breadth to test the generalizability of G×E effects across different domains of wellbeing. These outcome measures allow the thesis to evaluate whether genetic susceptibility manifests across a broader health continuum, which a singular clinical focus on cognitive and mental health might overlook.

Furthermore, as the inclusion criteria and analytical subsamples vary depending on the research question of each chapter (e.g., the longitudinal sample in Chapter 2 versus the

pandemic-specific sample in Chapter 4), the specific descriptive statistics, distributions, and sample sizes for these variables are provided within the methodology sections of each respective empirical chapter.

1.6.4 Genetic Biomarkers: Construction of PGSs and PCs in ELSA

In the ELSA study, genetic data has been provided for detailed analyses of a wide range of age-related traits and outcomes in association with genetic factors in a large and phenotypically well-characterised sample of older individuals. To ensure methodological consistency across international ageing studies, the construction of polygenic scores (PGSs) in ELSA adheres to the standardized predictive modeling protocols established by the Health and Retirement Study (HRS, 2024). Specifically, this study utilizes the pre-calculated scores provided by the ELSA genomic team, which were generated using an additive model based on external GWAS meta-analysis weights to facilitate cross-cohort harmonisation (Ajnakina & Steptoe, 2019; Health and Retirement Study, 2024).

1) Genotyping and Administration Procedures

The genome-wide genotyping was performed at University College London (UCL) Genomics in 2013-2014. Blood samples were taken from 7,597 ELSA participants of European ancestry who met eligibility criteria (e.g., absence of clotting disorders or anticoagulant medication). Genotyping was conducted using the Illumina HumanOmni2.5 BeadChips (HumanOmni2.5-4v1 and HumanOmni2.5-8v1.3), measuring approximately 2.5 million markers. Initial cross-batch correlation for allele frequencies exceeded 99%. After excluding ethnic outliers and duplicates, the initial GWAS-ready dataset comprised 7,412 participants and 2,230,767 SNPs.

2) GWAS Quality Control (QC) Protocols

Prior to the construction of polygenic scores, rigorous QC was performed using PLINK 1.9 at both the individual and SNP levels. The genomic data underwent a multi-stage quality control (QC) protocol at both the individual and variant levels to ensure the integrity of the genetic biomarkers. At the individual level, samples were excluded if there was a discrepancy between biologically inferred genetic sex and recorded phenotypic sex. To address potential bias from population structure and relatedness, cryptic kinship was assessed using Identity-by-Descent (IBD) estimates; consequently, one individual from each pair with an IBD value exceeding 0.2—a threshold representing the midpoint between second- and third-degree relatives—was removed at random. Furthermore, participants with extreme heterozygosity scores, defined as deviating more than three standard deviations from the sample mean, were excluded to mitigate risks of sample contamination or inbreeding artifacts.

Parallel to individual-level filtering, stringent single-nucleotide polymorphism (SNP) quality control was implemented. Variants were excluded if they demonstrated a call rate below 98%, a minor allele frequency (MAF) of less than 0.01, or a significant deviation from Hardy-Weinberg Equilibrium ($p < 10^{-4}$). For the specific purpose of population structure analysis, Linkage Disequilibrium (LD) pruning was conducted using a sliding window approach (10 SNPs, $r^2 > 0.10$). This process intentionally excluded high-LD genomic regions, including the Lactase gene (LCT), the Human Leukocyte Antigen (HLA) complex, and inversion regions on 8p23.1 and 17q21.31, to prevent localized genetic clusters from distorting global ancestry estimates. Following these rigorous filtration steps, the final analytical dataset comprised 7,183 participants and 1,374,524 high-quality, directly genotyped SNPs.

3) Principal Component Analysis (PCA) for Population Structure

To account for population stratification—systematic ancestry differences that can lead to spurious associations—Principal Component Analysis (PCA) was employed. 10 ancestry principal components (PCs) were generated. These PCs act as continuous variables representing an individual's genetic ancestry. The construction process involves two critical technical steps:

A. Linkage Disequilibrium (LD) Pruning

In the human genome, many genetic variants (SNPs) are not inherited independently; instead, they are often inherited together in "blocks." This phenomenon is known as Linkage Disequilibrium (LD). If a large number of highly correlated SNPs from a single genomic block are included in a PCA, they will disproportionately dominate the analysis, masking the true global ancestry of the individual. To prevent this bias, ELSA performed LD-pruning, a process that filters the genome to retain only "unlinked" (independent) SNPs. This was achieved using a sliding window approach with a threshold of $r^2 < 0.1$, effectively ensuring that no two SNPs in the final subset are highly correlated. Additionally, we excluded specific "high-LD" regions (e.g., the HLA complex on chromosome 6 and the LCT gene), which are known for their unusually long and complex correlation patterns that can distort ancestry estimates.

B. Mathematical Construction of PCs

Following LD-pruning, the PCs were generated via eigen-decomposition of the genotype covariance matrix (C):

$$C v_k = \lambda_k v_k$$

Where: C represents the covariance of genotypes across the independent SNP subset. λ_k is the eigenvalue representing the amount of genetic variance captured by the k^{th} component. v_k is the eigenvector used to calculate the individual's score (PC_{ik}). The resulting top 10 PCs capture the most significant axes of genetic variation within the sample. By including these 10 PCs as covariates in all regression models, this study effectively "filters out" the noise of ancestral background, ensuring that the interaction between the PGS and environmental factors reflects a robust phenotypic association rather than a confounding effect of population structure.

4) Polygenic Score (PGS) Construction and Distribution

A Polygenic Score (PGS) provides a single-value estimate of an individual's genetic propensity for a phenotype, constructed as a weighted sum of risk-associated alleles, where different components contribute differently to the final score based on their importance. The mathematical formula is defined as:

$$PGS_i = \sum_{j=1}^J W_j G_{ij}$$

Where:

G_{ij} (Individual Genotype): Represents the number of risk-increasing alleles (0, 1, or 2) an individual possesses at a specific genetic location (SNP_j).

W_j (Effect Weight): Represents the "strength" of that specific variant's association with the trait. These weights are derived from independent, large-scale Genome-Wide Association Studies (GWAS) involving hundreds of thousands of individuals. By using external weights, the genetic risk estimated for the ELSA sample is unbiased and biologically grounded in broader population findings.

A critical decision in PGS construction is the "p-value threshold"—essentially deciding how many genetic variants to include. While some researchers only include "significant" variants (e.g., $P < 5 \times 10^{-8}$), the ELSA team follows the HRS recommendation to use a threshold of $P = 1$. This means we include all available genetic variants regardless of their individual statistical significance. The rationale is that even variants with very weak individual signals collectively capture a substantial portion of the "missing heritability" of complex traits. Including the full spectrum of genomic variation maximizes the predictive power of the score and ensures maximum transparency, as it does not rely on arbitrary cut-off points.

Because a PGS is the aggregate of over a million independent genetic variants, its distribution is governed by the Central Limit Theorem. This theorem states that the sum of many independent random variables tends toward a Normal Distribution, regardless of the distribution of the individual variables. Within the ELSA sample, the calculated PGSs exhibit a robust normal distribution, which satisfies the normality requirements for the linear regression and longitudinal models used in the empirical chapters. To ensure comparability across different phenotypes (e.g., comparing the effect of genetic risk for depression vs. cognition), all PGS variables were Z-score standardized (Mean = 0, Standard Deviation = 1).

To investigate the "G" component of the G×E interactions, five specific PGSs were extracted from the ELSA genetic database: General Cognition, Depressive Symptoms (DS), Major Depressive Disorder (MDD), Anxiety Disorders (AD), and Loneliness. These scores were selected based on their high overlap with the phenotypic measures described above. The use of multiple PGSs for mental health (e.g., both DS and MDD) allows this thesis to test the robustness of genetic associations across different discovery

GWAS frameworks, thereby enhancing the reliability of the findings. Reflecting the variations in analytical subsamples across different research questions, the specific descriptive statistics, distributions, and standardized properties of these PGSs are presented in detail within the methodology section of each respective empirical chapter.

Chapter 2

Trajectories of Cognition and Mental Health in Older Age

2.1 Introduction

Ageing is a complex, multidimensional process, accompanied by marked changes in both cognitive function and mental health. Among older adults, memory decline and depressive symptoms are two of the most prevalent and consequential outcomes (Harada et al., 2013; Cole & Dendukuri, 2003). Cognitive decline not only reduces the capacity for independent living but also serves as an early indicator of neurodegenerative disorders such as Alzheimer's disease and other dementias (Salthouse, 2010; Petersen, 2004). Likewise, depression is a leading mental health challenge in late life, with profound consequences for quality of life, morbidity, and mortality (Beekman et al., 1999; Blazer, 2003). Together, these two domains constitute critical dimensions of healthy ageing and public health.

Cognitive ageing is characterized by considerable heterogeneity. While some individuals maintain stable performance into late life, others experience gradual or accelerated decline (Zaninotto et al., 2018; Glymour et al., 2012). Influences on these trajectories span genetic, biological, social, and environmental domains. For example, sex differences are well documented: women tend to perform better on episodic memory tasks, whereas men often outperform in visuospatial domains (De Frias et al., 2006; Pauls et al., 2013). Genetic risks are also salient, with the APOE $\epsilon 4$ allele consistently associated with steeper cognitive decline and elevated risk of dementia (Hong et al., 2015; Krell-Roesch et al., 2015). At the same time, environmental and lifestyle factors act as powerful modifiers. Higher education and cognitively stimulating activities are

protective, consistent with the concept of cognitive reserve (Stern, 2012), while chronic conditions (e.g., hypertension, diabetes, cardiovascular disease) and unhealthy behaviors (e.g., smoking, inactivity) accelerate decline (Livingston et al., 2017; Yaffe et al., 2014). As Dixon and Lachman (2019) emphasize, cognitive ageing is best understood as a multidomain, interactive process where protective and risk factors accumulate and interact across the life course.

Mental health, as defined by the World Health Organization (2022a, 2022b, 2022c), is both intrinsically valuable and foundational to wellbeing, autonomy, and social participation. In older adults, depressive symptoms are particularly consequential, associated with functional impairment, multimorbidity, and increased health care utilization (Ownby et al., 2006; Santabárbara et al., 2019). Mental health trajectories, like cognitive trajectories, are heterogeneous, ranging from persistent or chronic depressive symptoms to transient or improving patterns (Steffens et al., 2000). Social and economic inequalities, gender, multimorbidity, and poor self-rated health are consistently implicated in shaping these outcomes (Hyde et al., 2008; Kuehner, 2017; Mendes de Leon et al., 2001; Zaninotto et al., 2016).

Traditional analyses often rely on population averages, which obscure meaningful variation in ageing processes. Longitudinal studies and group-based trajectory modeling (Nagin & Odgers, 2010) have revealed distinct subgroups with divergent developmental paths, underscoring the complexity of ageing. Despite this, few studies have jointly examined trajectories of memory and depressive symptoms in nationally representative cohorts, leaving gaps for identifying clusters of risk and resilience, as well as for understanding the cumulative influence of life-course inequalities.

Although memory and depressive symptoms are modelled as parallel outcomes in this chapter, they are not conceptually independent. A growing body of evidence points to a bidirectional relationship between cognitive decline and depression in later life: depressive symptoms are both a risk factor for and a consequence of cognitive deterioration, with shared neurobiological substrates including dysregulation of the hypothalamic–pituitary–adrenal (HPA) axis and hippocampal atrophy under chronic stress exposure (Dotson et al., 2010; Pariante & Lightman, 2008). From a life-course perspective, these two dimensions function as interlinked markers of cumulative disadvantage, sharing common long-term upstream determinants—including early-life socioeconomic disadvantage and educational attainment—while markedly diverging in their sensitivity to more proximate, time-varying exposures. Specifically, as demonstrated in the empirical sections of this chapter, cognitive trajectories appear to reflect the accumulated cognitive reserve shaped by early-life human capital investments (Stern, 2012)—reserves that confer a degree of insulation against short-term health shocks among those on higher-functioning trajectories. Depressive symptoms, by contrast, exhibit greater volatility and acute sensitivity to late-life situational disruptions: across all five trajectory groups, deteriorations in self-rated health showed a graded and consistently significant contemporaneous association with elevated symptom burden, with the magnitude of this effect amplified among those already carrying higher levels of distress. The analytical separation of memory and depressive symptoms is therefore a methodological choice to preserve interpretive clarity, not an assumption of biographical independence. The convergences and divergences in their predictors—documented throughout this chapter—are themselves substantively informative, revealing the domain-specific timescales and pathways through which social conditions get under the skin to shape the heterogeneous landscape of late-life health.

Therefore, this chapter addresses these gaps by examining trajectories of memory and depressive symptoms among older adults in England, using data from the English Longitudinal Study of Ageing (ELSA) (Stephoe et al., 2013). Leveraging repeated assessments across multiple domains, rich socio-demographic and health measures, and long follow-up, the analysis identifies patterns of change in cognition and mental health and their key determinants, thereby providing a comprehensive overview of these two critical late-life outcomes. This descriptive foundation sets the stage for the subsequent chapters, which examine each outcome in greater depth by focusing on the influence of specific life events and the interplay between genetic risks and modifiable environmental factors. Therefore, this chapter intentionally focuses on sociodemographic and time-varying social and health determinants without including genetic predictors. Estimating social-environmental associations in the full longitudinal sample maximises statistical power and yields robust benchmark estimates of the socioeconomic and health correlates of memory and depressive symptom trajectories. Genetic analyses require additional sample selection and controls (for example, exclusion criteria and principal components to address population stratification); presenting them separately in Chapters 3 and 4 therefore ensures methodological transparency and direct comparability with the descriptive benchmarks reported here.

2.2 Methods

Study Design and Participants

Data were drawn from the English Longitudinal Study of Ageing (ELSA), a nationally representative longitudinal panel study of individuals aged 50 and over living in private households in England, which began in 2002 (Banks et al., 2021). Data are collected biennially through face-to-face interviews and self-completion questionnaires. The first wave of ELSA commenced in 2002 with an initial individual response rate of 67%. To maintain representativeness of the population aged 50 and over, the ELSA sample has been periodically refreshed—specifically at Waves 3, 4, 6, 7, and 9. To date, over 18,000 respondents have taken part in the biennial interviews.

The longitudinal trajectories in this chapter were estimated using six biennial waves of data from Wave 4 (2008/09) to Wave 9 (2018/19). To ensure a consistent ten-year follow-up period, the analytical sample was restricted to respondents who participated in both the Wave 4 and Wave 9 interviews. Refreshment samples introduced after Wave 4 were excluded to maintain a consistent ten-year longitudinal window. This decision also ensures that the analysis focuses strictly on the cohort aged 50 and older at baseline, avoiding the age-related heterogeneity that younger refreshment samples would have introduced. Of the 10,749 eligible participants at Wave 4, 8,557 (79.6%) completed the Wave 9 interview, representing a cumulative attrition rate of 20.4%—a figure within the expected range for a decade-long study of an ageing population (Steptoe et al., 2013). Within this final sample, 5,844 (68.3%) were present across all six waves.

While this selection criterion ensures longitudinal stability, it may introduce survivor bias or a 'healthy volunteer' effect, as individuals with poorer health are more likely to drop out or pass away before the final wave. To mitigate these concerns, official ELSA

longitudinal weights were applied to adjust for non-random attrition and differential non-response. Additionally, the Group-Based Trajectory Modelling (GBTM) employed a Full Information Maximum Likelihood (FIML) estimator. This approach leverages all available observations from individuals who missed intermediate waves (Waves 5–8) but met the endpoint criteria, thereby maximising statistical power and reducing estimation bias.

Measures

Outcomes

In this chapter, which establishes a ten-year longitudinal trajectory, the analysis is strictly concentrated on Memory and Depressive Symptoms (CES-D 8). These two indicators were selected as the primary representatives of cognitive and mental health due to their high instrumental fidelity and uninterrupted availability from Wave 4 through Wave 9. Prioritizing these two gold-standard measures ensures maximum statistical power and representative population trends, avoiding the significant sample attrition and selection bias associated with auxiliary psychosocial modules (e.g., loneliness and quality of life), which are primarily collected via self-completion questionnaires and often suffer from higher rates of non-random missingness.

Specifically, the cognitive function module in ELSA contains assessments of memory, executive function, processing speed, and numerical ability. Memory was consistently measured from Wave 1 to Wave 9, and executive function was available only from Wave 1 to Wave 5. To ensure comparability across the full study period and to reflect the most updated sample, this chapter uses memory—rather than executive function—as the primary cognitive outcome. In addition, memory plays a central role in age-related cognitive decline and is a key early indicator of neurodegenerative processes, making it

particularly relevant for research on ageing. Memory was assessed using a word-list learning test, in which participants were read a list of 10 words and asked to recall as many as possible both immediately and after a delay of approximately five minutes, during which they completed other survey questions. An overall memory score (ranging from 0 to 20) was calculated by summing the scores from the immediate and delayed recall tasks, with a maximum of 10 points for each.

Depressive symptoms (DS) were assessed using an abbreviated eight-item version of the validated Center for Epidemiologic Studies Depression (CES-D) Scale (Beekman et al., 1997; Radloff, 1977). Although not a diagnostic tool for clinical depression, the CES-D is widely used in population-based research to identify individuals at risk of depression. The short form includes eight binary (yes/no) items asking whether respondents experienced specific symptoms—such as feeling sad or having restless sleep—in the week prior to the interview. This version demonstrated comparable psychometric properties to the original 20-item scale (Karim et al., 2015). Respondents reporting four or more depressive symptoms were classified as having elevated depressive symptoms (Zaninotto, Iob, Demakakos, & Steptoe, 2022).

Time-invariant covariates

This chapter included the following time-invariant covariates: age, gender, education and father's occupation at age 14. Although age, gender and education were measured at each wave, they were treated as time-constant covariates due to minimal or systematic variability over time. Age at baseline was categorised into five-year bands (50–54, 55–59, ..., 95–99) to capture non-linear age effects and the accelerated health changes among the oldest-old. This high-resolution grouping improves functional flexibility to model non-proportional effects while retaining adequate sample sizes within strata for

statistical power. Educational attainment was re-coded into three categories: low (below secondary level), middle, and high (university level or above).

Early-life socioeconomic status was measured by the father's occupation when the respondent was 14 years old. While this indicator relies on retrospective self-reports and may be subject to some degree of recall bias, it remains one of the most robust and widely validated proxies for childhood SEP in longitudinal studies of ageing (Galobardes et al., 2006; Marmot et al., 2001). The primary advantage of using father's occupation lies in its ability to reflect the long-term material resources and social prestige of the household during a critical developmental period, which has been shown to set the foundation for late-life health trajectories (Kuehner, 2017; Zaninotto et al., 2016). Furthermore, compared to other childhood measures in ELSA, father's occupation offers a more complete data profile with fewer missing values, allowing for a more representative analysis of early-life influences. In this chapter, these occupations were categorized into three skill-level classes based on the International Standard Classification of Occupations (ISCO), providing a standardized framework to capture early-life socioeconomic gradients.

Time-varying covariates

Time-varying socioeconomic factors included **wealth**, which was operationalized as total net non-pension household wealth, aggregated at the household level. This measure encompasses financial assets, physical wealth (e.g., business assets, land, and valuables), and housing wealth (including both primary and secondary residences), net of any outstanding debts. Following established protocols in research using the English Longitudinal Study of Ageing (ELSA) (Frank et al., 2019; Steptoe et al., 2025), wealth was categorised into terciles for the analysis to capture relative socio-economic

standing. **Employment** status was operationalised using a tripartite classification: employed (including those self-employed or in paid work), unemployed, and retired. The primary health-related factor, **self-rated health (SRH)**, was measured using a five-point Likert scale. Respondents were asked to rate their general health on an ordinal scale with the following categories: excellent, very good, good, fair, and poor. In the analytical models, SRH was treated as a five-level categorical variable. This approach maintains the full granularity of the original measure, allowing the analysis to account for the complete health gradient and potential non-linear associations between subjective health perceptions and the identified trajectories of cognitive/depressive outcomes.

Physical disability was assessed based on limitations in activities of daily living (ADLs)—such as getting out of bed or walking across a room—and instrumental ADLs, including tasks like grocery shopping or preparing a hot meal. Respondents reporting difficulty with one or more of these activities were classified as physically disabled.

Multimorbidity was assessed based on six chronic conditions commonly reported in the ELSA study: arthritis, cancer, chronic lung disease, diabetes, hypertension, and stroke. For the analysis, this was operationalised as a binary indicator, where participants were categorised as having multimorbidity if they reported two or more of these conditions (coded as 1), while those with fewer than two conditions were coded as 0.

2.3 Statistical Analysis

The analysis was conducted in two main steps. First, group-based trajectory modelling (GBTM) was applied to identify distinct longitudinal patterns of memory and depressive symptoms (Nagin & Odgers, 2010). This method accounts for the dependency of repeated measures and assumes that the population consists of unobserved subgroups, each following a distinct developmental trajectory (Herle et al., 2020; Nguena Nguetack et al., 2020). Both linear and non-linear trajectories can be modelled by including quadratic or cubic growth terms, as needed. Missing data were addressed using full information maximum likelihood estimation.

For each participant, the model estimated the probability of belonging to each trajectory group and assigned individuals to the group with the highest posterior probability. To determine the optimal number of trajectory groups, unconditional models with two to seven groups were tested using data from six waves. Model selection was based on a combination of criteria, including the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample size-adjusted BIC (c-BIC), with lower values indicating better model fit. Additional indicators included the average posterior probability of group membership (values >0.7), the odds of correct classification (>5.0), minimum group size (avoiding groups smaller than 5% of the sample), the distinctiveness of trajectory shapes, and the interpretability of the resulting patterns (Nagin & Odgers, 2010; Nguena Nguetack et al., 2020). Once the optimal number of groups was determined, the most appropriate functional form for each trajectory—linear, quadratic, or cubic—was identified, and non-significant higher-order terms ($p \geq 0.05$) were excluded.

The choice of survey wave as the time metric, rather than chronological age, was deliberate and is consistent with established practice in GBTM studies using panel data with structured follow-up intervals (Nagin & Odgers, 2010; Jones & Nagin, 2007). Because ELSA collects data at biennial intervals with cohort-refreshment at Waves 3, 4, 6, 7, and 9, using age as the ranking variable would require aligning observations across a continuous, unequally spaced age range spanning approximately 50 to 100 years, substantially increasing model complexity and reducing the precision of trajectory estimation within the observed follow-up window (Waves 4–9, covering approximately 10 years). The wave-based specification captures the within-individual rate of change over the observed study period in a parsimonious manner. Age is nonetheless accounted for as a time-invariant covariate in the conditional models, ensuring that age-related differences in trajectory membership are identified.

In the second stage of the analysis, two approaches were used to examine factors associated with the identified trajectory groups. First, a multinomial logistic regression model was estimated using baseline covariates (Wave 4) to identify predictors of group membership. To enhance interpretability, average marginal effects (AMEs) were reported. These represent the average change in the predicted probability of each outcome category associated with changes in covariates, expressed in percentage points.

Second, the GBTM framework was extended by incorporating time-constant and time-varying covariates, allowing for the examination of predictors of memory and depressive symptom scores within each trajectory group over time. Including time-varying variables enabled trajectories to reflect additional influences beyond time itself. These covariate effects were estimated separately for each trajectory group, capturing group-specific associations. Estimates for time-varying variables represent the average

change in memory and depressive symptom scores associated with a one-unit change in the respective covariate. All data management and statistical analyses were conducted using Stata/SE v18.0 with the traj plugin. Due to participant confidentiality and data protection regulations, the raw data cannot be shared directly by the author. Researchers interested in accessing the data must apply through the UKDS following the standard procedures. The analysis code necessary to reproduce the results are available from the author upon reasonable request.

2.4 Results

Memory scores showed a small downward trend across the six waves: mean = 11.2 (SD = 3.30) in Wave 4, dipping to about 11.0 (SD = 3.72) by Wave 7 and further to ~10.6–10.8 (SD ≈ 3.73) in later waves, indicating a modest overall decline. Depressive symptoms also declined slightly over time, from mean = 1.37 (SD = 1.82) in Wave 4 to 1.21 (SD = 1.77) in Wave 9.

Table 2.1 Summary of longitudinal data on the scores of memory and depressive symptoms.

ELSA wave	Memory		Depressive symptoms	
	N	Mean (SD)	N	Mean (SD)
4	8367	11.2 (3.30)	8305	1.37 (1.82)
5	8041	11.1 (3.32)	7987	1.31 (1.82)
6	7884	11.2 (3.36)	7809	1.28 (1.81)
7	7579	11.0 (3.57)	7514	1.24 (1.78)
8	6809	10.6 (3.73)	6785	1.27 (1.78)
9	6784	10.8 (3.73)	6690	1.21 (1.77)

Notes: Data from the English Longitudinal Study of Ageing (ELSA), Waves 4–9 (2008–2018). Memory scores are based on a word-list recall task (range 0–20). Depressive symptoms are measured using the CES-D 8 scale (range 0–8). Values are sample means (SD). Higher memory scores indicate better cognitive performance; higher CES-D 8 scores indicate greater depressive symptom burden.

To summarise the development of memory and depressive symptoms over time and determine the optimal number of trajectory groups, a series of group-based trajectory models were fitted (with a specification of from two to seven trajectory groups). Based on the model fit statistics, (see Supplementary Table S2.1-4), the 3-group solution is selected for memory, as it offers a strong balance between model fit (low BIC/AIC) and

classification quality (entropy = 0.688), with clearly interpretable group sizes. For depressive symptoms, the 5-group model is chosen. It demonstrates improved fit over simpler models while maintaining meaningful group distinctions and avoiding redundancy, making it the most informative and parsimonious solution. The identified trajectory groups as a function of time (measured by study wave) are presented in Figure 2.1 and Figure 2.2.

Figure 2.1 illustrates three distinct groups representing respondents with stable levels of memory: 17.6% exhibited consistently low memory performance (“low and decreasing”), 51.4% maintained medium levels (“stable medium”), and 31.0% retained high levels (“stable high”) throughout the observation period. As detailed in Supplementary Table S2.5, the low and decreasing group started at 5.93 in Wave 4 and declined steadily to 4.20 by Wave 9. The stable medium group maintained scores around 10, with only a modest decline from 10.22 in Wave 4 to 9.47 in Wave 9. The stable high group consistently performed above 13, showing minimal reduction from 13.64 in Wave 4 to 13.53 in Wave 9. The low and decreasing group is characterised by a marked non-linear decline across waves — from a mean of 5.93 at Wave 4 to 4.20 at Wave 9 (a reduction of approximately 0.5 SD) — in contrast to the substantially smaller changes observed in the stable medium and stable high groups (−0.75 and −0.11 points respectively). These trajectories highlight the heterogeneity in memory performance, with the more pronounced decline observed among individuals in the lower memory group.

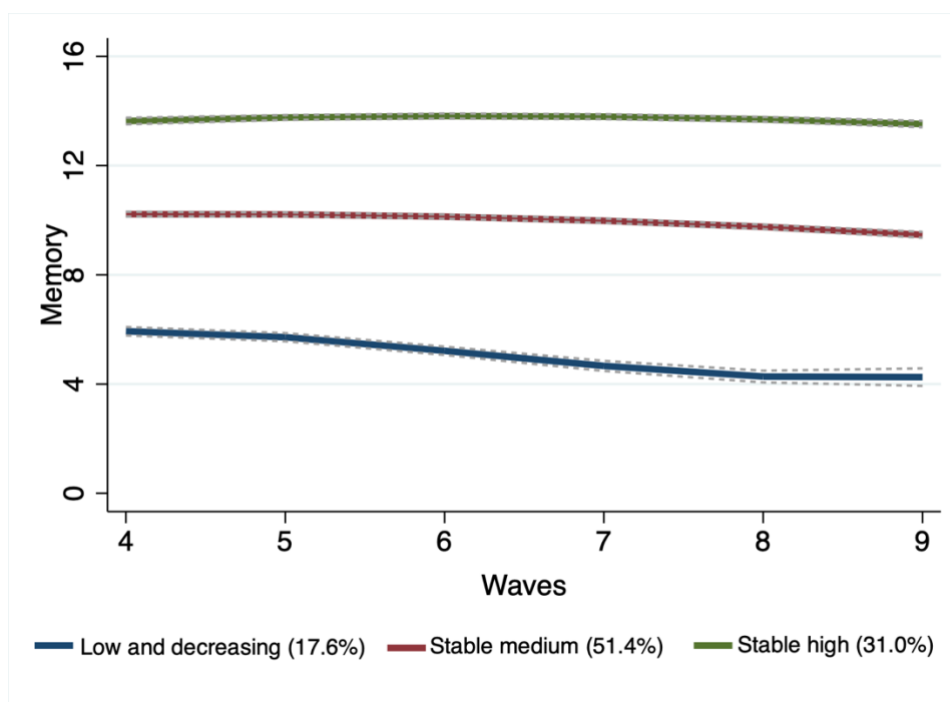


Figure 2.1 Memory trajectories

Trajectories obtained with an unconditional model that accounts only for time (expressed as the wave of the study).

Source: English Longitudinal Study of Ageing (ELSA) waves 4 (2008/09) – 9 (2018/19).

Figure 2.2 displays five trajectory groups based on depressive symptom levels. Three groups exhibited relatively stable patterns: 27.5% reported persistently low levels of depressive symptoms (“stable low”), 34.6% medium (“stable medium”), and 11.8% high (“stable high”). The remaining two groups demonstrated time-varying trajectories: 13.9% of respondents showed a steady increase in depressive symptoms (“increasing”), while 12.2% experienced a progressive decline (“decreasing”). As detailed in Supplementary Table S2.6, the “increasing” group began with an average score of 1.19 and rose to 3.20 by the final wave. In contrast, the “decreasing” group showed a reduction in depressive symptoms from an average of 3.63 to 1.06 over a ten-year period, representing a decline of approximately 2.6 points. Together, these patterns illustrate both stability and dynamic change in depressive symptom trajectories among older adults.

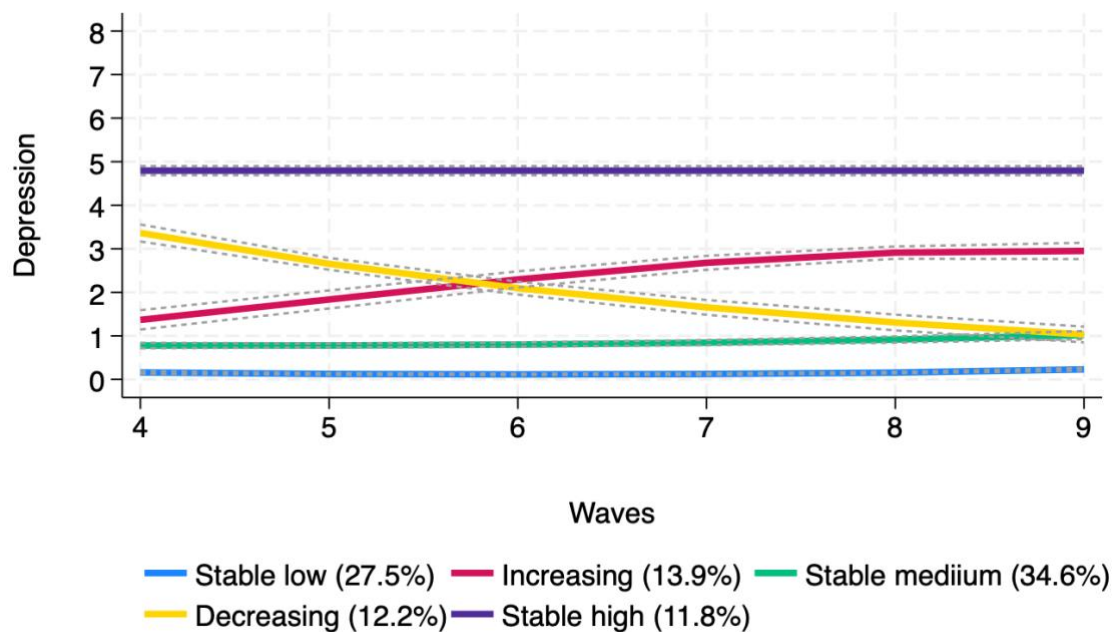


Figure 2.2 Depressive symptoms trajectories

Trajectories obtained with an unconditional model that accounts only for time (expressed as the wave of the study).

Source: English Longitudinal Study of Ageing (ELSA) waves 4 (2008/09) – 9 (2018/19).

Table 2.2 and 2.3 present the baseline characteristics of ELSA respondents, showing the distribution of key demographic and socioeconomic factors across the three memory trajectory groups and five depressive symptom trajectory groups, respectively. Overall, the percentage of women in the “stable high” and “stable medium” memory groups is higher than in the overall sample distribution. Also, among those in the “low and decreasing” memory group, there is a higher percentage of groups at aged 70+ (notably those in the 70–89 bands). Those in the “stable high” memory group tend to have a higher percentage of respondents in this group having medium/high levels of education. Additionally, a greater proportion of this group reported having a father in a high-skilled occupation at age 14, indicating a potential link between early-life socioeconomic status and later-life cognitive function.

Table 2.2 Baseline sample characteristics (%), by memory trajectories and overall.

	Low and decreasing	Stable medium	Stable high	Overall	P value
Age 50-54	1.5	3.1	5.4	3.4	<0.001
Age 55-59	5.2	11.9	19.8	13	
Age 60-64	9.8	21.7	31.2	22.3	
Age 65-69	14	22.5	23.9	21.5	
Age 70-74	18.9	18.8	12.7	17.2	
Age 75-79	20.8	12.5	4.8	11.8	
Age 80-84	17.4	6.7	1.7	7.1	
Age 85-89	8.9	2.2	0.5	2.8	
Age 90-94	1	0.2	0	0.3	
Age 95-99	2.7	0.4	0	0.7	
Female	51.7	53.1	54.5	53	<0.001
Low education	40.1	35.9	29.8	35	<0.001
Medium education	35.2	36.5	37.4	36	
High education	24.7	27.6	32.8	29	
Father's job: low	47.1	41	33	39.1	<0.001
Father's job: medium	31.4	33.1	30.3	32	
Father's job: high	21.5	26	36.8	28.9	

Notes: Baseline characteristics measured at Wave 4 (2008). Percentages within each trajectory group are shown. Education categories: low = no qualifications or below O-level; middle = O-level or equivalent; high = A-level or above. Father's occupation at age 14 classified using the National Statistics Socio-economic Classification. P-values from chi-squared tests for categorical variables and ANOVA for continuous variables. Models use ELSA longitudinal survey weights based on Wave 4.

Table 2.3 summarises the baseline characteristics of respondents by depressive symptom trajectory group. A notable pattern emerges regarding age distribution that younger participants at baseline are disproportionately represented in the 'stable high' trajectory. For instance, the youngest age band (50–54) accounts for 5.18% of the 'stable high' group, compared to 3.43% of the overall sample. Similarly, the 55–59 age group makes up 16.35% of the 'stable high' profile, surpassing their representation in any other trajectory. Women are overrepresented in all groups with elevated depressive symptoms, especially in the “stable high” categories. Clear socioeconomic gradients are also evident. Respondents in the “stable high” group are the least likely to have low

education (28.5%) and most likely to have higher educational attainment groups, including medium (38.0%) and high education (33.5%). A greater share of participants with low father's occupations are found in the “decreasing” and “stable high” groups (44.0% and 47.8%, respectively), again highlighting the potential long-term influence of early-life socioeconomic conditions.

Table 2.3 Baseline sample characteristics (%), by depressive symptoms trajectories and overall.

	Stable low	Increasing	Stable medium	Decreasing	Stable high	Overall	P value
Age 50-54	3.26	2.73	3.79	3	5.18	3.43	
Age 55-59	13.47	9.32	13.95	11.36	16.35	12.97	
Age 60-64	22.98	18.67	24.21	20.02	22.42	22.3	
Age 65-69	22.59	19.58	22.34	19.86	18.92	21.48	
Age 70-74	17.52	18.59	16.44	17.7	14.31	17.16	<0.001
Age 75-79	11.68	14.53	10.07	13.58	10.28	11.78	
Age 80-84	6.09	10.13	6.07	8.86	7.75	7.12	
Age 85-89	1.95	4.61	2.33	4.02	3.68	2.82	
Age 90-94	0.15	0.52	0.18	0.58	0.31	0.28	
Age 95-99	0.31	1.3	0.63	1.02	0.8	0.67	
Female	52.9	52.1	53.8	54.8	55.3	53.7	<0.001
Low education	42	38.3	33.6	30.5	28.5	35.2	<0.001
Medium education	34.5	35.6	36.7	37.5	38	36	
High education	23.5	26.1	29.7	32	33.5	28.8	
Father's job: low	36.1	41.2	37.9	44	47.8	39.1	<0.001
Father's job: medium	31.2	30.88	32.79	33.18	31.6	32	
Father's job: high	32.8	28	29.3	22.8	20.6	28.9	

Notes: Baseline characteristics measured at Wave 4 (2008). Percentages within each trajectory group are shown. Education categories: low = no qualifications or below O-level; middle = O-level or equivalent; high = A-level or above. Father's occupation at age 14 classified using the National Statistics Socio-economic Classification. P-values from chi-squared tests for categorical variables and ANOVA for continuous variables. Models use ELSA longitudinal survey weights based on Wave 4.

Following these descriptive tables, I present two complementary sets of regression-based results. Tables 2.4–2.5 report baseline determinants of trajectory membership using multinomial logistic regression; estimates are presented as average marginal

effects (AMEs) and therefore express changes in the probability (percentage points) of being assigned to each trajectory group relative to the reference category of determinants. Tables 2. 6–2.7 report trajectory-group-specific associations of time-varying variables with cognitive/depressive outcome scores: these coefficients are estimated conditional on trajectory group assignment and indicate the average within- - group change in the outcome associated with a one-unit change in the predictor. In other words, AMEs in Tables 2.4-2.5 address “who is more likely to belong to which trajectory group”, whereas group-specific coefficients in Table 2.6-2.7 address “among those assigned to a trajectory, how does a time-varying factor associate with the outcome over time”. All models use longitudinal survey weights based on ELSA Wave 4.

The AME results address baseline selection into trajectories: they show which baseline characteristics are associated with greater or lesser likelihood of following each longitudinal course. They do not estimate within-trajectory changes over time. Positive AMEs should be read as higher probability (percentage-point change) of membership; negative AMEs as reduced probability. Table 2.4 reports AMEs from the multinomial logit predicting memory-trajectory membership. Age effects are heterogeneous and non-monotonic at the oldest ages. Compared with the youngest cohort (aged 50–54), the probability of belonging to the 'stable medium' trajectory is significantly higher among several older age bands (70–74: AME = 0.174, SE = 0.074, $p < 0.01$; 75–79: AME = 0.145, SE = 0.076, $p < 0.05$; 85–89: AME = 0.207, SE = 0.118, $p < 0.05$; 90–94: AME = 0.502, SE = 0.069, $p < 0.001$; 95–99: AME = 0.305, SE = 0.159, $p < 0.05$). For instance, individuals aged 70–74 (AME = 0.174, $p < 0.01$) and 90–94 (AME = 0.502, $p < 0.001$) exhibit a 17.4 and 50.2 percentage point increase in the likelihood of 'stable medium' group membership, respectively. By contrast, Conversely, the probability of

membership in the 'stable high' trajectory—representing the most optimal memory profile—declines sharply with advancing age from the 70s onwards (70–74: AME = -0.213 , SE = 0.069 , $p < 0.001$; 75–79: AME = -0.281 , SE = 0.069 , $p < 0.001$; 80–84: AME = -0.350 , SE = 0.069 , $p < 0.001$; 85–89: AME = -0.389 , SE = 0.071 , $p < 0.001$; 90–94 and 95–99: AME = -0.417 , SE ≈ 0.066 , $p < 0.001$). These findings suggest a substantial 'shifting' effect: as age increases, the likelihood of maintaining the highest cognitive performance diminishes, with a corresponding migration into intermediate or declining profiles. The pronounced AME observed for the 90–94 age group in the 'stable medium' category (0.502) likely encapsulates complex selection effects and survivor bias, whereby the oldest-old survivors in the sample are more likely to exhibit moderate stability rather than the high-functioning profiles characteristic of their younger counterparts.

Gender significantly differentiates the probability of trajectory group membership. Compared with men, women exhibit a lower likelihood of belonging to the 'low and decreasing' memory trajectory, with an Average Marginal Effect (AME) of -0.068 (SE = 0.019 , $p < 0.001$). This indicates that, on average, being female is associated with a 6.8 percentage point decrease in the probability of following the most adverse memory trajectory. Conversely, women are substantially more likely to be classified in the 'stable high' trajectory (AME = 0.104 , SE = 0.021 , $p < 0.001$), representing a 10.4 percentage point increase in probability relative to men.

A significant educational gradient is evident across the identified memory trajectories. Relative to those with low educational attainment, individuals with middle education have a 7.9 percentage point lower probability of being assigned to the 'low and decreasing' trajectory (AME = -0.079 , SE = 0.023 , $p < 0.001$) and a 9.8 percentage

point higher probability of occupying the 'stable high' profile (AME = 0.098, SE = 0.026, $p < 0.001$). This gradient is even more pronounced for the highly educated, who show a 13.8 percentage point reduction in the probability of the most adverse decline (AME = -0.138, SE = 0.020, $p < 0.001$) and a 21.5 percentage point increase in the probability of maintaining high cognitive stability (AME = 0.215, SE = 0.027, $p < 0.001$). These AMEs underscore the role of education as a 'fundamental cause' of health inequalities; higher educational attainment substantially shifts the probability of assignment toward more favourable trajectories, likely by enhancing cognitive reserve and providing superior socio-economic buffers against age-related deterioration.

Early-life socioeconomic status (SES), proxied by father's occupation at age 14, emerges as a significant long-term predictor of late-life cognitive trajectories.

Individuals whose fathers held high-skilled occupations exhibit a substantial advantage in cognitive preservation; they are 10.1 percentage points more likely to be classified in the 'stable high' trajectory (AME = 0.101, $p < 0.001$) and significantly less likely to follow the 'low and decreasing' path (AME = -0.053, $p < 0.01$) compared to those from low-skilled backgrounds. Even middle-skilled father's occupations provide a protective buffer, increasing the probability of optimal memory stability by 5.2 percentage points (AME = 0.052, $p < 0.01$). These findings strongly support the cognitive reserve hypothesis, suggesting that advantaged early-life environments contribute to a more resilient neural or cognitive 'scaffolding' that persists into older age.

Table 2.4 Average marginal effects for a multinomial logistic regression model for
memory trajectories.

	Low and decreasing	Stable medium	Stable high
Age group (Ref: 50–54)			
55-59	--0.006 (0.014)	0.001 (0.078)	-0.002 (0.097)
60-64	-0.010 (0.041)	0.047 (0.074)	-0.037 (0.069)
65-69	0.019 (0.043)	0.086 (0.074)	-0.105 (0.069)
70-74	0.039 (0.043)	0.174** (0.074)	-0.213*** (0.069)
75-79	0.137*** (0.047)	0.145* (0.076)	-0.281*** (0.069)
80-84	0.256*** (0.055)	0.094 (0.081)	-0.350*** (0.069)
85-89	0.182* (0.100)	0.207* (0.118)	-0.389*** (0.071)
90-94	-0.085** (0.038)	0.502*** (0.069)	-0.417*** (0.066)
95-99	0.112 (0.148)	0.305* (0.159)	-0.417*** (0.066)
Gender (Ref: male)			
Female	-0.068*** (0.019)	-0.036 (0.026)	0.104*** (0.021)
Education (Ref: low education)			
Middle education	-0.079*** (0.023)	-0.019 (0.031)	0.098*** (0.026)
High education	-0.138*** (0.020)	-0.077** (0.031)	0.215*** (0.027)
Father's occupation at 14 (Ref: low)			
Middle	-0.022 -0.021	-0.029 -0.031	0.052** -0.026
High	-0.053** -0.021	-0.048 -0.031	0.101*** -0.027

Notes: AMEs report the average change in the probability (percentage points) of belonging to the indicated memory trajectory associated with the covariate (relative to the reference category). Robust standard errors in parentheses. Reference categories: Age 50–54; Gender = male; Education = low. Models use ELSA longitudinal weights based on Wave 4. Positive AME = higher probability of membership; negative AME = lower probability. Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 2.5 presents the AMEs from a multinomial logistic regression model predicting membership in the identified depressive symptom (DS) trajectories. Age shows non-monotonic and relatively nuanced associations across most bands. However, distinct patterns emerge at the oldest ages: individuals in the 85–89 band exhibit a 23.4 percentage point higher probability of belonging to the ‘increasing’ depressive trajectory (AME = 0.234, SE = 0.113, $p < 0.01$) compared to the youngest cohort (aged 50–54). In contrast, those in the 90–94 and 95–99 bands show significantly lower probabilities of membership in both the ‘increasing’ (AME = -0.119 , $p < 0.001$) and ‘stable high’ (AME ≈ -0.102 , $p < 0.01$) trajectories. Similar to the findings for cognitive memory, these seemingly paradoxical effects among the oldest-old likely reflect selective survival and cohort heterogeneity. Individuals with the most severe or increasing mental health burdens are less likely to survive into their 90s or remain in the study, resulting in a survivor sample characterized by more stable, albeit perhaps not optimal, mental health profiles.

Women are markedly less likely to follow a ‘stable low’ depressive trajectory (AME = -0.168 , SE = 0.025, $p < 0.001$), with a 16.8 percentage point lower probability compared to men. Instead, women exhibit a significantly higher likelihood of membership in the ‘stable medium’ (AME = 0.075, $p < 0.001$), ‘decreasing’ (AME = 0.087, $p < 0.001$), and ‘stable high’ (AME = 0.036, $p < 0.01$) trajectories.

Higher educational attainment is associated with substantial shifts toward more favourable mental health outcomes. Middle education (relative to low education) increases the probability of following a ‘stable low’ trajectory (AME = 0.088, SE = 0.030, $p < 0.001$) while decreasing the likelihood of falling into the ‘decreasing’ (AME = -0.056 , $p < 0.01$) or ‘stable high’ (AME = -0.033 , $p < 0.05$) symptom groups. High education demonstrates an even stronger protective effect, increasing the probability of

the most optimal ‘stable low’ profile by 16.1 percentage points (AME = 0.161, $p < 0.001$).

The influence of father’s occupation at 14 is more nuanced on depressive symptom trajectory. A high-skilled father’s background is associated with an 8.2 percentage point increase in the probability of belonging to the ‘stable medium’ trajectory (AME = 0.082, $p < 0.001$) and a corresponding 6.8 percentage point reduction in the likelihood of following a ‘decreasing’ symptom path (AME = -0.068, $p < 0.001$). Interestingly, early-life SES does not significantly differentiate the ‘stable low’ or ‘stable high’ distress profiles in the same way it stratifies memory.

Table 2.5 Average marginal effects for a multinomial logistic regression model for DS trajectories.

	Stable low	Increasing	Stable medium	Decreasing	Stable high
Age group (Ref: 50–54)					
55-59	-0.046 (0.037)	-0.026 (0.048)	0.049 (0.079)	0.024 (0.060)	-0.008 (0.041)
60-64	-0.073 (0.073)	-0.037 (0.043)	0.063 (0.062)	0.064 (0.049)	-0.018 (0.044)
65-69	-0.030 (0.073)	-0.028 (0.044)	0.054 (0.063)	0.035 (0.049)	-0.031 (0.045)
70-74	-0.033 (0.074)	-0.033 (0.044)	0.044 (0.063)	0.070 (0.050)	-0.048 (0.043)
75-79	-0.050 (0.076)	-0.019 (0.046)	0.066 (0.066)	0.076 (0.051)	-0.073* (0.043)
80-84	-0.053 (0.081)	0.060 (0.053)	-0.006 (0.068)	0.053 (0.054)	-0.055 (0.046)
85-89	-0.155 (0.115)	0.234** (0.113)	-0.091 (0.094)	0.090 (0.087)	-0.078 (0.048)
90-94	0.075 (0.258)	-0.119*** (0.041)	-0.222*** (0.059)	0.368 (0.253)	-0.102** (0.042)
95-99	0.267 (0.203)	-0.119*** (0.041)	-0.069 (0.172)	0.023 (0.126)	-0.102** (0.042)
Gender (Ref: male)					
Female	-0.168*** (0.025)	-0.029* (0.015)	0.075*** (0.023)	0.087*** (0.020)	0.036** (0.015)

	Stable low	Increasing	Stable medium	Decreasing	Stable high
Education (Ref: low education)					
Middle education	0.088***	0.013	-0.013	-0.056**	-0.033*
	-0.03	-0.019	-0.028	-0.023	-0.018
High education	0.161***	-0.004	-0.044	-0.058**	-0.055***
	-0.03	-0.018	-0.027	-0.024	-0.018
Father's occupation at 14 (Ref: low)					
Middle	-0.042	0.011	0.034	-0.010	0.007
	(0.029)	(0.018)	(0.027)	(0.024)	(0.018)
High	0.006	-0.004	0.082***	-0.068***	-0.015
	(0.031)	(0.018)	(0.029)	(0.023)	(0.017)

Notes: AMEs report the average change in the probability (percentage points) of belonging to the indicated depressive symptoms trajectory associated with the covariate (relative to the reference category). Robust standard errors in parentheses. Reference categories: Age 50–54; Gender = male; Education = low. Models use ELSA longitudinal weights based on Wave 4. Positive AME = higher probability of membership; negative AME = lower probability. Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 2.6 presents the group-specific coefficients for time-varying variables within the GBTM framework. Unlike the baseline multinomial models, these coefficients estimate the contemporaneous associations between longitudinal fluctuations in socioeconomic or health conditions and memory performance, conditional on an individual's assigned trajectory. This analysis evaluates how intra-individual changes in life circumstances relate to deviations from the average predicted path within each established trajectory.

Higher contemporaneous wealth is consistently and positively associated with memory scores across all trajectories. For individuals in the 'low and decreasing' memory trajectory group, the transition into the highest wealth quintile (relative to the lowest) is associated with a 0.519-point increase in memory score ($SE = 0.163$, $p < 0.001$). Similar positive associations are observed in the 'stable medium' (highest wealth: $coef = 0.459$, $p < 0.001$) and 'stable high' (highest wealth: $coef = 0.347$, $p < 0.01$) groups.

The relationship between Self-Rated Health (SRH) and memory varies significantly by trajectory group context. Compared with 'excellent' SRH, reporting 'good' or 'poor'

health is associated with a marked reduction in memory scores, primarily among those in the 'low and decreasing' group (Poor SRH: $\text{coef} = -0.598$, $\text{SE} = 0.303$, $p < 0.01$). Interestingly, this negative association is less pronounced or non-significant in the more optimal trajectory groups. This heterogeneity suggests that subjective health deterioration may have a more acute impact on memory performance for individuals who are already cognitively vulnerable, whereas those on 'stable high' paths may possess greater cognitive reserve to withstand physical health declines.

Being employed (relative to being retired) shows divergent within-group associations. In the 'low and decreasing' group, employment is linked to substantially higher memory scores ($\text{coef} = 0.960$, $\text{SE} = 0.208$, $p < 0.001$), while the effect is marginal in the 'stable medium' group ($\text{coef} = 0.137$, $p < 0.05$) and non-significant in the 'stable high' group. The exceptionally large coefficient of 0.960 in the disadvantaged group likely reflects a 'healthy worker' selection effect: within a declining cognitive group, only those with relatively higher functioning are able to maintain employment. Conversely, it may also point to the cognitive stimulation. For individuals on the verge of steep cognitive decline, the social interaction and mental demands of employment may offer a critical protective buffer—a 'use-it-or-lose-it' effect that is more vital for this vulnerable group than for those in higher-functioning, stable trajectories.

Table 2.6 Trajectories of memory and the effects of time-varying predictors on trajectory shapes, stratified by trajectory group.

	Low and decreasing	Stable medium	Stable high
Wealth tercile (Ref: lowest wealth)			
Middle wealth	0.240* (0.130)	0.227*** (0.072)	0.299** (0.125)
Highest wealth	0.519*** (0.163)	0.459*** (0.074)	0.347*** (0.119)
Self-rated health (Ref: excellent)			
Very good	-0.220 (0.249)	0.133 (0.100)	0.090 (0.117)
Good	-0.407* (0.240)	-0.031 (0.099)	-0.017 (0.125)
Fair	-0.213 (0.253)	-0.210* (0.115)	-0.068 (0.175)
Poor	-0.598** (0.303)	-0.183 (0.166)	0.043 (0.302)
Emplotment (Ref: retired)			
Employed	0.960*** (0.208)	0.137* (0.071)	-0.047 (0.096)
Unemployed	0.361 (0.220)	0.104 (0.116)	0.167 (0.164)
Disability (Ref: disable)			
Undisable	0.202 (0.131)	0.048 (0.074)	-0.035 (0.131)
Multimorbidity (Ref: yes >=2 chronic conditions)			
No < 2 chronic conditions	-0.736 (0.526)	-0.293 (0.316)	0.020 (0.566)

Notes: Coefficients are estimated from regressions of memory score on group-specific time-varying predictors. Each coefficient denotes the average within-trajectory change in memory score associated with a one-unit change in the predictor among observations assigned to that trajectory. Robust standard errors in parentheses. Models control for wave and baseline time-stable variables (gender, age groups, education) and use ELSA longitudinal weights based on Wave 4. Coefficients are on the memory score scale. Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

The coefficients for disability and multimorbidity (having two or more chronic conditions) are generally imprecise across groups. While the signs are largely negative, as expected, they do not reach robust statistical significance in most cells. This suggests that while these conditions are key markers of baseline health, their contemporaneous

fluctuations within a ten-year window may not be as immediately predictive of memory changes as subjective health perceptions or economic resources.

Table 2.7 reports group-specific coefficients estimating the contemporaneous associations between longitudinal fluctuations in socioeconomic and health predictors and depressive symptom scores (CES-D 8). In contrast to its strong role in cognitive trajectories, contemporaneous wealth tertiles show no consistent or precisely estimated associations with depressive symptoms within trajectories. Coefficients across most cells are small and statistically non-significant. This indicates that after conditioning on trajectory membership and other covariates, short-term within-person movement between wealth tertiles does not meaningfully predict immediate changes in CES-D scores. This lack of association may be due to the coarse measurement of wealth in tertiles, which may not capture the subtle financial fluctuations relevant to mental health, or the relative stability of wealth among ELSA participants over the follow-up period. Sensitivity analyses using equivalized wealth tertiles produced consistent results, reinforcing the conclusion that wealth's primary influence on mental health in later life may be more cumulative and structural (determining group membership) than immediate and episodic (Supplements Table S2.7 and S2.8).

Subjective health exhibits a graded and powerful contemporaneous association with depressive symptoms across all trajectories. Using 'excellent' health as the reference, the analysis reveals that even good SRH is associated with a slight increase in symptoms within the stable-low group (coef = 0.035, $p < 0.05$). The associations become substantially more pronounced as health deteriorates: reporting 'fair' SRH is linked to higher CES-D scores across the board, with the absolute magnitude of the effect increasing alongside the symptom level (e.g., stable-low: 0.124; stable-high: 1.287).

‘Poor’ SRH shows the largest absolute associations, particularly within the increasing (coef = 1.246, $p < 0.001$) and stable-high (coef = 1.580, $p < 0.001$) groups. These patterns suggest a vulnerability effect, where individuals already experiencing high symptom burdens are more psychologically sensitive to physical health shocks. However, these are contemporaneous associations; thus, while subjective health deterioration tracks closely with symptom burden, these estimates may be inflated by reporting bias or reverse causality (whereby depressed mood leads to a more pessimistic evaluation of one’s health).

The associations between employment status (vs. retired) and depressive symptoms are highly heterogeneous. Employment appears specifically protective for individuals in the increasing trajectory, where it is associated with a significant reduction in depressive symptoms (coef = -0.243 , SE = 0.091, $p < 0.01$). This is consistent with the social integration and activity hypotheses, which posit that the structured environment and social resources provided by work may buffer against rising distress. Conversely, the associations in other groups are either marginal (stable-medium: 0.078, $p < 0.05$) or non-significant. These divergent results necessitate a cautious interpretation, as they likely reflect complex selection processes regarding who remains in the workforce and the varying psychosocial quality of late-life employment.

Improved physical health is consistently linked to reduced depressive symptom burden. Not being disabled is associated with lower depressive scores in the ‘stable medium’ and ‘decreasing’ groups. Notably, having fewer than two chronic conditions (lower multimorbidity) is associated with substantially lower scores in these same groups (e.g., ‘stable medium’: coef = -0.688 , $p < 0.05$). These results underscore that while

trajectories remain stable, the contemporaneous relief of physical health burdens is significantly related to improved mental well-being within those paths.

Table 2.7 Trajectories of depressive symptoms and the effects of time-varying predictors on trajectory shapes, stratified by trajectory group.

	Stable low	Increasing	Stable medium	Decreasing	Stable high
Wealth tercile (Ref: lowest wealth)					
Middle wealth	-0.017 (0.017)	-0.026 (0.091)	-0.068 (0.044)	0.06 (0.085)	-0.166 (0.159)
Highest wealth	-0.018 (0.017)	-0.035 (0.092)	-0.073 (0.045)	0.053 (0.102)	-0.027 (0.238)
Self-rated health (Ref: excellent)					
Very good	0.016 (0.015)	0.007 (0.165)	-0.033 (0.058)	0.127 (0.184)	0.164 (0.603)
Good	0.035** (0.016)	0.203 (0.163)	0.033 (0.059)	0.227 (0.172)	0.538 (0.588)
Fair	0.124*** (0.031)	0.343** (0.174)	0.113 (0.070)	0.493*** (0.175)	1.287** (0.579)
Poor	0.209* (0.114)	1.246*** (0.304)	0.591*** (0.154)	0.709*** (0.203)	1.580*** (0.588)
Emplotment (Ref: retired)					
Employed	0.015 (0.015)	-0.243*** (0.091)	0.078* (0.045)	0.044 (0.122)	-0.152 (0.255)
Unemployed	0.04 (0.031)	-0.362*** (0.121)	0.128* (0.074)	0.197 (0.140)	0.15 (0.184)
Disability (Ref: disable)					
Undisable	-0.016 (0.021)	-0.121 (0.092)	-0.091* (0.048)	-0.177** (0.082)	-0.04 (0.159)
Multimorbidity (Ref: yes >=2 chronic conditions)					
No < 2 chronic conditions	-0.096 (0.076)	0.433 (0.304)	-0.688* (0.364)	-0.617* (0.330)	-0.576 (0.454)

Notes: Coefficients are estimated from regressions of depression score (CES-D 8) on group-specific time-varying predictors. Each coefficient denotes the average within-trajectory change in depression score associated with a one-unit change in the predictor among observations assigned to that trajectory. Robust standard errors in parentheses. Models control for wave and baseline time-stable variables (gender, age groups, education) and use ELSA longitudinal weights based on Wave 4. Coefficients are on the depression score scale. Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

2.5 Discussion

This study identified substantial heterogeneity in the longitudinal trajectories of memory and depressive symptoms among older adults in England using six waves of data from the English Longitudinal Study of Ageing (ELSA). For memory, three distinct groups emerged: a low and decreasing group showing gradual decline, a stable medium group with relatively consistent scores, and a stable high group maintaining superior performance over time. For depressive symptoms, five groups were observed, ranging from stable low levels to persistently high symptoms, with additional patterns of increasing and decreasing symptoms across waves. These findings align with prior studies—Su and Xiao (2022) reported heterogeneous cognitive trajectories in a Chinese cohort, while de la Torre-Luque et al. (2019) and Xiang (2019) identified chronic and fluctuating subgroups for depressive symptoms in late-life depression. Similarly, Tampubolon and Maharani (2017) demonstrated that depressive symptom trajectories vary across national contexts. Extending this body of work, my findings from ELSA highlight that dynamic groups, particularly those with declined memory and increasing depressive symptoms, represent a non-trivial share of the population. This underscores the importance of continuous monitoring and targeted interventions, rather than assuming homogeneity in late-life cognitive and mental health outcomes.

Time-constant socioeconomic and demographic factors were strongly associated with trajectory group membership: individuals with higher education were substantially more likely to maintain high memory function and low depressive symptoms, whereas older age, female gender, and early-life disadvantage (father's occupation at 14) predicted less favourable outcomes. An apparent discrepancy emerged between the descriptive patterns and the regression-based results in my findings. While baseline tabulations

suggested that individuals in the “stable high” depressive symptom group were relatively more likely to hold higher levels of education, the regression analyses—adjusting for age, gender, and other covariates—revealed that education was in fact protective against persistent depressive symptoms. This divergence likely reflects several factors. First, compositional effects may have played a role: individuals with higher education and wealth who nevertheless experienced depressive symptoms may represent a particularly resilient subgroup that is small in number but highly visible in descriptive distributions (Gondek, Lacey, Blanchflower, & Patalay, 2022). Second, reporting differences may also explain the unexpected socioeconomic patterns in depressive symptoms. Individuals with higher education or greater health awareness may be more willing to acknowledge mental health difficulties in surveys, consistent with evidence that non-disclosure is common in primary care settings, often due to stigma or concerns about treatment (Bell et al., 2011). Third, the regression models better capture the joint distribution of sociodemographic variables, revealing the underlying protective role of socioeconomic resources once confounding is addressed. Taken together, these findings underscore the importance of combining descriptive and multivariate perspectives: descriptive patterns highlight surface-level heterogeneity, whereas regression analyses disentangle net associations and reveal the broader socioeconomic gradients shaping late-life mental health.

Further analyses incorporating time-varying predictors revealed that poor self-rated health, disability, and multimorbidity were consistently associated with more depressive symptoms, which aligns with evidence from primary care populations, where chronic illness and multimorbidity have been shown to substantially increase the risk of depression (Gunn et al., 2012). Self-rated health exhibited a graded and powerful contemporaneous association with depressive symptoms across all trajectories, with the

magnitude of the effect increasing alongside baseline symptom level. This pattern suggests a vulnerability effect: individuals already experiencing high symptom burdens may be more psychologically sensitive to physical health shocks, or the association may be inflated by reporting bias and reverse causality, whereby depressed mood leads to more pessimistic health evaluations. Disentangling these possibilities is beyond the scope of the present analysis but represents an important direction for future research.

However, poorer self-rated health was associated with better memory. These patterns likely reflect selection and reserve mechanisms. High-memory individuals may represent a “healthy survivor” subgroup with greater cognitive resilience, consistent with the cognitive reserve hypothesis, which suggests that lifelong cognitive engagement enable individuals to maintain cognitive function despite health challenges (Stern, 2009).

Higher contemporaneous wealth was consistently and positively associated with memory scores across all trajectory groups, suggesting that economic resources continue to matter for cognitive function even conditional on an individual's underlying trajectory. For depressive symptoms, however, wealth showed no consistent within-trajectory association, implying that its primary influence on mental health in later life may be more cumulative and structural—determining trajectory membership—than immediate and episodic.

Notably, being in paid work was associated with better memory outcomes but, in some trajectory groups, with worse depressive symptom outcomes. This divergence is revealing. For memory, the association likely reflects both a “healthy worker” selection effect—whereby only those with relatively preserved cognition remain able to work—and a “use-it-or-lose-it” mechanism, wherein the cognitive demands and social

engagement of employment provide ongoing stimulation (Meng et al., 2017). For depressive symptoms, however, the picture is more complex. While employment can confer social integration, purpose, and economic security, it may also entail stress, role strain, and exposure to negative workplace experiences, particularly for those already experiencing elevated symptoms or facing precarious employment conditions (Wahrendorf et al., 2013). These countervailing forces may explain why employment does not uniformly protect against depressive symptoms in the same way it appears to protect cognition. This divergence underscores that resilience processes operate differently across health domains. Assuming a common ageing pathway risks obscuring the distinct mechanisms through which socioeconomic and health factors shape late-life cognition and mental health. Future research should explicitly model these domain-specific pathways rather than treating cognitive and mental health outcomes as interchangeable markers of successful ageing.

This study has several strengths compared with prior research. First, it uses nationally representative longitudinal data from ELSA, with repeated measures of both memory and depressive symptoms across more than a decade. This design enables the identification of fine-grained heterogeneity in trajectories and strengthens generalisability beyond regional or clinical samples used in earlier studies (e.g., Andreescu et al., 2008; Bergdahl et al., 2005). Second, the dual-outcome approach—modelling memory and depressive symptoms in parallel—represents an important advance over most prior work that has focused on a single domain. My findings show that these two outcomes, while shaped by overlapping sociodemographic and health factors, respond differently to life-course risks and resources: for example, paid work appears protective for memory but a potential stressor for depressive symptoms. This divergence underscores the need to examine cognitive and mental health trajectories side

by side rather than assuming uniform ageing pathways. Finally, by incorporating both early-life and adult socioeconomic indicators, the analysis highlights the long-term influence of life-course social conditions on later-life health.

Nonetheless, several limitations warrant consideration. First, although Group-Based Trajectory Modelling (GBTM) leverages all available observations, non-random attrition remains a potential source of bias. Respondents with poorer health or faster cognitive decline are more likely to drop out, which may lead to an underestimation of the true rate of decline in the general population (Haviland et al., 2011; Musliner et al., 2016). Second, while strong socioeconomic gradients were identified, the observational nature of this study limits causal inference compared with quasi-experimental designs that exploit exogenous policy changes (Glymour, 2008; Banks & Mazzonna, 2012). Regarding outcome measures, cognitive assessment was restricted to episodic memory (word-list recall). While this is a sensitive marker of age-related change, it may not capture the heterogeneity across other domains, such as executive function or processing speed (Clouston et al., 2020). This narrow focus is particularly relevant for interpreting sex differences, as women typically outperform men on verbal memory tasks despite facing similar or higher dementia risk (Sundermann et al., 2018; Mielke et al., 2014). Relying solely on episodic memory therefore risks mischaracterising the true landscape of cognitive ageing between sexes. Future research should incorporate a broader battery of cognitive tests and employ sex-stratified sensitivity analyses to validate these patterns (Stern, 2012). Finally, depressive symptoms were assessed using the CES-D 8, a brief screening instrument. While highly reliable for longitudinal tracking, it does not substitute for clinical diagnostic interviews and may not fully capture the complexity of subclinical presentations (Blazer, 2003; Luppa et al., 2012).

In conclusion, this chapter demonstrates the existence of distinct longitudinal trajectories of memory and depressive symptoms, underscoring the heterogeneity of ageing processes. The identification of both favourable and unfavourable pathways highlights the role of sociodemographic and health-related disparities in shaping late-life outcomes. These findings provide a conceptual and empirical basis for subsequent analyses, enabling a more refined understanding of cognitive and mental health dynamics in later life.

2.6 Supplements

Table S2.1 Comparison of goodness of fit criteria for group-based trajectory modelling
models of longitudinal memory

N groups	AIC	BIC	cBIC	Entropy	gp1(%)	gp2(%)	gp3(%)	gp4(%)	gp5(%)	gp6(%)	gp7(%)
1	-149172	-149194	-149191		100						
2	-141596	-141640	-141633	0.674	38.267	61.733					
3	-138367	-138434	-138424	0.688	17.609	51.391	31.000				
4	-137111	-137200	-137186	0.671	11.395	28.711	45.244	14.649			
5	-136557	-136669	-136652	0.651	7.254	14.802	35.251	34.385	8.308		
6	-136338	-136472	-136451	0.657	7.846	2.571	16.073	35.847	30.533	7.130	
7	-136102	-136258	-136234	0.635	3.228	9.017	19.706	2.669	34.916	24.816	5.649

Table S2.2 Comparison of goodness of fit criteria for group-based trajectory modelling
models of longitudinal memory

Ngroups	AIC	BIC	cBIC	Entropy	gp1(%)	gp2(%)	gp3(%)	gp4(%)	gp5(%)	gp6(%)	gp7(%)
	-	-	-								
1	102286	102304	102301		100						
2	-83191	-83231	-83225	0.769	65.843	34.157					
3	-80311	-80374	-80364	0.667	40.697	40.207	19.096				
4	-79802	-79886	-79873	0.606	30.806	37.411	21.214	10.570			
5	-79458	-79564	-79548	0.562	27.020	14.179	34.242	12.446	12.113		
6	-79265	-79394	-79375	0.552	7.594	27.836	12.454	30.715	10.796	10.604	
7	-79145	-79297	-79274	0.562	6.846	29.382	9.202	29.387	3.732	11.757	9.694

For depression, the 5-group model is chosen. It demonstrates improved fit over simpler models while maintaining meaningful group distinctions and avoiding redundancy, making it the most informative and parsimonious solution.

Table S2.3 Estimates of unconditional group-based trajectory models for memory

Group	Parameter	Estimate	SE	P value
1	Intercept	-0.53868	2.3462	0.8184
	Linear	3.83295	1.19133	0.0013
	Quadratic	-0.70367	0.19387	<0.001
	Cubic	0.03736	0.01015	<0.001
2	Intercept	9.55353	0.26907	<0.001
	Linear	0.30869	0.08709	<0.001
	Quadratic	-0.03538	0.0067	<0.001
	Intercept	12.35672	0.34411	<0.001
3	Linear	0.47289	0.10983	<0.001
	Quadratic	-0.0381	0.00835	<0.001

Table S2.4 Estimates of unconditional group-based trajectory models for depressive symptoms

Group	Parameter	Estimate	SE	P value
1	Intercept	0.72312	0.48786	0.1383
	Linear	-0.95559	0.15763	<0.001
	Quadratic	0.07916	0.01195	<0.001
2	Intercept	-1.56545	0.29545	<0.001
	Linear	0.61018	0.07863	<0.001
	Quadratic	-0.03511	0.00542	<0.001
3	Intercept	0.00691	0.21133	0.9739
	Linear	-0.11654	0.06403	0.0687
	Quadratic	0.01307	0.00458	0.0043
4	Intercept	2.15847	0.09889	<0.001
	Linear	-0.23635	0.02016	<0.001
5	Intercept	1.56752	0.0112	<0.001

Table S2.5 Average memory scores over time across classes

Group	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9
Low and decreasing	5.93	5.70	5.20	4.63	4.22	4.20
Stable medium	10.22	10.21	10.13	9.98	9.76	9.47
Stable high	13.64	13.77	13.82	13.80	13.70	13.53

Table S2.6 Average depressive symptom scores over time across classes

Group	Wave4	Wave5	Wave6	Wave7	Wave8	Wave9
Stable low	0.11	0.12	0.07	0.12	0.12	0.17
Increasing	1.19	2.02	2.39	2.81	3.12	3.20
Stable medium	0.82	0.94	0.78	0.89	0.97	1.13
Decreasing	3.63	2.88	2.20	1.54	1.21	1.06
Stable high	4.87	5.15	5.01	4.95	4.84	4.95

Table S2.7 Trajectories of memory and the effects of time-varying predictors on trajectory shapes, stratified by trajectory group (using equivalized wealth)

	Low and decreasing	Stable medium	Stable high
Wealth tercile (Ref: lowest wealth)			
Middle wealth	0.297** (0.130)	0.274*** (0.073)	0.153 (0.127)
Highest wealth	0.506*** (0.161)	0.456*** (0.074)	0.260** (0.120)
Self-rated health (Ref: excellent)			
Very good	-0.219 (0.249)	0.134 (0.100)	0.093 (0.117)
Good	-0.408* (0.240)	-0.033 (0.099)	-0.019 (0.125)
Fair	-0.205 (0.253)	-0.214* (0.115)	-0.077 (0.176)
Poor	-0.599** (0.302)	-0.183 (0.166)	0.004 (0.304)
Employment (Ref: retired)			
Employed	0.972*** (0.208)	0.150** (0.071)	-0.045 (0.097)
Unemployed	0.379* (0.220)	0.119 (0.116)	0.162 (0.165)
Disability (Ref: disable)			
Undisable	0.206 (0.131)	0.054 (0.074)	-0.030 (0.131)
Multimorbidity (Ref: yes ≥ 2 chronic conditions)			
No < 2 chronic conditions	-0.748 (0.525)	-0.292 (0.317)	0.022 (0.562)

Table S2.8 Trajectories of depressive symptoms and the effects of time-varying predictors on trajectory shapes, stratified by trajectory group (using equivalized

	Stable low	Increasing	Stable medium	Decreasing	Stable high
Wealth(eq) tercile (Ref: lowest wealth)					
Middle wealth	-0.021 (0.017)	-0.094 (0.092)	-0.030 (0.044)	0.134 (0.087)	-0.048 (0.160)
Highest wealth	-0.011 (0.016)	-0.030 (0.094)	-0.035 (0.045)	0.182* (0.098)	0.145 (0.220)
Self-rated health (Ref: excellent)					
Very good	0.017 (0.015)	0.009 (0.165)	-0.031 (0.058)	0.135 (0.184)	0.189 (0.595)
Good	0.036** (0.016)	0.205 (0.164)	0.037 (0.059)	0.246 (0.172)	0.577 (0.577)
Fair	0.127*** (0.031)	0.342** (0.175)	0.121* (0.070)	0.515*** (0.176)	1.349** (0.568)
Poor	0.212* (0.114)	1.240*** (0.305)	0.598*** (0.155)	0.748*** (0.204)	1.633*** (0.577)
Emplotment (Ref: retired)					
Employed	0.015 (0.015)	-0.243*** (0.091)	0.078* (0.045)	0.044 (0.122)	-0.152 (0.255)
Unemployed	0.04 (0.031)	-0.362*** (0.121)	0.128* (0.074)	0.197 (0.140)	0.15 (0.184)
Disability (Ref: disable)					
Undisable	-0.017 (0.021)	-0.122 (0.092)	-0.093* (0.048)	-0.183** (0.082)	-0.064 (0.158)
Multimorbidity (Ref: yes >=2 chronic conditions)					
No < 2 chronic conditions	-0.095 (0.076)	0.446 (0.303)	-0.693* (0.365)	-0.630* (0.329)	-0.581 (0.452)

Chapter 3

The Impact of Compulsory Schooling on Cognition and Its Interaction with Genetic Scores

3.1 Introduction

Cognitive abilities, both fluid (e.g., memory and processing speed) and crystallized (e.g., vocabulary and numeracy) abilities (Horn & Cattell, 1967), are important predictors of socioeconomic attainment, aspects of decision-making, health, and longevity (McArdle, 2011; Der, 2009; Mata, 2007; Shipley, 2006). Age-related declines in cognitive abilities are linked to impairments in older adults' everyday functions, such as financial decisions, retirement plans and health care decisions. Hence, identifying factors that protect against compromised cognitive functioning in older age can help inform the design of interventions to prevent cognitive declines. Cognitive function in older age is the result of a dynamic process throughout the life course based on a combination of genetic, biological, social and environmental factors. It has been increasingly accepted that the aetiology of most common diseases involves not only discrete genetic and environmental factors but also interactions between genetic susceptibility and environmental exposures. (Tucker-Drob & Briley, 2014).

Understanding the long-term effects of education on later-life cognition requires distinguishing and integrating several complementary mechanisms. A central theoretical account is cognitive reserve: higher educational attainment is thought to foster more efficient, flexible and compensatory patterns of neural processing that allow some individuals to tolerate greater neuropathology before clinical deficits emerge (Stern 2002; Stern 2009). This concept is distinct from brain reserve, which emphasises

structural capacity (for example greater synaptic density or brain volume) (Valenzuela & Sachdev 2006). Education can therefore operate through at least two empirically distinguishable pathways: by raising baseline cognitive performance (a level effect) and/or by enhancing compensatory mechanisms that delay clinical manifestation of underlying pathology (a buffer or tolerance effect) (Stern 2009; Scarmeas et al. 2006). Beyond neurobiology, education shapes cognition via socio-behavioural pathways. Higher attainment facilitates entry into cognitively complex occupations and fosters stimulating lifestyles, providing sustained "cognitive exercise" across the life course (Richards & Deary 2005; Scarmeas 2013). Additionally, education-driven advantages in health behaviours and healthcare access mitigate vascular risks implicated in cognitive decline (Cutler & Lleras-Muney 2010; Barnes & Yaffe 2011). Observed education–cognition associations likely reflect a composite of biological buffering, lifelong stimulation, and socio-economic mediation.

The association between educational attainment and cognitive function in older age has already been well-established in the empirical literature (Arvanitakis et al., 2004; McArdle et al., 2002; Zahodne et al., 2011). To address the methodological challenges of reverse causation and omitted variables, recent studies have explored the causality between educational attainment and cognitive function using quasi-experimental methods by exploiting exogenous policy shocks that lead to differences in educational attainment. One leading example of these policy changes is compulsory schooling laws (CSLs), which establish a minimum school-leaving age and require school-age children to attend school for a certain period (Banks & Mazzonna, 2012; Crespo, 2014; Glymour, 2008; Mazzonna, 2012; Mazzonna, 2014; Nguyen, 2016; Schneeweis, 2014). Most of these studies support the hypothesis that the increase in schooling is associated with better cognitive functioning in later life, though some report effects in the opposite

direction (Schneeweis et al., 2014). Glymour et al. (Glymour, 2008) investigated state compulsory schooling laws in the US using an instrumental variable (IV) approach and found that each additional year of schooling improved memory scores by 0.18 standard deviations. Similar findings were reported using a regression discontinuity design based on the 1947 compulsory schooling reform in England, 1947 (Banks & Mazzonna, 2012). They found that the effect of an additional year of education was approximately 0.5 of a standard deviation for males and around 0.4 for females on memory scores and ranged between 0.37 and 0.63 of a standard deviation for males on executive functioning scores. Nevertheless, another IV analysis (Schneeweis et al., 2014) exploited the compulsory schooling reforms in six European countries and found no causal effects of education on the level of fluency, numeracy, and orientation-to-date tasks, where respondents were asked to report the current day, month, year, and day of the week..

Some studies also suggested that genes can moderate the relationship between education and cognition (McArdle & Prescott, 2010; Seeman et al., 2005). Using the U.S. Health and Retirement Study (HRS) data, McArdle and Prescott (McArdle & Prescott, 2010) found steeper memory declines for APOE4 carriers compared to APOE4 non-carriers only among those with 8 years of education or less (Shanahan & Hofer, 2005). Wang et al. (2012) reported the relationships between APOE genotype and education with dementia among 3436 participants aged 65 and over in Northern Europe. The findings suggested that early-life higher education may buffer the negative effect of genetic factors on late-life dementia. As one possible mechanism, they mentioned the cognitive reserve hypothesis, and proposed highly intellectual environments might allow people to cope better with brain pathology than others (Stern, 2009). The interaction effect of APOE4 and educational attainment is contingent upon sex because men and women show different levels in both cognitive function and educational attainment (Ishioka et

al., 2016). However, the evidence on the interaction effect between APOE4, education, and sex on cognition especially for different cognitive domains is still scarce. Further research needs to improve causal inference for the interaction effect of APOE4 and educational attainment on cognition (Smith et al., 2018).

Gene-environment interaction ($G \times E$) processes impacting cognitive function likely play out over the entire life course, with special vulnerability during early life (Reynolds et al., 2014; Scarr & McCartney, 1983). The availability of genetic data in large population surveys has resulted in an explosion of research on gene-environment interactions. Understanding the effects of critical social and environmental exposures in sensitive periods on later life cognition and how these vary by underlying genetic susceptibility can help inform targeted interventions going forward. For the “E” part in $G \times E$, the environment is not randomly assigned in most cases, may be confounded by other non-genetic variables, and could be correlated with the genetic variant in question, which is called gene-environment correlation (rGE). This study will incorporate quasi-experimental designs into the $G \times E$ framework to leverage exogenous environmental variation to reduce the influence of gene-environment correlation and strengthen the ability to demonstrate causal inferences.

This study aims to examine the causal effects of education on cognitive function and how the impact might vary by genetic susceptibility combined with a quasi-experimental design, to improve causal inference for the gene-environment interactions that shape cognitive decline. I examined the 1947 reform of compulsory schooling in England, which raised the minimum school-leaving age from 14 to 15 for individuals born after a specific cut-off in March 1933. This policy provides a natural experiment to study the causal effects of education on later-life cognitive outcomes and its interaction with

genetic risk. I built my analyses based on the research design of Banks and Mazzonna (Banks & Mazzonna, 2012), which exploited the reform of compulsory schooling in England in 1947, and used a fuzzy regression discontinuity design to compare individuals born before and after the cut-off, to examine the effect of an additional year of education on memory and executive functioning scores by sex. This study further incorporated the genetic risks measured by APO4 status and polygenic scores to examine how the effect of an additional year of education on cognitive function interacted with genetic risks.

3.2 Data

This study uses Wave 1 to Wave 9 (2002-2003 to 2018–2019) from the English Longitudinal Study of Ageing (ELSA). These data include consistent measurement of cognitive function, including memory and executive function. I also use the ELSA Special License data which includes the “month-of-birth” variable and the Genetic data. I then use the year- and month-of-birth variables to construct the month-year birth cohorts to identify those exposed to the compulsory schooling reform. Since memory was measured across nine waves in ELSA but verbal fluency, which is one of the components of executive function, was only collected from Wave 1 to Wave 5, I create two analytical samples. The first sample is comprised of participants with the month of birth, genetic, and memory data from the ELSA Wave 1 to Wave 9. The second sample includes participants with measures of the month of birth, genetic, and executive function data from the ELSA Wave 1 to Wave 5. The broader study design and sampling procedures were detailed in Section 1.6.2.

The analytical sample is restricted to participants who provided valid phenotypic data during the pandemic and were included in the ELSA Genetic database (genetic sample $N = 7,183$). As detailed in Section 1.6.4, these genetic data consist of high-density genotypes of European ancestry, from which Polygenic Scores (PGSs) were derived following the harmonized HRS methodology (Ware et al., 2017)

Measures

Polygenic Scores

This essay tests whether the effect of education on cognitive function varies by a person’s genetic risks as measured by APOE4 status, polygenic scores for general

cognition (2018) and Alzheimer's disease (AD) (2019). Cognitive function is a polygenic trait, but the APOE gene is an exceptional case. The APOE gene is a major risk allele for late-onset Alzheimer's disease and explains 13% of the heritability in AD and the presence of each risk allele translates into a fivefold increase in the risk of developing AD (Battista et al., 2016). Therefore, I include both the APOE4 status and polygenic scores related to cognitive function. Continuous polygenic scores are standardized in the regression analyses.

The apolipoprotein E (APOE) genotype was based on the two single nucleotide polymorphisms (SNPs) for APOE, rs429358 and rs7412 that comprise the ϵ 2, ϵ 3 and ϵ 4 alleles. APOE4 carrier status was determined based on the presence or absence of one or more APOE ϵ 4 alleles (APOE4). In the sample of 7181 individuals, 25.3% are APOE4 carriers.

The PGSs for general cognition (2018) in this project were created by ELSA using results from a 2018 Genome-Wide Association Study (GWAS) including 300,486 individuals across 57 population-based cohorts brought together by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), CHARGE), the Cognitive Genomics Consortium (COGENT) consortia, and UK Biobank. A total of SNPs was included in the meta-analysis summary statistics, 1,348,174 SNPs overlapped with the ELSA genetic database and were included in the PGS for the general cognition phenotype. In the sample of 7181 individuals, the mean of the PGS for GC is 1352024.8 and the standard error is 74.0.

The PGSs for Alzheimer's disease (AD) (2019) were created using results from a large genome-wide association study of clinically diagnosed AD and AD-by-proxy (71,880 cases, 383,378 controls), in which participants were obtained from multiple sources,

including raw data from case-control samples collected by PGC-ALZ and ADSP, summary data from the case-control samples in the IGAP, and raw data from the population-based UKB sample which was used to create a weighted AD-by-proxy phenotype. Adjustment covariates within each contributing cohort included age, sex, and genetic principal components. The PGS for AD contains 1,712,973 SNPs with a mean of 5597.3 and a standard error of 0.08.

Cognitive function

The cognitive function module in the ELSA contains measures of memory, executive function, speed of processing and numerical ability domains. This study examines memory, consistently measured from Wave 1 to Wave 9, and executive function outcomes, measured from Wave 1 to Wave 5.

Memory was measured using a word-list learning test in which a list of 10 words was presented orally to study participants who were then asked to recall as many words as possible immediately after the reading of the list had been completed and then again after around a 5 min delay during which they completed other survey questions. I computed an overall memory score (ranging from 0 to 20) by adding the points of the immediate and delayed recall tests (maximum of 10 points for immediate and 10 points for delayed recall). The memory score has been standardised by subtracting their mean and dividing by their standard deviation.

The executive function measure was constructed by summing the two standardized scores of verbal fluency and letter cancellation. The verbal fluency test is a test of how quickly participants can name as many different animals as possible in one minute (semantic verbal fluency). The overall score ranges from 0 to 60. The letter cancellation test is a test of attention, visual search, and mental speed. The participant is handed a

clipboard with a page of random letters of the alphabet set out in rows and columns and is asked to cross out as many target letters (P and W) as possible within one minute. The page comprises 26 rows and 30 columns, and there are 65 target letters in all.

Respondents are asked to work across and down the page as though they were reading and to perform the task both as quickly and as accurately as possible. After one minute they are asked to mark the point on the page where they stopped. The total number of letters searched provides a measure of the speed of processing, whilst the number of target letters (P and W) missed by the respondent provides a measure of accuracy. These procedures are outlined in the ELSA cohort documentation (Banks et al., 2006; Steptoe et al., 2013).

3.3 Analytical Strategy

This study exploited the variation in the minimum school leaving age from 14 to 15 in March 1947 in England. The Education Act 1944 raised the minimum school-leaving age from 14 to 15 years old in England, Scotland and Wales, and this took effect from April 1, 1947, which means that students born before April 1, 1933 could drop out at age 14, whilst those born after that date could not leave until they turned 15, generating a discontinuity in the relationship between education and the date of birth at the April 1, 1933 cut-off point.

The study considers different subsamples for various descriptive and empirical analyses. First, I consider birth cohorts. I select my sample from zero year to five years on either side of the cut-off date of April 1, 1933. Second, I consider educational attainment, including the full sample, the subsample of those who left full-time education before the age of 19 (age left <19), and the subsample of those who left full-time education before the age of 16 (age left <16). The subsample (age left <19) excludes individuals who went to college but still contains many "always takers" who would have left full-time education after 15 regardless of the 1947 reform. The threshold of age 19 was chosen because it corresponds to the normative completion age for upper-secondary education (A-levels or equivalent) in the British system, and thus represents the natural institutional boundary between secondary and tertiary education for the cohorts under study. Individuals who continued beyond age 19 entered higher education — a pathway whose uptake was determined by prior academic attainment and family background rather than by the compulsory leaving age, and which was therefore unaffected by the 1947 reform. The age-19 cut-off is consequently not arbitrary but reflects the structural divide between those whose educational careers fell within the reach of compulsory

schooling legislation and those whose trajectories were governed by the selective entry requirements of tertiary institutions. The subsample (age left <16) is the most restrictive and contains the biggest concentration of "compliers" who would have left full-time education after 15 if and only if they were born after the cut-off date. The cut-off of age 15 — operationalised here as leaving before age 16 — corresponds directly to the minimum school-leaving age imposed by the 1947 Education Act, the reform being exploited as the quasi-experimental instrument in this analysis. This sample therefore contains the most directly treated individuals, and since the local average treatment effect (LATE) is the average of the unit-level causal effects for the "compliers", the last sample (age left <16) is the most treated and important subsample in this analysis, as further confirmed by the first-stage F-statistics reported in Table 3.5.

Table 3.1 shows the distribution of the full-time education leaving age by sex in our sample for the cohorts born five years on either side of the cut-off date for the 1947 reform. The most obvious large change in educational attainment across the reform dates is from the age of 14 to 15, especially for males. The change in educational attainment for females is slightly different with a big increase in the proportion of women leaving school between 16 and 18.

Table 3.1 Distribution of age of leaving full-time education by sex and date of birth
(individuals born five years on either side of the cut-off date of April 1, 1933) (N, %)

Age of Leaving School	Female		Male	
	Born before	Born after	Born before	Born after
	April 1, 1933	April 1, 1933	April 1, 1933	April 1, 1933
Left school at or before 14	1253 (54.2%)	173 (4.7%)	1121 (58.4%)	171 (5.7%)
Left school at 15	284 (12.3%)	1754 (48.1%)	138 (7.2%)	1545 (51.3%)
Left school between 16-18	596 (25.8%)	1246 (34.2%)	506 (26.4%)	851 (28.2%)
Left school at or over 19	179 (7.7%)	472 (12.9%)	153 (8.0%)	447 (14.8%)
Total	1958(100%)	3111(100%)	1958(100%)	3111(100%)

Notes: Sample restricted to ELSA respondents born within five years on either side of the April 1, 1933 compulsory schooling reform cut-off. N and column percentages shown by sex and birth cohort (before vs. after the cut-off). The sharp decline in the proportion leaving school at or before age 14 after the cut-off illustrates the first-stage discontinuity exploited in the regression discontinuity design.

Figure 3.1 visually shows the discontinuity in educational outcomes for birth cohorts by sex for individuals born five years on either side of the cut-off date. The points are the fractions of individuals who left full-time education by age 14 for a particular month of birth; the vertical line indicates the cut-off date of April 1, 1933; the fitted lines are a local linear fit. Before the cut-off date of April 1, 1933, around 55% of the individuals left full-time education by age 14, whilst after the cut-off date this fraction dropped to approximately 5%.

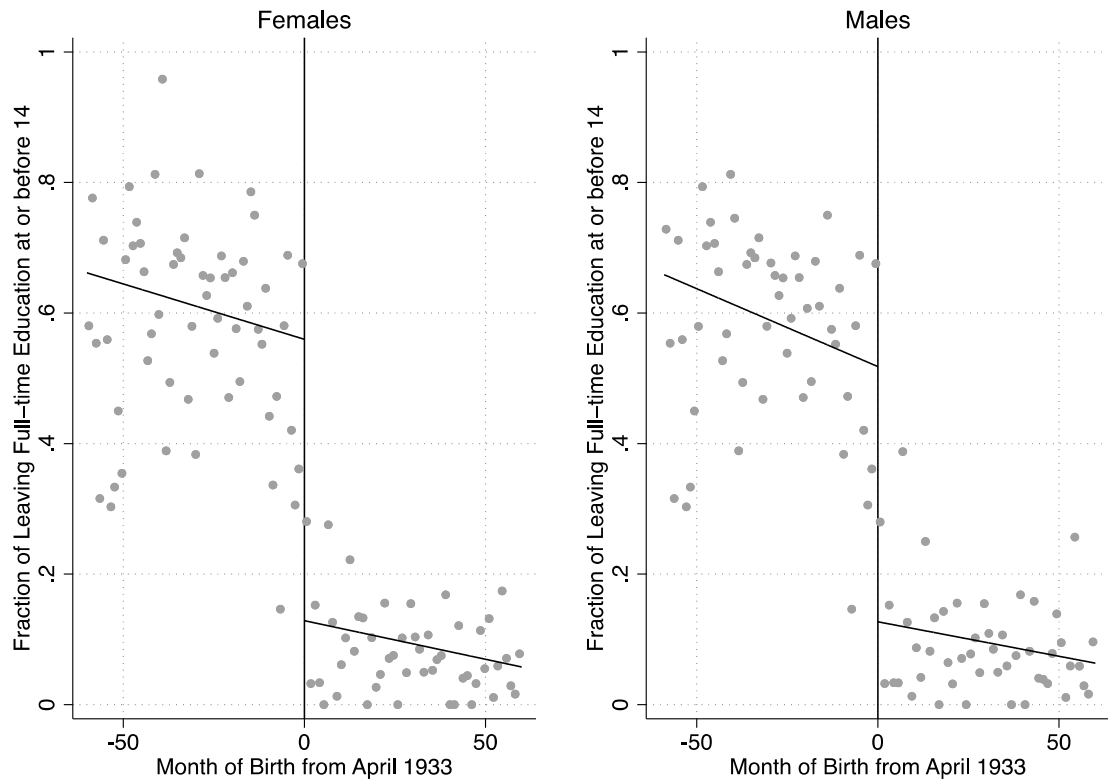


Figure 3.1 Fraction leaving school by age 14 by month of birth for males and females

Notes: Each point represents the fraction of individuals leaving full-time education by age 14 within a given birth month cohort. The vertical dashed line marks the April 1, 1933 compulsory schooling reform cut-off. Fitted lines are local linear regressions estimated separately on each side of the cut-off. Sample restricted to individuals born within five years of the cut-off date. The sharp discontinuity at the cut-off provides the identifying variation for the fuzzy regression discontinuity design.

The graph demonstrates the essence of the fuzzy RDD we adopt to estimate the effects of educational attainment since the discontinuous relationship between the month of birth and the fraction of individuals left full-time education by age 14 is clear. Even though not all the individuals born before the cut-off date left full-time education by age 14, their probability of leaving full-time education by age 14 is significantly higher than that of individuals who were born after the cut-off date. Since the probability of treatment changes discontinuously at the cut-off point, it is possible to determine the treatment effect in the case of partial compliance by comparing outcomes of individuals in a narrow range on either side of the cut-off point and scaling the difference in outcomes by the difference in the probability of treatment.

Table 3.2 shows the comparison of cognitive outcome scores and other characteristics by sex for those who left full-time education before the age of 19 (the upper set of rows) and those who left before 16 (the lower set of rows). The third column for each sex reports p-values from t-tests on the equality of means (p-values from Chi-squared test for APOE4 status). This is a descriptive comparison, which is important for our analysis as differences would have cast doubt on the “continuity hypothesis”, the main assumption of our estimation strategy. For both sexes and both educational subsamples, memory scores and executive function scores are significantly different between individuals born before and after the cut-off date.

Notably, for both sexes in the subsample of those who left full-time education before the age of 19, adult height is significantly larger among those born after the cut-off point than those born before. Therefore, adult height must be included in the following regression models as a proxy for unobserved early-life cohort conditions that may confound the relationship between birth cohort and both educational attainment and cognitive function (Case & Paxson, 2008). Height at adulthood reflects cumulative nutritional status, disease exposure, and socioeconomic circumstances during childhood and adolescence—factors that co-vary with birth cohort and are not fully captured by calendar-year cohort trends alone. The cohorts in this analysis straddle the immediate aftermath of the Great Depression (born 1928–1938), a period of economic disruption with potentially heterogeneous effects on childhood environments. As Table 3.2 shows, individuals born after the April 1933 cut-off are slightly taller on average than those born before it, suggesting that cohort-level differences in early-life conditions are present in the data. By controlling for adult height, the models absorb this residual cohort-level variation, reducing the risk that discontinuities in cognitive outcomes at the

cut-off are confounded by differences in early-life conditions rather than driven by the schooling reform itself.

Table 3.2 Cognition scores and comparison by sex and age of leaving full-time education (for individuals born 5 years before and after the 1933 cut-off)

	Female			Male		
	Before	After	P-value	Before	After	P-value
Left school before 19						
Memory (std.)	-0.45	-0.18	<0.001	-0.54	-0.35	<0.001
Executive function	-0.38	-0.12	<0.001	-0.66	-0.39	<0.001
Height	153.9	157.0	<0.001	168.3	170.3	0.009
PGS for GC	1351993	1351972	0.901	1351815	1351745	0.714
PGS for AD	5597	5597	0.829	5597	5597	0.001
APOE4 present	24.6%	23.9%	0.525	22.7%	22.8%	0.933
N	2222	3301		1805	2664	
Left school before 16						
Memory (std.)	-0.54	-0.29	<0.001	-0.68	-0.47	<0.001
Executive function	-0.49	-0.20	<0.001	-0.82	-0.52	<0.001
Height	155.1	155.4	0.709	169.3	168.9	0.656
PGS for GC	1351657	1351752	0.639	1351364	1351548	0.406
PGS for AD	5597	5598	0.209	5597	5597	0.001
APOE4 presented	25.4%	25.0%	0.762	23.0%	20.5%	0.095
N	1626	2055		1299	1813	

Note: Each domains of cognitive scores are standardised z-scores derived from the full analytical sample (N=44,832). The composite executive function standardized score is constructed by summing the two standardised scores of verbal fluency and letter cancellation. A mean of zero represents the total sample average; negative values indicate performance below this benchmark. The negative means for the 5-year bandwidth cohorts reflect the greater average age and specific educational profiles of these individuals relative to the broader ELSA population.

There are a number of studies that have explored the causal effect of educational attainment on cognitive function/ health outcomes using changes in compulsory schooling laws (CSLs) (Banks & Mazzonna, 2012; Crespo, 2014; Glymour, 2008; Mazzonna, 2012; Mazzonna, 2014; Nguyen, 2016; Clark & Royer, 2013) and two that exploited the 1947 British School-leaving Age Reform (Banks & Mazzonna, 2012); (Clark & Royer, 2013). This study extends that literature by exploring how the causal effect of education on cognitive function varies by a person's genetic risks. Since the relevant compulsory schooling changes depended on month-year of birth, a fuzzy regression discontinuity design (RDD) (Imbens & Lemieux, 2008; Lee & Lemieux, 2010) will be used to identify the causal effect of the policy-induced additional year of education change and how this interacts with genetic risk for cognitive decline. Specifically, this study will estimate the following two equations:

$$\widehat{E}_{ic} = \alpha_0 + \alpha_1 D_{ic} \times a(G_{ic}) + \alpha_2 D_{ic} + \alpha_3 a(G_{ic}) + f(R_{ic}) + \alpha_4 X_{icw} + \mu_{icw} \quad (3.1)$$

$$CF_{icw} = \beta_0 + \beta_1 \widehat{E}_{ic} \times a(G_{ic}) + \beta_2 \widehat{E}_{ic} + \beta_3 a(G_{ic}) + g(R_{ic}) + \beta_4 X_{icw} + \omega_{icw} \quad (3.2)$$

where CF_{icw} is a measure of cognitive function for individual i in birth cohort c in the ELSA wave w ; E_{ic} is a dummy variable indicating whether a respondent leaves compulsory schooling after age 15; D_{ic} is a dummy variable indicating whether a respondent belongs to a post-reform cohort; R_{ic} is a respondent's birth month cohort relative to the relevant cut-off (March 1933 for the 1947 reform). R_{ic} is positive when the respondent is born after the reform and negative when born before the reform; the Function $f(R_{ic})$ and $g(R_{ic})$ capture the associations between the birth cohort with educational attainment and cognitive outcomes; the vector X_{icw} contains age, age squared, adult height, year of the survey and month-of-birth dummies and μ_{icw} , ω_{icw} is

the residual. E_{icw} and $E_{icw} \times a(G_{ic})$ are endogenous gene-environment interaction terms and D_{ic} and $D_{ic} \times a(G_{ic})$ are used as instruments; $a(G)_{ic}$ represents the individual's genotype, which is linear combinations of genetic risks, as measured by APOE4 status, PGS for GC or PGS for AD in separate models, and the first ten principal components (PC_1-PC_{10}) of the genetic data to correct for population stratification (Price et al., 2006).

$$a(G_{ic}) = a_0 + a_1(APOE4_i/PGS\ GC_i/PGS\ AD_i) + a_2PC_i \quad (3.3)$$

First, I included one of PGSs and its interaction term in the model separately, to investigate the individual effect of the single education and PGSs interaction on cognitive outcomes (Model1-3). Then I added three PGSs and their interaction terms simultaneously in a model to estimate the effect of multiple education and PGSs interactions (Model 4). Specifically:

In Model 1 [APOE4 Model]:

$$a(G_{ic}) = a_0 + a_1(APOE4_i) + a_2PC_i \quad (3.4)$$

In Model 2 [PGS GC Model]:

$$a(G_{ic}) = a_0 + a_1(PGS\ GC_i) + a_2PC_i \quad (3.5)$$

In Model 3 [PGS AD Model]:

$$a(G_{ic}) = a_0 + \frac{a}{b_1} (PGS\ AD_i) + a_2PC_i \quad (3.6)$$

In Model 4 [Three-PGSs Model]:

$$a(G_{ic}) = a_0 + a_1(APOE4_i) + a_2(PGS\ GC_i) + a_3(PGS\ AD_i) + a_4PC_i \quad (3.7)$$

Correlations between these polygenic scores and other variables in this study are shown in Table 3.3.

Table 3.3 Correlations between Study Variables

Variables	Memory	Executive Function	Leaving School before 15	PGS GC	PGS AD	APOE4	Age	Sex	Height
Memory	1								
Executive Function	0.388*	1							
Leaving School before 15	0.270*	0.222*	1						
PGS GC	0.083*	0.070*	0.019*	1					
PGS AD	-0.018*	-0.024*	-0.004	0.117*	1				
APOE4	-0.032*	-0.004	0.028*	0.025*	0.037*	1			
Age	-0.420*	-0.323*	-0.447*	0.011*	- 0.011*	-0.031*	1		
Sex	-0.088*	-0.074*	-0.008*	-0.001	- 0.010*	-0.007	- 0.008*	1	
Height	0.030*	0.039*	0.035*	0.006	0.020*	-0.001	-0.070	0.235*	1

Significance: * $p < 0.05$.

To estimate the effect of the reform on schooling, I use local linear regression around the discontinuity choosing the optimal bandwidth in a cross-validation procedure suggested by Imbens and Lemieux (Imbens & Lemieux, 2008), which is detailed in the supplemental chapter (Supplements Section S3.1). In brief, the cross-validation

procedure selects the optimal bandwidth that performs the best in an out-of-sample predictive test. The strategy substantially reduces the chances that bias will be introduced by using a smaller portion of the sample. To evaluate the effect of the exogenous increase in education on cognitive outcomes, I make use of the 2SLS estimator, using the same bandwidth and polynomial in the treatment and outcome regression, to obtain directly the correct standard errors that are robust and clustered at the individual level.

Moreover, although this analysis used a nationally representative sample of older individuals living in England, the genotyping work was conducted in 2013-2014. The results may be subject to mortality selection (Imbens & Lemieux, 2008; Lee & Lemieux, 2010). Due to mortality selection in longitudinal social surveys, the sample of respondents who were genotyped in the genetic data was healthier, better educated, and longer-lived as compared with birth-cohort peers in the general ELSA data. As shown in Table 3.4 in the following section, the included genotyped sample in this analysis is generally older, more educated, and healthier. Thus, I applied inverse probability weighting (IPW) to my analytical data to account for the mortality selection and differential participation in the genetic sample (Domingue, Belsky, et al., 2017).

In vector X, confounders such as age are typically used as covariates to control for confounding effects. The inclusion of height helps control for any unobserved cohort effects that cannot be captured by cohort trends. The cohort affected by the reform was born after March 1933, i.e., in the immediate aftermath of the Great Depression, and it is unclear if and how the Great Depression affected the childhood environment (and thus education and health) of the cohorts in our analytical sample. For instance, we find some indication that individuals born after the cut-off point are slightly taller than older

cohorts. Thus, I control for adult height in the regressions to avoid any bias due to the lack of balance.

Even with potential confounding variables controlled in the regression, this approach can only control for the additive effects of covariates; it does nothing to control for the potential confounding effects that these covariates might have on the interaction itself (Garatachea et al., 2014). To properly control for confounding effects in the $G \times E$ research, I also controlled for all *Covariate* $\times$ *Gene* and *Covariate* $\times$ *Environment* effects.

Even though our identification is not strictly RDD, I carried out recommended robustness checks for RDD. In RDD, data manipulation near the cut-off point by sorting behaviour or anticipation effects is examined by the McCrary test (Aiken et al., 1991). However, because the running variable R_{ic} , the respondent's birth month cohort relative to the relevant cut-off, in this analysis is time specifications, the discontinuity of this type of running variable is not testable. If the RDD results are just mechanical correlations, or only affected by other unobservable factors, it is possible to produce the RDD results at other points. To rule this out, I conducted placebo tests for educational attainment moving the discontinuity point to 12, 24, 36, 48 and 60 months on both sides of the April 1, 1933 cut-off point. I also investigated the sensitivity of the results using alternative bandwidths, categorical PGSs terciles and exclusion of inverse probability weights (IPW). All analyses were performed in STATA SE v18. Due to participant confidentiality and data protection regulations, the raw data cannot be shared directly by the author. Researchers interested in accessing the data must apply through the UKDS following the standard procedures. The analysis code necessary to reproduce the results are available from the author upon reasonable request.

3.4 Results

This study included respondents of European ancestries in the ELSA who had information about month and year of birth, full-time education leaving age, APOE4 status, polygenic scores for general cognition and Alzheimer's disease, at least one of memory (from Wave 1 to 9) and executive function (from Wave1 to 5) scores and excluded respondents with cognitive pathology including dementia or Alzheimer's disease diagnosis, given that cognitive pathology could confound the associations under investigation.

Table 3.4 compares the characteristics of the genotyped analytical sample with those of ELSA participants who were excluded because they lacked genetic data. The purpose of this comparison is twofold. First, it serves as a representativeness check: systematic differences between included and excluded participants indicate that the genotyped subsample is not a random draw from the broader ELSA population, which has implications for the external validity of the findings. Second, it provides the empirical basis for applying inverse probability weighting (IPW): where meaningful differences exist—such as in age, educational attainment, or cognitive function—these reflect mortality selection and differential participation, which IPW is designed to partially mitigate (Domingue et al., 2017). As Table 3.4 confirms, the genotyped sample is on average older, more educated, and performs better on cognitive assessments than the excluded sample, consistent with survivorship bias in longitudinal genetic studies. This pattern motivates the use of IPW throughout the analyses and warrants caution in generalising findings beyond this selected population.

Table 3.4 Descriptive statistics of variables of the included and non-included samples

	Included (N=44832)		Non-included (N=38406)		P Value (T-test/ Chi-squared test)
(Mean, SD)					
Age	62	8.8	59	10.2	<0.001
Height (Ware et al.)	161.9	29.27	158.0	36.15	<0.001
Memory	10.3	3.62	10.2	3.85	0.003
Memory (standardised)	0.01	0.97	-0.01	1.03	0.003
Executive Function	0.00	1.44	-0.07	1.56	<0.001
PGS for GC	1350000	6280.3			
PGS for GC (standardised)	0	1.0			
PGS for AD	5597	6.9			
PGS for AD (standardised)	0	1.0			
(Percentage)					
Apoe4 Present		25%			
Sex, Female		55%	58%		<0.001
Ethnicity, White		100%	93%		<0.001
Left School at or over 19		14%	12%		
Left School between 16-18		34%	29%		
					<0.001
Left School at 15		36%	40%		
Left School at or before 14		17%	18%		

The optimal bandwidth was selected using the cross-validation procedure proposed by Imbens and Lemieux (2008), which balances the bias–variance trade-off by minimizing the mean squared prediction error. To address potential seasonality in month of birth—where systematic fluctuations in birth rates or early-life conditions associated with specific months might introduce spurious trends near the threshold (Buckles & Hungerman, 2013)—I restricted the cross-validation search to bandwidths that are exact multiples of 12 months (e.g., 12, 24, 36... 120 months). This ensures that each temporal window contains a balanced representation of all birth months, effectively neutralizing seasonal noise.

Furthermore, birth cohort effects (secular trends in health and cognition) are rigorously controlled for through two mechanisms in this chapter: 1) the local linear functional form of the running variable (distance in months from the April 1933 cut-off), which captures smooth variations across cohorts, and 2) the inclusion of month-of-birth fixed effects in the covariate vector X_{icw} to absorb idiosyncratic monthly variations.

As Imbens and Lemieux (Imbens & Lemieux, 2008) suggested using the same bandwidth for both outcome and treatment equation and using the smallest bandwidth selected by the cross-validation procedures, the optimal bandwidth in this analysis is equal to 48 months for males and females (Supplements Table S3.1). In the following tables of empirical results, I investigate the results of bandwidths at 48 months around the discontinuity.

Table 3.5 shows the effect of the compulsory school reform on school-leaving age, conditional on three samples of educational attainment within the optimal bandwidth of 48 months. The first set of rows shows the estimated effect on school leaving age for the full sample. The estimated results of the reform are significant only for females in the

full sample and subsample of those who left full-time education before 19. The third set of rows shows the results for the most “treated” sample, which includes only individuals who left full-time education before the age of 16. The effect of the reform on education is large and statistically significant for both males and females for each model. The relevance of the instruments is tested in the first-stage regression. The F-test results reported in the last rows of each set show weak instrument problems in the full sample, the subsample of <19 years, and Model 4 in the subsample of <16 years, which includes all three PGSs and their interaction terms. Therefore, in the next step to explore the 2SLS effect of an additional year of schooling at age 15 on cognitive function by PGSs, I only report the results for the subsample of <16 years.

Table 3.5 First stage results- effect of the reform on years of education

Memory	Males				Females			
	Model	Model	Model	Model	Model	Model	Model	Model
	1	2	3	4	1	2	3	4
Full sample								
1947 Reform	0.404*	0.318*	0.356*	0.389*	0.323*	0.312*	0.332*	0.306*
	**	**	**	**	**	**	**	**
N	4,026	4,026	4,026	4,026	4,922	4,922	4,922	4,922
F-test	16.37	1.362	12.92	0.297	13.31	7.342	13.57	3.623
Age<19								
1947 Reform	0.381*	0.305*	0.355*	0.361*	0.331*	0.333*	0.361*	0.316*
	**	**	**	**	**	**	**	**
N	3604	3604	3604	3604	4449	4449	4449	4449
F-test	13.65	0.81	9.258	0.208	13.71	6.424	14.61	3.049
Age<16								
1947 Reform	0.738*	0.669*	0.686*	0.708*	0.511*	0.556*	0.570*	0.539*
	**	**	**	**	**	**	**	**
N	2,542	2,542	2,542	2,542	2,969	2,969	2,969	2,969
F-test	43.92	22.88	47.05	4.157	22.51	21.45	33.7	9.102
Executive								
Function	Males				Females			
	Model	Model	Model	Model	Model	Model	Model	Model
	1	2	3	4	1	2	3	4
Full sample								
1947 Reform	0.418*	0.325*	0.361*	0.389*	0.359*	0.337*	0.351*	0.343*
	**	**	**	**	**	**	**	**
N	2,641	2,641	2,641	2,641	3,098	3,098	3,098	3,098

F-test	16.6	0.686	13.4	0.0476	14.4	6.77	14.86	3.087
Age<19								
1947 Reform	0.402*	0.307*	0.362*	0.367*	0.364*	0.358*	0.378*	0.349*
	**	**	**	**	**	**	**	**
N	2374	2374	2374	2374	2817	2817	2817	2817
F-test	14.26	0.354	7.466	0.0335	15.69	5.754	15.63	2.578
Age<16								
1947 Reform	0.735*	0.684*	0.691*	0.718*	0.562*	0.585*	0.596*	0.595*
	**	**	**	**	**	**	**	**
N	1,712	1,712	1,712	1,712	1,913	1,913	1,913	1,913
F-test	47.88	13.72	48.41	2.467	26.18	17.65	34.86	7.301

Notes: Significance level- * 10% level, ** 5% level *** 1% level. This table presents first-stage estimates of the effect of being born after the 1931 cut-off point on an individual's probability of leaving compulsory school after the age of 14 within the optimal bandwidth of 48 months. The dependent variable is an indicator variable that takes on the value 1 if an individual left compulsory school at or after the age of 15. Model 1 Includes APOE4 status as the genetic index; Model 2 Includes PGSs for cognitive function as the genetic index; Model 3 includes PGSs for Alzheimer's disease as the genetic index; Model 4 includes all four of these genetic indexes. All models use inverse probability weights. Heteroscedasticity standard errors are reported in parenthesis and clustered at the individual level.

First, I estimate the effect of educational attainment on memory and executive function, controlling for polygenic scores but excluding gene–environment interactions, to allow comparison with earlier studies using longer follow-up data. Figure 3.2 shows that an additional year of schooling at age 15 generally leads to an increase of 0.43-0.45 in the memory Z-score among males, which is marginally significant and consistent with previous findings on the direction and magnitude of the effect of educational attainment on memory (Glymour, 2008; Mazzonna, 2014; Banks & Mazzonna, 2012). For females, an additional year of schooling had a negative effect (around -0.75) on executive function, which is in line with Schneeweis's findings of education's negative effect on fluency of executive function (Schneeweis et al., 2014) but inconsistent with Banks et al.'s findings of no significant effect of education on executive function for females

(Banks & Mazzonna, 2012). We find no evidence of an effect of additional years of education for either executive function for males or on memory for females in our genotyped sample, which is inconsistent with the beneficial effect in Banks et al's study.

Moreover, I check the effect of years of education on cognition using the full sample as Banks et al did. My estimates of the effect of the reform on cognitive outcomes (Table S3.2) are almost identical to Banks et al.'s findings using the ELSA wave 1-3, indicating that the inconsistency might be a function of the genotyped sample as well as different follow-up period from the ELSA wave 1 to wave 9. The genotyped subsample is substantially smaller than the full population, and the 2SLS estimator-particularly when interacted with genetic variables-requires larger samples to achieve adequate precision. Null results in the genotyped sample may therefore reflect underpowering rather than a genuine absence of effect. Therefore, the lack of a statistically significant main effect in this subsample may reflect underpowering rather than the absence of a true causal link. Another concern is selection bias: as Table 3.4 demonstrates, genotyped participants are older, more educated, and healthier than non-genotyped ELSA members. This means the estimated effects pertain to a survivor and participant-selected population and may not generalise to the broader birth cohort or to individuals with lower socioeconomic status or poorer health. IPW partially addresses this but cannot fully restore representativeness if selection is driven by unobserved factors. Consequently, the null main-effect results should be interpreted with caution.

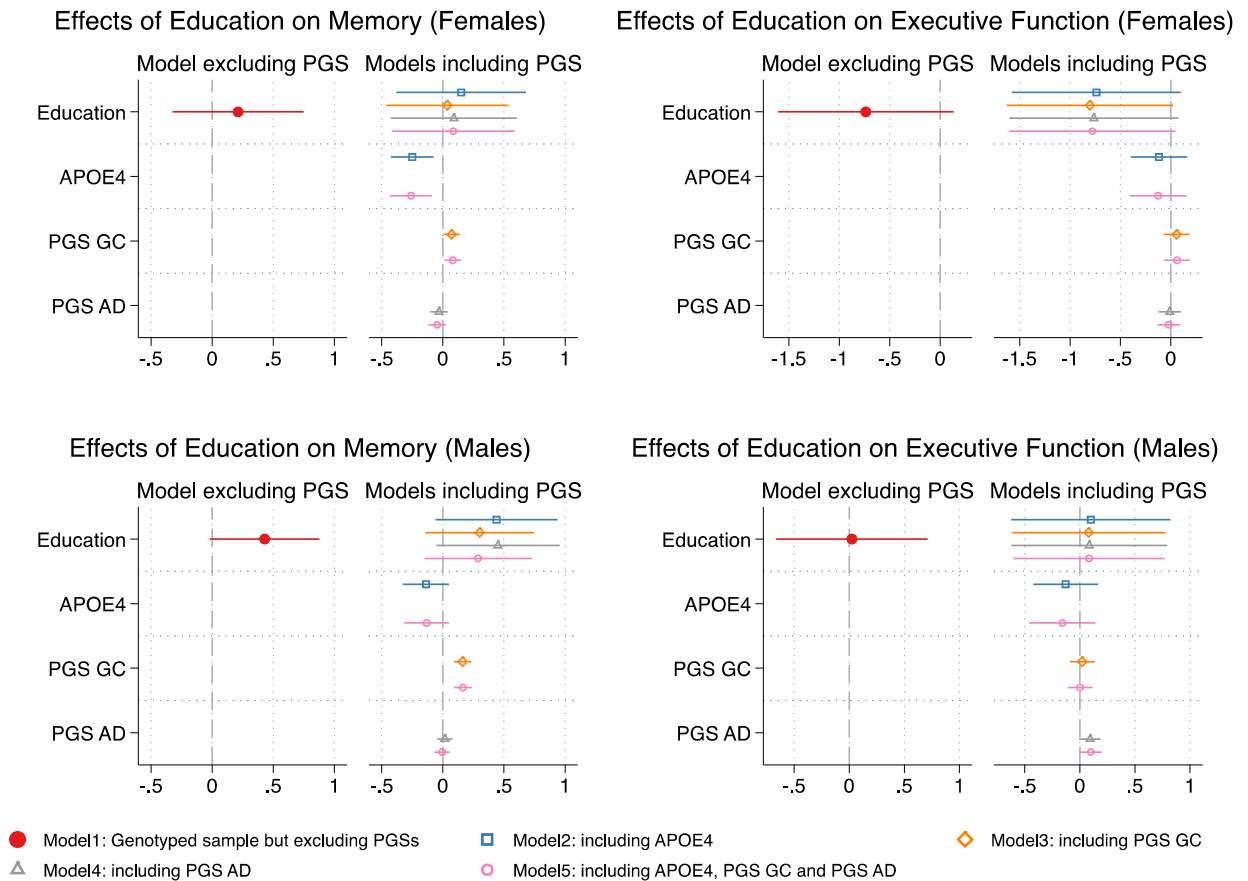


Figure 3.2 Two-stage least squares effect of an additional year of schooling at age 15 on Cognition controlling for polygenic scores but excluding gene–environment interactions

Notes: Significance level- 5% level. This plot presents second-stage estimates of the effect of an additional year of education on an individual’s cognitive function within the optimal bandwidth of 48 months. Model 1 Includes APOE4 status as the genetic index; Model 2 Includes PGSs for cognitive function as the genetic index; Model 3 includes PGSs for Alzheimer’s disease as the genetic index; Model 4 includes all four of these genetic indexes. All models use inverse probability weights. Heteroscedasticity standard errors are reported in parenthesis and clustered at the individual level.

Next, I present the effects of educational attainment interacting with PGSs, on memory and executive function in Figure 3.3. The detailed regression outputs corresponding to these estimates are provided in the Supplementary Materials (Table S3.7 and Table

S3.8). The results in the top right panel show that an additional year of schooling at age 15 did not change executive function for females overall but interacted with APOE4 status. When including APOE4 status in the individual APOE4 model and the three-PGSs model, the effect of an additional year of schooling on executive function was moderated by APOE4 status. There is a decrease of around 2.5 SD in the executive function score with an additional year of schooling at age 15 among APOE4 carriers compared to less educated APOE4 carriers for females (Table S3.8). The individual PGS GC model and individual PGS AD model showed negative borderline associations between an additional year of schooling at age 15 and executive function. There is no evidence of an interaction between PGSs and education on cognition for males in our genotyped sample. As shown in Figure S3.1, the placebo test results of the discontinuity point to 12, 24, 36, 48 and 60 months on both sides of the cut-off point are generally not significant compared to the April 1, 1933 cut-off point, suggesting that the April 1, 1933 cut-off point is robust. Moreover, the sensitivity analyses (Supplements Table S3.3-5) using alternative bandwidths, categorical PGSs terciles and exclusion of inverse probability weights (IPW) affirm the significant and sustained interaction effect of education and APOE4 status on executive function for females.

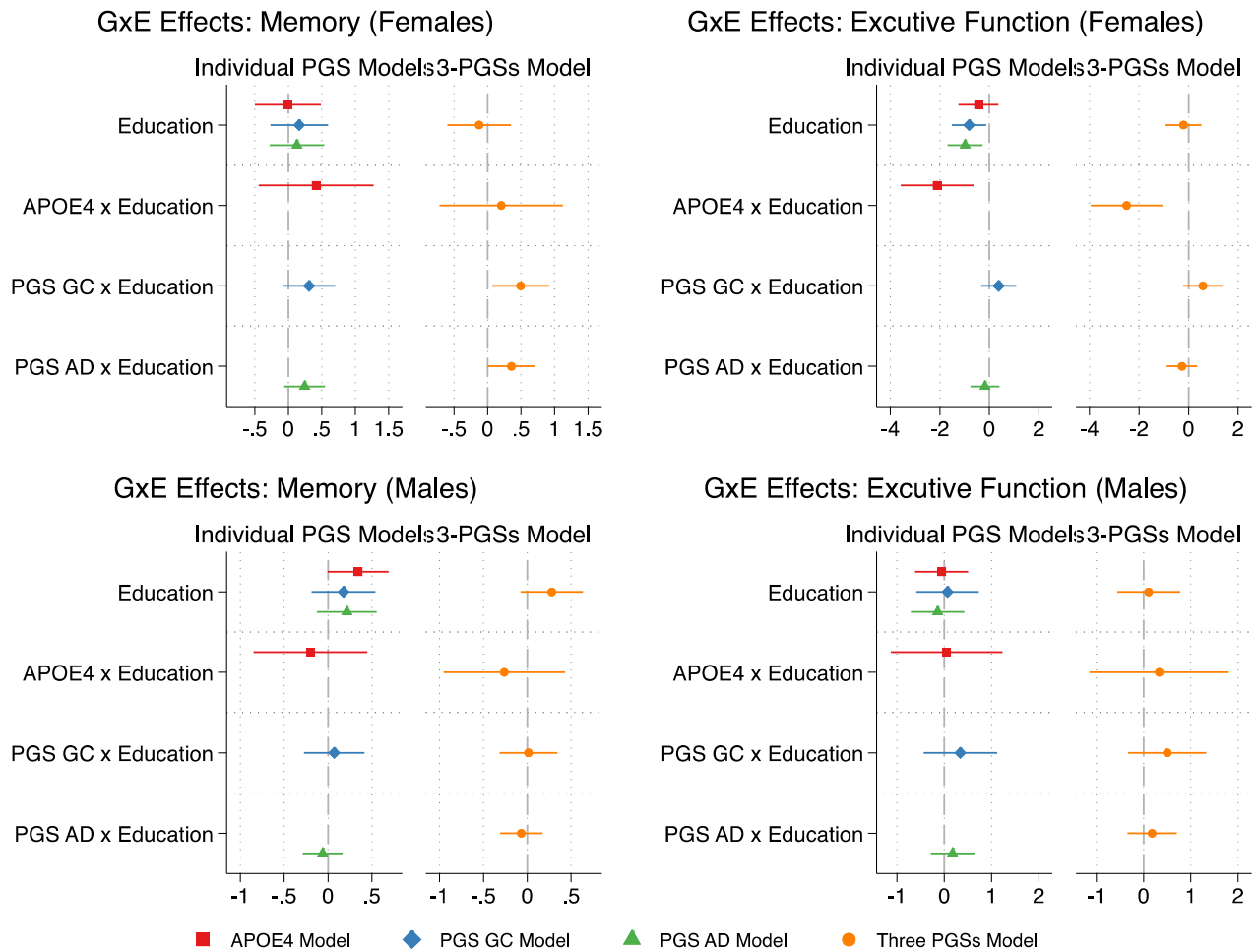


Figure 3.3 Two-stage least squares effect of an additional year of schooling at age 15 on memory and executive function including interactions with individual PGS vs three PGSs

Notes: Significance level- 5% level. This plot presents second-stage estimates of the effect of an additional year of education on an individual's cognitive function within the optimal bandwidth of 48 months. Model 1 Includes APOE4 status as the genetic index; Model 2 Includes PGSs for cognitive function as the genetic index; Model 3 includes PGSs for Alzheimer's disease as the genetic index; Model 4 includes all four of these genetic indexes. All models use inverse probability weights. Heteroscedasticity standard errors are reported in parenthesis and clustered at the individual level.

Because standard instrumental variable (IV) models such as *ivreg2* do not provide a straightforward way to compute model fit metrics like R^2 , it is not possible to directly estimate ΔR^2 or Cohen's f^2 for gene-environment interaction terms using the IV framework. To address this limitation, I re-estimated the same models using ordinary least squares (OLS) to obtain approximate measures of effect size (Supplements Table S3.6). While OLS does not account for endogeneity, it enables the calculation of ΔR^2 (change in explained variance) and Cohen's f^2 (a measure of effect magnitude) (2013), which provide useful benchmarks for interpreting the strength of $G \times E$ effects. Across models for both memory and executive function, the interaction terms added minimal explanatory power. ΔR^2 values ranged from less than 0.001 to 0.006, and the corresponding Cohen's f^2 values were between 0.000 and 0.006—well below the conventional threshold for a “small” effect ($f^2 = 0.02$). These very small effect sizes are consistent with prior findings that $G \times E$ interactions typically explain only modest variance in behavioral outcomes (Dudbridge, 2013; Keller, 2014).

Given the limited variance explained by the gene-environment interaction terms (as reflected in the small ΔR^2 and Cohen's f^2 values), I complement the regression estimates with margins plots in Figure 3.4. The corresponding numerical results are reported in Supplementary Table S3.9. The visualizations illustrate how the effect of an additional year of schooling varies by genetic profile, offering a more intuitive grasp of interaction patterns that may not be apparent from the coefficients alone. In particular, the modifying effect of APOE4 status on the relationship between education and executive function among females is clearly depicted, highlighting how education may differentially influence cognitive outcomes depending on genetic vulnerability. The X-axis is the compulsory-school-leaving age, and the Y-axis is the executive function score. The solid line represents the APOE4 carrier, and the dash line represents the

APOE4 non-carrier. For non-APOE4 carriers, an additional year of schooling at age 15 had no significant effect on executive function among females.

The composite executive function standardized score is constructed by summing the two standardised scores of verbal fluency and letter cancellation (EF full-sample SD = 1.44), such that the unit of measurement reflects combined z-score units rather than a single standardised scale. The 2SLS coefficient for the APOE4 $\times$ education interaction term is -2.505 (Table S3.8), indicating that among female APOE4 carriers, an additional year of compulsory schooling at age 15 is associated with a reduction of approximately 2.5 units on the composite executive function score, equivalent to approximately 1.7 standard deviations of the full-sample distribution. Among APOE4 non-carriers, predicted executive function scores are close to zero regardless of schooling exposure (margin = -0.004 for those leaving school at or before age 14; margin = -0.441 for those leaving after age 15), and the difference between the two groups is modest (-0.44 units, or approximately 0.3 SD). Among APOE4 carriers, however, the pattern diverges sharply: those who left school at or before age 14 exhibit a higher predicted executive function score (margin = $+0.989$, $p = 0.006$), while those exposed to an additional year of compulsory schooling score substantially lower (margin = -1.560 , $p < 0.001$). The within-carrier difference between these two schooling groups amounts to approximately -2.5 units (-1.7 SD) on the composite executive function score, consistent with the 2SLS interaction coefficient.

This produces a clear crossover interaction: at low levels of educational attainment, APOE4 carriers actually outperform non-carriers on executive function by approximately 1.0 unit (0.7 SD), but this advantage reverses sharply following exposure to an additional year of compulsory schooling, leaving carriers approximately 1.1 units

(0.8 SD) below non-carriers. This crossover pattern is consistent with the differential susceptibility framework (Belsky & Pluess, 2009), under which the APOE4 allele acts as a plasticity allele conferring heightened sensitivity to both positive and negative environmental conditions, rather than functioning as a simple risk allele. The robustness of this interaction is confirmed across multiple sensitivity analyses using alternative bandwidths, categorical PGS terciles, and models excluding inverse probability weights (Supplementary Tables S3.3–S3.5).

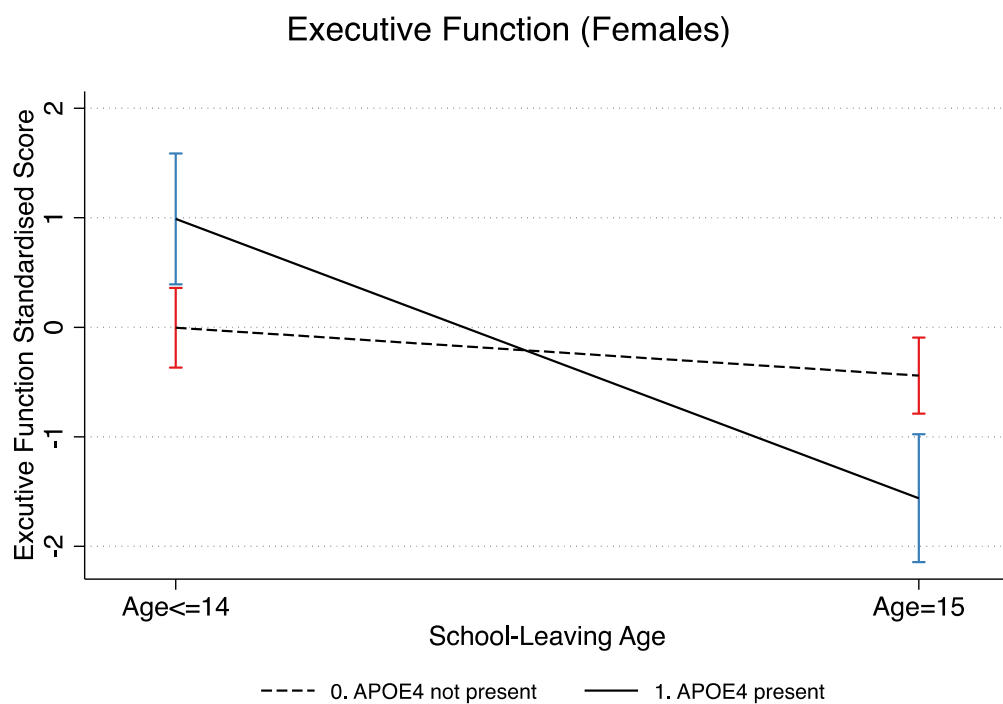


Figure 3.4 Effect of an additional year of schooling at age 15 on executive function by APOE4 status for females with 95% CIs

3.5 Discussion

This study investigated the impact of education on cognitive function and how this interacts with genetic risk, using a nationally representative sample of older people in England. Endogenous measures of environment and low statistical power are believed to be the main challenges for the $G \times E$ analyses (Dawson, 2014). This study improved previous $G \times E$ studies by combining a quasi-experimental design with genetic susceptibility using a large, nationally representative cohort of individuals from the same ancestry in the ELSA. I used a fuzzy regression discontinuity design and exploited the reform of compulsory schooling in England in 1947 to identify the causal effect of the policy-induced additional year of education change on later life cognition and how this interacted with a person's genetic risks.

The results provide cautious evidence that education exerts heterogeneous effects on cognitive domains and interacts with genetic risk. A positive and significant causal effect of education on the old age memory was observed among less-educated males, consistent with prior findings (Glymour, 2008; Mazzonna, 2012; Banks & Mazzonna, 2012). In contrast, for less-educated females, an additional year of compulsory schooling was associated with lower executive function, measured by verbal fluency and letter cancellation. This unexpected finding aligns with Schneeweis et al. (2014), though it contrasts with Banks et al (2012). The discrepancy may be due to differences in the ELSA sample used—specifically, my analysis focused on genotyped participants, which may reflect a healthier or more advantaged population. Further, when including PGS-education interactions, a significant effect emerged between APOE4 status and educational attainment, but only for females in one cognitive domain: female APOE4 carriers had higher levels of executive function if leaving compulsory school before 15

and lower levels of executive function with an additional year of compulsory schooling compared to APOE4 non-carriers.

The substantive meaning of the crossover interaction warrants careful unpacking. The 2SLS coefficient for the APOE4 $\times$ education interaction term is -2.505 , equivalent to approximately 1.7 standard deviations of the full-sample executive function distribution. Among non-carriers, predicted executive function scores are close to zero regardless of schooling exposure, and the difference between the two educational groups is modest (-0.44 units, approximately 0.3 SD) — indicating that the 1947 reform had no meaningful impact on executive function for the genetically typical woman in this cohort. Among APOE4 carriers, the pattern diverges sharply: those who left school at or before age 14 outperform non-carriers by approximately 1.0 unit (0.7 SD), while those exposed to an additional year of compulsory schooling score approximately 1.1 units (0.8 SD) below non-carriers.

The interaction effect resonates with the differential susceptibility (DS) model, which proposes that plasticity varies across individuals: some are more genetically susceptible to both positive and negative environmental exposures, while others remain resilient across all situations (Belsky & Pluess, 2009; Cheesman et al., 2022). My findings are consistent with those of Seeman et al. (2005). Using a sample of 865 participants aged 70–79 in the MacArthur Successful Aging Study, they found that APOE4 carriers with more than 8 years of education experienced greater cognitive decline, whereas for those with fewer than 8 years of education, time trends in cognitive performance were not affected by APOE4. The findings of this study also suggested that education may accelerate the detrimental effects of the APOE4 allele on cognition. On the contrary, some studies found that higher educational attainment buffers the negative effect of

APOE4 on cognition (Wang, 2012), and still other studies showed no interaction between education and APOE4 on cognitive decline (Van Gerven, 2012). While the findings in previous G×E research are partially contradictory, the cumulative evidence suggests that education may not only shape later-life cognition directly but could also interact with genetic vulnerability in complex ways.

The restriction of this interaction to female APOE4 carriers and to executive function specifically — rather than to males or to episodic memory — points to a unified neurobiological explanation. Women who carry the APOE4 allele face a disproportionately elevated lifetime risk of Alzheimer's disease relative to male carriers, a disparity rooted in sex-differentiated biological mechanisms including oestrogen-APOE4 interactions and differential immune response to amyloid pathology (Altmann et al., 2014). Critically, APOE4-related neurodegeneration in women is characterised by earlier and more pronounced cortical thinning in frontal and parietal regions — the neural substrate of executive function — relative to the hippocampal-entorhinal networks that mediate episodic memory (Fleisher et al., 2005; De Frias et al., 2006). This means that, at the ages observed in ELSA, the frontal-executive system is precisely the domain under greatest APOE4-related biological pressure in women, and therefore the domain most susceptible to gene-environment modulation. The absence of a parallel interaction for episodic memory or for males is thus not null evidence — it reflects the domain- and sex-specific timing of APOE4's neurobiological effects and constitutes indirect validation of the proposed mechanism.

The direction of the interaction — higher executive function among less-educated female APOE4 carriers, lower among those exposed to additional schooling — requires explanation at two intersecting levels. At the neurobiological level, the antagonistic

pleiotropy hypothesis proposes that APOE4 confers cognitive advantages at earlier life stages that subsequently reverse. Evidence from the 20-year Whitehall II cohort study of 5,561 participants found that APOE4 heterozygous carriers showed better performance on reasoning and phonemic fluency — precisely the executive function domains captured in this thesis — up to approximately midlife, with this advantage reversing and accelerating into decline thereafter in late life (Gharbi-Meliani et al., 2021). For less-educated female APOE4 carriers, this early-life executive function advantage may have been preserved and amplified through the cognitively demanding domestic and community roles they entered immediately upon leaving school — including complex household management, childcare coordination, and dense neighbourhood networks — which imposed sustained demands on planning, cognitive flexibility, and verbal processing. Because APOE4 carriers are characterised by heightened environmental plasticity, this form of ecologically embedded stimulation may have produced larger executive function benefits among carriers than among non-carriers. At the social-structural level, the critical question for those who received an additional year of compulsory schooling is what that year translated into in terms of subsequent life-course experience. Compulsory schooling reforms improve cognitive outcomes primarily through the pathway of occupational complexity: it is sustained occupational engagement — not formal schooling per se — that consolidates cognitive reserve over the life course (Ellwood, 2023). This pathway, however, operated far more weakly for women in post-war England: additional schooling did not substantially alter women's occupational trajectories, with most continuing into domestic roles or low-complexity clerical employment regardless of whether they left school at 14 or 15 (Banks & Mazzonna, 2012). Consistent with this, evidence from a comparable French compulsory schooling reform found that women eligible for the reform showed no cognitive gains in

older age and instead exhibited higher levels of depressive symptoms in midlife, suggesting that mandatory additional schooling in contexts where it does not translate into occupational advancement may carry costs rather than benefits for women (Fichera & Gathergood, 2019). In the absence of the occupational complexity pathway, the additional schooling year generated no lasting cognitive reserve — while simultaneously delaying entry into the domestic roles that had provided ecologically embedded executive stimulation for less-educated carriers. For female APOE4 carriers, whose heightened environmental sensitivity amplifies both the benefits of cognitively stimulating environments and the costs of their absence, this failure of the educational pathway to generate downstream cognitive engagement may have been particularly consequential.

These mechanisms are not mutually exclusive, and the present data cannot adjudicate definitively between them. The interaction should therefore be interpreted as preliminary evidence for sex- and domain-specific moderation of educational effects by APOE 4 genotype. While the evidence base for sex-specific APOE 4 $\times$ education interactions remains limited, the convergence of neurobiological, epidemiological, and social-historical explanations offered above suggests that this finding is theoretically plausible. Future research with larger, sex-stratified samples and richer measures of life-course occupational and social engagement will be needed to disentangle these mechanisms before strong causal claims can be made.

There are several limitations in this study. First, the sample size in this study is similar to previous studies (Plomin et al., 2022) which is relatively small, resulting in low statistical power for detecting interaction effects. A larger sample size and replication of the current findings are needed to elucidate how education interacts with genetic

susceptibility for cognitive function. Another limitation of this study is the sample selection. On one hand, the sample in the ELSA study is highly selected, and is from more affluent socioeconomic groups than the general population, and (intentionally) for individuals older than 50 years (the minimum age at recruitment). On the other hand, the genotyping work was conducted in 2013-2014. Due to mortality selection in longitudinal social surveys, the sample of respondents who were genotyped in the genetic survey was healthier, better educated, and longer-lived compared to the main ELSA sample. Although I applied inverse probability weighting, these factors still carry a risk of introducing biases. On the genetic side, ELSA hardly has any non-European respondents, so I only included European ancestries. Therefore, the findings using genetic data constructed from European ancestry GWAS cannot be generalized to other ancestral populations (Gharbi-Meliani et al., 2021). In addition, the PGSs used in this analysis were constructed from GWAS that estimate the association between genetic variants and the conditional mean of a phenotype, which may not separate genetic effects on phenotypic variability from effects on the mean/level of trait values. Therefore, this study cannot detect interaction effects that are driven by loci that affect plasticity. Evaluating both mean effects (mPGS) and variance effects (vPGS) in further studies is warranted (Johnson et al., 2021). On the environmental side, since race can shape individual educational attainment outcomes (Albin, 1993; Schultz et al., 2008; Taylor et al., 2017), limiting our analysis to European ancestries restricts the scope of potential education effects that can be observed. In addition, while the regression discontinuity design reduces endogeneity, education itself is influenced by genetic and non-genetic factors. Conditioning on an environmental variable such as education can introduce collider bias (Akimova et al., 2022), while attrition and non-response may further distort associations (Fry, 2017). Moreover, SNP effects may differ across

education levels, meaning that average $G \times E$ estimates risk masking meaningful heterogeneity (Zhang, 2024; Tropf, 2017). Lastly, although I detect a statistical gene-by-education interaction, I do not have the means to characterise the biological mechanisms underlying the interaction. Future studies are needed to examine the mechanisms of these interactions.

In conclusion, this study contributes to the literature on social determinants of cognitive ageing by combining quasi-experimental variation in education with genetic susceptibility. The results suggest that education influences later-life cognition and may interact with genetic risk in ways that vary by sex and cognitive domain. However, given the limitations of sample size, representativeness, and polygenic measurement, these findings should be interpreted as preliminary. Future research should seek replication in larger and more diverse cohorts, employ more refined genetic measures, and explore the mechanisms through which education and genetics jointly shape cognitive trajectories in older age.

3.6 Supplements

Section S3.1 Optimal bandwidth in the cross-validation procedure

The optimal bandwidth is selected with a “leave one out” procedure proposed by Imbens and Lemieux (2008). For each observation on the left side of the cut-off point, I run a linear regression using only observations with the value of X (the treatment determining variable) on the left of X_i ($X_i - h \leq X < X_i$), while for each observation on the right side of the cut-off point, I use only those on the right side of X_i ($X_i \leq X < X_i + h$). I repeat this procedure for each i in order to obtain the whole set of predicted value of Y which can be compared with the actual value of Y . Formally, the cross-validation criterion is defined as

$$F_Y(h) = \frac{1}{N} \sum_{i=1}^{N_h} \{Y_i - \hat{Y}[X_i]\}^2,$$

The optimal bandwidth is that value of h which minimises the criterion function.

Table S3.1 Optimal bandwidth in the cross-validation procedure

	Educational Attainment	Memory	Executive Function
Males	49	33	36
Females	52	49	45

Table S3.2 Effect of education on cognitive function, replication of Bank & Mazzonna,

2012

	Males			Females		
	Bandwidth=12	Bandwidth	Bandwidth	Bandwidth	Bandwidth	Bandwidth
	months	=24 months	=36 months	=12 months	=24 months	=36 months
Age<16						
Memory	0.246	0.481**	0.536***	0.947	0.530	0.349*
	(0.479)	(0.225)	(0.198)	(0.659)	(0.264)	(0.145)
Exec. Func.	0.093	0.371*	0.396**	-5.802	-0.454	-0.128
	(0.636)	(0.275)	(0.209)	(7.334)	(0.705)	(0.383)
N	457	933	1,405	504	985	1,466
Age<19						
Memory	-0.137	0.361	0.408	1.176	0.121	0.327
	(0.509)	(0.250)	(0.2321)	(3.038)	(0.480)	(0.260)
Exec. Func.	-0.627	0.111	-0.040	-5.575	-0.510	0.215
	(0.708)	(0.400)	(0.349)	(5.737)	(0.752)	(0.409)
N	566	1,125	1,679	653	1,285	1,923

Notes. *p = 10%; **p = 5%; ***p = 1%; Standard errors in parenthesis are robust to heteroscedasticity and clustered at the individual level.

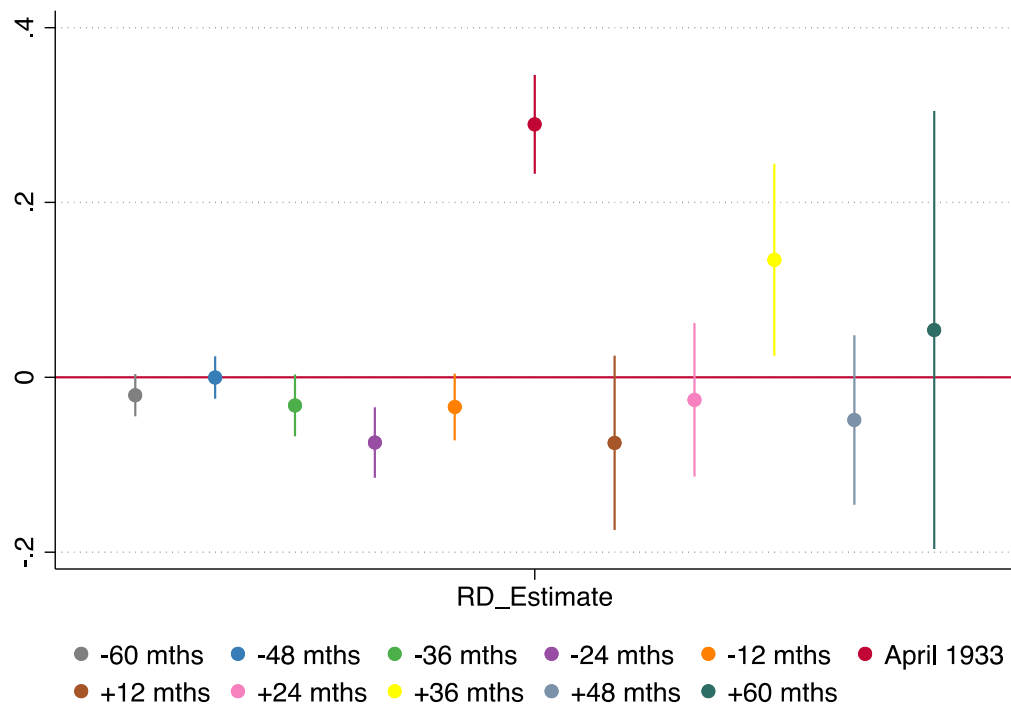


Figure S3.1 Placebo Tests

Table S3.3 2SLS estimates of the effect of an additional year of schooling at age 15 on
cognitive function by APOE4 status and PGSs, Bandwidths=36/50

Males					Females			
Bandwidth=36								
Memory	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1947 Reform	0.694***	0.25	0.524**	0.632***	0.171	0.197	0.061	-0.477
APEO4 x 1947 Reform	-0.23			-0.724*	1.275***			1.600***
PGS GC x 1947 Reform		0.139		-0.071		0.650**		0.944**
PGS AD x 1947 Reform			-0.098	-0.067			0.205	0.726***
Executive	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Function	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1947 Reform	0.205	0.414	0.595	1.138**	-0.8	-1.214**	-1.360**	-0.864
APEO4 x 1947 Reform	0.592			-0.002	-1.959**			-1.734*
PGS GC x 1947 Reform		0.635		0.762		-0.097		-0.584
PGS AD x 1947 Reform			0.08	0.21			-0.476	-0.189
Bandwidth=60								
Memory	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1947 Reform	0.138	-0.049	-0.012	0.057	-0.018	0.107	0.136	0.023
APEO4 x 1947 Reform	-0.013			-0.142	0.234			0.052
PGS GC x 1947 Reform		-0.028		-0.108		0.271		0.294
PGS AD x 1947 Reform			-0.088	-0.06			0.211	0.315**
Executive	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Function	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1947 Reform	0.138	-0.049	-0.012	0.057	-0.569	-0.623*	-0.717**	-0.577*
APEO4 x 1947 Reform	-0.013			-0.142	-0.665			-0.924*
PGS GC x 1947 Reform		-0.028		-0.108		0.649**		0.828**
PGS AD x 1947 Reform			-0.088	-0.06			-0.194	-0.031

Notes. *p = 10%; **p = 5%; ***p = 1%

Table S3.4 2SLS estimates of the effect of an additional year of schooling at age 15 on
cognitive function by APOE4 status, PGS terciles

	Males			Females		
Memory	Model 2	Model 3	Model 4	Model 2	Model 3	Model 4
1947 Reform	-0.235	0.127	0.293	0.514	0.361	-0.152
APEO4 x 1947 Reform			-0.542			0.040
2nd PGS GC x 1947 Reform	0.123		0.308	-0.036		-0.118
3rd PGS GC x 1947 Reform	-0.266		-0.225	0.414		1.108*
2nd PGS AD x 1947 Reform		-0.364	-0.307		0.320	0.275
3rd PGS AD x 1947 Reform		-0.273	-0.539		0.452	0.547
Executive Function	Model 2	Model 3	Model 4	Model 2	Model 3	Model 4
1947 Reform	-0.75	-0.403	1.986	-1.130**	-1.023*	-1.585**
APEO4 x 1947 Reform			-3.513			-3.201***
2nd PGS GC x 1947 Reform	-0.306		-0.67	0.553		0.321
3rd PGS GC x 1947 Reform	-0.992		-2.675	0.219		0.888
2nd PGS AD x 1947 Reform		-0.634	-0.843		0.151	2.109*
3rd PGS AD x 1947 Reform		-0.926	-3.523		0.295	1.104

Notes. *p = 10%; **p = 5%; ***p = 1%

Table S3.5 2SLS estimates of the effect of an additional year of schooling at age 15 on
cognitive function by APOE4 status, UNWEIGHTED

	Males				Females			
Memory	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1947 Reform	-0.097	-0.355	-0.128	-0.476	0.495	0.482	0.401	0.43
APEO4 x 1947 Reform	-0.056			-0.873	0.226			-0.034
PGS GC x 1947 Reform		-0.255		-0.825		0.272		0.348
PGS AD x 1947 Reform			-0.114	-0.129			0.280*	0.368**
Executive Function	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1947 Reform	-0.785	-1.873	-0.988*	-1.635	-0.222	-0.753	-0.828*	-0.386
APEO4 x 1947 Reform	-0.561			-3.62	-2.419***			-2.629***
PGS GC x 1947 Reform		-1.315		-2.699		0.029		0.084
PGS AD x 1947 Reform			-0.044	-0.575			0.17	0.265

Notes. *p = 10%; **p = 5%; ***p = 1%

Table S3.6 Approximate Effect Sizes (ΔR^2 and f^2) for Gene $\times$ Education Interactions on Cognitive Outcomes

	Males				Females			
	R^2	R^2 Full	ΔR^2	Cohen's f^2	R^2	R^2	ΔR^2	Cohen's f^2
	Reduced				Reduced	Full		
Memory								
PGS GC x 1947 Reform	0.2929	0.295	0.0019	0.0027	0.2175	0.219	0.0015	0.0019
PGS AD x 1947 Reform	0.2721	0.282	0.0097	0.0135	0.2248	0.225	0.0001	0.0001
APEO4 x 1947 Reform	0.3068	0.307	0.0001	0.0001	0.3179	0.32	0.002	0.0029
Executive Function								
	R^2	R^2 Full	ΔR^2	Cohen's f^2	R^2	R^2	ΔR^2	Cohen's f^2
	Reduced				Reduced	Full		
PGS GC x 1947 Reform	0.3063	0.309	0.0026	0.0038	0.2909	0.296	0.0049	0.0070
PGS AD x 1947 Reform	0.2981	0.304	0.0054	0.0078	0.2905	0.291	0.0003	0.0004
APEO4 x 1947 Reform	0.2273	0.228	0.0003	0.0004	0.2784	0.287	0.0083	0.0116

Table S3.7 Two-stage least squares effect of an additional year of schooling at age 15 on memory and executive function including interactions with individual PGS vs three PGSs (males)

Males				
Memory	Model 1	Model 2	Model 3	Model 4
1947 Reform	0.344*	0.177	0.215	0.279
APEO4 x 1947 Reform	-0.200			-0.262
PGS GC x 1947 Reform		0.071		0.014
PGS AD x 1947 Reform			-0.061	-0.069
Executive Function	Model 1	Model 2	Model 3	Model 4
1947 Reform	-0.056	0.069	-0.142	0.108
APEO4 x 1947 Reform	0.048			0.331
PGS GC x 1947 Reform		0.341		0.502
PGS AD x 1947 Reform			0.176	0.18

Table S3.8 Two-stage least squares effect of an additional year of schooling at age 15 on memory and executive function including interactions with individual PGS vs three PGSs (females)

Females				
Memory	Model 1	Model 2	Model 3	Model 4
1947 Reform	-0.007	0.164	0.128	-0.126
APEO4 x 1947 Reform	0.416			0.204
PGS GC x 1947 Reform		0.311		0.493**
PGS AD x 1947 Reform			0.244	0.356*
Executive Function	Model 1	Model 2	Model 3	Model 4
1947 Reform	-0.437	-0.815**	-0.976***	-0.205
APEO4 x 1947 Reform	-2.113***			-2.505***
PGS GC x 1947 Reform		0.379		0.581
PGS AD x 1947 Reform			-0.174	-0.263

Table S3.9 Predicted Margins of Executive Function Score by APOE4 Status and
School-Leaving Age (Females)

APOE4 Status	School-Leaving Age	Margin	Std. Err.	z	p	95% CI
Not present	Left school at or before 14	-0.004	0.221	-0.02	0.986	[-0.437, 0.430]
Not present	Left school after 15	-0.441	0.211	-2.09	0.037	[-0.855, -0.027]
APOE4 present	Left school at or before 14	0.989	0.363	2.73	0.006	[0.278, 1.700]
APOE4 present	Left school after 15	-1.560	0.355	-4.40	0.000	[-2.256, -0.865]

Chapter 4

Gene-Environment Interaction during a Pandemic: How Polygenic Scores Shape Mental Health Responses to COVID-19

4.1 Introduction

The COVID-19 pandemic, with its rapid spread and social distancing regulations, has been a unique stressor and has affected the mental health and well-being of people all around the world. (Santomauro et al., 2021; O'Connor, 2021; Fancourt, 2021; Santomauro, 2021). The COVID-19 pandemic led to a 27.6% increase in cases of major depressive disorder (MDD) and a 25.6% increase in cases of anxiety disorders (AD) worldwide in 2020. Overall, the pandemic was estimated to have caused 137.1 additional disability-adjusted life years (DALYs) per 100 000 population for MDD and 116.1 per 100 000 population for AD (Sacco et al., 2023).

Older adults are at higher risk of death and serious illness due to the pandemic (Dowd et al., 2020) and are more vulnerable to the adverse impact of stressors such as social isolation, lack of health care and social support than other age groups (Choi, 2022). Recent longitudinal studies demonstrate that older individuals did not adapt well to the new psychosocial stressors introduced by the pandemic and experienced more symptoms of depression and anxiety, greater loneliness, and poorer psychological well-being, and quality of life compared with pre-pandemic levels (Kämpfen et al., 2020) (Steptoe, 2021; Zaninotto et al., 2022). Inequalities in experiences of mental health disorders persisted during the COVID-19 pandemic among older adults (e.g., sex, ethnicity, socioeconomic status, and physical disabilities), highlighting the need for

targeted psychological interventions to restore the well-being of older people, both in the COVID-19 pandemic and for the future (Duan & Zhu, 2020; Yang et al., 2020).

The COVID-19 pandemic acted as a profound “macro-stressor” that disrupted mental health through a complex interaction of environmental, social, and psychological pathways. Central to this disruption was the experience of perceived social isolation. According to the Interpersonal Theory of Suicide (Joiner, 2005) and the foundational research of Cacioppo and Hawkley (2009), social connection is a fundamental biological need. This is supported by recent meta-analytic evidence from Santini et al. (2020) in *The Lancet Public Health*, which identifies social disconnectedness as a primary driver of depression and anxiety in older populations during periods of physical distancing. The mandated lockdowns did not merely result in a lack of contact; they severed the “belongingness” essential for emotional regulation, converting physical distancing into a significant psychological deficit.

Beyond the immediate loss of social anchors, the pandemic imposed a state of chronic existential threat. As articulated by Holmes et al. (2020) in *The Lancet Psychiatry*, the constant salience of mortality triggered severe health anxiety. This is further detailed by Aknin et al. (2022) in *Nature Human Behaviour*, who demonstrate that the prolonged nature of the crisis—characterised by a lack of perceived control and high uncertainty—significantly depleted individuals' psychological resilience. For older adults, this existential stress was uniquely compounded by Social Role Loss. Continuity Theory suggests that well-being in later life depends on maintaining established social roles; the suspension of activities such as grandparenting or volunteering stripped individuals of their self-efficacy, a mechanism recently highlighted by Vahia et al. (2020) in *JAMA* as a critical factor for the mental health of older adults during the COVID-19 crisis.

Crucially, these environmental pressures were distributed along gendered lines, as explained by Social Role Theory (Eagly & Wood, 2012). The domestic sphere became a primary site of stress, where women often bore a disproportionate “double burden” of increased unpaid labour and emotional labour. This is empirically supported by Etheridge and Spantig (2022), whose longitudinal analysis confirms that women’s mental health declined significantly more than men’s due to the increased domestic and caregiving strain. Combined with women’s documented higher levels of interpersonal sensitivity (Kuehner, 2017), this systemic pressure created a gender-specific vulnerability. As argued by Pierce et al. (2020) in *The Lancet Psychiatry*, the pandemic did not act as a uniform shock but rather amplified existing social inequalities, resulting in deeply stratified psychological outcomes.

Because non-pharmaceutical COVID interventions have not been randomized, multiple studies rely on two-way fixed effects (TWFE) and difference-in-differences (DiD) research design to assess the causal effect of the overall pandemic or mitigation policies on mental health outcomes. Zaninotto et al. (Zaninotto et al., 2022) used a two-way fixed-effects linear regression model to examine changes in mental health and well-being before and during the initial and later phases of the COVID-19 pandemic. They analysed data from 5146 older adults in England who provided data before the COVID-19 pandemic (2018/2019) and on two occasions in 2020 (June/July and November/December) and found that the prevalence of clinically significant depressive symptoms increased accompanied by increased loneliness, increased anxiety and deterioration in the quality of life. They also found that higher socioeconomic groups had better mental health overall but responded to the COVID-19 pandemic with more negative changes.

Another study used difference-in-differences (DiD) estimation to disentangle the effect of lockdown and compared the mental health of 20839 people in Australia in the state of Victoria who were exposed to lockdown at the time of the 2020 interviews (treatment group) with those living elsewhere in Australia (who were living relatively free of restrictions; control group) (Butterworth et al., 2022). The study showed a small but statistically significant effect of lockdown on mental health, with a greater decline for residents of Victoria in 2020 than for those in the rest of Australia. They also found that the mental health effects of lockdowns differ by population subgroups and for some might have exaggerated existing inequalities in mental health.

In general, an adverse mental health effect of the pandemic was found in older adults. Most current findings of the TWFE and DiD research design are specific to one specific country/region context and therefore might not be generalisable to other settings. Nevertheless, there is a commonality in the types of lockdown policy responses implemented around the world, involving stay-at-home orders, school and business closures, cancellation of public events, and restrictions on movement. Therefore, researchers could adapt this approach to new data and settings to provide causal estimates of the COVID-19 mental health effects.

However, important questions remain in our understanding of poorer mental health in older people after the COVID-19 pandemic. Psychiatric disorders are multifactorial in aetiology, influenced by both genetic and environmental factors (Plomin et al., 1994). Behavioural genetics research also suggests accounting for the interaction effect between genetic and environmental factors rather than independent main effects on psychiatric disorders (Assary et al., 2018; Uher, 2014). However, these early hypothesis-driven studies adopted a bottom-up approach which requires a strong biological

hypothesis to select appropriate candidate genes, but knowledge regarding specific biological mechanisms underlying these mental disorders is limited. Therefore, the risk of selecting inappropriate candidate genes is high (Bosker et al., 2011; Collins et al., 2012). In addition, candidate gene studies have low statistical power, unsuccessful replications, and publication bias toward positive findings (Karg et al., 2011; Munafo et al., 2009; Risch et al., 2009; Uher & McGuffin, 2010), which are “wrong or misleading” (Hewitt, 2012).

The non-replication of candidate gene studies is due to the limited understanding of the genetic architecture of complex traits. Psychiatric disorders are complex and highly polygenic traits, which are influenced by multiple genetic variants with small effects throughout the genome rather than a few selected candidate genes of large effects (Donnelly, 2008). Many risk loci have emerged from genome-wide association studies (GWAS) of, among others, major depression, anxiety, schizophrenia, bipolar disorder, and attention-deficit/hyperactivity disorder from the Psychiatric Genomics Consortium (PGC) and other efforts (Sullivan et al., 2018). Furthermore, GWAS studies have begun to employ polygenic scores (PGS), in which the aggregate effects of multiple genetic variants are summarized in a single score, specifically, the number of trait-associated alleles possessed by each individual in the target sample, weighted by their effect sizes from a discovery sample (Dudbridge, 2013). As polygenic scores aggregate the information from many genetic variants, they are more strongly associated with the outcomes than any single genetic variant.

GWAS heritability estimates vary from less than 10% for anxiety disorders to more than 20% for bipolar disorder and schizophrenia), compared with heritability estimates of around 37% for major depressive disorder (MDD) (Sullivan et al., 2000), 23% for

depressive symptoms (Jansson et al., 2004), 32% for generalised anxiety disorders (Hettema et al., 2001) and 40% for loneliness in adults (Distel et al., 2010) (Goossens et al., 2015). GWAS heritability estimates are lower than those derived from twin studies, in part attributed to many common variants of extremely small effect sizes that elude detection, rare variants to which GWAS are insensitive, additionally rare structural variations, etc. Yet a further possibility is that GWAS, which are based on combining data from multiple periods, countries, and birth cohorts, will introduce heterogeneity in genetic effects, as a function of environmental factors to which some but not all individuals are exposed, implying a gene-environment interaction ($G \times E$) (Conley, 2016a; Tropf et al., 2017). Therefore, it should ultimately be possible to account for additional missing heritability in mental disorders by examining how environmental exposures modify genetic risk (gene-environment interaction - $G \times E$).

Using polygenic scores in the $G \times E$ framework, many studies have examined gene-environment interactions and depression. Colodro-Conde et al. (Colodro-Conde et al., 2018) found that the PGS for major depressive disorder (MDD) significantly interacted with stressful life events (SLEs) to increase the risk of depression. Iob and Steptoe (Iob et al., 2021) similarly found that the associations between adverse childhood experiences (ACEs) and depressive symptoms were larger among older adults with a higher PGS for MDD (OR 1.13, 95% CI 1.04–1.23), in line with the diathesis–stress model (Monroe & Simons, 1991; Robins & Block, 1989). In contrast, one study showed that individuals exposed to severe childhood trauma were at the greatest risk if they also had a *low* PGS for depression (Mullins, 2016). Others reported independent but not interactive associations of ACEs and PGSs with depression (Musliner et al., 2015; Peyrot et al., 2018). A few $G \times E$ studies employed polygenic scores to examine other mental disorders. One study derived a polygenic score of environmental sensitivity from the

psychiatric genetics consortium in 1,026 monozygotic twin pairs and found that parenting had little effect on emotional problems for those with a low PGS. However, in those with a high PGS, negative parenting was associated with increased emotional problems, while positive parenting was associated with decreased emotional symptom scores (Keers et al., 2016). This finding implies that individuals with the greatest environmental sensitivity may be more likely to develop mental disorders in adverse environments but also benefit more from favourable environments, supporting the differential-susceptibility model (Belsky & Pluess, 2009). Another study explored both depression and well-being outcomes and reported significant independent effects of the PGS for depression and peer victimisation in predicting the risk of depressive symptoms and well-being in early adulthood, but no clear interaction effects (Armitage et al., 2022).

Another challenge in $G \times E$ research is that the sensitivity genotype is correlated with environmental exposure (gene-environment correlation [r_{GE}]). The r_{GE} issue can be overcome by using natural experiments or other quasi-experimental designs (Fletcher, 2013). Domingue et al. (Domingue, 2017) tested whether higher subjective well-being polygenic scores buffered survivors' risk of developing depressive symptoms following the death of their spouse from 8,588 non-Hispanic white adults in the Health and Retirement Study (HRS). This study found that individuals who carry more well-being-associated alleles tended to develop fewer additional depressive symptoms following the death of a spouse. While spousal death is not exogenous, the genetic buffering was not affected by any gene-environment correlation between the genetics of subjective well-being and exposure to the death of a spouse; HRS respondents' subjective well-being polygenic scores were unrelated to their risk of losing a spouse during follow-up.

Fletcher (Fletcher, 2014) exploited a quasi-experimental variation in the environmental

exposure (whether a respondent was interviewed before or after the 9/11 attacks) using Wave 3 of the National Longitudinal Study of Adolescent Health (Add Health). This study showed that the effects of this external stressor were moderated by two candidate genes. These research designs can be adapted to other causal inferences in gene-environment interaction studies for other specific life stressors, such as the COVID-19 pandemic.

To summarise, polygenic scores can more strongly capture the additive genetic influences of complex traits such as mental disorders; quasi-experimental designs can more effectively isolate exogenous variation in observational data to minimize the impact of rGE . Applying polygenic scores together with quasi-experimental designs can thus be used to improve causal inference for $G \times E$ research. In order to expand knowledge about mental disorders and the gene-environment interaction effects among older adults in the COVID-19 pandemic, this study uses two-way fixed effects model regression to estimate the causal effects of the COVID-19 pandemic on mental health outcomes among older people using the panel data from the English Longitudinal Study of Ageing survey (ELSA), and employed polygenic scores in the ELSA genetic data to examine whether and how the effect of the pandemic was moderated by genetic risks.

4.2 Methods

Data and Participants

This chapter utilizes data from the English Longitudinal Study of Ageing (ELSA), specifically focusing on the transition from the most recent pre-pandemic assessment (Wave 9, 2018/19) to the two dedicated COVID-19 Substudies (collected in June/July and November/December 2020). ELSA is a nationally representative panel study of individuals aged 50 and older; comprehensive details regarding its sampling frame, recruitment history, and refreshment waves are provided in Section 1.6.2.

The analytical sample is restricted to participants who provided valid phenotypic data during the pandemic and were included in the ELSA Genetic database (genetic sample $N = 7,183$). As detailed in Section 1.6.4, these genetic data consist of high-density genotypes of European ancestry, from which Polygenic Scores (PGSs) were derived following the harmonized HRS methodology. The response rate for the COVID-19 substudies remained high at 75% ($n = 7,040$ in Substudy 1), ensuring that the sample remained representative of the older English population during the pandemic.

Measures

Outcomes

In this chapter, I examined five outcome measures of mental health, including depressive symptoms, major depressive disorder, loneliness, anxiety, and quality of life. To maintain structural efficiency and avoid redundancy, the detailed psychometric rationale, longitudinal invariance testing, and clinical thresholds for these instruments are established in Section 1.6.3. A summary of the measures utilized in this chapter is provided in Table 4.1. Notably, while the first four outcomes were measured across both

pre-pandemic and pandemic waves, the GAD-7 (anxiety) was introduced exclusively during the ELSA COVID-19 substudies. Consequently, the anxiety measure is utilized here to capture acute psychological responses during the pandemic period specifically, acknowledging that the absence of a pre-pandemic baseline precludes the same longitudinal fixed-effects comparison applied to the other indicators.

Table 4.1 Measures of mental disorders

Mental Health Outcome	Measure	Questions	Range	Cut-off Point to Identify Cases
Depressive Symptoms	8-item Centre for Epidemiological Studies Depression (CESD-8) Scale	Eight symptoms of depression during the past week: depression, sadness, lack of happiness, loneliness, restless sleep, inability to get going, lack of energy and that everything was an effort. (1 point for each)	0-8, high scores indicating more depressive symptoms	≥ 4 (Steffick, 2000)
Major Depressive Disorder	CESD-8 ≥ 4	--	--	--
Loneliness	3-item revised University of California (UCLA) Loneliness Scale	How often feel isolated from others, feel lack companionship, feel left out. Each item was scored on a 3-point scale from hardly ever/never to often.	1-12, higher scores indicating greater loneliness	≥ 5
Quality of Life	12-item version of the Control, Autonomy, Self-realisation, and Pleasure (CASP) Scale	Seven symptoms of anxiety in the past 2 weeks. Each symptom is rated on a 4-point scale from not at all to nearly every day.	1-48, higher scores indicating <i>poorer</i> quality of life	≤ 35
Anxiety	7-item Generalised Anxiety Disorder Scale (GAD-7)	12 Likert-scaled questions measuring older people's control and autonomy, pleasure as well as self-realisation. Each symptom is rated on a 4-point scale from not at all to nearly every day.	1-28, high scores indicating more anxiety symptoms	≥ 10

Note: Clinical thresholds and their justification, including the CESD-8 ≥ 4 cut-off for major depressive disorder, are detailed in Section 1.6.3 (Steffick, 2000).

Polygenic Scores

I used four polygenic scores for mental-health-related phenotypes in the ELSA Genetic data, including the polygenic scores (PGS) for depressive symptoms (DS), for major depressive disorder (MDD)-2018, for anxiety disorders (AD) (factor score) and for Loneliness. GWAS summary statistics are provided in Table 4.2. In the regression analyses, all PGS variables were Z-score standardized (Mean = 0, Standard Deviation = 1). Correlations between these polygenic scores, with mental health measures as well as demographic and socioeconomic characteristics are shown in Table 4.3.

Table 4.2 The summary statistics for PGS for DS, MDD, AD and Loneliness

PGS	Sample Size	Mean	SE	Range
Depressive symptoms	7183	5224.7	0.16	5170.3-5283.8
Major depressive disorder-2018	7183	202.2	0.26	-283.8--116.9
Anxiety disorders (factor score)	7183	7449.6	0.30	7332.1-7556
Loneliness	7183	34737.4	0.81	34514.4-35006.3

Note: PGS values shown unstandardised; PGSs were standardised (mean = 0, SD = 1) for regression analysis.

Table 4.3 Correlations between Study Variables

Variables	Depression	Loneliness	QoL	Anxiety	PGS DS
Depression	1				
Loneliness	0.538*	1			
QoL	-0.612*	-0.545*	1		
Anxiety	0.675*	0.437*	-0.501*	1	
PGS DS	0.058*	0.036*	-0.044*	0.064*	1
PGS MDD	0.081*	0.049*	-0.049*	0.091*	0.510*
PGS AD	-0.005	-0.016	0.005	0	0.209*
PGS Loneliness	0.028*	0.02	-0.026*	0.033*	0.075*
Age	0.047*	0.050*	-0.163*	-0.060*	-0.021
Wealth	-0.152*	-0.108*	0.195*	-0.119*	-0.059*
Couple Relationship	0.135*	0.273*	-0.136*	0.012	-0.024

Variables	PGS MDD	PGS AD	PGS Loneliness	Age	Wealth	Couple Relationship
PGS MDD	1					
PGS AD	0.261*	1				
PGS Loneliness	0.105*	0.153*	1			
Age	-0.012	-0.032*	-0.009	1		
Wealth	-0.075*	-0.023*	-0.01	-0.071*	1	
Couple Relationship	0.004	0.004	0.007	0.224*	-0.261*	1

Note: Pearson correlations. Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. All correlations computed on analytic sample ($N = 2,494$).

4.3 Statistical Analysis

Identification Strategy: The Two-Way Fixed Effects (TWFE) Framework

To rigorously assess the moderating effect of genetic susceptibility during the COVID-19 pandemic, this study employs a Two-Way Fixed Effects (TWFE) framework. This choice represents a deliberate transition from standard individual fixed effects (FE) models to a more robust quasi-experimental identification strategy. The core methodological challenge in evaluating Gene-by-Environment ($G \times E$) interactions over time lies in addressing unobserved confounding.

The advantage of using individual fixed effects (FE) is its ability to absorb all time-invariant heterogeneity. This includes stable baseline traits, socioeconomic background, and ancestral population structure. However, a standard individual FE model remains inherently vulnerable to omitted variable bias caused by macro-temporal shocks. In the absence of explicit temporal controls, any population-wide change in mental health—driven by the collective experience of lockdowns, seasonal variations, or the global media cycle—would be mistakenly conflated with the specific $G \times E$ interaction.

While this model accounts for "between-person" differences, it remains inherently vulnerable to omitted variable bias caused by macro-temporal shocks. In the absence of explicit temporal controls, any population-wide change in mental health—driven by the collective experience of the pandemic—would be mistakenly conflated with the effect of individual genetic interactions.

To ensure internal validity, two-way fixed effects (TWFE) estimation can be used for an intervention analysis with a common entry time into treatment to adjust for both unit-specific and time-specific confounders. Addressing the critique regarding the absence of

a traditional, non-exposed control group—a condition unfeasible during a universal pandemic—this study redefines the identification of the causal effect through the framework of Heterogeneous Treatment Effects (HTE) (Abadie & Cattaneo, 2018; Wooldridge, 2010). In this quasi-experimental setup, the COVID-19 pandemic is conceptualised as a universal exogenous shock (the "treatment"), while the PGSs serve as the moderator of susceptibility. Causal identification in this framework does not rely on a binary cross-sectional split of the sample into "exposed" and "non-exposed" groups. Instead, it utilises a multi-dimensional counterfactual architecture:

First, the model utilises within-person temporal variation, treating each participant's pre-pandemic observations as their own stochastic counterfactual (Donald & Lang, 2007). This longitudinal approach benchmarks individual changes against their own pre-shock trajectory, effectively eliminating bias from stable personal characteristics. Second, by operationalising the gradient of genetic liability, the model identifies differential sensitivity to the same macro-environmental shock. In this context, individuals at the lower end of the genetic risk distribution serve as a functional control group, representing the common population-level response to the pandemic. The coefficient of interest thus identifies the excess risk gradient—the degree to which higher genetic risk exacerbates the impact of the universal shock relative to this baseline trend.

Accordingly, the TWFE model controls for both individual and wave-level fixed effects. Interaction terms between the pandemic exposure (comparing pre-pandemic Wave 9 to the two waves of the COVID-19 sub-study) and PGSs are included to identify the moderating role of genetic risk in the relationship between the pandemic and mental health outcomes. To examine the moderating role of different genetic predispositions, I estimate the TWFE model separately for each of the four polygenic scores (PGSs for

depressive symptoms, major depressive disorder, anxiety disorders and loneliness). This parallel specification allows this chapter to isolate the specific interaction between the pandemic and each distinct genetic risk factor.

To ensure the robustness of our estimates, I also controlled for population stratification by incorporating the first ten genetic principal components into the model. Following the methodological rigour suggested by Keller (2014), we specifically included interaction terms between these PCs and the pandemic indicator to complement the focal interaction term between the PGS and the pandemic. This specification accounts for potential time-varying effects of ancestral or geographic backgrounds, ensuring that the observed genetic moderation is not a proxy for differential sensitivity or exposure to the pandemic that might be correlated with population structure. By accounting for these interactions, the model effectively isolates the specific psychological response attributable to the PGSs, net of any broader, ancestry-linked trajectories during the crisis.

Consequently, this approach strengthens the causal claim that the estimated effects reflect true genetic susceptibility rather than residual confounding from ancestral stratification. Specifically, for each mental health outcome, we estimate the following model:

$$MH_{it} = \beta_0 + \beta_1 Post_{it} + \beta_2 (Post_{it} \times PGS_{ki}) + \beta_3 (Post_{it} \times PC_{1-10,i}) + \alpha_i + \lambda_t + \epsilon_{it} \quad (4.1)$$

where i denotes the individual and t denotes the survey wave; MH_{it} represents a measure of mental health outcomes for individual i at time t ; $Post_{it}$ is a dummy variable indicating whether a respondent belongs to a post-COVID-19 cohort (Wave 9 vs. COVID-19 sub-study Wave 1/2), set to 0 for the pre-pandemic baseline (ELSA Wave 9, 2018/19) and 1 for the COVID-19 substudies (Waves 1 and 2, 2020); PGS_{ki} represents one of the four standardized polygenic scores (where $k \in \{\text{Depressive Symptoms, MDD, AD, Loneliness}\}$). Each PGS is entered into the model separately to estimate its specific interaction with the pandemic ($Post_{it} \times PGS_{ki}$); $PC_{1-10,i}$ is a vector of the first ten genetic principal components. I included the interaction term $Post_{it} \times PC_{1-10,i}$ to control for potential time-varying effects of population stratification; α_i represents the individual fixed-effects; λ_t represents the time (wave) fixed-effects, ϵ_{it} is the residual. All analyses are stratified by sex to account for potential gender differences in the genetic moderation of pandemic-related distress.

Crucially, λ_t captures individual fixed effects, which absorb all time-invariant heterogeneity, including stable baseline traits and ancestral structure. As λ_t is collinear with the time-invariant main effects of PGS_{ki} and the ancestry Principal Components ($PC_{1-10,i}$), these main effects are mathematically absorbed and not explicitly estimated. The identification thus focuses solely on the focal interaction term β_2 , capturing how the psychological response to the pandemic shock varies across the gradient of genetic liability.

For outcome variables, I included four countable mental health measures in the TWFE regression, including depressive symptoms (CESD-8), loneliness, anxiety, and quality of life. In addition, since I had both PGSs for depressive symptoms and PGSs for major depressive disorder, I also created one binary measure of major depression using a cut-

off point of four or more depressive symptoms to identify likely cases of clinical depression ($\text{CESD-8} \geq 4$).

However, it should be noted that the data structure for anxiety symptoms (measured via GAD-7) differs from the other outcomes. While depressive symptoms, loneliness, and quality of life were measured at both the pre-pandemic baseline (Wave 9) and the subsequent COVID-19 sub-study waves, anxiety data were only collected during the two pandemic-era sub-studies. Consequently, the analysis for anxiety is restricted to a comparison between the two pandemic periods (COVID Wave 1 vs. Wave 2) to observe the anxiety symptoms, rather than a pre-to-post pandemic longitudinal change as specified in the primary model (Equation 4.1)

As shown in Table 4.4, the selection of regression models for each outcome was guided by the distributional properties of the outcome variables. All count outcomes were modelled using count regression rather than OLS to respect the non-negative, discrete, and right-skewed nature of these distributions. For loneliness, anxiety, and quality of life, the variance-to-mean ratios were close to unity in the baseline data, and likelihood ratio tests comparing Poisson and negative binomial specifications did not reject the null of equidispersion; these outcomes were therefore modelled using standard Poisson regression. For depressive symptoms (CESD-8, range 0–8), the variance substantially exceeded the mean at baseline (mean = 1.2, SD = 1.6, implying variance ≈ 2.56), indicating overdispersion that violates the equidispersion assumption of standard Poisson regression; negative binomial regression, which incorporates a dispersion parameter to accommodate excess variance, was therefore used for this outcome. For the binary major depression outcome, I used modified Poisson regression with a robust error variance approach (Zou, 2004), which directly estimates relative risk rather than the

odds ratio — preferable when the outcome is not rare (prevalence $\approx 15\%$ in this sample) and avoids the non-collapsibility problem of the odds ratio in models with covariates. For the Poisson and negative binomial regressions, rate ratio estimates and 95% confidence intervals are presented.

Table 4.4 Statistical Analysis for Different Outcomes

Mental Health Outcome	Scale	Statistical Analysis
Depressive Symptoms	CESD-8	Negative Binomial Regression
Major Depressive Disorder	CESD-8 $\geq 4, 1$ or CESD-8 $< 4, 0$	Modified Poisson Regression
Loneliness	T-ILS	
Anxiety	GAD-7	Poisson Regression
Quality of Life	CASP-12	

Most missing data in this study came from two aspects. First, the ELSA COVID-19 longitudinal sample (ELSA Wave1, COVID sub-study Wave1 and Wave2) consists of 5146 individuals, of whom 2494 individuals (48.5%) participated in the ELSA genetic survey. Therefore, data on mental health and genetic risk are available for 2494 participants. Moreover, due to mortality selection in longitudinal social surveys (Ware et al., 2017), the sample of respondents who were genotyped in the genetic data was healthier and longer-lived as compared with birth-cohort peers in the general ELSA data (as shown in Table 3.4 in the previous chapter). Therefore, in one of the sensitivity analyses, I estimated the same models using the weighted sample. The inverse probability weighting (IPW) was applied in the sensitivity analyses to account for the selection bias and missingness in the genetic sample. Second, due to a survey error, the last item of the CESD-8 questionnaire for depressive symptoms was not administered

for around 75% of the sample in the first wave of the ELSA COVID-19 sub-study. I assumed that these CESD-8 item data were missing completely at random (MCAR). I used multiple imputation (MI) by chained equations with all variables of the analysis included as predictors of the imputation models as well as auxiliary variables, including age, sex, education, wealth, employment, etc. 20 imputations were used and then the regression estimates were pooled across the imputed datasets using Rubin's rules (Appendix) (Rubin, 1987).

I conducted four types of sensitivity analyses. First, I estimated the same models using the sample of participants with complete data on all variables without imputation. Second, I reran the models including the interaction terms between the COVID-19 and socioeconomic characteristics (age, wealth, and partnership) in addition to the COVID-19 and PGSs interaction terms. Third, I investigated the sensitivity of the results using inverse probability weighted data. Fourth, I estimated the same models using terciles of PGSs rather than continuous PGSs. All analyses were conducted in Stata/SE v18.0. Due to participant confidentiality and data protection regulations, the raw data cannot be shared directly by the author. Researchers interested in accessing the data must apply through the UKDS following the standard procedures. The analysis code necessary to reproduce the results are available from the author upon reasonable request.

4.4 Results

This study included respondents of European ancestries from ELSA Wave 9 as well as the two waves of the COVID-19 Sub-study. Table 4.5 reports the descriptive statistics of the included sample (2494 observations) and non-included sample at the ELSA Wave 9. The mean age in this sample is 73.4. 55.0% of the sample is male. The included genotyped sample are generally older, wealthier and healthier. The genetic characteristics of the included sample are close to the original genetic sample.

Table 4.5 Comparison of the sample included vs. non-included from the ELSA COVID-19 longitudinal study

	Included		Non-included		P Value (T-test/ Chi-
	(N=2494)		(N=2652)		squared test)
(Mean, SD)					
Age	73.4	6.5	64.5	9.2	<0.001
Depressive symptoms CESD-8	1.2	1.6	1.3	1.8	0.006
Loneliness	3.9	1.4	4.1	1.5	<0.001
Quality of Life CASP-12	39.3	5.5	39.0	6.0	0.030
Anxiety GAD-7	9.5	3.7	10.3	4.3	<0.001
PGS for Depressive symptoms	5224.4	13.6			
PGS for Depressive symptoms (standardised)	0.0	1.0			
PGS for Major depressive disorder-2018	-202.5	22.3			
PGS for Major depressive disorder-2018 (standardised)	-0.0	1.0			
PGS for Loneliness	34514.4	70.3			

	Included (N=2494)	Non-included (N=2652)	P Value (T-test/ Chi- squared test)
PGS for Loneliness (standardised)	-0.0	1.0	
PGS for Anxiety disorders	7450.2	25.4	
PGS for Anxiety disorders (standardised)	-0.0	1.0	
(Percentage)			
Sex, Female	55.0%	58.8%	0.005
Ethnicity, White	100.0%	92.8%	<0.001
Couple Relationship, Partnered	68.3%	67.3%	0.130
Education			
Low	37.9%	22.9%	
Middle	39.9%	50.1%	<0.001
High	22.3%	27.0%	
Long-standing -illness	50.5%	50.5%	<0.001
Wealth			
Lowest	12.2%	19.2%	
Middle	28.9%	29.1%	<0.001
Highest	58.9%	51.7%	
Work Status			
Retired	83.6%	47.7%	
Employed	12.3%	42.8%	<0.001
Other	4.1%	9.5%	

Note: Means (SD) or proportions shown. Differences tested with t-tests or chi-square. Non-included respondents are ELSA participants not in genotyped analytic sample. Use IPW in sensitivity analyses.

Figure 4.1 and Table 4.6 present the changes in mental health outcomes before and during COVID-19, with higher percentages of depression, loneliness and anxiety, and lower percentages of good quality of life in both sexes, which all indicate poorer mental

health outcomes. There is a significant sex difference in the mental health outcomes, with poorer mental health scores among men than women. Some recent molecular and empirical studies have also suggested sex-dependent or sex-specific effects on mental health. Therefore, I stratified by sex in the following analyses.

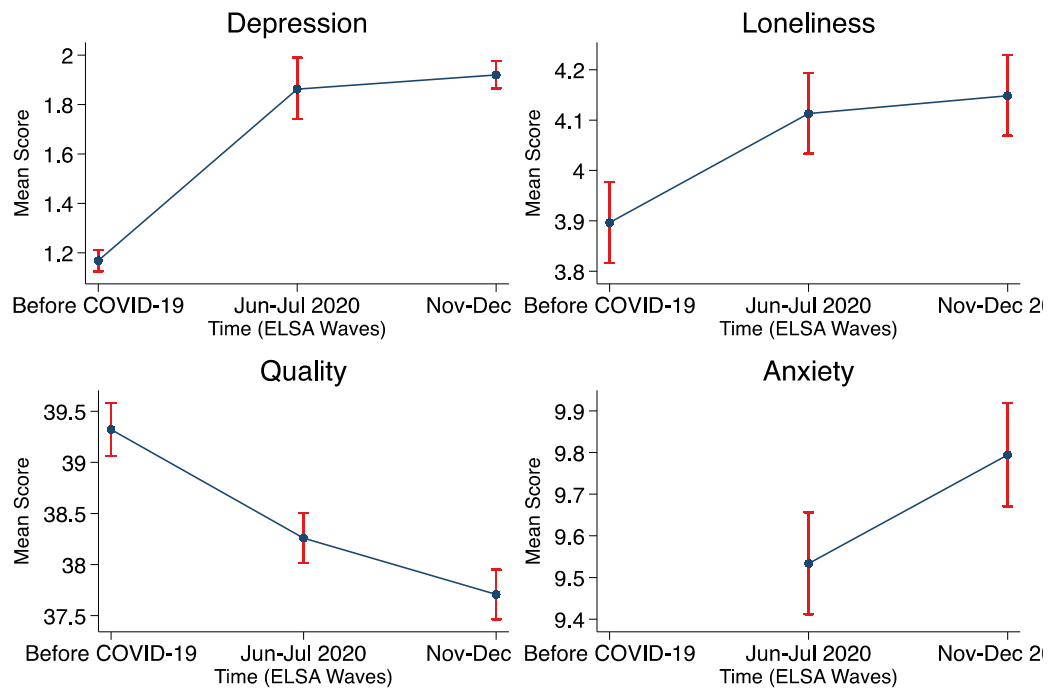


Figure 4.1 Mental Health Trajectories before and after the COVID-19 Pandemic

Note: Sample = ELSA analytic sample (genotyped subset with valid outcomes; N = 2,494). Points are sample means for each wave; vertical bars are 95% confidence intervals. Waves: Wave 9 (pre-pandemic, 2018–19); COVID Wave1 (June–July 2020); COVID Wave2 (Nov–Dec 2020).

Table 4.6 Descriptive changes in mental health before and during COVID-19 by sex

	ELSA Wave9		ELSA COVID Wave1		ELSA COVID Wave2	
	Male	Female	Male	Female	Male	Female
Depression, %	5.2	11.4	14.0	23.9	14.5	26.2
Chi2 Test P Value by Sex		<0.001		0.009		<0.001
Chi2 Test P Value by Wave						<0.001
Loneliness, %	21.7	27.3	24.5	36.8	25.1	37.0
Chi2 Test P Value by Sex		0.002		<0.001		<0.001
Chi2 Test P Value by Wave						<0.001
Quality of Life, %	81.3	80.3	78.7	73.0	75.0	69.0
Chi2 Test P Value by Sex		0.548		0.001		0.001
Chi2 Test P Value by Wave						<0.001
Anxiety, %			27.8	40.9	30.4	44.5
Chi2 Test P Value by Sex				<0.001		<0.001
Chi2 Test P Value by Wave						0.021

I examined the within-person effects of the COVID-19 pandemic on mental-health outcomes and then report how those effects are moderated by polygenic scores. The analysis employs a series of two-way fixed-effect regressions to exploit within-individual variation. To account for potential structural differences in how the pandemic and genetic risks affect mental health across genders, all models are estimated separately for males and females rather than merely adjusting for sex as a covariate. Figures 4.2 and 4.3 provide a comprehensive suite of coefficient plots delineating the results from these interaction models. Each figure is organized into five distinct panels, facilitating a systematic comparison across diverse psychological and well-being dimensions: (A) depressive symptoms (D) (CES-D 8 scale), (B) the binary probability of major

depressive disorder (MDD) (caseness threshold of CESD-8 ≥ 4), (C) loneliness, (D) quality of life (QoL) (CASP-19 scores), and (E) anxiety symptoms. Within each panel, the visualisation contrasts the baseline pandemic main effect (the Post COVID-19 indicator) against specific G×E interaction terms (e.g., Post × PGS MDD). This multi-panel structure is designed to allow for a direct comparative assessment of how the same exogenous environmental shock—the COVID-19 pandemic—is differentially moderated by various genetic risks across different outcomes. The x-axis represents exponentiated coefficients (Incidence Rate Ratios, IRRs), where a value of 1.0 denotes the null hypothesis. This facilitates the immediate identification of whether a specific polygenic score amplifies or attenuates the pandemic’s impact. To ensure the robustness of these findings, sensitivity analyses are summarised in Supplement S4.1. Full regression estimates, including standard errors, are documented in Supplementary Tables S4.2–S4.11. To quantify the explanatory power of the gene-environment (G×E) interactions, I calculated approximate effect sizes ΔR^2 and Cohen’s f^2 for each model using linear fixed-effects regressions as an approximation (Table S4.12). These metrics reflect how much additional variance in mental health outcomes is explained by the interaction terms.

Main Effects of the COVID-19 Pandemic

Consistent with the descriptive analyses, the pandemic had an adverse overall effect on mental health. These are the main pandemic effects (Post COVID-19) rather than the G×E interactions: across sexes the pandemic is associated with more depressive symptoms, major depressive disorder and loneliness and with lower QoL, with anxiety increases more consistently observed among females.

Specifically, For females, the pandemic is associated with statistically significant adverse changes: depressive symptoms increase (Post COVID-19: IRR = 1.562, SE = 0.062; 95% CI 1.441–1.683; Table S4.7), indicating a 56.2% higher expected CESD-8 score post-pandemic, and the probability of major depressive disorder rises substantially (Post COVID-19: IRR = 2.275, SE = 0.204; 95% CI 1.877–2.758; Table S4.8), indicating a 2.28-fold higher risk of caseness; loneliness increases (Post COVID-19: IRR = 1.085, SE = 0.011; Table S4.9) and quality-of-life scores decline (Post COVID-19: IRR = 0.952, SE = 0.004; Table S4.10). These female main effects are robust across the sensitivity checks reported in Supplement S4.1.

For males the adverse shift is more outcome-specific: the Post COVID-19 indicator is not consistently significant for continuous depressive symptoms or major depressive disorder (DS: IRR = 1.016, SE = 0.068, Table S4.2; MDD: IRR = 0.846, SE = 0.114, Table S4.3), loneliness shows no consistent post effect (Post COVID-19: IRR = 0.989, SE = 0.008; Table S4.4), whereas quality of life exhibits a small but statistically significant post change (Post COVID-19: IRR = 1.015, SE = 0.003; Table S4.5) and anxiety increases at the Nov–Dec 2020 wave (Nov–Dec 2020: IRR = 1.018, SE = 0.010; Table S4.6).

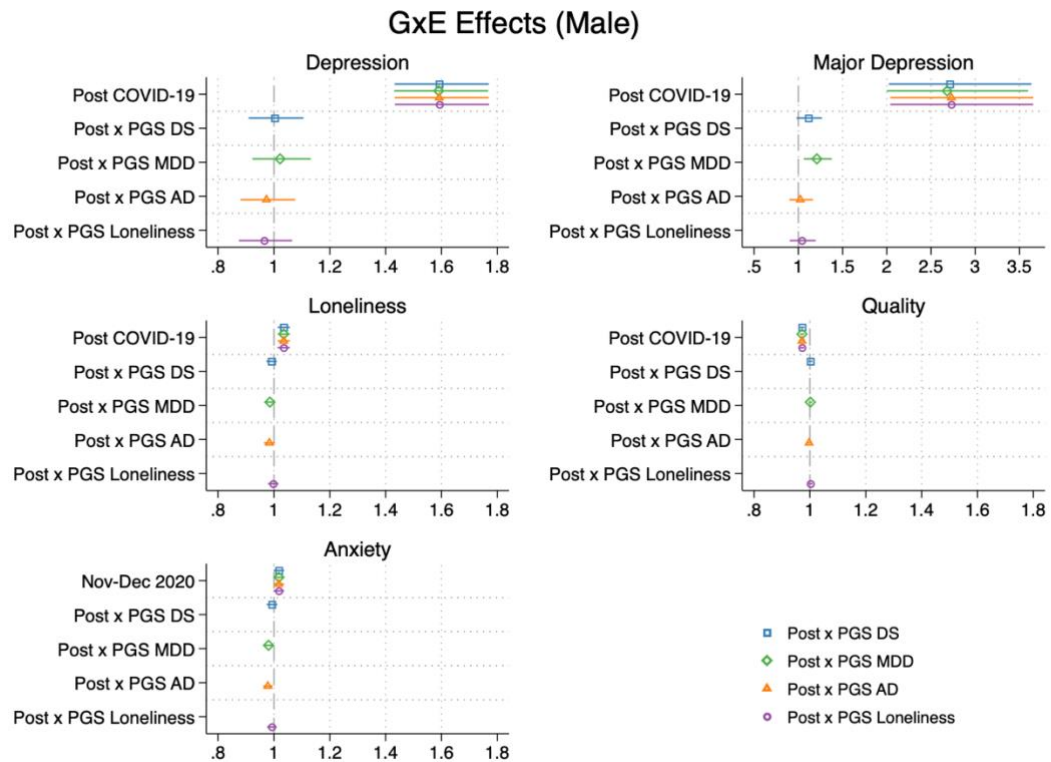


Figure 4.2 Post × PGS Interaction Effects on Mental-Health Outcomes (Males)

Note: “Post COVID-19” denotes observations collected in the two ELSA COVID substudies (June–July 2020 and November–December 2020). The Post indicator equals 0 for the pre-pandemic baseline (ELSA Wave 9, 2018–19) and 1 for the two COVID substudy waves. For anxiety models that use a wave-specific indicator, the Nov–Dec 2020 occasion is modelled separately. Coefficient plot of exponentiated coefficients (IRRs) from TWFE models. Each symbol represents an IRR for Post or Post×PGS from separate models for PGS DS, PGS MDD, PGS AD, PGS Loneliness. Error bars = 95% CIs. See Supplementary Tables S4.2–S4.7 for full coefficients and sample sizes. Exponentiated coefficients (IRR) reported; value 1 = null.

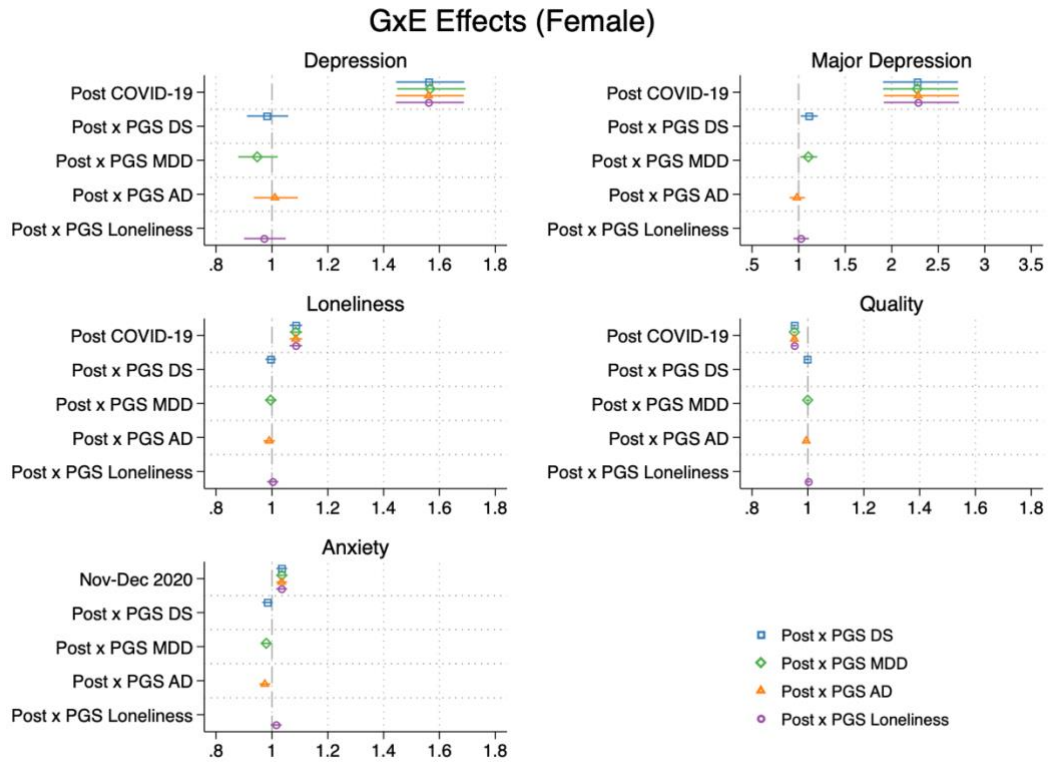


Figure 4.3 Post $\times$ PGS Interaction Effects on Mental-Health Outcomes (Females)

Note: As Figure 4.2 but for the female sample. See Tables S4.7–S4.12 for full results.

These estimates describe the direct effects of the pandemic (the COVID-19 pandemic main effects), not the COVID $\times$ PGS interaction; the observed sex differences in these main effects further justify sex-stratified analyses. Overall, the pandemic's impact manifests as a universally detrimental shift in mental health trajectories for both men and women. However, the magnitude of these changes exhibits a distinct gender gradient: while both sexes experienced significant declines, the effect sizes for females are notably larger, especially regarding the increased depressive symptoms and the increased probability of major depressive disorder.

PGS × Pandemic (G×E) Interactions

The G×E interaction effects are heterogeneous across PGSs, mental health outcomes and sex. The clearest interaction signal emerges for major depressive disorder in males.

Figure 4.2 and Table S4.3 report a Post × PGS MDD interaction of IRR = 2.683 (SE = 0.402; approximate 95% CI 1.90–3.47; $p < 0.01$), indicating a substantial increase in the within-person post effect for men with higher genetic risks for MDD.

For females the Post × PGS MDD interaction on MDD is positive and statistically detectable but markedly smaller: Table S4.8 reports IRR = 1.104 (SE = 0.046; $p < 0.01$). The female effect therefore indicates modest genetic amplification of pandemic-related MDD risk, yet $\Delta R^2/f^2$ for this cell (Table S4.12) are an order of magnitude smaller than for males. In short, both sexes show PGS MDD moderation of major depressive disorder, but the male effect is substantially larger in both relative and absolute terms.

Results for continuous depressive symptoms exhibit less consistency across models.

While the interaction between the pandemic and PGS DS among males (Table S4.2) suggests a significant relative increase in depressive symptoms reported (IRR = 1.592, SE = 0.086), the incremental variance explained remains negligible ($\Delta R^2 \approx 0.001$; Table S4.12). When these relative estimates are converted into absolute CES-D 8 scores through marginal analysis, the resulting shifts in predicted symptom counts appear substantively minor. To illustrate this, marginal analysis shows that for a male individual at the mean level of genetic risk (PGS DS = 0) the onset of the pandemic is associated with an increase in the expected CES-D-8 count of approximately 0.465 units ($dy/dx = 0.4648$; SE = 0.0539; 95% CI = 0.3591–0.5704). For an individual at +1 SD of genetic risk for DS (roughly the upper-middle of the distribution) the predicted increase is approximately 0.469 units ($dy/dx = 0.4694$; SE = 0.0725; 95% CI = 0.3273–0.6115).

The absolute difference between these two marginal effects is therefore only ≈ 0.0047 CES-D-8 units ($0.4694 - 0.4648 \approx 0.0047$), which is a negligible amount. Given that the standard deviation of the CES-D 8 score in this analytic sample is 1.6) the pandemic-associated increase at the mean corresponds to $\approx 0.465/1.6 \approx 0.29$ SD, whereas the difference in that increase between mean and +1 SD individuals corresponds to $\approx 0.0047/1.6 \approx 0.003$ SD. Thus, although the Post $\times$ PGS interaction is statistically detectable in the model, the increment in average symptom counts attributable to higher genetic liability is functionally trivial on the CES-D-8 scale.

Among females, Post $\times$ PGS interactions for depressive symptoms are generally closer to unity (Table S4.7) and lack robustness across different specifications. In sensitivity checks, the directional patterns for depressive symptoms mirror those for major depressive disorder, yet the statistical evidence is considerably weaker. This is consistent with the theoretical expectation that acute environmental shocks may trigger clearer G $\times$ E signals in extreme clinical outcomes than in modest shifts across a continuous symptom distribution.

Anxiety models use a Nov–Dec 2020 indicator for the second pandemic wave. The Nov-Dec $\times$ PGS AD interaction is statistically detectable in both sexes but of vanishing practical magnitude. In males (Table S4.6) Nov-Dec $\times$ PGS AD IRR = 0.981 (SE = 0.009; $p < 0.05$) and in females (Table S4.11) IRR = 0.975 (SE = 0.010; $p < 0.01$). Both IRRs are very close to 1 and Table S4.12 shows essentially zero incremental variance explained ($\Delta R^2 \approx 0$, $f^2 \approx 0$). Margins convert these IRRs into absolute predicted-probability differences well below one percentage point. Therefore, although the sign is consistent and reproducible across some robustness checks, the interaction is practically negligible at the population level.

Across PGS specifications I find little evidence of meaningful $G \times E$ moderation for loneliness or QoL. Some interaction coefficients for loneliness (for example, male Post $\times$ PGS Loneliness reported as $IRR \approx 1.035$ in Table S4.4) achieve statistical significance because standard errors are small, but the relative magnitudes are tiny, and margins imply minimal absolute changes; $\Delta R^2/f^2$ values for these interactions are orders of magnitude smaller than conventional thresholds for practical importance (Table S4.12). QoL models similarly show Post $\times$ PGS estimates clustered around 1 in both sexes (Tables S4.5 and S4.10) and negligible explained variance. In short, the coefficient plots (Figures 4.2–4.3) and full regressions (Tables S4.4, S4.5, S4.9, S4.10) together indicate no consistent, policy-relevant moderation of pandemic effects on loneliness or QoL by the PGSs tested.

Across $G \times E$ outcomes, the values of f^2 were generally small (e.g., $f^2 < 0.01$), suggesting that while the $G \times E$ interactions are statistically observable, their practical effect sizes are limited (Table S4.12). This aligns with the broader literature indicating small effect sizes in polygenic score $\times$ environment studies (Dudbridge, 2013; Keller, 2014). The highest observed effect sizes were in models where PGS MDD interacted with the pandemic to predict major depressive disorder among males, supporting the earlier graphical findings (Figures 4.2-3).

The interaction effect of PGS MDD and Post COVID-19 on depression for males is demonstrated in Figure 4.4. The X-axis is the standardized polygenic score for major depressive disorder, and the Y-axis is the probability of major depressive disorder (CESD-8 ≥ 4). The solid line represents the time after the COVID-19 pandemic and the dash line represents the time before the pandemic.

For males, sensitive individuals with a higher PGS for major depressive disorder and those who are less sensitive with a lower PGS MDD do not differ from one another with respect to depressive symptoms before the COVID-19 pandemic. Differences in overall sensitivity, however, lead to a departure such that individuals with a higher PGS MDD report significantly higher levels of depressive symptoms in increasingly stressful environments after the outbreak of the COVID-19 pandemic compared to those with a lower PGS MDD.

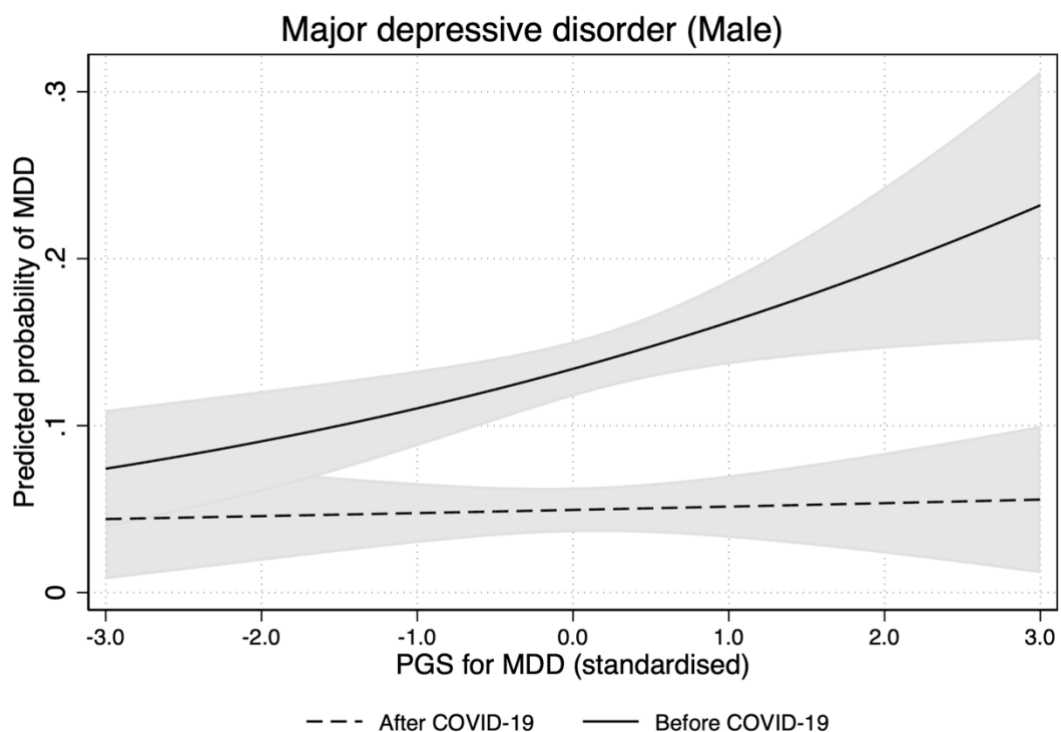


Figure 4.4 Marginal effects of PGS MD x Post COVID-19 with 95% CIs

Note: Predicted probability of major depressive disorder ($CESD-8 \geq 4$) at grid points of standardised PGS MDD. Dashed = pre-pandemic; solid = post-pandemic. Shaded bands are 95% CIs computed using the delta method.

In conclusion, the results indicate selective genetic moderation of pandemic-related mental-health change. The most robust and interpretable G×E signal is the amplification of pandemic effects on major depressive disorder by PGS MDD in men — an effect that is statistically robust, translates into a meaningful absolute increase in the probability of MDD for higher genetic risks, and exhibits the largest interaction effect size in $\Delta R^2/f^2$ terms. Other detected interactions are either very small in absolute terms or explain negligible additional variance; hence, while genetic heterogeneity contributes to differential vulnerability, its practical contribution to population-level change following COVID-19 is limited.

4.5 Discussion

To my knowledge this is the first study to assess the interactive associations of the COVID-19 pandemic and polygenic scores for psychopathology with trajectories of mental health, using a nationally representative sample of older people in the UK. My analysis revealed several important findings. First, levels of depression, loneliness and anxiety increased and ratings of high quality of life decreased significantly during the COVID-19 pandemic. Women reported worse mental health than men before and after the COVID-19 pandemic and exhibited more negative responses to the pandemic. My findings are consistent with previous studies suggesting that older individuals did not adapt well to the stressors introduced by the pandemic and experienced more symptoms of mental health disorders compared with pre-pandemic levels (Zaninotto, 2022; Steptoe, 2021).

Importantly, my study extends prior research by providing suggestive evidence that genetic vulnerability moderated the adverse mental health impact of the COVID-19 pandemic on older people. Specifically, I found that higher polygenic risks for major depressive disorder (PGS MDD) amplified the pandemic's effect on major depressive disorder among men (Post $\times$ PGS MDD: IRR = 2.683; SE = 0.402; 95% CI 1.90–3.47), with margins indicating appreciably larger absolute increases in MDD probability for men in higher PGS strata, in line with the diathesis-stress model emphasizing the toxicity of stressful environments (Monroe & Simons, 1991; Robins & Block, 1989). It is important to note that the findings presented in this chapter reflect associations identified through observational panel analysis and should not be interpreted as causal relationships in the strict counterfactual sense. Although the TWFE specification controls for all time-invariant individual-level confounders, it cannot fully exclude time-

varying confounding from unmeasured factors that changed concurrently with the pandemic onset. Accordingly, findings are described throughout in terms of associations and within-person change rather than causal effects.

My findings of the $G \times E$ interplay for major depressive disorder support those from Colodro-Conde et al. (Colodro-Conde et al., 2018) and Arnau-Soler et al. (Arnau-Soler, 2019), with an independent sample and study design. Colodro-Conde et al. (Colodro-Conde et al., 2018) used scores of stressful life events rather than a specific stressful life event in a sample of 5221 individuals from 3083 families in Australia. Their findings pointed to an extra risk for individuals with combined polygenic vulnerability and a high number of reported SLEs beyond the additive contributions of these factors to the liability for depression. Arnau-Soler et al. (Arnau-Soler, 2019) tested the stress-diathesis model using an independent sample of 4919 unrelated white British participants from Generation Scotland, and assessed the differences between women and men, using self-reported depressive symptoms and recent SLEs. They also found that high PGS MDD increased the risk of depression in participants reporting a high number of SLEs.

At the same time, these interaction effects should be interpreted cautiously. Only a subset of the tested $G \times E$ models yielded significant interactions in my study.

Furthermore, I performed multiple sensitivity checks (complete-data estimation, inclusion of SES interactions, IPW for selection into genotyped sample, and alternative PGS codings such as terciles). The male PGS_MDD $\times$ Post interaction is directionally consistent across many specifications and remains statistically significant in several; other interactions are less stable. Therefore, the general pattern does not provide conclusive support for the diathesis-stress model. The heterogeneity in $G \times E$ effects is well documented and may reflect differences in statistical power, timing of stress

exposure, symptom severity, or the psychosocial buffering factors unequally distributed across subgroups. Thus, while the current results suggest plausible gene-environment interactions—especially for MDD—they also underscore the complexity of this interplay and the importance of contextual and biological moderators. Further research with larger, diverse samples and harmonized measures is needed to elucidate these relationships.

The sex-stratified results - larger pandemic main effects among women but stronger PGS MDD moderation on major depressive disorder among men - deserve careful, non-speculative interpretation. Possible explanations include baseline differences (ceiling/floor effects), sex-dependent biological expression of genetic liability, and gendered differences in coping and help-seeking. For example, higher baseline symptom levels among women might diminish the marginal detectability of Post×PGS effects on mean scores, while men with lower baseline prevalence might show sharper relative increases that are more likely to cross clinical thresholds. These hypotheses require replication and direct tests (biomarkers, measures of social support and help seeking) before causal attribution. Moreover, these sex-specific findings are contrary to previous research. Colodro-Conde et al. reported PGS MDD × SLEs effects for depression only in women, and Arnau-Soler et al. similarly found interactions only among women when using continuous SLEs scores. However, when SLEs were categorized into levels to investigate diathesis effects, Arnau-Soler et al. detected a significant effect among men who reported no stressful events in the prior six months. Such inconsistencies underscore the need for replication in larger cohorts and more harmonized measures of stress and mental health. While theoretical frameworks like diathesis–stress remain valuable, they require a more nuanced operationalisation that accounts for biological mechanisms and heterogeneity across subgroups.

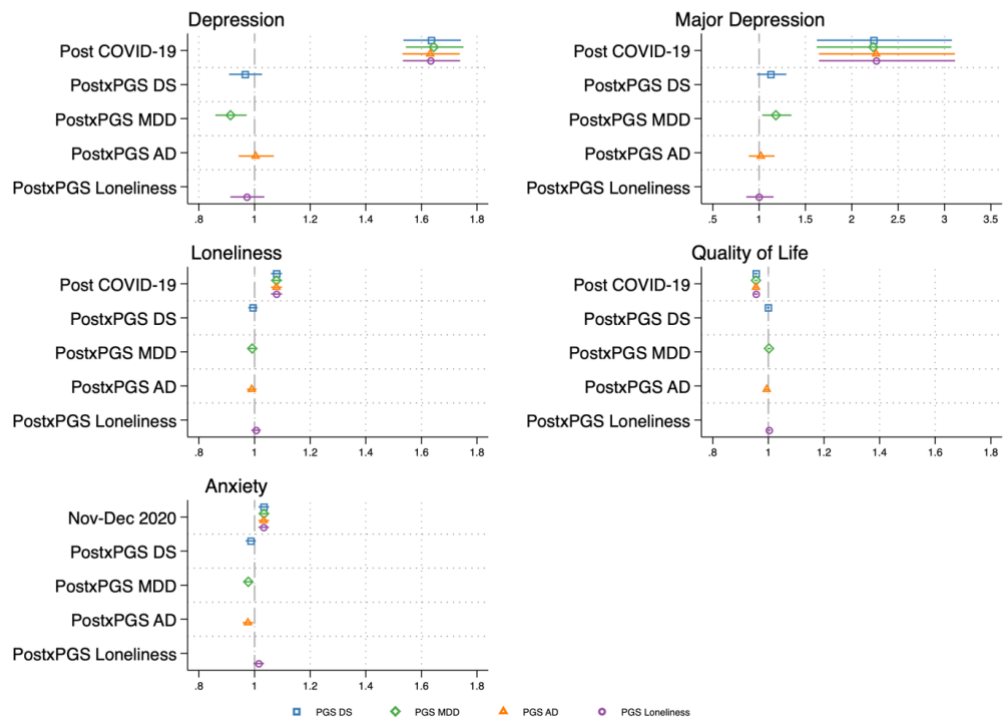
There are several limitations in this study. First, principally, the sample size in this study is 2494, which is relatively small in $G \times E$ studies, resulting in low statistical power for detecting interaction effects. A larger sample size and replication of the current findings are needed to elucidate how the pandemic interacts with genetic vulnerability for depression. Another limitation of this study is that the sample in the ELSA study is highly selected and the ELSA genotyping work was conducted in 2013-2014, which will induce mortality selection bias. While this study uses inverse probability weighting to adjust for non-random selection, observational COVID-19 research is particularly susceptible to collider bias (Holmberg, 2022; Griffith, 2020). Analysing the ELSA data, compared to the wider cohort of the pandemic-exposed individuals were highly selected for a range of genetic, behavioural, cardiovascular, demographic, and anthropometric traits. Moreover, if SNP effects vary across stress levels, average $G \times E$ estimates risk obscuring meaningful heterogeneity (Zhang, 2024; Tropf, 2017). On the genetic side, due to a tiny minority of non-European respondents in the ELSA survey, I only included European ancestries. Therefore, the findings using genetic data constructed from European ancestry GWAS cannot be generalized to other ancestral populations (Gharbi-Meliani et al., 2021). In addition, the PGSs employed were based on GWAS that capture average effects on phenotypes, not genetic influences on plasticity or variance. As such, potential vPGS effects could not be detected (Johnson et al., 2021). On the environmental side, since race can shape individual life and health outcomes (Albin, 1993; Schultz et al., 2008; Taylor et al., 2017), limiting this analysis to European ancestries restricts the scope of potential effects of the pandemic that can be observed. Lastly, this study focuses on the mental health impact of the COVID-19 pandemic among older people. The use of a specific stressful life event may not generalize to other measures of exposure to stressors or populations of other age groups.

In conclusion, mental health deteriorated among older people during the COVID-19 pandemic and a higher PGS MDD increased the risk of depression after the pandemic among males. My findings add to the growing literature on the mental health consequences of the COVID-19 pandemic by showing both overall declines in well-being and suggestive evidence that genetic risks for depression moderated these effects among older men. While these findings are preliminary, they highlight the importance of incorporating genetic risks into analyses of population health shocks. Future research should replicate these findings in larger and more diverse samples, integrate variance-sensitive polygenic measures, and examine mechanisms through which genetic risks interacts with acute and chronic stressors. Such work will be crucial for identifying at-risk populations and informing targeted interventions to mitigate the long-term mental health burden of the pandemic.

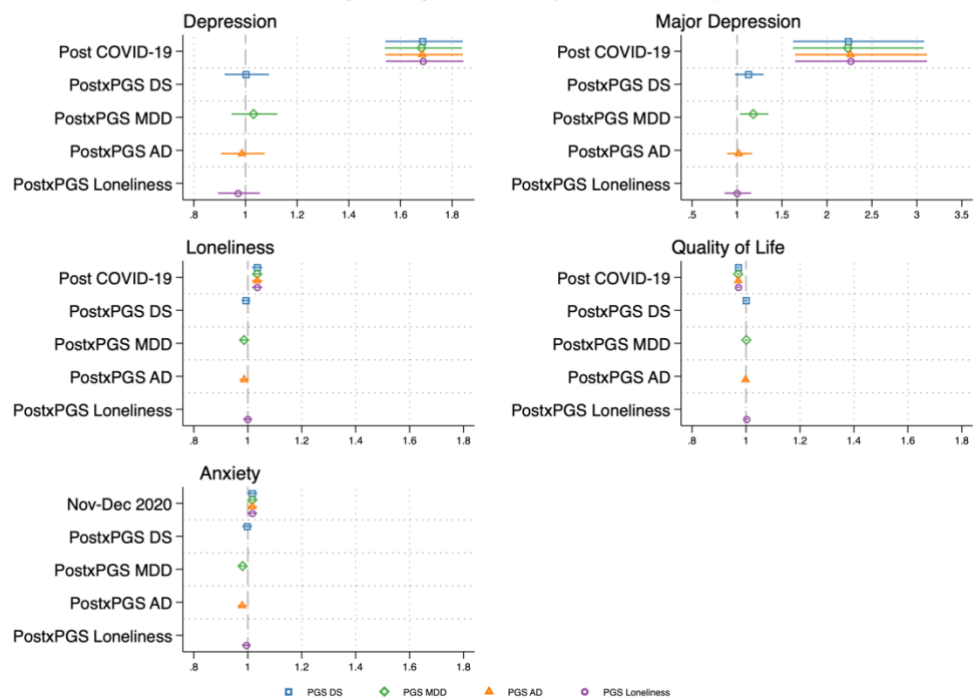
4.6 Supplement

Section S4.1 Sensitivity Analyses Results

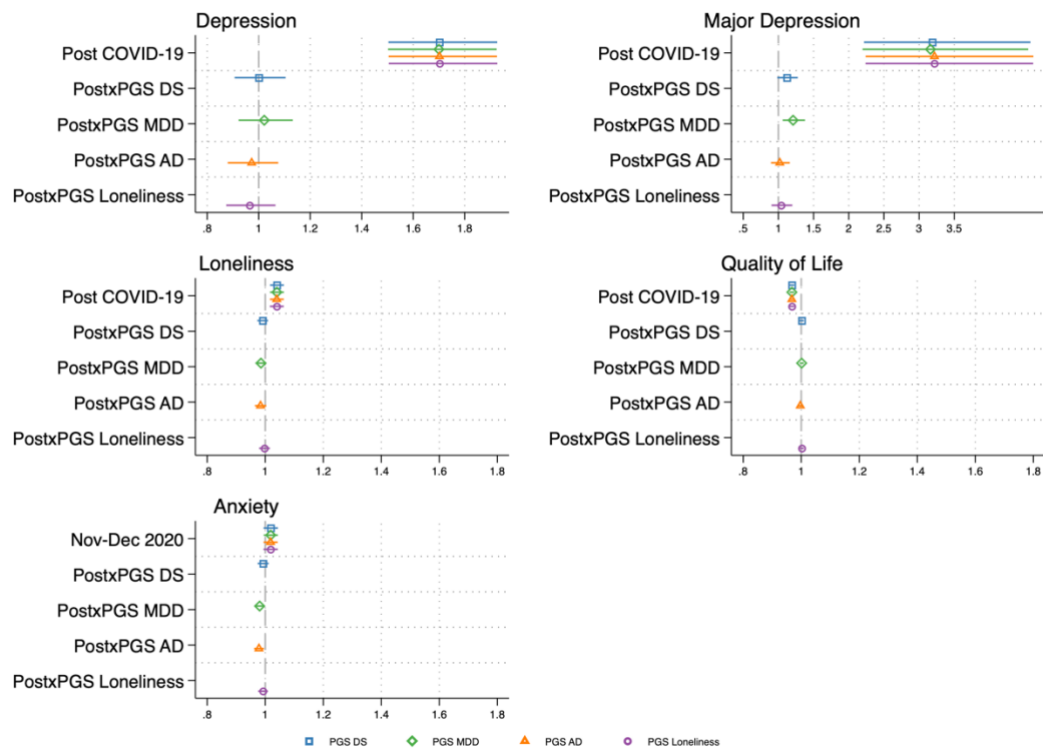
Sensitivity Analyses- Complete Data (M)



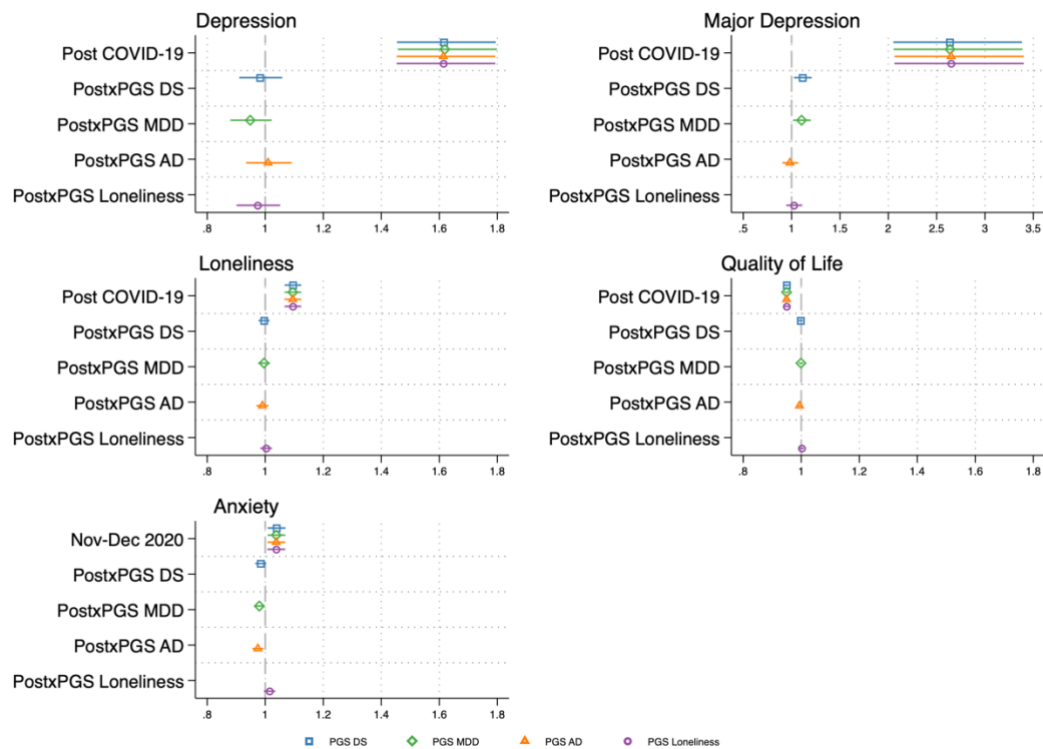
Sensitivity Analyses- Complete Data (F)



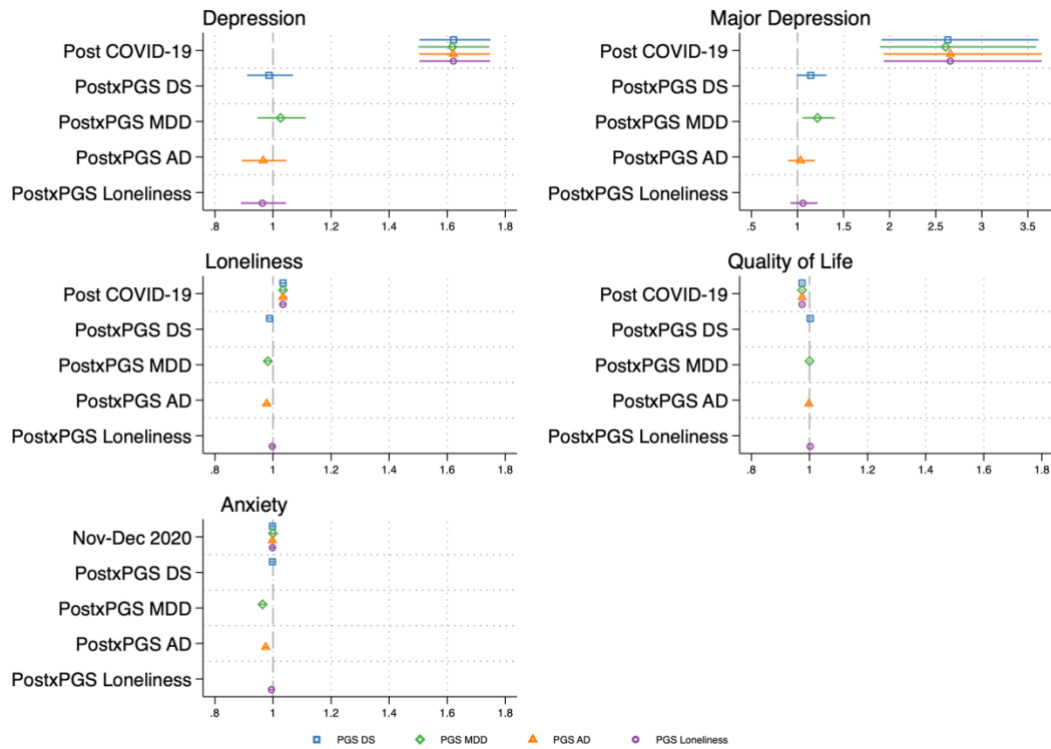
Sensitivity Analyses-including SES Controls (M)



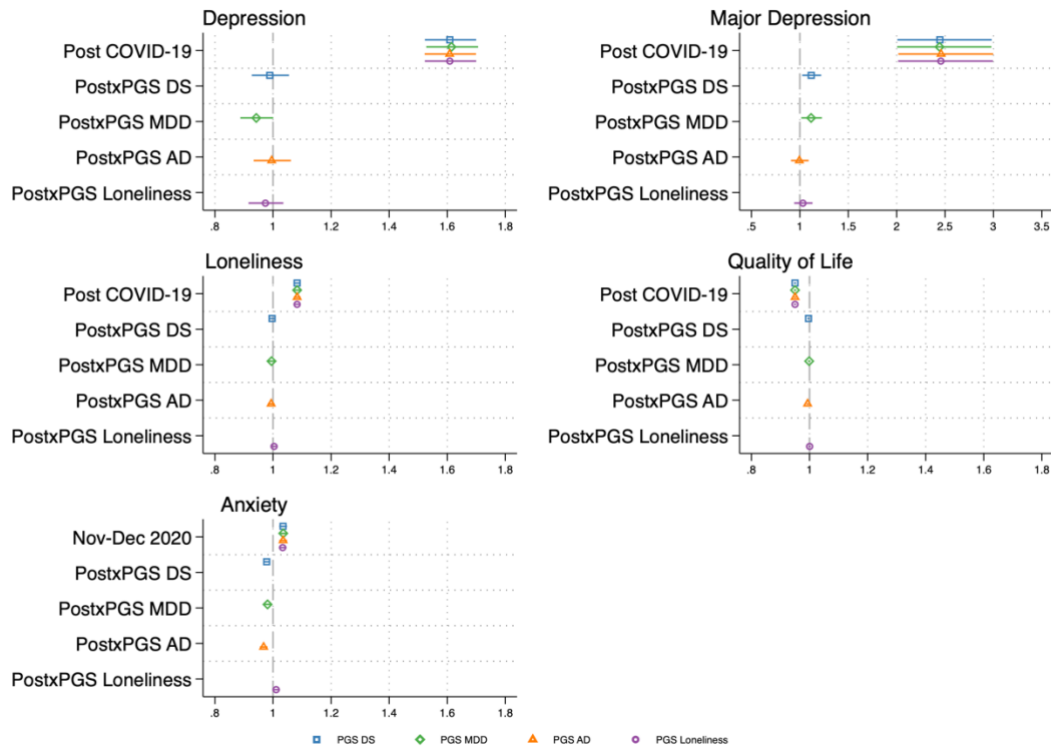
Sensitivity Analyses-including SES Controls (F)



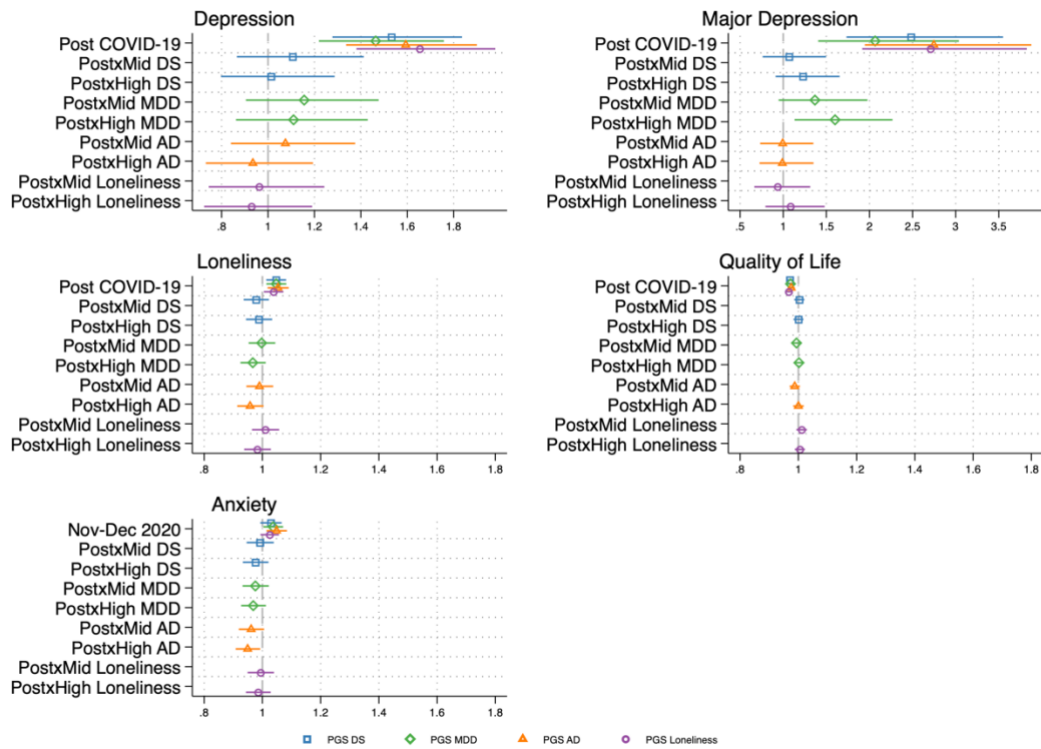
Sensitivity Analyses - Weighted (M)



Sensitivity Analyses - Weighted (F)



Sensitivity Analyses - PGS Terciles (M)



Sensitivity Analyses - PGS Terciles (F)

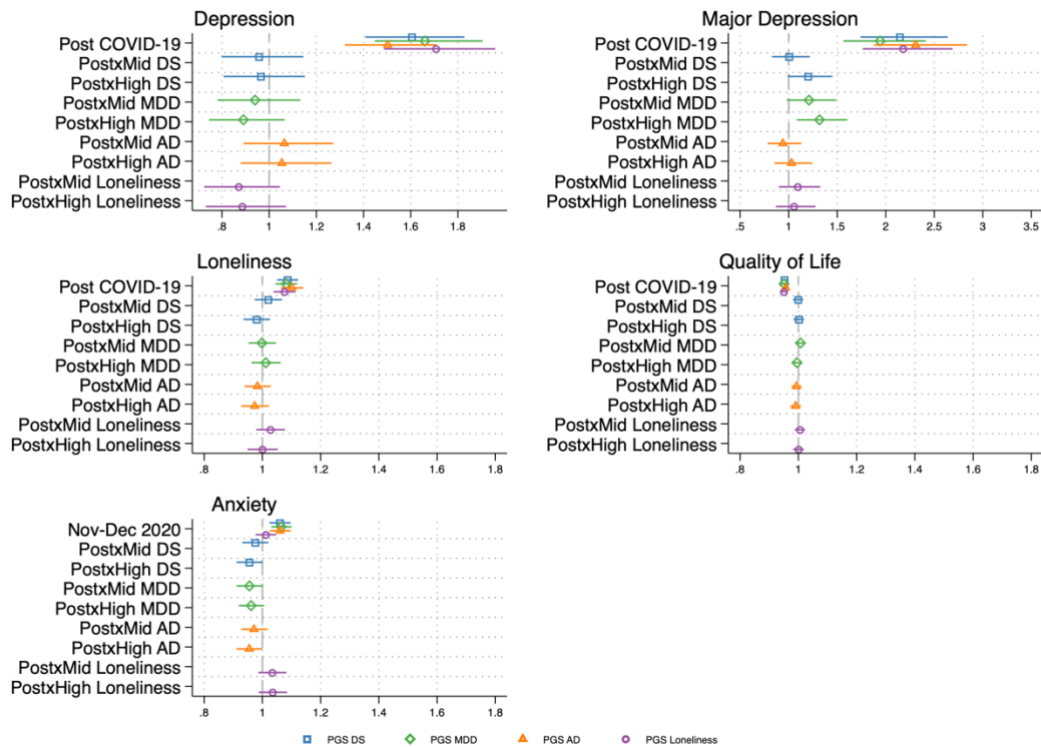


Table S4.2 Full Regression Estimates for Depressive Symptoms Among Males

Depressive symptoms	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	1.592*** (0.086)	1.590*** (0.086)	1.592*** (0.086)	1.593*** (0.086)
COVID	1.016 (0.068)	1.017 (0.068)	1.016 (0.068)	1.017 (0.068)
3rd Wave	1.003 (0.050)	1.022 (0.053)	0.973 (0.050)	0.965 (0.048)
PC1	0.993 (4.188)	0.907 (3.815)	1.183 (4.987)	1.092 (4.602)
PC2	0.094 (0.389)	0.105 (0.436)	0.095 (0.392)	0.088 (0.363)
PC3	19.903 (83.684)	21.202 (89.264)	19.626 (82.460)	21.845 (92.002)
PC4	0.117 (0.519)	0.111 (0.494)	0.108 (0.482)	0.149 (0.666)
PC5	0.071 (0.308)	0.070 (0.300)	0.074 (0.320)	0.069 (0.299)
PC6	1.937 (8.388)	1.671 (7.257)	2.220 (9.578)	1.984 (8.534)
PC7	3.026 (12.929)	2.817 (12.048)	2.776 (11.885)	3.009 (12.860)
PC8	2.798 (11.448)	2.873 (11.710)	2.557 (10.404)	2.280 (9.329)
PC9	4.963 (20.451)	5.137 (21.193)	5.246 (21.640)	5.777 (23.848)
PC10	0.118 (0.517)	0.123 (0.532)	0.107 (0.461)	0.100 (0.430)
Constant	1.193 (0.137)	1.193 (0.138)	1.192 (0.138)	1.193 (0.138)
Observations	3,369	3,369	3,369	3,369

General note for Tables S4.2–S4.11: Each table reports full regression coefficients, standard errors (in parentheses), and sample sizes (Observations). For count models coefficients are exponentiated and presented as IRRs. Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table S4.3 Full Regression Estimates for Major Depressive Disorder Among Males

Major depressive disorder	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	2.712*** (0.405)	2.683*** (0.402)	2.731*** (0.407)	2.730*** (0.407)
COVID	0.846 (0.114)	0.846 (0.114)	0.846 (0.114)	0.846 (0.114)
3rd Wave	1.114* (0.073)	1.210*** (0.080)	1.023 (0.067)	1.040 (0.074)
PC1	0.873 (5.079)	0.678 (3.888)	0.993 (5.790)	1.014 (5.907)
PC2	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)
PC3	28.836 (162.802)	34.786 (195.346)	27.674 (157.725)	23.659 (134.049)
PC4	20.758 (116.830)	18.459 (102.885)	40.331 (226.867)	32.266 (183.402)
PC5	0.127 (0.712)	0.117 (0.658)	0.116 (0.651)	0.104 (0.584)
PC6	27,505.894* (149,222.783)	11,119.460* (60,956.689)	37,086.295* (202,501.729)	41,184.188* (223,879.694)
PC7	84.560 (456.513)	54.253 (296.175)	104.646 (566.128)	98.457 (533.393)
PC8	16.013 (79.045)	18.166 (89.699)	23.893 (116.778)	29.113 (143.031)
PC9	1.393 (7.738)	2.128 (11.824)	1.076 (5.970)	0.997 (5.495)
PC10	0.018 (0.096)	0.015 (0.079)	0.011 (0.061)	0.012 (0.063)
Constant	0.051*** (0.007)	0.051*** (0.007)	0.051*** (0.007)	0.051*** (0.007)
Observations	3,369	3,369	3,369	3,369

Table S4.4 Full Regression Estimates for Loneliness Among Males

Loneliness	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	1.035*** (0.011)	1.035*** (0.011)	1.035*** (0.011)	1.035*** (0.011)
COVID	0.989 (0.008)	0.989 (0.008)	0.989 (0.008)	0.989 (0.008)
3rd Wave	0.991 (0.009)	0.986 (0.009)	0.984 (0.010)	0.997 (0.010)
PC1	0.691 (0.522)	0.712 (0.534)	0.731 (0.552)	0.685 (0.518)
PC2	0.874 (0.718)	0.844 (0.691)	0.878 (0.720)	0.884 (0.726)
PC3	0.866 (0.696)	0.844 (0.678)	0.817 (0.656)	0.863 (0.693)
PC4	0.572 (0.422)	0.583 (0.431)	0.537 (0.398)	0.555 (0.409)
PC5	2.985 (2.402)	3.007 (2.419)	2.957 (2.359)	3.002 (2.421)
PC6	1.164 (1.041)	1.229 (1.111)	1.218 (1.092)	1.118 (1.002)
PC7	0.800 (0.678)	0.832 (0.706)	0.767 (0.652)	0.797 (0.675)
PC8	3.106 (2.549)	3.066 (2.527)	2.996 (2.463)	2.952 (2.437)
PC9	0.167** (0.131)	0.162** (0.126)	0.171** (0.134)	0.171** (0.134)
PC10	1.103 (0.906)	1.122 (0.921)	1.095 (0.899)	1.137 (0.935)
Constant				
Observations	3,369	3,369	3,369	3,369

Table S4.5 Full Regression Estimates for Quality of Life Among Males

Quality of life	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	0.972*** (0.004)	0.972*** (0.004)	0.972*** (0.004)	0.972*** (0.004)
COVID	1.015*** (0.003)	1.015*** (0.003)	1.015*** (0.003)	1.015*** (0.003)
3rd Wave	1.001 (0.004)	1.002 (0.004)	0.997 (0.004)	1.002 (0.004)
PC1	1.473 (0.428)	1.473 (0.427)	1.505 (0.438)	1.472 (0.427)
PC2	1.114 (0.381)	1.118 (0.382)	1.114 (0.381)	1.115 (0.381)
PC3	0.923 (0.307)	0.924 (0.307)	0.911 (0.303)	0.918 (0.304)
PC4	1.718* (0.553)	1.720* (0.554)	1.726* (0.554)	1.712* (0.550)
PC5	1.263 (0.437)	1.264 (0.437)	1.251 (0.434)	1.256 (0.436)
PC6	0.815 (0.251)	0.810 (0.251)	0.832 (0.259)	0.819 (0.252)
PC7	0.674 (0.220)	0.673 (0.220)	0.673 (0.219)	0.674 (0.220)
PC8	1.170 (0.390)	1.175 (0.391)	1.180 (0.393)	1.188 (0.396)
PC9	1.408 (0.431)	1.410 (0.431)	1.407 (0.430)	1.392 (0.427)
PC10	1.124 (0.308)	1.120 (0.307)	1.109 (0.304)	1.124 (0.308)
Constant				
Observations	3,369	3,369	3,369	3,369

Table S4.6 Full Regression Estimates for Anxiety Among Males

Anxiety	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x Nov-Dec 2020	0.993 (0.010)	0.981** (0.009)	0.981** (0.009)	0.992 (0.009)
Nov-Dec 2020	1.018* (0.010)	1.018* (0.009)	1.017* (0.009)	1.017* (0.010)
PC1	1.216 (0.823)	1.264 (0.851)	1.341 (0.905)	1.225 (0.828)
PC2	0.449 (0.367)	0.426 (0.347)	0.459 (0.375)	0.450 (0.367)
PC3	1.329 (1.016)	1.271 (0.967)	1.244 (0.946)	1.361 (1.040)
PC4	0.192** (0.158)	0.200** (0.164)	0.178** (0.145)	0.194** (0.158)
PC5	2.760 (2.346)	2.773 (2.347)	2.745 (2.322)	2.820 (2.389)
PC6	0.279 (0.224)	0.309 (0.247)	0.301 (0.241)	0.270 (0.216)
PC7	4.181 (3.681)	4.366* (3.830)	3.919 (3.457)	4.132 (3.636)
PC8	1.915 (1.441)	1.903 (1.422)	1.847 (1.376)	1.787 (1.337)
PC9	1.905 (1.352)	1.793 (1.271)	1.982 (1.404)	1.992 (1.417)
PC10	1.585 (1.158)	1.574 (1.153)	1.583 (1.162)	1.615 (1.188)
Constant				
Observations	2,246	2,246	2,246	2,246

Table S4.7 Full Regression Estimates for Depressive Symptoms Among Females

Depressive symptoms	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	0.982 (0.038)	0.947 -0.036	1.011 -0.04	0.972 (0.038)
COVID	1.562*** (0.062)	1.566*** (0.063)	1.561*** (0.062)	1.561*** (0.062)
3rd Wave	0.971 (0.051)	0.970 (0.051)	0.970 (0.051)	0.971 (0.051)
PC1	3.092 (9.969)	3.230 (10.432)	2.842 (9.237)	2.889 (9.332)
PC2	0.636 (2.049)	0.520 (1.678)	0.693 (2.231)	0.649 (2.086)
PC3	1.633 (5.121)	1.288 (4.003)	1.850 (5.758)	1.899 (5.919)
PC4	0.283 (0.890)	0.341 (1.075)	0.289 (0.908)	0.295 (0.929)
PC5	0.021 (0.070)	0.021 (0.070)	0.019 (0.063)	0.022 (0.072)
PC6	0.876 (2.851)	0.887 (2.891)	0.897 (2.926)	0.937 (3.052)
PC7	3.310 (10.392)	4.356 (13.686)	3.188 (10.062)	3.002 (9.439)
PC8	0.974 (3.172)	0.920 (2.999)	1.022 (3.329)	0.898 (2.929)
PC9	3.696 (11.808)	3.384 (10.837)	3.638 (11.642)	3.672 (11.749)
PC10	312.858* (995.478)	304.847* (971.491)	310.143* (988.657)	268.087* (857.873)
Constant	1.640*** (0.156)	1.648*** (0.158)	1.638*** (0.156)	1.640*** (0.156)
Observations	4,113	4,113	4,113	4,113

Table S4.8 Full Regression Estimates for Major Depressive Disorder Among Females

Major depressive disorder	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	1.111** (0.046)	1.104** (0.046)	0.982 (0.042)	1.024 (0.042)
COVID	2.275*** (0.204)	2.274*** (0.204)	2.284*** (0.205)	2.284*** (0.205)
3rd Wave	0.874 (0.080)	0.874 (0.081)	0.874 (0.081)	0.874 (0.081)
PC1	0.035 (0.128)	0.039 (0.143)	0.060 (0.225)	0.053 (0.197)
PC2	0.094 (0.325)	0.150 (0.523)	0.056 (0.193)	0.066 (0.226)
PC3	199.969 (729.125)	227.036 (825.182)	129.901 (475.523)	132.769 (485.583)
PC4	10.116 (34.277)	8.163 (27.768)	8.993 (30.411)	8.790 (29.932)
PC5	1.097 (3.649)	1.098 (3.668)	1.408 (4.677)	1.240 (4.158)
PC6	0.521 (2.057)	0.457 (1.802)	0.469 (1.867)	0.419 (1.663)
PC7	101.528 (336.251)	85.880 (284.958)	112.504 (374.412)	122.260 (405.552)
PC8	13.897 (42.579)	14.575 (45.243)	10.306 (31.759)	11.366 (34.983)
PC9	20.875 (72.104)	31.141 (107.468)	27.446 (94.843)	27.384 (94.530)
PC10	7.627 (26.889)	8.429 (29.609)	8.293 (29.289)	9.245 (32.628)
Constant	0.113*** (0.009)	0.113*** (0.009)	0.113*** (0.009)	0.113*** (0.009)
Observations	4,113	4,113	4,113	4,113

Table S4.9 Full Regression Estimates for Loneliness Among Females

Loneliness	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	0.996 (0.010)	0.996 (0.010)	0.99 (0.010)	1.003 (0.010)
COVID	1.085*** (0.011)	1.085*** (0.011)	1.085*** (0.011)	1.085*** (0.011)
3rd Wave	0.993 (0.008)	0.993 (0.008)	0.993 (0.008)	0.993 (0.008)
PC1	0.623 (0.536)	0.623 (0.535)	0.659 (0.560)	0.616 (0.531)
PC2	1.100 (0.879)	1.080 (0.858)	1.072 (0.858)	1.124 (0.901)
PC3	1.852 (1.598)	1.838 (1.576)	1.831 (1.582)	1.871 (1.603)
PC4	1.488 (1.124)	1.502 (1.134)	1.419 (1.075)	1.467 (1.107)
PC5	0.714 (0.570)	0.714 (0.570)	0.736 (0.588)	0.704 (0.563)
PC6	2.429 (2.126)	2.447 (2.140)	2.536 (2.218)	2.445 (2.136)
PC7	0.276 (0.231)	0.279 (0.234)	0.259 (0.216)	0.275 (0.230)
PC8	0.783 (0.715)	0.785 (0.715)	0.798 (0.731)	0.804 (0.738)
PC9	4.530* (3.900)	4.454* (3.839)	4.428* (3.813)	4.466* (3.859)
PC10	4.096* (3.434)	4.071* (3.407)	4.010* (3.355)	4.119* (3.457)
Constant				
Observations	4,113	4,113	4,113	4,113

Table S4.10 Full Regression Estimates for Quality of Life Among Females

Quality of life	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x COVID	0.998 (0.004)	0.999 (0.004)	0.999 (0.004)	1.002 (0.004)
COVID	0.952*** (0.004)	0.952*** (0.004)	0.952*** (0.004)	0.952*** (0.004)
3rd Wave	1.014*** (0.003)	1.014*** (0.003)	1.014*** (0.003)	1.014*** (0.003)
PC1	1.108 (0.387)	1.100 (0.383)	1.151 (0.402)	1.098 (0.382)
PC2	1.010 (0.305)	1.012 (0.307)	0.986 (0.300)	1.026 (0.309)
PC3	0.626 (0.190)	0.629 (0.192)	0.623 (0.188)	0.631 (0.190)
PC4	1.283 (0.377)	1.285 (0.377)	1.252 (0.368)	1.274 (0.376)
PC5	1.482 (0.449)	1.480 (0.449)	1.501 (0.453)	1.473 (0.447)
PC6	1.233 (0.395)	1.235 (0.395)	1.267 (0.407)	1.233 (0.395)
PC7	1.418 (0.422)	1.418 (0.423)	1.366 (0.404)	1.411 (0.419)
PC8	0.996 (0.327)	0.998 (0.328)	0.993 (0.324)	1.008 (0.330)
PC9	0.724 (0.228)	0.720 (0.227)	0.716 (0.225)	0.719 (0.226)
PC10	0.682 (0.235)	0.681 (0.234)	0.677 (0.233)	0.685 (0.236)
Constant				
Observations	4,113	4,113	4,113	4,113

Table S4.11 Full Regression Estimates for Anxiety Among Females

Anxiety	PGS DS	PGS MDD	PGS AD	PGS Loneliness
PGS x Nov-Dec 2020	0.984* (0.010)	0.980** (0.009)	0.975** (0.010)	1.015 (0.010)
Nov-Dec 2020	1.035*** (0.010)	1.035*** (0.010)	1.035*** (0.010)	1.034*** (0.010)
PC1	4.522* (3.493)	4.504* (3.466)	5.136** (3.957)	4.329* (3.318)
PC2	1.117 (0.907)	0.998 (0.815)	1.059 (0.861)	1.211 (0.984)
PC3	0.590 (0.461)	0.576 (0.447)	0.619 (0.481)	0.625 (0.485)
PC4	0.271 (0.224)	0.282 (0.232)	0.240* (0.199)	0.253* (0.209)
PC5	1.666 (1.337)	1.679 (1.345)	1.749 (1.398)	1.567 (1.259)
PC6	1.199 (0.991)	1.250 (1.036)	1.358 (1.122)	1.209 (0.999)
PC7	0.628 (0.533)	0.660 (0.556)	0.535 (0.452)	0.620 (0.522)
PC8	1.369 (1.142)	1.344 (1.113)	1.428 (1.186)	1.478 (1.234)
PC9	1.075 (0.855)	1.011 (0.801)	1.012 (0.801)	1.020 (0.807)
PC10	0.258* (0.207)	0.253* (0.202)	0.240* (0.192)	0.267* (0.213)
Constant				
Observations	2,742	2,742	2,742	2,742

Table S4.12 Approximate Effect Sizes (ΔR^2 and f^2) for Gene $\times$ COVID Interactions on
Mental Health Outcomes

		Females		Males	
		ΔR^2	Cohen's f^2	ΔR^2	Cohen's f^2
Depressive symptoms	PGS DS x COVID	5.88E-07	6.20E-07	5.88E-07	6.20E-07
	PGS MDD x COVID	5.88E-07	6.20E-07	5.88E-07	6.20E-07
	PGS AD x COVID	1.83E-06	1.93E-06	0.0000398 7	0.0000416
	PGS Loneliness x COVID	0.0001047 7	0.0001104 5	5.66E-08	5.90E-08
Major depressive disorder	PGS DS x COVID	3.39E-05	3.54E-05	0.0002254 1	0.0002328 6
	PGS MDD x COVID	0.0001337 3	0.0001398 1	0.0016723 4	0.0017302
	PGS AD x COVID	0.0002708 6	0.0002832 2	5.45E-05	5.63E-05
	PGS Loneliness x COVID	0.0003376 2	0.0003530 4	0.0003376 2	0.0003530 4
Loneliness	PGS DS x COVID	0.0000384 2	0.0000399	0.0004733 2	0.0004789 1
	PGS MDD x COVID	0.0000551 2	0.0000572 5	0.0012093 1	0.0012245
	PGS AD x COVID	0.0004902 1	0.0005093 5	0.0016899 5	0.0017120 2
	PGS Loneliness x COVID	8.04E-05	8.35E-05	5.19E-05	5.24E-05
Quality of life	PGS DS x COVID	0.0001193 6	0.0001291 1	9.69E-05	0.0001001 4

		Females		Males	
		ΔR^2	Cohen's f^2	ΔR^2	Cohen's f^2
	PGS MDD x COVID	4.64E-06	5.01E-06	0.0001160	0.0001200
				7	2
	PGS AD x COVID	0.0013554	0.0014681	0.0003902	0.0004036
		7	7	8	7
	PGS Loneliness x COVID	0.0001323	0.0001431	0.0001890	0.0001955
Anxiety	PGS DS x COVID	0	0	0	0
				0.0001890	0.0001955
	PGS MDD x COVID	0	0	9	4
	PGS AD x COVID	0	0	0	0
	PGS Loneliness x COVID	0	0	0	0

Note: Effect-size table reports ΔR^2 and Cohen's f^2 computed from linear FE approximations.

Chapter 5

Conclusion

Summary of Findings

This thesis examined how genetic risks interact with social and environmental exposures to shape cognitive function and mental health in later life, drawing on data from the English Longitudinal Study of Ageing (ELSA). Across three empirical studies employing distinct methodologies, the findings converge on a central argument: ageing trajectories are neither uniform nor biologically predetermined, but emerge from the dynamic interplay between individual genetic endowments and the structural, social, and historical conditions of the life course.

The first empirical study applied group-based trajectory modelling to characterise heterogeneity in cognitive and mental health pathways among older adults in England from 2008 to 2018. Three distinct memory trajectories were identified—"low and decreasing" (17.6%), "stable medium" (51.4%), and "stable high" (31.0%)—alongside five depressive symptom trajectories: "stable low" (27.5%), "increasing" (13.9%), "stable medium" (34.6%), "decreasing" (12.2%), and "stable high" (11.8%).

Socioeconomic gradients were pronounced throughout. Higher educational attainment was associated with a 13.8 percentage point reduction in the probability of belonging to the most adverse memory trajectory and a 21.5 percentage point increase in the probability of maintaining the most stable cognitive profile. Early-life socioeconomic conditions, proxied by father's occupational status at age 14, independently predicted late-life trajectory membership—children of high-skilled fathers were 10.1 percentage points more likely to maintain high cognitive stability—underscoring the long reach of

childhood disadvantage into late-life health. Women were more likely than men to follow the stable high memory trajectory, but also more likely to experience elevated and persistent depressive symptoms, reflecting well-documented sex differences in both cognitive and mental health ageing. Contemporaneous analyses within trajectory groups further showed that wealth, employment, and self-rated health all predicted within-person deviations from expected cognitive and mental health scores, with effects concentrated among those already following adverse trajectories—suggesting that socioeconomic resources provide critical protection precisely for those most biologically vulnerable.

The second empirical study exploited the 1947 compulsory schooling reform in England as a natural experiment to estimate the causal impact of education on late-life cognition and how this effect is moderated by genetic risk. Using a fuzzy regression discontinuity design, the analysis found that an additional year of compulsory schooling improved memory scores among less-educated males by approximately 0.43–0.45 standard deviations, consistent with prior evidence from the same reform (Banks & Mazzonna, 2012). The novel contribution, however, was the identification of a crossover interaction between APOE4 status and educational attainment among females for executive function. Among female APOE4 carriers, those who left school at or before age 14 showed markedly higher executive function than non-carriers (margin = +0.989), while those exposed to an additional year of compulsory schooling scored substantially lower (margin = -1.560)—a within-carrier difference of approximately 2.5 units (1.7 SD) on the composite executive function score. This crossover pattern is consistent with the differential susceptibility framework, in which the APOE4 allele confers heightened environmental sensitivity rather than simply increasing risk.

The third empirical study examined the mental health consequences of the COVID-19 pandemic using a two-way fixed-effects design. The pandemic was associated with broadly adverse mental health outcomes for both men and women: depression, loneliness, and anxiety increased, while quality of life declined. Women experienced larger main effects—a 56.2% higher expected CES-D 8 score and a 2.28-fold higher probability of major depressive disorder post-pandemic. The key G×E finding concerned men: a higher polygenic score for major depressive disorder (PGS MDD) substantially amplified the pandemic's effect on MDD risk (Post × PGS MDD: IRR = 2.683; 95% CI 1.90–3.47), with marginal analyses confirming appreciably larger absolute increases in MDD probability across higher genetic risk strata. This finding is consistent with the diathesis-stress model, under which acute environmental stressors reveal and amplify latent genetic vulnerabilities that remain masked in more stable conditions.

Taken together, these three studies demonstrate that genetic risks do not operate independently of social context, and that the patterning of health in older age reflects the accumulated interplay of biology, environment, and historical circumstance. The methodological contribution—combining polygenic scores with quasi-experimental designs within a life-course framework—advances on prior observational G×E research by reducing confounding and gene–environment correlation.

Intellectual honesty requires, however, that these contributions be situated alongside a limitation that runs through all three empirical chapters: the effect sizes are uniformly small. In Chapter 3, OLS-approximated ΔR^2 values for G×E interaction terms range from less than 0.001 to 0.006 (Cohen's $f^2 = 0.000$ –0.006), well below the conventional threshold for even a "small" effect ($f^2 = 0.02$). In Chapter 4, the most robust finding—the male PGS MDD × pandemic interaction on MDD (IRR = 2.683)—is the largest

effect observed, yet it accounts for only a negligible fraction of total variance, and most other G×E interactions are practically trivial in absolute magnitude. These small effect sizes are not a failure of design; they are consistent with the broader G×E literature and reflect the polygenic and environmentally complex nature of the outcomes (Keller, 2014; Dudbridge, 2013). But they mean that the findings should be read as proof of concept and preliminary directional evidence, not as definitive estimates on which strong policy or clinical decisions should rest without replication.

Theoretical Contributions

Beyond its empirical findings, this thesis makes substantive contributions to three intersecting theoretical traditions: the social determinants of health, the life-course perspective, and gene–environment interaction theory.

Social determinants of health. The thesis engages directly with a well-established theoretical tradition arguing that health is fundamentally shaped by social conditions rather than biology alone. From the Black Report's documentation of socioeconomic health gradients (Townsend & Davidson, 1982) through Marmot's work on the social gradient in health (Marmot et al., 1991; Marmot, 2005), to Link and Phelan's fundamental cause theory (Link & Phelan, 1995), this literature holds that socioeconomic status functions as a "fundamental cause" of health inequalities because it confers flexible resources—knowledge, money, power, and social connections—that can be deployed to reduce risk or minimise the consequences of disease, regardless of the proximate mechanisms in play. The empirical findings of Chapter 2 directly instantiate this framework: education and wealth are strong predictors of both cognitive and mental health trajectory membership, with effects persisting across time, and father's occupation at age 14 independently predicts late-life outcomes even after

controlling for adult socioeconomic position—extending the fundamental cause argument backward into early life and connecting it to the literature on the long shadow of childhood conditions (Ben-Shlomo & Kuh, 2002; Kuh & Ben-Shlomo, 2004).

Crucially, however, the thesis goes beyond documenting social gradients to ask whether genetic endowments modify them—a question that fundamental cause theory has not traditionally engaged with. The findings offer a theoretically significant complication: the same social intervention can produce heterogeneous biological returns depending on genetic profile. Among female APOE4 carriers, an additional year of compulsory schooling was associated with lower executive function, while among non-carriers and males it was neutral or beneficial. This suggests that social determinants of health operate through biologically heterogeneous pathways, and that the effectiveness of social interventions cannot be evaluated without attention to genetic variation in responsiveness. Rather than undermining the social determinants framework, this extends it: it calls for a biosocial account in which genetic susceptibility is itself socially patterned and socially expressed, and in which policies targeting social exposures may not close health inequalities uniformly across genetically diverse populations.

Life-course theory. This thesis situates its analyses within Elder's life-course paradigm (Elder et al., 2003) and the life course cube (Bernardi et al., 2019), which organises individual trajectories along three intersecting dimensions: historical time, institutional contexts, and linked lives. Each empirical chapter can be read as an application of this framework. Chapter 2 documents the cumulative effects of socioeconomic exposures across time, showing how early-life disadvantage and adult resources combine to shape late-life trajectories—consistent with the life-course principles of cumulative advantage and linked lives. Chapter 3 exploits a specific historical turning point—the 1947 compulsory schooling reform—as an exogenous shock to the institutional context of

education, demonstrating that macro-level institutional change can have lasting biological consequences decades later. Chapter 4 uses the COVID-19 pandemic as a further historical shock, showing how acute disruption to the social environment amplifies latent genetic vulnerabilities in sex-differentiated and outcome-specific ways.

The theoretical contribution is to demonstrate that genetic risks are not life-course constants but life-course variables: their expression depends on the historical moment, the institutional environment, and the social resources available to individuals at critical junctures. This moves beyond static $G \times E$ models toward a dynamic, life-course conception of gene–environment interplay—in which the question is not simply "does environment moderate genetic risk?" but "when, for whom, and under what institutional conditions does this moderation occur?"

Gene–environment interaction theory. The $G \times E$ literature has been shaped by two competing theoretical frameworks. The diathesis-stress model (Monroe & Simons, 1991) proposes that genetic variants increase vulnerability to adverse environmental exposures, such that high-risk genotypes produce worse outcomes under stress while genetic differences are masked in benign environments. The differential susceptibility model (Belsky & Pluess, 2009) proposes instead that certain genetic variants confer heightened sensitivity to environments in both directions—functioning as plasticity alleles rather than simple risk alleles, making carriers both more harmed by adversity and more benefited by positive conditions. These frameworks make different predictions about interaction shape and carry different implications for intervention.

The findings of this thesis provide differentiated support for each. The COVID-19 results—where PGS MDD amplified pandemic-related MDD risk among men—are most consistent with the diathesis-stress model: genetic predisposition for depression is

expressed most strongly under acute stress. The APOE4 $\times$ education crossover interaction in Chapter 3—where APOE4 carriers outperform non-carriers at low education but underperform at higher education—is most consistent with the differential susceptibility model: APOE4 appears to confer sensitivity to both adverse and beneficial environmental conditions. Together, the findings suggest that different genetic variants may operate through different mechanisms, and that a single unified framework is insufficient to capture the range of G $\times$ E patterns observable in polygenic, population-level data.

The thesis also directly engages with a central methodological critique of the G $\times$ E literature. Keller (2014) argued that most G $\times$ E studies have failed to adequately control for gene–environment correlation (rGE): if a genetic variant predicts both the environmental exposure and the outcome, apparent interactions may be artefactual confounds rather than true biological moderation. By using quasi-experimental variation—the 1947 schooling reform and the COVID-19 pandemic—this thesis substantially reduces this concern. Both shocks provide exogenous environmental variation uncorrelated by design with individual genetic profiles, allowing interaction terms to be interpreted with greater causal confidence than is possible in observational designs (Fletcher & Conley, 2013; Biroli et al., 2024). This methodological integration of the quasi-experimental toolkit with sociogenomic analysis constitutes a concrete response to Keller's critique and a model for future G $\times$ E research.

Finally, the thesis contributes to ongoing debates about the interpretive status of polygenic scores. PGSs capture additive genetic effects on phenotype means but exclude rare variants, gene–gene interactions, and variance-related genetic effects. They are better understood as probabilistic indices of average genetic liability than as measures of individual genetic destiny. The modest effect sizes observed throughout—G $\times$ E terms

explaining less than 1% of variance in most models—are consistent with the "fourth law of behavioral genetics" (Chabris et al., 2015), which holds that variation in any complex trait is associated with many genetic variants of very small effect. This positions PGS-based G×E research as a tool for detecting systematic population-level patterns of differential susceptibility, not for individual prediction or clinical application in its current form.

Strengths and Limitations

This thesis makes several methodological and substantive contributions to the emerging field of sociogenomics and the sociology of health ageing.

Methodologically, the most important contribution is the integration of polygenic scores with quasi-experimental designs. Chapter 3 combined a fuzzy regression discontinuity approach—exploiting exogenous variation in schooling induced by a national policy reform—with genetic risk measures including APOE4 status and polygenic scores for general cognition and Alzheimer's disease. This design reduces the gene–environment correlation (r_{GE}) that plagues conventional G×E analyses and provides a stronger basis for causal inference about how education moderates genetic expression. Chapter 4 similarly leveraged the COVID-19 pandemic as a near-exogenous environmental shock, using two-way fixed effects to control for time-invariant individual characteristics.

These quasi-experimental strategies represent a substantive advance over the observational designs that dominate the G×E literature, addressing longstanding methodological critiques about confounding and reverse causality (Fletcher & Conley, 2013; Biroli et al., 2024). Chapter 2 provided methodological complementarity by employing group-based trajectory modelling to characterise outcome heterogeneity

before any genetic analysis, ensuring that the social environmental landscape was well characterised.

Substantively, the thesis generates three novel contributions. First, it demonstrates that the protective effect of education on cognitive ageing is not genetically uniform: for female APOE4 carriers, an additional year of compulsory schooling was associated with lower executive function, suggesting that social advantage can, under certain genetic conditions, produce unexpected costs rather than uniform benefits. This challenges simplistic policy narratives about education as an unambiguously protective factor and enriches theoretical frameworks such as differential susceptibility (Belsky & Pluess, 2009). Second, the thesis shows that environmental shocks such as COVID-19 can activate latent genetic vulnerabilities in ways that are sex-differentiated and outcome-specific, with the clearest signal appearing for major depressive disorder among men—a result consistent with diathesis-stress theory and with the broader literature linking PGS MDD to stress reactivity (Colodro-Conde et al., 2018; Arnau-Soler et al., 2019). Third, Chapter 2 documents the persistence of early-life socioeconomic influences on late-life cognitive and mental health trajectories, extending the life-course paradigm (Elder et al., 2003; Ben-Shlomo & Kuh, 2002) by showing that father's occupation at age 14 independently predicts cognitive trajectory membership decades later, even after controlling for adult education and wealth.

Theoretically, the thesis broadens the scope of sociogenomics by embedding genetic analysis within a life-course framework. By drawing on Elder's life-course paradigm and the life course cube (Bernardi et al., 2019), genetic risks are situated within historical time, institutional settings, and linked lives, rather than treated as decontextualised biological constants. The COVID-19 pandemic analyses in particular illustrate how a shared historical event generates heterogeneous biological responses depending on

genetic predisposition—an illustration of the "historical time and place" principle of life-course theory operating at the genomic level.

Several important limitations must be acknowledged, and given the examiners' request for transparency, the magnitude of effect sizes warrants particular attention. The $G \times E$ interaction terms across all three empirical chapters explain only a very small fraction of variance in outcomes. In Chapter 3, OLS-approximated ΔR^2 values for interaction terms ranged from less than 0.001 to 0.006 (Cohen's $f^2 = 0.000\text{--}0.006$), well below the conventional "small" effect threshold of $f^2 = 0.02$. In Chapter 4, similar patterns obtained, with the notable exception of the male PGS MDD $\times$ Post interaction on MDD probability, which showed the largest—but still modest—effect in relative terms (IRR = 2.683). These small effect sizes are consistent with the broader $G \times E$ literature (Keller, 2014; Dudbridge, 2013) and reflect the polygenic nature of the outcomes under study: no single genetic variant or score can be expected to explain large proportions of variance in complex behaviours and health outcomes. Nevertheless, this means the findings should be regarded as providing preliminary, directional evidence rather than definitive causal estimates, and replication in larger and more diverse samples is essential before strong conclusions are warranted.

Statistical power is a related and serious constraint. The genotyped ELSA subsample, while appropriate for the analytical designs employed, limits the precision of interaction estimates, particularly when stratified by sex. The sample's non-random composition—genotyped participants are on average older, more educated, and healthier than non-genotyped ELSA members (as documented in Table 3.4)—introduces survivorship and participation bias that inverse probability weighting can only partially address. The ELSA sample also draws exclusively from European-ancestry participants, both because of the sample's demographic composition and because the polygenic scores were

constructed from European-ancestry GWAS. This restricts generalisability to non-European populations and precludes examination of how genetic associations may differ across ancestry groups (Gharbi-Meliani et al., 2021).

The polygenic scores themselves carry inherent limitations. Constructed from GWAS of additive effects on phenotype means, the PGSs used in this thesis capture common variant contributions to average outcomes but do not detect rare variants, gene–gene interactions, gene–environment correlations, or variance-related genetic effects. The emerging approach of variance polygenic scores (vPGS), which specifically target loci associated with phenotypic plasticity, offers a promising extension that this thesis was not in a position to implement (Johnson et al., 2021). Collider bias is also a concern in both Chapters 3 and 4: conditioning on education (Chapter 3) or on being a pandemic survivor with valid mental health data (Chapter 4) can induce spurious associations among variables that share common causes (Akimova et al., 2022). While the quasi-experimental designs and sensitivity analyses provide some robustness against these concerns, they cannot be fully eliminated.

A further limitation concerns the temporal scope of the analyses. Chapter 3 examines a specific historical reform cohort (born 1928–1938) whose educational experiences may not generalise to contemporary contexts where the content, quality, and social returns to schooling are markedly different. Chapter 4 uses data from only three waves spanning 2018–2020, capturing the acute pandemic period but unable to assess longer-term mental health trajectories as the pandemic receded or as vaccination and social restrictions evolved.

Finally, while the quasi-experimental designs employed in Chapters 3 and 4 substantially reduce confounding relative to conventional observational studies, the

findings of Chapter 4 in particular should be interpreted as associations consistent with genetic moderation of pandemic-related distress rather than as evidence of strictly causal pathways, given the absence of an unexposed control group and the possibility of time-varying confounding concurrent with the pandemic onset.

Policy Implications

Despite their modest effect sizes, the findings of this thesis carry meaningful implications for social and health policy, and for how policymakers think about the relationship between education, societal shocks, and health inequality.

The results from Chapter 2 and Chapter 3 together make a strong case for educational investment as a lever for reducing health inequalities in later life. The trajectory analyses in Chapter 2 show that higher education is associated with a 21.5 percentage point increase in the probability of maintaining the most favourable cognitive trajectory and a corresponding reduction in adverse outcomes, across both cognitive and mental health domains. These associations are complemented by the causal estimates in Chapter 3, which provide quasi-experimental evidence that an additional year of compulsory schooling improves memory outcomes among less-educated males. Policies aimed at raising educational quality and attainment in disadvantaged communities—or expanding access to adult education for those who received limited initial schooling—could yield downstream benefits for cognitive ageing at the population level, with the potential to reduce inequalities that are currently stratified by early-life socioeconomic conditions.

However, the $G \times E$ findings from Chapter 3 introduce an important qualification.

Among female APOE4 carriers, an additional year of compulsory schooling was associated with lower executive function, suggesting that the benefits of education are not genetically uniform and that increased schooling can, under specific genetic

conditions, have unexpected consequences. This does not argue against educational investment—the population-level benefits remain clear—but it does caution against assuming that universal educational interventions will produce uniformly beneficial outcomes. As genetic screening becomes increasingly feasible, targeted educational support or tailored pedagogical approaches for individuals at elevated genetic risk could be considered, though this must be approached with significant ethical caution to avoid genetic stigmatisation.

The Chapter 4 findings have direct relevance for mental health policy and pandemic preparedness. The demonstration that men with higher genetic liability for depression experienced substantially amplified risk of MDD following the onset of COVID-19 suggests that mental health crisis responses should incorporate genetic and psychosocial vulnerability into their design. More practically, this implies that mental health services during periods of societal stress should not rely solely on population-level interventions but should actively identify and reach out to groups with known risk factors—whether biological, socioeconomic, or social—who may face disproportionate burdens during crises. The sex-specific pattern of effects (larger main effects for women; stronger genetic amplification for men) further suggests that mental health responses need to be sensitive to differential pathways of vulnerability rather than adopting a one-size-fits-all model.

More broadly, the consistent finding of socioeconomic gradients in trajectory membership, combined with the gene–environment interactions documented in Chapters 3 and 4, points to a policy conclusion that bridges structural and biological perspectives: reducing socioeconomic inequalities in early life—in parental occupation, educational access, and economic resources—is likely to reduce the conditions under which genetic vulnerabilities are most powerfully expressed. Gene–environment interactions do not

operate in a social vacuum; they are embedded in the structural inequalities that sociology has long documented.

Generalisability Beyond the United Kingdom

All three empirical chapters use ELSA data, and careful consideration of generalisability is warranted before findings are applied to other national contexts.

The compulsory schooling reform exploited in Chapter 3 is specific to the English historical context of 1947, and its cognitive effects may not translate directly to other educational systems or cohorts. Studies exploiting similar reforms in continental Europe have produced mixed results (Schneeweis et al., 2014; Crespo et al., 2014), suggesting that the cognitive returns to an additional year of schooling depend on the broader institutional environment—including school quality, occupational returns to education, and the social meaning of additional credentials within a specific historical moment. The $APOE4 \times$ education interaction identified in this thesis has not been consistently replicated internationally, and the sex-specific pattern may reflect features of post-war English social structure that shaped how women experienced additional compulsory education. The historical context—immediate post-war England, where women's labour market participation and occupational complexity were constrained—is unlikely to generalise directly to contemporary educational systems or other national contexts where women's education translates more directly into occupational attainment.

The COVID-19 findings in Chapter 4 may be more internationally comparable, as the pandemic constituted a near-universal environmental shock, but the severity of mental health effects varied substantially across countries depending on the stringency of lockdown measures, social safety net provision, housing conditions, and pre-existing population health. The UK experienced some of the longest and most stringent

lockdowns in Europe, which may have amplified the mental health impacts observed here relative to less restrictive contexts. Moreover, the findings concern older adults (aged 50+) in England: younger populations, or those in contexts with stronger extended family support networks, may show different patterns of G×E interaction.

Future cross-national comparative analyses using harmonised genetic and mental health data—such as those available through the SHARE network, the Health and Retirement Study (HRS), and other coordinated ageing studies—would substantially strengthen the evidence base. Such work would enable direct tests of whether the G×E effects documented here are specific to the UK context or reflect more universal patterns of gene–environment interplay in ageing.

Directions for Future Research

Building on these findings, several avenues for future work emerge as priorities.

First, replication in larger and more diverse samples is essential. The effect sizes documented across all three empirical chapters are modest, and the genotyped ELSA subsample is demographically constrained. Collaborative, cross-cohort studies—drawing on the HRS, SHARE, and UK Biobank, among others—would increase statistical power, enable sex-stratified analyses with adequate precision, and allow tests of whether the APOE4 × education interaction for executive function in women replicates in independent samples.

Second, incorporating multi-ancestry GWAS and diverse population cohorts would address the portability problem of polygenic scores. The PGSs used in this thesis were constructed from European-ancestry samples, limiting their predictive validity in non-European populations (Duncan et al., 2019). Future G×E studies should prioritise

ancestry-diverse samples and use PGSs derived from multi-ancestry GWAS to ensure that findings about gene–environment interplay are not systematically biased by ancestral homogeneity.

Third, the integration of variance polygenic scores (vPGS) and related methodological innovations offers a promising extension of this work. Conventional PGSs, constructed from GWAS of mean effects, cannot detect genetic effects on phenotypic plasticity or differential sensitivity. vPGS, which capture loci associated with variance in outcomes rather than means, could provide a more precise instrument for studying differential susceptibility (Johnson et al., 2021). Applying vPGS within the quasi-experimental frameworks used in this thesis would offer a direct test of the differential susceptibility model that goes beyond the indirect evidence currently available.

Fourth, mechanistic exploration is needed. The statistical interactions documented in this thesis identify the existence of $G \times E$ effects but cannot characterise the biological or psychosocial pathways through which they operate. Future research should investigate mechanisms linking genetic variants to cognitive and mental health outcomes—including neuroinflammatory processes, synaptic plasticity, stress-axis regulation, and epigenetic modification. Epigenetic data in particular would allow examination of how social conditions—education, socioeconomic stress, pandemic exposure—become biologically embedded across the life course, potentially mediating the $G \times E$ effects documented here (Goossens et al., 2015).

Fifth, policy evaluation studies are needed to translate these findings into actionable knowledge. Could adult literacy and numeracy programmes in mid-life buffer genetic risks for cognitive decline? How might mental health services be redesigned for genetically vulnerable groups during crises such as COVID-19? Natural experiments

evaluating such interventions—analogous to the schooling reform design used in Chapter 3—would provide the causal evidence needed to guide policy, moving beyond the characterisation of G×E effects toward their modification.

Finally, interdisciplinary work on the ethical, social, and policy implications of sociogenomic research is urgently needed. Findings that link genetic variants to health trajectories risk being misinterpreted in deterministic or stigmatising ways if divorced from their social context. Future work should engage not only with methodological rigour but also with the frameworks of equity, accountability, and inclusivity that are essential for ensuring that advances in sociogenomics contribute to health justice rather than to new forms of biological essentialism (Dolan et al., 2024; Halley et al., 2024).

Concluding Remarks

This thesis has argued that health in older age is the outcome of cumulative, dynamic interactions between genes and environments, unfolding across historical time, institutional contexts, and individual life trajectories. By combining polygenic scores with quasi-experimental designs and embedding both within a life-course framework, it has demonstrated that ageing outcomes are neither biologically fixed nor socially determined, but emerge from their ongoing interplay.

The findings underscore three points of practical and theoretical significance. First, socioeconomic inequalities in early life—in education, parental occupation, and economic resources—cast long shadows over cognitive and mental health trajectories in later life, and these inequalities structure the conditions under which genetic vulnerabilities are expressed. Second, the benefits of social interventions such as education are not genetically uniform: the same environmental exposure can have protective effects for some individuals and adverse effects for others, depending on their

genetic profile. Third, acute societal shocks reveal and amplify latent genetic predispositions, producing health inequalities that have both structural and biological dimensions.

These findings, while modest in effect size and preliminary in nature, point toward a more nuanced understanding of healthy ageing than either purely biological or purely social accounts can provide. Policy responses to cognitive decline and mental health deterioration must therefore address both structural inequalities and biological heterogeneity—not as competing priorities, but as complementary dimensions of a shared challenge. Investing in education, reducing socioeconomic disadvantage, and strengthening mental health infrastructure are not alternatives to understanding genetic risk; they are the structural conditions through which genetic risks are most consequentially expressed, and through which they can most effectively be mitigated.

Ultimately, the contribution of this thesis lies in bridging disciplinary boundaries—between sociology, genomics, and epidemiology—to demonstrate that a just and comprehensive understanding of ageing requires sustained attention to both structure and biology, and to the interactions between them.

References

- Adams, M. J., Streit, F., Meng, X., Awasthi, S., Adey, B. N., Choi, K. W., & Jonsson, L. (2025). Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell*, 188(3), 640 – 652.
- Aiken, L. S., West, S. G., & Reno, R. R. (1991). Multiple regression: Testing and interpreting interactions. Sage.
- Akimova, E. T., Breen, R., & Ermisch, J. (2022). Gene – environment interplay and collider bias in sociogenomics. *Proceedings of the National Academy of Sciences*, 119(10), e2113023119.
- Albin, R. L. (1993). Antagonistic pleiotropy, mutation accumulation, and human genetic disease. *Genetica*, 91(1), 279 – 286.
- Andreescu, C., Chang, C. C. H., Mulsant, B. H., & Ganguli, M. (2008). Twelve-year depressive symptom trajectories and their predictors in a community sample of older adults. *International Psychogeriatrics*, 20(2), 221 – 236.
- Andrews, S. J., Das, D., Anstey, K. J., & Easteal, S. (2015). Association of Alzheimer's genetic risk factors with cognitive decline: The PATH through life project. *Alzheimer's & Dementia*, 11(7), P343.

- Anstey, K., & Christensen, H. (2000). Education, activity, health, blood pressure and apolipoprotein E as predictors of cognitive change in old age: A review. *Gerontology*, 46(3), 163 – 177.
- Arenaza-Urquijo, E. M., Gonneaud, J., Fouquet, M., Perrotin, A., Mézenge, F., Landeau, B., Egret, S., De la Sayette, V., Desgranges, B., & Chételat, G. (2015). Interaction between years of education and APOE ϵ 4 status on frontal and temporal metabolism. *Neurology*, 85(16), 1392 – 1399.
- Armitage, J. M., Wang, R. A. H., Davis, O. S. P., & Haworth, C. M. A. (2022). A polygenic approach to understanding resilience to peer victimisation. *Behavior Genetics*, 52(1), 1 – 12.
- Arnaud-Soler, A., Adams, M. J., Clarke, T.-K., MacIntyre, D. J., Milburn, K., Navrady, L., Hayward, C., McIntosh, A., Thomson, P. A., & Generation Scotland. (2019). A validation of the diathesis-stress model for depression in Generation Scotland. *Translational Psychiatry*, 9, Article 25.
- Arvanitakis, Z., Wilson, R. S., Bienias, J. L., Evans, D. A., & Bennett, D. A. (2004). Diabetes mellitus and risk of Alzheimer disease and decline in cognitive function. *Archives of Neurology*, 61(5), 661 – 666.

- Assary, E., Vincent, J. P., Keers, R., & Pluess, M. (2018). Gene-environment interaction and psychiatric disorders: Review and future directions. *Seminars in Cell & Developmental Biology*, 77, 133 - 143.
- Atalay, K., Barrett, G. F., & Staneva, A. (2019). The effect of retirement on elderly cognitive functioning. *Journal of Health Economics*, 66, 37 - 53.
- Avramopoulos, D. (2018). Recent advances in the genetics of schizophrenia. *Complex Psychiatry*, 4(1), 35 - 51.
- Banks, J., Nazroo, J., & Steptoe, A. (2006). The dynamics of ageing: Evidence from the English Longitudinal Study of Ageing 2002 - 2003 (Wave 2). Institute for Fiscal Studies.
- Banks, J., & Mazzonna, F. (2012). The effect of education on old age cognitive abilities: Evidence from a regression discontinuity design. *The Economic Journal*, 122(560), 418 - 448.
- Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*, 115(42), E9765 - E9772.
- Barnes, H., & Parry, J. (2004). Renegotiating identity and relationships: Men and women's adjustments to retirement. *Ageing & Society*, 24(2), 213 - 233.

- Barr, P. B., Driver, M. N., Kuo, S. I. C., Stephenson, M., Aliev, F., Linner, R. K., Marks, J., Anokhin, A. P., Bucholz, K., Chan, G., Edenberg, H. J., Edwards, A. C., Francis, M. W., Hancock, D. B., Harden, K. P., Kamarajan, C., Kaprio, J., Kinreich, S., Kramer, J., . . . Dick, D. M. (2022). Clinical, environmental, and genetic risk factors for substance use disorders: Characterizing combined effects across multiple cohorts. *Molecular Psychiatry*, 27(11), 4633 – 4641.
- Battista, A. M., Heinsinger, N. M., & Rebeck, G. W. (2016). Alzheimer's disease genetic risk factor APOE- ϵ 4 also affects normal brain function. *Current Alzheimer Research*, 13(11), 1200 – 1207.
- Bell, R. A., Franks, P., Duberstein, P. R., Epstein, R. M., Feldman, M. D., Garcia, E. F. y, & Kravitz, R. L. (2011). Suffering in silence: Reasons for not disclosing depression in primary care. *The Annals of Family Medicine*, 9(5), 439 – 446.
- Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, 135(6), 885 – 908.
- Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, 31(2), 285 – 293.

- Ben-Shlomo, Y., Cooper, R., & Kuh, D. (2016). The last two decades of life course epidemiology, and its relevance for research on ageing. *International Journal of Epidemiology*, 45(4), 973 – 988.
- Bergdahl, E., Gustavsson, J. M. C., Kallin, K., Wågert, P. V. H., Lundman, B., Bucht, G., & Gustafson, Y. (2005). Depression among the oldest old: The Umeå 85+ study. *International Psychogeriatrics*, 17(4), 557 – 575.
- Bernardi, L., Huinink, J., & Settersten, R. A. (2019). The life course cube: A tool for studying lives. *Advances in Life Course Research*, 41, 100258. <https://doi.org/10.1016/j.alcr.2018.11.004>
- Biddinger, K. J., Jurgens, S. J., Maamari, D., Gaziano, L., Choi, S. H., Morrill, V. N., Halford, J. L., Khera, A. V., Lubitz, S. A., & Ellinor, P. T. (2022). Rare and common genetic variation underlying the risk of hypertrophic cardiomyopathy in a national biobank. *JAMA Cardiology*, 7(7), 715 – 722.
- Biroli, P., Galama, T., von Hinke, S., van Kippersluis, H., Rietveld, C. A., & Thom, K. (2024). The economics and econometrics of gene – environment interplay. *The Review of Economic Studies*, rdaf034.
- Bishop, S. L., Thurm, A., Robinson, E., & Sanders, S. J. (2021). Prevalence of returnable genetic results based on recognizable phenotypes among children with autism spectrum disorder. *medRxiv*. <https://doi.org/10.1101/2021.05.28.21257736>

- Blazer, D. G. (2003). Depression in late life: Review and commentary. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 58(3), 249 – 265.
- Boardman, J. D., Barnes, L. L., Wilson, R. S., Evans, D. A., & de Leon, C. F. M. (2012). Social disorder, APOE-E4 genotype, and change in cognitive function among older adults living in Chicago. *Social Science & Medicine*, 74(10), 1584 – 1590.
- Bogdan, R., Baranger, D. A. A., & Agrawal, A. (2018). Polygenic risk scores in clinical psychology: Bridging genomic risk to individual differences. *Annual Review of Clinical Psychology*, 14(1), 119 – 157.
- Bolano, D., & Studer, M. (2020). The link between previous life trajectories and a later life outcome: A feature selection approach (LIVES Working Paper 82). LIVES Swiss Centre of Expertise in Life Course Research. <https://doi.org/10.12682/lives.2296-1658.2020.82.2>
- Bosker, F. J., Hartman, C. A., Nolte, I. M., Prins, B. P., Terpstra, P., Posthuma, D., van Veen, T., Willemsen, G., DeRijk, R. H., de Geus, E. J., Hoogendijk, W. J., Sullivan, P. F., Penninx, B. W., Boomsma, D. I., Snieder, H., & Nolen, W. A. (2011). Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Molecular Psychiatry*, 16(5), 516 – 532.

- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169(7), 1177 – 1186.
- Butterworth, P., Schurer, S., Trinh, T.-A., Vera-Toscano, E., & Wooden, M. (2022). Effect of lockdown on mental health in Australia: Evidence from a natural experiment analysing a longitudinal probability sample survey. *The Lancet Public Health*, 7(5), e427 – e436.
- Börsch-Supan, A., & Schuth, M. (2013). Early retirement, mental health and social networks. In A. Börsch-Supan, M. Brandt, H. Litwin, & G. Weber (Eds.), *Active ageing and solidarity between generations in Europe* (pp. 337 – 348). De Gruyter.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5), 1127 – 1160.
- Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, 297(5582), 851 – 854.
- Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., & Braithwaite, A. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, 301(5631), 386 – 389.

- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1 – 22.
- Celidoni, M., Dal Bianco, C., & Weber, G. (2017). Retirement and cognitive decline: A longitudinal analysis using SHARE data. *Journal of Health Economics*, 56, 113 – 125.
- Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, 24(4), 304 – 312.
- Chang, Y. L., Zhuo, Y. Y., & Luo, D. H. (2021). Education moderates the negative effect of apolipoprotein E ϵ 4 on response inhibition in older adults. *Journal of Alzheimer's Disease*, 82(3), 1147 – 1157.
- Chaudhury, S., Brookes, K. J., Patel, T., Fallows, A., Guetta-Baranes, T., Turton, J. C., Guerreiro, R., Bras, J., Hardy, J., Francis, P. T., Croucher, R., Holmes, C., Morgan, K., & Thomas, A. J. (2019). Alzheimer's disease polygenic risk score as a predictor of conversion from mild cognitive impairment. *Translational Psychiatry*, 9(1), 154.
- Cheesman, R., Borgen, N. T., Lyngstad, T. H., Eilertsen, E. M., Ayorech, Z., Torvik, F. A., Andreassen, O. A., Zachrisson, H. D., & Ystrom, E. (2022). A population-wide gene – environment interaction

- study on how genes, schools, and residential areas shape achievement. *npj Science of Learning*, 7, 29.
- Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2018). A guide to performing polygenic risk score analyses. *bioRxiv*, 416545.
- Choi, S. W., & O'Reilly, P. F. (2020a). PRSice-2: Polygenic risk score software for biobank-scale data. *GigaScience*, 8(7), giz082.
- Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020b). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9), 2759 – 2772.
- Clare, L., Wu, Y. T., Teale, J. C., MacLeod, C., Matthews, F., Brayne, C., & Woods, B. (2017). Potentially modifiable lifestyle factors, cognitive reserve, and cognitive function in later life: A cross-sectional study. *PLOS Medicine*, 14(3), e1002259.
- Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6), 2087 – 2120.
- Clayton, E. W. (2003). Ethical, legal, and social implications of genomic medicine. *New England Journal of Medicine*, 349(6), 562 – 569.
- Clouston, S. A. P., Richards, M., Cadar, D., Hurst, L., & Joshi, H. (2015). Educational inequalities in health behaviors at midlife: Is there a role for early-life cognition? *Journal of Health and Social Behavior*, 56(3), 323 – 340.

- Clouston, S. A. P., Smith, D. M., Mukherjee, S., Zhang, Y., Hou, W., Link, B. G., & Richards, M. (2020). Education and cognitive decline: An integrative analysis of global longitudinal studies of cognitive aging. *The Journals of Gerontology: Series B*, 75(9), e151 – e160.
- Coe, N. B., von Gaudecker, H.-M., Lindeboom, M., & Maurer, J. (2012). The effect of retirement on cognitive functioning. *Health Economics*, 21(8), 913 – 927.
- Collins, A. L., Kim, Y., Sklar, P., O'Donovan, M. C., & Sullivan, P. F. (2012). Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychological Medicine*, 42(3), 607 – 616.
- Colman, I., & Ataullahjan, A. (2010). Life course perspectives on the epidemiology of depression. *The Canadian Journal of Psychiatry*, 55(10), 622 – 632.
- Colodro-Conde, L., Couvy-Duchesne, B., Zhu, G., Coventry, W. L., Byrne, E. M., Gordon, S., Wright, M. J., Montgomery, G. W., Madden, P. A. F., Ripke, S., Eaves, L. J., Heath, A. C., Wray, N. R., Medland, S. E., & Martin, N. G. (2018). A direct test of the diathesis-stress model for depression. *Molecular Psychiatry*, 23(7), 1590 – 1596.
- Conley, D. (2016a). Socio-genomic research using genome-wide molecular data. *Annual Review of Sociology*, 42, 275 – 299.

- Conley, D. (2016b). Socio-genomic research using genome-wide molecular data. *Annual Review of Sociology*, 42, 258 – 274.
- Conley, D. (2017). The challenges of GxE: Commentary on "Genetic endowments, parental resources and adult health: Evidence from the Young Finns Study." *Social Science & Medicine*, 188, 201 – 203.
- Coyle, J. T. (2003). Use it or lose it: Do effortful mental activities protect against dementia? *New England Journal of Medicine*, 348(25), 2489 – 2490.
- Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, 10(3), 131 – 138.
- Crespo, L., López-Noval, B., & Mira, P. (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review*, 43, 36 – 46.
- Dattani, S., Sham, P. C., Jermy, B. S., & Peyrot, W. J. (2023). Common and rare variant associations with latent traits underlying depression, bipolar disorder, and schizophrenia. *Translational Psychiatry*, 13, 46.
- Davies, G., Harris, S. E., Reynolds, C. A., Payton, A., Knight, H. M., Liewald, D., Lopez, L. M., Luciano, M., Gow, A. J., & Corley, J. (2014). A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing. *Molecular Psychiatry*, 19(1), 76 – 87.

- Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C. A., Kirin, M., Lahti, J., van der Lee, S. J., Le Hellard, S., Liu, T., Marioni, R. E., Oldmeadow, C., Postmus, I., Smith, A. V., Smith, J. A., Thalamuthu, A., . . . Deary, I. J. (2015). Genetic contributions to variation in general cognitive function: A meta-analysis of genome-wide association studies in the CHARGE consortium (n = 53,949). *Molecular Psychiatry*, 20(2), 183 – 192.
- Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, 29(1), 1 – 19.
- De Frias, C. M., Nilsson, L.-G., & Herlitz, A. (2006). Sex differences in cognition are stable over a 10-year period in adulthood and old age. *Aging, Neuropsychology, and Cognition*, 13(3 – 4), 574 – 587.
- De la Torre-Luque, A., de la Fuente, J., Prina, M., Sanchez-Niubo, A., Haro, J. M., & Ayuso-Mateos, J. L. (2019). Long-term trajectories of depressive symptoms in old age: Relationships with sociodemographic and health-related factors. *Journal of Affective Disorders*, 246, 329 – 337.
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, 86(1), 130 – 147.

- Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., Penke, L., Rafnsson, S. B., & Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, 92(1), 135 - 152.
- Distel, M. A., Rebollo-Mesa, I., Abdellaoui, A., Derom, C. A., Willemssen, G., Cacioppo, J. T., & Boomsma, D. I. (2010). Familial resemblance for loneliness. *Behavior Genetics*, 40(4), 480 - 494.
- Dixon, R. A., & Lachman, M. E. (2019). Risk and protective factors in cognitive aging: Advances in assessment, prevention, and promotion of alternative pathways. In R. A. Dixon, L. Bäckman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 1 - 28). Oxford University Press.
- Dolan, D. D. (2022). Three decades of ethical, legal, and social implications ("ELSI") in genomics. *Cell Genomics*, 2(6), 100162.
- Dolan, D. D., Pacia, D. M., Johnston, J., Lee, S.-S., & Cho, M. K. (2024). Expanding the agenda for a more just genomics. *Hastings Center Report*, 54(S2), S2 - S13.
- Domingue, B. W., Belsky, D. W., Harrati, A., Conley, D., Weir, D. R., & Boardman, J. D. (2017a). Mortality selection in a genetic sample and implications for association studies. *International Journal of Epidemiology*, 46(4), 1285 - 1294.

- Domingue, B. W., Liu, H., Okbay, A., & Belsky, D. W. (2017b). Genetic heterogeneity in depressive symptoms following the death of a spouse: Polygenic score analysis of the US Health and Retirement Study. *American Journal of Psychiatry*, 174(10), 963 – 970.
- Domingue, B. W., Trejo, S., Armstrong-Carter, E., & Tucker-Drob, E. M. (2020). Interactions between polygenic scores and environments: Methodological and conceptual challenges. *Sociological Science*, 7, 465 – 486.
- Donnelly, P. (2008). Progress and challenges in genome-wide association studies in humans. *Nature*, 456(7223), 728 – 731.
- Dotson, V. M., Beydoun, M. A., & Zonderman, A. B. (2010). Recurrent depressive symptoms and the incidence of dementia and mild cognitive impairment. *Neurology*, 75(1), 27 – 34.
- Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X., Liu, Y., & Mills, M. C. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences*, 117(18), 9696 – 9698.
- Duan, L., & Zhu, G. (2020). Psychological interventions for people affected by the COVID-19 epidemic. *The Lancet Psychiatry*, 7(4), 300 – 302.
- Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genetics*, 9(3), e1003348.

- Dudbridge, F., Pashayan, N., & Yang, J. (2018). Predictive accuracy of combined genetic and environmental risk scores. *Genetic Epidemiology*, 42(1), 4 – 19.
- Duncan, L., Shen, H., Gelaye, B., Meijssen, J., Ressler, K., Feldman, M., & Domingue, B. (2019). Analysis of polygenic risk score usage and performance across diverse human populations. *Nature Communications*, 10, 3328.
- Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., Goldman, D., & Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences*, 98(12), 6917 – 6922.
- Elder, G. H., Johnson, M. K., & Crosnoe, R. (2003). The emergence and development of life course theory. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the life course* (pp. 3 – 19). Springer.
- Elder, G. H. (2018). *Children of the Great Depression*. Routledge.
- Ellwood, P. (2023). Does schooling have lasting effects on cognitive function? Evidence from compulsory schooling laws. *Demography*, 60(4), 1139 – 1165.
- ELSA Project. (2023). Genetic data and polygenic scores in ELSA. UK Data Service. <https://www.elsa-project.ac.uk/genetics>

- Fichera, E., & Gathergood, J. (2019). Long-term effects of compulsory schooling on physical, mental and cognitive ageing: a natural experiment. *Journal of Health Economics*, 63, 1 – 9.
- Fletcher, J. M., & Conley, D. (2013). The challenge of causal inference in gene – environment interaction research: Leveraging research designs from the social sciences. *American Journal of Public Health*, 103(S1), S42 – S45.
- Fletcher, J. M. (2014). Enhancing the gene-environment interaction framework through a quasi-experimental research design: Evidence from differential responses to September 11. *Demography*, 51(4), 1187 – 1214.
- Fletcher, J. (2017). Differential susceptibility of peer influences: Gene-environment interactions and gene-environment correlations. In N. Mossige & T. Ødegård (Eds.), *Gene – environment transactions in developmental psychopathology* (pp. 1 – 18). Springer.
- Fletcher, J. M., & Venkataramani, A. S. (2018). Early life exposures, gene-environment interactions, and cognition in old age. Unpublished working paper.
- Flint, J., & Kendler, K. S. (2014). The genetics of major depression. *Neuron*, 81(3), 484 – 503.
- Frank, P., Ajnakina, O., Steptoe, A., & Cadar, D. (2019). Genetic susceptibility, socio-economic status and the risk of dementia: A

12-year follow-up of participants from the English Longitudinal Study of Ageing. *BMJ Open*, 9(4), e025309.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., & Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, 186(9), 1026 – 1034.

Galama, T. J., Lleras-Muney, A., & van Kippersluis, H. (2018). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. *Oxford Research Encyclopedia of Economics and Finance*.
<https://doi.org/10.1093/acrefore/9780190625979.013.253>

Gall, T. L., Evans, D. R., & Howard, J. (1997). The retirement adjustment process: Changes in the well-being of male retirees across time. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52(3), P110 – P117.

Galobardes, B., Shaw, M., Lawlor, D. A., & Lynch, J. W. (2006). Indicators of socioeconomic position (part 2). *Journal of Epidemiology and Community Health*, 60(2), 95 – 101.

Galton, F. (1875). The history of twins, as a criterion of the relative powers of nature and nurture. *Fraser's Magazine*, 12(71), 566 – 576.

- Garatachea, N., Emanuele, E., Calero, M., Fuku, N., Arai, Y., Abe, Y., Murakami, H., Miyachi, M., Yvert, T., & Verde, Z. (2014). APOE gene and exceptional longevity: Insights from three independent cohorts. *Experimental Gerontology*, 53, 16 – 23.
- Gatz, M., Reynolds, C. A., Fratiglioni, L., Johansson, B., Mortimer, J. A., Berg, S., Fiske, A., & Pedersen, N. L. (2006). Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry*, 63(2), 168 – 174.
- Génin, E. (2020). Missing heritability of complex diseases: Case solved? *Human Genetics*, 139(1), 103 – 113.
- Gharbi-Meliani, A., Dugravot, A., Sabia, S., Regy, M., Fayosse, A., Schnitzler, A., Kivimäki, M., Singh-Manoux, A., & Dumurgier, J. (2021). The association of APOE ϵ 4 with cognitive function over the adult life course and incidence of dementia: 20 years follow-up of the Whitehall II study. *Alzheimer's Research & Therapy*, 13(1), 1 – 11.
- Glymour, M. M., Kawachi, I., Jencks, C. S., & Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology & Community Health*, 62(6), 532 – 537.
- Glymour, M. M., & Manly, J. J. (2018). Compulsory schooling laws as quasi-experiments for the health effects of education: Reconsidering

mechanisms to understand inconsistent results. *Social Science & Medicine*, 214, 67 – 69.

Gondek, D., Lacey, R. E., Blanchflower, D. G., & Patalay, P. (2022).

How is the distribution of psychological distress changing over time? Who is driving these changes? Analysis of the 1958 and 1970 British birth cohorts. *Social Psychiatry and Psychiatric Epidemiology*, 57(5), 1007 – 1016.

Goossens, L., Van Roekel, E., Verhagen, M., Cacioppo, J. T., Cacioppo,

S., Maes, M., & Boomsma, D. I. (2015). The genetics of loneliness: Linking evolutionary theory to genome-wide genetics, epigenetics, and social science. *Perspectives on Psychological Science*, 10(2), 213 – 226.

Griffin, B., & Hesketh, B. (2008). Post-retirement work: The individual

determinants of paid and volunteer work. *Journal of Occupational and Organizational Psychology*, 81(1), 101 – 121.

Griffith, G. J., Morris, T. T., Tudball, M. J., Herbert, A., Mancano,

G., Pike, L., & Hemani, G. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature Communications*, 11(1), 5749.

Grossman, M. (1972). On the concept of health capital and the demand

for health. *Journal of Political Economy*, 80(2), 223 – 255.

- Gunn, J. M., Ayton, D. R., Densley, K., Pallant, J. F., Chondros, P., Herrman, H. E., & Dowrick, C. F. (2012). The association between chronic illness, multimorbidity and depressive symptoms in an Australian primary care cohort. *Social Psychiatry and Psychiatric Epidemiology*, 47(2), 175 – 184.
- Guo, G., Cai, T., Guo, R., Wang, H., & Harris, K. M. (2010). The dopamine transporter gene, a spectrum of most common risky behaviors, and the legal status of the behaviors. *PLOS ONE*, 5(2), e9352.
- Haas, W. H., III, & Serow, W. J. (2002). The baby boom, amenity retirement migration, and retirement communities: Will the golden age of retirement continue? *Research on Aging*, 24(1), 150 – 164.
- Halley, M. C., Olson, N. W., Ashley, E. A., Goldenberg, A. J., & Tabor, H. K. (2024). A just genomics needs an ELSI of translation. *Hastings Center Report*, 54(S2), S126 – S135.
- Hamad, R., Elser, H., Tran, D. C., Rehkopf, D. H., & Goodman, S. N. (2018). How and why studies disagree about the effects of education on health: A systematic review and meta-analysis of studies of compulsory schooling laws. *Social Science & Medicine*, 212, 168 – 178.
- Harden, K. P. (2021). *The genetic lottery: Why DNA matters for social equality*. Princeton University Press.

- Harden, K. P., Turkheimer, E., & Loehlin, J. C. (2007). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavior Genetics*, 37(2), 273 – 283.
- Harden, K. P., & Koellinger, P. D. (2020). Using genetics for social science. *Nature Human Behaviour*, 4, 567 – 576.
- Haviland, A. M., Jones, B. L., & Nagin, D. S. (2011). Group-based trajectory modeling extended to account for nonrandom participant attrition. *Sociological Methods & Research*, 40(2), 367 – 390.
- Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *American Journal of Psychiatry*, 158(10), 1568 – 1578.
- Hewitt, J. K. (2012). Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics*, 42(1), 1 – 2.
- Hofer, S. M., Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., & Easteal, S. (2002). Change in cognitive functioning associated with APOE genotype in a community sample of older adults. *Psychology and Aging*, 17(2), 194 – 208.
- Holmberg, M. J., & Andersen, L. W. (2022). Collider bias. *JAMA*, 327(13), 1282 – 1283.
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107 – 129.

- Horwitz, B. N., & Neiderhiser, J. M. (2011). Gene - environment interplay, family relationships, and child adjustment. *Journal of Marriage and Family*, 73(4), 804 - 816.
- Howe, L. J., Lawson, D. J., Davies, N. M., St. Pourcain, B., Lewis, S. J., Davey Smith, G., & Hemani, G. (2019). Genetic evidence for assortative mating on alcohol consumption in the UK Biobank. *Nature Communications*, 10(1), 5039.
- Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., & Palviainen, T. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, 54(5), 581 - 592.
- Hsu, Y. C., Lin, M. C., Su, M. H., Cheng, C. F., Pan, Y. J., Fan, C. C., & Wang, S. H. (2025). Education as a modifier of genetic influence on cognitive ability in older adults. *Behavior Genetics*. Advance online publication.
- Hultsch, D. F., Hertzog, C., Small, B. J., & Dixon, R. A. (1999). Use it or lose it: Engaged lifestyle as a buffer of cognitive decline in aging? *Psychology and Aging*, 14(2), 245 - 263.
- Hunter, D. J. (2005). Gene - environment interactions in human diseases. *Nature Reviews Genetics*, 6(4), 287 - 298.

- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615 – 635.
- Iob, E., Ajnakina, O., & Steptoe, A. (2021). The interactive association of adverse childhood experiences and polygenic susceptibility with depressive symptoms and chronic inflammation in older adults: A prospective cohort study. *Psychological Medicine*, 52(16), 3908 – 3919.
- Ishioka, Y. L., Gondo, Y., Fuku, N., Inagaki, H., Masui, Y., Takayama, M., Abe, Y., Arai, Y., & Hirose, N. (2016). Effects of the APOE ϵ 4 allele and education on cognitive function in Japanese centenarians. *Age*, 38(5), 495 – 503.
- Jansson, M., Gatz, M., Berg, S., Johansson, B., Malmberg, B., McClearn, G. E., Schalling, M., & Pedersen, N. L. (2004). Gender differences in heritability of depressive symptoms in the elderly. *Psychological Medicine*, 34(3), 471 – 479.
- Johnson, W., Geyer, C., Thavikulwat, A. C., & Balbona, J. V. (2021a). Polygenic scores for plasticity: A new tool for studying gene – environment interplay. *Development and Psychopathology*, 33(2), 314 – 325.
- Johnson, R. A., Sotoudeh, R., & Conley, D. (2021b). Polygenic scores for plasticity: A new tool for studying gene-environment interplay. *bioRxiv*. <https://doi.org/10.1101/2020.08.30.274530>

- Jones, R. N., Manly, J., Glymour, M. M., Rentz, D. M., Jefferson, A. L., & Stern, Y. (2011). Conceptual and measurement challenges in research on cognitive reserve. *Journal of the International Neuropsychological Society*, 17(4), 593 – 601.
- Jones, B. L., & Nagin, D. S. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods & Research*, 35(4), 542 – 571.
- Jung, H.-U., Kim, D. J., Baek, E. J., Chung, J. Y., Ha, T. W., Kim, H.-K., Kang, J.-O., Lim, J. E., & Oh, B. (2023). Gene-environment interaction explains a part of missing heritability in human body mass index. *Communications Biology*, 6(1), 324.
- Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: Evidence of genetic moderation. *Archives of General Psychiatry*, 68(5), 444 – 454.
- Keers, R., Coleman, J. R., Lester, K. J., Roberts, S., Breen, G., Thastum, M., Bogels, S., Schneider, S., Heiervang, E., Meiser-Stedman, R., Nauta, M., Creswell, C., Thirlwall, K., Rapee, R. M., Hudson, J. L., Lewis, C., Plomin, R., & Eley, T. C. (2016). A genome-wide test of the differential susceptibility hypothesis reveals a genetic predictor of differential response to

- psychological treatments for child anxiety disorders. *Psychotherapy and Psychosomatics*, 85(3), 146 – 158.
- Keller, M. C. (2014). Gene $\times$ environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biological Psychiatry*, 75(1), 18 – 24.
- Kendler, K. S. (1996). Parenting: A genetic-epidemiologic perspective. *American Journal of Psychiatry*, 153(1), 11 – 20.
- Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry*, 163(1), 109 – 114.
- Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsen, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., & Masson, G. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424 – 428.
- Kuan, V., Denaxas, S., Patalay, P., Nitsch, D., Mathur, R., Gonzalez-Izquierdo, A., Sofat, R., Partridge, L., Roberts, A., Wong, I. C. K., Hingorani, M., Chaturvedi, N., Hemingway, H., & Hingorani, A. D. (2023). Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: A population-based study. *The Lancet Digital Health*, 5(1), e16 – e27.

- Kuh, D. (2007). A life course approach to healthy aging, frailty, and capability. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 62(7), 717 – 721.
- Kwon, J. M., & Goate, A. M. (2000). The candidate gene approach. *Alcohol Research & Health*, 24(3), 164 – 168.
- Kämpfen, F., Kohler, I. V., Ciancio, A., Bruine de Bruin, W., Maurer, J., & Kohler, H.-P. (2020). Predictors of mental health during the COVID-19 pandemic in the US: Role of economic concerns, health worries and social distancing. *PLOS ONE*, 15(11), e0241895.
- Lai, W., Li, Y., Tian, X., & Li, S. (2017). Air pollution and cognitive function: Evidence from crop production cycles (SSRN Working Paper 3039935). <https://doi.org/10.2139/ssrn.3039935>
- Langella, S., Barksdale, N. G., Vasquez, D., Aguillon, D., Chen, Y., Su, Y., & Quiroz, Y. T. (2023). Effect of apolipoprotein genotype and educational attainment on cognitive function in autosomal dominant Alzheimer's disease. *Nature Communications*, 14(1), 5120.
- Le Marchand, L., & Wilkens, L. R. (2008). Design considerations for genomic association studies: Importance of gene-environment interactions. *Cancer Epidemiology, Biomarkers & Prevention*, 17(2), 263 – 267.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), 281 – 355.

- Lee, D., Brooks-Gunn, J., McLanahan, S. S., Notterman, D., & Garfinkel, I. (2013). The great recession, genetic sensitivity, and maternal harsh parenting. *Proceedings of the National Academy of Sciences*, 110(34), 13780 – 13784.
- Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine*, 12, 44.
- Liu, T., Li, C., Zhang, R., Millender, E. F., Miao, H., Ormsbee, M., & Kelly, T. N. (2023). A longitudinal study of polygenic score and cognitive function decline considering baseline cognitive function, lifestyle behaviors, and diabetes among middle-aged and older US adults. *Alzheimer's Research & Therapy*, 15(1), 196.
- Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., & Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet*, 390(10113), 2673 – 2734.
- Lorant, V., Delière, D., Eaton, W., Robert, A., Philippot, P., & Anseau, M. (2003). Socioeconomic inequalities in depression: A meta-analysis. *American Journal of Epidemiology*, 157(2), 98 – 112.
- Lovden, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020). Education and cognitive functioning across the life span. *Psychological Science in the Public Interest*, 21(1), 6 – 41.

- Luppa, M., Sikorski, C., Luck, T., Ehreke, L., Konnopka, A., Wiese, B., & Riedel-Heller, S. G. (2012). Age- and gender-specific prevalence of depression in latest-life: A systematic review and meta-analysis. *Journal of Affective Disorders*, 136(3), 212 - 221.
- Maier, R., Visscher, P. M., Robinson, M. R., & Wray, N. R. (2018). Embracing polygenicity: A review of methods and tools for genetic prediction. *PLOS Genetics*, 14(8), e1007580.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., & Chakravarti, A. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747 - 753.
- Marden, J. R., Mayeda, E. R., Walter, S., Vivot, A., Tchetgen Tchetgen, E. J., Kawachi, I., & Glymour, M. M. (2016). Using an Alzheimer disease polygenic risk score to predict memory decline in Black and White Americans over 14 years of follow-up. *Alzheimer Disease & Associated Disorders*, 30(3), 195 - 202.
- Maroille, T., & Tarailo-Graovac, M. (2019). Uncovering missing heritability in rare diseases. *Genes*, 10(4), 275.
- Marmot, M., Ryff, C. D., Bumpass, L. L., Shipley, M., & Marks, N. F. (1997). Social inequalities in health: Next questions and converging evidence. *Social Science & Medicine*, 44(6), 901 - 910.

- Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4), 584 – 591.
- Mazzonna, F. (2014). The long lasting effects of education on old age health: Evidence of gender differences. *Social Science & Medicine*, 101, 129 – 138.
- Mazzonna, F., & Peracchi, F. (2017). Unhealthy retirement? *Journal of Human Resources*, 52(1), 128 – 151.
- McAllister, K., Mechanic, L. E., Amos, C., Aschard, H., Blair, I. A., Chatterjee, N., Conti, D., Gauderman, W. J., Hsu, L., & Hutter, C. M. (2017). Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *American Journal of Epidemiology*, 186(7), 753 – 761.
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38(1), 115 – 142.
- McArdle, J. J., & Prescott, C. A. (2010). Contemporary modeling of gene $\times$ environment effects in randomized multivariate longitudinal studies. *Perspectives on Psychological Science*, 5(5), 606 – 621.

- McGue, M., Elkins, I., Walden, B., & Iacono, W. G. (2005). Perceptions of the parent-adolescent relationship: A longitudinal investigation. *Developmental Psychology*, 41(6), 971 – 984.
- McIntosh, A. M., Gow, A., Luciano, M., Davies, G., Liewald, D. C., Harris, S. E., Corley, J., Hall, J., Starr, J. M., Porteous, D. J., Tenesa, A., Visscher, P. M., & Deary, I. J. (2013). Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biological Psychiatry*, 73(10), 938 – 943.
- Merolla, D. M. (2013). The net Black advantage in educational transitions: An education careers approach. *American Educational Research Journal*, 50(5), 895 – 924.
- Merolla, D. M. (2018). Completing the educational career: High school graduation, four-year college enrollment, and bachelor's degree completion among Black, Hispanic, and White students. *Sociology of Race and Ethnicity*, 4(2), 281 – 297.
- Miech, R. A., & Shanahan, M. J. (2000). Socioeconomic status and depression over the life course. *Journal of Health and Social Behavior*, 41(2), 162 – 176.
- Mielke, M. M., Vemuri, P., & Rocca, W. A. (2014). Clinical epidemiology of Alzheimer's disease: Assessing sex and gender differences. *Mayo Clinic Proceedings*, 89(7), 754 – 761.

- Mills, M. C., & Rahal, C. (2019). A scientometric review of genome-wide association studies. *Communications Biology*, 2(1), 9.
- Mills, M. C., & Tropf, F. C. (2020). Sociogenomics: What it is and how it can be used. *Annual Review of Sociology*, 46, 553 – 581.
- Monroe, S. M., & Simons, A. D. (1991). Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders. *Psychological Bulletin*, 110(3), 406 – 425.
- Mortensen, E. L., & Høgh, P. (2001). A gender difference in the association between APOE genotype and age-related cognitive decline. *Neurology*, 57(1), 89 – 95.
- Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, 9, e48376.
- Muller, M., Sigurdsson, S., Kjartansson, O., Jonsson, P. V., Garcia, M., von Bonsdorff, M. B., Gunnarsdottir, I., Thorsdottir, I., Harris, T. B., & van Buchem, M. (2014). Birth size and brain function 75 years later. *Pediatrics*, 134(4), 761 – 770.
- Munafò, M. R., Durrant, C., Lewis, G., & Flint, J. (2009). Gene × environment interactions at the serotonin transporter locus. *Biological Psychiatry*, 65(3), 211 – 219.
- Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M., & Davey Smith, G. (2018). Collider scope: When selection bias can substantially

- influence observed associations. *International Journal of Epidemiology*, 47(1), 226 – 235.
- Musliner, K. L., Seifuddin, F., Judy, J. A., Pirooznia, M., Goes, F. S., & Zandi, P. P. (2015). Polygenic risk, stressful life events and depressive symptoms in older adults: A polygenic score analysis. *Psychological Medicine*, 45(8), 1709 – 1720.
- Musliner, K. L., Munk-Olsen, T., Eaton, W. W., & Zandi, P. P. (2016). Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. *Journal of Affective Disorders*, 192, 199 – 207.
- Nagin, D. S. (2014). Group-based trajectory modeling: An overview. *Annals of Nutrition and Metabolism*, 65(2 – 3), 205 – 210.
- Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, 6, 109 – 138.
- Najar, J., Thorvaldsson, V., Kern, S., Skoog, J., Waern, M., Zetterberg, H., & Zettergren, A. (2023). Polygenic risk scores for Alzheimer's disease in relation to cognitive change: A representative sample from the general population followed over 16 years. *Neurobiology of Disease*, 189, 106357.
- Nash, A. (2018). National population projections: 2018-based. Office for National Statistics.

- National Institute for Health and Care Excellence. (2022). Dementia, disability and frailty in later life: Mid-life approaches to delay or prevent onset (NICE Guideline NG16).
- Nguyen, T. T., Tchetgen, E. J. T., Kawachi, I., Gilman, S. E., Walter, S., Liu, S. Y., Manly, J. J., & Glymour, M. M. (2016). Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. *Annals of Epidemiology*, 26(1), 71 – 76.
- Niti, M., Yap, K.-B., Kua, E.-H., Tan, C.-H., & Ng, T.-P. (2008). Physical, social and productive leisure activities, cognitive decline and interaction with APOE- ϵ 4 genotype in Chinese older adults. *International Psychogeriatrics*, 20(2), 237 – 251.
- Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1), 152 – 175.
- Owen, M. J., Legge, S. E., Rees, E., Walters, J. T. R., & O'Donovan, M. C. (2023). Genomic findings in schizophrenia and their implications. *Molecular Psychiatry*, 28, 3638 – 3647.
- Pariante, C. M., & Lightman, S. L. (2008). The HPA axis in major depression: classical theories and new developments. *Trends in Neurosciences*, 31(9), 464 – 468.

- Pauls, F., Petermann, F., & Lepach, A. C. (2013). Gender differences in episodic memory and visual working memory including the effects of age. *Memory*, 21(7), 857 – 874.
- Pereira, R. D., Rietveld, C. A., & van Kippersluis, H. (2025). The interplay between maternal smoking and genes in offspring birth weight. *Journal of Human Resources*, 60(2), 400 – 433.
- Peyrot, W. J., Van der Auwera, S., Milaneschi, Y., Dolan, C. V., Madden, P. A. F., Sullivan, P. F., Strohmaier, J., Ripke, S., Rietschel, M., Nivard, M. G., Mullins, N., Montgomery, G. W., Henders, A. K., Heath, A. C., Fisher, H. L., Dunn, E. C., Byrne, E. M., Air, T. A., & Penninx, B. (2018). Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5,765 subjects from the Psychiatric Genomics Consortium. *Biological Psychiatry*, 84(2), 138 – 147.
- Phelan, J. C., Link, B. G., & Feldman, N. M. (2014). The genomic revolution and beliefs about essential racial differences: A backdoor to eugenics? *American Sociological Review*, 79(2), 209 – 227.
- Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science*, 264(5166), 1733 – 1739.
- Plomin, R., Gidziela, A., Malanchini, M., & von Stumm, S. (2022). Gene–environment interaction using polygenic scores: Do polygenic scores for psychopathology moderate predictions from environmental risk to

behavior problems? *Development and Psychopathology*, 35(3), 1107 – 1117.

- Podewils, L. J., Guallar, E., Kuller, L. H., Fried, L. P., Lopez, O. L., Carlson, M., & Lyketsos, C. G. (2005). Physical activity, APOE genotype, and dementia risk: Findings from the Cardiovascular Health Cognition Study. *American Journal of Epidemiology*, 161(7), 639 – 651.
- Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256), 748 – 752.
- Reynolds, C. A., Finkel, D., & Zavala, C. (2014). Gene by environment interplay in cognitive aging. In D. Finkel & C. A. Reynolds (Eds.), *Behavior genetics of cognition across the lifespan* (pp. 169 – 199). Springer.
- Rhemtulla, M., & Tucker-Drob, E. M. (2012). Gene-by-socioeconomic status interaction on school readiness. *Behavior Genetics*, 42(4), 549 – 558.
- Riaz, M., Huq, A., Ryan, J., Orchard, S. G., Tiller, J., Lockery, J., Woods, R. L., Wolfe, R., Renton, A. E., Goate, A. M., Sebra, R., Schadt, E., Brodtmann, A., Shah, R. C., Storey, E., Murray, A. M., McNeil, J. J., & Lacaze, P. (2021). Effect of APOE and a polygenic

- risk score on incident dementia and cognitive decline in a healthy older population. *Aging Cell*, 20(6), e13384.
- Ridley, M. (2003). *Nature via nurture: Genes, experience, and what makes us human*. Fourth Estate.
- Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L., Hoh, J., Griem, A., Kovacs, M., Ott, J., & Merikangas, K. R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. *JAMA*, 301(23), 2462 - 2471.
- Robins, C. J., & Block, P. (1989). Cognitive theories of depression viewed from a diathesis-stress perspective: Evaluations of the models of Beck and of Abramson, Seligman, and Teasdale. *Cognitive Therapy and Research*, 13(4), 297 - 313.
- Rohwedder, S., & Willis, R. J. (2010). Mental retirement. *Journal of Economic Perspectives*, 24(1), 119 - 138.
- Rowe, D. C., Jacobson, K. C., & Van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development*, 70(5), 1151 - 1162.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

- Rusted, J., & Carare, R. O. (2015). Are the effects of APOE ϵ 4 on cognitive function in nonclinical populations age- and gender-dependent? *Neurodegenerative Disease Management*, 5(1), 37 – 48.
- Røysamb, E., & Tambs, K. (2016). The beauty, logic and limitations of twin studies. *Norsk Epidemiologi*, 26(1 – 2).
<https://doi.org/10.5324/nje.v26i1-2.2014>
- Sabia, S., Kivimaki, M., Kumari, M., Shipley, M. J., & Singh-Manoux, A. (2010). Effect of apolipoprotein E ϵ 4 on the association between health behaviors and cognitive function in late midlife. *Molecular Neurodegeneration*, 5(1), 1 – 7.
- Sacco, M. A., Ricci, P., & Aquila, I. (2023). The effects of COVID-19 pandemic on public mental health: State of art and future perspectives. *International Journal of Environmental Research and Public Health*, 20(9), 5639.
- Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., & Aravkin, A. Y. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, 398(10312), 1700 – 1712.
- Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype-environment effects. *Child Development*, 54(2), 424 – 435.

- Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science*, 174(4016), 1285 - 1295.
- Schikowski, T., Vossoughi, M., Vierkötter, A., Schulte, T., Teichert, T., Sugiri, D., Fehsel, K., Tzivian, L., Bae, I.-s., & Ranft, U. (2015). Association of air pollution with cognitive functions and its modification by APOE gene variants in elderly women. *Environmental Research*, 142, 10 - 16.
- Schmitz, L. L., & Conley, D. (2016). The impact of late-career job loss and genotype on body mass index. NBER Working Paper 22080.
- Schneeweis, N., Skirbekk, V., & Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion? *Demography*, 51(2), 619 - 643.
- Schuebel, K., Gitik, M., Domschke, K., & Goldman, D. (2016). Making sense of epigenetics. *International Journal of Neuropsychopharmacology*, 19(11), pyw058.
- Schultz, M. R., Lyons, M. J., Franz, C. E., Grant, M. D., Boake, C., Jacobson, K. C., Xian, H., Schellenberg, G. D., Eisen, S. A., & Kremen, W. S. (2008). Apolipoprotein E genotype and memory in the sixth decade of life. *Neurology*, 70(19 Part 2), 1771 - 1777.
- Schwartz, G. L., & Glymour, M. M. (2023). Bridging the divide: Tackling tensions between life-course epidemiology and causal inference. *Annual Review of Developmental Psychology*, 5(1), 355 - 374.

- Seeman, T. E., Huang, M.-H., Bretsky, P., Crimmins, E., Launer, L., & Guralnik, J. M. (2005). Education and APOE-E4 in longitudinal cognitive decline: MacArthur studies of successful aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60(2), P74 - P83.
- Seeman, T. E., Miller-Martinez, D. M., Stein Merkin, S., Lachman, M. E., Tun, P. A., & Karlamangla, A. S. (2011). Histories of social engagement and adult cognition: Midlife in the U.S. study. *The Journals of Gerontology: Series B*, 66(Suppl. 1), i141 - i152.
- Selzam, S., Ritchie, S. J., Pingault, J.-B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing within- and between-family polygenic score prediction. *The American Journal of Human Genetics*, 105(2), 351 - 363.
- Shanahan, M. J., & Hofer, S. M. (2005). Social context in gene - environment interactions: Retrospect and prospect. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60(Special Issue 1), 65 - 76.
- Shen, H., & Feldman, M. W. (2020). Genetic nurturing, missing heritability, and causal analysis in genetic statistics. *Proceedings of the National Academy of Sciences*, 117(41), 25646 - 25654.
- Smith, J. A., Kho, M., Zhao, W., Yu, M., Mitchell, C., & Faul, J. D. (2018). Genetic effects and gene-by-education interactions on

- episodic memory performance and decline in an aging population. *Social Science & Medicine*, 212, 51 – 62.
- So, H. C., & Sham, P. C. (2017). Improving polygenic risk prediction: From analytical challenges to clinical applications. *Genome Medicine*, 9, 10.
- Sowan, W., Rutin, R., & Cohen, M. (2023). Chronic stressors, coping strategies, and depressive symptoms: A comparison across older age groups. *Stress and Health*, 39(5), 1037 – 1046.
- Step toe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: The English Longitudinal Study of Ageing. *International Journal of Epidemiology*, 42(6), 1640 – 1648.
- Step toe, A., Biddulph, J. P., & De Oliveira, C. (2025). Social isolation, wealth, and cognitive health over the life course: Evidence from the English Longitudinal Study of Ageing. *Molecular Psychiatry*. Advance online publication.
- Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 8(3), 448 – 460.
- Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, 47(10), 2015 – 2028.
- Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, 11(11), 1006 – 1012.

- Stern, Y., Chetelat, G., Habeck, C., Arenaza-Urquijo, E. M., Vemuri, P., Estanga, A., Bartres-Faz, D., Cantillon, M., Clouston, S. A. P., Elman, J. A., Gold, B. T., Jones, R., Kempermann, G., Lim, Y. Y., Van Loenhoud, A., Martinez-Lage, P., Morbelli, S., Okonkwo, O., Ossenkoppele, R., & Vuoksima, E. (2019). Mechanisms underlying resilience in ageing. *Nature Reviews Neuroscience*, 20(4), 246.
- Stern, Y., Barnes, C. A., Grady, C., Jones, R. N., & Raz, N. (2023). A framework for concepts of reserve and resilience in aging. *Neurobiology of Aging*, 124, 100 - 103.
- Su, J., & Xiao, X. (2022). Factors leading to the trajectory of cognitive decline in middle-aged and older adults using group-based trajectory modeling: A cohort study. *Medicine*, 101(47), e31817.
- Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry*, 157(10), 1552 - 1562.
- Sullivan, P. F., Agrawal, A., Bulik, C. M., Andreassen, O. A., Børghlum, A. D., Breen, G., Cichon, S., Edenberg, H. J., Faraone, S. V., Gelernter, J., Mathews, C. A., Nievergelt, C. M., Smoller, J. W., & O'Donovan, M. C. (2018). Psychiatric genomics: An update and an agenda. *American Journal of Psychiatry*, 175(1), 15 - 27.
- Sundermann, E. E., Biegon, A., Rubin, L. H., Lipton, R. B., Mowrey, W., Landau, S. M., & Maki, P. M. (2018). Does the female advantage in

verbal memory contribute to underestimating Alzheimer's disease pathology in women versus men? *Alzheimer's & Dementia*, 14(7), 900 - 910.

Tabery, J. (2023). *Beyond versus: The struggle to understand the interaction of nature and nurture*. MIT Press.

Tampubolon, G., & Maharani, A. (2017). When did old age stop being depressing? Depression trajectories of older Americans and Britons 2002 - 2012. *The American Journal of Geriatric Psychiatry*, 25(11), 1187 - 1195.

Taylor, W. D., Boyd, B., Turner, R., McQuoid, D. R., Ashley-Koch, A., MacFall, J. R., Saleh, A., & Potter, G. G. (2017). APOE ϵ 4 associated with preserved executive function performance and maintenance of temporal and cingulate brain volumes in younger adults. *Brain Imaging and Behavior*, 11(1), 194 - 204.

Thomas, D. (2010). Gene-environment-wide association studies: Emerging approaches. *Nature Reviews Genetics*, 11(4), 259 - 272.

Thomasson, H. R., Edenberg, H. J., Crabb, D. W., Mai, X. L., Jerome, R. E., Li, T. K., Wang, S. P., Lin, Y. T., Lu, R. B., & Yin, S. J. (1991). Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *American Journal of Human Genetics*, 48(4), 677 - 681.

- Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, 19(9), 581 – 590.
- Trejo, S., & Domingue, B. W. (2018). Genetic nature or genetic nurture? Introducing social genetic parameters to quantify bias in polygenic score analyses. *Biodemography and Social Biology*, 64(3 – 4), 187 – 215.
- Trejo, S., Belsky, D. W., Boardman, J. D., Freese, J., Harris, K. M., Herd, P., Sicinski, K., & Domingue, B. W. (2018). Schools as moderators of genetic associations with life course attainments: Evidence from the WLS and Add Health. *Sociological Science*, 5, 513 – 540.
- Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., van der Most, P. J., de Vlaming, R., Bakshi, A., Briley, D. A., Rahal, C., Hellpap, R., Iliadou, A. N., Esko, T., Metspalu, A., Medland, S. E., Martin, N. G., Barban, N., Snieder, H., Robinson, M. R., & Mills, M. C. (2017). Hidden heritability due to heterogeneity across seven populations. *Nature Human Behaviour*, 1(10), 757 – 765.
- Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E., & Fask, D. (2011). Emergence of a gene $\times$ socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychological Science*, 22(1), 125 – 133.

- Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin*, 140(4), 949 - 979.
- Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., & Benjamin, D. J. (2021). Problems with using polygenic scores to estimate genetic effects on complex traits. *Frontiers in Genetics*, 12, 626800.
- Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 59.
- Uher, R., & McGuffin, P. (2010). The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update. *Molecular Psychiatry*, 15(1), 18 - 22.
- Uher, R. (2014). Gene - environment interactions in severe mental illness. *Frontiers in Psychiatry*, 5, 48.
- Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era: Concepts and misconceptions. *Nature Reviews Genetics*, 9(4), 255 - 266.
- Wang, H.-X., Gustafson, D. R., Kivipelto, M., Pedersen, N. L., Skoog, I., Winblad, B., & Fratiglioni, L. (2012). Education halves the risk

of dementia due to apolipoprotein ϵ 4 allele: A collaborative study from the Swedish Brain Power initiative. *Neurobiology of Aging*, 33(5), 1007.e1 – 1007.e7.

Wang, M., Zhan, Y., Liu, S., & Shultz, K. S. (2008). Antecedents of bridge employment: A longitudinal investigation. *Journal of Applied Psychology*, 93(4), 818 – 830.

Ware, E., Schmitz, L., Faul, J., Gard, A., Smith, J., Mitchell, C., Weir, D., & Kardina, S. (2017). Method of construction affects polygenic score prediction of common human traits. *bioRxiv*.
<https://doi.org/10.1101/106062>

Whalley, L. J., Dick, F. D., & McNeill, G. (2006). A life-course approach to the aetiology of late-onset dementias. *The Lancet Neurology*, 5(1), 87 – 96.

World Health Organization. (2019). Dementia. WHO.

World Health Organization. (2022a). Depressive disorder (depression). WHO.

World Health Organization. (2022b). Mental disorders. WHO.

World Health Organization. (2022c). Mental health. WHO.

Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., & Carlson, C. S. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature*, 570(7762), 514 – 518.

- Xiang, X. (2019). Seven-year trajectories of depressive symptoms and their predictors among older Americans. *Journal of Aging and Health*, 32(7 – 8), 626 – 642.
- Yaffe, K., Hoang, T. D., Byers, A. L., Barnes, D. E., & Friedl, K. E. (2014). Lifestyle and health-related risk factors and risk of cognitive aging among older veterans. *Alzheimer's & Dementia*, 10(Suppl.), S111 – S121.
- Yang, Y., Li, W., Zhang, Q., Zhang, L., Cheung, T., & Xiang, Y.-T. (2020). Mental health services for older adults in China during the COVID-19 outbreak. *The Lancet Psychiatry*, 7(4), e19.
- Zahodne, L. B., Glymour, M. M., Sparks, C., Bontempo, D., Dixon, R. A., MacDonald, S. W., & Manly, J. J. (2011). Education does not slow cognitive decline with aging: 12-year evidence from the Victoria Longitudinal Study. *Journal of the International Neuropsychological Society*, 17(6), 1039 – 1046.
- Zaninotto, P., Iob, E., Demakakos, P., & Steptoe, A. (2022). Immediate and longer-term changes in the mental health and well-being of older adults in England during the COVID-19 pandemic. *JAMA Psychiatry*, 79(2), 151 – 159.
- Zhang, Q., Privé, F., Vilhjálmsón, B. J., & Speed, D. (2024). Improving polygenic prediction in ancestrally diverse populations. *Nature Communications*, 15, 2098.

- Zhou, H., Deng, J., Li, J., Wang, Y., Zhang, M., & He, H. (2003). Study of the relationship between cigarette smoking, alcohol drinking and cognitive impairment among elderly people in China. *Age and Ageing*, 32(2), 205 – 210.
- Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*, 159(7), 702 – 706.
- van Calker, D., & Serchov, T. (2021). The “missing heritability” problem in psychiatry: Is the interaction of genetics, epigenetics and transposable elements a potential solution? *Neuroscience & Biobehavioral Reviews*, 126, 23 – 42.
- von Hinke, S., & Sørensen, E. N. (2023). The long-term effects of early-life pollution exposure: Evidence from the London smog. *Journal of Health Economics*, 90, 102827.