Genetic and Environmental Correlates of Psychological Distress: evidence from the British Birth Cohorts

Thesis submitted in accordance with the requirements of the University College London for the Degree of Doctor in Philosophy by:

Esme Elsden

Centre for Longitudinal Studies

University College London

Primary Supervisor: Professor David Bann

Secondary Supervisors: Professor Praveetha Patalay, Dr Liam Wright

Declaration

I, Esme Elsden, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis explored the association between a polygenic score (PGS) for psychological distress and observed (phenotypic) psychological distress across ages, cohorts, and social environments. Utilising data from the 1958 National Child Development Study (1958c) and the 1970 British Cohort Study (1970c), this thesis examined how the PGS interacted with environmental factors correlated to distress outcomes.

Study 1 examined whether the PGS is associated with phenotypic distress from ages 23-50 in 1958c. Linear regressions assessed the strength of the association changed across adulthood and whether the variance explained by the PGS fluctuated by age. The PGS was positively correlated with psychological distress. Findings further suggest a small increase in the association between the PGS and psychological distress with age.

Study 2 compared associations between the PGS and psychological distress in two cohorts—1958c and 1970c – between ages 23-50. Using multilevel modelling, this study explored whether there are cohort and sex differences in the association between PGS and psychological distress during adulthood. Results indicated the association was similar across cohorts, while the variance in distress explained by the PGS was greater in the 1958c than the 1970c. Females exhibited stronger associations between the PGS and distress, compared to males.

Study 3 examined the interplay between the PGS and an indicator of childhood socioeconomic status (father's social class), testing whether these factors independently or interactively associated with psychological distress in adulthood and whether the estimates differ by sex. Findings suggested that both PGS and father's social class independently contributed to psychological distress, but there was no evidence of an interaction, indicative of additive rather than multiplicative effects.

Collectively, these studies suggest that associations between genes and psychological distress are not static but may dynamically interact—with age, cohort, and social conditions. The findings further suggest that while genetic liability plays a role in influencing psychological distress across contexts, its expression is modified by both temporal and social factors, supporting a complex biosocial model of distress.

Impact statement

This thesis contributes to our understanding of how genetic propensity to psychological distress interacts with environmental factors across different life stages and social contexts.

For genetic researchers, this work demonstrates the importance of incorporating longitudinal data when studying genetic influences on psychological distress. Traditional genetic research often relies on large but non-representative cohorts without sufficient longitudinal coverage. By showing that the association between polygenic scores and psychological distress increases with age, this thesis challenges the assumption that genetic effects remain static throughout life, encouraging more dynamic models of genetic influence in future research.

To conduct this research, I generated polygenic scores (PGS) for psychological distress from raw genetic data in both birth cohorts. These PGS' are a new resource for lifecourse researchers. The widespread availability of these scores will enable research that examines genetic influences, alongside to control for genetic factors when investigating environmental determinants. This methodological advance provides another way to better understand psychological distress aetiology.

Cohort researchers, who have historically focused on environmental determinants without examining genetic factors, will benefit from this interdisciplinary approach. The integration of polygenic scores with longitudinal cohort data provides a methodological framework for researchers seeking to understand how genetic and environmental factors interact across the lifecourse.

The cross-cohort comparison between individuals born in 1958 and 1970 illuminates how societal changes may correlate with the expression of genetic liability to psychological distress. The observation that the 1958 cohort had more variance in distress attributable to genetic factors compared to the 1970 cohort stimulates new research questions about how specific historical and environmental changes might moderate genetic influences.

For policymakers and public health officials, this research adds supporting evidence that societal-level interventions and policy changes can potentially modify the association between genetic factors and psychological distress in a population. This is due to the overall low variance explained by genetic factors (low R^2) in psychological distress. Therefore, this research could be beneficial for devising evidence-based policy.

This research was presented at multiple peer-reviewed conferences, including:

- May 2023 Bologna, Italy – European Social Science Genetics Network Meeting
- June 2023 Murcia, Spain – Behaviour Genetics Association Meeting
- October 2023 Munich, Germany – Society for Longitudinal and Lifecourse Studies Conference

Acknowledgements

First, I would like to thank my supervisors: Professor David Bann, Professor Praveetha Patalay and Dr Liam Wright, for their patience and continued inspiration to be a better researcher. I'd also like to acknowledge the contributions of time and effort from both Dr Victoria Garfield and Dr Gemma Shireby, who enabled me to succeed with the genetic analyses.

This research would not have been possible without the invaluable contribution from participants from the 1958 National Child Development Study and the 1970 British Cohort Study. I would like to thank the UK Data Service and the Centre for Longitudinal Studies for enabling access to such rich data.

There is no way I could have got to this point without the lateral learning and support of my fellow PhD students: Dr James Lathe, Sorcha Ni Chobhthaigh, Dr Tom Steare, Dr Tracy Odigie, Dr Giulia Piazza and Dr Charlotte Constable Fernandez. To name a few!

I dedicate this thesis to my parents, Andy Elsden and Professor Brenda Roe. I can only hope I become more and more like you as I age.

This thesis is especially dedicated to my dad, Andy Elsden (1953-2022), who taught me that the greatest gift is to be interested. That it is the small acts of kindness which make the world go round; the holding of a hand and unequivocally supporting every curiosity I've ever had. As much as this thesis was a major life accomplishment, to love and be loved are the incomparable reasons for living.

This thesis highlights that we are continuously becoming ourselves. I have the privilege of multiple decades-long friendships, watching people I love do just that. Thank you Ravir Grewal, Dr Esther Lee, May Xiang, Giulia Baggaley Simpson, Emma Ratcliffe and Dr Rhona Dawson. And to Leila Parsa – we've come a long way since sneaking books under desks in Mrs Burrow's class.

Thank you to the best gift that UCL has given me – endless laughter with Jake Cohen, Manya Eversley, Eleri Fanshawe and Dr Rachel Funk.

Finally, a thank you to Dr Oliver Yule-Smith, for everything.

"All things originate from one another, and vanish into
one another
According to necessity;
They give to each other justice and recompense for injustice
In conformity with the order of Time."

Anaximander

Table of Contents

1 Introduction 13

2 Research Questions 37

3 Description of the Data and Analytical Strategies 38

4 Study 1: Age differences in the association between polygenic scores and psychological distress: evidence from the 1958 British Birth Cohort 69

5 Study 2: Cohort differences in the association between polygenic scores and psychological distress: evidence from the 1958 and 1970 British Birth Cohorts 102

6 Study 3: Investigating gene-environment interplay between polygenic and father's social class on adult psychological distress during adulthood 141

7 Discussion 192

8 Conclusion 205

9 References 206

List of Tables

Table 1: Summary Of Primary Theoretical Frameworks And Hypothesis For Each Study 29

Table 2: Summary Of Sample Sizes 38

Table 3: Genetic Data Quality Control Summary Of The Removal Of Individuals And Snps For The 1958C And 1970C 40

Table 4: Summary Of The Cronbach's Alpha For 9-Item Malaise Inventory Score At Ages 23-50 In Both The 1958C And 1970C 42

Table 5: Malaise Inventory Score Items Used At Different Ages In The 1958 And 1970C, Responses Were Asked "Yes" Or "No" For Presence Of Symptoms 43

Table 6: Number Of Snps Included At Each P-Value Threshold For The Polygenic Scores Used In This Thesis 46

Table 7: Mean And Standard Deviation Of The Malaise Inventory Score Of Participants In The 1958C And 1970C With Genetic Data 52

Table 8: Correlation Matrix Of Malaise Inventory Score Variables (Ages 23-50) The Polygenic Score For Psychological Distress In The 1958C And 1970C 55

Supplementary Table 1: Summary Of Baselmans Et Al 2018 Gwas' Cohorts Meta-Analysed And Their Corresponding Phenotype 66

Supplementary Table 2: Summary Of Baselmans Et Al 2018 Gwas Phenotype Measures And Comparison To Malaise Inventory Score 68

Table 9: Summary Table Of Key Literature Of Genetic Contributions To Psychological Distress Outcomes By Age 74

Table 10: Mean And Standard Deviation Of The Malaise Inventory Score In The 1958C Of Participants With Genetic Data 80

Table 11: Correlation Matrix Of Malaise Inventory Score Variables At Ages 23, 33, 42 And 50 And The Polygenic Scores For Psychological Distress And Major Depressive Disorder In The 1958C 81

Table 12: Linear Regression Associations Between Polygenic Score For Psychological Distress And Observed Psychological Distress Across Ages 23-50 In The 1958C (N=6,312, Threshold=0.005) 83

Supplementary Table 3: Linear Regression Associations Across All Thresholds Of The Polygenic Score For Psychological Distress And Observed Psychological Distress Across Ages 23-50 In The 1958C (N = 6,312) 94

Supplementary Table 4: Linear Regression Associations Across All Thresholds Of The Polygenic Score For Major Depressive Disorder And Observed Psychological Distress Across Ages 23-50 In The 1958C (N = 6,312) 96

Supplementary Table 5: Linear Regression Associations Between Polygenic Score For Major Depressive Disorder And Observed Psychological Distress For Ages 23-50 In The 1958C (N=6,312, Threshold=0.005) 98

Supplementary Table 6: Complete Case Sample Per Age Linear Regression Associations Between Polygenic Score For Psychological Distress And Observed Psychological Distress Across Ages 23-50 In The 1958C (N= 3,748, Threshold = 0.005) 100

Supplementary Table 7: Stratified By Sex Linear Regression Associations Between Polygenic Score For Psychological Distress And Observed Psychological Distress Score For Ages 23-50 In The 1958C (Female N=3,172; Male N=3,140, Threshold=0.005) 101

Table 13: Mean And Standard Deviation Of The Malaise Inventory Score By Sex Of Those With Genetic Data In The 1958C And 1970C (Female N=; Male N=) 114

Table 14: Linear Regression Associations Between Polygenic Score For Psychological Distress With Overlapping Snps And 9-Item Malaise Inventory Score At Ages 23-50 In 1958C And 1970C (Pgs Threshold=0.005, 1958C N=6,312; 1970C N=5,423) 117

Table 15: Results From The Pooled And Stratified Multilevel Mixed Effects Models Interacting Mean-Centred Polygenic Score And Cohort For 1958C And 1970C (N=10,713) 121

Table 16: Results From The Pooled And Stratified Multilevel Mixed Effects Models Interacting Mean-Centred Polygenic Score And Sex For 1958C And 1970C (N=10,648) 125

Table 17: Linear Regression Associations Between Polygenic Score For Psychological Distress With Overlapping Snps And Observed Psychological Distress Using The 24-Item Malaise Inventory Score At Ages 23, 26, 30, 33 And 42 In 1958C And 1970C (Pgs Threshold=0.005; Imputed 1958C N = 5,939; Imputed 1970C N= 4,045) 131

Supplementary Table 8: Linear Regression Associations Across All Thresholds Of The Polygenic Score For Psychological Distress With Overlapping Snps Between 1958C And 1970C And Observed Psychological Distress Across Ages 23-50 In The 1958C (N=6,312) 137

Supplementary Table 9: Linear Regressions Associations Across All Thresholds Of The Polygenic Score For Psychological Distress With Overlapping Snps Between 1958C And 1970C And Observed Psychological Distress Across Ages 26-46 In The 1970C (N=5,423) 139

Table 18: Summary Table Of Key Evidence From Studies That Examined The Independent And Interaction Effects Of Polygenic Scores And Socioeconomic Circumstance Variables And Their Association With Psychological Distress Outcomes 149

Table 19: Results From The Multilevel Mixed-Effects Model Interacting Mean-Centred Polygenic Score At The 0.005 P-Value Threshold And 2-Category Father's Social Class For 1958C And 1970C Pooled (N=8,923) 161

Table 20: Results From The Multilevel Mixed-Effects Model Interacting Mean-Centred Polygenic Score At The 0.005 P-Value Threshold And 6-Category Father's Social Class For 1958C And 1970C Pooled (N=8,923) 164

Table 21: Results From The Multilevel Mixed Effects Model Interacting Mean-Centred Polygenic Score At The 0.005 P Value Threshold And Ridit Score Of The 6-Category Father's Social Class For 1958C And 1970C Pooled (N=8,923) 167

Table 22: Results From The Cohort-Stratified Multilevel Mixed Effects Model Interacting Mean-Centred Polygenic Score At The 0.005 P Value Threshold And Father's Social Class For 1958C And 1970C (1958C N=5,242; 1970C N=3,681) 169

Table 23: Results From The Sex-Stratified Multilevel Mixed Effects Model Interacting Mean-Centred Polygenic Score At The 0.005 P Value Threshold And Father's Social Class For Females (N=4,300) And Males (N=4,623) 171

Table 24: Summary Of Potential Studies Included For Meta-Analysis With Key Data Extracted 173

Table 25: Results From The Multilevel Mixed Effects Model Interacting Mean-Centred Polygenic Score At The $5X10^{-8}$ And 1 P-Value Thresholds And 2-Category Father's Social Class For 1958C And 1970C (N=8,976) 177

Table 26: Results From The Multilevel Mixed Effects Model Interacting Mean-Centred Polygenic Score At The 1 And $5X10^{-8}$ P Value Thresholds And 6-Category Father's Social Class For 1958C And 1970C (N=8,923) 179

Supplementary Table 10: Mean And Standard Deviation Of The Malaise Inventory Score Of Those With Genetic Data By Father's Social Class In 1958C And 1970C 188

Supplementary Table 11: Mean And Standard Deviation Of The Malaise Inventory Score Of Those With Genetic Data By 6-Category Father's Social Class In 1958C And 1970C 189

Supplementary Table 12: Post-Hoc Power Calculations Research Questions 1 And 2 From Study 3 190

Table 27: Summary Of Primary Theoretical Frameworks And Hypothesis For Each Study 193

List of Figures

Figure 1: Adapted From “Harris, K. M. & Mcdade, T. W. 2018. The Biosocial Approach To Human Development, Behavior, And Health Across The Lifecourse. Rsf: The Russell Sage Foundation Journal Of The Social Sciences Apr 2018, 4 (4) 2 26; Doi:10.7758/Rsf.2018.4.4.01” 14

Figure 2 A-C: From “Figure 6.1 Gene-Environment Interaction Models. Source: Adapted From Liu And Guo [38] In Mills, Barban & Tropf (2020) Introduction To Statistical Genetic Data Analysis” – Lines Indicate The Characterisation Of The Environment. 23

Figure 3: From Nivard Et Al. 2014 (107): “Figure 3. (A) Proportions Of Variance Explained By Genetic Factors (H2, ——), Comm2on Environment (C2, - - -), And Unique Environment (E2, ····) At Each Age As Derived From The Model. (B) Variance Components V_A, V_C, And V_E At Each Age As Derived From The Model.” 26

Figure 4: From Gondek Et Al. 2021 (124): “Figure 2. Age Profile Of The Mean Number Of Psychological Distress Symptoms – Cohort-Stratified And Pooled Across Cohorts.” 31

Figure 5: A) Mean And B) Standard Deviation Of The Malaise Inventory Score At Each Age Point, Shown As A Line Graph For 1958c And 1970c 53

Figure 6: Histogram Frequency Percentage Plot Of The 9-Item Malaise Inventory Score At Ages 23, 26, 30, 33, 34, 42, 46 And 50 In 1958c And 1970c 54

Figure 7: Missing Pattern In The 1958c Of The Malaise Inventory Score Variables At Ages 23, 33, 42 And 50 And The Polygenic Score 58

Figure 8: Missing Pattern In The (A) 1958c At Ages 23, 33, 42 And 50 And (B) 1970c At Ages 25, 30, 34, 42 And 46 Of The Malaise Inventory Score And The Polygenic Score 62

Figure 9: Plot Of The Linear Regression Results Of The Associations Between The Polygenic Score For Psychological Distress At 0.005 P-Value Threshold And The Malaise Inventory Score At Ages 23-50 In The 1958c 84

Supplementary Figure 1: Specification Curve Plot Of The Variance Explained By The Polygenic Score For Psychological Distress In The Linear Regression Model At Each Age At Each Potential P-Value Threshold In The 1958c 92

Supplementary Figure 2: Specification Curve Plot Of The Variance Explained By The Polygenic Score For Major Depressive Disorder In The Linear Regression Model At Each Age At Each Potential P-Value Threshold In The 1958c 93

Supplementary Figure 3: Plot Of The Linear Regression Results Of The Associations Between The Polygenic Score For Major Depressive Disorder At 0.05 P-Value Threshold And Observed Psychological Distress At Ages 23-50 In The 1958c 99

Figure 10: Mean And Standard Deviation Of Malaise Inventory Score By Sex For Ages 23-50 In 1958c And 1970c (N=9,620) 115

Figure 11: Plot Of The Beta Coefficients From The Linear Regression Results Of The Association Between The Polygenic Score For Psychological Distress With Overlapping Snps At 0.005 P-Value Threshold And Psychological Distress Outcomes In 1958c And 1970c Between 1981-2012 118

Figure 12: Plot Of The Incremental Adjusted R-Squared From The Linear Regression Results Of The Association Between The Polygenic Score For Psychological Distress With Overlapping Snps At 0.005 P-Value Threshold And Psychological Distress Outcomes In 1958c And 1970c Between 1981-2012 119

Figure 13: Predictive Margins Of Psychological Distress Across Mean-Centred Polygenic Score For 1958c And 1970c Cohorts In Model 1 122

Figure 14: The Marginal Effect Of The Polygenic Score On Psychological Distress In The 1958c And 1970 By Age In Model 2 And 3 123

Figure 15: Predictive Margins Of Psychological Distress Across Mean-Centred Polygenic Score For Males Versus Females In Model 1 126

Figure 16: Specification Curve Plots Of The Variance Explained By The Polygenic Score For Psychological Distress In The Regression Model At Each Age At Each Potential P-Value Threshold In A) 1958c And B) 1970c 128

Figure 17: Specification Curve Plots Of The Beta Coefficient From Linear Regressions Between Polygenic Score And Observed Psychological Distress, By Each Age At Each Potential P-Value Threshold In A) 1958c And B) 1970c 129

Figure 18: Mean And Standard Deviation Of Malaise Inventory Score By Father's Social Class Coded As Manual Versus Non-Manual At Ages 23-50 In The 1958c And 1970c (N=9,620) 159

Figure 19: Predictive Margins Of Psychological Distress For 6-Categories Of Father's Social Class Across The Polygenic Score 165

Figure 20: A) Forest Plot Of Fixed-Effect Meta-Analysis Of Gxe Studies With 1958c From Study 3, B) With 1958c From Keers 2017 Childhood Ses Composite, C) With 1958c From Keers 2017 Adulthood And Childhood Ses Composite 174

Figure 21: A) Forest Plot Of Random-Effect Meta-Analysis Of Gxe Studies With 1958c From Study 3, B) With 1958c From Keers 2017 Childhood Ses Composite, C) With 1958c From Keers 2017 Adulthood And Childhood Ses Composite. 175

Supplementary Figure 4: Histograms Of Polygenic Score Density By Father's Social Class And By Cohort In The 1958c And 1970c 187

Supplementary Figure 5: Estimated Power Of Main Effect (0.30) And Interaction Effect (-0.01) Across Sample Sizes Of 2,000-9,500. Lines Indicate The Effect Size (Red = 0.03, Blue = -0.01) 191

Abbreviations

GWAS – Genome-Wide Association Study

PGS – Polygenic Score

SNP – Single Nucleotide Polymorphism

1958c – 1958 National Child Development Study cohort

1970c – 1970 British Cohort Study cohort

*r*GE – Gene-Environment Correlation

GxE – Gene-environment Interaction

SES – Socioeconomic Status

SEP – Socioeconomic Position

EUR – European Ancestry

DSM – Diagnostic and Statistical Manual of Mental Disorders

ICD – International Classification of Diseases

ALSPAC – Avon Longitudinal Study of Parents and Children

HRS – Health and Retirement Study

ELSA – English Longitudinal Study of Ageing

APC – Age, Period, Cohort

MAR – Missing at Random

RGSC - Registrar General's Social Class

1 Introduction

The introduction of this thesis will begin by positioning the research within its core disciplinary frameworks—psychiatric, genetic, and lifecourse epidemiology—each of which contributes to a comprehensive biosocial epidemiology perspective on the aetiology of psychological distress. Following this, the thesis introduction will define psychological distress, the primary outcome, through a historical lens, examining its evolving definition and highlighting its significance as a global public health concern.

Next, the genetic epidemiology of psychological distress will be explored by discussing relevant key concepts such as heritability. The lifecourse epidemiology approach will then be outlined, emphasising the value of longitudinal data in capturing trends of the genetic and social contribution to psychological distress across adulthood. Finally, evidence for the social gradient of psychological distress will be discussed.

Section 2 will then introduce each study's main objective and research question. Within each study's introduction, a more targeted literature review section will be included.

1.1 Situating the thesis: what is biosocial epidemiology?

This thesis is interdisciplinary in scope and aim. It fulfils the requirements of a biosocial epidemiology thesis by synthesising across psychiatric epidemiology, genetic epidemiology and lifecourse epidemiology.

Psychiatric epidemiology uses statistics to understand the occurrence and distribution of mental and behavioural disorders across people, over space and time (1). As a discipline, it focuses on the causes and consequences of these disorders. With an aim to better understand their aetiology, nosology and to develop more effective intervention strategies for treatment and prevention (1).

Genetic epidemiology focuses on how genetic factors are associated with human traits, such as physical and mental health (2). It examines how genetic variations, such as single nucleotide polymorphisms (SNPs), contribute to individual differences in traits and diseases. Importantly, genetic epidemiology also considers gene-environment interaction, recognising that genes do not operate in isolation but are expressed within the context of environmental exposures.

Lifecourse epidemiology examines how early-life exposures and experiences shape long-term health trajectories. Central concepts include accumulation of risk, where repeated exposures accumulate, contributing to a risk burden for health outcomes; critical or sensitive periods, whereby specific stages in life, such as early childhood or adolescence, are pivotal in determining health outcomes (3). An example of accumulation could be when someone is exposed to disadvantaged socioeconomic circumstances in adolescence and begins to experience psychological distress as an outcome, with compounding exposures that accrue and maintain the individual's high level of distress. This creates a gradient of good to poor mental health outcomes corresponding to the gradient in socioeconomic status (4). The social gradient in psychological distress can begin in childhood, when parents' advantageous societal status transfers to their children and typically persists over the children's lives (5–7).

This is seen as an accumulation across the lifecourse. Still, there are also singular acute socioeconomic stressor events in a person's life that can be pivotal to the development of distress, such as job loss or the bereavement of a spouse (8–10).

Biosocial epidemiology, by integrating these three disciplines, is an approach which aims to understand human development, behaviour, and health (11). A key part of biosocial epidemiology focuses on the dynamic interplays between biology, experiences, and behaviours over the lifecourse. Figure 1 shows the conceptual model of biosocial dynamics across the lifecourse. The complex interplay between genetics, psychosocial processes, and wider environmental context informs the onset and progression of mental health outcomes across the lifecourse.

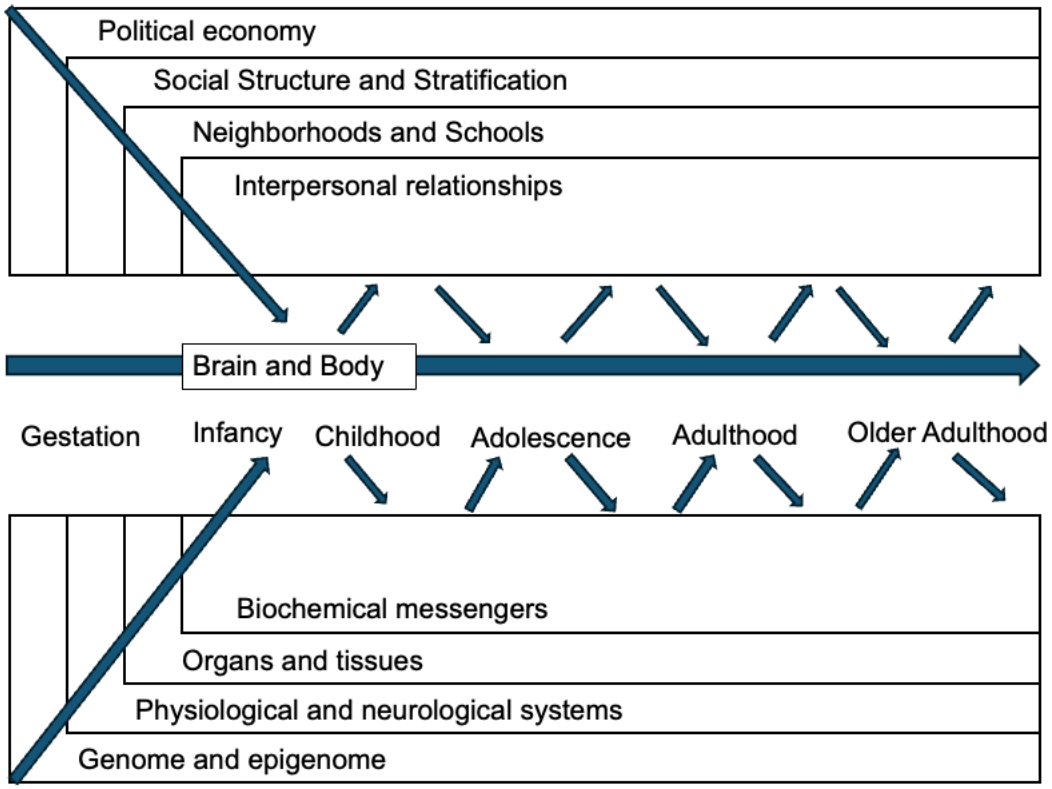


Figure 1: Adapted from "Harris, K. M. & McDade, T. W. 2018. The Biosocial Approach to Human Development, Behavior, and Health Across the Lifecourse. RSF: The Russell Sage Foundation Journal of the Social Sciences Apr 2018, 4 (4) 2 26; DOI:10.7758/RSF.2018.4.4.01"

This thesis applies concepts from psychiatric, lifecourse, and genetic epidemiology to explore how genetic liability and early-life socioeconomic circumstances are correlated with psychological distress across different cohorts. This integrative approach leverages biosocial epidemiology to explore the independent and interactive contribution of genetic factors and environmental exposures over time. In the context of mental health, a biosocial approach consistently positions psychological distress as not solely a product of biology or environment but emerges from their interplay.

Study 1 examines age-specific associations between polygenic scores for psychological distress and phenotypic expressions of distress. This focus aligns with lifecourse epidemiology by exploring the changing genetic contribution that manifests differently across adulthood. Study 2 explores cohort differences in the association between polygenic scores and psychological distress. By comparing two cohorts, this study examines how environmental shifts, and societal change may alter the importance of the genetic contribution to psychological distress. Study 3 investigates the gene-environment interaction between polygenic score and father's social class, focusing on how early-life socioeconomic conditions might moderate genetic vulnerability to psychological distress. This synthesis of genetic and social factors within a developmental framework is informed by biosocial epidemiology.

1.2 Defining Psychological Distress: Psychiatric Epidemiology

1.2.1 Historical overview

Psychological distress is a phenotype that combines the symptoms of major depressive disorder and anxiety disorders (12–14). Historically, this overlap was conceptualised as *melancholia*, characterised by a pervasive sadness and associated physical symptoms. This term then evolved into *generalised depression* and, more recently, into the broader term *psychological distress* (15–20). This transition reflects the changing understanding of mood and anxiety disorders, driven by shifts in psychiatric theory and practice. Despite changes in nomenclature and more refined diagnostic tools, traces of how we conceptualise psychological distress can be seen in descriptions as far back as 2000 BC (15,20).

Over time, depression and anxiety came to be seen as distinct disorders, and they have been clinically diagnosed and treated independently of one another (21). This is reflected in the refinement of diagnostic criteria in modern times, particularly through the development of the DSM and ICD, which led to the clinical separation of depression and anxiety. However, their overlapping symptoms continue to be studied under the umbrella of distress (22,23). Moreover, growing research on sub-types of depressive and anxious symptoms shows symptom overlap that then heightens how their comorbidity and complicate diagnosis (14). As diagnostic categories continue to evolve, the concept of psychological distress remains central to understanding the shared and unique features of these conditions (12).

Consequently, given the substantial nosological overlap and high rates of comorbidity observed, the DSM-5 has reconceptualised these constructs through the introduction of a novel integrated depression-anxiety diagnostic classification (22). This revision underscores the relevance of studying these conditions in tandem, particularly at the population-level where subclinical manifestations are prevalent.

1.2.2 Current Definition

As discussed, psychological distress is an expansive construct that includes symptoms of both major depressive disorder and anxiety disorder (12,13). It is primarily characterised as a state of emotional anguish comprised of depressive

symptoms such as a lack of interest, feelings of sadness or hopelessness, and anxiety symptoms such as restlessness and feelings of worry (24). Somatic symptoms such as insomnia, headaches, backache, and lack of energy may also be present. It is a highly heterogeneous emergent state, as it is comprised of thousands of symptom combinations. No single symptom is necessary for a person to be classified as distressed (25). The symptoms of psychological distress can vary greatly, indicating that the disorder is complex and influenced by multiple factors.

There is evidence that suggests different presentations of psychological distress may share similar underlying causes. For instance, overlapping depressive and anxiety symptoms such as fatigue, sadness and impaired executive functioning have been identified (26,27). Studies have found genetic links between major depressive disorder and generalised anxiety disorder (28,29). Psychological distress comprises a continuous broader unidimensional phenotype, meaning biological vulnerability can surpass diagnostic boundaries between disorders, as the genes identified are equally found in those who do and do not express the phenotype (30). Investigating psychological distress as a highly comorbid condition can help further our understanding of its complex impact on societal health. This thesis used this conceptualisation for psychological distress.

1.2.3 Global Public Health Importance of Psychological Distress

The global prevalence of depressive and anxiety disorders is projected to be the leading cause of morbidity by 2030 and currently affects ~500 million people (31,32). These disorders are the most common mental health conditions and account for most of the disease burden associated with mental health (33). As noted, they are also highly comorbid, with nearly 60% of those who report anxious feelings having depressive symptoms and vice versa (34), and symptoms of these conditions are often studied under the umbrella of psychological distress in epidemiological and population health studies.

Beyond the individual burden, psychological distress has substantial economic implications, contributing to decreased productivity and increased healthcare costs globally (35). Furthermore, the burden is disproportionately felt in low- and middle-income countries, where access to mental health care is often limited, exacerbating social and health inequalities (36). These conditions are also associated with various physical health outcomes, highlighting the need for an integrative biosocial approach to addressing their impact across adulthood (37).

Considering the complexities surrounding psychological distress—its heterogeneous symptoms, comorbidity between depression and anxiety, and its significant societal burden—this thesis adopted a biosocial epidemiological approach to understand better the biological and environmental factors that influence its progression across adulthood. As a result, this thesis used the broader phenotype of psychological distress as the outcome of interest.

1.3 Defining genetic concepts: Genetic Epidemiology

Psychological distress emerges from a complex interaction of symptoms, where genotypes interact with the environment leading to the expression of phenotypes

(2,38). Research in genetic epidemiology highlights that genetic variations contribute to individual differences in susceptibility to distress, suggesting that certain inherited SNP variants may increase vulnerability (2,39,40). By examining these genetic contributions, we can better understand the potential biological underpinnings of psychological distress and explore how genetic factors interact with environmental exposures to influence its manifestation and persistence.

1.3.1 Heritability: Broad and Narrow

Heritability is defined as the "*proportion of total variance in a population for a particular measurement, taken at a particular time or age, that is attributable to variation in additive genetic or total genetic contribution*" (41). To differentiate, narrow heritability refers to the proportion of variance explained solely by additive genetic factors. In contrast, broad heritability encompasses all genetic variance, including non-additive factors like gene-gene interactions (epistasis) and dominance effects. As this thesis employs polygenic scores, which capture additive genetic effects, references to heritability in this context refer to narrow-sense heritability. Elsewhere in the thesis, the term 'heritability' is used in line with the specific method employed (e.g., twin-based, SNP-based), and the relevant distinctions are made clear as appropriate.

Heritability was historically evaluated by comparing the similarity between monozygotic (identical twins, who share 100% of their genes) and dizygotic (non-identical twins, who share 50% of their genes on average) twins (42,43). However, with the advent of the genome-wide association study (GWAS), it became possible to estimate heritability from people's genotypes (44). This is done by measuring the genetic variation of single nucleotide polymorphisms (SNPs), which gives SNP-heritability. On average, two people may differ at every 1 in 1000 base pairs of the DNA sequence (45). This means that between person A and person B, at the same point in each DNA sequence, the base pair of A, T, G or C can differ. This accumulates along the DNA and results in phenotypic diversity in a population. A key advantage of this approach is that it avoids certain biases inherent in twin and family designs (e.g. shared environment assumptions), and it enables researchers to estimate genetic contributions in large cohorts of unrelated individuals.

1.3.2 Genome-Wide Association Studies & Polygenic Scores

Genome-Wide Association Studies

Genome-wide association studies, or GWAS, is a hypothesis-free approach used to identify genetic variants associated with particular diseases or traits (44). In GWAS, researchers examine the genomes of large samples of individuals to identify SNPs that are statistically more common in people with a certain disease or trait than in those without it (44). The purpose of GWAS is not only to find these associated SNPs, but also to provide a basis for further investigation into nearby genetic variants that may directly influence trait development. This is due to linkage disequilibrium where nearby genetic variants located close together on a chromosome tend to be inherited together. Due to the number of associations tested, the usual p-value threshold of 0.05 is replaced by a threshold of 5×10^{-8}. This is set to control for multiple testing and to reduce type 1 errors. This is a Bonferroni correction approximation, where, dividing

the conventional significance level by 1 million independent tests (0.05/1,000,000 = 5 x 10^{-8}).

The theoretical underpinning of the GWAS method advocates a 'common disease, common variant' hypothesis, that common genes are likely to influence common disorders (46). GWAS are usually limited to commonly varying SNPs with minor allele frequency (MAF – the extent to which the less common allele occurs in the population) of at least 1% (47). Though not included in GWAS, rare variants (MAF <1%) can have large effects on phenotypes at a population level (48,49). Therefore, GWAS' may underestimate the total genetic contribution to phenotypes.

Detecting associations with common SNPs requires large sample sizes to reach sufficient statistical power, especially for traits influenced by many genetic factors with small effect sizes. The first GWAS of depressive symptom was by Noordam et al. in 2015 (50), and they found 10 SNPs at the genome-wide significance level. Whereas the most recent GWAS of depressive symptom by Baselmans et al. 2018 found 239 (51). A narrower trait of major depressive disorder identified zero genome-wide significant SNPs in the first GWAS conducted in 2009 by Sullivan et al.(52), whereas the most recent, in 2024, by McIntosh et al., found 697 associations (53).This is due to the sample size available for analysis jumping from N=3,540 to N=5,050,033.

Initial research utilised twin study designs and estimated the heritability of distress to be 41%; these studies were predominantly based in the United States and other high-income countries (38,42). However, molecular studies of psychological distress have yielded more conservative estimates of SNP heritability. A study by the Psychiatric Genomics Consortium (PGC) in 2013 established major depressive disorder SNP heritability as 21%, utilising nine cohorts and a total sample of 18,422 (54). A more recent SNP heritability study estimated 3% to 11.3% depending on population and specificity of distress phenotype; their sample comprised 1,154,267 people across four cohorts of European ancestry (55,56).

SNP heritability estimates are thought to be smaller because GWAS' typically only capture the additive effects of common SNPs (57). Additionally, the genetic component in the twin model might be an overestimation, due to the conflation of the common shared environment. The difference between estimates is called the "missing heritability" problem (57,58). Both methods are flawed in their respective estimation of genetic effects on a trait. A more conservative estimate for heritability can be obtained from a within-family molecular design that considers both genetic nurture and shared environment to estimate the heritability of psychological distress (59,60). The most recent family GWAS estimated lower heritability for depression of ~2.5%, which is a large reduction from the estimate of 21% from the PGC 2013 GWAS (61). Despite this, exposures with low-moderate prediction can be useful in population research (62,63).

Polygenic Scores

One of the useful outputs of a GWAS is the potential to use results to generate a polygenic score (PGS). PGS' are summary variables that capture genetic liability to a trait. A weighted PGS is derived by summing the effect sizes from a GWAS multiplied by alleles frequency (64). This weighted approach enhances the predictive accuracy of the PGS for traits, providing a quantitative measure of an individual's genetic

propensity to a trait. In a weighted PGS, each allele's contribution to the score is adjusted by its effect size (beta) from a GWAS results, while an unweighted score counts the number of effect alleles. Weighted scores are generally more predictive and were chosen in this thesis for greater precision (65).

While PGS can predict traits to a degree, limitations exist. Predictive power can be affected by factors like selection bias. Where those who live longer contribute to more complete genetic data, those who do not may have more extreme expressions of the phenotypes, possibly reducing variance in scores (66,67). PGS-heritability estimates are smaller still than SNP-heritability (68). For example, the PGS for psychological distress only explains 1.64% of the variance in distress (51). This is potentially because PGS capture only a portion of commonly measured SNPs as they are derived from a GWAS. As PGS research advances, scores may become more predictive due to larger sample sizes, increasing SNP resolution, and improvements in multi-ancestry GWAS data (69,70). However, it remains important to consider that PGS-heritability is a conservative estimate and likely represents a lower bound of the genetic contribution to distress (71). At the same time, recent within-family studies suggest that polygenic scores may also be upwardly biased in certain contexts – particularly traits like educational attainment –due to indirect genetic effects, assortative mating, and residual population stratification (72).

Polygenic score heritability quantifies the proportion of phenotypic variance explained by measured genetic variants at a given time point, whereas polygenic penetrance refers to the association (slope) between the PGS and the absolute level of psychological distress—i.e., the expected difference in distress per unit (or 1 SD) higher PGS (44–46). Penetrance varies dynamically across development due to changing gene-environment interactions, neurobiological maturation, and environmental exposures, while heritability may remain stable (44,47). Conley (2016) defined polygenic penetrance as "*the association between a polygenic score (PGS) and its associated phenotype*" (36, p.1). Consequently, polygenic scores may demonstrate fluctuating predictive accuracy across ages not due to changes in genetic architecture, but due to developmental variation in genetic risk expression (46,48)**.** Low penetrance periods may lead to underestimation of genetic effects, while high penetrance windows may reveal stronger associations, making single time-point analyses potentially misleading (46)**.** Longitudinal examination of penetrance patterns therefore provides insights into how genetic predisposition manifests across the lifecourse (47)**.** Within this thesis, the derivation of incremental adjusted R^2 is a proxy for PGS heritability while the beta coefficients reflect polygenic penetrance.

1.3.3 Partitioning variance: Genes and the Environment

Within behavioural genetics, the "biometric approach" has focused on partitioning population variation into additive components of genetic and environmental variance (77). As a result, what we consider the genetic contribution to the variance has changed via the advent of the DNA revolution and the re-leveraging of twin studies via genotyping as outlined. The environment partition of the variance was originally divided into the "shared" and "non-shared" (78,79). These terms are derived from twin studies where the assumption was made that the shared environment, i.e. the familial home and childhood circumstances, were the same for each twin and therefore

controlled for (80). The non-shared environment then encompasses the macrosocial environment experienced separately to their twin, e.g., the friendship groups formed, the jobs worked, the higher education institutions attended. This also included the stochastic non-systematic contributions to phenotypes.

Some assumptions are made when dichotomising the shared versus non-shared environment. These include the equal environments assumption that the environments of identical twins are no more similar than those of fraternal twins. This may not always be the case, as identical twins may be treated more similarly than fraternal twins, which could inflate heritability estimates (81). An additional assumption is that the shared environment is devoid of genetic contribution. Yet, genetic nurture plays a part –when a parent's genes influence their child's outcomes by shaping the nurturing environment, rather than through direct genetic transmission (82). This may mean the passive gene-environment correlation inflates the genetic estimate, as outlined in the section below (60). This entanglement highlights the difficulty faced in partitioning the variance attributable to genes versus environments.

Shared environments contribute to early developmental foundations, influencing access to resources and opportunities, while non-shared environments, such as cohort-specific historical contexts or age-related life transitions, introduce variability in how these foundations manifest. Therefore, this thesis tests the relationship between polygenic scores and broader environments such as cohort and early life socioeconomic circumstances. Resultantly, the father's social class is more akin to a "shared" environmental variable while cohort is more conceptually aligned with the "non-shared" environment. However, it is key to note that as the shared and non-shared environments cannot be disentangled within an unrelated individual's design, these are merely conceptualisations.

1.3.4 Gene-Environment Interplay

Gene-environment interplay refers to Gene-Environment Correlation (*r*GE) and Gene-Environment Interaction (GxE), which will be explained in more detail below.

Historically, gene-environment interplay was conceptualised as a query about nature versus nurture (77). The debate of nature versus nurture was articulated by Francis Galton in 1883, who utilised the term, taken from Shakespeare's The Tempest when Prospero describes Caliban as *"devil, a born devil, on whose nature/nurture can never stick*" (77,83). Galton declared that "*there is no escape from the conclusion that nature prevails enormously over nurture*" (p.241) (84). By 1918 Galton's colleague Ronald A. Fisher then consolidated the "biometric approach," which focused on partitioning population variation into additive components of genetic and environmental variance (77). Fisher acknowledged that GxE could complicate this partitioning. For Fisher, however, the interplay between genes and environment was secondary to quantifying the heritable component, and GxE was treated as a nuisance rather than an important consideration. As a result, he used statistical transformations of variables to maintain additivity.

In contrast, Lancelot Hogben, writing in 1933, whose counter approach focused on the developmental tradition of biology, emphasised the inherent interplay between genes and environments as integral to developmental processes (77,85). Hogben argued

that phenotypic traits are the product of complex interactions between genetic and environmental factors, which cannot be reduced to additive components. Thereby making the equation shift from:

Genes + Environment = Phenotype $\rightarrow$ Genes x Environment = Phenotype

His critique of Fisher's approach highlighted the "*lack of singularity",* i.e. it neither being a question of one or the other in defining the contributions of genes or environment. Using empirical examples, such as changes in environmental conditions in the growth of fruit flies, Hogben demonstrated how genotype-specific responses to environmental variation invalidated simple partitioning methods (77). He advocated for considering GxE as fundamental to understanding variability, challenging the frameworks of the time, and reaffirming the "*interdependence of nature and nurture*" (77,86).

Gene-environment correlations occur when individuals with specific genetic liability may actively seek out or create environments that align with their genotypes (87). This process introduces a bidirectionality between genetics and the environment, as people may find themselves in environs that upregulate genotypic expression into a certain phenotype and genotypes may correlate to the environs we seek out (88). There are three types of *r*GE: active, passive and evocative. Active *r*GE occurs when individuals actively seek out environments that align with their genetic tendencies, such as a child with a genetic liability for extroversion choosing social activities (87). Passive *r*GE arises from the overlap between the genes a child inherits, and the environment provided by their parents, such as a child of book-loving parents being exposed to book-rich environments (82). Evocative *r*GE is when an individual's genetically influenced traits elicit specific responses from others, such as a child's temperament evoking particular parenting styles (82). These interactions are not static, therefore how gene-environment correlations operate may change over time.

Gene-environment interactions are thought to shape psychological distress phenotypes (89,90). Individuals with a genetic liability to psychological distress may exhibit varying levels of vulnerability depending on their exposure to different environmental factors (91). Examining these interactions helps to understand how genetic liability may be expressed differently across cohorts, potentially influencing the observed cohort differences in psychological distress.

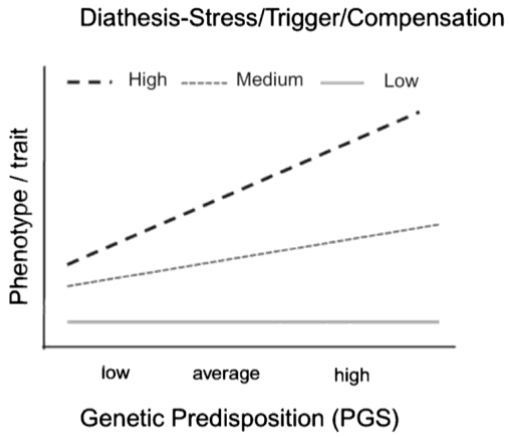
As the proportion of variation in psychological distress attributed to genetic factors depends on the population and its context, it is important to remember this can change over time as genes are expressed through and interact with the environment which is investigated in Study 1 (41). In understanding the potential cohort differences in the association of genetic factors with psychological distress as per Studies 2 and 3, various theoretical models have been proposed to explain the interplay between genetic liability and environmental influences that may be relevant.

There are five main theoretical models behind gene-environment interactions: the diathesis-stress model, the bioecological (social compensation) model, the differential susceptibility model, the biological sensitivity to context model and the social control model (92,93).

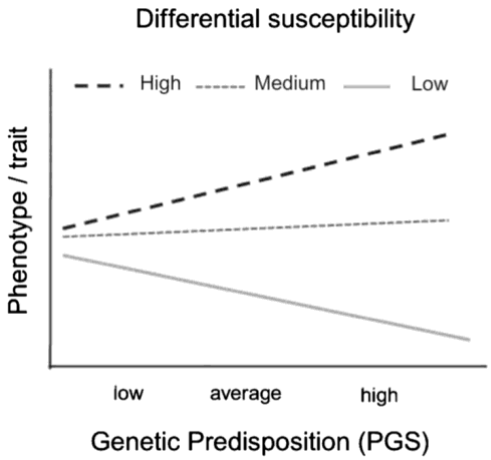
The diathesis-stress model posits that genetic liability for certain traits remain dormant until triggered by environmental stressors (94) (Figure 2a). Inspired by research by Caspi (2012 & 2002), this model suggests that individuals may exhibit varying levels of vulnerability to environmental stressors based on their genotype (95,96). Stressful life events tend to be negative experiences, which means genetic factors are triggered and the distress phenotype is expressed in this model. Therefore, in those without the negative environmental conditions, there is an absence of an association of genotype with phenotype.

Conversely, the bioecological or social compensation model suggests that genetic influences are maximised in stable, adaptive environments, particularly in higher socioeconomic contexts (97). A study summarising results from ten countries showed that populations and birth cohorts with higher social mobility shared higher heritability of education and lower environmental influences (98). Although the phenotype was not psychological distress, this study showed that changes in the wider societal context may impact the expression of the genotype to the phenotype. This result is reminiscent of the diathesis-stress model prediction for how genotype-environment interactions function.

a)



b)



c)

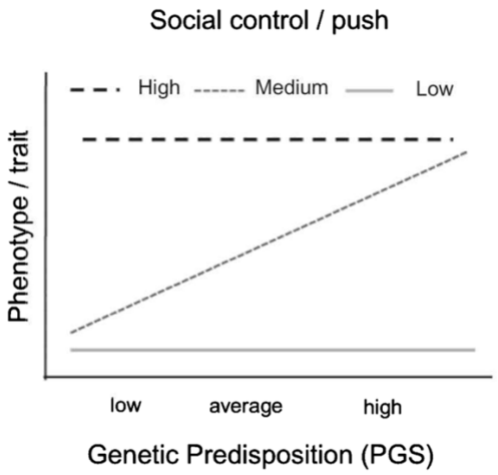


Figure 2 a-c: From “Figure 6.1 Gene-environment interaction models. Source: Adapted from Liu and Guo [38] in Mills, Barban & Tropf (2020) Introduction to Statistical Genetic Data Analysis” – lines indicate the characterisation of the environment.

The differential susceptibility model, proposed by Belsky and Pluess (2009), posits that individuals vary in their susceptibility to both positive and negative environmental influences based on genetic factors (Figure 2b) (91). While the diathesis-stress model concentrates on the adverse effects of the environment, the differential susceptibility model maintains that individuals vary in their levels of susceptibility, with some individuals more genetically sensitive to both positive and negative environments, while others remain resilient across all types of environments (91). This model challenges the traditional notion of vulnerability by emphasising individuals' plasticity and responsiveness to environmental conditions. The prediction is similar to the biological sensitivity to context model, which outlines how natural selection favours genotypes that exhibit a wide range of phenotypes in response to environmental conditions. This adaptability is crucial for species' survival (93).

Lastly, the social control model suggests that genetic factors may be filtered or buffered by social norms and structural constraints (Figure 2c) (99). A study conducted by Shanahan and Hofer in 2005 outlined four mechanisms by which social context may influence the significance of genetic factors (99). The environment may *“(a) trigger or (b) compensate for a genetic predisposition, (c) control the expression of a genetic predisposition, or (d) enhance a genetic predisposition (referring to the accentuation of “positive” genetic predispositions)”* (99). In predicting the relationship between genotype and environment interaction for psychological distress, it may be that stressful life events trigger distress; supportive social networks compensate and buffer against distress; sociocultural norm changes and wider societal stigma for distress stops the control or suppression of reporting distress while access to mental health services can enhance resilience to distress.

Each model predicts how genes and the environment interact for a trait differently. It may be that for a trait such as psychological distress, each is important in understanding the varied aetiology and nosology of distress. Therefore, these

theoretical models are not mutually exclusive in a population and often overlap in their explanations (100). Moreover, the precise mechanisms underlying gene-environment interactions remain largely under investigated, leading to what is often referred to as a "black box" explanation, i.e. the inner machinations of the process are opaque, while the input and output are visible. Nonetheless, these models provide valuable frameworks for understanding how genetic liability may interact with environmental factors to shape psychological distress outcomes.

The diathesis-stress model has largely been the dominant model in psychiatry (101). However, increasingly the differential susceptibility hypothesis has been tested (91). The hypothesis posits that individuals most adversely affected by negative environments may also be those most likely to benefit from positive environments, it is a more flexible theoretical approach to potential socioeconomic differences in psychological distress over time. The differential susceptibility hypothesis is closely related to the concept of biological sensitivity to context (93).

Biological sensitivity to context originates from evolutionary developmental biology, where natural selection favours genotypes that exhibit a wide range of phenotypes in response to environmental conditions. An example of this is research by Armitage et al 2024, where theory could be applied to explain the cohort difference phenomena between the 2001 and 1991 populations (102). They posited that loneliness is becoming more prevalent among younger populations, which could lead to increased biological stress reactivity, as it may be mediated by biological and social processes, including genetics (103). Therefore, it could be argued that the increased distress levels in younger cohorts as a direct response to the environment.

I used the social control model in Study 2 to test the role cohort may have in changing the relationship between polygenic scores for distress and observed psychological distress. In Study 3, I used the differential susceptibility framework to test whether there was an interaction between father's social class as a proxy for childhood socioeconomic circumstance and polygenic scores for distress. The different studies of my thesis and research questions are further explained in Section 2 below.

The gene-environment interplay debate has moved from the 20th century into the 21st century, with key developments confirming that the debate is beyond "versus" and rather relies on both. Finally, another key figure in the nature/nurture debates was Richard Lewontin. A query from his diaries summarises his position towards GxE: "*why is it that most geneticists do not understand that the phenotype is a developmental process?"* (p.134) (104). Consequently, time acts as the third partition:

$$\text{Genes x Environment} = \text{Phenotype} \rightarrow \text{Genes x Environment x Time} = \text{Phenotype}$$

This position was recently articulated by Boyce et al. 2020 (90). They reemphasised the importance of time in GxE, noting that timing—spanning critical developmental periods to evolutionary scales—shapes how genes and environments interact. The timing of environmental exposures, such as childhood trauma, influences the manifestation of traits like resilience or vulnerability, underscoring the significance of temporal windows in GxE research. Much of this framing lifts directly from concepts that are well-established and tested within lifecourse epidemiology (3).

1.3.5 Genetic contribution to distress over the lifecourse

As established, prior studies on the genetic contribution to psychological distress have traditionally used cross-sectional designs to explore between-individual variation. An extensive meta-analysis of 2,748 twin studies spanning 50 years identified increasing heritability with age, reaching a peak of 50% for distress at age 65 and older (42). This raises the question of whether the genetic contribution is stable or changes across the lifecourse and whether this trend remains across generations.

To answer this question Sallis et al. 2017 investigated whether the association between genotype and phenotype was consistent by age within the same individuals. They utilised the Avon Longitudinal Study of Parents and Children (ALSPAC) to identify adolescent distress symptom heritability within the same individuals across time. Their findings showed that heritability estimates ranged between 2-17%, fluctuating across adolescence up to 18, with a peak at age 13 (105). However, their study was limited by the narrow population chosen of only children and adolescents born in the southwest of England. Therefore, this study alone could not confirm stable genetic contribution as individuals develop. However, a recent systematic review synthesising 18 longitudinal behavioural genetic studies concluded that stability in psychological distress is likely attributed to genetic factors during childhood development through to adolescence (106). However, this systematic review was therefore also limited to childhood and adolescence.

Further research which spanned across the lifespan also demonstrated a similar trend of stable genetic contribution to psychological distress over time with data from the U.S. on individuals born between 1919 and 1950 (75). Another study that has tested heritability across age is Nivard et al. 2014, which utilised longitudinal data of 49,524 twins to assess the relative contribution of genetics and environment to distress across the lifecourse (107). The environmental estimate contribution increased up to adulthood in twins aged between 3 and 63 years in the study, with the genetic contribution remaining stable after adolescence (Figure 3 (b)) (107).

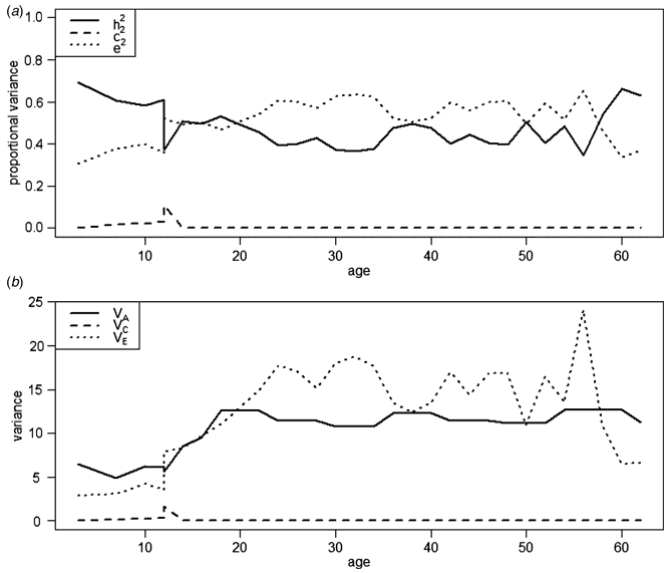


Figure 3: From Nivard et al. 2014 (107): “Figure 3. (a) Proportions of variance explained by genetic factors (h2, ——), comm2on environment (c2, - - -), and unique environment (e2, ····) at each age as derived from the model. (b) Variance components V_A, V_C, and V_E at each age as derived from the model.”

Their method relied on the ACE model of a twin design, which decomposes the variance in a trait into additive genetic factors (A), shared or common environmental influences (C), and unique environmental influences (E) (43). By studying the differences in these components between monozygotic (identical) and dizygotic (fraternal) twins, researchers can estimate the relative contributions of genetics and environment to observed variations in traits.

This study utilised structural equation modelling to analyse the mean trend, sex differences, and covariance structure of distress across time between twins. The stability of genetic and environmental effects were investigated by age via heritability estimates derived at each age and genetic and environmental correlations calculated between ages. Further studies replicated these results with stability mainly attributable to additive genetic factors, whereas change was related to environmental influences (108–111). However, there has been no synthesis of these results. Furthermore, this study was limited by its use of the twin design which may be overestimating the contribution of additive genetic factors due to equal environments assumption.

The temporal stability of genetic contributions to phenotypes across the lifecourse remains insufficiently investigated. Age-dependent genetic variations could manifest through uniformly stronger or weaker effects of the same genes, activation of entirely different genes at distinct developmental stages, or non-uniform changes in effect magnitude of identical genes over time. These potential variations in genetic influence have important implications for research methodologies employing polygenic score (PGS) by age interaction analyses, which this thesis addresses.

1.4 Psychological distress across the lifecourse: Lifecourse Epidemiology

Lifecourse epidemiological theories attempt to explain how exposures affect outcomes over time (3). These include critical or sensitive period models, accumulation of risk models, and chain of risk models (3). Socially critical periods in human development include transitions in schooling, entry and exit to the labour market, leaving the parental home, establishing a home, cohabitation, marriage or divorce, transition to parenthood, job insecurity, economic inactivity, and the onset of chronic illness. Change in lifecourse psychological distress can be explained by several theories. However, for the purposes of this thesis there is one lifecourse approach that is utilised to help conceptualise the environment. This is the Bioecological Model (Bronfenbrenner, 1999), which emphasises the influence of multiple layers of environmental systems on human development, from immediate surroundings to broader societal contexts (97,112).

All three studies utilise a lifecourse epidemiology approach, since the developmental timing of genetic and environmental influences is relevant to the understanding of psychological distress. Nevertheless, each study has a different primary theoretical

lens. Study 1 applies Baltes' Selection, Optimisation, and Compensation (SOC) model to investigate gene-age effects in the 1958c (1). Previous paragraphs in this section outlined pertinent lifecourse theories of how the environment is conceptualised, particularly Bronfenbrenner's distinction between distal and proximal influences (2). In this framework, Study 2 treats birth cohort membership as a distal environment and uses social control theory to conceptualise it (3). Study 3 is also a gene x environment study, where father's social class is a proximal environmental factor and uses the differential susceptibility hypothesis to conceptualise it (4).

The Selection, Optimisation, and Compensation (SOC) model holds that biological potential and genome-based plasticity decline with age, while individuals increasingly rely on cultural and social resources to maintain functioning as they age (1). Baltes et al. (1997) state: "*With age, the genetic material, associated genetic mechanisms, and genetic expressions become less effective and less able to generate or maintain high levels of functioning. Evolution and biology are not good friends of old age*." If applied to psychological distress, it implies that the genetic contribution to psychological distress should diminish with age, as environmental compensation plays a larger role. However, the present study does not assess total genetic influence, but rather the observable effects of a polygenic score that captures only a part of common liability to distress. Additionally, developmental theories of gene-environment correlation suggest that genetic effects may intensify with age. SOC offers one plausible theoretical prediction rather than the only expected trajectory. Therefore, the theory, when applied here, would hypothesise improved mental health in later life, proposing that individuals become more skilled at adapting and compensating.

While SOC is typically applied to functional ageing rather than genetic variance, its premise that biological reserves and genome-based plasticity decline across adulthood provides a rationale for testing changes in the expression of genetic effects. If compensatory environmental resources increasingly buffer distress with age, observable polygenic effects should attenuate. In this study, I therefore treat PRS-age interaction as an empirical test of whether compensatory social and cultural resources increasingly mask genetic influences on distress as individuals age.

The Social Control theory posits that social institutions, norms, and structural conditions regulate behaviour and health outcomes across historical time (3). If genetic effects on psychological distress differ between cohorts, this would suggest that shifting societal contexts moderate the expression of genetic influences. In Study 2, I therefore treat PRS-cohort interaction as an empirical test of whether societal contexts moderate the expression of genetic influences between cohorts.

The differential susceptibility hypothesis proposes that some individuals are more susceptible to negative and positive environmental influences due to their genetic liability and early-life experiences (4). In this thesis, the effect of genetic propensity for distress may vary by advantaged versus disadvantaged fathers' social class, with higher genetic scores amplifying the impact of both advantaged and disadvantaged environments. . In Study 3, I therefore treat PRS-SES interaction as an empirical test of whether genetic propensity for distress may vary by advantaged versus disadvantaged fathers' social class.

Both the social control theory and the differential susceptibility hypothesis are explored in more detail in Section 1.3.4 of the Introduction where theories of gene-environment interplay are described.

It may be that all these theories go some way to explain average population lifecourse processes. The current thesis is focused on testing the *environmental influences and susceptibility* lifecourse epidemiology theories. To summarise the evidence, an individual's distress level is not stable over their lifetime, but there are population trends in onset and prevalence with various theories underpinning the lifecourse development of distress (114–116).

Table 1: Summary of primary theoretical frameworks and hypothesis for each study

Study	Primary theoretical framework	Hypothesis
1	Baltes' Selection, Optimisation and Compensation (SOC) model	The effect of the polygenic score for psychological distress will diminish with age**,** reflecting age-related declines in biological plasticity and increasing reliance on compensatory social and cultural resources.
2	Social Control Theory	Genetic effects on psychological distress will differ by cohort, reflecting historical differences in distal social environments. Specifically, variation in social norms and institutional structures across cohorts is expected to moderate the expression of genetic influences.
3	Differential susceptibility hypothesis	Genetic effects on psychological distress will be moderated by paternal social class as a proximal environmental factor, with individuals carrying higher genetic scores showing greater sensitivity (for better or worse) to socio-economic conditions.

The analysis of age, period, and cohort (APC) effects is a challenging methodological issue in lifecourse epidemiology. The central difficulty lies in the fact that age, period, and cohort are perfectly collinear (Age = Period – Cohort), making it mathematically impossible to estimate their independent linear effects without imposing strong constraints or assumptions (5). This issue has sparked considerable controversy. On one side, Yang and colleagues developed hierarchical APC models that attempt to model all three dimensions simultaneously by introducing random effects at the cohort and period level (6). On the other side, Bell and colleagues argue that such models cannot truly resolve the identification problem and instead rely on untestable assumptions that may produce misleading inferences (7). Recent contributions have emphasised that researchers must be explicit about their research goals and the assumptions underpinning APC modelling (8). For applications to psychological distress research, one approach may be to embrace the entanglement of APC processes—recognising that ageing always occurs within historical time—rather than to attempt strict decomposition. This perspective shifts emphasis from "solving" the identification problem to framing whether age, cohort, or period patterns are most relevant to the research question at hand.

Due to the identification problem, a pattern interpreted as an age or cohort effect could alternatively reflect a shared period influence. While it is therefore not possible to fully eliminate period-based interpretations, in this thesis, period effects are not treated as the primary explanatory focus for two reasons. First, the study period does not contain a single, discrete historical disruption (e.g. policy shock or natural disaster) that would plausibly affect all age groups simultaneously in a way that could explain the observed trend. Second, while population-wide shocks can produce short-term spikes in psychological stress, their effects are often transient or unevenly distributed. In contrast, age and cohort frameworks capture long-term developmental and generational patterns, which are more analytically suited to this context.

The following sections outline age, cohort and period effects in succession.

1.4.1 Age Effects

The age effect refers to the alteration in an outcome that occurs uniformly across cohorts as they grow older (3,122). In contrast, a cohort effect is a phenomenon that characterises individuals born at a specific point in time and remains unaffected by the ageing process (122). Lastly, a period effect is a change that occurs at a particular moment in time, affecting all age groups and cohorts uniformly (122). The exact collinearity among these three variables (Age = Year - Birth Year) makes it challenging to estimate their effects; it is impossible to estimate linear components of these effects without strong assumptions about at least one.

The onset of distress can occur at any age, although typical patterns are observed in the general population. In the United Kingdom (U.K.), the mean of distress scores is shown to peak during early adulthood, remain stable during adulthood, increase again in mid-life, and then decline in later years (Figure 4). It has been shown that in mid-life there is an increase in distress severity, which was confirmed using other longitudinal U.K.-based datasets and studies (123–128). Previous research has shown that when the overall trend in the average population is examined in more detail, it reveals well-known trajectories (129). These have been previously identified as stable high,

increasing, and stable low (115). It is important to note that some people have the opposite trajectory to the average, where they experience good mental health in adolescence and mid-adulthood but suffer the most during early adulthood (130).

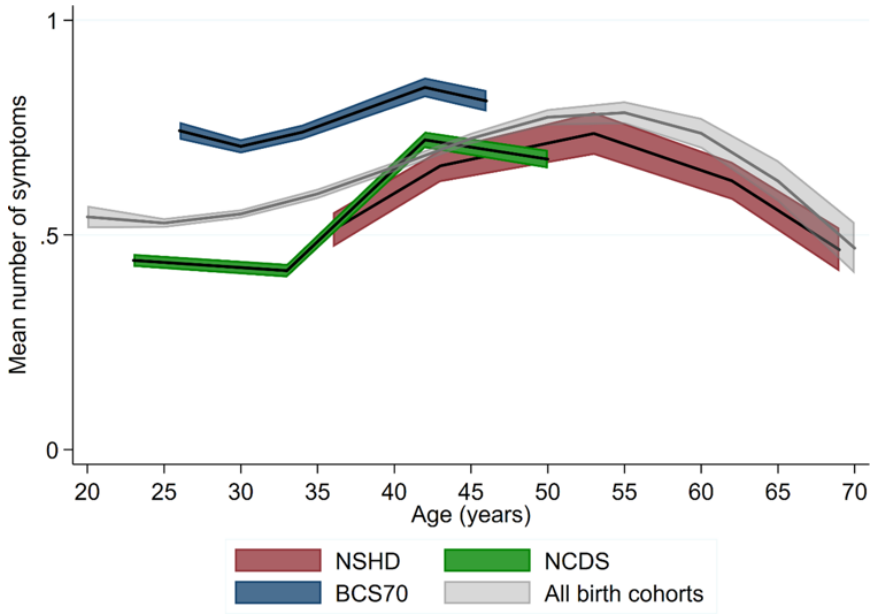


Figure 4: From Gondek et al. 2021 (124): "Figure 2. Age profile of the mean number of psychological distress symptoms – cohort-stratified and pooled across cohorts."

While age effects shape the patterns of psychological distress across adulthood, it is not chronological age alone that drives these changes; rather, they reflect a complex interaction of life-stage-related factors. In early adulthood, transitions such as entry into the workforce, relationship formation, and establishing independence can introduce stressors that elevate distress levels (131). Midlife often brings unique challenges, such as career pressures, caregiving responsibilities, and potential health declines, all of which contribute to a documented increase in distress severity (132). Later in life, factors like retirement, social isolation, and physical health concerns can influence psychological distress, though some theories, like the Socioemotional Selectivity Theory, suggest that enhanced emotional regulation may reduce distress in older adulthood (133,134). These factors highlight how age serves as a proxy for the developmental and environmental contexts that individuals encounter at different life stages, shaping mental health trajectories over time.

Study 1 of this thesis examined whether the relationship between polygenic score for psychological distress and observed psychological distress varies across different ages. This offers an original contribution to the existing body of research by focusing on genetic contribution as a dynamic factor that may alter with age-related changes in psychological distress. Unlike prior studies that have primarily focused on environmental or psychosocial influences on distress across adulthood, this study integrates genetic data to explore how polygenic scores are associated with distress at different ages across adulthood.

Prior research established the age effects of psychological distress following a U-shape across the lifespan (Figure 4) (126,135,136). This followed that people became increasingly distressed in middle age with increasing mean scores, which tapered off

as they reached older age. Further research with an extended age range found that the U-shape pattern more resembled an S-shape (123). By investigating these potential age-related variations in genetic contribution, Study 1 provides insights into whether polygenic influences on distress follow a pattern consistent with observed age-related distress trends. However, more recent research utilising younger cohorts has disproven this observed trend, finding that younger cohorts have higher mean scores at earlier age points (102,137). This introduces the question as to whether cohort-related variations in the genetic contribution to psychological distress.

1.4.2 Cohort Effects

Cohort effects refer to variations in an outcome specific to groups of individuals born around the same time, influenced by shared social, environmental, and economic conditions during their formative years (3). Cohort-specific psychosocial stressors are tied to unique societal conditions that affect one generation differently, and do not affect older generations similarly at comparable ages (128,138). For example, younger generations have high levels of educational debt compared to the relatively debt-free education experienced by earlier generation (139). In contrast, age effects would refer to changes common to people of a certain age across cohorts (e.g., career pressures in midlife or increased health concerns in later years).

Policy changes and economic conditions influence cohorts differently. For example, people entering the workforce during an economic recession may face long-term job instability, whereas another cohort entering during a period of growth may experience more stability (140). These effects are specific to the time the cohort came of age rather than to an age group itself. While certain experiences, like entering the workforce, are age-graded, the context in which they occur differs by cohort. For example, midlife pressures in a cohort that experienced significant economic growth may differ from those in a cohort facing high housing costs and stagnant wages, leading to different levels and sources of distress (138,141).

Explanations for cohort differences and the rising rates of distress during adulthood in the UK have been credited to a range of determinants, both distal and proximal that may be specific to each cohort. On an absolute scale, living standards have improved over the past 150 years. An example of this is in 1860, in England, workers averaged 66 hours per week and had a life expectancy of 40 (142,143). By 2010, on average, people in the U.K. worked 31 hours and had a life expectancy of 80 (142,143). For instance, comparing cohorts born in 1958 and 1970 in the United Kingdom reveals marginal differences in key socioeconomic indicators at birth. The Gini coefficient grew slightly from 25 to 28, and unemployment increased by 0.5% (144,145). As absolute living standards rose between the cohorts, life expectancy grew by two years from 70 to 72, and nearly double of adolescents were in full-time education (142,146).

Despite overall improvements in living standards, the economic position of the lowest-income groups has deteriorated compared to the rest of the population. Between 1970 and 1990, the proportion of individuals in households earning below 60% of the median income doubled, reflecting a growing income gap that has significantly impacted the most economically vulnerable populations (147). This increasing economic inequality has manifested as a polarisation within cohorts, where wealth and

access to resources are increasingly concentrated among the more affluent, leaving lower-income groups further behind.

Polarisation within cohorts leads to a steeper socioeconomic gradient within each generation, meaning that the economic differences within a single cohort have become more pronounced over time (p.148) (139). The cumulative disadvantages associated with low socioeconomic status—such as limited access to higher education, healthcare, and job opportunities—compound over a lifetime, impacting psychological distress. Additionally, within-cohort inequalities may exacerbate psychological distress by amplifying social comparisons and reducing social cohesion, factors that are known to influence psychological distress (7,148,149).

Study 2 of this thesis investigates whether the polygenic score for psychological distress varies across cohorts, addressing a gap in current research by examining how genetic contribution may be moderated by the distinct socio-economic and broader environmental contexts experienced by different birth cohorts. This study is novel in that it compares whether the expression of polygenic scores varies between cohorts, potentially due to shifting societal conditions or genetic factors. It extends beyond age-specific effects to consider how cohort differences might shape the manifestation of psychological distress in individuals.

Additionally, Study 3 investigates the interaction between the polygenic score for psychological distress and early-life socioeconomic circumstances, as measured by father's social class. This study is original in its aim to test cohort and sex differences in the GxE between polygenic scores and father's social class across adulthood. By examining these interactions, the thesis provides a more comprehensive understanding of how psychological distress is influenced by polygenic score, cohort-specific environmental conditions, and time.

1.4.3 Period Effects

Period effects refer to changes in outcomes that are attributable to the specific calendar period or historical time in which individuals are observed, irrespective of their age or birth cohort (9). This necessitates the assumption that psychosocial stressors are tied to a unique period where societal conditions affect all generations in the same way. For example, a global pandemic may or may not increase the psychological distress of an entire population across all age groups, irrespective of (10,11).

Although this thesis's findings were interpreted using a age-cohort lens, it is possible that unmeasured period effects could contribute to the observed patterns. However, in this thesis, period effects are not explicitly modelled. This is primarily due to it not being the main research question of interest. Period effects are best assessed when a clear natural experiment or policy discontinuity exists—using, for instance, difference-in-differences or regression discontinuity designs—but no such discrete, identifiable period exists in the context of this study to justify that analysis. Instead, this thesis focuses on age and cohort effects as reflections of broader social contextual shifts over time, a perspective aligning with cohort analysis that embraces the intertwinement of age-period-cohort processes, thereby assuming no long-term period trend (5,8,12).

1.4.4 Gender/Sex effects

There are gender differences in mental health observed across many populations, known as the gender gap in distress (13). Across multiple countries, women tend to report higher psychological distress levels than men (14–16).

There are multiple theories that attempt to explain why women could experience higher psychological distress compared to men. The biopsychosocial model for the gender gap in psychological distress emphasises that no single mechanism explains phenomena sufficiently, but rather the interplay of biological, psychological, and social mechanisms that work together (17,18). Across these domains, (i) biological/sex-linked factors include neuroendocrine dynamics across puberty, perinatal periods and perimenopause, differences in HPA-axis reactivity, and immune–inflammatory pathways (19,20); (ii) psychological processes include differences in emotion regulation, rumination and stress appraisal (18); and (iii) social and structural factors include gendered exposures to violence, discrimination, caregiving burden, work–family conflict and economic precarity (21). These domains interact over the life course. Through the lens of lifecourse epidemiology, the predominant explanation is the cumulative disadvantage theory, which posits that early life advantages or disadvantages shape an individual's trajectory over time, leading to differential outcomes in later life (22,23). Women face greater structural strain due to gender inequalities, which accumulates as a form of life course disadvantage that may impact psychological distress outcomes (24).

As outlined, one mechanism may be a biological basis for sex differences in distress (19,20,25). Prior evidence investigated biological sex differences contributing to the female gap in distress outcomes using GWAS (26,27). The most recent and largest sex-stratified meta-analysis GWAS of depression found more genetic variants contribute to MDD risk in females (~13,244 variants) versus males (~7,111 variants) meaning greater polygenicity in females (28). There is a potential for gene-environment entanglement here. As if women are exposed to cumulative disadvantage to then express the phenotype, it is not the XX chromosome giving rise to SNP variants that contribute to polygenic distress burden. Rather, the lived experience of social and structural inequality across the lifecourse may interact with biological systems in ways that magnify the genetic signal observed in women, such that polygenicity could reflect the interplay between social exposures and genomic architecture rather than a purely intrinsic sex-linked difference.

However, it has been shown that the female gap in mental health is not ubiquitous. A study identified the female gap in positive mental health was not present in 26 out of 73 countries indicating variation by cultural context (29). The same female adolescents were found to have higher psychological distress across all 73 countries. There is evidence for both social and biological factors driving differences in distress levels between genders, and these factors may be context specific and thus may differ by time and place. Therefore, it would be pertinent to test for gender differences as supplementary analyses in each of the empirical chapters of this thesis.

1.5 Social Gradient of Psychological Distress

Socioeconomic status (SES) and socioeconomic position (SEP) are terms used to describe an individual's or group's standing within a society's stratified social structure (167,168). SES is typically current social and economic conditions (e.g., education, income, occupation), whereas SEP encompasses broader, lifelong dynamics, including historical and structural factors (169). Theoretical foundations for these concepts come from sociologists who focused on the acquisition and distribution of capital and status in society – including Karl Marx, Max Weber, and Pierre Bourdieu (170–172).

Marx emphasised the structural relationship between individuals and the "*means of production*" focusing on class conflict and material inequality as defining features of societal stratification (171). Weber expanded this framework, arguing that stratification occurs across multiple dimensions, including class, status, and power, and that individuals actively shape their “life chances” through skills and social capital (172). Bourdieu introduced the concept of "capital" in its various forms—economic, social, and cultural—to explain how individuals and groups maintain and reproduce social advantage (170). Together, these theories highlight the multifaceted and dynamic nature of SEP, its material and symbolic components.

SEP is thought to contribute to health disparities through two main mechanisms: social causation and health selection (173). The social causation hypothesis posits that socioeconomic disadvantage exposes individuals to adverse environmental conditions that lead to poorer health outcomes (173). These exposures include material deprivation (e.g., poor housing, inadequate nutrition), psychosocial stress, and poor access to education. In contrast, the health selection hypothesis posits that health influences SEP, with poor health limiting educational attainment, employment opportunities, and income potential (173). For example, chronic illness or disability in childhood can restrict academic performance, resulting in lower occupational status and income in adulthood. While both mechanisms operate to some extent, for health outcomes like psychological distress, where the cumulative effects of disadvantage persist across the lifecourse, evidence supports social causation as the primary driver of health inequalities (174,175).

There is a wide body of evidence that demonstrates a social gradient in psychological distress, where individuals in lower socioeconomic positions report higher levels of distress than those in higher positions (4,5,176). This gradient persists across adulthood, reflecting both material and psychosocial inequalities (177). Financial insecurity is associated with elevated risk of psychological distress (176,178,179). Psychosocial stressors such as social isolation, discrimination, and perceived lack of control further exacerbate distress as part of socioeconomic disadvantage (12,180,181). Longitudinal studies highlight the cumulative effects of disadvantage, showing that chronic exposure to adverse conditions amplifies distress over time (182). Importantly, this gradient is not static; it evolves across the lifecourse, with early life shaping adult mental health outcomes and adult circumstances exerting additional, compounding effects (3,11,183).

The relationship between socioeconomic circumstances and psychological distress may be influenced by cohort and age effects (3,123,183). Cohort effects capture how

specific generations' social and economic conditions shape their mental health outcomes. For example, individuals born during economic hardship or limited access to education may experience a steeper social gradient in distress than later-born cohorts who benefit from improved living standards (3,122).

Emerging research has identified that, as with other complex behavioural traits, socioeconomic status has genetic underpinnings (184,185). However, it should be clear that environmental factors mediate the genetic contribution, as individuals with similar genetics can achieve vastly different socioeconomic outcomes depending on their social contexts. For example, access to advantageous environments can amplify or mitigate genetic factors. Importantly, this research does not suggest that socioeconomic disadvantage is biologically determined; rather, it underscores the complex interplay between genetics and social environments in shaping life chances (186).

Study 3 examines the interaction between polygenic scores for psychological distress and early-life socioeconomic circumstances, as measured by the father's social class. This study is original in its dual focus on genetic vulnerability and socioeconomic background across generational contexts, addressing how early-life conditions unique to each cohort might amplify or mitigate the expression of genetic liability to psychological distress. By exploring these interactions, the thesis offers a more comprehensive understanding of how polygenic score for psychological distress are influenced not only by age but also by cohort-specific environmental conditions.

Environments, when conceptualised through age, cohort, and father's social class, offer a developmental framework for understanding the interplay between socioeconomic factors and psychological distress. Drawing on developmental theory from Hogben and Lewontin, this perspective emphasises how shared environments (e.g., familial socioeconomic conditions like father's social class) and non-shared environments (e.g., individual experiences unique to individuals) interact dynamically with genetic contribution to shape mental health outcomes

2 Research Questions

Study 1: Age differences in the association between polygenic scores and psychological distress: evidence from the 1958 British Birth Cohort

Main objective: to examine whether the polygenic score for psychological distress is associated with phenotypic psychological distress at ages 23, 33, 42 and 50 in the 1958 National Child and Development Study (1958c).

Research Question 1: Does the magnitude of association between polygenic score and adult psychological distress outcomes differ from age 23 to 50?

Research Question 2: Does the amount of variance explained in adult psychological distress outcomes by polygenic scores differ from age 23 to 50?

Study 2: Cohort differences in the association between polygenic score and psychological distress: evidence from the 1958 and 1970 British Birth Cohorts

Main objective: to examine whether the polygenic score for psychological distress is associated with phenotypic psychological distress using the 1958c from the previous chapter and the 1970 British Cohort Study (1970c).

Research Question 1: Does the association between polygenic score and adult psychological distress outcomes differ at all age points in the pooled 1958c and 1970c?

Research Question 2: Does the association between polygenic score and adult psychological distress outcomes differ by cohort?

Research Question 3: Does the association between polygenic score and adult psychological distress outcomes differ by sex?

Study 3: Investigating gene-environment interplay between polygenic score and father's social class on adult psychological distress during adulthood

Main objective: to examine whether father's social class interacts with polygenic scores in its association with adulthood psychological distress in the 1958c and 1970c.

Research Question 1: Does the association between polygenic scores for distress and adult psychological distress outcomes vary according to father's social class?

Research Question 2: Do the independent associations or interactions between polygenic score for distress and father's social class on adulthood psychological distress differ by cohort or by sex?

Research Question 3: Using meta-analysis, is there robust evidence of an interaction between polygenic score for psychological distress and socioeconomic circumstances on adulthood psychological distress outcomes?

3 Description of the Data and Analytical Strategies

3.1 Introduction to the Data

3.1.1 1958 National Child Development Study (1958c)

The 1958 National Child Development Study is a nationally representative longitudinal birth cohort study following the lives of 17,415 individuals born in a single week of March 1958, with 9,100 in the data collection of 2013-14 at age 55 (187). 1958c is nationally representative of the population born in England, Scotland, and Wales in that year. Immigrants born in the reference week were added to the target sample at ages 7, 11 and 16, which resulted in a total sample of 18,558 cohort members. Its original aim was to investigate factors associated with positive and negative health outcomes among mothers and children. There have been nine points of data collection at birth and from ages 7, 11, 16, 23, 33, 41-42, 46-47, 50, 55, and the next survey of the 1958c will include data collected between 2020 and 2023. Data are collected using a mix of face-to-face interviews, and telephone interviews with a biomedical data collection at age 44-45, and another ongoing for age 62.

3.1.2 1970 British Birth Cohort Study (1970c)

The 1970s British Birth cohort is a nationally representative longitudinal birth cohort study of children born in a single week in 1970 in England, Scotland, and Wales (188). Immigrants born in the reference week were added to the target sample at ages 5, 10 and 16. This resulted in 79 new participants at age 5, 294 at age 10 and 65 at age 16. Data collection was conducted at birth and then at ages 5, 10, 16, 26, 30, 34, 38, 42, 46, and 51 between 1970 and 2022 (188). Data collection has been done via face-to-face interviews, postal surveys, and a biomedical data collection via a nurse visit that occurred at ages 46-48 (188).

3.1.3 Sample Size

Inclusion criteria for the samples used throughout the thesis necessitated individuals who contributed data for at least one time point for the malaise inventory score, who were of European ancestry and who had genetic data for a polygenic score to be derived.

Table 2: Summary of Sample Sizes

Study	Total for main analysis	Sensitivity (complete case)	Cohort-stratified	Sex-stratified
1	6312	1958c: 3748	N/A	Female=3172 Male=3140
2	10713	1970c: 2446	1958c=6312 1970c=4401	Female=5259 Male=5454
3	8923	N/A	1958c=5242 1970c=3681	Females=4300 Males=4623

3.1.4 Genetic Data

In the 1958c, genetic data were collected from the biomedical survey between 2002 and 2004 at age 44-46 blood samples were taken from 9,377 individuals (187). The quality-controlled imputed set of genotypes contained 7,545,708 variants and 6,396 individuals (genome build: GRCh38, imputation with TOPMed). In the 1970c, genetic data were collected during the age 46 biomedical data collection, blood samples were taken from 8,581 individuals, and genetic data collected from 5,905 samples (188). The quality-controlled imputed set of genotypes contained 8,640,849 variants and 5,598 individuals (genome build: GRChr38, imputation with TOPMed).

Both the 1958c and 1970c utilise the same quality control process. The benefit of this harmonisation enables a more robust comparison. Quality control of the 1958c and 1970c genetic data was conducted using PLINK 1.9/2.0 and R v 4.1.2. Table 3 details the removal of individuals and SNPs depending on the process of quality control outlined below. More information about genetic data collection and quality control can be found at: https://cls-genetics.github.io/docs/NCDS.html, https://cls-genetics.github.io/docs/BCS70.html, Shireby et al. 2024 and Bridges et al. 2023 (189,190). Access to (quality controlled and imputed) genetic data is granted via formal application, from the U.K. Data Service (191).

Quality Control

Quality control removed individuals if they met the following criteria:

- >2% missing data – individuals with over 2% missing genotype data are removed. Missing data can lead to biased and less reliable results, affecting statistical power and imputation quality.
- Discordant sex between predicted genetic sex and sex reported in the phenotypic data, excluding females with an F value > 0.2 and males with an F value < 0.8.
- Excessive heterozygosity – individuals with outlying rates of heterozygosity as defined by more than 3 standard deviations from the mean are removed.
- Closely related individuals in the sample (king-cutoff 0.0884), where one individual from each pair of related samples was excluded based on the King greedy related algorithm.
- Non-European ancestry as determined by merging the 1958c combined genotypes with data from 1000 genomes Phase 3, linkage disequilibrium pruning the overlapping single nucleotide polymorphisms (SNPs) such that no pair of SNPs within 1000 bp had $r2 > 0.20$ and inspecting the first genetic principal components to assess using a threshold of >4 standard deviations from the mean to exclude outliers i.e. non-European ancestries.

Quality control removed SNPs if they met the following criteria:

- Minor allele frequency of <1% – SNPs occurring in less than 1% of the population are excluded because they are rare, more prone to genotyping errors, and have reduced power for identifying SNP-phenotype associations.

- Hardy-Weinberg equilibrium (HWE) of P <$1x10^{-6}$- deviation from HWE is a good indicator of genotyping error, and can indicate evolutionary selection, whereby genetic drift, selection pressures or migration influence the allele frequencies in a population.
- >3% missingness – SNPs which are missing in over 3% of individuals are removed.

Table 3: Genetic data quality control summary of the removal of individuals and SNPs for the 1958c and 1970c

	1958c	1970c
>2% missing data	0 individuals	136 individuals
Discordant sex	0 individuals	15 individuals
Excessive heterozygosity	0 individuals	46 individuals
Closely related individuals	23 individuals	35 individuals
Non-European ancestry	72 individuals	83 individuals
Minor allele frequency of <1%	0 SNPs	0 SNPs
Hardy-Weinberg equilibrium deviation	60 SNPs	0 SNPs
>3% missingness	0 SNPs	0 SNPs

3.2 Measures

3.2.1 Psychological Distress

Across both 1958c and 1970c multiple measures of mental health were used, including the Malaise Inventory Score, Short Form Health Survey, the Warwick Edinburgh Mental Wellbeing Scale, the Generalised Anxiety Disorder 2-item and the Patient Health Questionnaire (192–196). This thesis used the malaise inventory score as it was the closest in conceptualisation to psychological distress that was also most repeatedly measured and therefore the most comparable between the two birth cohorts. The Malaise Inventory is used across multiple British Birth Cohorts: the 1970c, the 2000 Millennium Cohort study, and 1958c. It has also been tested and validated in different populations, including in New Zealand (197).Thereby strengthening its utility in mental health research.

The Malaise Inventory was developed by Rutter and colleagues in 1970 and derived from the Cornell Medical Index Health Questionnaire (192,198). The questionnaire consists of 24 questions pertaining to the participants' current or recent state. These questions require a simple "yes" or "no" response and do not require participants to recall their symptoms or state within a specific period, such as the last two weeks. Scores then range from 0-24 in the full format. In case of missing items, prorated imputation was conducted if up to 2 items were missing for the 9-item score in both cohorts. This adheres to prior prorated imputation methods; imputation is only done if no more than 20% of the items are missing. For nine items, that would be no more than 2.8 missing was rounded down to 2. Missing items are imputed from the individual’s mean answer to the observed items.

The Malaise Inventory measures psychological distress from items that capture both somatic and genitive-affective symptoms, which are characteristic of depressive symptoms (199). The measure was developed as a unidimensional construct. The Malaise Inventory was previously found to be reliable in 1958c participants by having fair internal consistency with a Cronbach’s Alpha of a=0.77 at age 23 and a=0.80 at age 33 (199); 0.70 to 0.80 is deemed good (200–202). At older ages, it has stronger internal consistency compared to earlier time points; therefore, it may better capture the construct at these ages. In this thesis, both cohorts have a similar trend in internal consistency with stronger internal consistency in older time points compared to earlier. However, in 1958c alpha is initially larger than 1970c and then switches at the later ages to 1970c, having marginally larger Cronbach’s Alpha values (Table 3). Yet, the confidence intervals of the Cronbach’s Alpha estimate overlap, implying comparable reliability of the Malaise Inventory across cohorts at similar ages.

The Malaise Inventory score’s external criterion/concurrent validity was tested by ROC analyses between psychiatric morbidity diagnoses within the 1958c and the Isle of Wight Epidemiological Survey (199). For the 24-item Malaise score, the area under the curve was 0.79 between ages 16-23 and 0.77 for ages 23-33, which are considered fair values (199). In another study by Hirst et al., the authors assessed criterion/concurrent validity using Goodman and Kruskal's gamma with two other measures of stress: (1) the symptom scale and (2) the use of medication scale. They found a positive moderate association between the malaise inventory score and the

symptom scale at 0.57 and 0.50 for the medication scale (203). Measurement invariance of the malaise inventory was previously tested and showed scalar invariance across time within the same population and between cohorts (204). This means that individuals across different time points and birth cohorts interpret and respond to the items in a comparable way, allowing for valid comparisons of mean levels of psychological distress. For the purposes of this thesis, scalar invariance supports the assumption that differences in malaise scores across age and between cohorts reflect true differences in distress, rather than differences in how the measure functions across groups or time.

Table 4: Summary of the Cronbach's Alpha for 9-item Malaise Inventory Score at ages 23-50 in both the 1958c and 1970c

1958c		1970c	
Age	Cronbach's Alpha (95% CI)	**Age**	Cronbach's Alpha (95% CI)
23	0.71 (0.69-0.72)	**26**	0.70 (0.68-0.71)
33	0.76 (0.74-0.77)	**30**	0.73 (0.71-0.75)
42	0.76 (0.75-0.77)	**34**	0.77 (0.76-0.78)
50	0.80 (0.79-0.81)	**42**	0.76 (0.74-0.78)
		46	0.82 (0.80-0.83)

Psychological distress was measured using the Malaise Inventory as the 24-item at ages 23, 33, 42 in 1958c and ages 16, 26, and 30 in the 1970c. The reduced 9-item was given at age 50 in the 1958c and from age 34 onwards in the 1970c. The 9-item version is a subset of the 24-item version (Table 5). The questionnaire asks a variety of items about whether you have experienced a variety of somatic and cognitive symptoms. This thesis utilised the 9-item version sum scores for all analyses as the outcome variable, as this enabled the inclusion of age 50 in the 1958c and ages 34-46 in the 1970c.

Aligning items across instruments shows that the Malaise Inventory shares considerable overlap with the depressive and anxiety symptom domains assessed in the GWAS (e.g., low mood, worry and anxiety). At the same time, the Malaise includes items on irritability, which extend beyond the constructs captured in the GWAS measures. This partial overlap implies that our results should not be interpreted as reflecting a one-to-one correspondence with depressive symptom GWAS outcomes. Instead, associations with the Malaise may index a broader psychological distress (low mood/anxiety) construct. This could attenuate correlations relative to what might be observed with perfectly aligned measures, but the direction of any bias is uncertain. Consequently, effect sizes should be interpreted with caution, and future work using a GWAS utilising better harmonised measures would help clarify the degree to which construct differences shape observed genetic associations.

Table 5: Malaise Inventory Score items used at different ages in the 1958 and 1970c, responses were asked "yes" or "no" for presence of symptoms

Item	Symptom	Questions	9-item (ages 23, 33, 42, 50 in 1958c)	24-item (ages 23, 33, 42 1958c, ages 34+ 1970c)
1	Backache	Do you often have backache?		X
2	Fatigue	Do you often feel tired most of the time?	X	X
3	Low Mood	Do you often feel depressed?	X	X
4	Headache	Do you often have bad headaches?		X
5	Worry	Do you often get worried about things?	X	X
6	Sleep	Do you usually have great difficulty in falling or staying asleep?		X
7	Awake	Do you usually wake unnecessarily early in the morning?		X
8	Health Anxiety	Do you wear yourself out worrying about your health?		X
9	Rage	Do you often get into a violent rage?	X	X
10	Annoy	Do people annoy and irritate you?		X
11	Twitching	Have you at times had a twitching face, head, or shoulders?		X
12	General Anxiety	Do you suddenly become scared for no good reason?	X	X
13	Situational Anxiety	Are you scared to be alone when there are not friends near you?		X
14	Irritable	Are you easily upset or irritated?	X	X
15	Social Phobia	Are you frightened of going out alone or meeting people?		X
16	Tension	Are you constantly keyed up and jittery?	X	X
17	Indigestion	Do you suffer from indigestion?		X
18	Stomach	Do you suffer from stomach upset?		X
19	Appetite	Is your appetite poor?		X
20	Nerves	Does every little thing get on your nerves and wear you out?	X	X
21	Panic	Does your heart race like mad?	X	X
22	Eye Pain	Do you often have bad pain in eyes?		X
23	Rheumatism	Are you trouble with rheumatism?		X
24	Breakdown	Have you ever had a nervous breakdown?		X

3.2.2 Polygenic scores

Polygenic scores (PGS) were created for psychological distress using PRSice-2 v 2.3.5 (205). The PGS was weighted using the Baselmans et al. 2018 GWAS for depressive symptom measurement (1,067,913 individuals) (51). This GWAS was selected from the GWAS catalog, with a trait identifier of EFO_0007006. It was the most recently conducted under this trait, with the largest sample size (N=1,067,913) and the highest number of SNP associations found (n=239). It had the most extensive age range, which included early adults to older adults. This was a necessary characteristic in the selection of the GWAS as the PGS assessed the genetic contribution over adulthood. Therefore, the PGS needed to be trained on a cross-section of ages across the lifespan, otherwise the testing of age effects would be limited.

The Baselmans et al. 2018 GWAS of depressive symptoms was derived from a meta-analysis of multiple datasets, which used various continuous measures for the phenotype, such as the General Health Questionnaire and Centre for Epidemiologic Studies - Depression (Supplementary Table 1) (206,207). These measures are all examples of questionnaires that are attempting to assess a continuous construct of depressive symptoms. Therefore, this GWAS phenotype was the closest to align with the Malaise Inventory Score, as it asked comparable questions assessing low mood and other depressive symptoms (Supplementary Table 2).

None of the GWAS samples in the GWAS utilised the Malaise Inventory Score. The 1958c and 1970c were not included in any of the GWAS (Supplementary Table 1). If they were included, it would bias the results towards overestimation.

This GWAS was originally reported using the human genome build GRCh37. Therefore, the summary statistics were lifted over to the GRCh38 build to match the 1958c and 1970c data genome build, with 4,310,707 of 4,980,155 SNPs successfully converted. After clumping and thresholding Study 1 polygenic score included all available SNPs (n=111,772, p value threshold = 1; n=124, p value threshold = 5×10^{-8}). For Study 2 and 3 another polygenic score was made to aid comparison between the 1970c and 1958c, whereby the included SNPs were restricted to overlapping variants in both cohorts' polygenic scores (n=93,962, p value threshold = 1; n=186, p value threshold = 5×10^{-8}) (Table 6).

This restriction was necessary because the two cohorts were genotyped on different platforms and imputed separately, resulting in partly distinct SNP sets. Using only the overlapping variants ensured that the polygenic scores in each cohort were constructed from the same set of genetic markers, thereby making results directly comparable. Although this approach reduced the number of SNPs by 2144 and hence statistical power, it avoided introducing bias due to differential SNP availability between cohorts and increased the interpretability of cross-cohort comparisons.

The PGS was created using the clumping and thresholding method (208). Clumping is achieved by removing highly correlated variants, with the linkage disequilibrium (LD) cut-off parameter set at r^2=0.1 and a window of 250kb (209). For the PGS' used in this thesis, clumping parameters were chosen for European-ancestry LD structure and the

polygenic architecture of depressive symptoms. A 250-kb window reflects the typical distance over which LD decays to background levels in European populations balancing the removal of correlated SNPs with the retention of independent loci (210). The r^2 threshold of 0.1 was chosen to minimise redundancy while capturing weak, polygenic signal characteristics of depressive symptoms, aligning with parameters from recent GWAS and previous studies (211,212).

For each study, PGS' were created by weighting the effect sizes of the single-nucleotide polymorphisms (SNPs) associated with the distress trait from the initial GWAS at 14 *p*-value thresholds (5×10^{-08}, $1x10^{-08}$, 0.0001, 0.00015, 0.0002, 0.0005, 0.001, 0.005, 0.009, 0.2, 0.3, 0.4, 0.5, 1). The number of SNPs included at each of threshold is reported in Table 6. PGS' were standardised to have a mean of 0 and a standard deviation of 1; thus, a higher polygenic score represents higher genetic liability to the trait. PGS were also standardised with both cohorts combined, and the same SNPs included for Studies 2 and 3.

There is no set methodology for threshold choice when using the clumping and thresholding method for PGS creation; previous studies have defined the optimal p-value threshold as that which explains the most phenotypic variation for a trait. Maximising R^2 enabled capturing aggregate genetic liability relevant to prediction across adulthood, rather than isolating causal variants. Furthermore, a permissive p-value threshold increases the risk of false positives but decreases the risk of false negatives. Therefore, at a more stringent threshold (e.g. 5×10^{-8}), the variance explained may be lower than with a more liberal threshold (e.g. at a threshold of 1) (213). Prior research used thresholds between 0.005 and 1 for a similar phenotype (212). For Studies 1 and 2, specification curve plots were created with the incremental R^2 for each threshold to help illustrate the most predictive threshold as 0.005 for both 1958c and 1970c across all ages (see Section 5.3.4 Figure 17). Alongside helping to identify the optimal threshold, the specification curve plots were used as a sensitivity analysis as the choice of threshold may alter the magnitude of estimates by age and cohort.

Table 6: Number of SNPs included at each P-value threshold for the polygenic scores used in this thesis

	Study 1		Study 2 and 3
Threshold	Distress PGS in 1958c	MDD PGS in 1958c	Distress PGS with overlapping SNPs
$5.00x10^{-08}$	194	83	186
$1.00x10^{-08}$	746	415	705
0.0001	1555	1095	1469
0.00015	1793	1334	1693
0.0002	1994	1558	1882
0.0005	2793	2561	2625
0.001	3829	3820	3599
0.005	26310	47624	24166
0.009	36168	71539	32907
0.1	38338	76817	34849
0.2	55706	121923	49981
0.3	68816	158417	61051
0.4	79399	188883	69610
0.5	88110	214468	76472
1	111772	292495	93962

NB: PGS = polygenic score, SNPs = single nucleotide polymorphisms

3.2.3 Father's Social Class

Socioeconomic position has been defined in many ways, as discussed in the Introduction. Fathers' social class was selected as the environmental exposure variable. This is based on the Registrar-General's Social Classes, which classify social class by occupational group: I (professional), II (managerial and technical), IIIN (skilled non-manual), IIIM (skilled manual), IV (partly skilled), and V (unskilled) and measured when participants were age 11 (214).

Father's social standing was categorised into a manual versus non-manual classification to ensure an equal sample size for each group, as well as to align with previous research conducted on these cohorts using the malaise inventory score (215). The non-manual category comprises classes I-III from the Registrar General's social class, which includes professional, managerial, and skilled non-manual labourers. The manual category includes classes IV-VI from the Registrar General's social class, including the manual, partially skilled and unskilled labourers. The economically inactive category contained N=372 participants, which included fathers of participants who were unemployed, retired and disabled; these were dropped from the variable (216). They were dropped due to their economic inactivity, as they did not fit into the conceptualisation of the economically active ordinal nature of the variable. Therefore, it would be inappropriate to amalgamate them into the category V (unskilled). The small sample size also precluded them from being a separate category.

Mother's social class was not used as in both cohorts' social class was classified using the Registrar General's Social Class (RGSC), which was primarily based on male occupations (215). This reflects the social norms of the time whereby expansion of women's economic activity in the workplace had reached 57% amongst women of prime working age between 25-and 57 by 1975, compared to 90% employment for men (217). Although there was increasing gender equality, in the 1950s and 1970s, fathers were still more likely to be most families' primary earners. Their occupation and social class were often considered representative of the family's overall socioeconomic status (215,218). Therefore, only father's social class was used in the present study.

Further justification for why the present study used the father's social class as an indicator of childhood socioeconomic position and broader childhood circumstances is outlined below.

It should be recognised, however, that an alternative approach commonly used in analyses of these cohorts is the dominance method, whereby the higher of the two parents' occupational classes is taken, or the mother's occupation is substituted when no father is present (37). This approach may more fully capture family resources in two-parent households and avoid underrepresentation of single-mother families. This thesis study retained a father-only measure.

This decision was motivated by the need for cross-cohort comparable measures of childhood social class as a proxy for distinguishing between advantaged and disadvantaged circumstances. Historically, the male head of household's occupation was treated as the primary household indicator of social class. For this thesis, in both

the 1958c and 1970c the age 11 father's social class was utilised, which is derived from 1958c variable n1687 (RG Social Class 1970 version), and the 1970c variable c3.4 (RG Social Class 1980 version). These were then standardised according to the RG Class 1990 version to enable comparison between the cohorts using the CLOSER harmonisation work package 2 variable fclrg90.

The main factor that precluded the dominance approach was measurement non-equivalence: mother's social class variables span incompatible classification systems (GRO 1966, RG 1970, RG 1980) that cannot be reliably harmonised across cohorts. Father's social class was consistently standardised to RG 1990 (CLOSER variable fclrg90), enabling robust cross-cohort comparison. Using different measurement strategies across cohorts would introduce further measurement error.

The details below outline the exact variables investigated as potential candidates for utilisation of the dominance approach and the subsequent issues with measurement and minimal sample size increase. If the dominance approach was utilised, the following variables could have been used from the 1958c and 1970c to include mother's social class:

1958c:

1. At birth, n660 Mother's father's social class using the GRO (1951)
2. At birth, n492/n236 Mother's social class via her husband's social class using the GRO (1951)
3. At birth, the variable n1687 showed that 510 households did not have a male head , i.e. single-mother households. By age 11, the variable n1685 indicates that 702 households had no male head.
4. At age 11, n1225 asks mother's occupation (GRO 1966), 6,974 have missing values.

1970c:

1. At birth, there are two variables a0014 (Father's social class in 1970) and a0018 (Mother's social class in 1970). These are derived from asking about their occupation in the questionnaire and then converting it to the RG social class 1970 version.
2. At age 10, the variable c3.11 mother's social class as per RG Social Class 1980 version. By age 10 there were 534 households that gave insufficient data or no data on father or mother's social class and a further 1,002 missing completely.

By age 11, n1225 asks the mother's occupation as based on the GRO 1966 classification system. These two variables (n1225 & n1687) are correlated at r=0.35, showing a moderate relationship. However, the gain in cases is small, as when tabulating between them, there is potential to include 577 extra participants. Yet due to only being able to include those who contributed genetic data, the actual gain would only be 201 cases.

A further consideration is not only the gain in participants from this approach, but also the balance with whether they are analogous schemas and categories. Therefore, it must be considered whether there is feasible comparability between different social class indicators at various time points and across genders.

The reliance on the male head of household as the wider household indicator of class is reflected in the lack of the mother's social class at birth being attributed to their own occupation being recorded, but rather that of their own father's in the 1958c. If the 1958c variable, n660 - father of the mother's social class at birth, was utilised alongside the 1970c variable a0018, it would be hard to ensure consistent comparability between the 1951 GRO and the 1970 RG social class. This is due to revisions to the Registrar General's Social Class (RGSC) as to how occupations were classified across the 1970, 1980, and 1990 versions, let alone the 1951 precursor. Under the 1970 scheme, occupations were ordered primarily by their "standing in the community," with little explicit reference to occupational competence. As a result, coal miners were often placed in the lower classes (IV–V), reflecting low prestige, despite the technical demands of their work (38). Similarly, farmers were inconsistently allocated depending on their perceived status, with larger landowners more likely to be placed in Class II and smallholders positioned lower. Clerical and retail workers were coded relatively highly, occupying Class III on the basis of their social reputation. In the 1980 revision, however, social class was redefined in terms of occupational skill rather than prestige (39) (p.77-108). This reclassification saw miners reallocated to Class III manual, recognising the training and competence required for the occupation, while farmers were more consistently placed in Class II as managers or technical workers. By contrast, clerks and shop assistants were downgraded to Class IV, as their tasks were deemed to require limited skill.

The 1990 revision, based on the new Standard Occupational Classification, made the skill framework more explicit, defining class in terms of the competence required to perform an occupation (40). This change consolidated the earlier reallocation of manual occupations upward in status, while also maintaining the downward movement of routine non-manual roles relative to 1970. For example, agricultural labourers were firmly classified as Class V, while farm managers remained Class II, marking a clearer distinction between managerial responsibility and unskilled work. Clerks remained in Class IV, but supervisory staff or retail managers could be placed in Class II, demonstrating the sharper mapping between occupational authority, competence, and class position. Overall, the cumulative effect of the 1980 and 1990 revisions was to move the classification system away from a prestige-based hierarchy and toward a skill- and competence-based framework, though continuity was preserved in order to maintain comparability with earlier time series (39,40).

As a result, the decision was made to keep using the CLOSER work package 2 variable fclrg90 to ensure robust comparison between cohorts for fathers' social class due to the harmonisation and standardisation efforts made to update the RG 1970 and 1980 schemas to the 1990.

3.2.4 Covariates

Covariates included across the thesis in varying combinations include the 10 first principal components of genetic ancestry, sex as male (0) and female (1), age, age squared and cohort as 1958c (0) and 1970c (1).

Sex was included to reduce the confounding that comes from a non-random sample and potential additional response from females over males, as prior survey research has shown females were less likely to drop out of cohort studies (222).

Cohort coded as 1958c (0) and 1970c (1) was used as a covariate to account for cohort differences.

Age and age squared were included as covariates to ensure that the reported associations reflect differences not attributable to age-related changes

All studies include the 10 principal components of genetic ancestry, which control for population stratification (223,224). These help to account for the non-random distribution of alleles across different geographic regions and ancestral groups (225). Population structure is based on the grouping of individuals by geographical location and ancestry. When individuals cluster based on shared ancestry or geographic origin, it can create systematic differences in genetic variation. This can then confound associations by creating spurious correlations between SNPs and phenotypes. This can lead to incorrectly attributing an association to a specific genetic variant when the relationship stems from shared ancestry. Principal component analysis takes the genetic data of a given population and reduces it to a smaller number of variables (principal components) that capture the main patterns of genetic variation in a population. Adjusting for principal components helps ensure that polygenic score associations are due to genetic effects rather than population structure (223).

Specific reporting of which covariates are included will be in the specific analytical strategy of each study.

3.3 Descriptive analyses of 1958c and 1970c

3.3.1 Mean, Standard Deviations and Histograms

These analyses provide a visual and descriptive understanding of the mean level difference in psychological distress and its variability in both cohorts by age. They also enable a descriptive comparison of the cohorts.

Table 7 summarises the mean and standard deviation of the 9-item psychological distress scores in the two cohorts 1958c and 1970c at ages 23, 26, 30, 34, 42, 46 and 50. The magnitude of mean scores varies by ages and cohorts (Figure 5a).

In 1970c cohort, the mean scores increase with age, with a decrease at age 30. It ranges from 1.43 to 1.74, with the highest mean observed at age 42 (Mean: 1.74, SD: 1.89) and the lowest at age 30 (Mean: 1.43, SD: 1.65). In the 1958c cohort, the mean scores range from 0.94 to 1.49, with the highest mean observed at age 42 (Mean: 1.49, SD: 1.74) and the lowest at age 23 (Mean: 1.15, SD: 1.50). There is a no7 difference in mean scores between the two cohorts at the same ages: the1970c had higher distress scores at comparative ages to 1958c.

The standard deviation increases across ages and between cohorts (Figure 5b). In 1970c, it increases with age from 1.65 (age 26) to 2.06 (age 46). In 1958c, it ranged from 1.50 to 1.89, with the highest standard deviation observed at age 50 (SD: 1.89) and the lowest at age 23 (SD: 1.46). As well as 1970c having higher distress scores compared to 1958c, participants also vary more in their scores.

In Figure 6, a histogram plot depicts the percentage frequency distribution of 9-item malaise inventory scores at ages 23, 26, 30, 33, 34, 42, 46 and 50 in 1958c and 1970c. Notably, individuals at age 30 exhibit the highest frequency of scoring 0 compared to other age groups, meaning they have fewer people with distress symptoms than all other ages. This may indicate levels of distress may reduce at this age point.

Table 7: Mean and standard deviation of the Malaise Inventory Score of participants in the 1958c and 1970c with genetic data

Age	Mean	SD
1958c		
23	1.15	1.50
33	0.94	1.46
42	1.49	1.74
50	1.42	1.89
1970c		
26	1.63	1.65
30	1.43	1.65
34	1.52	1.77
42	1.74	1.89
46	1.68	2.06

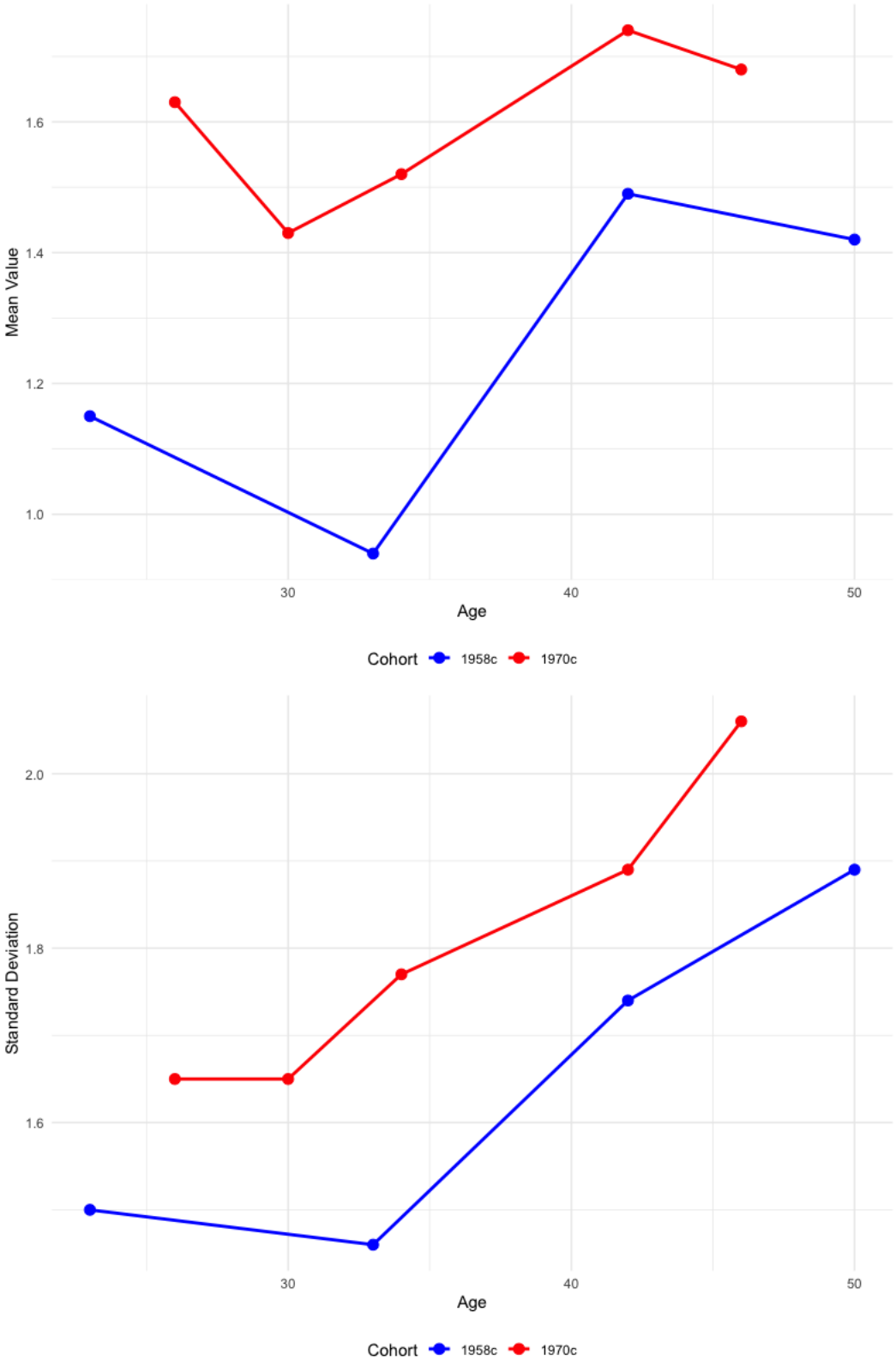
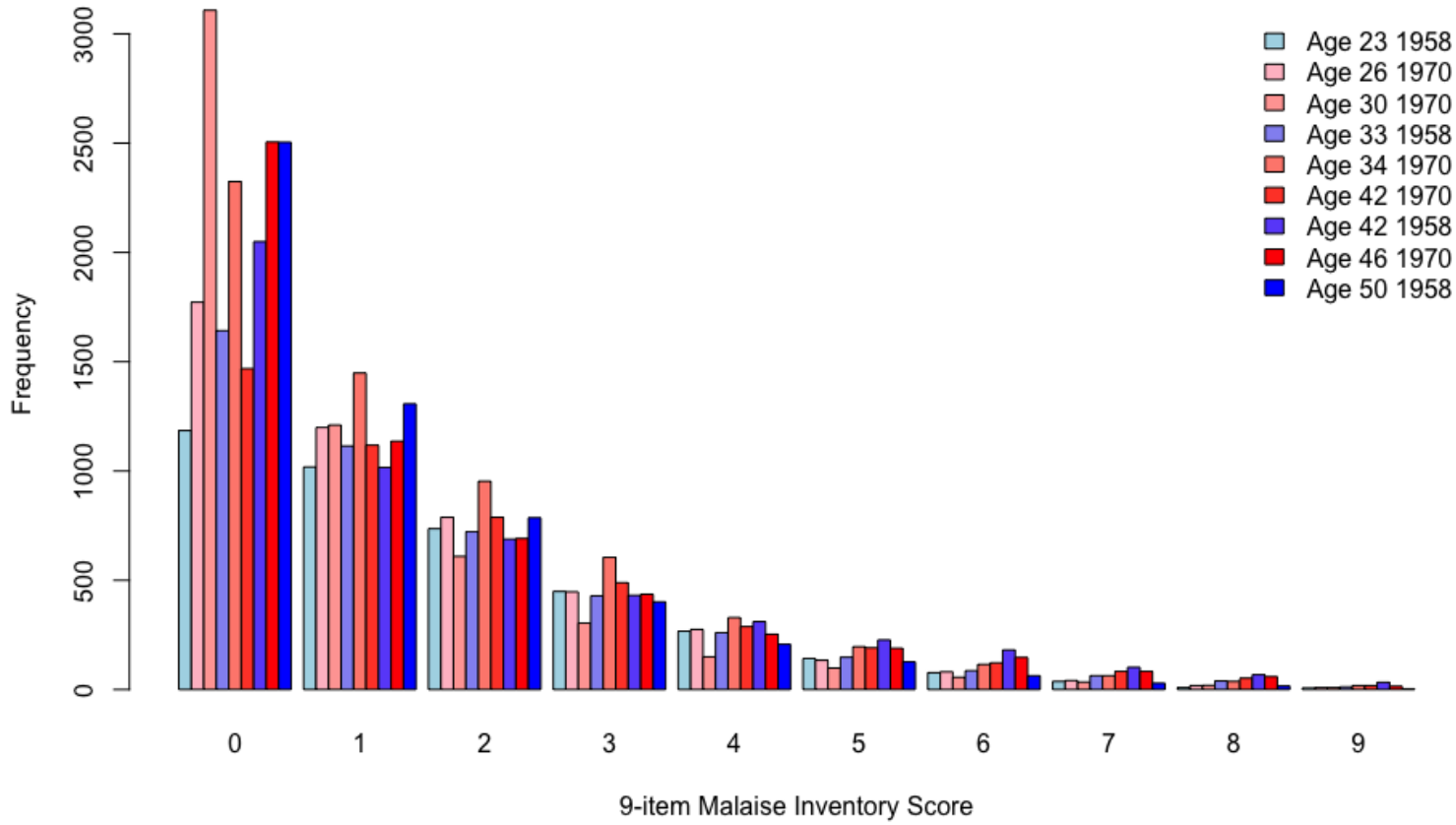


Figure 5: a) Mean and b) standard deviation of the Malaise Inventory Score at each age point, shown as a line graph for 1958c and 1970c

Figure 6: Histogram frequency percentage plot of the 9-item Malaise Inventory Score at ages 23, 26, 30, 33, 34, 42, 46 and 50 in 1958c and 1970c



3.3.2 Correlations

To better understand how the exposure and outcome variables relate to one another at different ages and cohorts, pairwise correlations were computed. The polygenic scores have an increasing correlation with malaise scores by age in the 1958c, while there is more variation in the 1970c. This may indicate there is a cohort difference in the relationship between the polygenic score and observed psychological distress.

Table 8 presents a correlation matrix of psychological distress variables at ages 23, 26, 30, 34, 42, 46 and 50, utilising the 9-item Malaise Inventory score and polygenic scores for psychological distress in both cohorts. The strength of positive correlation varied across time points, with higher correlations observed among those in closer temporal proximity than those more temporally distant. It therefore suggests that people differ from one another consistently across cohorts. However, there were still correlations even at temporary distant ages (e.g., rho = 0.40 for ages 23-50 in the 1958c and rho = 0.45 for ages 26-46 in 1970c). This finding is consistent as a feature of longitudinal data (226).

Table 8: Correlation matrix of Malaise Inventory Score variables (ages 23-50) the polygenic score for psychological distress in the 1958c and 1970c

1958c	Age 23	Age 33	Age 42	Age 50	PGS	
Age 23	1					
Age 33	0.47	1				
Age 42	0.43	0.50	1			
Age 50	0.40	0.44	0.56	1		
PGS	0.11	0.12	0.13	0.14	1	
1970c	Age 26	Age 30	Age 34	Age 42	Age 46	PGS
Age 26	1					
Age 30	0.53	1				
Age 34	0.49	0.54	1			
Age 42	0.46	0.49	0.55	1		
Age 46	0.45	0.47	0.53	0.62	1	
PGS	0.10	0.07	0.10	0.13	0.11	1

PGS: polygenic score

3.4 Missingness Patterns and Strategy

Across this thesis, missing data arose due to item non-response and attrition. Analysing only complete cases would have reduced statistical power and may have biased results if missingness was related to key exposures or outcome. Therefore, multiple imputation was used where appropriate to minimise information loss and to reduce the risk of biased estimates under the assumption of missing at random (MAR).

Multiple imputation across all waves, the analysed sample is held constant across ages by design; therefore, age-specific changes in R^2 are not driven by changing inclusion of participants across time points. Nonetheless, selective patterns in observed data that inform the imputation (e.g., individuals with lower socioeconomic position or poorer mental health being less likely to provide outcome data at later waves) can make the effective sample more homogeneous with age. Such changes in outcome variance can influence R^2 even when the analysis N is constant. Accordingly, the age patterning of R^2 should be interpreted as reflecting both genuine developmental change and potential shifts in outcome heterogeneity under a MAR assumption that may not fully hold.

Patterns of missing information were examined in both cohorts to evaluate the potential effects of incomplete data on the results. Predictors of who provided and did not provide data are important, as disparities could bias the estimated relationship between explanatory variables and outcomes. A notable advantage of both cohorts is the availability of data collected prospectively, which enables the identification of characteristics of participants in later data collection phases. This information helps assess the likelihood of differential biases between cohorts.

3.4.1 Study 1

In Study 1, using the 1958c, of the 6,312 participants who provided genetic data, only 3,748 completed all malaise inventory scores at ages 23-50. The missingness strategy employed multiple imputations by multiple chained equations to impute the missing malaise inventory scores at ages 23-50. Leaving an imputed sample of 6,312 for the 1958c.

Patterns of Missingness

Figure 7 shows the data is non-monotonic in missingness pattern; with the two most common missingness patterns being (1) non-response to all variables or (2) non-response to the malaise inventory score. 34% of the participants dropped out between age 23 and age 50 (N=4,256), with those missing at age 50 experiencing higher psychological distress at age 23 (Mean = 1.40, SD=1.72) compared to those who completed follow-up at age 50 (Mean = 1.24, SD=1.58). Of the possible 6,312 participants who gave genetic data, only 4,325 also completed all the malaise inventory scores at ages 23-50.

To explore the pattern of missingness in the 1958c, multiple logistic regressions were run:

- to assess whether attrition at age 50 was associated with worse psychological distress scores at age 23.
- whether having worse psychological distress scores was associated with not participating in genetic data collection

Logistic Regression Results:

- Logistic regression showed that individuals with higher values of the malaise inventory score, or poor psychological distress scores at age 23 had higher odds of missingness at age 50 by 67% (OR = 1.67, 95% CI [1.42-1.95], p = 0.001).
- Higher distress scores were also associated with 20% higher odds of not contributing genetic data (OR = 1.20, 95% CI [1.03-1.41], p = 0.024).

The consequences of these patterns of missingness are explored in the strategy for addressing missingness below.

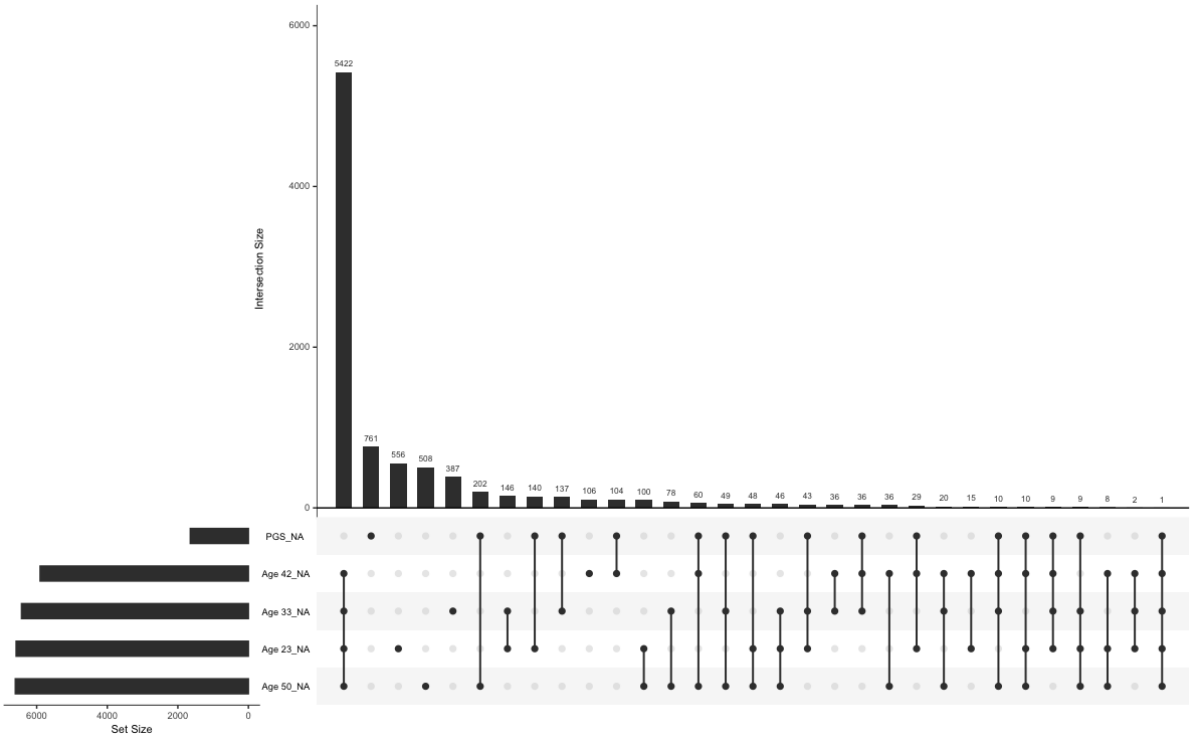


Figure 7: Missing pattern in the 1958c of the Malaise Inventory Score variables at ages 23, 33, 42 and 50 and the polygenic score

Strategy to address missing data

Several strategies could be used to address missing data, from simple imputation methods to just conducting complete-case analysis; however, the latter can bias estimates if the missingness is related to the observed or unobserved data (227–229). Due to the attrition in the analytical sample the missingness strategy used multiple imputation, to replenish the sample size and enable better representativeness.

Multiple Imputation of Chained Equations (MICE) was performed using available case data, where a separate conditional distribution was specified for each imputed variable (230). Multiple imputation sequentially generates replacement values for missing observations, using these to create *multiply imputed* datasets that reflect imputation uncertainty (227).

However, the validity of MICE assumes that data missingness is at random (MAR), meaning that data missingness is random after accounting for observed data (229). In Study 1, of the 6,312 participants who provided genetic data, only 3,748 completed all malaise inventory scores at ages 23-50, with 1,987 participants missing outcome values across all time points. Previous research has documented that attrition in the 1958 cohort is associated with higher levels of mental illness (231,232). As a result, the outcome variable missingness is related to higher values of malaise scores at earlier time points. This also affects the exposure, whereby those with higher polygenic score for psychological distress may then drop out of the study if they express the phenotype of higher psychological distress, which was explored in the missingness section above. Therefore, the missingness strategy employed multiple imputations by multiple chained equations.

The missing rate (i.e., those with missing data for all measures as a fraction of the total target population) was 78%. As per Rubin's formula for estimating the efficiency of an estimate based on imputed data, 78 imputations were used as they would yield at least 99% efficiency (233). The imputation was done using Stata 18.5 (234). Given the suitability of multiple imputations by chained equations for imputing missing values for categorical and continuous measures within the same imputation model, continuous variables were imputed with predictive mean matching and categorical variables with logistic regression. Auxiliary variables were used, as they can help improve the plausibility of meeting the missing at random assumption (227). Auxiliary variables included sex, father's social class, the polygenic score for psychological distress, crowding in childhood, whether breastfed, whether parents divorced during participants' childhood, family difficulties due to mental illness, poor maternal mental health, low birth weight, well-being alongside internalising and externalising behaviours during childhood (232,235,236).

Because MICE relies on the MAR assumption, it is worth considering its validity in this modelling context. If the deterioration in mental health occurs immediately before dropout, but cannot be inferred from earlier measurements, then the data may be missing not at random. However, our inclusion of extensive auxiliary variables that correlate with both the outcome and missingness patterns helps go some way to mitigate potential bias.

3.4.2 Study 2 and 3

In Studies 2 and 3, using both the 1958c and 1970c. The same imputed sample as the 1958c in Study 1 was included. The 1970s had 4,401 participants who provided genetic data; only 2,446 completed all Malaise Inventory scores at ages 26-46. In study 2, research question 1, which utilised cross-sectional regressions, the missingness strategy employed multiple imputations by multiple chained equations to impute the missing Malaise Inventory scores at ages 26-46 in the 1970s. Leaving an imputed sample of 4,401 for the 1970c.

For the analyses that utilised multilevel models, missing data were handled through full information maximum likelihood, which means using all available case data at the sum score level (not item level). Compared to a complete case analysis, this increased statistical power and precision. This resulted in a pooled sample of 10,713 of 1958c and 1970c for Study 2 and 8,923 in Study 3. Between Study 2 and Study 3, with the inclusion of the father's social class variable in the model as an interaction, there is a loss of 1,790 people who do not have a row observation of the polygenic score nor the father's social class. 984 are lost from the 1958 cohort and 806 are lost from the 1970 cohort.

Patterns of Missingness

Figure 8 demonstrates that data missingness in the 1970 cohort is similar to that in the 1958 cohort; again, non-monotonic in pattern; with the two most common being (1) non-response to all variables and (2) non-response to the malaise inventory score. Therefore, having similar patterns of missingness and no evidence of differential attrition, supports the comparability of the two cohorts.

34% of the participants dropped out between age 23 and age 50 (N=3,748), with those missing at age 50 experiencing higher psychological distress at age 23 (Mean = 1.40, SD=1.72) compared to those who completed follow-up at age 50 (Mean = 1.24, SD=1.58). As per1958c, in the1970c, 41% of the participants dropped out between age 26 and age 46 (N=2,312), with those missing at age 46 experiencing higher psychological distress at age 26 (Mean = 1.92, SD=2.27) compared to those who completed follow-up at age 46 (Mean = 1.71, SD=2.07).

To explore the pattern of missingness in the 1958c and 1970c, multiple logistic regressions were run:

- to assess whether attrition at the oldest time point (age 50 in 1958c and age 46 in 1970c) was associated with worse psychological distress scores at the youngest age point (age 23 in 1958c and age 26 in 1970c).
- whether having worse psychological distress scores was associated with not participating in genetic data collection in both cohorts.

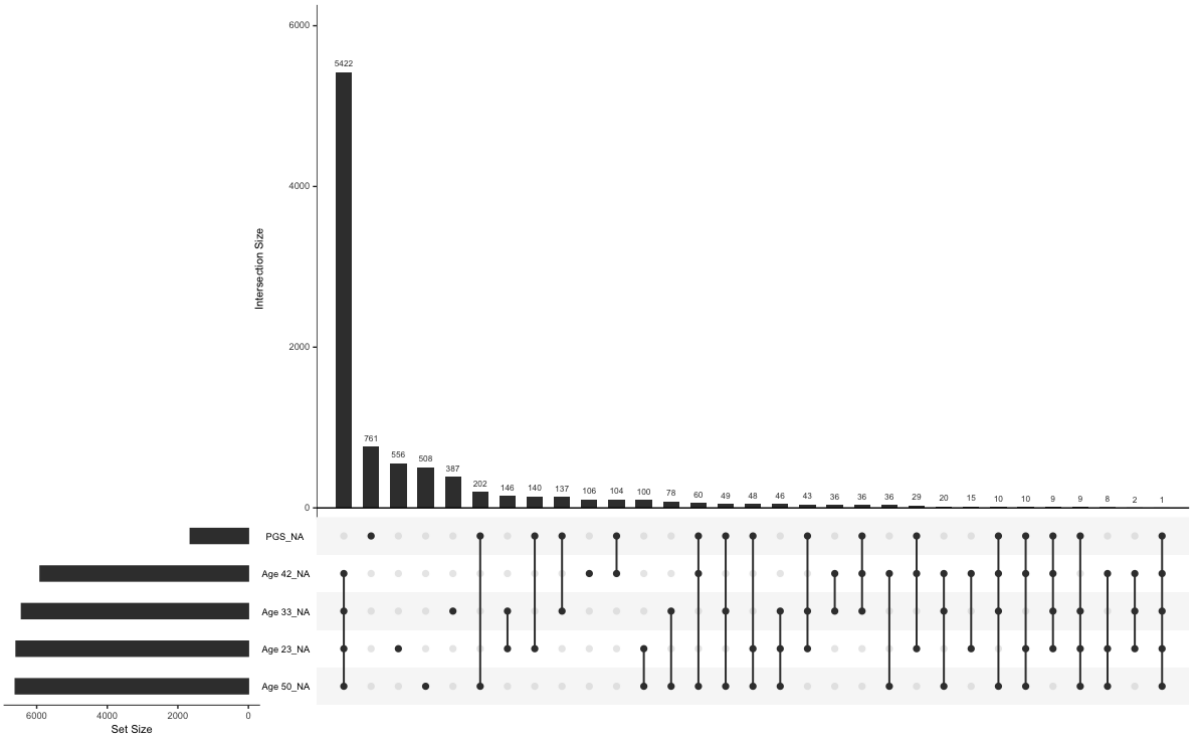
Logistic Regression Results:

- In both cohorts, logistic regressions showed that high distress scores at the youngest ages 23 and 26 had higher odds of missingness at age 46 by 57% in 1970c (OR = 1.57, 95% CI [1.35-1.82], $p < 0.001$) and at age 50 by 67% in the 1958c (OR = 1.67, 95% CI [1.42-1.95], $p = 0.001$).

- In both cohorts, higher distress scores were also associated with 20% higher odds of not contributing genetic data in the 1958c (OR = 1.20, 95% CI [1.03-1.41], p = 0.024) and by 52% in the 1970c (OR = 1.52, 95% CI [1.30-1.77], p < 0.001).

The consequences of these patterns of missingness are explored in the strategy for addressing missingness below.

a)



b)

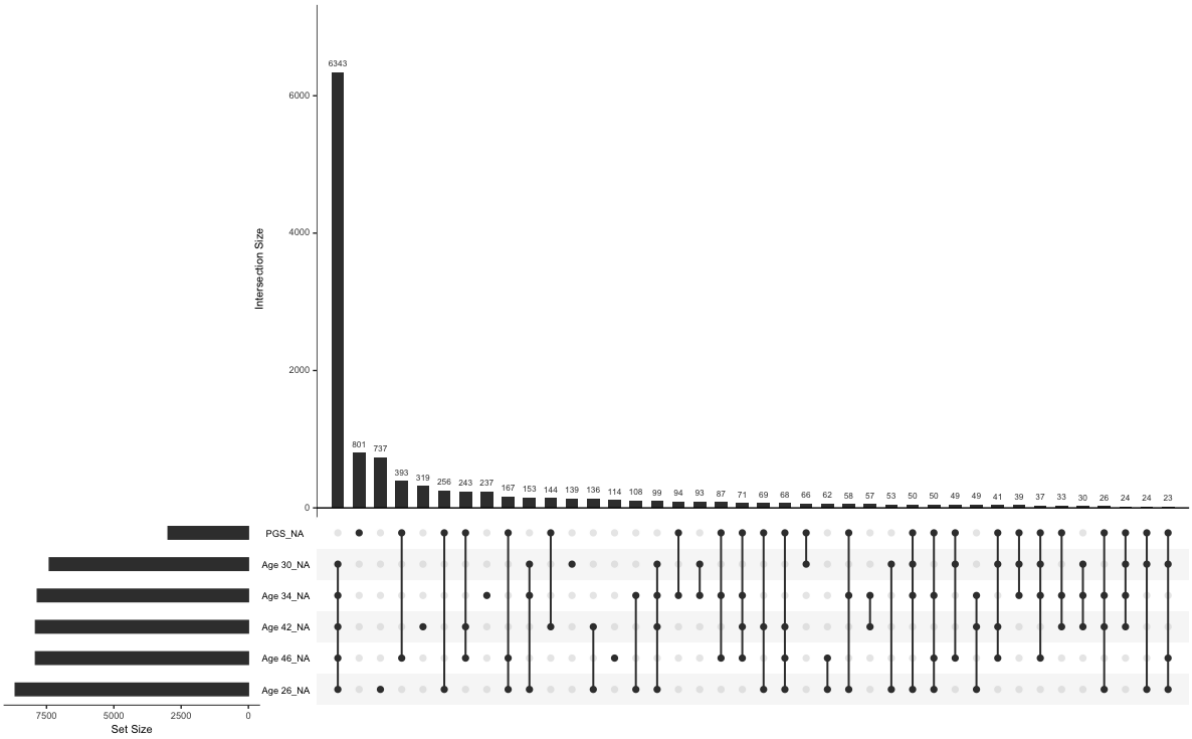


Figure 8: Missing pattern in the (a) 1958c at ages 23, 33, 42 and 50 and (b) 1970c at ages 25, 30, 34, 42 and 46 of the Malaise Inventory Score and the polygenic score

Strategy to address missing data

The main limitation in both 1958c and 1970c was participants' incomplete data across all waves for the malaise inventory score and genetic material (Figure 10). In 1958c, 25.8% of the initial sample participated in all time points, meaning a sample size of 4,497 out of 17,415 (232). Similarly, in 1970c, 19.8% of participants participated in all time points and included genetic information, which is a sample size of 3,423 out of 17,284 (235). For a complete case sample with observed data at all time points for all variables, the sample size is 3,748 in 1958c and 2,446 in 1970c.

The age 16 malaise inventory score in 1970c was not included in the current study due to increased attrition and to make the ages comparable. Missingness was explored in both cohorts and found to have a similar pattern, whereby those with the higher distress scores at the first time point were more likely not to be included in the following time points.

As research question 1 is a cross-sectional analysis, multiple imputations using chained equations in both cohorts were employed as per the study 1 strategy. For research questions 2 and 3, multilevel models were used as they can handle missing data effectively through maximum likelihood estimation, using all available data points without excluding participants with incomplete records, thereby increasing statistical power and precision (p.269) (229).

3.5 Statistical Analyses

Analyses were performed in Stata version 18 and R v 4.3.2.

3.5.1 Linear Regression

Throughout the thesis linear regression models are used, with an ordinary least squares estimator:

$$Y_i = \alpha + \beta . X_i + \sum_{g=11}^{g=1} \beta_g C_{gi} + \varepsilon_i$$

Where i = individual, Y = outcome, α = intercept, β = polygenic score coefficient, X = polygenic score, g = control variables (1 to 11), C = control variable coefficient, and ε = the idiosyncratic error term. Therein, the fitted value of y is equal to the coefficient of β multiplied by the value of X, plus the control variable coefficients times their values, plus the intercept and error term.

Strengths

Results from a linear regression provide information about the magnitude and direction of the relationship between an outcome and an exposure. Model estimates can be interpreted as the change in the outcome variable for a one-unit change in the predictor, holding all other variables constant. Another strength is that linear regression is a simple model that is robust to violations of normality with large sample sizes due to the central limit theorem.

Limitations

Due to the different equations for each age, there is a limitation in how to interpret the confidence intervals, whether they overlap or not. This is because each linear regression error term may be correlated, especially if it is the same people repeatedly measured for a psychological distress phenotype. This is because the prior state of an individual will inform the next time point. Multilevel models mixed-effects models can to account for within-individual correlation.

3.5.2 Multilevel Mixed Model

Multilevel mixed-effects models were used to investigate change or stability between phenotypic and genotypic psychological distress. Linear mixed models with maximum likelihood estimation are nonparametric regression models for handling grouped, nested and hierarchical data. They enable full use of all data points in longitudinal data, adjust for the correlation between repeated measures, weight estimates for missing data between waves, and increase statistical power and precision.

The fixed part of the model includes exposure and covariates that are interpreted the same as an ordinary regression. The random part is added for the intercept, slope, to provide a flexible way to model the variability and correlation among the repeated measurements within the same person.

Strengths:

Multilevel mixed models are specifically designed to handle hierarchical or nested data structures, such as repeated measurements within individuals (237). These models appropriately adjust for the correlation between repeated measures within the same individual, which improves the validity of the statistical inferences by correctly specifying the error structure (237). These models allow the specification of various covariance structures to best fit the data, which enhances the model's ability to represent the underlying relationships among repeated measures accurately (237). By including random effects, these models can account for individual variability in baseline levels (random intercepts) and in the rate of change over time (random slopes), offering a more flexible approach that can capture individual differences in trajectories of psychological distress (237). Multilevel models can handle missing data effectively through maximum likelihood estimation, using all available data points without excluding participants with incomplete records, thereby increasing statistical power and precision (237).

Limitations:

Estimation in multilevel models, especially with maximum likelihood, can be computationally demanding, particularly with large datasets or complex random effect structures, which might lead to convergence issues (237). These models assume that random effects are normally distributed and that the relationship between predictors and the outcome is linear (237). Violations of these assumptions can lead to biased estimates or incorrect inferences (237). Like other regression models, multilevel mixed models can be influenced by outliers or influential data points, particularly within the random effects. Including multiple random effects and complex covariance structures can lead to overfitting, particularly with small sample sizes or when the number of repeated measures is limited (237). However, the current studies 2 and 3 have large sample sizes, more than 3 repeated outcome measures, and well-specified covariance structures. Hence these issues are unlikely to impact the analyses.

3.6 Supplementary Materials

Supplementary Table 1: Summary of Baselmans et al 2018 GWAS' cohorts meta-analysed and their corresponding phenotype

GWAS/Cohort	Author	Cohort	Phenotype
Cohort	Bycroft et al. 2017	UK Biobank	1) Over the past two weeks, how often have you felt down, depressed, or hopeless? (2) Over the past two weeks, how often have you had little interest or pleasure in doing things?
GWAS	Okbay et al. 2016	U.K. Biobank Genetic Epidemiology Research on Adult Health and Aging (GERA) Psychiatric Genetics Consortium (PGC) (9 cohorts[1]):	UKB: (1) Over the past two weeks, how often have you felt down, depressed, or hopeless? (2) Over the past two weeks, how often have you had little interest or pleasure in doing things? GERA: (1) if patient had at least 2 diagnoses of depression on separate days in a Kaiser Permanente Northern California facility between January 1, 1995, and March 15, 2013. PGC: (1) if patient diagnosed with major depressive disorder. Table S2 in [4] lists the exact clinical measures from the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) used to determine the disease status.
Cohort	Brice, buck & Prentice-Lane	British Household Panel Survey/ Understanding Society	General Health Questionnaire: Have you recently: 1. Been able to concentrate on what you're doing? 2. Felt you were playing a useful part in things? 3. Felt capable of making decisions? 4. Felt able to enjoy day-to-day activities? 5. Been able to face your problems? 6. Been feeling reasonably happy, all things considered? 7. Lost sleep over worry? 8. Felt constantly under strain? 9. Felt couldn't overcome difficulties? 10. Been feeling unhappy and depressed?

			11. Been losing confidence in self? 12. Been thinking of self as worthless?
GWAS	Hyde et al. 2016	23andMe	Self-report prior clinical diagnosis of, or treatment for, major depression
GWAS	Hek et al. 2013	Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) (17 cohorts[2])	Centre for Epidemiological Studies Depression Scale: I was bothered by things that usually don't bother me. I had trouble keeping my mind on what I was doing. I felt depressed. I felt that everything I did was an effort. I felt hopeful about the future. I felt fearful. My sleep was restless. I was happy. I felt lonely. I could not “get going.”

NB: Ref column denotes the reference number in the Baselmans et al GWAS. Both PGC and CHARGE cohorts were checked for inclusion of the 1958c and 1970c, they were excluded.

[1]Okbay et al. 2016 utilised the cohorts from the Ripke et al. 2013 GWAS which utilised 9 cohorts from PGC these include: (1) Bonn/Mannheim, Germant, (2) Genetic Association Information Network from NL, (3) Genetics of Recurrent Early-On Depression from USA, (4) GSK from Germany, (5) MDD2000 and (6) QIMR from Australia, (7) Max Planck Institute of Psychiatry from Germany, (8) RADIANT MDD study is mixed from clinical cohorts GENDEP, DeCC and DeNt and the (9) STAR*D trial from USA.

[2]Hek et al. 2013 utilised 17 cohorts from CHARGE these include: (1,2) the Atherosclerosis Risk In Communities 1 and 2 studies (ARIC1 and ARIC2), (3) the Cardiovascular Health Study, (4) the Framingham Heart Study, (5-7) the Rotterdam Study I, II, and III (RS-I, RS-II and RS-III), (8) the Baltimore Longitudinal Study of Aging, (9) The Erasmus Rucphen Family study, (10) the Health, Aging and Body Composition study, (11) Invecchiare in Chianti, (12) Helsinki Birth Cohort Study, (13) Multi-Ethnic Study of Atherosclerosis, (14) Nurses' Health Study, (15) Rush Memory and Aging Project (MAP), (16) Religious Orders Study and (17) SardiNIA study.

Supplementary Table 2: Summary of Baselmans et al 2018 GWAS phenotype measures and comparison to Malaise Inventory Score

Construct	Item	The General Health Questionnaire	Item	Centre for Epidemiological Studies Depression Scale:	Item	Diagnosis/One-off Question	Item	Malaise Inventory Score
Low mood	10	Been feeling unhappy and depressed?	3	I felt depressed.		Over the past two weeks, how often have you felt down, depressed, or hopeless?	2	Do you often feel depressed?
Worry/anxiety	7/8	Lost sleep over worry?/ Felt constantly under strain?	6	I felt fearful.			3/5	Do you often get worried about things? /Do you suddenly become scared for no good reason?
Somatic Symptoms			7	My sleep was restless.			1/9	Do you feel tired most of the time?/ Does your heart often race like mad?
Irritability							4/6/7/9	Do you often get into a violent rage? / Are you easily upset or irritated? /Are you constantly keyed up and jittery? / Does every little thing get on your nerves and wear you out?
Anhedonia	4	Felt able to enjoy day-to-day activities?	4/?	I felt that everything I did was an effort/I could not get going		Over the past two weeks, how often have you had little interest or pleasure in doing things?"		
Self-esteem/Worthlessness	2/11/12	Felt you were playing a useful part in things? / Been losing confidence in self?/ Been thinking of self as worthless?	9	I felt lonely.				
Concentration	1	Been able to concentrate on what you're doing?	2	I had trouble keeping my mind on what I was doing.				
Positive Affect	6	Been feeling reasonably happy, all things considered?	5/8	I felt hopeful about the future/I was happy.				
Decision making	5/?/9	Been able to face your problems?/Felt capable of making decisions?/ Felt couldn't overcome difficulties?	1	I was bothered by things that usually don't bother me.				

NB: These are not definitive as there are multiple cohorts used within the meta-analysis GWAS which would be beyond the scope of the current table, the measures chosen to reflect the broad consistency of phenotype chosen

4 Study 1: Age differences in the association between polygenic scores and psychological distress: evidence from the 1958 British Birth Cohort

Main objective: to examine whether the polygenic score for psychological distress is associated with phenotypic psychological distress at ages 23, 33, 42 and 50 in the 1958 National Child and Development Study (1958c).

Research Question 1: Does the magnitude of association between polygenic score and adult psychological distress outcomes differ from age 23 to 50?

Research Question 2: Does the amount of variance explained in adult psychological distress outcomes by polygenic scores differ from age 23 to 50?

4.1 Introduction

As previously outlined in the Introduction section, psychological distress is a complex construct that includes symptoms of both major depressive disorder and anxiety disorder (12,13). There is a well-established lifecourse patterning in increasing levels of distress as people enter middle age. The association between genetic liability to distress and observed psychological distress might vary with age. Therefore, the age differences in polygenic contribution to psychological distress are less well established and are the focus of this chapter.

4.1.1 Age differences in the polygenic contribution to psychological distress

Psychological distress is a polygenic trait, as there are many genes that contribute to it with small effect sizes (39). The phenotype for distress is varied and complex across adulthood, and as people age the polygenic contribution to distress may alter. Polygenic contribution to distress phenotypes has been shown to increase from childhood through to adolescence (238–241). A systematic review from 2022 identified 131 molecular genetic studies of common mental disorders in childhood and adolescence. Eight studies of the 131, assessed genetic associations with distress by age (239,242–249). Their methods varied with some using longitudinal data and others having a broad age range at one time point. The sample sizes ranged from 466 to 42,998, and the variance explained by polygenic contributions to distress also ranged from 0.2% to 5%. This variability highlights that depending on population and context, the contribution of genetic factors to distress can differ. This may vary across the lifecourse as people age beyond adolescence.

One molecular study from the U.K. identified by the 2022 systematic review addressed how polygenic contribution may change by age; their population age range was from 10 to 24 years old – following adolescents into young adulthood. The beta coefficient changed from a 0.21 increase in distress scores per 1 standard deviation increase in polygenic at age 10; this increased further to 0.58 by age 24 (250). The same magnitude of association was found in data from the Netherlands in participants aged 7-15 (240). Furthermore, in the U.K. study, distress polygenic scores explained 0.32% of the variance in distress scores at age 11, increasing to 1.30% at age 18. These estimates may differ due to age-related variation in the contribution of genetic propensity to psychological distress, with early-onset and adult-onset symptoms potentially having different configurations of genetic propensity, phenotypic quality, environment, and their interaction.

A meta-analysis of 42,998 twins aged between 6- to 17-years-old from the U.K., Netherlands, Sweden, Norway, and Finland identified that adult major depressive disorder polygenic scores were associated with childhood psychopathology in all contexts (239). The evidence is consistent that the transition of distress from childhood to adulthood has some genetic element. This evidence replicates the trend identified in psychiatric epidemiological studies that previously have not used genetic data. Whereby they find childhood distress, in turn, predicts stability in adulthood distress (114).

Four molecular studies have explored genetic contribution to distress in older adults above the age of 50 (251–254). They all found a positive association between

polygenic scores and observed distress outcomes. However not all tested how this relationship changed with age.

Two studies utilised data from the English Longitudinal Study of Ageing (ELSA), one study examined the association between polygenic scores and psychological distress symptoms in a cohort of 3,231 individuals aged 50- to 95-year-olds over a 15-year period of follow-up (251). They identified participants with higher polygenic scores had a stronger association with psychological distress scores than those with lower polygenic scores. The other study found that common genetic variants for distress were associated with a greater number of symptoms at onset but not with their rate of change over the following 14-years (253). Whereas, evidence from the Health and Retirement Study, which uses the same panel design as ELSA but based in the United States, found that polygenic scores predicted chronic distress symptoms in sample of 2,071 (252).

Evidence from a sample of 2,279 in the Longitudinal Age Study of Amsterdam found that higher polygenic scores were associated with higher distress symptoms as people aged (254). However, the dichotomised Centre for Epidemiologic Studies-Depression scores limited the scope of their investigation as distress symptoms are on a continuum, restricting the inclusion of genotypes and phenotypes of participants who fell below the threshold. Furthermore, the study had a short follow up of 14 years. Compared to data spanning across adulthood, they would only be capturing a small amount of change, thus the power to detect the small effect is lowered.

In one study, heritability estimates were found to be stable throughout mid-life (255). Their study design used data from 2,153 twins assessed over 8 years between ages 30t to 38. They found that genetic influence on distress was stable. By using structural equation modelling, the stability became an assumption of the model. As their latent construct assumed that genes were a stable liability to distress phenotype over time while environmental influences contributed to variability.

In summary, the disparate nature of studies focusing solely on childhood and adolescence, late adulthood, or only mid-life means it is hard to make conclusions across different life stages. This is due to limited number of data sources that have genetic data and collected longitudinal psychological distress data in the same people over many years (256).

4.1.2 Theoretical model for the gene-age interaction of the current study

The Selection, Optimisation, and Compensation (SOC) model holds that biological potential and genome-based plasticity decline with age, while individuals increasingly rely on cultural and social resources to maintain functioning age (1). Baltes et al. (1997) state: "*With age, the genetic material, associated genetic mechanisms, and genetic expressions become less effective and less able to generate or maintain high levels of functioning. Evolution and biology are not good friends of old age*." If applied to psychological distress, it implies that the genetic contribution should diminish with age, as environmental compensation plays a larger role. Therefore, the theory, when applied, would hypothesise improved emotional well-being in later life, proposing that individuals become more skilled at adapting to and reconstructing losses as they age.

Furthermore, in the context of this study, it is important to distinguish between two related but distinct concepts: polygenic score heritability and polygenic penetrance. These were previously outlined in Section 1.3.2 of the Introduction. Polygenic score heritability refers to the proportion of variance in psychological distress explained by measured genetic variants at a given time point, whereas penetrance reflects the extent to which carrying these variants is expressed phenotypically as elevated distress (44,45). While heritability provides a summary of genetic contribution to population variance, penetrance captures the manifestation of genetic predisposition as the expected difference in the phenotype. Penetrance can remain constant, but heritability varies if the population variation of the trait changes (which may occur if other causes of distress vary in their effect). Within this study, incremental adjusted R^2 values are taken as indicators of PGS heritability, while age-specific beta coefficients are interpreted as reflecting polygenic penetrance. This distinction helps evaluate gene–age interactions in psychological distress, clarifying whether observed changes reflect variation in the strength of genetic expression across adulthood.

4.1.3 Prior findings from the 1958 National Child Development Study (1958c)

Previously, a polygenic score of major depressive disorder has been created in the 1958c, with a study which assessed whether the same genetic liability to emotional problems was the same at different ages (257). Their findings highlighted a potential difference in genetic architecture underpinning emotional problems from childhood to mid-life. They utilised the Bristol Social Adjustment Guide (BSAG) at ages 7 and 11, and at age 16, they used the Rutter Parent and Teacher questionnaire. Their conclusions that the phenotypic-genotypic association increases by age between childhood and adulthood do not consider that this may instead be due to a mismatch in phenotypic measurement and genome-wide association study (GWAS) used for the polygenic score. They did not use the same measure across time, so the increase in association between the major depressive disorder polygenic score to the Malaise Inventory Score (ages 23-42), compared to the BSAG (ages 7-11) and Rutter (age 16) could be due to the change in construct.

When solely focusing on the study's adulthood results, at age 23 (OR: 1.02, 95% CI [0.97, 1.08], p = 0.420), 33 (OR: 1.06, 95% CI [1.00, 1.12], p = 0.048) and 42 (OR: 1.06, 95% CI [1.00, 1.11], p = 0.034), the association between the MDD risk alleles and Malaise Inventory Score was only significant at ages 33 and 42. They also did not control for population stratification, making it harder to infer robust genotypic-phenotypic associations by age. Finally, they used a major depressive disorder case/control GWAS, with 33,332 cases and 27,888 controls, which is considered underpowered by contemporary GWAS research of this trait (54). Therefore, replication of the study, with a longer follow-up, utilising a continuous outcome measure, and an updated GWAS would help to better identify differences in genotypic and phenotypic psychological distress by age.

Overall, as summarised in Table 9, existing studies all found independent associations between polygenic scores and psychological distress at multiple ages. There were inconsistent results regarding evidence for whether there was stability or change in the estimates as samples aged across adulthood.

4.1.4 Summary

To summarise, research using twin and molecular designs has provided evidence that genetics plays a role in psychological distress outcomes. Whether the polygenic influence changes over a more extended age span in the same people is unclear. Therefore, this thesis used genetic data from the 1958 National Child Development Study to examine if the association between genetic factors and psychological distress outcomes at different ages. The most recent and powerful GWAS for broad-spectrum depressive symptoms was used to generate polygenic scores that align closely with psychological distress outcomes measured in the 1958c (51,211).

Table 9: Summary table of key literature of genetic contributions to psychological distress outcomes by age

Author (Year)	Country	Dataset (Type)	Time of Data	Sample Size	Sample age range	Depressive Symptoms Measure	Same Participants over time?	Genetic Contribution
Molecular studies								
Akingbuwa (2020)	Meta-Analysis	Multiple: U.K., Netherlands, Sweden, Norway, Finland	1985-2002	42,998	6-17	Polygenic Score from Wray 2018 MDD	No	Stable
Riglin (2018)	England	The 1958 National Child Development Study (Cohort)	1958-2000	5,257	7-42	Malaise Inventory Score	Yes	Increasing
Schultebraucks (2021)	U.S.	Health and Retirement Study (Panel)	1992-2010	2,071	50-95	CES-D	Yes, compares between.	Stable
Torvik (2019)	Norway	Norwegian Twin Register	2006-2015	11,727	18-45	Clinical diagnosis	No	Stable
Nes (2007)	Norway	Norwegian Twin Register	1967/74-1998	4,393	18-31	Clinical diagnosis	No	Stable
Twin Studies								
Hannigan (2017)	Systematic Review	Multiple	Multiple	~64,000	0-18	Multiple for emotional problems	No	Stable
Nivard 2015	Netherlands	Netherlands Twin Register	1991-2009	49,524	3-63	SxAnxDep were assessed repeatedly with a	No	Stable

						maximum of eight assessments over a 25-year period		
Gillespie (2004)	Australia	Australian Twin Register	1980-1996	5,141	20-96	14-item self-report DSSI/sAD scale	Yes	Stable
Baselmans (2018)	Netherlands	Netherlands Twin Register	1994-201	43,427	7-99	Achenbach's System of Empirical Based Assessment (ASEBA); Child Behaviour Checklist	No	Stable

NB: Due to the heterogenous methods used in all the evidence, genetic contribution column was derived from the results of each study including a statistically significant positive association between the distress phenotype and the genetic component of their research design.

4.2 Method

4.2.1 Data

The study used data from the 1958 National Child Development Study (1958c) (187,188), which is described in the thesis Methods Section 3.1.

4.2.2 Measures

4.2.2.1 Psychological Distress

The malaise inventory score was the outcome variable. The 9-item version was used for the current chapter. For details regarding the development, harmonisation, measurement invariance, internal validity, and external validity of the Malaise Inventory, they are presented in the Methods Section 3.2.1.

4.2.2.2 Polygenic scores (PGS)

Methods for the construction of the two polygenic scores used in the current study are outlined in Methods Section 3.2.2. The polygenic score was standardised to have a mean of 0 and a standard deviation of 1. The main analysis exposure is the standardised polygenic score for psychological distress at a threshold of 0.005.

4.2.2.3 Covariates

Covariates included the 10 first principal components and sex as male (0) and female (1). Sex is included to reduce the potential bias due to a non-random sample. This is due to potential additional response from females versus males, as prior survey research has shown females were less likely to drop out of cohort studies (222). The 10 principal components are included to control for any residual population stratification (224).

4.2.3 Analytical Strategy

4.2.3.1 Research Question 1: Does the magnitude of association between polygenic score and adult psychological distress outcomes differ from age 23 to 50?

4.2.3.2 Research Question 2: Does the amount of variance explained in adult psychological distress outcomes by polygenic scores differ from age 23 to 50?

Research Question 1 and 2 were answered using repeated linear regression.

The outcome, the malaise inventory score, is continuous. Therefore, linear regressions were used. If it were categorised at the typical cut point of 4 out of 9 items and above, there would be reduced information and statistical power (258,259). This is due to sub-threshold effects whereby people who are below the arbitrary cut-offs tend to fare worse in the chronicity of distress but are undetected (260).

The association between the polygenic score (x) and the malaise inventory score (y) was tested at different ages on the absolute scale. The benefit of the absolute scale is that it gives a practical significance to the results, whereby a 1 unit increase in the exposure of the polygenic score may correspond to the coefficient change in the Malaise Inventory score.

Linear regressions were performed to investigate the relationship between the polygenic score for psychological distress and the malaise inventory score at four different time points - ages 23, 33, 42 and 50. Each model controlled for 10 principal components and sex as covariates. Research question 1 is addressed by estimating the beta coefficient in these analyses and research question 2 by using the incremental adjusted R squared.

Both beta coefficient and R squared results were plotted to visually inspect potential differences by age.

Multiple imputation was employed to handle missing data of malaise inventory score for the linear regression analyses assuming that data was missing at random (261). Multiple imputations utilised auxiliary variables with 78 imputed datasets to address missingness. Auxiliary variables included sex, father's social class, the polygenic score for psychological distress, crowding in childhood, whether breastfed, whether parents divorced during participants' childhood, family difficulties due to mental illness, poor maternal mental health, low birth weight, well-being alongside internalising and externalising behaviours during childhood (232,235,236) (See section 3.4).

4.2.3.3 Supplemental and Sensitivity Analyses

1) *Testing polygenic score for psychological distress at different p-value thresholds:*

Multiple analyses using different p-value thresholds were conducted to check if the main findings were robust to threshold selection. This is important because the choice of threshold affects which SNPs are included in the polygenic score:

During PGS construction, clumping first removes SNPs, retaining only independent signals. The remaining SNPs are weighted according to the base GWAS data risk alleles (213). Thresholding then determines which of these SNPs are included in the final score based on their p-value significance.

Therefore, the higher the threshold, the more SNPs included in the score. Table 5 summarises the number of SNPs included at the different p-value thresholds. Previous research with similar outcomes by Kwong et al. (2020) found that thresholds between 0.005 and 1 were the best predictors of psychological distress, as measured by R-squared (212). Based on this literature, thresholds were tested ranging from the genome-wide significance level (5×10^{-8}) up to 1, with particular attention to the 0.005-1 range identified in previous research.

2) *Comparison to a more specific phenotype:*

To examine whether age-related differences in polygenic associations with psychological distress might be driven by age differences in specific types of distress symptoms, analyses were conducted comparing the broader psychological distress phenotype with a more specific MDD phenotype. This comparison helps determine whether the observed age patterns are consistent across different mental health phenotypes or specific to general psychological distress The most recent and largest GWAS summary statistic for major depressive disorder was used to create an MDD-PGS (211)). The threshold was set at 0.005 for MDD-PGS consistent with the psychological distress PGS, to enable direct comparison.

3) *Complete Case Analysis:*

A complete case analysis was run as a sensitivity check to check robustness of estimates. After using multiple imputation under the assumption of data being Missing at Random it allows for comparison with imputed results. If the complete case and imputed results are similar, this strengthens confidence in the findings as each analysis has different assumptions. Whereas differences in estimates may suggest issues with the imputation model, missing data mechanism or that the complete case is biased.

4) *Sex Stratified Analyses:*

Sex-stratified analyses were conducted to examine potential differences in how genetic liability to psychological distress operates across males and females. This approach aligns with the gene-environment interaction framework discussed in the introduction, as biological sex represents both a physiological context (e.g., hormonal

differences) and a social context (e.g., gender-based expectations) that may moderate genetic expression. Previous research has demonstrated sex and gender differences in both the prevalence and symptom presentation of psychological distress, suggesting that genetic influences might similarly vary by sex, potentially due to sex-specific biological vulnerabilities or gender-specific differential exposure to environmental stressors.

4.3 Results

4.3.1 Descriptives

4.3.1.1 Mean, Standard Deviations and Histograms

In Thesis Methods Section 3.3 Figure 8, a histogram plot depicts the frequency distribution of psychological distress scores on the 9-item malaise inventory at ages 23, 33, 42, and 50. Notably, individuals at age 33 exhibit the highest frequency of scoring 0 compared to other age groups, meaning they have fewer people with distress symptoms than all other ages.

Table 10 summarises the 9-item psychological distress scores at ages 23, 33, 42, and 50, incorporating key statistical parameters. The magnitude of mean scores is highest at age 42 (Mean: 1.45, SD: 1.69), followed by age 50 (Mean: 1.39, SD: 1.85), age 23 (Mean: 1.12, SD: 1.46), and age 33 (Mean: 0.89, SD: 1.42). This indicates increasing psychological distress scores from age 33 to 42, with a subsequent decrease at age 50. The standard deviation at age 50 attains the highest value, suggesting greater variability in psychological distress scores.

Table 10: Mean and standard deviation of the Malaise Inventory Score in the 1958c of participants with genetic data

Age	Mean	SD
23	1.12	1.46
33	0.89	1.42
42	1.45	1.69
50	1.39	1.85

4.3.1.2 Correlations

Table 11 presents a correlation matrix of psychological distress variables at ages 23, 33, 42, and 50, utilising the 9-item Malaise Inventory score and polygenic scores for Major Depressive Disorder (MDD) and psychological distress. Between the malaise scores, the strength of positive correlation varied across time points, with higher correlations observed among those in closer temporal proximity compared to those more temporally distant. Notably, the psychological distress score at age 23 exhibited less correlation over time with the measures at ages 42 and 50. The polygenic score for psychological distress was increasingly correlated with the malaise score by age, whereas the polygenic score for MDD became increasingly correlated up to age 42, followed by a decline at age 50. Overall, the PD-PG exhibited consistently stronger correlations with the malaise score than the MDD-PGS, suggesting it make better capture the distress construct.

Table 11: Correlation matrix of Malaise Inventory Score variables at ages 23, 33, 42 and 50 and the polygenic scores for psychological distress and major depressive disorder in the 1958c

	Age 23	Age 33	Age 42	Age 50	MDD-PGS	PD-PGS
Age 23	1					
Age 33	0.52	1				
Age 42	0.44	0.53	1			
Age 50	0.42	0.49	0.58	1		
MDD-PGS	0.08	0.09	0.10	0.07	1	
PD-PGS	0.14	0.14	0.16	0.18	0.48	1

MDD-PGS: Major depressive disorder polygenic score. PD-PGS: Psychological distress disorder polygenic score

4.3.2 Regression Results

4.3.2.1 Research Question 1: Does the magnitude of association between polygenic score and adult psychological distress outcomes differ from age 23 to 50?

Beta Coefficients

Results from the linear regression analyses demonstrate a positive association between the polygenic scores for psychological distress and malaise inventory scores at all ages (Table 12; Figure 9). The size of the association increases with age, at the youngest age 23 in 1981, the beta coefficient was 0.21 (95% CI: 0.17-0.25) and at the oldest age point of 50, the beta coefficient was 0.32 (95% CI: 0.27-0.37).

4.3.2.2 Research Question 2: Does the amount of variance explained in adult psychological distress outcomes by polygenic scores differ by age?

R-Squared

The R-squared (R^2) values, which represent the amount of variance in the malaise inventory score explained by the polygenic score for psychological distress. The R^2 values increase with age (Table 12; Figure 9). Starting at 1.83% at age 23 rising to 2.86% by age 50, indicating a strengthening in explanatory power by age.

Table 12: Linear regression associations between polygenic score for psychological distress and observed psychological distress across ages 23-50 in the 1958c (N=6,312, Threshold=0.005)

	Mean (SD)	R^2	Beta Coefficient	Standard Error	Lower 95%CI	Upper 95%CI	*P*
Age 23	1.12 (1.46)	1.83%	0.21	0.02	0.17	0.25	<0.001
Age 33	0.89 (1.42)	2.02%	0.21	0.02	0.17	0.24	<0.001
Age 42	1.45 (1.69)	2.46%	0.27	0.02	0.23	0.32	<0.001
Age 50	1.39 (1.85)	2.86%	0.32	0.03	0.27	0.37	<0.001

R^2: amount of variance explained by the PGS. PD: psychological distress. Analyses were adjusted for sex and the first 10 principal components of ancestry.

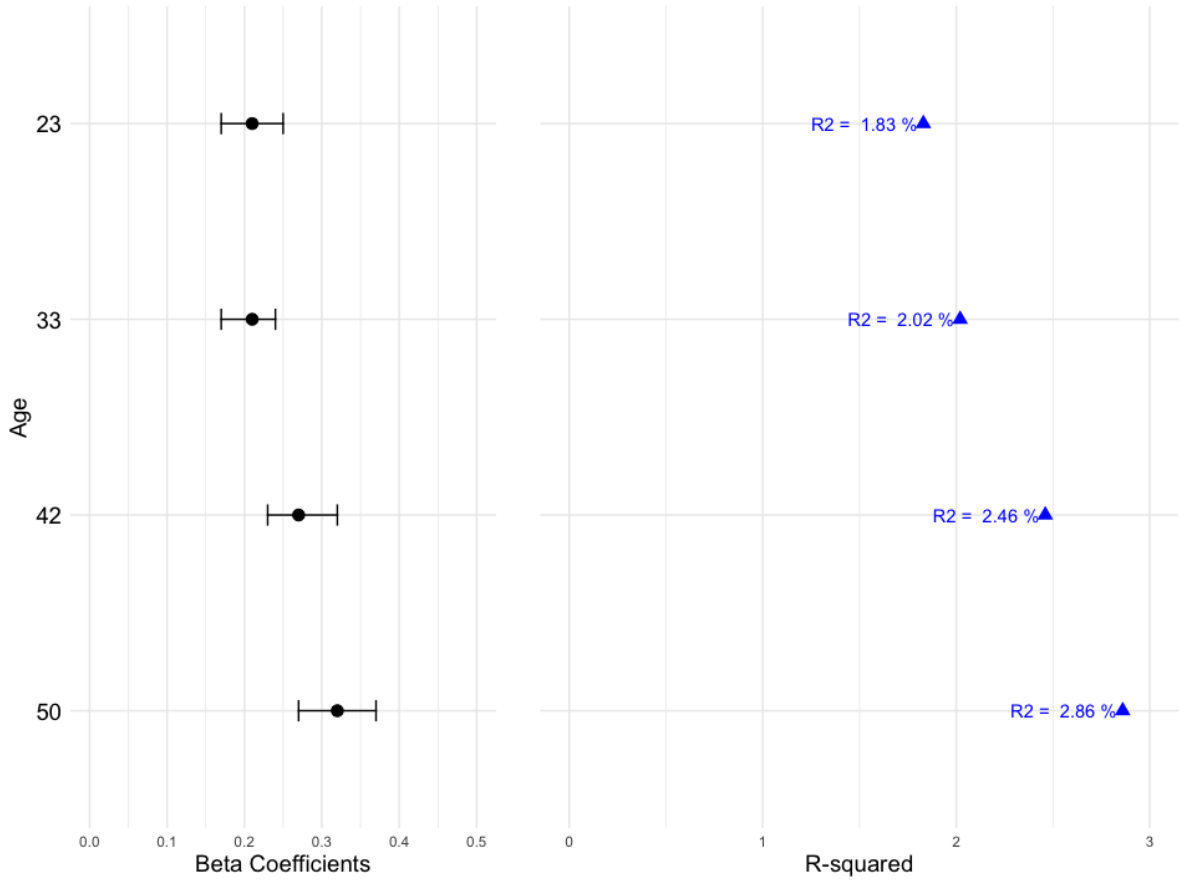


Figure 9: Plot of the linear regression results of the associations between the polygenic score for psychological distress at 0.005 p-value threshold and the Malaise Inventory score at ages 23-50 in the 1958c

4.3.2.3 Supplemental and Sensitivity Results

1) *Testing polygenic score for psychological distress at different p-value thresholds:*

To assess the impact of the selected p-value threshold on the estimates, specification curve plots were used, and the results are depicted in Supplementary Figures 1 and 2. The regression outcomes across various thresholds are presented in Supplementary Tables 3 & 4. The variance explained by genetic contribution increased as the threshold changed from 0.0001 to 0.005, confirming our results were not artifacts of threshold selection (Supplementary Figures 1 and 2). Beyond the 0.005 threshold, additional SNPs provided minimal gain in explanatory power despite increasing the risk of false positives (Table 5). This analysis validates the selection of the 0.005 threshold for main analyses and confirms that age-related increases in genetic associations remain consistent across different thresholding approaches.

2) *Comparison to a more specific phenotype:*

The estimates were comparable to the main findings, confirming robustness (Supplementary Table 4 & 5; Supplementary Figure 3). Both scores showed increasing beta coefficients with age, with the 0.005 threshold consistently emerging as optimal. The MDD-PGS produced smaller effect sizes than the PD-PGS and exhibited a different pattern of variance explained across ages, suggesting that phenotype specificity matters when studying genetic influences on mental health. In the following studies the PD-PGS will continue to be used to test the genotypic-phenotypic association.

3) *Complete Case Analysis:*

Both complete case and main analyses shared comparable age-related increases in variance explained and beta coefficients (Supplementary Table 6). In the complete case analysis, R^2 values increased from 1.80% at age 23 to 2.70% at age 50, with similar patterns in the main analysis (1.83% to 2.86%). This consistency strengthens confidence that the findings are robust and not artifacts of the multiple imputation methodology or missing data patterns.

4) *Sex Stratified Analyses:*

Sex-stratified analyses revealed stronger genetic associations with psychological distress in females compared to males across all ages (Supplementary Table 7). At age 23, among males, a 1 SD increase in polygenic score was associated with a 0.014 increase in distress (beta = 0.14, 95% CI [0.10, 0.19], p= 0.001), while females showed a higher beta coefficient of 0.27 (95% CI [0.21, 0.32], p= 0.001). Similarly, the R^2 trend showed that women had higher values at each age point. Overall, these sex-stratified analyses suggest that the association between polygenic scores and psychological distress varies by sex, subsequent studies in this thesis will continue to test for sex differences.

4.4 Discussion

4.4.1 Summary of Main Findings

The study investigated age-related differences in the strength of the association between genotypic and phenotypic psychological distress, and the extent to which polygenic scores explain variance in psychological distress from ages 23 to 50 in the 1958 cohort. This research adds to the recent literature of longitudinal studies examining the contribution of molecular genetics to distress-related phenotypes by age. This study found a positive association between polygenic for psychological distress and psychological distress increased slightly with age. The variance explained by polygenic scores also increased slightly across age. This modest but increasing genetic contribution suggested that while genetic factors contribute to psychological distress at all ages, their influence may become more pronounced with age.

4.4.2 Comparison to the literature and explanation of findings

These findings are consistent with previous research on psychological distress over similar time periods (109,110,254,257,262). The beta coefficient increased from 0.21 (95% CI: 0.17-0.25) at age 23 to 0.32 (95% CI: 0.27-0.37) at age 50, with corresponding increases in variance explained from 1.83% to 2.86%. This pattern aligns with theoretical models of genetic amplification, wherein genetic influences on complex traits become more pronounced throughout the life course, despite underlying genetic stability (263).

The Selective Optimisation with Compensation theory predicted that biological potential and genome-based plasticity decline with age, while individuals increasingly rely on cultural and social resources to maintain functioning age. As individuals age, they may increasingly select fewer but more meaningful roles and relationships, optimise their investment in effective coping strategies (e.g., therapy, exercise, social connection), and compensate for losses in resilience by adopting new supports or routines. In this way, SOC predicts that psychological distress should become less genetically determined over time, as individuals rely more on cultural and social resources to maintain emotional well-being. However, in this thesis, both PGS heritability (R-squared) and polygenic penetrance (beta coefficients) rose slightly with age, even though mean levels of distress declined between the penultimate and oldest age points. This suggests that while the average experience of distress improves in later adulthood, genetic influences may become more salient among those who continue to experience distress, contrary to the expectations of SOC model.

The observed age-related strengthening of genetic associations supports gene-environment correlation mechanisms rather than models positing that environmental exposures would overwhelm genetic influences over time. This pattern suggests that as individuals progress through adulthood, they increasingly select, modify, and construct environments that are correlated with their underlying genetic propensities for psychological distress (87,264). Importantly, these findings were observed despite widening standard deviations in malaise inventory scores (from 1.46 at age 23 to 1.85 at age 50), indicating that genetic factors explain an increasing proportion of phenotypic variance even as overall variability in distress increases

In a key study of twins from Australia, it was found that the genetic contribution to distress was stable across the lifecourse (111). This may be because the peak of the genetic contribution to the distress phenotypes was expressed by age 20. As people developed throughout childhood and adolescence, they are becoming more "themselves" as their genes are expressed through the environment (87). In early childhood, the environment is largely shaped by parents and confers passive genetic effects by genetic nurture (82). Across development, adolescents and young adults can actively select their environments, leading to an increase in active gene-environment correlation. If active selection of a person's environment exceeds the declining influence of passive gene-environment correlation (i.e., children leave the familial home), then heritability estimates may seem to rise at this transitional point.

Contrastingly, if environmental variance increases more than genetic influences, then heritability measures may seem to decrease. However, if the association between genetic factors and distress increases at the same rate as total environmental variance, then heritability measures will remain constant over time. After which, throughout adulthood, there may be a stable genetic contribution to psychological distress (107,265). If the contribution to psychological distress is stable across adulthood, then what may impact an individual's distress outcomes could be more due to environmental factors.

This study found that the explained variance increased between ages 23 and 50. The use of the incremental adjusted r-squared in this study helps to infer a proxy for heritability as it gives the proportion of variance predicting phenotypic psychological distress from genotypic psychological distress. The results are in keeping with other evidence from a 50-year-spanning meta-analytic review of twin design contributions to the heritability of distress were shown to increase over the lifecourse (42).

One possible explanation for the slight increase in variance explained by age could be through the cumulative effects of consistently expressed genes overshadowing occasion-specific environmental effects, or through the expression of a new set of genes across the lifecourse that boosts the genetic proportion of the variance (111,266). Additionally, a reduction in environmental variance could also result in increased heritability estimates, which could be achieved from reduced error variance due to older individual's ability to accurately reflect and report on their distress.

The finding of increasing variance explained by polygenic scores contrasts with some previous evidence. Multiple studies have found a decreasing genetic contribution to the variance of psychological distress. Research from the Netherlands Twin Registry following twins from the age of 3 to 63, found that during adulthood, there was a decrease in heritability due to an increase in environmental variance (107). This was replicated in a more recent study with the same data with an age range of 7 to 99 years old (108).

In another context, with the same study design, evidence from the Norwegian Twin Registry indicated that the genetic influence on the variability of distress between ages 25 and 45 followed a pattern: increasing between 25 and 27, then gradually declined until 33, subsequently rising until 37, and thereafter sharply decreased (109). Within their auto-regressive model that tested whether individual-specific genetics or

environments depended on previous observations, they found that genetic contribution effects were stable across all ages.

The present investigation addresses critical methodological limitations in the existing literature by employing consistent phenotypic measurement across timepoints, utilising contemporary polygenic scoring approaches derived from well-powered GWAS, appropriately controlling for population stratification, and examining a substantial period of the adult life course within a single prospective birth cohort. While previous molecular genetic studies have documented increasing heritability from childhood to young adulthood or examined older adults, the present findings bridge the gap in understanding how genetic influences operate across the full span of adulthood.

4.4.3 Strengths & Limitations

This study adds to the growing literature by utilising longitudinal outcomes and molecular genetic data, especially in exploring how phenotypes may shift over time and how genetic liability to these phenotypes can change accordingly. The use of the U.K. British Birth cohort offers a potentially more representative snapshot of the population born in the U.K. during that period, with the ability to integrate rich social data and genetic information. Hence, the advantage of this study is that in the main analysis, it is the same participants included across each age. Sensitivity analyses using more conservative and more liberal p-value thresholds for polygenic score showed that the estimates were largely comparable no matter the threshold specified. Sensitivity analyses using a more specific phenotype of major depressive disorder for the GWAS showed that while the estimates were largely comparable, the polygenic score for major depressive disorder explained less variance in distress compared to the broader distress phenotype, suggesting that the latter was a more appropriate choice.

A number of limitations should be noted when interpreting these findings.

First, the sample may suffer from selection bias. The nature of distress as a phenotype results in a healthy survivor bias, which exerts selection pressure over time. We know that attrition in this sample is highest at age 50, whereby people who were categorised with high psychological distress at age 23 had 67% higher odds of not participating at age 50. This could mean there is less variability as the people with the highest distress scores are not captured due to attrition, which could create an underestimation of the explained variance.

Secondly, participants were drawn from only a representative sample of people born in Britain in 1958. This limited the sample to European ancestry, which means the results are not generalisable to populations of differing ancestry (267). As heritability and genetic contribution to a phenotype are specific to the population in a given time, it would be beneficial for future studies to explore these findings in other birth cohorts; this work formed the basis of the follow-on study in this PhD thesis.

Finally, there are inherent challenges posed by autocorrelation and missingness within the dataset which the linear regressions alone do not account for. The nature of longitudinal studies introduces complexities such as autocorrelation, where

observations at one-time point are correlated with observations at nearby time points. Moreover, missing data can introduce bias and compromise the generalisability of the findings. Although multiple imputation was used to address this. Multilevel mixed-effect model approach can be used to account for both fixed effects, such as age-related trends, and random effects, representing individual variability, offering a more robust framework for handling autocorrelation and missing data issues. This was then used as the approach for Studies 2 and 3.

4.4.4 Conclusion

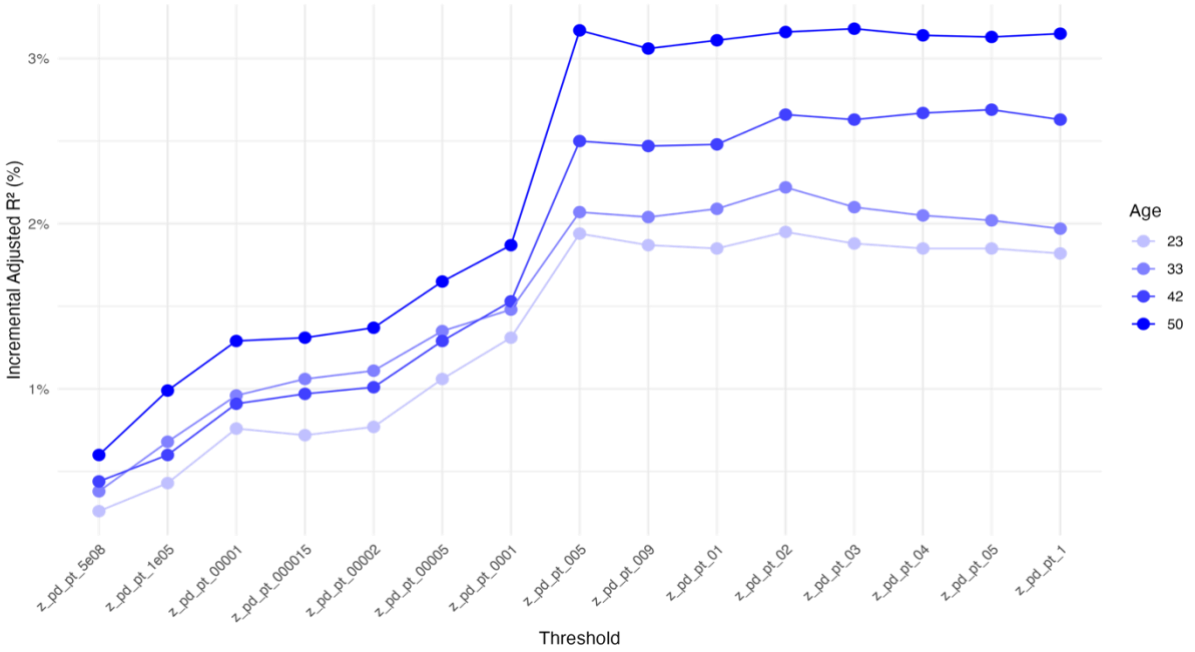
The study findings suggest that there are age-related differences in the association between a polygenic score for distress and phenotypic psychological distress from ages 23 to 50 in the 1958 cohort. The findings revealed a positive association between polygenic scores and psychological distress that persisted across adulthood, with the strength of this association increasing slightly with age. The variance explained by polygenic scores also increased slightly across age.

The finding of increasing genetic influence with age may be specific to this 1958 birth cohort and the unique historical period in which they lived. As such, examining whether this pattern exists in cohorts born in different time periods is essential to determine if this represents a universal developmental pattern or a cohort-specific phenomenon. This directly motivates the next study's research question, which uses the 1970 British Cohort Study with comparable longitudinal follow-up and genetic data to examine whether the relationship between polygenic scores and observed distress follows similar age-related patterns across different historical contexts.

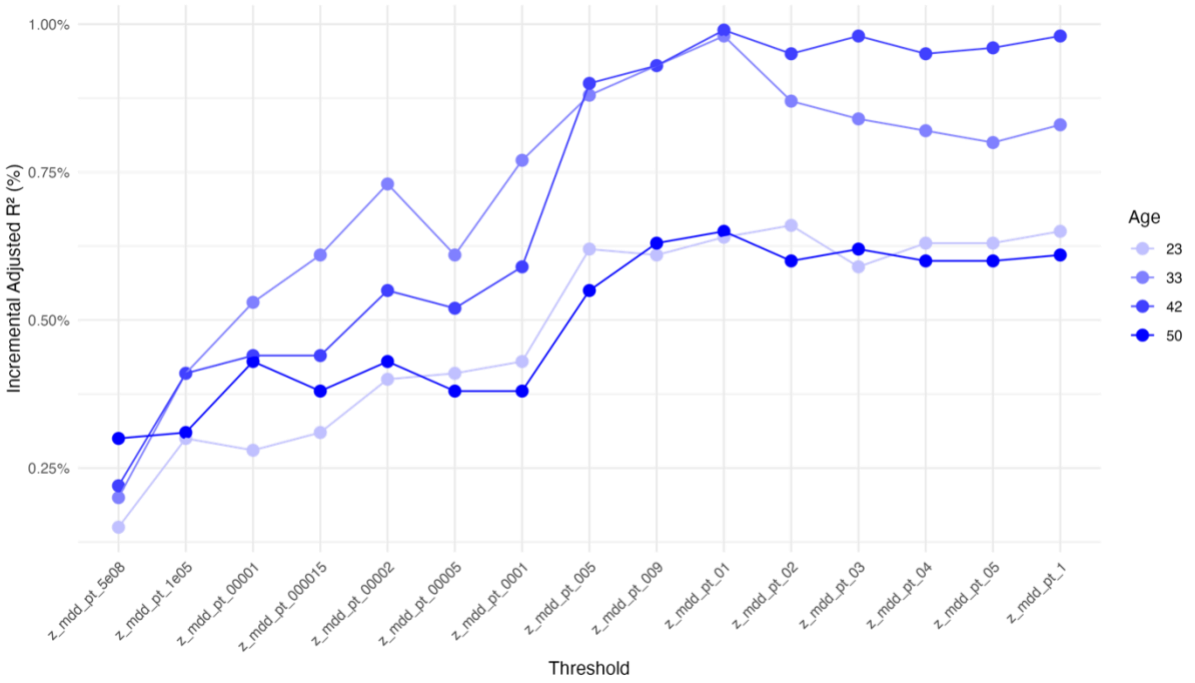
4.5 Study 1 Summary

- The genetic contribution to psychological distress increases as the 1958c sample ages from 23 to 50.
- The variance explained by polygenic scores also increased by age.
- Sex-stratified analyses found the polygenic score association with psychological distress by age was stronger in females compared to males.
- Sensitivity analyses showed that the estimates were largely comparable no matter the threshold specified or whether it was complete case or multiple imputation.
- It would be beneficial for future studies to replicate these analyses in other cohorts; this work formed the basis of the follow-up study in this PhD thesis.

4.6 Supplementary Material



Supplementary Figure 1: Specification curve plot of the variance explained by the polygenic score for psychological distress in the linear regression model at each age at each potential p-value threshold in the 1958c



Supplementary Figure 2: Specification curve plot of the variance explained by the polygenic score for major depressive disorder in the linear regression model at each age at each potential p-value threshold in the 1958c

Supplementary Table 3: Linear regression associations across all thresholds of the polygenic score for psychological distress and observed psychological distress across ages 23-50 in the 1958c (N = 6,312)

Age	Threshold	R^2	Effect Size	Lower 95% CI	Upper 95%CI	*P*
23	$5.00x10^{-08}$	0.26%	0.06	0.02	0.10	0.001
	$1.00x10^{-08}$	0.43%	0.09	0.05	0.13	0.001
	0.00001	0.76%	0.12	0.08	0.16	0.001
	0.000015	0.72%	0.12	0.08	0.16	0.001
	0.00002	0.77%	0.12	0.08	0.16	0.001
	0.00005	1.06%	0.15	0.11	0.19	0.001
	0.0001	1.31%	0.17	0.13	0.20	0.001
	0.005	1.94%	0.20	0.17	0.24	0.001
	0.009	1.87%	0.20	0.16	0.24	0.001
	0.01	1.85%	0.20	0.16	0.24	0.001
	0.02	1.95%	0.20	0.17	0.24	0.001
	0.03	1.88%	0.20	0.16	0.24	0.001
	0.04	1.85%	0.20	0.16	0.24	0.001
	0.05	1.85%	0.20	0.16	0.24	0.001
	1	1.82%	0.20	0.16	0.23	0.001
33	$5.00x10^{-08}$	0.38%	0.08	0.04	0.12	0.001
	$1.00x10^{-08}$	0.68%	0.11	0.07	0.15	0.001
	0.00001	0.96%	0.14	0.10	0.17	0.001
	0.000015	1.06%	0.14	0.11	0.18	0.001
	0.00002	1.11%	0.15	0.11	0.19	0.001
	0.00005	1.35%	0.16	0.13	0.20	0.001
	0.0001	1.48%	0.17	0.14	0.21	0.001
	0.005	2.07%	0.21	0.17	0.24	0.001
	0.009	2.04%	0.20	0.17	0.24	0.001
	0.01	2.09%	0.21	0.17	0.24	0.001
	0.02	2.22%	0.21	0.18	0.25	0.001
	0.03	2.10%	0.21	0.17	0.24	0.001
	0.04	2.05%	0.20	0.17	0.24	0.001
	0.05	2.02%	0.20	0.17	0.24	0.001
	1	1.97%	0.20	0.16	0.24	0.001
42	$5.00x10^{-08}$	0.44%	0.10	0.06	0.15	0.001
	$1.00x10^{-08}$	0.60%	0.12	0.08	0.17	0.001
	0.00001	0.91%	0.16	0.11	0.20	0.001
	0.000015	0.97%	0.16	0.12	0.21	0.001
	0.00002	1.01%	0.17	0.12	0.21	0.001
	0.00005	1.29%	0.19	0.15	0.23	0.001
	0.0001	1.53%	0.21	0.17	0.25	0.001
	0.005	2.50%	0.27	0.23	0.31	0.001
	0.009	2.47%	0.27	0.23	0.31	0.001
	0.01	2.48%	0.27	0.23	0.31	0.001
	0.02	2.66%	0.28	0.24	0.32	0.001
	0.03	2.63%	0.28	0.23	0.32	0.001
	0.04	2.67%	0.28	0.24	0.32	0.001

	0.05	2.69%	0.28	0.24	0.32	0.001
	1	2.63%	0.28	0.23	0.32	0.001
50	$5.00x10^{-08}$	0.60%	0.14	0.09	0.19	0.001
	$1.00x10^{-08}$	0.99%	0.18	0.13	0.23	0.001
	0.00001	1.29%	0.21	0.16	0.26	0.001
	0.000015	1.31%	0.21	0.16	0.26	0.001
	0.00002	1.37%	0.22	0.17	0.26	0.001
	0.00005	1.65%	0.24	0.19	0.29	0.001
	0.0001	1.87%	0.26	0.21	0.30	0.001
	0.005	3.17%	0.33	0.28	0.38	0.001
	0.009	3.06%	0.33	0.28	0.37	0.001
	0.01	3.11%	0.33	0.28	0.38	0.001
	0.02	3.16%	0.33	0.28	0.38	0.001
	0.03	3.18%	0.33	0.28	0.38	0.001
	0.04	3.14%	0.33	0.28	0.38	0.001
	0.05	3.13%	0.33	0.28	0.38	0.001
	1					

Threshold: *P*-value threshold for the PGS (i.e., threshold for the number of significant SNPs from the original GWAS included into the PGS). R^2: amount of variance explained by the PGS. PD: psychological distress. Analyses were adjusted for sex and the first 10 principal components of ancestry.

Supplementary Table 4: Linear regression associations across all thresholds of the polygenic score for major depressive disorder and observed psychological distress across ages 23-50 in the 1958c (N = 6,312)

Age	Threshold	R^2	Effect Size	Lower 95% CI	Upper 95%CI	*P*
23	$5.00x10^{-08}$	0.15%	0.03	-0.01	0.07	0.111
	$1.00x10^{-08}$	0.30%	0.07	0.03	0.11	0.001
	0.00001	0.28%	0.06	0.02	0.10	0.001
	0.000015	0.31%	0.07	0.03	0.11	0.001
	0.00002	0.40%	0.08	0.04	0.12	0.001
	0.00005	0.41%	0.08	0.04	0.12	0.001
	0.0001	0.43%	0.09	0.05	0.12	0.001
	0.005	0.62%	0.11	0.07	0.15	0.001
	0.009	0.61%	0.11	0.07	0.15	0.001
	0.01	0.64%	0.11	0.07	0.15	0.001
	0.02	0.66%	0.11	0.07	0.15	0.001
	0.03	0.59%	0.11	0.07	0.15	0.001
	0.04	0.63%	0.11	0.07	0.15	0.001
	0.05	0.63%	0.11	0.07	0.15	0.001
	1	0.65%	0.11	0.07	0.15	0.001
33	$5.00x10^{-08}$	0.20%	0.05	0.01	0.09	0.010
	$1.00x10^{-08}$	0.41%	0.08	0.04	0.12	0.001
	0.00001	0.53%	0.10	0.06	0.13	0.001
	0.000015	0.61%	0.11	0.07	0.14	0.001
	0.00002	0.73%	0.12	0.08	0.15	0.001
	0.00005	0.61%	0.11	0.07	0.14	0.001
	0.0001	0.77%	0.12	0.08	0.16	0.001
	0.005	0.88%	0.13	0.09	0.17	0.001
	0.009	0.93%	0.14	0.10	0.17	0.001
	0.01	0.98%	0.14	0.10	0.18	0.001
	0.02	0.87%	0.13	0.09	0.17	0.001
	0.03	0.84%	0.13	0.09	0.17	0.001
	0.04	0.82%	0.13	0.09	0.16	0.001
	0.05	0.80%	0.12	0.09	0.16	0.001
	1	0.83%	0.13	0.09	0.17	0.001
42	$5.00x10^{-08}$	0.22%	0.06	0.02	0.11	0.001
	$1.00x10^{-08}$	0.41%	0.10	0.06	0.14	0.001
	0.00001	0.44%	0.10	0.06	0.15	0.001
	0.000015	0.44%	0.10	0.06	0.15	0.001
	0.00002	0.55%	0.12	0.08	0.16	0.001
	0.00005	0.52%	0.11	0.07	0.16	0.001
	0.0001	0.59%	0.12	0.08	0.17	0.001
	0.005	0.90%	0.16	0.11	0.20	0.001
	0.009	0.93%	0.16	0.12	0.20	0.001
	0.01	0.99%	0.17	0.12	0.21	0.001
	0.02	0.95%	0.16	0.12	0.21	0.001
	0.03	0.98%	0.17	0.12	0.21	0.001
	0.04	0.95%	0.16	0.12	0.21	0.001
	0.05	0.96%	0.16	0.12	0.21	0.001
	1	0.98%	0.17	0.12	0.21	0.001

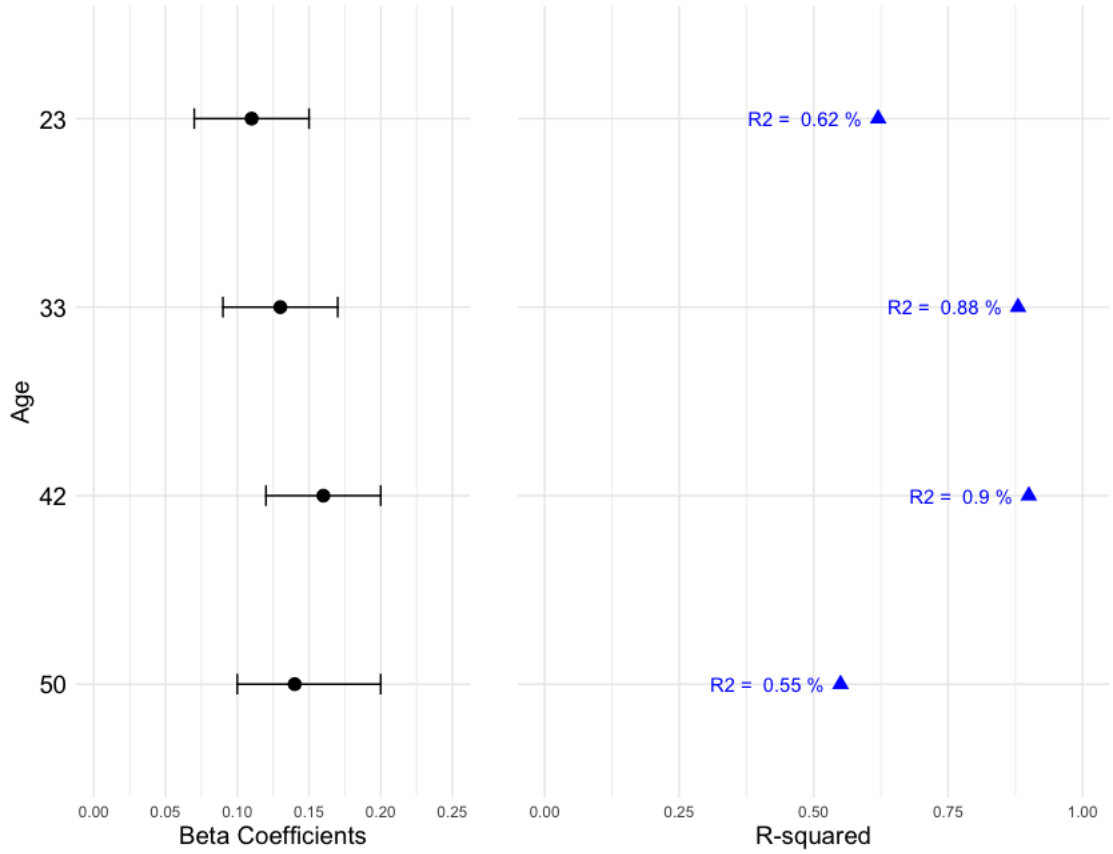
50	$5.00x10^{-08}$	0.30%	0.09	0.04	0.14	0.001
	$1.00x10^{-08}$	0.31%	0.09	0.04	0.14	0.001
	0.00001	0.43%	0.11	0.06	0.16	0.001
	0.000015	0.38%	0.10	0.05	0.15	0.001
	0.00002	0.43%	0.11	0.06	0.16	0.001
	0.00005	0.38%	0.10	0.05	0.15	0.001
	0.0001	0.38%	0.10	0.05	0.15	0.001
	0.005	0.55%	0.13	0.08	0.18	0.001
	0.009	0.63%	0.14	0.09	0.19	0.001
	0.01	0.65%	0.14	0.09	0.19	0.001
	0.02	0.60%	0.14	0.09	0.19	0.001
	0.03	0.62%	0.14	0.09	0.19	0.001
	0.04	0.60%	0.14	0.09	0.19	0.001
	0.05	0.60%	0.14	0.09	0.19	0.001
	1	0.61%	0.14	0.09	0.19	0.001

Threshold: *P*-value threshold for the PGS (i.e., threshold for the number of significant SNPs from the original GWAS included into the PGS). R^2: amount of variance explained by the PGS. PD: psychological distress. Analyses were adjusted for sex and the first 10 principal components of ancestry.

Supplementary Table 5: Linear regression associations between polygenic score for major depressive disorder and observed psychological distress for ages 23-50 in the 1958c (N=6,312, Threshold=0.005)

	Mean (SD)	R^2	Beta Coefficient	Standard Error	Lower 95%CI	Upper 95%CI	*P*
Age 23	1.12 (1.46)	0.62 %	0.11	0.02	0.07	0.15	<0.001
Age 33	0.89 (1.42)	0.88 %	0.13	0.02	0.09	0.17	<0.001
Age 42	1.45 (1.69)	0.90 %	0.16	0.02	0.12	0.20	<0.001
Age 50	1.39 (1.85)	0.55 %	0.14	0.03	0.10	0.20	<0.001

R^2: amount of variance explained by the PGS. Analyses were adjusted for sex and the first 10 principal components of ancestry.



Supplementary Figure 3: Plot of the linear regression results of the associations between the polygenic score for major depressive disorder at 0.05 p-value threshold and observed psychological distress at ages 23-50 in the 1958c

Supplementary Table 6: Complete case sample per age linear regression associations between polygenic score for psychological distress and observed psychological distress across ages 23-50 in the 1958c (N= 3,748, Threshold = 0.005)

Age	Mean (SD)	R^2	Beta Coefficient	Std Error	95%CI	*P*
23	1.25(1.59)	1.80%	0.19	0.02	0.15-0.24	<0.001
33	1.02(1.55)	1.90%	0.18	0.02	0.14-0.23	<0.001
42	1.52(1.79)	2.30%	0.25	0.02	0.19-0.30	<0.001
50	1.49(1.94)	2.70%	0.30	0.03	0.28-0.36	<0.001

R^2: amount of variance explained by the PGS. Analyses were adjusted for sex and the first 10 principal components of ancestry.

Supplementary Table 7: Stratified by sex linear regression associations between polygenic score for psychological distress and observed psychological distress score for ages 23-50 in the 1958c (Female N=3,172; Male N=3,140, Threshold=0.005)

	Sex	Mean (SD)	R^2	Beta Coefficient	Standard Error	Lower 95%CI	Upper 95%CI	*P*
Age 23	Male	0.73 (1.17)	1.30%	0.14	0.03	0.10	0.19	<0.001
	Female	1.47 (1.60)	2.70%	0.27	0.04	0.21	0.32	<0.001
Age 33	Male	0.62 (1.18)	1.65%	0.16	0.03	0.11	0.21	<0.001
	Female	1.14 (1.57)	2.30%	0.24	0.03	0.19	0.30	<0.001
Age 42	Male	1.19 (1.58)	2.20%	0.25	0.03	0.23	0.36	<0.001
	Female	1.68 (1.73)	2.60%	0.29	0.04	0.11	0.25	<0.001
Age 50	Male	1.10 (1.64))	2.10%	0.26	0.03	0.20	0.32	<0.001
	Female	1.66 (1.99)	3.40%	0.38	0.04	0.31	0.45	<0.001

R^2: amount of variance explained by the PGS. PD: psychological distress. Analyses were adjusted for the first 10 principal components of ancestry.

5 Study 2: Cohort differences in the association between polygenic scores and psychological distress: evidence from the 1958 and 1970 British Birth Cohorts

Main objective: to examine whether the polygenic score for psychological distress is associated with phenotypic psychological distress using the 1958c from the previous chapter and the 1970 British Cohort Study (1970c).

Research Question 1: Does the association between polygenic score and adult psychological distress outcomes differ at all age points in the pooled 1958c and 1970c?

Research Question 2: Does the association between polygenic score and adult psychological distress outcomes differ by cohort?

Research Question 3: Does the association between polygenic score and adult psychological distress outcomes differ by sex?

5.1 Introduction

5.1.1 Summary from Previous Study

In the previous study, the association between polygenic scores for distress and observed distress increased as the 1958c sample aged from 23 to 50. Yet, the findings may not be generalisable. This is due to the unique historical context, societal changes, and evolving environmental factors experienced by the 1958c cohort. These factors contribute to cohort effects. Cohort effects occur when individuals born in the same year may have distinct experiences that influence their outcomes (in this case, psychological distress) compared to those born in other periods.

There is evidence of a change in observed psychological distress by cohort, with the 1970c cohort having higher mean levels of distress from their early 20s (124,268). The observed cohort difference might be due to the contribution of genetics and the environment having changed; this is explained further below. Therefore, this chapter compares associations between the polygenic score for psychological distress and phenotypic psychological distress in adulthood, utilising data from both 1958c and 1970c.

5.1.2 Cohort and sex differences in psychological distress

Depressive and anxiety disorders affect an estimated ~500 million people worldwide; the global prevalence is projected to be the leading cause of morbidity by 2030 (31,32). The global prevalence of distress has exhibited a concerning upward trend over time. This escalation has been particularly pronounced in the United States (U.S.A) and the United Kingdom (U.K), where rates of psychological distress are higher in younger generations (31,123,269). As evidenced by research using the 1958c and 1970c (124). Higher distress in younger cohorts has been observed in multiple countries including the U.S., Canada, Germany, and Japan (269–274).

There have been profound changes in society across the 20th and 21st centuries. As a result, the environments in which individuals develop have changed over time, resulting in divergent socioeconomic circumstances for those born in 1900, 1950, or 2000. Social and economic factors can change more rapidly over time than population genetics (275). Known socioeconomic determinants of distress include, but are not limited to, socioeconomic inequalities, changing social norms, and time expenditure, as explored above in shaping the wider environment of a generation (128). This suggests environmental changes have played a significant role in the rising levels of psychological distress.

Sex differences are also found in psychological distress outcomes by birth cohort (15,30,31). In the U.K. context, evidence suggests that the gender gap in distress has widened across successive generations, particularly among more recent cohorts of women (15,29). This widening gap may reflect both changes in structural inequalities and shifting sociocultural expectations over time (32–34). For example, younger cohorts of women entered adulthood during periods of increasing labour market participation and educational attainment, yet these gains may not have translated into equivalent improvements in mental health due to persisting or newly emerging strains such as work-family conflict, labour market precarity and gendered caregiving

responsibilities (21,35,36). Analysing how differences between males and females vary by cohort is therefore important, as cross-cohort differences may be stronger in females due to greater change in adversities.

5.1.3 How do cohort effects differ to period and age effects in psychological distress?

Prior research has examined potential age, cohort, and period effects, aiming to understand why there may be differences in psychological distress over time (122). Age effects pertain to changes that occur with an individual's advancing age, whereas period effects relate to changes experienced by all individuals, irrespective of their age, in a specific period (for example, during a war) (122). Cohort effects refer to differences experienced by groups of individuals born during the same time (122). In the context of the aeti9ology of distress, puberty is an example of an age effect, increasingly social isolation may be a period effect, and stress from a delayed entry to the labour market due to a recession could be a cohort effect.

Recent cohort comparison research in the U.K., provides an example of potential cohort effects. Children born in 2000-2002 had an earlier onset of emotional problems compared to those born in 1991-1992. They also had higher distress scores by age 9 (102). These cohorts were only 10 years apart and yet showed differences in the prevalence of distress. Nevertheless, this phenomenon could also be attributed to a period effect.

The study hypothesised multiple factors that may influence these specific cohort differences between 2000-2002 and 1991-1992, not limited to: "*young people's lifestyles, body image, use of digital technology, family life, school life, academic pressure, social relationships, and broader cultural factors such as increasing societal inequalities" (p.7)* (102). A study using the same 2000-2002 cohorts identified correlates of psychological distress, including low-income family, chronic illness, peer and sibling relationships, and parental health (283). Each of these factors may change between cohorts and, therefore, become part of a cohort effect due to shifting socioeconomic norms. Such differences highlight the importance of considering cohort effects when studying psychological distress.

5.1.4 Societal changes that may impact psychological distress outcomes

Changes in U.K. society between 1958, 1970 the 2010's follows a similar pattern to other anglophone societies. Research attributes the mounting distress to widening inequality (7,148,284–288). Furthermore, the shift from manual to non-manual labour, meant educational access became a key polarising factor. Deindustrialisation, particularly in the U.K., accentuates regional divides, as those benefitting from the knowledge economy concentrate wealth in affluent areas, perpetuating intergenerational inequality (285,286). Systematic reviews and meta-analyses of observational studies have found a risk ratio of 1.19 (95% CI: 1.07-1.31) for distress in populations with larger income inequality relative to populations with lower inequality (289–292). Income inequality operates as a distal determinant that affects the broader environment of the population. However, it should be noted that much of the evidence is ecological in nature, examining population-level associations rather than individual level effects, which limits causal inference.

As each cohort aged, they experienced different socioeconomic circumstances at similar life stages. The 1958 cohort began with higher rates of homeownership and a lower proportion of women in the workforce than the 1970 cohort (p.74 and p.194) (139). Moreover, while the 1958 cohort experienced lower levels of economic inequality in their early adulthood, the Gini index rapidly increased over the 80s and 90s, peaking at 38.8 in 2000, meaning as the 1970 cohort entered the labour market income inequality was. higher (293). Throughout this period, public expenditure was constricted, with investment in building housing and infrastructure such as railways, hospitals, and schools peaking in 1970 (294). Subsequent decades led to a superficial boom in economic growth, from which the 1958 cohort was more likely to benefit. Hence, the 1970 cohort faced a more polarised landscape, with reduced social mobility and increased reliance on dual-income households (p.100-103) (139).

For instance, as stated, individuals from the 1970 cohort tended to stay longer in higher education. Still, they faced greater polarisation in educational attainment, as those with the highest qualifications increased by ~5-7% and those without qualifications increased by 2% (p.38) (139). This was more marked in the general population, where the percentage of people with full-time education at age 16 increased from 11.1% in 1960 to 20.2% by 1970 (146). This resulted in the number of students obtaining a bachelors doubling (146). The extension in the transition to adulthood and access to the labour market delayed entry into homeownership and childrearing, coinciding with an increase in women's economic activity (295). Coupled with longer work hours during their 30s, these macrosocial trends characterised the experience of the 1970 cohort (295). All these factors coalesced to lower leisure time alongside stagnating wages and a need for dual-income households to have the same material living standards as the previous generation, potentially decreasing satisfaction with the quality of life (296,297). Overall, there has been material improvement in living standards, but other psychosocial factors have worsened quality of life may have reduced (296,298). These shifts in societal norms and economic dynamics may contribute to heightened levels of psychological distress observed among individuals born in 1970 compared to their counterparts born in 1958.

Both cohorts experienced a period of major labour market transition throughout the 1980s; for those born in 1970, these effects were not felt until they reached the end of the 1980s. Unemployment across the period in which both cohorts were economically active rose steadily, and with a shift in job quality, there was potential for changing perceptions around purpose derived from work due to underemployment (145,299,300). This period was characterised by rapid deindustrialisation, as the proportion of manual work in various sectors diminished substantially, resulting in a shift from an economy reliant on manufacturing to one focused on services (145,301). Consequently, skills that were once valued in the job market were no longer relevant. Mass reskilling programmes did not facilitate the expected reabsorption of workers from the manufacturing sector into the service sector (302).

Alongside labour market swings was an expansion of women in the workforce plus the need for dual-income households, leading to less time in the home and less time for engaging with community-based activities (149,296). This might have led to a lowering in community and social cohesion over time, albeit expanding the inclusion of women into contributing to the UK economy and having more financial autonomy (303). Rising

female workforce participation is a positive step for gender equality and economic growth, but it does not automatically enhance social cohesion due to the role women play in kin-keeping and society's social functioning. Social cohesion, defined by trust, shared values, and belonging, can falter due to widening socioeconomic disparities (304). Unequal distribution of opportunities can worsen disparities in education, healthcare, and housing, fostering division (305). Marginalisation and discrimination further weaken trust. Economic shifts, like the move from manual to non-manual labour, also disrupt community ties for both genders, leading to isolation, while technological advances and globalisation may exacerbate feelings of disconnection (288). These changes could have influenced each cohort in distinct ways, given that they were at various ages when they took place. It is possible that there are interactions between the period, age, and cohort effects (122).

5.1.5 Theoretical model for the gene-cohort interaction of the current study

In understanding the potential cohort differences in the association of genetic factors with psychological distress, various theoretical models have been proposed to explain the interplay between genetic liability and environmental influences. There are five main theoretical models behind gene-environment interactions: the diathesis-stress model, the bioecological (social compensation) model, the differential susceptibility model, the biological sensitivity to context model and the social control model (92,93) (as defined and reviewed in Introduction Section 1).

Each model predicts how genes and the environment interact for a trait differently. It may be that for a trait such as psychological distress, each is important in understanding the varied aetiology and nosology of distress. Therefore, these theoretical models are not mutually exclusive in a population and often overlap in their explanations (100). Moreover, the precise mechanisms underlying gene-environment interactions remain largely unspecified, leading to what is often referred to as a "black box" explanation, i.e. the inner machinations of the process are opaque while the input and output are visible. Nonetheless, these models provide valuable frameworks for understanding how genetic liability may interact with environmental factors to shape psychological distress outcomes.

The diathesis-stress model has largely been the dominant model in psychiatry (101). However, the differential susceptibility hypothesis offers a more flexible theoretical approach to understanding cohort differences in psychological distress. This hypothesis suggests that individuals most adversely affected by negative environments may also benefit the most from positive environments (91). The differential susceptibility hypothesis is closely related to the concept of biological sensitivity to context (93). The theory originates from evolutionary developmental biology, where natural selection favours genotypes that exhibit a wide range of phenotypes in response to environmental conditions. They posit this adaptability is needed for the survival of the species as a whole (101).

In the context of the current thesis study, which aimed to explore cohort differences in the association of genotypic and phenotypic psychological distress, the most relevant theoretical framework is the social control model (99). As no characteristics of the environment are being tested, rather, the birth year acts as an exogenous environment between the two cohorts, enabling comparison of the relative strength of the

association between genotype and phenotype. This model aligns with the study's focus on examining how genotype-phenotype associations may differ in response to environmental alterations such as social norms and structural constraints. Due to the significant changes in the economic, social, and physical environment during this time, it is plausible that the relationship between genotype and its associated phenotype has also altered. Investigating whether there is a cohort difference in the association between genotype and phenotype of psychological distress in the 1970 and 1958 cohorts enables future studies to better understand what factors might be contributing to the difference.

5.1.6 Prior findings of gene-cohort and gene-sex interactions with psychological distress

There were three previous studies that have investigated gene-cohort differences. One out of the three found an increase in heritability estimates of distress while two found no increasing trend in the genetic penetrance.

The study that found increasing heritability of various mental health diagnostic categories was based in Denmark by Athanasiadis (2022) which utilised a sample of 6,691,426 individuals born between 1901-2013 (306). Whereas, Conley (2016) assessed changes in genetic penetrance across birth cohorts in the U.S.A, thereby utilising birth year as a proxy for wider environmental change (75). They defined genetic penetrance as "*the association between a polygenic score (PGS) and its associated phenotype*" (36, p.1). They found that genotypic penetrance of depression had a consistent association trend across cohorts of individuals born between 1919 and 1955 in the U.S.A. This contrasts with other behavioural traits they tested, such as education, which have increasing trends in heritability and polygenic penetrance.

Both studies are using slightly different metrics of genetic contributions to distress. Penetrance and heritability can have opposing trends. The penetrance of a genetic variant might increase over time (more people with the variant are showing the trait), while the heritability of the trait may remain stable or even decrease. Instead of directly comparing penetrance and heritability, they can be used to understand different aspects of the relationship between genetics and a specific trait (in this case, psychological distress). While the heritability of distress (the proportion of distress variability explained by genetics) might be increasing over time in a population, the association between the polygenic genotype (multiple genetic factors) and the phenotype of distress may remain stable. These studies provide a foundation for investigating gene-cohort interactions in psychological distress.

The other study out of the three, by Machlitt-Northern et al. (2022), used the 1958c in a cohort comparison with Understanding Society (USoc) and the Millennium Cohort Study (MCS) (307).They identified that the polygenic score of MDD was predictive of gene-environment correlations in each cohort, with USoc having stronger estimates comparatively. A limitation of Machlitt-Northern et al. (2022) is that they did not test the association between the genotypic data via the PGS and the phenotypic data. Data was collected at different intervals and ages that did not overlap between the three cohorts used. Therefore, it would be better to investigate differences and variations between cohorts in the genetic contribution to psychological distress where the

distress measure is the same, as well as being at comparable age points in the cohorts – which this study does.

It could be possible that sex differences moderate the penetrance of polygenic scores to psychological distress by birth cohort. This possibility may be due to the observed gender gap in psychological distress outcomes in both 1958c and 1970c; whereby women have been shown to have worse distress outcomes in midlife (124). In Study 1 of this thesis, it was found that the beta coefficient estimates were higher for females compared to males in 1958c when the main analysis was stratified. This is inconsistent with the study by Conley et al., which tested whether there were sex differences in the genetic liability to depression that changed between cohorts (75). They found that there was no significant difference between the sexes in the association between genetic propensity to depression, sex, and cohort. However, Conley et al. undertook this research using the Health and Retirement Study, which is a panel study of older adults, therefore the results might not generalise to the 1958c and 1970c.

Some evidence suggesting a gene-environment interaction effect for distress and sex has come from the candidate gene literature. The serotonin transporter gene (5-HTTLPR) showed gene-environment interaction effects for depressive symptoms in females compared to males (308). However, one study found the 5-HTTLPR genotype interacted with county-level environmental factors (e.g. deprivation) to influence depressive symptoms in adolescent males but not females (309). Yet it is important to note that the candidate gene literature is not robust to replication (310). However, it means it is still pertinent to interrogate sex differences as the literature is not conclusive.

In summary, while there are some contradictory findings, multiple studies suggest sex-specific gene-environment interactions contribute to psychological distress. Therefore, this study will further explore whether there are cohort and sex differences in genetic liability to psychological distress.

5.1.7 Summary

In summary, a comparative analysis across different birth cohorts within the same country would be a helpful addition to the literature to better understand the link between genetic propensity to distress and observed psychological distress. The current study examined the relationship between genotypic and phenotypic psychological distress at different age points (RQ1), between cohorts (RQ2) and by sex (RQ3) in 1958c and 1970c.

5.2 Method

5.2.1 Data

The study used data from the 1958 National Child Development Study (1958c) and the 1970 Birth Cohort Study (1970c) (187,188), both of which are described in the thesis Methods Section 3.1.

5.2.2 Measures

5.2.2.1 Psychological Distress

The malaise inventory score was the outcome variable in both cohorts. The 9-item version was used for the current chapter. For details regarding the development, harmonisation, measurement invariance, internal validity, and external validity of the Malaise Inventory, see Methods Section 3.2.1.

5.2.2.2 Polygenic score (PGS)

Polygenic score was made using the same method of clumping and thresholding as in Study 1, however a new score was generated to only include SNPs that were found in both the 1958c and 1970c (Methods Section 3.2.2). The polygenic score was standardised across both cohorts (rather than separately) to have a mean of 0 and a standard deviation of 1. The main analysis exposure is the standardised polygenic score for psychological distress at a threshold of 0.005, which has overlapping SNPs between the 1958c and 1970c.

5.2.2.3 Covariates

Covariates included the 10 first principal components and sex as male (0) and female (1). For research questions 1 and 2, sex is included to reduce the confounding that comes from a non-random sample and potential additional response from females over males, as prior survey research has shown females were less likely to drop out of cohort studies (222). The 10 principal components are included to control for any residual population stratification (224).

5.2.3 Analytical Strategy

The thesis Methods Section 3 describes scale-level descriptives, patterns of missingness attrition and the strategy to address this is outlined.

5.2.3.1 Research Question 1*:* Does the association between polygenic score and adult psychological distress outcomes differ at all age points in the pooled 1958c and 1970c?

The outcome, malaise inventory score, is continuous; therefore, linear regressions were used. The association between the polygenic score (x) and the malaise inventory score (y) was tested at different ages on the absolute scale in 1958c and 1970c.

Linear regressions were performed to investigate the relationship between the polygenic score for psychological distress and the malaise inventory score at different time points - ages 23, 26, 30, 33, 34, 42 46 and 50. Each model controlled for 10 principal components and sex as covariates.

Effect sizes were compared visually by plotting the coefficient and incremental R^2 estimates. Beta coefficients were used to assess the strength of the association between polygenic scores and observed psychological distress. Incremental R^2 were calculated by deducting the R^2 from a model without the PGS from the full model with the PGS at each age. This was used to assess the proportion of variance in psychological distress attributable to PGS, thereby representing PGS heritability.

Multiple imputation was employed to handle missing data of the malaise inventory score for the linear regression analyses assuming that data was missing at random (261). Multiple imputations utilised auxiliary variables with 78 imputed datasets to address missingness. Auxiliary variables included sex, father's social class, the polygenic score for psychological distress, crowding in childhood, whether breastfed, whether parents divorced during participants' childhood, family difficulties due to mental illness, poor maternal mental health, low birth weight, well-being alongside internalising and externalising behaviours during childhood (232,235,236).

5.2.3.2 Research Question 2: Does the association between polygenic score and adult psychological distress outcomes differ by cohort?

To answer this research question, a two-stage approach was taken. The first analysis was run with both cohorts in a pooled dataset while the second analysis used a cohort-stratified approach.

Model 1 used a multilevel mixed-effects model to test whether there was an interaction between the polygenic score and cohort in the pooled dataset. The advantage of a multilevel mixed-effect model is that it may give more precise estimates compared to linear regressions by utilising repeated measures. To improve interpretation, the polygenic score was mean-centred.

The fixed part of the model included the polygenic score interaction with the cohort variable as well as covariates. Covariates of sex, age, age squared and the 10 principal components were included. A random intercept was also included to account for nesting of observations within individuals.

Model 2 (1958c) and 3 (1970c) were separate mixed-effects models fitted for each cohort to examine the interaction between the polygenic score and mean-centred age, with covariates included. Covariates were the 10 principal components and age squared.

From model 2 and 3, for each cohort, marginal effects of the PGS were calculated at each age, with age being mean-centred, within the follow-up range (e.g., ages 23 to 50). The marginal effect is the absolute change in predicted level of psychological distress associated with a one standard deviation increase in the polygenic score, holding all other variables constant.

To formally compare the interaction coefficient between cohorts, z-statistic was computed as the difference between coefficients divided by the square root of the sum of their squared standard errors: $(b1 - b2)/\sqrt{(SE1^2 + SE2^2)}$. This approach tests whether the difference between cohorts' coefficients is statistically significant, with values exceeding ±1.96 indicating significant differences at $p < 0.05$.

5.2.3.3 Research Question 3: Does the association between polygenic score and adult psychological distress outcomes differ by sex?

To answer this research question, a two-stage approach was taken. The first analysis was run with both cohorts in a pooled dataset while the second analysis used a cohort-stratified approach.

Model 1 used a multilevel mixed-effects model to test whether there was an interaction between the polygenic score and sex in the pooled dataset. To improve interpretation, the polygenic score was mean-centred. The fixed part of the model included the polygenic score interaction with sex, plus covariates. Covariates were the 10 principal components, cohort, age and age squared. A random intercept was also included to account for nesting of observations within individuals.

Model 2 (1958c) and 3 (1970c) were separate mixed-effects models fitted for each cohort to examine the interaction between the polygenic score and sex, plus covariates. Covariates were the 10 principal components, age and age squared.

From model 2 and 3, for each cohort, the marginal effect of the polygenic score on psychological distress was plotted for male and females. The marginal effect is the absolute change in predicted level of psychological distress associated with a one standard deviation increase in the polygenic score, holding all other variables constant.

To formally compare the sex interaction coefficient between cohorts, z-statistic was computed. This approach tests whether the difference between cohorts' coefficients is statistically significant, with values exceeding ±1.96 indicating significant differences at $p < 0.05$.

5.2.3.4 Supplemental and Sensitivity Analyses

Testing polygenic score for psychological distress at different p-value thresholds:

Specification curve plots were used to assess the impact of the selected p-value threshold on main analysis estimates. This analysis was important because the threshold selection determines which SNPs are included in the final score. During PGS construction, clumping first removes SNPs, retaining only independent signals (213). Then, thresholding excludes SNPs that do not meet the selected p-value significance, with higher thresholds including more SNPs (Table 5 summarises SNP counts at each threshold). Including this analysis strengthens confidence in the findings remaining constant across threshold specifications.

Testing the 24-item Malaise Inventory Score:

A sensitivity analysis using the 24-item Malaise Inventory Score was run. This analysis tested associations between the 24-item score measured at ages 23, 33, 42 in 1958c and ages 26 and 30 in the 1970c and the polygenic score. This was to test whether the associations were of a similar direction as with the 9-item score.

5.3 Results

5.3.1 Descriptives

5.3.1.1 Mean, Standard Deviations and Histograms

Figure 10 shows the overall trend of females have higher distress scores than males in both cohorts. Table 12 summarises the mean and standard deviation of the 9-item malaise inventory scores in the two cohorts 1958c and 1970c at ages 23, 26, 30, 33, 34, 42, 46, and 50 stratified by females versus males.

Males consistently had lower mean psychological distress scores across both cohorts. For example, in the 1958 cohort the mean distress score at age 23 for females was 1.54 (SD = 1.66), while for males, it was notably lower at 0.75 (SD = 1.19).

The 1970c cohort had higher mean psychological distress scores at equivalent ages as the 1958c cohort for both sexes. For example, at age 42, the mean distress score for females in the 1970c was 1.85 (SD = 1.90), compared to 1.80 (SD = 1.79) in the 1958c cohort. This suggests that the baseline levels of psychological distress may have increased for females in the 1970c.

The standard deviations of the psychological distress scores further illustrated the variability in distress levels within each sex. Across both sexes, males consistently showed smaller standard deviations compared to females.

Table 13: Mean and standard deviation of the Malaise Inventory Score by sex of those with genetic data in the 1958c and 1970c (Female N=; Male N=)

		Female			Male		
	Age	Mean	SD	N	Mean	SD	N
1958c	23	1.54	1.66	2,768	0.75	1.19	2,683
	33	1.22	1.63	2,859	0.64	1.20	2,750
	42	1.80	1.79	2,104	1.22	1.63	2,787
	50	1.73	2.04	2,810	1.12	1.68	2,718
1970c	26	1.97	1.71	1,829	1.20	1.48	1,480
	30	1.67	1.72	2,107	1.15	1.51	1,850
	34	1.76	1.84	1,998	1.24	1.65	1,775
	42	1.85	1.90	2,111	1.30	1.69	1,789
	50	1.90	2.13	2,176	1.44	1.94	1,979

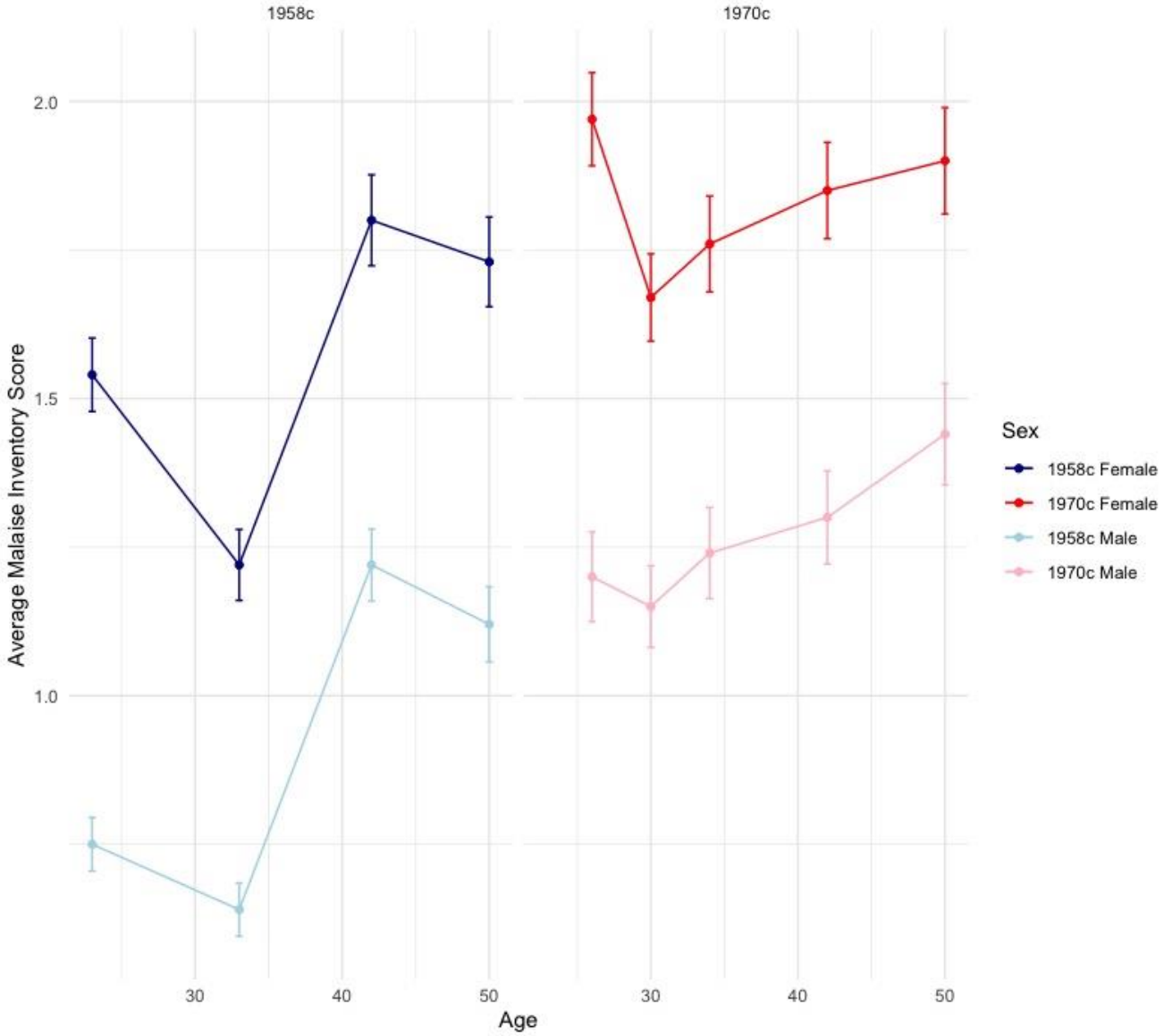


Figure 10: Mean and standard deviation of Malaise Inventory Score by sex for ages 23-50 in 1958c and 1970c (N=9,620)

5.3.2 Regression Results

5.3.2.1 Research Question 1: Does the association between polygenic score and adult psychological distress outcomes differ at all age points in the pooled 1958c and 1970c?

Beta Coefficients

Results from the linear regression analyses demonstrate positive and increasing associations between polygenic scores for psychological distress and malaise inventory scores from ages 23 to 50 in both 1958c and 1970c (Table 14; Figure 11).

In the 1958c cohort, at the youngest age 23 in 1981, the beta coefficient was 0.21 (95% CI: 0.17-0.25), indicating a positive association between genotype and phenotype of psychological distress. Similarly, in the 1970c cohort, at the youngest age 26 in 1996, the beta coefficient was 0.23 (95% CI: 0.17-0.28). The only age point that both cohorts overlap is at age 42. The beta coefficient was marginally higher in the 1958c at 0.27 (95% CI: 0.23-0.32) compared to 0.26 (95% CI: 0.21-0.33) in the 1970c, though confidence intervals overlapped. At the oldest age point of 46 in 1970c and 50 in 1958c, respectively, there is only a difference of 0.02 in beta coefficients being higher in 1958c.

R-Squared

The R-squared (R^2) values, which represent the amount of variance in the malaise inventory score explained by the polygenic score for psychological distress, increased slightly by age in both cohorts. The overall trend shows a modest contribution of genetic factors to psychological distress throughout adulthood (Table 14; Figure 12).

In the 1958c cohort, at age 23, the R^2 value was 1.83% (95% CI: 1.11%-2.52%) while in the 1970c cohort, at age 26, the R^2 value was slightly lower at 1.80% (95% CI: 1.09%-2.69%). At the overlapping age of 42, the R^2 value in the 1970c cohort is smaller than the 1958c, at 1.85% (95% CI: 1.08%-2.75%) compared to 2.46% (95% CI: 1.64%-3.22%)). At the oldest age points, the R^2 value for the 1970c cohort at age 46 was 2.16% (95% CI: 1.39%-3.19%) whereas, for the 1958c cohort at age 50, it was larger at 2.86% (95% CI: 1.94%-3.68%)

Table 14: Linear regression associations between polygenic score for psychological distress with overlapping SNPs and 9-item Malaise Inventory Score at ages 23-50 in 1958c and 1970c (PGS Threshold=0.005, 1958c N=6,312; 1970c N=5,423)

Year	Age	Cohort	Mean (SD)	R^2	R^2 95% CI	Beta Coefficient	Beta 95%CI
1981	23	1958c	1.15 (1.50)	1.83%	(1.11%-2.52%)	0.21	(0.17-0.25)
1996	26	1970c	1.63 (1.65)	1.80%	(1.09%-2.69%)	0.23	(0.17-0.28)
2000	30	1970c	1.43 (1.65)	1.44%	(0.82%-2.27%)	0.21	(0.16-0.26)
1991	33	1958c	0.94 (1.46)	2.02%	(1.25%-2.70%)	0.21	(0.17-0.24)
2004	34	1970c	1.52 (1.77)	2.10%	(1.32%-3.12%)	0.26	(0.21-0.32)
2012	42	1970c	1.74 (1.89)	1.85%	(1.08%-2.75%)	0.26	(0.21-0.33)
2000	42	1958c	1.49 (1.74)	2.46%	(1.64%-3.22%)	0.27	(0.23-0.32)
2016	46	1970c	1.68 (2,06)	2.16%	(1.39%-3.19%)	0.31	(0.24-0.36)
2008	50	1958c	1.68 (2.06)	2.86%	(1.94%-3.68%)	0.33	(0.27-0.37)

R^2: amount of variance explained by the PGS. Analyses were adjusted for sex and the first 10 principal components of ancestry.
NB: blue indicates the 1958c, and red indicates the 1970c

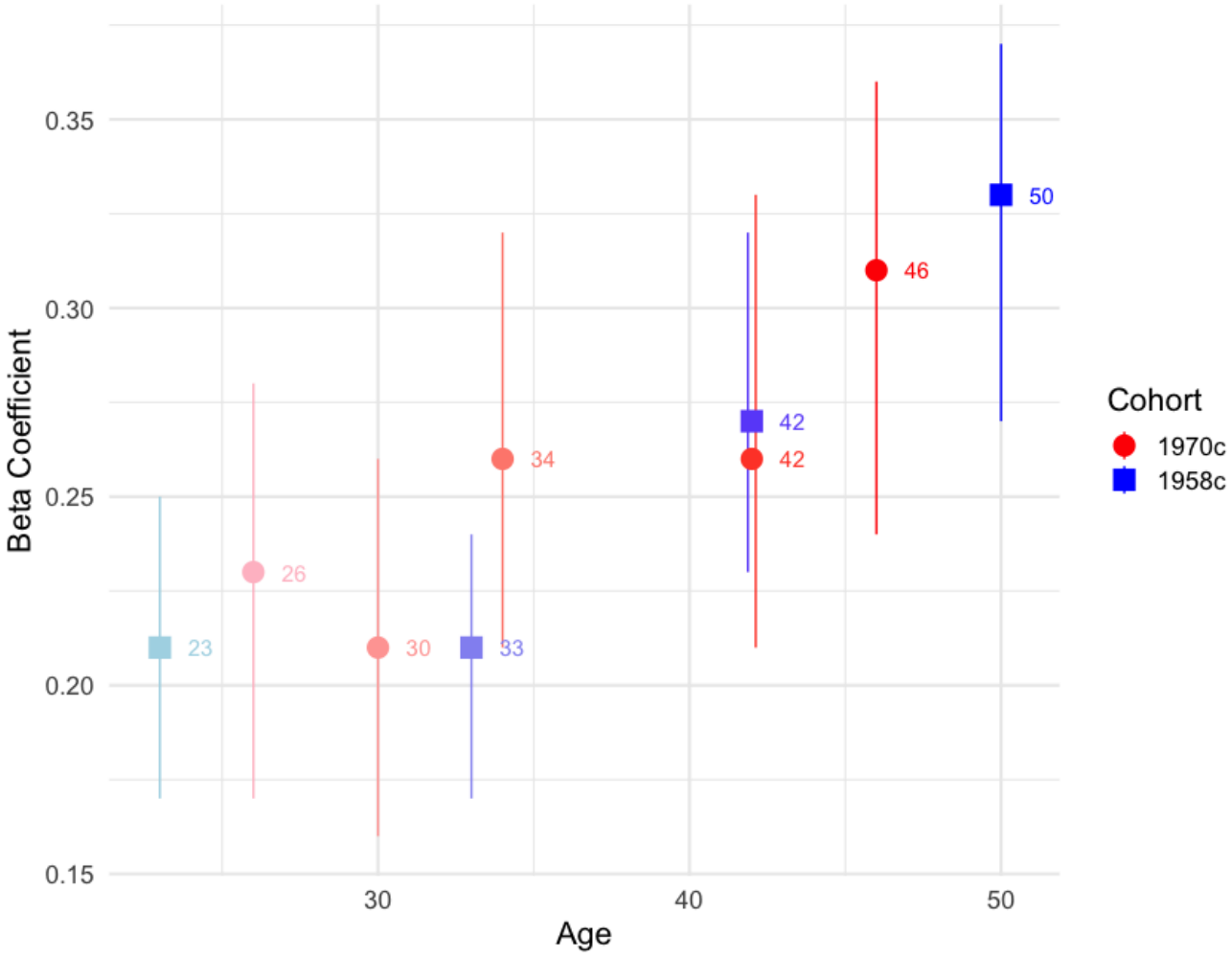


Figure 11: Plot of the beta coefficients from the linear regression results of the association between the polygenic score for psychological distress with overlapping SNPs at 0.005 p-value threshold and psychological distress outcomes in 1958c and 1970c between 1981-2012

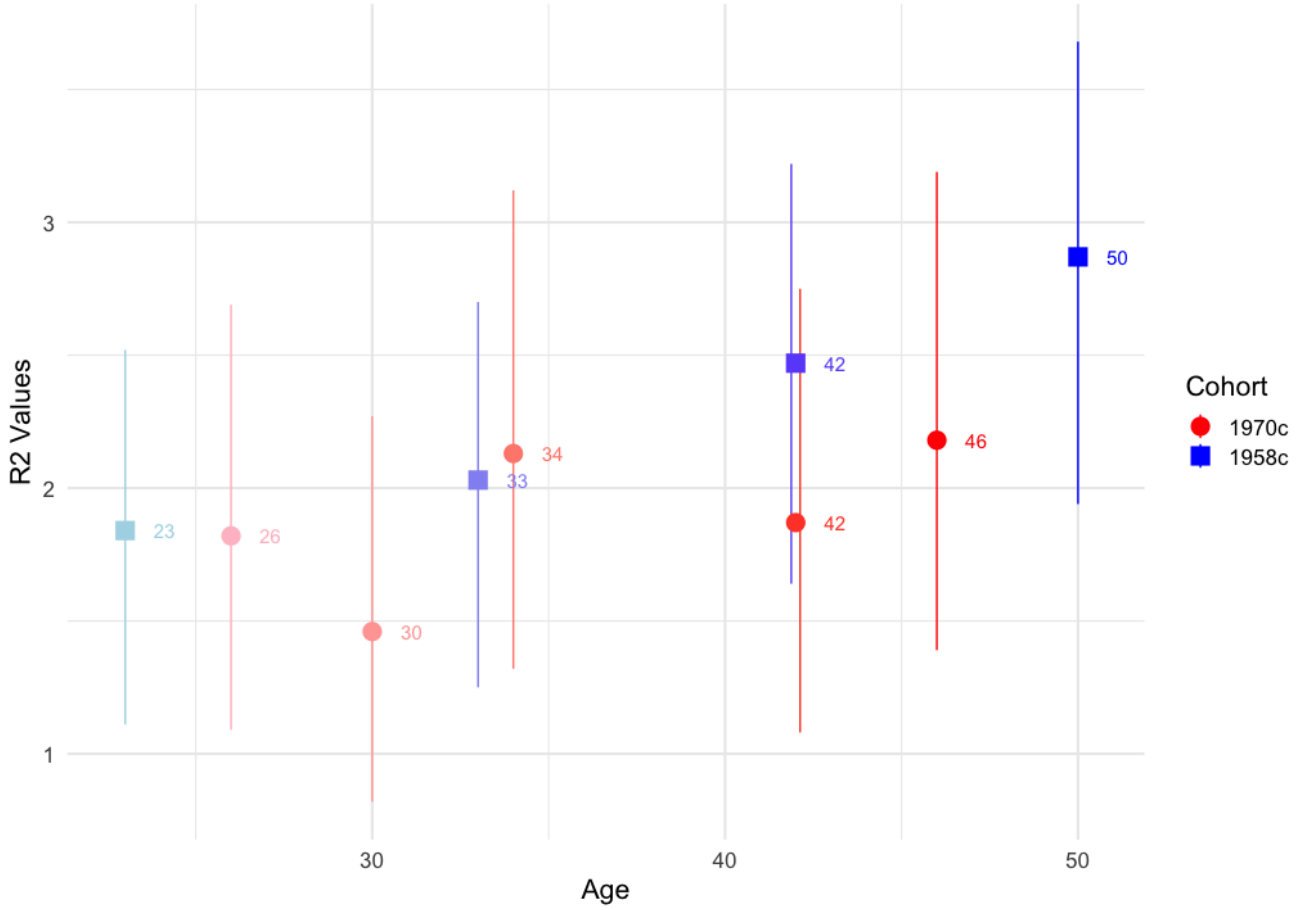


Figure 12: Plot of the incremental adjusted r-squared from the linear regression results of the association between the polygenic score for psychological distress with overlapping SNPs at 0.005 p-value threshold and psychological distress outcomes in 1958c and 1970c between 1981-2012

5.3.2.2 Research Question 2: Does the association between polygenic score and adult psychological distress outcomes differ by cohort?

Pooled Analysis: Model 1

The results from Model 1 indicate that, on average, individuals born in the 1958 cohort have lower psychological distress scores compared to those born in the 1970 cohort, as shown by a negative beta coefficient of -0.33 (95% CI [-0.39 - -0.28], p = 0.001) (Table 15). This implies a downward shift in the linear prediction for the 1958 cohort. The polygenic score shows a positive association with psychological distress (β = 0.26, 95% CI [0.22-0.29], p = 0.001), indicating that higher polygenic score values correspond to higher distress scores.

There was no evidence of an interaction between the polygenic score and cohort in the pooled analysis (β = -0.006, 95% CI [-0.06-0.05], p = 0.831), suggesting that the effect of genetic liability for psychological distress does not differ between the two cohorts.

Figure 13 shows predicted levels of psychological distress across the polygenic score distribution. The slopes are near parallel for each cohort, meaning the effect of the polygenic score is potentially similar thus does not differ.

Cohort-stratified: Model 2 and 3

The results from models 2 and 3 indicate evidence for an interaction between polygenic scores and age in each cohort when stratified (1958c: β = 0.004, 95% CI [0.003-0.005], p = 0.001; 1970c: β = 0.004, 95% CI [0.001-0.006], p = 0.012).

Predictive margins plots from models 2 (1958c) and 3 (1970c) show converging genetic effects between the cohorts by age (Figure 14). The 1970 cohort had a higher marginal genetic effect at younger ages, but both cohorts converged at older age points.

A z-statistic was computed to compare the interaction coefficients from Models 2 and 3. The estimated difference in interaction coefficients between cohorts was 0.001 (SE = 0.006). The z-statistic for this difference was 0.38 (p= 0.703). This indicates no evidence of an overall difference in the effect of polygenic score on psychological distress between the 1958 and 1970 cohorts.

Table 15: Results from the pooled and stratified multilevel mixed effects models interacting mean-centred polygenic score and cohort for 1958c and 1970c (N=10,713)

	Beta Coefficient	Bootstrapped 95% CI	P value
Model 1: Pooled			
1970c is the reference category			
1958c	**-0.33**	**-0.39 - -0.28**	**0.001**
PGS	**0.26**	**0.22-0.29**	**0.001**
PGS*1958c	-0.006	-0.06-0.05	0.831
Model 2+3: Cohort-stratified			
1958 Model (N=6,312)			
PGS	**0.24**	**0.23-0.26**	**0.001**
Age	**0.01**	**0.01-0.02**	**0.001**
PGS*Age	**0.0041**	**0.003-0.005**	**0.001**
1970 Model (N=4,401)			
PGS	**0.25**	**0.23-0.27**	**0.001**
Age	**0.007**	**0.005-0.01**	**0.001**
PGS*Age	**0.0035**	**0.001– 0.006**	**0.012**

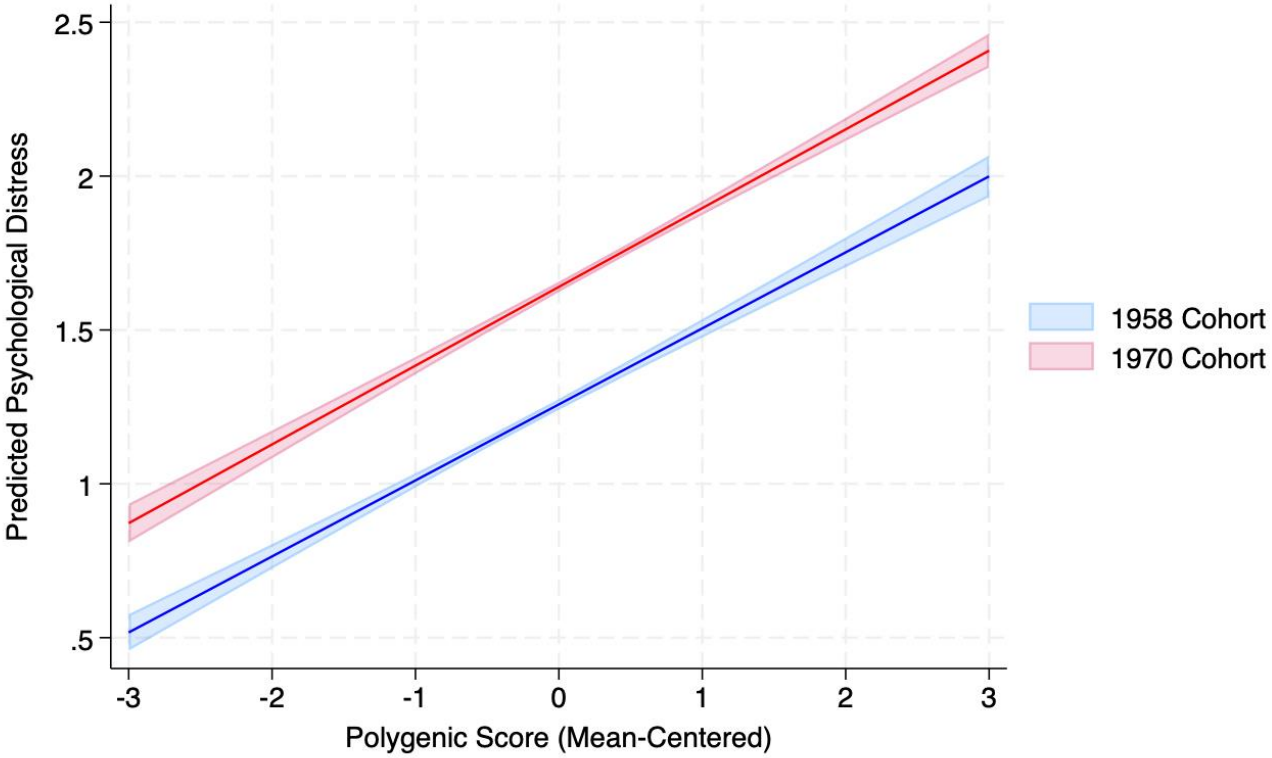


Figure 13: Predictive margins of psychological distress across mean-centred polygenic score for 1958c and 1970c cohorts in model 1

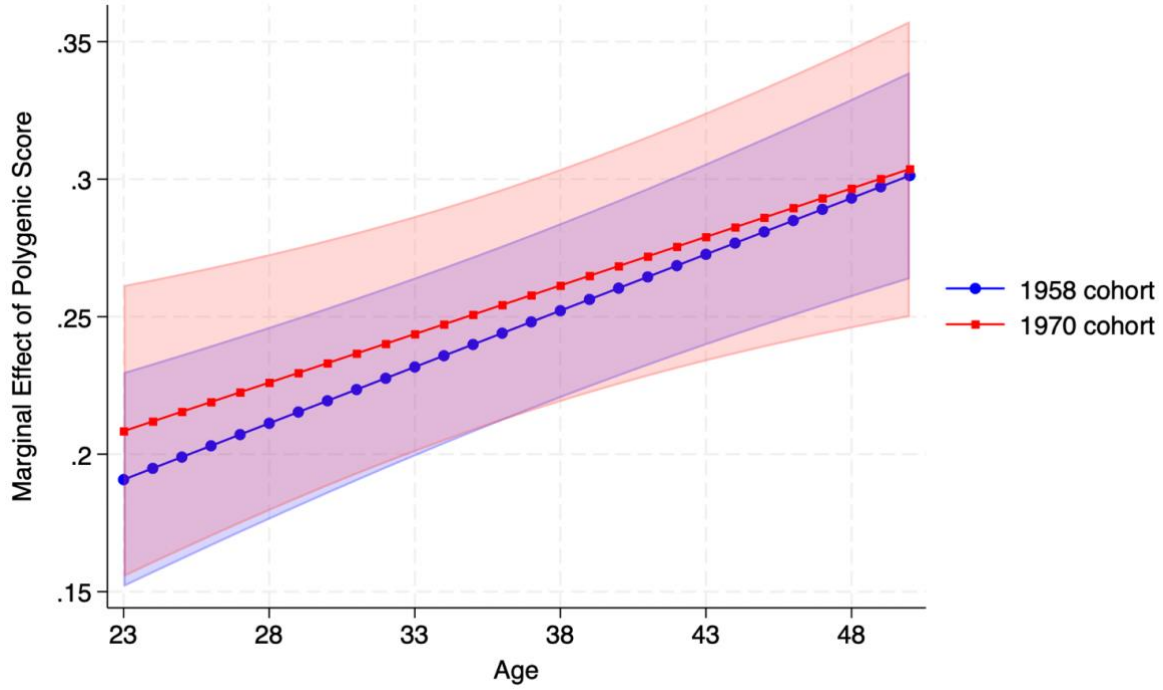


Figure 14: The marginal effect of the polygenic score on psychological distress in the 1958c and 1970 by age in model 2 and 3

5.3.2.3 Research Question 3: Does the association between polygenic score and adult psychological distress outcomes differ by sex?

Pooled Analysis: Model 1

Model 1 multilevel model shows that being female is associated with higher observed psychological distress outcomes compared to males, as indicated by a positive beta coefficient for females (β = 0.60, 95% CI [0.55-0.66]), p = 0.001) (Table 16)).

In model 1 pooled analysis, there was evidence for an interaction between the polygenic score and sex (β = 0.06, 95% CI [0.01-0.11], p = 0.018), suggesting that genetic liability's effect on psychological distress was stronger among females.

Figure 15 shows females have consistently higher predicted level of psychological distress across the polygenic score distribution. The slope for females is slightly steeper, meaning the polygenic score may have a stronger effect.

Cohort-stratified: Model 2 and 3

Model 2 (1958c) and 3 (1970c), tested for sex differences within each cohort. The independent effect results remained largely the same as model 1 (Table 16).

In model 2 there was some evidence of an interaction between the polygenic score and sex in 1958c (β = 0.09, 95% CI [0.01, 0.11], p = 0.005), indicating that the effect of genetic liability for psychological distress was greater among females than males. However, there was no evidence of an interaction in the 1970c in model 3 (β = 0.02, 95% CI [-0.06, 0.11], p = 0.584).

The z-statistic computed to test the difference between the interaction coefficients showed no evidence of a difference (β = 0.07, SE = 0.05, Z = 1.26, p = 0.208). This suggests that while there may be some cohort differences in how genetic liability to psychological distress operates across sexes, the observed differences between cohorts may reflect variability rather than a true difference.

Table 16: Results from the pooled and stratified multilevel mixed effects models interacting mean-centred polygenic score and sex for 1958c and 1970c (N=10,648)

Fixed Effect			
	Beta Coefficient	Bootstrapped 95% CI	P value
Model 1: Pooled			
Male is the reference category			
Female	**0.60**	**0.55-0.66**	**0.001**
PGS	**0.22**	**0.18-0.26**	**0.001**
PGS*Female	**0.06**	**0.01-0.11**	**0.018**
Model 2 + 3: Cohort-stratified			
1958 Model (N=6,312)			
Male (ref)			
Female	**0.66**	**0.60-0.71**	**0.001**
PGS	**0.20**	**0.16-0.25**	**0.001**
PGS*Female	**0.09**	**0.01-0.11**	**0.005**
1970 Model (N=4,401)			
Male (ref)			
Female	**0.53**	**0.45-0.62**	**0.001**
PGS	**0.24**	**0.18-0.30**	**0.001**
PGS*Female	0.02	-0.06 - 0.11	0.584

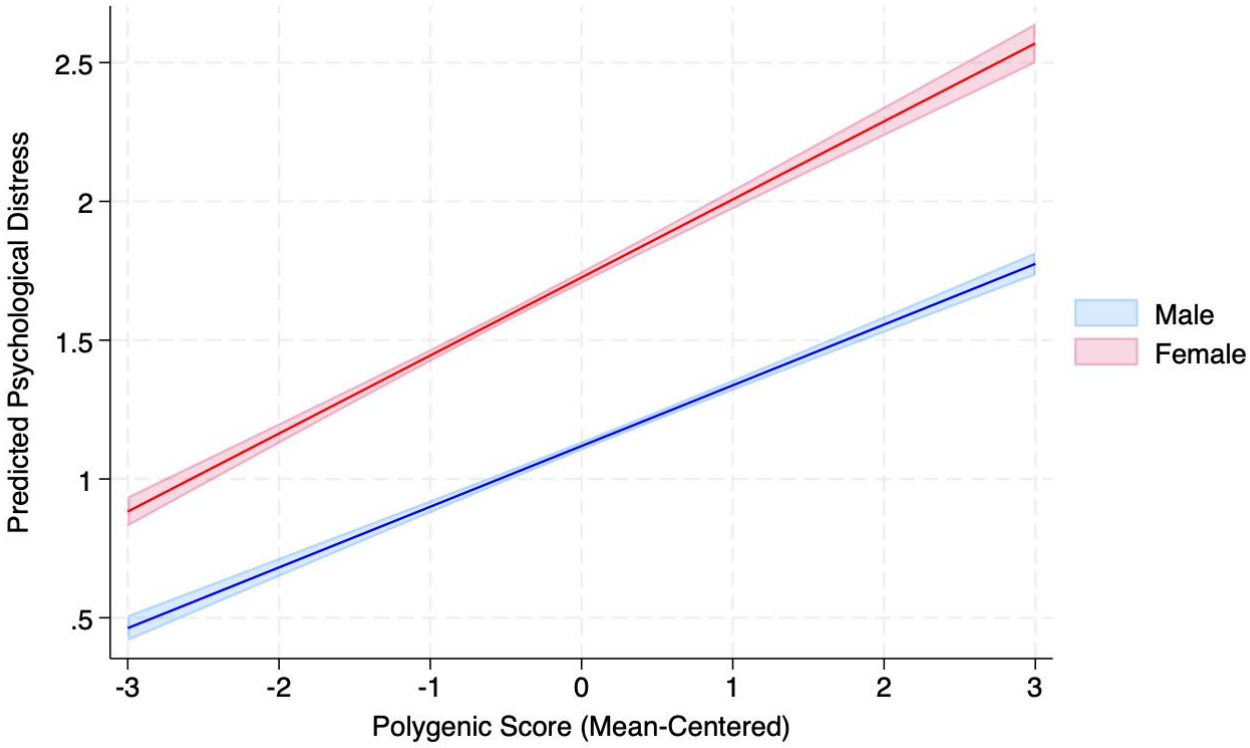


Figure 15: Predictive margins of psychological distress across mean-centred polygenic score for males versus females in model 1

5.3.2.4 Supplemental and Sensitivity Analyses

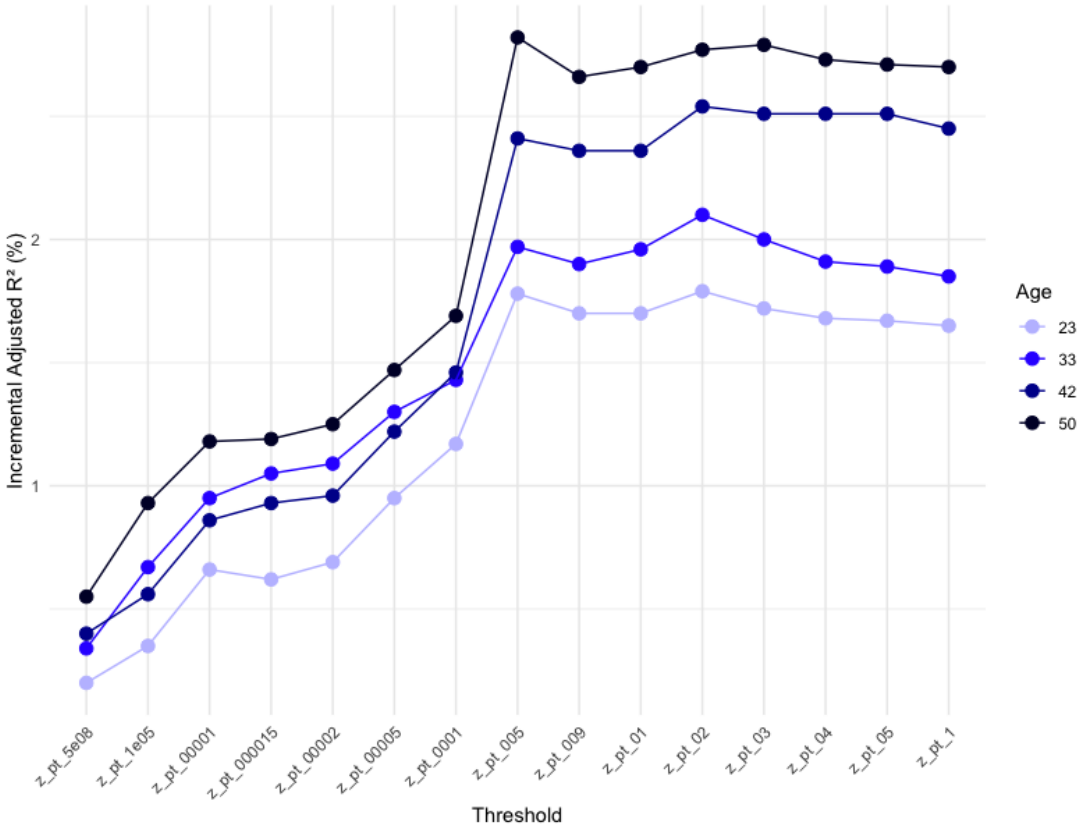
Testing polygenic score for psychological distress at different p-value thresholds:

Specification curve plots were used to assess the impact of the selected p-value thresholds on the estimates (Figures 16 and 17). Both the beta coefficient and R^2 remained similar to the main analyses, with estimates stabilising as the thresholds reached closer to 1 (Supplementary Tables 8 and 9).

At the most stringent genome-wide significance threshold (5.00×10^{-8}), the explanatory power was modest but still detectable, with R^2 values ranging from 0.07% to 0.55% and effect sizes between 0.05 and 0.14 across ages and cohorts. At the threshold of 0.005 (used in the main analyses), explanatory power substantially increased, with R^2 values ranging from 1.44% to 2.86% and effect sizes between 0.20 and 0.31, representing an improvement over the genome-wide threshold. The most inclusive threshold of 1 (including all SNPs) showed similar results to the 0.005 threshold, with R^2 values of 1.39% to 2.70% and identical effect sizes in many cases, confirming that the 0.005 threshold captures most of the polygenic signal without introducing excessive noise.

Including this analysis strengthens confidence in utilising the PGS at the 0.005 threshold, and that the findings remain relatively consistent across threshold specifications.

a)



b)

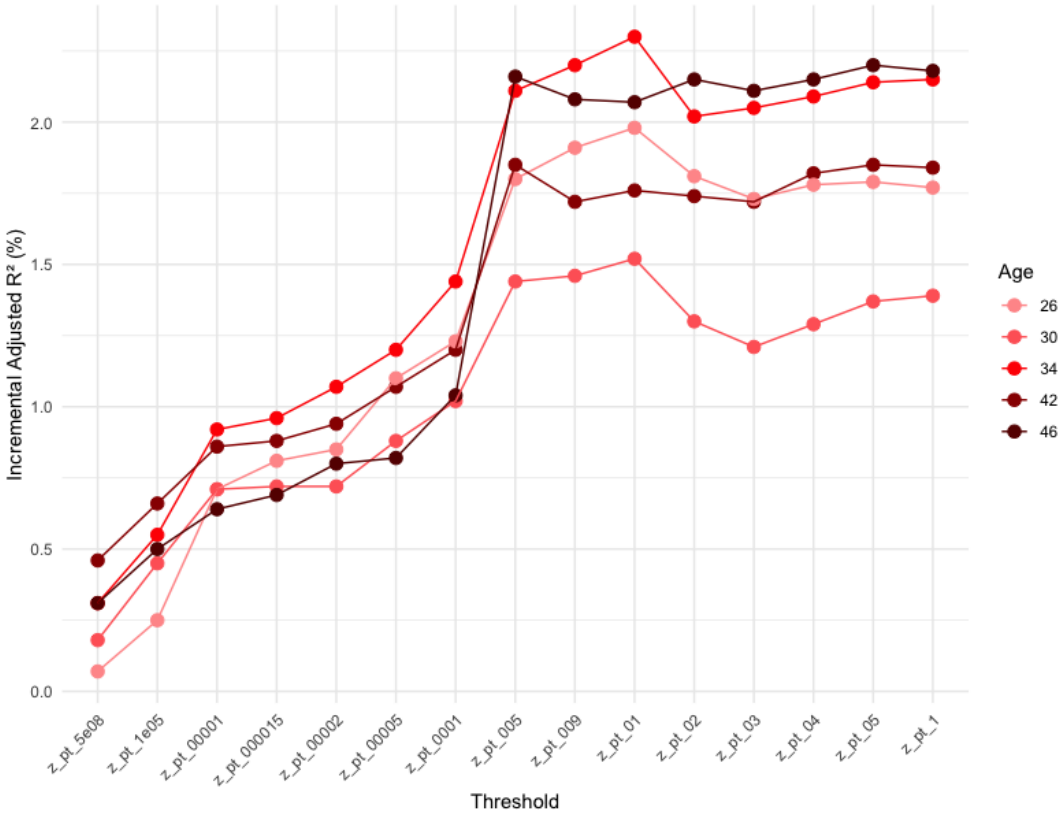
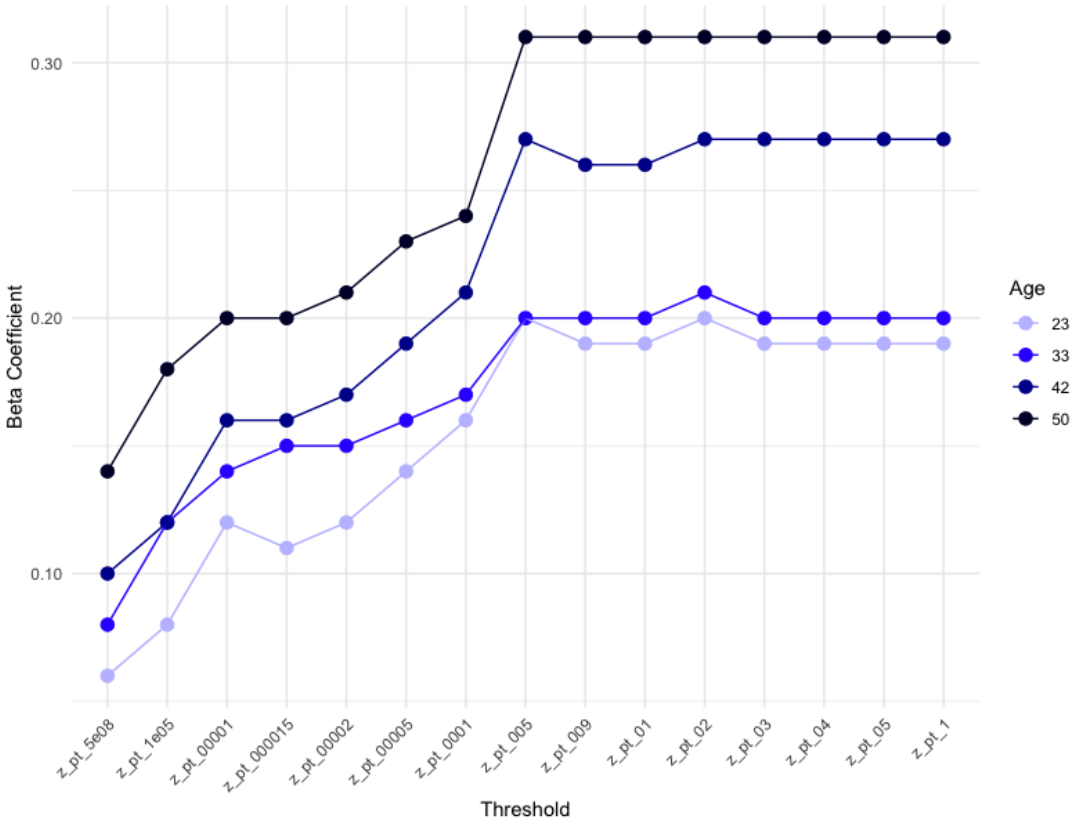


Figure 16: Specification curve plots of the variance explained by the polygenic score for psychological distress in the regression model at each age at each potential p-value threshold in **a)** 1958c and **b)** 1970c

a)



b)

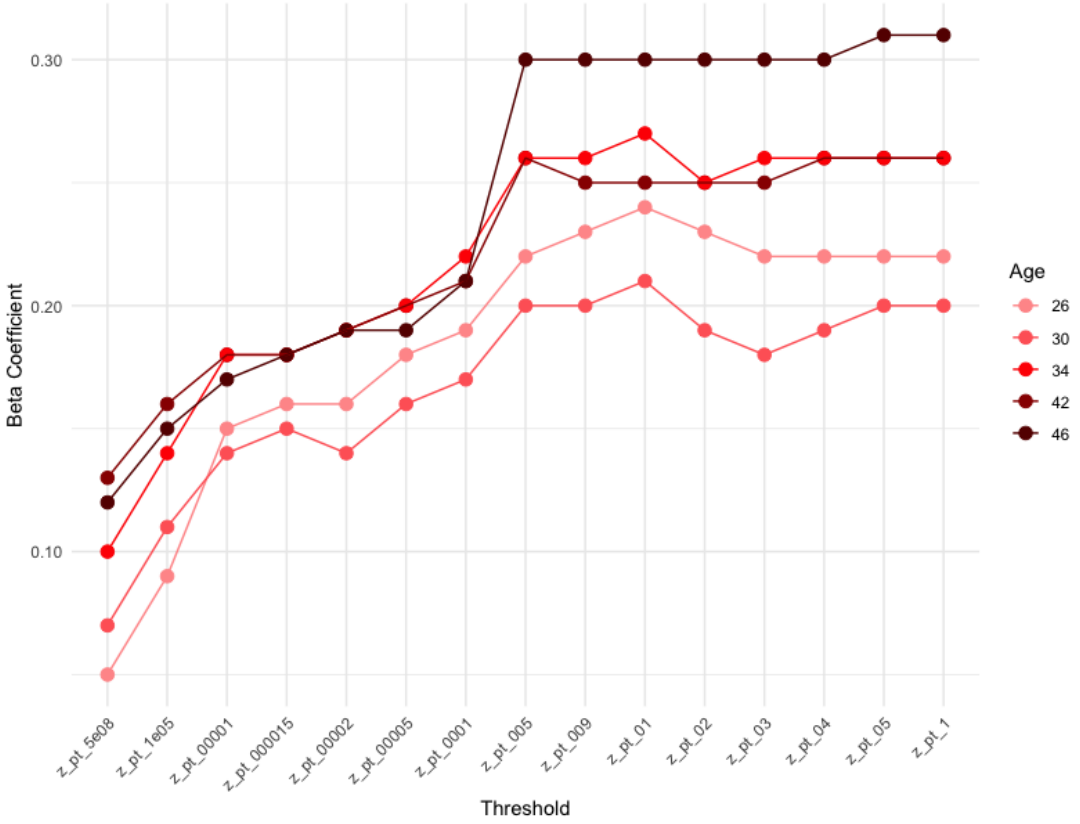


Figure 17: Specification curve plots of the beta coefficient from linear regressions between polygenic score and observed psychological distress, by each age at each potential p-value threshold in **a)** 1958c and **b)** 1970c

Testing the 24-item Malaise Inventory Score:

Beta Coefficient
Results from the linear regression analyses demonstrate positive and consistent associations between polygenic scores for psychological distress and 24-item Malaise Inventory scores from ages 23 to 42 in both the 1958c and 1970c cohorts (Table 17).

R-squared
The R-squared (R^2) values, which represent the amount of variance in the Malaise Inventory explained by the polygenic score for psychological distress, also increased with age in both cohorts (Table 17). This indicates that the polygenic score explains more of the variance in psychological distress across adulthood for both cohorts, similarly to the 9-item results (Table 17).

Due to the lack of standardisation of the outcome variable, as polygenic penetrance is the absolute change in Malaise score per standard deviation increase in polygenic, the beta coefficients cannot be directly compared between the 9- and 24-item Malaise Inventory Score. Therefore, R-squared is the focus of comparison. Overall, the 24-item Malaise Inventory has a similar range and overlapping confidence intervals for the R-squared values compared to the 9-item. Specifically, R-squared ranged from 1.20%-2.50% for the 24-item version versus 1.44%-2.86% for the 9-item version. Both versions showed consistent positive associations across all ages and cohorts, with similar age-related increases in variance explained. This consistency provides strong evidence that the 9-item abbreviated scale captures the polygenic signal as effectively as the full 24-item version, while also extending the age range examined.

Table 17: Linear regression associations between polygenic score for psychological distress with overlapping SNPs and observed psychological distress using the 24-item Malaise Inventory Score at ages 23, 26, 30, 33 and 42 in 1958c and 1970c (PGS Threshold=0.005; Imputed 1958c N = 5,939; Imputed 1970c N= 4,045)

Year	Age	Cohort	N	Mean (SD)	R^2	Beta Coefficient	Beta 95%CI
1981	23	1958c	5,939	2.73 (2.94)	1.20%	0.37	(0.29-0.43)
1996	26	1970c	4,045	3.85 (3.38)	2.50%	0.47	(0.36-0.56)
2000	30	1970c	4,045	3.52 (3.48)	2.40%	0.43	(0.33-0.52)
1991	33	1958c	5,939	2.43 (2.98)	1.30%	0.39	(0.32-0.47)
2000	42	1958c	5,939	3.60 (3.63)	2.46%	0.58	(0.49-0.66)

R^2: amount of variance explained by the PGS. Analyses were adjusted for sex and the first 10 principal components of ancestry.
NB: blue indicates the 1958c, and red indicates the 1970c

5.4 Discussion

5.4.1 Summary of Main Findings

This study examined whether the polygenic score for psychological distress is associated with observed psychological distress by age and cohort, using the National Child and Development Study (1958c) from the previous chapter and the British Cohort Study (1970c). Findings show that the genotype-phenotype association for psychological distress remains persistent from early (~ age 23) to middle (~ age 50) adulthood in both cohorts. Each cohort shows an increasing association with age and greater explanatory power of genotypic effects. There was some evidence of a cohort difference at younger age points. However, the cohorts converged by late adulthood. Females showed higher levels of psychological distress than males, and the 1970 cohort exhibited higher distress levels compared to the 1958 cohort. There was some evidence of an interaction between genetic factors and sex in the pooled analyses. Once stratified, a potential gene-sex interaction was found in the 1958c, however the z-statistic computed to test the difference between the interaction coefficients showed no evidence of a difference

5.4.2 Comparison to the literature and explanation of findings

The persistence of the genotype-phenotype association across adulthood suggests a tentative trend of increasing strength in the explanatory power of both cohorts' genotypes at later ages. For the 1970c, less of the variance in distress is explained by polygenic scores, meaning differences may be more attributable to environmental factors than in the 1958c.

This is potentially due to changing environmental factors, i.e., a depressogenic environment emerges which the 1970c experienced more of during their middle age. The shift from an industrial to a service-based economy in the late 20th century, coupled with increased educational attainment and women's labour market participation, may have contributed to a more distressing environment, particularly for the 1970 cohort (301) (139). As absolute living standards increased, the relative quality of life decreased, and inequalities increased for both cohorts, but more so within the 1970 cohort (296). These macroeconomic and social changes in how society is structured could have had a knock-on effect on the prevalence of psychological distress in the UK population.

The positive association between genotypic and phenotypic distress at all ages is consistent with previous research (75,109,254,262,307). These suppose that the contribution of genotypes to phenotypes as people age is relatively consistent after development across childhood and adolescence (311). Finding that this association increases with age in both cohorts, may be due to people accruing life experiences and developing behavioural traits where their genotypes interact with and are expressed through the environments they actively select for themselves (264).

The estimates increasing by age but not by cohort was inconsistent with one previous study by Athanasiadis et al. 2022 (306). This research used the Danish Registry Data

and found higher heritability estimates for mental conditions, such as depression and anxiety, in younger cohorts. This could be due to the environment in Denmark being more homogenous, so over time, the variation in the phenotype is more attributable to the genotype (41,312). When utilising the social context theory of gene-environment interaction, it has been observed that in environments characterised by high levels of social control (i.e. an environ that constrains behaviour), the heritability tends to either decrease or remain stable, whereas in contexts with low levels of social control, it tends to increase (99). The higher psychological distress scores in the 1970 cohort suggest that environmental and socio-economic factors play a prominent role. The lack of evidence for an interaction between polygenic score and cohort indicates that genetic effects on psychological distress are stable across cohorts, but environmental differences likely contribute to the observed disparities. It may be that 1970c, as a cohort, experienced an environment that had changing social control, i.e. the social norms and structural underpinnings of society altered, which affected the expression of the distress phenotype.

It is possible that the higher heritability estimates in older age, rather than cohort, are attributable to several factors. The risk for mental health conditions may increase towards the end of the lifespan, or the accumulation of environmental exposures throughout life in the older cohorts increases the heritability estimates (183,262,277,313). At the same time, in younger ages, the accumulation of diagnoses is an ongoing process that is interrupted by right censoring (314). In prior research, this has shown a cohort difference where successive generational cohorts are characterised by progressively earlier onset ages, typically associated with more robust genetic signals and higher heritability estimates (241,315).

The larger estimates between ages 42-50 in the 1958c and 1970c could feasibly be due to the applied GWAS being derived from a predominantly middle-aged sample. The GWAS summary statistics did not report the mean age, but the cohorts included mean participants' ages ranging between 18 and 85, including the Understanding Society Cohort. As the sample size was ~1,000,000 for the GWAS, the majority of the sample came from the U.K Biobank; where the average age is 56 years old (SD: 8.09) (316). This might mean the SNPs identified in that genome-wide association study were specific to mid-life expression of psychological distress, as different genes can be expressed at different ages and their environment interaction, and behavioural influences may also vary with time.

In the gene-sex pooled analysis, evidence for an interaction was identified. Furthermore in the specific cohort-stratified analyses, there was there was evidence for an interaction in the 1958c.This was not consistent with prior literature, which did not consistently identify sex differences (317). However, the z-statistic computed to test the difference between the interaction coefficients showed no evidence of a difference. Prior literature supports the findings that there was no evidence for an interaction in the 1970c, from the cohort-stratified analyses. Therefore, the observed differences in statistical significance between models may be more likely due to sample variability rather than a true sex differences between cohorts**.** These findings might be due to environmental or societal factors specific to women that have amplified the gene-sex interaction. This is further explored in limitations below.

5.4.3 Strengths & Limitations

Strengths of this study included the use of psychological distress data collected at multiple ages in each cohort – utilising the same questionnaire for measurement at each time point. Data collection spans from 1981 to 2016 across two generationally different cohorts (Baby Boomers versus Generation X), enabling cross-cohort comparison. Similarly, how the polygenic scores were constructed by being standardised across the cohorts, restricted to overlapping SNPs, and neither cohort being present in the discovery sample of the GWAS used also enabled robust cross-cohort comparison.

Another strength of the current study is the construction of the environment variable. Previous GxE studies suffered from potential variations in gene-environment interactions that might differ between cohorts due to unique historical and environmental contexts. For instance, environmental factors such as socio-economic conditions, lifestyle changes, and exposure to different stressors can vary between cohorts, potentially influencing the expression of genetic liability differently. This study's reliance on birth year as an exogenous source of environmental variation helps mitigate some of these concerns by enabling the estimation of gene-cohort interaction. This method circumvents the problem of environmental endogeneity, where the environment could be influenced by genetic factors

However, this study was not without limitations. First, comparing only two cohorts may introduce bias, as errors in either cohort can distort observed trends. Including three or more cohorts, as suggested by prior research, would provide a more accurate identification of trends (318). However the thesis was restricted by lack of existing data with comparable measures, as the 1946 cohort changes mental health questionnaire across adulthood and there is no data of the 2000 cohort across adulthood yet (124).

Second, random measurement errors in the exposure variable (polygenic scores) can dilute effect sizes, while errors in the outcome variable (psychological distress) can widen standard errors and confidence intervals, reducing estimate precision. Non-random errors can bias results in either direction (318). However, prior work has found that there is measurement invariance in the outcome of distress between these cohorts (204).

Third, in both cohorts, there are a higher number of women remaining in the study compared to men by age, meaning the results are potentially confounded by gender-based differences in study retention rates. Therefore, the analyses controlled for this using sex difference analysis. In that analysis, differential associations were observed between both cohorts' genotypic and phenotypic psychological distress across the sexes. Females consistently exhibited higher distress levels than males. This finding may reflect broader societal influences exacerbating female distress despite similar genetic liability.

Fourth, it is important to note that the polygenic scores are assumed to capture genetic liability factors for psychological distress that are stable across time. This assumption is based on the premise that SNPs remain relatively constant and do not vary between cohorts. However, this GWAS approach may not isolate cohort-specific genetic signals due to the inclusion of multiple generations and cohorts (319).Yet the current study

utilised the polygenic score as a valid tool for examining genetic liability to psychological distress by cohort. As it still effectively captures the overarching polygenic scores for psychological distress that span different time periods.

A limitation of the use of an exogenous environmental variable is, while this approach provides evidence indicating wider environmental contribution, it is limited in being able to pinpoint which specific environmental characteristics interact with polygenic liability to psychological distress. The current study does not test for differential susceptibility or other specific environmental features that may vary across cohorts. Future research should identify and incorporate environmental variables to provide a better understanding of gene-environment interactions and to determine how different environments may modulate polygenic liability for psychological distress.

5.4.4 Conclusion

This study provides evidence that the association between polygenic scores for psychological distress and observed psychological distress remains persistent from early (~ age 23) to middle (~ age 50) adulthood in both cohorts. Each cohort shows an increasing association and greater explanatory power of genotypic effects by age. There was some evidence of a cohort difference at younger age points. However, the cohorts converged by late adulthood. There was also some evidence of a sex difference when the cohorts were pooled. Relative to other factors, the genetic contribution explained only a small proportion of the variance in psychological distress, suggesting that other factors beyond genetic liability likely contribute to psychological distress in these individuals.

Future research should investigate specific environmental stressors that might interact with genetic liability. The thesis introduction (Section 1.5) outlined the importance of broader environmental exposures in relation to distress outcomes over time, one prominent component being socioeconomic position. Longitudinal studies could further elucidate how these factors evolve and interact over time. The next chapter will explore the interaction between father's social class and polygenic score for psychological distress in the 1958c and 1970c cohorts.

5.5 Study 2 Summary

- Both the 1958c and 1970c show an increasing association and greater explanatory power of genotypic effects between ages 23-50, suggesting that genetic factors may become more pronounced with age.
- There was some evidence of a cohort difference at younger ages. However, the cohorts converged by older adulthood.
- Pooled analyses found the polygenic score association with psychological distress was stronger in females compared to males.
- The genetic contribution explained only a small proportion of the variance in psychological distress.
- These findings highlight the importance of considering both genetic and environmental factors in understanding psychological distress.
- The results motivate the next study of the thesis, which will explore the interaction between childhood social class and genetic liability to psychological distress in the 1958c and 1970c cohorts.

5.6 Supplementary Materials

Supplementary Table 8: Linear regression associations across all thresholds of the polygenic score for psychological distress with overlapping SNPs between 1958c and 1970c and observed psychological distress across ages 23-50 in the 1958c (N=6,312)

Age	Threshold	R^2	Beta Coefficient	Lower 95% CI	Upper 95% CI	*P*
23	$5.00x10^{-08}$	0.20%	0.06	0.03	0.09	0.338
	$1.00x10^{-08}$	0.35%	0.08	0.06	0.11	0.008
	0.00001	0.66%	0.12	0.09	0.14	<0.001
	0.000015	0.62%	0.11	0.09	0.14	<0.001
	0.00002	0.69%	0.12	0.09	0.15	<0.001
	0.00005	0.95%	0.14	0.12	0.17	<0.001
	0.0001	1.17%	0.16	0.13	0.19	<0.001
	0.005	1.83%	0.20	0.17	0.23	<0.001
	0.009	1.70%	0.19	0.17	0.22	<0.001
	0.01	1.70%	0.19	0.17	0.22	<0.001
	0.02	1.79%	0.20	0.17	0.22	<0.001
	0.03	1.72%	0.19	0.17	0.22	<0.001
	0.04	1.68%	0.19	0.16	0.22	<0.001
	0.05	1.67%	0.19	0.16	0.22	<0.001
	1	1.65%	0.19	0.16	0.22	<0.001
33	$5.00x10^{-08}$	0.34%	0.08	0.05	0.11	0.037
	$1.00x10^{-08}$	0.67%	0.12	0.09	0.14	0.003
	0.00001	0.95%	0.14	0.11	0.17	<0.001
	0.000015	1.05%	0.15	0.12	0.17	<0.001
	0.00002	1.09%	0.15	0.12	0.18	<0.001
	0.00005	1.30%	0.16	0.14	0.19	<0.001
	0.0001	1.43%	0.17	0.15	0.20	<0.001
	0.005	2.02%	0.20	0.18	0.23	<0.001
	0.009	1.90%	0.20	0.17	0.23	<0.001
	0.01	1.96%	0.20	0.18	0.23	<0.001
	0.02	2.10%	0.21	0.18	0.24	<0.001
	0.03	2.00%	0.20	0.18	0.23	<0.001
	0.04	1.91%	0.20	0.17	0.23	<0.001
	0.05	1.89%	0.20	0.17	0.22	<0.001
	1	1.85%	0.20	0.17	0.22	<0.001
42	$5.00x10^{-08}$	0.40%	0.10	0.07	0.14	0.002
	$1.00x10^{-08}$	0.56%	0.12	0.09	0.15	<0.001
	0.00001	0.86%	0.16	0.13	0.19	<0.001
	0.000015	0.93%	0.16	0.13	0.19	<0.001
	0.00002	0.96%	0.17	0.14	0.20	<0.001
	0.00005	1.22%	0.19	0.16	0.22	<0.001
	0.0001	1.46%	0.21	0.18	0.24	<0.001
	0.005	2.46%	0.27	0.24	0.30	<0.001
	0.009	2.36%	0.26	0.23	0.29	<0.001
	0.01	2.36%	0.26	0.23	0.29	<0.001
	0.02	2.54%	0.27	0.24	0.30	<0.001

	0.03	2.51%	0.27	0.24	0.30	<0.001
	0.04	2.51%	0.27	0.24	0.30	<0.001
	0.05	2.51%	0.27	0.24	0.30	<0.001
	1	2.45%	0.27	0.24	0.30	<0.001
50	$5.00x10^{-08}$	0.55%	0.14	0.10	0.17	0.001
	$1.00x10^{-08}$	0.93%	0.18	0.14	0.21	<0.001
	0.00001	1.18%	0.20	0.17	0.24	<0.001
	0.000015	1.19%	0.20	0.17	0.24	<0.001
	0.00002	1.25%	0.21	0.17	0.24	<0.001
	0.00005	1.47%	0.23	0.19	0.26	<0.001
	0.0001	1.69%	0.24	0.21	0.28	<0.001
	0.005	2.86%	0.31	0.28	0.35	<0.001
	0.009	2.66%	0.31	0.27	0.34	<0.001
	0.01	2.70%	0.31	0.27	0.34	<0.001
	0.02	2.77%	0.31	0.28	0.35	<0.001
	0.03	2.79%	0.31	0.28	0.35	<0.001
	0.04	2.73%	0.31	0.27	0.34	<0.001
	0.05	2.71%	0.31	0.27	0.34	<0.001
	1	2.70%	0.31	0.27	0.34	<0.001

Threshold: *P*-value threshold for the PGS (i.e., threshold for the number of significant SNPs from the original GWAS included into the PGS). R^2: amount of variance explained by the PGS. PD: psychological distress. Analyses were adjusted for sex and the first 10 principal components of ancestry.

Supplementary Table 9: Linear regressions associations across all thresholds of the polygenic score for psychological distress with overlapping SNPs between 1958c and 1970c and observed psychological distress across ages 26-46 in the 1970c (N=5,423)

Age	Threshold	R^2	Beta Coefficient	Lower 95% CI	Upper 95% CI	*P*
26	$5.00x10^{-08}$	0.07%	0.05	0.00	0.11	0.176
	$1.00x10^{-08}$	0.25%	0.09	0.03	0.14	0.014
	0.00001	0.71%	0.15	0.09	0.20	0.001
	0.000015	0.81%	0.16	0.10	0.21	0.001
	0.00002	0.85%	0.16	0.10	0.21	<0.001
	0.00005	1.10%	0.18	0.12	0.24	<0.001
	0.0001	1.23%	0.19	0.13	0.25	<0.001
	0.005	1.80%	0.22	0.17	0.28	<0.001
	0.009	1.91%	0.23	0.18	0.29	<0.001
	0.01	1.98%	0.24	0.18	0.29	<0.001
	0.02	1.81%	0.23	0.17	0.28	<0.001
	0.03	1.73%	0.22	0.17	0.28	<0.001
	0.04	1.78%	0.22	0.17	0.28	<0.001
	0.05	1.79%	0.22	0.17	0.28	<0.001
	1	1.77%	0.22	0.17	0.28	<0.001
30	$5.00x10^{-08}$	0.18%	0.07	0.02	0.12	0.002
	$1.00x10^{-08}$	0.45%	0.11	0.06	0.16	<0.001
	0.00001	0.71%	0.14	0.09	0.20	<0.001
	0.000015	0.72%	0.15	0.09	0.20	<0.001
	0.00002	0.72%	0.14	0.09	0.20	<0.001
	0.00005	0.88%	0.16	0.11	0.21	<0.001
	0.0001	1.02%	0.17	0.12	0.22	<0.001
	0.005	1.44%	0.20	0.15	0.25	<0.001
	0.009	1.46%	0.20	0.15	0.25	<0.001
	0.01	1.52%	0.21	0.15	0.26	<0.001
	0.02	1.30%	0.19	0.14	0.24	<0.001
	0.03	1.21%	0.18	0.13	0.23	<0.001
	0.04	1.29%	0.19	0.14	0.24	<0.001
	0.05	1.37%	0.20	0.14	0.25	<0.001
	1	1.39%	0.20	0.15	0.25	<0.001
34	$5.00x10^{-08}$	0.31%	0.10	0.05	0.16	0.030
	$1.00x10^{-08}$	0.55%	0.14	0.08	0.19	0.001
	0.00001	0.92%	0.18	0.12	0.23	<0.001
	0.000015	0.96%	0.18	0.12	0.24	<0.001
	0.00002	1.07%	0.19	0.13	0.24	<0.001
	0.00005	1.20%	0.20	0.14	0.26	<0.001
	0.0001	1.44%	0.22	0.16	0.27	<0.001
	0.005	2.10%	0.26	0.20	0.31	<0.001
	0.009	2.20%	0.26	0.21	0.32	<0.001
	0.01	2.30%	0.27	0.22	0.33	<0.001
	0.02	2.02%	0.25	0.20	0.31	<0.001
	0.03	2.05%	0.26	0.20	0.31	<0.001
	0.04	2.09%	0.26	0.20	0.31	<0.001

	0.05	2.14%	0.26	0.21	0.32	<0.001
	1	2.15%	0.26	0.21	0.32	<0.001
42	$5.00x10^{-08}$	0.46%	0.13	0.07	0.19	<0.001
	$1.00x10^{-08}$	0.66%	0.16	0.10	0.22	<0.001
	0.00001	0.86%	0.18	0.12	0.24	<0.001
	0.000015	0.88%	0.18	0.12	0.24	<0.001
	0.00002	0.94%	0.19	0.13	0.25	<0.001
	0.00005	1.07%	0.20	0.14	0.26	<0.001
	0.0001	1.20%	0.21	0.15	0.27	<0.001
	0.005	1.85%	0.26	0.20	0.31	<0.001
	0.009	1.72%	0.25	0.19	0.31	<0.001
	0.01	1.76%	0.25	0.19	0.31	<0.001
	0.02	1.74%	0.25	0.19	0.31	<0.001
	0.03	1.72%	0.25	0.19	0.31	<0.001
	0.04	1.82%	0.26	0.20	0.32	<0.001
	0.05	1.85%	0.26	0.20	1.85	<0.001
	1	1.84%	0.26	0.20	0.32	<0.001
46	$5.00x10^{-08}$	0.31%	0.12	0.06	0.18	<0.001
	$1.00x10^{-08}$	0.50%	0.15	0.09	0.21	<0.001
	0.00001	0.64%	0.17	0.11	0.23	<0.001
	0.000015	0.69%	0.18	0.11	0.24	<0.001
	0.00002	0.80%	0.19	0.13	0.25	<0.001
	0.00005	0.82%	0.19	0.13	0.25	<0.001
	0.0001	1.04%	0.21	0.15	0.28	<0.001
	0.005	2.16%	0.30	0.24	0.36	<0.001
	0.009	2.08%	0.30	0.24	0.36	<0.001
	0.01	2.07%	0.30	0.24	0.36	<0.001
	0.02	2.15%	0.30	0.24	0.36	<0.001
	0.03	2.11%	0.30	0.24	0.36	<0.001
	0.04	2.15%	0.30	0.24	0.36	<0.001
	0.05	2.20%	0.31	0.24	0.37	<0.001
	1	2.18%	0.31	0.24	0.37	<0.001

Threshold: *P*-value threshold for the PGS (i.e., threshold for the number of significant SNPs from the original GWAS included into the PGS). R^2: amount of variance explained by the PGS. PD: psychological distress. Analyses were adjusted for sex and the first 10 principal components of ancestry.

6 Study 3: Investigating gene-environment interplay between polygenic score and father's social class on adult psychological distress during adulthood

Main objective: to examine whether father's social class interacts with polygenic scores in its association with adulthood psychological distress in the 1958c and 1970c.

Research Question 1: Does the association between polygenic scores for distress and adult psychological distress outcomes vary according to father's social class?

Research Question 2: Do the independent associations or interactions between polygenic score for distress and father's social class on adulthood psychological distress differ by cohort or by sex?

Research Question 3: Using meta-analysis, is there robust evidence of an interaction between polygenic score for psychological distress and socioeconomic circumstances on adulthood psychological distress outcomes?

6.1 Introduction

The present study's main objective is to examine the interplay between polygenic scores for psychological distress and father's social class. This investigation builds upon previous chapters which explored the relationship between genotypic and phenotypic psychological distress by age and cohort. These previous chapters have discussed relevant literature summarising the polygenic basis for psychological distress. Studies 1 and 2 showed an increasing association between polygenic scores and psychological distress by age. However, polygenic scores only accounted for a small percentage of the variance in distress scores, suggesting that other potential environmental correlates may better explain distress outcomes. This chapter extends this by first examining whether father's social class interacts independently or synergistically with genetic liability to impact distress outcomes across adulthood and then addressing the question by cohort and by sex. Finally, evidence is synthesised in a meta-analysis.

This introduction is structured by first focusing on the socioeconomic determinants of psychological distress, why socioeconomic circumstances may modify genetic effects, and prior evidence from gene-environment studies.

6.1.1 Socioeconomic determinants of psychological distress

Environmental risk factors can be physical (e.g., climate), chemical (e.g., pollution), biological (e.g., pathogens), behavioural factors (e.g., lifestyle choices like smoking or exercise) and psychosocial (e.g., social networks, socioeconomic position) (320). Socioeconomic position refers to aspects of an individual's social and economic position in society, which impacts their access to resources and social networks alongside their maintenance of them (168). Within a population, individuals accrue or lack cultural, social, and economic capital that amalgamate to inform their advantaged or disadvantaged socioeconomic circumstances (170). Socioeconomic determinants of psychological distress include, but are not limited to, lower income levels, unemployment, poor housing conditions, food insecurity, experiencing discrimination, adverse childhood experience and lower educational attainment (277,321).

There is a well-documented social gradient in mental health outcomes whereby people who are in disadvantaged socioeconomic circumstances tend to have worse mental health outcomes (277) (5,7,135,322). Health selection and social causation are the primary frameworks to explain these inequalities (173,323). Social causation posits that socioeconomic circumstances influence psychological distress, while health selection suggests the reverse—distress impacts socioeconomic circumstances (173,323). For example, a review in 2015 summarised across 34 studies that there was no hegemony of either hypothesis across the literature (173). However, they found research that utilised indicators such as education and income favoured the social causation hypothesis.

Research has largely focused on documenting the presence of inequalities, with less attention paid to when and how they develop. However, as per lifecourse epidemiology, timing is critical to how the mechanisms driving the social gradient in psychological distress may vary across life stages (3). For instance, during the transition to adulthood—a period marked by shifts from parental socioeconomic

circumstance to forming one's own socioeconomic circumstance —health selection effects may be particularly salient (324). Negative mental health episodes can disrupt pivotal events like educational milestones and early-career decisions, thereby influencing socioeconomic circumstance trajectories (238,321). Later in working-age life, poor mental health may still affect socioeconomic circumstance, however individuals often have greater resources to buffer these impacts. Financial savings, support from long-term partners, or job tenure can provide stability in the face of mental health challenges.

For instance, Chandola (2003) who used the Whitehall II study refuted evidence for health selection in mid-life contexts, highlighting the importance of a nuanced lifecourse approach to understanding social gradient of psychological distress (175). Social causation relies on socioeconomic circumstances influencing psychological distress, this is more in-keeping with how lifecourse epidemiology conceptualises the accumulation of risk factors across adulthood compound to influence health outcomes (183,323). The current chapter, due to its focus on adulthood, posits the social causation framework as the underpinning theory for why socioeconomic status influences psychological distress.

As outlined, there are differing aspects to what constitutes our socioeconomic circumstances. These include but are not limited to income, education, and occupation. A recent systematic meta-review found that the social gradient is consistent across all domains but may be steeper for lower social positions via deprivation (e.g., poverty), socioeconomic status, income, and subjective social status (324). However, there was less robust evidence synthesis for an individual's educational attainment, occupation, wealth, and social capital. They pointed out that most reviews employed individual-level measures instead of interpersonal or community-level measures. The evidence summarised thus far has all pertained to the socioeconomic circumstances of an individual rather than parental circumstances.

Prior research shows the social gradient begins in childhood, where children and adolescents from disadvantaged backgrounds are more likely to develop mental health issues than their peers from advantaged backgrounds (183). This means the socioeconomic circumstances that a parent may provide can impact the subsequent health of their child, reflecting a more social causation framework for health inequality development (182). Disadvantaged circumstances can then compound throughout the lifespan (6,183). Hence, the present study focused on childhood socioeconomic circumstances as the environmental risk factor due to its correlation with psychological distress and its potential role in modulating genetic vulnerabilities (91,325,326). Specifically, father's social class was used for various reasons, further detailed below.

6.1.2 Father's social class as a measure of socioeconomic circumstance

Childhood social class, as measured by the father's occupation, served as the primary proxy for childhood socioeconomic position in this Chapter (214). This measure effectively captures the broader construct of socioeconomic circumstance, which is correlated and interconnected with other domains such as income and education (216,327,328). It follows that the educational attainment one achieves relates to the occupational status gained and income renumerated. Although education and income are also important, social class is correlated with and can serve as a proxy for these other dimensions of socioeconomic position. It also benefits from the temporal

ordering that it removes the ability for reverse causality to be introduced to a research design.

As childhood social class sets the foundation for educational and economic opportunities, health exposures, and life chances, over time, the quality and relevance of socioeconomic circumstances proxies have evolved (329). Traditional measures such as manual vs non-manual labour have become less reflective of contemporary 21st century economies dominated by service and knowledge-based jobs (145). Despite these changes, for cohorts born in the middle of the 20th century father's social class remains a robust indicator of childhood socioeconomic circumstance as it captures the structural conditions and constraints influencing an individual's early life (329,330).

One challenge in understanding the relationship between socioeconomic position and psychological distress is measuring socioeconomic position consistently across generations, as the factors that constitute position and shape the social gradient of health can change over time (331). For example, while homeownership might be a significant marker of economic stability for one generation, income may be more influential for another, particularly in economic shifts such as inflation or changes in the labour market (135). The transition from a predominantly industrial economy to a knowledge-based economy in the 1980s further complicates the comparison of occupational types (i.e. manual versus non-manual) across this period (139). As does the widening access to higher education: the percentage of people with full-time education at age 16 increased from 11.1% in 1960 to 20.2% by 1970 (146). This then impacts who had access to non-manual jobs within the knowledge economy. It also makes it harder to ascertain a social gradient as educational attainment is less variable in the population. This makes it easier for occupational classes in the pre-1980s to be comparable for use in the current study. There is evidence that the class structure based on occupational classes did not change considerably between these two cohorts, enabling comparison (329,332,333).

As outlined, the health selection and social causation framework posit opposing directions of the same pathway between socioeconomic circumstances and psychological distress. However, the relationship is complex and potentially bidirectional (186). Adults who develop psychological distress may experience a decline in their social class due to dropping out of the workforce or becoming economically inactive, which could in turn limit their educational attainment or income (289,334). This can create a problem of reverse causality. Disadvantaged socioeconomic circumstances also impact cohort study attrition rates, as people with higher stress tend to drop out of the studies (335). The loss of this sample makes it harder to estimate the relationship accurately. Consequently, measuring parent social class at an early age helps to mitigate some of these issues. Therefore, using father's social class was used in the present study as a main indicator of early life conditions.

Occupational class may provide a more comprehensive measure of socioeconomic position by incorporating occupational prestige, which captures aspects of social standing that income or education alone may not fully reflect (214). Occupational class tends to be more stable over time compared to income, which can fluctuate yearly (4). This means having one measure of it at an early time point is indicative of the broader circumstances (336,337). This stability provides a more reliable indicator of long-term

socioeconomic position, without losing the nuance of the gradient. This study used prior harmonisation efforts to ensure comparability of socioeconomic position across cohorts (338). Specifically, the Registrar General's 1990 social class framework was used to assess socioeconomic position in both the 1958 and 1970 cohorts, recording father's social class when participants were around 10 to 11 years old (216,338).

6.1.3 Gene-environment interplay in psychological distress

Psychological distress is thought to be influenced by both genetic and environmental factors, which can operate through both independent and interactive pathways (253,254,325,339–342). While these factors can each independently contribute to psychological distress, genes are necessarily expressed within environmental contexts, leading to complex gene-environment interplay (343). For example, genetic vulnerabilities to anxiety or depression might be more strongly expressed under adverse environmental conditions while remaining relatively dormant in a supportive environment (91). This aligns with prior research in the 1958 National Child Development Study, which suggests that childhood is a sensitive period in which environmental factors, including socioeconomic conditions, may influence the expression of genetic vulnerabilities related to psychological distress (91,325).

Independent associations would occur if both polygenic scores and father's social class separately influence psychological distress, regardless of each other. There are a few explanations for interactions, defined as the effect of an exposure changes depending upon the presence of the other exposure (344). A synergistic interaction would occur if individuals with both a high polygenic score and lower father's social class experience psychological distress at a rate much higher than would be expected from only adding the separate coefficients of high polygenic scores and low social class (344).

Previous psychological literature has focused on testing the diathesis-stress theory of gene x environment interaction (101). They found conflicting evidence between candidate gene and polygenic-based studies which utilised stressful life events as the environment (310). This research was based on the premise that exposures were stressful life events that were acute rather than chronic stressors (345). The ability to test not just adverse experiences but also advantaged and disadvantaged environments meant that the differential susceptibility theory could be tested within psychiatric genomic studies (91). A review of studies that utilised differential susceptibility theory added to the evidence that greater sensitivity was associated with a greater risk of psychopathology in adverse contexts, simultaneously sensitivity increased the protective nature of positive environments by decreasing risk (339,346). (99). Each GxE theory goes some way to explain the observed patterns in the relationship between genetic effects and macrosocial environments.

As described in Introduction Section 1.3, the differential susceptibility hypothesis posits that individuals with a high genetic liability for psychological distress may be more sensitive to adverse childhood environments, potentially leading to increased distress in adulthood (91). Therefore, research questions 1 and 2 use this framework to explore how a father's social class interacts with genetic liability for psychological distress across adulthood. According to the social context theory, the role of wider social norms and broader cultural environment means factors such as cohort

membership might create the conditions for a synergistic gene-environment interaction (99). Therefore, research questions 3 and 4 test for cohort and sex differences in gene-environment interactions, which adhere to testing the social control/context theory

6.1.4 Prior work on gene-environment interactions between polygenic scores and socioeconomic factors and association with psychological distress

Candidate gene studies were historically used to examine gene x environment interactions (310). These studies predominantly tested for the interaction of stressful life events and the 5-HTTLPR gene, which is thought to affect the serotonin transporter works (310). The underpinning theory was to test the assumptions of the diathesis-stress theory (94). However, these findings were not widely replicated (347).

Candidate GxE studies have several limitations that may contribute to the lack of replication. The candidate gene approach assumes that one gene strongly relates to distressed phenotypes. However, the specific biological mechanisms underlying psychiatric disorders remain unclear (310). Another key discovery is that most behavioural and, therefore, psychiatric traits are influenced by many thousands of SNPs with small effects rather than by a few gene variants with large effects, meaning they are polygenic (39). Given these limitations, GxE research regarding psychological distress outcomes is transitioning towards a polygenic approach in larger, better-powered samples (348).

Several studies have investigated the interactions between polygenic scores, adult socioeconomic position (across multiple domains), and psychological distress, focusing on gene-environment (GxE) interaction (Table 16). Overall, the studies all found independent associations between polygenic scores, socioeconomic position (across multiple domains) and psychological distress outcomes. There were inconsistent results regarding evidence for synergistic interactions.

Six of the studies tested adulthood socioeconomic circumstances of the participants. Two out of the seven studies identified evidence of synergistic interactions between polygenic scores and socioeconomic circumstances, whereby distress was higher in those in more disadvantaged circumstances and higher polygenic scores (325,340). In the same data that the current study uses, Keers et al. (2017), demonstrated that socioeconomic status was an independent predictor of psychological distress across adulthood (325). They found evidence for an interaction between lifecourse socioeconomic circumstances and polygenic scores in contributing to exponential risk in psychological distress outcomes. Qi et al. (2024), also in a U.K. context, identified an interaction between the index of multiple deprivation and polygenic scores impacting worse distress scores (340). Two of the studies did not test for interactions (341,349), while the final two studies tested for synergistic interactions but did not find evidence for them (253,254). In summary, across the six studies, there is conflicting evidence for gene-environment interactions when using socioeconomic circumstances.

Three of the studies tested childhood socioeconomic circumstances of the participants. (325,341,342). They all identified independent associations between childhood socioeconomic circumstances and distress outcomes. However, only Keers

et al. (2017) tested for synergistic interactions with polygenic scores, which were not significant for solely childhood environment but rather when childhood and adulthood socioeconomic circumstances were combined (325). This combined approach potentially introduces issues of reverse causality with the adulthood measures, as psychological distress could feasibly influence adult socioeconomic position.

Across all studies, key limitations include their reliance on retrospective self-reports of childhood socioeconomic circumstances which may underestimate the effect (350). Most studies did not use cohort comparisons to see if the independent associations and interactions remain stable in different populations, nor did they stratify by sex differences. Therefore, the current study seeks to fill this gap.

6.1.5 Cohort and sex differences in gene-environment interaction studies

There are observed disparities in distress outcomes between the 1958 and 1970 cohorts, which also exhibit a gender gap, with women generally experiencing higher distress levels than men; the 1970 cohort showed higher average distress scores than the 1958 cohort (124). A study that triangulated three U.K. cohorts found the gender gap in distress to be persistent across all age groups (351). The social gradient at the individual level may differ considerably depending on the gender equality mechanisms in a given population's specific context. For example, the social gradient may not appear if the socioeconomic metric is education level if using data from the millennial generation as access to higher education equalised (352).

Sex and cohort differences may moderate the interaction between polygenic scores and father's social class in predicting psychological distress or be independently associated (164,306). Theoretically, sex differences in this interaction may be due to the changing environment. For example, different cohorts experienced varying levels of gender role rigidity (p.90) (353). Older cohorts may show stronger effects of father's social class on daughters due to greater economic dependence on their family or spouse (354). While more recent cohorts might show more similar patterns between sexes as gender roles evolved (218,355). Cohort differences, too, may be driven by societal changes in social mobility, educational opportunities, and labour market structures across time (4,333).

The impact of father's social class on psychological outcomes may vary across cohorts due to shifting economic conditions such as deindustrialisation and changes in intergenerational wealth transfer patterns (356,357). Additionally, cohort differences could reflect varying exposure to major societal stressors (e.g., economic recessions, technological changes) that may amplify or dampen the effects of both genetic liability and social class on psychological distress.

Both sex and cohort differences were found in previous chapters of this thesis. In Study 1 (Section 4), beta coefficient estimates were higher for females compared to males when the main analysis was stratified. Then in Study 2 (Section 5), results also indicated that females tended to experience higher levels of observed psychological distress than males, irrespective of their genetic liability in 1958c. Study 2 further investigated cohort differences, which were not different due to genetic liability, meaning other environmental factors might contribute to the distress gap between the cohorts. To the best of my knowledge, no prior literature was found that tested

polygenic scores – socioeconomic circumstance interactions comparing cohorts and by sex.

Previous chapters of the thesis found that women and participants in the 1970c have a stronger association between their polygenic scores and observed psychological distress. Based on the current evidence, it is plausible that there will be independent associations between the exposure of polygenic scores and father's social class with the outcome of psychosocial distress; and that these differ by cohort and sex. The social control model (as outlined in Introduction Section 1.3) states that genetic factors may be filtered or buffered by social norms and structural constraints (99). The current study leverages the social context to explore how father's social class interacts with genetic liability for psychological distress in the two cohorts and by sex.

Given the consistent challenge of limited statistical power in GxE research, a meta-analytic approach may be a useful method to obtain the necessary sample size to detect small effect size interactions. By synthesising across multiple cohorts, this approach would help add to the evidence with robust estimates of potential interaction effects between polygenic scores and socioeconomic circumstances.

6.1.6 Summary

In summary, this chapter seeks to build on the findings of and gaps within previous research by examining how father's social class interacts with genetic liability for psychological distress across adulthood. By using two British birth cohorts, this study will assess whether the interactions between polygenic scores and father's social class differ by cohort. Then, given sex differences in psychological distress, it assessed whether the relationship differs by sex.

Table 18: Summary table of key evidence from studies that examined the independent and interaction effects of polygenic scores and socioeconomic circumstance variables and their association with psychological distress outcomes

Author (Year)	Country	Dataset (Type)	Time of Data	Sample Size	Sample age range	Depressive Symptoms Measure	SES Measure	Same Participants	Independent Effect of exposures	Synergistic Interaction Effect of exposures
Kosciusko (2023)	England	English Longitudinal Study of Ageing (Panel)	2004 - 2019	6,202	50-95	Centre for Epidemiologic Studies-Depression Scale (CES-D)	Wealth; Years of Schooling	No	Yes Additive/independent associations between PGS and SES factors with depressive symptoms	No
Lam (2019)	Australia	Twins Research Australia (Twin Cohort)	2014 - 2017	3,662	18+	Kessler-6 Psychological Distress	Australian Socioeconomic Index; IRSD; Income	Yes	Yes - Additive/independent associations between income, IRSD, SES index	Not tested
Keers (2017)	U.K.	The 1958 National Child Development Study (Cohort)	1981 - 2008	7,075	23-50	Malaise Inventory Score	Composite score including social class, employment status, financial hardship, and tenure of accommodation taken at ages 7, 11, 16, 23, 33, 42, and 50.	Yes	Yes Additive/independent associations between PGS and both childhood and concurrent SES factors with depressive symptoms	Yes, for Childhood + adulthood SES interaction with PGS.
Stringa (2020)	Netherlands	Longitudinal Aging Study Amsterdam (Panel)	1992 - 2013	2,279	55-95	CES-D	Partner status; social network size; emotional support;	No	Yes - additive/independent associations were found between PRS, social factors and Depression.	No

Qi 2024	U. K	UK Biobank (Panel)	2006 - 2010	74,425	40-69	Generalized Anxiety Disorder (GAD)-7 scale and the Patient Health Questionnaire (PHQ)-9 (	Index of Multiple Deprivation (IMD)	No	Yes - additive/independent associations were found between PRS and IMD	Yes
Agerbo 2021	Denmark	iPSYCH2012 (Panel)	1981 - 2005	17,098	16-40	Diagnosed with major depression	Educational Attainment, occupational status, marital status, maternal educational attainment, paternal labour market affiliation, and maternal marital status	No	Yes - additive/independent associations were found between PRS, SES factors and Depression	Not tested
Hoang 2023	USA	Health & Retirement Study (Panel)	1992 - 2018	7,357	51-94	CES-D	Childhood SES index: social (parenting), economic (fathers' occupation) and human capital (parents' education)	No	Yes - associations were found between PRS, social factors and depression	Not tested

6.2 Method

6.2.1 Data

The study used data from the 1958 National Child Development Study (1958c) and the 1970 Birth Cohort Study (1970c) (187,188) described in Section 3.1 of the thesis Methods.

6.2.2 Measures

6.2.2.1 Psychological Distress

The malaise inventory score was the outcome variable in both cohorts. The 9-item version was used for the current chapter. For details regarding the development, harmonisation, measurement invariance, internal validity, and external validity of the Malaise Inventory, they are presented in the Methods Section 3.2.1.

6.2.2.2 Polygenic score (PGS)

Polygenic score was the same used in Study 2 (Methods Section 3.2.2). The polygenic score was standardised across both cohorts (rather than separately) to have a mean of 0 and a standard deviation of 1. The main analysis exposure is the standardised polygenic score for psychological distress at a threshold of 0.005, which has overlapping SNPs between the 1958c and 1970c.

6.2.2.3 Father's Social Class

Socioeconomic position has been defined in many ways, as discussed in the Introduction (Section 1). Father's social class is based on the Registrar-General's Social Classes, which classifies social class by occupational group: I (professional), II (managerial and technical), IIIN (skilled non-manual), IIIM (skilled manual), IV (partly skilled), and V (unskilled) (214).

For research question 1, father's social class was categorised into a manual versus non-manual classification for ease of interpretation of the interaction. The non-manual category comprises classes I-III from the registrar general's social class, which includes professional, managerial, and skilled non-manual labourers. The manual category includes classes IV-VI from the registrar general's social class, including the manual, partially skilled and unskilled labourers. The economically inactive category contained n=372 participants, which included fathers of participants who were unemployed, retired and disabled, these participants were coded as missing as they were economically inactive. The 6-category version is kept for the analysis in research question 2.

Mother's social class was not used as in both cohort's social class was classified using the Registrar General's Social Class (RGSC), which was primarily based on male occupations (216,327). This reflects the social norms of the time whereby expansion of women's economic activity in the workplace had reached 57% amongst women of prime working age between 25-and 57 by 1975, compared to 90% employment for

men (217). In the 1950s and 1970s, fathers were likelier to be most families' primary breadwinners. Their occupation and social class were often considered representative of the family's overall socioeconomic status (215,218). Therefore, only father's social class was used in the present study.

Socioeconomic position is partly heritable, with twin-based heritability estimates ranging from 34% to 47%, and SNP-based heritability is approximately 18% (49,50). This implies that paternal occupational class may capture not only environmental influences but also genetic liability shared between parents and offspring (41). In the 1958c and 1970c, father's social class was weakly correlated with the polygenic score, r=0.006, which may mean it is unlikely to confound associations. However, prior evidence has shown significant associations between polygenic scores for major depressive disorder and socioeconomic outcomes (51,52). There is also a known genetic correlation between mental health disorders and socioeconomic status traits (53). However, the ability to disentangle which SNP variants contribute causally to both traits, as opposed to those that overlap due to correlation, remains limited. Therefore, paternal SES should be interpreted as a proxy for early-life environment that is not free from genetic influence, and our results should be considered in light of this.

6.2.2.4 Covariates

The covariates included the ten first principal components and sex as male (0) and female (1). The 10 principal components were included to control for residual population stratification (224). The cohort variable was included as a covariate in research question 1's and research question 2's models.

Parental mental health was not controlled for in Study 3. Although parental mental health is strongly associated with offspring psychological distress, it occupies an ambiguous position in the causal structure. On the one hand, parental mental health partly lies on the causal pathway between genetic liability and offspring outcomes; adjusting for it would block indirect "genetic-nurture" pathways and underestimate the total contributing association between the polygenic score and adult distress (41). On the other hand, not adjusting for it may therefore overstate direct genetic effects (41,42). For father's social class, parental mental health additionally acts as a potential confounder of the social class–offspring distress association. This creates a trade-off: excluding parental mental health risks residual confounding, while including it risks adjusting on the causal pathway. Furthermore, the measurement of parental mental health was not uniform across the two cohorts (43). Parental mental health measurement varied across cohorts, assessment ages (7 years-1958c; 10 years-1970c), and dichotomisation methods, despite harmonisation to a binary 'poor/not poor' indicator, there was an arbitrary matched 3% prevalence rate. In light of these considerations, we elected not to control for parental mental health in the main models, but acknowledge that this decision may bias the estimates—either by overstating direct genetic effects or by leaving unmeasured familial confounding. This limitation should be taken into account when interpreting the findings.

6.2.3 Analytical Strategy

The missingness patterns, biases, and methods for addressing them are described in more detail in Methods Section 3.4 of the thesis.

6.2.3.1 Research Question 1: Does the association between polygenic scores for distress and adult psychological distress outcomes vary according to father's social class?

To answer research question 1 multiple conceptualisations of father's social class were tested. These include: binarised father's social class variable (manual versus non-manual), the 6-category Registrar-General's Social Class and a ridit score of the 6-category variable.

Binary: manual v non-manual

A mixed-effects model was run to examine the interaction between polygenic scores for psychological distress and dichotomised father's social class and how these interactions affect psychological distress.

A stepwise approach was taken. The first model utilised a pooled dataset of 1958c and 1970c and included fixed effects for binarised father's social class with covariates including cohort, age, age squared and the first 10 principal components to adjust for population stratification. Confidence intervals were bootstrapped to account for the outcome being zero-inflated and, therefore, a non-normal distribution. Random intercepts for individuals were included to account for repeated measurements. The model was implemented in Stata using the `mixed` command. Model 2 then added the polygenic score to identify independent associations. The final model 3 included an interaction between the PGS and the binarised father's social class.

A likelihood ratio test was used to compare Model 2 (the base model without the interaction term) to Model 3 (which includes the interaction term), to determine whether the inclusion of the interaction term significantly improved the model fit. The likelihood ratio test assesses whether the additional complexity introduced by the interaction term is justified by a statistically significant improvement in the model's ability to explain the data (363).

This test is particularly useful in nested models, where one model is a subset of the other. If the p-value from the likelihood ratio test is below the chosen threshold (e.g., 0.05), it indicates that the interaction term meaningfully contributes by explaining variability in the dependent variable.

To complement the likelihood ratio test, the change in Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) was also recorded between Model 2 and Model 3. These metrics provide an additional assessment of model fit, with lower values indicating better fit (364). AIC prioritises goodness of fit while penalising model complexity, and BIC adds a stricter penalty for the number of parameters. Together, these criteria offer a balance between improving explanatory power and avoiding overfitting.

6-category Father's Social Class

A mixed-effects model was run to examine the interaction between polygenic scores for psychological distress and the 6-category father's social class and how these interactions affect psychological distress.

A stepwise approach was taken. The first model utilised a pooled dataset of 1958c and 1970c and included fixed effects for 6-category father's social class with covariates including cohort, age, age squared and the first 10 principal components to adjust for population stratification. Confidence intervals were bootstrapped to account for non-normal distribution of the outcome. Random intercepts for individuals were included to account for repeated measurements. The model was implemented in Stata using the `mixed` command. Model 2 added the polygenic score to identify independent associations. The final model 3 included an interaction between the PGS and the 6-category father's social class. The marginal effects of the interaction were calculated and plotted.

The F-statistic was employed to evaluate the overall significance of the regression model, assessing whether the included 6-category father's social class predictor explains a significant proportion of variance in the dependent variable compared to a null model with no predictors (365). For example, when examining the interaction term, the F-statistic provides a measure of whether the variation explained by the interaction is significantly greater than what would be expected by chance (365).

The change in Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) was also recorded between Model 2 and Model 3. These metrics provide an additional assessment of model fit, with lower values indicating better fit (364). AIC prioritises goodness of fit while penalising model complexity, and BIC adds a stricter penalty for the number of parameters. Together, these criteria offer a balance between improving explanatory power and avoiding overfitting.

Ridit Score

A mixed-effects model was run to examine the interaction between polygenic scores for psychological distress and a ridit score derived from the 6-category father's social class.

The ridit score represents the relative rank or cumulative proportion of individuals in each category, offering a continuous gradient of socioeconomic position rather than a binary or categorical classification (366,367).This approach was included to account for the possibility that the binary manual versus non-manual classification might oversimplify socioeconomic gradients and fail to capture subtle differences in the effects of father's social class on the outcome.

A stepwise approach was taken. The first model utilised a pooled dataset of 1958c and 1970c and included fixed effects for ridit score with covariates including cohort, age, age squared and the first 10 principal components to adjust for population stratification. Confidence intervals were bootstrapped to account for non-normal distribution of the outcome. Random intercepts for individuals were included to

account for repeated measurements. The model was implemented in Stata using the `mixed` command. Model 2 added the polygenic score to identify independent associations. The final model 3 included an interaction between the PGS and the ridit score.

The change in Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) was also recorded between Model 2 and Model 3. These metrics provide an additional assessment of model fit, with lower values indicating better fit (364). AIC prioritises goodness of fit while penalising model complexity, and BIC adds a stricter penalty for the number of parameters. Together, these criteria offer a balance between improving explanatory power and avoiding overfitting.

6.2.3.2 Research Question 2: Do the independent associations or interactions between polygenic score for distress and father's social class on adulthood psychological distress differ by cohort or by sex?

To answer research question 2 by cohort, separate mixed-effects models were fitted for each cohort to examine the interaction between the polygenic score and father's social class. Covariates of sex, age, age squared, and the 10 principal components were included.

To answer research question 2 by sex, separate mixed-effects models were fitted for each sex to examine the interaction between the polygenic score and father's social class. Covariates of cohort, age, age squared, and the 10 principal components were included.

To formally compare the interaction coefficient between cohorts and sexes, z-statistic was computed as the difference between coefficients divided by the square root of the sum of their squared standard errors: $(b1 - b2)/\sqrt{(SE1^2 + SE2^2)}$. This approach tests whether the difference between coefficients is statistically significant, with values exceeding ±1.96 indicating significant differences at $p < 0.05$.

6.2.3.3 Research Question 3: Using meta-analysis, is there robust evidence of an interaction between polygenic score for psychological distress and socioeconomic circumstances on adulthood psychological distress outcomes?

To answer research question 5, a meta-analysis of gene x environment interaction studies was conducted. The meta-analysis objective was to investigate whether there is evidence for an interaction or independent associations between proxies of socioeconomic circumstances, polygenic scores for psychological distress and observed psychological distress (368).

Search Strategy

The advanced search function on Ovid was used to include databases: Ovid MEDLINE (1946–present November 2024) and Epub Ahead of Print, In-Process and other non-indexed; EMBASE (1974–present November 2024); American Psychology

Association (APA) PsychArticles full text, APAPsychINFO (1806–present November 2024). The search strategy was then re-run on PubMed for further studies not captured via previous databases.

Search terms were included in all fields:
polygenic score or polygenic risk score or polygenic index or polygenic*
AND
psychological distress or depressive symptoms or depress* or anxiety symptoms or MDD or GAD or generalised anxiety disorder or major depressive disorder
AND
social class or socioeconomic or socioeconomic circumstances or socioeconomic status or socioeconomic position
AND
interaction or gene x environment interaction or GxE or G x E

Selection process
In total, 139 potential publications were identified. Following the screening of titles and abstracts, 5 publications remained. The 5 full papers were screened and 2 were included alongside the results from the two studies in the current chapter.

Inclusion Criteria for Studies:

- Interactions between polygenic score for psychological distress and socioeconomic circumstance indicators.
    - Polygenic scores do not have to have the same GWAS and can have broad depression, MDD, depressive symptom, anxiety symptoms or psychological distress as phenotype
    - Socioeconomic circumstance indicators can also be broad but must be related to occupational status, income, social class, or education.
    - Socioeconomic circumstances do not have to be limited to parental over-participant.
- Quantitative effect sizes (e.g., regression coefficients, odds ratios) for the interaction term with standard errors, confidence intervals, or p-values.
- Adult populations (18+).
- Independent datasets (i.e., studies that are not overlapping or drawn from the same source as current study).

Meta-Analytical Approach

Beta coefficient, confidence intervals and sample size were extracted from each study. Both a fixed and random effect model were tested. The random-effect model accounts for heterogeneity due to varying socioeconomic circumstance measures and population demographics (369). The pooled interaction effect size with 95% confidence intervals was reported and heterogeneity was assessed using I^2 statistics.

6.2.3.4 Supplemental and Sensitivity Analyses

Testing polygenic score for psychological distress at different p-value thresholds:

Model 3 from research question 1 and research question 2 were re-run using the most conservative p-value threshold of the genome-wide significance $5x10^{-8}$ and the least conservative with a p-value threshold of 1. This was done to assess whether estimates differed greatly between the extremes of the thresholds and the threshold selected (0.005), thereby examining the robustness of the results. This is important because the choice of threshold may affect which SNPs are included in the polygenic score. Table 5 summarises the number of SNPs at the different p-value thresholds.

6.3 Results

6.3.1 Descriptives

6.3.1.1 Mean, Standard Deviations and Histograms

Figure 18 shows mean scores varied across ages and cohorts, indicating that they fluctuated at different ages without a distinct increasing trend. Supplementary Table 10 summarises the mean and standard deviation of the 9-item malaise inventory scores in the two cohorts 1958c and 1970c at ages 23, 26, 30, 33, 34, 42, 46 and 50 stratified by manual versus non-manual father's social class.

Participants from a non-manual background had consistently lower mean scores of psychological distress across both cohorts. For example, in the 1958 cohort the mean distress score at age 23 for individuals from a manual background was 1.19 (SD = 1.51), while for those from a non-manual background, it was notably lower at 1.02 (SD = 1.39). This trend of lower distress scores for the non-manual group persisted across all age points within both cohorts.

The 1970c cohort had higher mean psychological distress scores at equivalent ages to the 1958c cohort. For example, at age 42, the mean distress score for the manual group in 1970c was 1.49 (SD = 1.75), compared to 1.47 (SD = 1.68) in the 1958c cohort. This suggests that the baseline levels of psychological distress may have increased for those in the 1970c.

The standard deviations of psychological distress scores further illustrate the variability in distress levels within each group. Across both cohorts, participants from non-manual backgrounds consistently showed smaller standard deviations compared to those from manual backgrounds, indicating less variability in distress levels among individuals from higher socioeconomic backgrounds. For example, in the 1958c cohort at age 33, the standard deviation for the manual group was 1.51, whereas it was lower for the non-manual group at 1.25.

In contrast, the 1970c cohort exhibited larger standard deviations of psychological distress scores at equivalent ages compared to the 1958 cohort, indicating greater variability in distress levels among the 1970c participants. This pattern is seen at age 50, where the standard deviation for the manual group in 1970c was 2.06, considerably higher than the 1.85 observed in the 1958c cohort. Similarly, the non-manual group in 1970c at age 50 had a standard deviation of 1.94, compared to 1.75 in the 1958c cohort.

Supplementary Figure 4, the histograms illustrate that, while the overall distribution of polygenic scores is similar across cohorts and social classes, there are subtle differences in the density and spread of scores. The wider distribution seen in the manual backgrounds, particularly in the 1970c cohort, indicates a broader range of distress outcomes in these groups

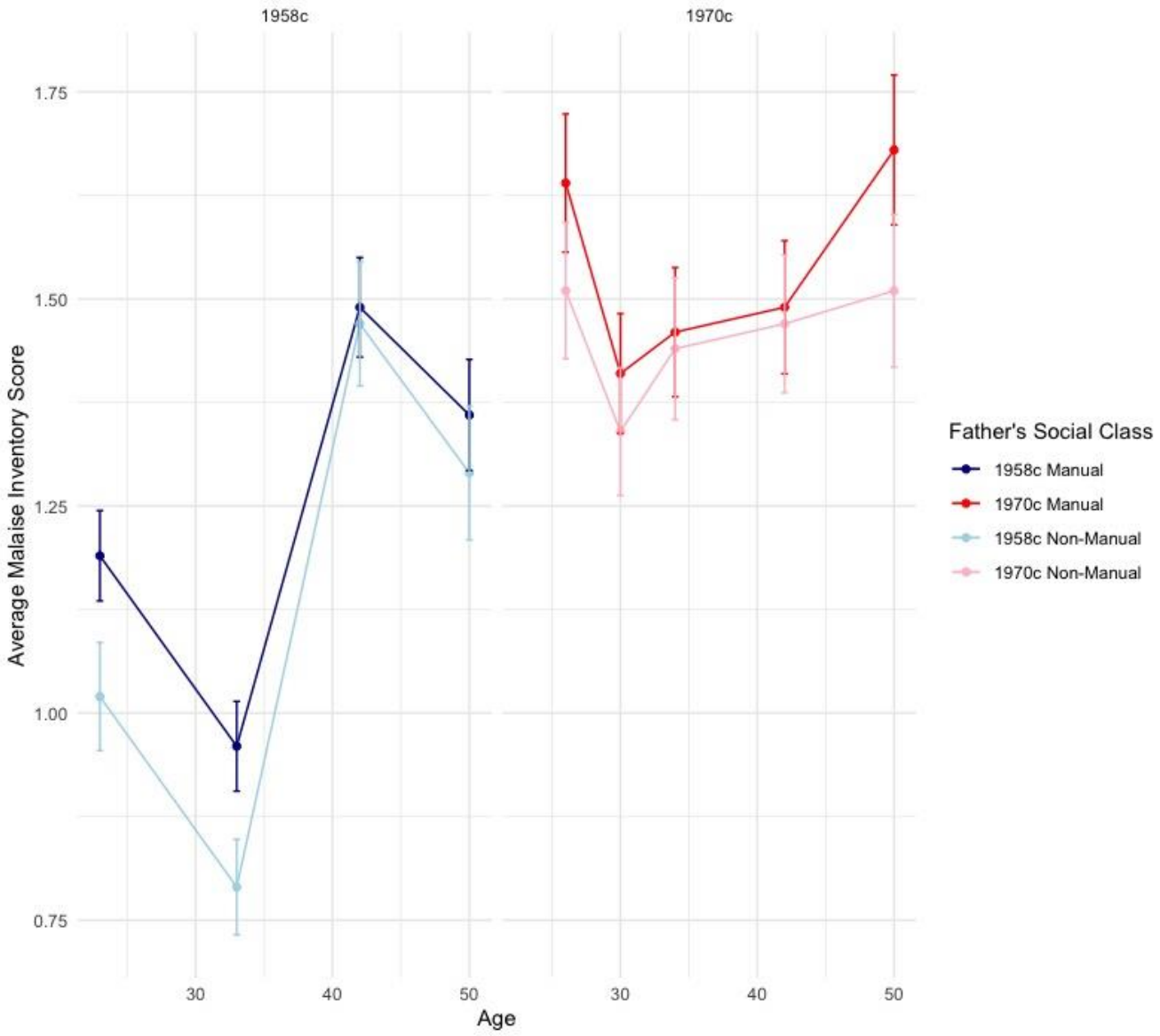


Figure 18: Mean and standard deviation of Malaise Inventory Score by father's social class coded as manual versus non-manual at ages 23-50 in the 1958c and 1970c (N=9,620)

6.3.2 Regression Results

6.3.2.1 Research Question 1: Do polygenic scores for psychological distress interact with father's social class to influence observed psychological distress in adulthood?

Binary manual v non-manual

The results from the multilevel mixed-effects model examining the interaction between polygenic score and father's social class indicate that both factors are predictors of psychological distress, with a positive direction of association (Table 19).

Independent Effects

Individuals with a father from a manual background exhibited higher distress scores compared to those from a non-manual background, as evidenced by a positive beta coefficient of 0.21 (95% CI: 0.18–0.23). This suggests that early socioeconomic disadvantage is associated with increased levels of psychological distress in adulthood. With the addition of the polygenic score for psychological distress to model 2, it also showed a positive association with distress outcomes, with a beta coefficient of 0.22 (95% CI: 0.21–0.23). The addition of the polygenic score did not attenuate father's social class as a predictor of psychological distress outcomes. The consistent positive effect across the sample suggests that genetic liability plays a role in psychological distress outcomes, independent of socioeconomic background.

Interaction Results

There was no strong evidence of an interaction between participant with fathers with a manual background and polygenic score with a small beta coefficient of -0.02, plus overlapping confidence intervals and null p value (95% CI: -0.05 to 0.01, p = 0.238). These findings suggest that while both childhood socioeconomic position and genetic liability independently contribute to psychological distress, there is no interaction in their association with distress. In other words, the influence of polygenic score on distress is relatively stable across different socioeconomic backgrounds. The present study indicates that the effects of early-life socioeconomic conditions and genetic liability on mental health may be additive rather than multiplicative.

Model Fit Statistics

To test whether model 3 was a better fit for the data than model 2, a likelihood-ratio test was conducted. The results of the likelihood-ratio test indicated that the interaction between polygenic score and cohort did not significantly improve the model fit, $\chi 2(1) = 0.35$, $p=0.552$. The Akaike Information Criterion (AIC) was 124975.6 for model 2 and 124977.3 for model 3. The AIC values showed that the null model (without the interaction term) had a lower AIC than the full model (with the interaction term). A lower AIC indicates that the full model fits the data better, considering both complexity and goodness of fit. Similarly, the Bayesian Information Criterion (BIC) for model 2 was 125136.6, compared to 125146.8 for model 3. The lower BIC of model 2 i.e. the model without the interaction demonstrates a better fit.

Table 19: Results from the multilevel mixed-effects model interacting mean-centred polygenic score at the 0.005 p-value threshold and 2-category father's social class for 1958c and 1970c pooled (N=8,923)

	Beta Coefficient	Bootstrap 95% CI	P value
Model 1: Father's Social Class only			
Non-manual is the reference category			
Manual	**0.21**	**0.18-0.23**	**0.001**
Model 2: Father's Social class + PGS			
Manual	**0.20**	**0.18-0.23**	**0.001**
PGS	**0.22**	**0.21-0.23**	**0.001**
Model 3: Father's Social class x PGS			
Manual	**0.20**	**0.14–0.26**	**0.001**
PGS	**0.23**	**0.19–0.26**	**0.001**
PGS*Manual	-0.02	-0.05–0.01	0.238

6-category Father's Social Class

Independent Effects

The results demonstrate a consistent main effect of father's social class on psychological distress across all three models (Table 20). Across all models, a clear social gradient is observed, with predicted psychological distress increasing as social class declines. Compared to the reference category (I Professional), the effect sizes for distress are largest in the lowest classes, such as 0.41 for VI Unskilled (95% CI [0.34, 0.49], p = 0.001), followed by V Partially Skilled at 0.29 (95% CI [0.23, 0.35], p = 0.001). When the polygenic score is introduced in Model 2, the main effect of father's social class remains significant across all categories, and the effect sizes show minimal attenuation. Similarly, the polygenic score itself predicts higher distress outcomes (= 0.23, 95% CI [0.19, 0.25], p = 0.001), indicating that both socioeconomic position and genetic liability independently contribute to psychological distress. In Model 3, which includes interaction terms, the main effects for father's social class and polygenic score persist with only slight reductions in effect sizes. This suggests that neither the inclusion of polygenic score nor its interaction with social class substantially diminishes the direct effect of father's social class on distress.

Interaction Effects

There was limited evidence of an interaction between polygenic score and father's social class. Of the interaction terms tested in Model 3, only PGS × IV Skilled Manual showed an interaction relative to the most advantaged social class with a beta coefficient of 0.10 (95% CI [0.04, 0.16], p = 0.001), suggesting that genetic liability has a slightly stronger effect on psychological distress in this group. However, there was no evidence of an interaction for other categories. This indicates that the moderating role of socioeconomic position is not consistent across social class categories. Despite the addition of interaction terms, the main effects of father's social class and polygenic score did not attenuate, reinforcing their independent contributions to psychological distress and suggesting that socioeconomic position has only a limited moderating influence on the genetic risk for distress. However, interpretation of this interaction should be cautious as comparative to other categories it had the largest sample size (Supplementary Table 11).

The F-statistic tests indicate differences in the interaction effects between polygenic scores and social class depending on how social class is categorised. For the 6-category model, the test of the five interaction terms did not reach significance ($\chi 2$ (5) = 9.95, p = 0.0767), suggesting no evidence of interactions across social class categories. Which is similar to the binary model (manual vs. non-manual social class), which also showed no evidence of interaction ($\chi 2$ (1) = 0.35, p = 0.552).

The predictive margins plot illustrates a consistent positive association between polygenic scores and the predicted outcome across all social class categories, indicating that genetic liability influences the outcome regardless of social class (Figure 19). However, a clear social gradient is evident, with higher social classes (e.g., I Professional and II Managerial) having lower predicted outcomes compared to lower social classes (e.g., V Partially Skilled and VI Unskilled) across all levels of PGS. Notably, VI Unskilled consistently exhibits the highest predicted outcomes across the

range of PGS, reflecting the main effect of social disadvantage. However, the slope of the line, which indicates the strength of the genetic effect, is steepest for IV Skilled Manual, the only category with an interaction between PGS and social class. This suggests that while VI Unskilled has higher overall distress, the amplification of genetic effects is most pronounced among IV Skilled Manual, where lower social advantage may uniquely exacerbate genetic risk. In contrast, the buffering effect of higher social classes, such as I Professional, is reflected in their more attenuated slopes. The error bars suggest that while there is overlap between some adjacent categories, the overall trend reflects differences in predicted outcomes.

Model Fit Statistics

To test whether model 3 was a better fit for the data than model 2, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were compared. For model 2 the AIC was 130821.4 and for model 3 it was 130821.5. The AIC values showed that the null model (without the interaction term) had a lower AIC than the full model (with the interaction term). A lower AIC indicates that the full model fits the data better, considering both complexity and goodness of fit. Similarly, the BIC for model 2 was 131016.8, compared to 131059.4 for model 3. The lower BIC of model 2 i.e. the model without the interaction demonstrates a better fit

Table 20: Results from the multilevel mixed-effects model interacting mean-centred polygenic score at the 0.005 p-value threshold and 6-category father's social class for 1958c and 1970c pooled (N=8,923)

	Beta Coefficient	Bootstrap 95% CI	P value
Model 1: Father's Social Class only			
I Professional is the reference category			
II Managerial	**0.12**	**0.07-0.18**	**0.001**
III Skilled non-manual	**0.06**	**0.01-0.12**	**0.039**
IV Skilled manual	**0.27**	**0.22-0.32**	**0.001**
V Partially Skilled	**0.29**	**0.23-0.35**	**0.001**
VI Unskilled	**0.41**	**0.34-0.49**	**0.001**
Model 2: Father's Social class + PGS			
I Professional is the reference category			
II Managerial	**0.12**	**0.07-0.18**	**0.001**
III Skilled non-manual	**0.06**	**0.01-0.13**	**0.026**
IV Skilled manual	**0.27**	**0.21-0.31**	**0.001**
V Partially Skilled	**0.29**	**0.20-0.32**	**0.001**
VI Unskilled	**0.41**	**0.32-0.47**	**0.001**
PGS	**0.23**	**0.19-0.25**	**0.001**
Model 3: Father's Social class x PGS			
I Professional is the reference category			
II Managerial	**0.13**	**0.07–0.18**	**0.001**
III Skilled non-manual	**0.07**	**0.01-0.13**	**0.001**
IV Skilled manual	**0.26**	**0.21-0.31**	**0.001**
V Partially Skilled	**0.27**	**0.21-0.33**	**0.001**
VI Unskilled	**0.40**	**0.33-0.47**	**0.001**
PGS	**0.18**	**0.12–0.24**	**0.001**
PGS*II Managerial	0.03	-0.03-0.09	0.414
PGS*III Skilled non-manual	0.03	-0.03-0.10	0.344
PGS*IV Skilled manual	**0.10**	**0.04-0.16**	**0.001**
PGS*V Partially Skilled	0.06	-0.02-0.13	0.126
PGS*VI Unskilled	-0.04	-0.12-0.04	0.316

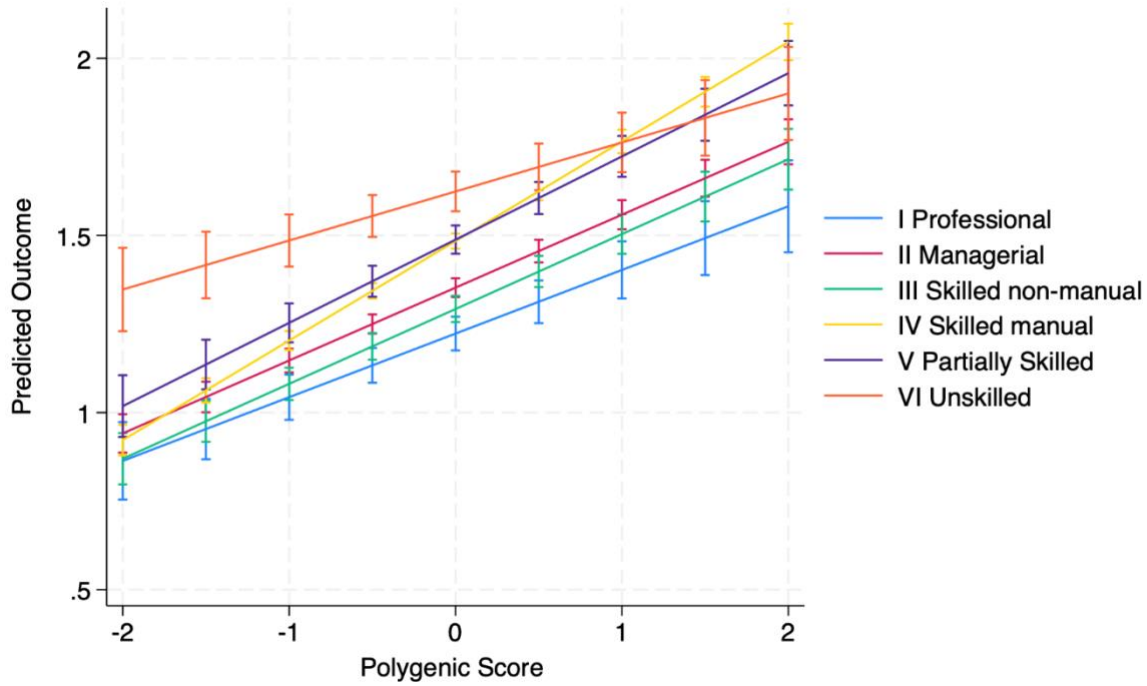


Figure 19: Predictive margins of psychological distress for 6-categories of father's social class across the polygenic score

Ridit Score

The ridit score analysis provides additional insight into the relationship between father's social class, polygenic scores, and their interaction on the predicted outcome (Table 21). In Model 1, the ridit score, which represents a continuous measure of social class rank, shows an independent association with psychological distress (beta = 0.36, 95% CI [0.31, 0.40], p = 0.001), indicating that, on average, individuals from lower social classes experience worse outcomes.

In Model 2, the inclusion of the polygenic score slightly attenuates the effect of the ridit score but both ridit and polygenic score remain independent predictors.

In Model 3, which tests the interaction between ridit and polygenic score, the main effects of both ridit (beta = 0.32, 95% CI [0.27, 0.37], p = 0.001) and PGS (beta = 0.22, 95% CI [0.19, 0.25], p = 0.00) persist, but there is no evidence of an interaction (beta = 0.03, 95% CI [-0.02, 0.08], p = 0.188).

Model Fit Statistics

To test whether model 3 was a better fit for the data than model 2, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were compared. For model 2 the AIC was 130823.4 and for model 3 it was 130825. The AIC values showed that the null model (without the interaction term) had a lower AIC than the full model (with the interaction term). A lower AIC indicates that the full model fits the data better, considering both complexity and goodness of fit. Similarly, the BIC for model 2 was 130984.8, compared to 130994.9 for model 3. The lower BIC of model 2 i.e. the model without the interaction demonstrates a better fit

Table 21: Results from the multilevel mixed effects model interacting mean-centred polygenic score at the 0.005 p value threshold and ridit score of the 6-category father's social class for 1958c and 1970c pooled (N=8,923)

	Beta Coefficient	Bootstrap 95% CI	P value
Model 1: Father's Social Class only			
Ridit	**0.36**	**0.31-0.40**	**0.001**
Model 2: Father's Social class + PGS			
Ridit	**0.32**	**0.27-0.37**	**0.001**
PGS	**0.24**	**0.22-0.25**	**0.001**
Model 3: Father's Social class x PGS			
Ridit	**0.32**	**0.27-0.37**	**0.001**
PGS	**0.22**	**0.19–0.25**	**0.001**
PGS*Ridit	0.03	-0.02-0.08	0.188

6.3.2.2 Research Question 2: Do the independent associations or interactions between polygenic scores for distress and father's social class on adulthood psychological distress differ by cohort or by sex?

Cohort

The results from the multilevel mixed-effects model examining the cohort-specific effects of the polygenic score and father's social class indicate that both factors are predictors of psychological distress across cohorts, with a positive association direction (Table 22).

Independent Effects

The polygenic score for psychological distress also demonstrated a positive association with distress outcomes within both cohorts, with a beta coefficient of 0.25 (95% CI: 0.21–0.29) for the 1970 cohort and 0.24 (95% CI: 0.20–0.28) for the 1958 cohort. These results reinforce the idea that genetic liability to psychological distress remains a predictor across different birth cohorts, suggesting that the correlation of genetic factors with distress is consistent by age and by cohort. Participants with a father from a manual background demonstrated a positive association with distress outcomes within both cohorts, with a beta coefficient of 0.20 (95% CI: 0.12–0.28) for the 1958 cohort and 0.22 (95% CI: 0.10–0.33) for the 1970 cohort. These results indicate that early socioeconomic disadvantage is associated with increased levels of psychological distress in adulthood in both birth cohorts.

Interaction Results

There was no evidence of an interaction in both cohorts. The beta coefficient was -0.03 (95% CI [-0.08 to 0.02], p = 0.307) in 1958 and 0.01 (95% CI [-0.10 to 0.12], p = 0.848). This means the interaction of polygenic score and father's social class on distress is relatively stable regardless of the cohort.

A z-statistic was computed to compare the interaction coefficients from the 1970c and 1958c models. The estimated difference in interaction coefficients between cohorts was -0.02 (SE = 0.06)**.** The z-statistic for this difference was -0.38 (p= 0.703)**.** This indicates no evidence of an overall difference in the effect of polygenic score on psychological distress between the 1958 and 1970 cohorts**.**

Table 22: Results from the cohort-stratified multilevel mixed effects model interacting mean-centred polygenic score at the 0.005 p value threshold and father's social class for 1958c and 1970c (1958c N=5,242; 1970c N=3,681)

1958c (N=5,242)	Beta Coefficient	Bootstrap 95% CI	P value
Non-manual is the reference category			
Manual	**0.20**	**0.12–0.28**	**0.001**
PGS	**0.24**	**0.20–0.29**	**0.001**
PGS*Manual	-0.03	-0.10-0.06	0.519
1970c (N=3,681)	Beta Coefficient	Bootstrap 95% CI	P value
Non-manual is the reference category			
Manual	**0.22**	**0.10–0.33**	**0.001**
PGS	**0.25**	**0.19-0.31**	**0.001**
PGS*Manual	0.01	-0.10-0.12	0.848

Sex

The results from the multilevel mixed-effects model examining the sex-specific effects of the polygenic score and father's social class indicate that both factors are predictors of psychological distress across sexes (Table 23).

Female Independent Effects

For females, the polygenic score was positively associated with psychological distress, with a beta coefficient of 0.27 (95% CI [0.22–0.33], p = 0.001). This suggests that genetic liability plays a role in psychological distress among women, with higher polygenic scores correlating with greater distress. Additionally, having a father from a manual labour background was associated with increased psychological distress, as indicated by a beta coefficient of 0.30 (95% CI [0.18–0.41], p = 0.001). This highlights the impact of early socioeconomic disadvantage on distress outcomes in adulthood for females.

Female Interaction Results

There was no evidence of an interaction between the polygenic score and father's social class (PGS*Manual) for females, with a small beta coefficient of 0.03 (95% CI [-0.09 to 0.14], p = 0.668). This result suggests that the combined effect of genetic liability and socioeconomic background do not interact when predicting psychological distress amongst women.

Male Independent Effects

For males, the polygenic score also showed a positive association with psychological distress, with a beta coefficient of 0.22 (95% CI [0.17–0.27], p = 0.001). This indicates that, similar to females, genetic liability is a consistent predictor of psychological distress among males. Additionally, having a father from a manual labour background was associated with increased psychological distress, as indicated by a beta coefficient of 0.14 (95% CI [0.06–0.22], p = 0.001). This suggests that early socioeconomic disadvantage is associated with increased levels of psychological distress in adulthood for males, but the impact is less pronounced than in females.

Male Interaction Results

There was no evidence of an interaction between the polygenic score and father's social class for males, with a beta coefficient of -0.02 (95% CI [-0.10 to 0.05], p = 0.542). This finding implies that, as with females, the effect of genetic liability on psychological distress does not differ based on socioeconomic background among males.

A z-statistic was computed to compare the interaction coefficients from the female and male models. The estimated difference in interaction coefficients between sexes was 0.007 (SE = 0.04). The z-statistic for this difference was 0.17 (p= 0.866). This indicates no evidence of an overall difference in the effect of polygenic score on psychological distress between the sexes.

Table 23: Results from the sex-stratified multilevel mixed effects model interacting mean-centred polygenic score at the 0.005 p value threshold and father's social class for females (N=4,300) and males (N=4,623)

Females (N=4,300)	Beta Coefficient	Bootstrapped 95% CI	p value
Non-manual is the reference category			
Manual	**0.30**	**0.18-0.41**	**0.001**
PGS	**0.27**	**0.22-0.33**	**0.001**
PGS*Manual	0.03	-0.09-0.14	0.668
Males (N=4,623)	Beta Coefficient	Bootstrapped 95% CI	p value
Non-manual is the reference category			
Manual	**0.14**	**0.06-0.22**	**0.001**
PGS	**0.22**	**0.17-0.27**	**0.001**
PGS*Manual	-0.02	-0.10-0.05	0.542

6.3.2.3 Research Question 4: Using meta-analysis, is there robust evidence of an interaction between polygenic score for psychological distress and socioeconomic circumstances on adulthood psychological distress outcomes?

Random-effects and fixed-effects meta-analyses were conducted to evaluate the interaction effects between polygenic scores and socioeconomic status on psychological distress (Figures 20 & 21). A study by Kosciusko et al. (2023), which utilised the English Longitudinal Study of Ageing dataset and identified SES as years of schooling, was excluded from the meta-analysis due to effect size and standard errors being zero. Each analysis included different combinations of studies utilising data from the two birth cohorts of 1958c and 1970 and UK Biobank (Table 24).

Meta-Analysis (a): Study 3 (1958c)

This meta-analysis included results from Study 3 (1958c and 1970c) and Qi et al. (2024, UK Biobank, using the Index of Multiple Deprivation). In the fixed effect model, the pooled interaction effect was 0.03 (95% CI [0.02, 0.04], p = 0.001); the Qi et al. study contributed the largest weight (99.9%) due to its large sample size. Similarly, in the random effect model, the pooled interaction effect was 0.02 (95% CI [-0.02, 0.05], p = 0.289), with low heterogeneity, meaning 12.5% of the variation in the effect sizes is attributable to differences between studies (I^2 = 12.5%, τ^2 = 0.0004, p = 0.3189). The Qi et al. study contributed the largest weight (75.7%) due to its large sample size.

Meta-Analysis (b): Keers 2017 (Childhood SES Composite)

As samples need to be independent, this meta-analysis replaced the Study 3 (1958c) model with Keers et al. (2017) results to test what a different study using the same sample would contribute to the analysis. They used a composite SES indicator limited to childhood. In the fixed effect model, the pooled interaction effect remained at 0.03 (95% CI [0.03, 0.03], p = 0.001), with no heterogeneity detected, meaning the effect sizes across the included studies are consistent, and any variation in the effect sizes is likely due to random sampling error rather than true differences between studies (I^2 = 0%, τ^2 < 0.0001, p = 0.5025). The Qi et al. study again contributed the majority of the weight (99.6%). In the random effect model, the pooled interaction effect remained at 0.03 (95% CI [0.02, 0.04], p = 0.001), with no heterogeneity detected, meaning the effect sizes across the included studies are consistent, and any variation in the effect sizes is likely due to random sampling error rather than true differences between studies (I^2 = 0%, τ^2 < 0.0001, p = 0.5025). The Qi et al. study again contributed the majority of the weight (89.4%).

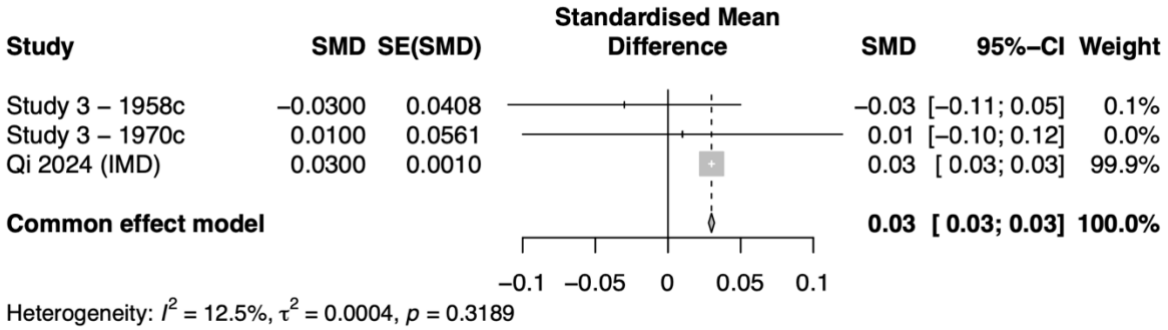
Meta-Analysis (c): Keers 2017 (Childhood and Adulthood SES Composite)

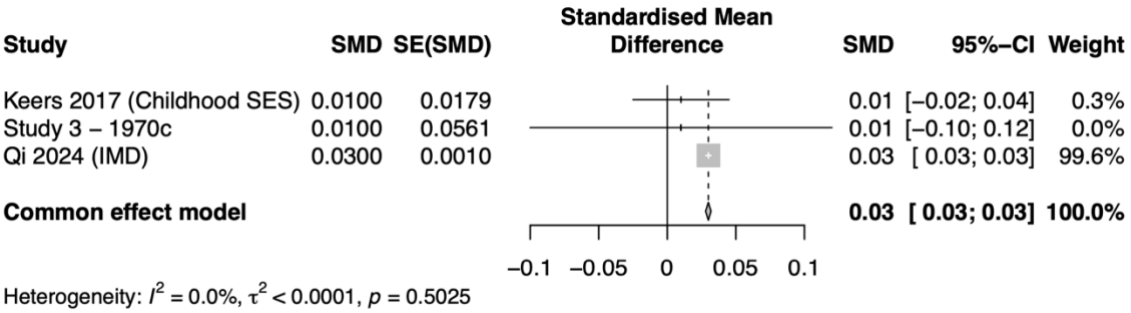
To test whether the way the exposure mattered to meta-analysing the findings, this version of the meta-analysis replaced Keers et al.'s childhood-only SES composite with a composite including both childhood and adulthood SES. In the random effect model, the pooled interaction effect was again 0.03 (95% CI [0.03, 0.03], p = 0.001), with no heterogeneity (I^2 = 0%, τ^2 = 0, p = 0.9385). Similar to previous analyses, the Qi et al. study contributed the largest weight (93.3%). In the fixed effect model, the pooled interaction effect was the same as the random effect model.

Table 24: Summary of potential studies included for meta-analysis with key data extracted

Study	Data	SES variable	Sample Size	PGS*SES Effect Size	95% CI	P value
Study 3	1958c	Father's social Class	5,242	-0.03	-0.10-0.06	0.519
Keers 2017	1958c	Composite of SES indicators across childhood	7,075	0.01	-0.03-0.04	0.691
Keers 2017	1958c	Composite of SES indicators across childhood and adulthood	7,075	0.03	0.01-0.06	0.001
Study 3	1970c	Father's social Class	3,681	0.01	-0.10-0.12	0.848
Qi 2024	Biobank	Index of Multiple Deprivation (IMD)	74,425	0.03	0.023-0.027	0.001

NB: A positive effect size means that the effect of a higher PGS (each standard deviation increase) on psychological distress is stronger amongst those with more disadvantaged SES circumstances, compared to those from more advantaged background.





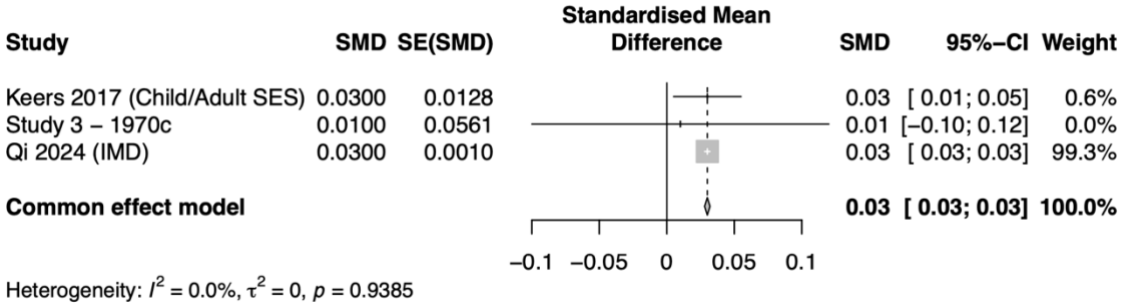
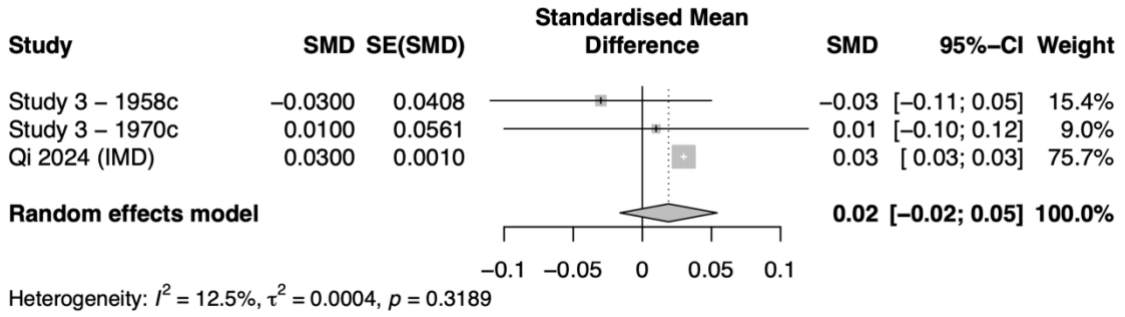
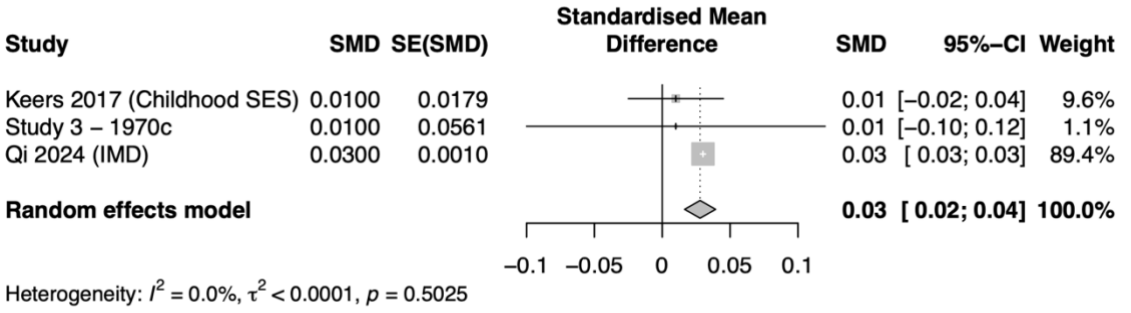


Figure 20: a) Forest plot of fixed-effect meta-analysis of GxE studies with 1958c from Study 3, b) with 1958c from Keers 2017 childhood SES composite, c) with 1958c from Keers 2017 adulthood and childhood SES composite

Effect sizes represent the difference in the association between polygenic scores and psychological distress when comparing disadvantaged versus advantaged socioeconomic circumstances, with positive effect sizes indicating a stronger genetic effect in disadvantaged environments.





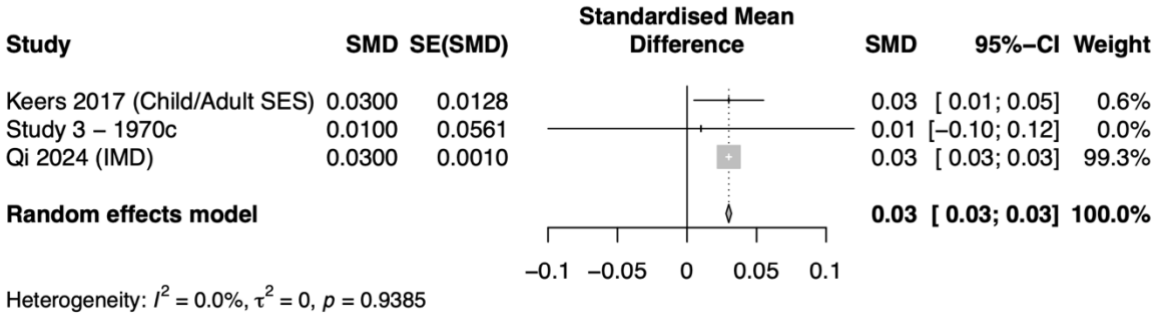


Figure 21: a) Forest plot of random-effect meta-analysis of GxE studies with 1958c from Study 3, b) with 1958c from Keers 2017 childhood SES composite, c) with 1958c from Keers 2017 adulthood and childhood SES composite.

Effect sizes represent the difference in the association between polygenic scores and psychological distress when comparing disadvantaged versus advantaged socioeconomic circumstances, with positive effect sizes indicating a stronger genetic effect in disadvantaged environments.

6.3.2.4 Supplemental and Sensitivity Analyses

Testing polygenic score for psychological distress at different p-value thresholds:

There was no evidence of an interaction term between the polygenic score and father's social class (PGS*Manual) across all thresholds, suggesting that the lack of a multiplicative interaction between genetic liability and socioeconomic background is robust to changes in the p-value threshold.

The sensitivity analysis was conducted to evaluate whether the interaction estimates between polygenic scores and father's social class changed when using polygenic scores with different p-value thresholds for SNP inclusion (Table 25). Specifically, models were re-run using the most conservative p-value threshold of genome-wide significance (5x10^-8) and the least conservative threshold (p = 1), and the results are compared to those obtained using the threshold selected for the main analysis (0.005).

For the conservative threshold ($p=5x10^{-8}$), the beta coefficient for the polygenic score at the conservative threshold was 0.11 (95% CI: 0.14–0.27). At this highly stringent threshold, the effect size of the polygenic score is reduced, indicating that only the most statistically significant SNPs were included, which may exclude many SNPs with smaller effects that contribute to the overall genetic liability for psychological distress.

At the less conservative threshold (p=1), the beta coefficient for the polygenic score at the least conservative threshold was 0.25 (95% CI: 0.21–0.28). At this liberal threshold, the effect size of the polygenic score is higher, reflecting the inclusion of a broader set of SNPs, including those with weaker associations. This threshold captures more of the genetic variance but may also introduce noise due to including SNPs with less robust associations.

The beta coefficient for the polygenic score is lower at the conservative threshold (0.11) compared to the least conservative threshold (0.25). This suggests that using a stringent threshold captures only the strongest genetic signals, resulting in a smaller effect size. In contrast, a liberal threshold captures a wider array of genetic contributions, leading to a larger effect size. Despite the differences in effect size, the direction of the association between polygenic score and psychological distress remains consistent across thresholds, with polygenic score positively associated with distress. However, the magnitude of this association varied, highlighting the importance of threshold selection in polygenic score analysis.

Research Question 1 Main Analysis:

Table 25: Results from the multilevel mixed effects model interacting mean-centred polygenic score at the $5x10^{-8}$ and 1 P-value thresholds and 2-category father's social class for 1958c and 1970c (N=8,976)

P-value threshold = $5x10^{-8}$			
	Beta Coefficient	Bootstrap 95% CI	P value
Non-manual is the reference category			
Manual	**0.21**	**0.07–0.14**	**0.001**
PGS	**0.11**	**0.14–0.27**	**0.001**
PGS*Manual	-0.02	-0.08–0.05	0.614
P-value threshold = 1			
	Beta Coefficient	Bootstrap 95% CI	P value
Non-manual is the reference category			
Manual	**0.20**	**0.13–0.27**	**0.001**
PGS	**0.25**	**0.21-0.28**	**0.001**
PGS*Manual	-0.02	-0.09-0.05	0.564

Research Question 2: Testing the 6-category father's social class

The same test was run for the supplementary analysis of the 6-category father's social class (Table 26). The finding of an interaction in the IV Skilled manual category remained the same as the main analysis at the liberal threshold. Whereas the results changed at the conservative threshold. Models were re-run using the most conservative p-value threshold of genome-wide significance ($p=5x10^{-8}$),) and the least conservative threshold (p = 1), and the results are compared to those obtained using the threshold selected for the main analysis (p=0.005).

For the conservative threshold ($p=5x10^{-8}$), the beta coefficient for the polygenic score was 0.04 (95% CI [0.12–0.23], p = 0.001). At this highly stringent threshold, the effect size of the polygenic score is notably reduced. This conservative approach may exclude many SNPs with smaller effect sizes that collectively contribute to the overall genetic liability for psychological distress. The main effects of father's social class remained consistent, with psychological distress increasing as social class declined. Notably, the largest main effects were observed for VI Unskilled (β = 0.42, 95% CI [0.33–0.48], p = 0.001), supporting a clear social gradient. However, evidence for interaction effects between polygenic score and social class was limited at this threshold, with only PGS × II Managerial (β = 0.07, 95% CI [0.02–0.13], p = 0.005) and PGS × V Partially Skilled (β = 0.07, 95% CI [0.01–0.13], p = 0.016) showing statistical significance. These results suggest that the moderating influence of social class on genetic risk may be minimal under this threshold and different to the category identified in the main analysis.

At the less conservative threshold (p=1), the beta coefficient for the polygenic score was 0.17 (95% CI [0.12–0.23], p = 0.001). At this liberal threshold, the effect size of the polygenic score is higher, reflecting the inclusion of a broader set of SNPs, including those with weaker associations. This approach captures more of the genetic variance but may also introduce noise due to the inclusion of less robust SNPs. The main effects of father's social class remained robust, with VI Unskilled (β = 0.41, 95% CI [0.33–0.48], p = 0.001) again exhibiting the largest effect sizes. PGS × IV Skilled Manual was the only category showing a statistically significant interaction (β = 0.10, 95% CI [0.04–0.16], p = 0.001), replicating the main analysis results.

Table 26: Results from the multilevel mixed effects model interacting mean-centred polygenic score at the 1 and $5x10^{-8}$ p value thresholds and 6-category father's social class for 1958c and 1970c (N=8,923)

P value threshold = 1	Beta Coefficient	Bootstrap 95% CI	P value
Model 3: Father's Social Class x PGS			
I Professional is the reference category			
II Managerial	**0.14**	**0.08–0.19**	**0.001**
III Skilled non-manual	**0.06**	**0.01-0.12**	**0.001**
IV Skilled manual	**0.26**	**0.21-0.32**	**0.001**
V Partially Skilled	**0.27**	**0.21-0.33**	**0.001**
VI Unskilled	**0.41**	**0.33-0.48**	**0.001**
PGS	**0.17**	**0.12–0.23**	**0.001**
PGS*II Managerial	0.05	-0.03-0.09	0.117
PGS*III Skilled non-manual	0.04	-0.03-0.10	0.201
PGS*IV Skilled manual	**0.10**	**0.04-0.16**	**0.001**
PGS*V Partially Skilled	0.05	-0.02-0.13	0.139
PGS*VI Unskilled	-0.02	-0.12-0.04	0.634
P value threshold = $5x10^{-8}$	Beta Coefficient	Bootstrap 95% CI	P value
Model 3: Father's Social Class x PGS			
I Professional is the reference category			
II Managerial	**0.13**	**0.08–0.19**	**0.001**
III Skilled non-manual	**0.07**	**0.01-0.12**	**0.001**
IV Skilled manual	**0.28**	**0.21-0.32**	**0.001**
V Partially Skilled	**0.29**	**0.21-0.33**	**0.001**
VI Unskilled	**0.42**	**0.33-0.48**	**0.001**
PGS	**0.04**	**0.12–0.23**	**0.001**
PGS*II Managerial	**0.07**	**0.02-013**	**0.005**
PGS*III Skilled non-manual	0.01	-0.05-0.07	0.702
PGS*IV Skilled manual	0.05	-0.01-0.10	0.055
PGS*V Partially Skilled	**0.07**	**0.01-0.13**	**0.016**
PGS*VI Unskilled	0.04	-0.03-0.12	0.234

6.4 Discussion

6.4.1 Summary of Findings

This study investigated the interplay between polygenic scores for psychological distress and father's social class in shaping psychological distress outcomes across adulthood. The findings revealed consistent independent effects of both polygenic scores and fathers' social class on psychological distress, with no evidence for gene-environment interactions detected in the primary analyses. Therefore, the associations of fathers' social class and polygenic scores with psychological distress were relatively stable regardless of the cohort or sex and followed a social gradient. While initial analyses using father's social class showed no clear gene-environment interaction, the meta-analysis incorporating multiple studies and SES measures found a small interaction effect (0.03, 95% CI [0.02, 0.04]). Although heterogeneity was minimal (low I^2 and τ^2 values), these results should be interpreted cautiously given the small number of studies included, as heterogeneity metrics like I^2 can be unreliable with few studies (370). Therefore, this suggests that socioeconomic disadvantage may slightly amplify genetic liability to psychological distress when larger sample sizes are tested and synthesised.

6.4.2 Comparison to the literature and explanation of findings

The findings from the current study contribute to the evidence of gene-environment research, with improved methods of using a polygenic score rather than a candidate gene design. Only one prior study tested for synergistic interactions with polygenic scores and childhood environment and their effect on psychological distress (325).

The consistent finding that polygenic scores and father's social class are independent predictors of distress confirms what is suggested from prior literature (Table 16*)*. The independent associations observed between polygenic scores and psychological distress are consistent with prior research demonstrating the polygenic nature of a complex phenotype (253,254,340–342,349). Similarly, the social gradient in psychological distress aligns with the well-documented relationship between socioeconomic disadvantage and poorer mental health (5). However, the absence of evidence for gene-environment interactions in the primary analyses contrasts with some studies that have reported synergistic effects.

The lack of interaction observed between the binarised father's social class and polygenic scores for distress was consistent with the work of Keers et al. (2017), which showed no evidence of an interaction between childhood socioeconomic circumstances and polygenic scores in 1958c. Their study was limited by using a polygenic score composed of 8 SNPs rather than including the many SNPs with small effects contributing to the complex distress trait. The present analysis had a more predictive polygenic score, demonstrating that both polygenic scores and father's social class, are independent predictors of psychological distress across adulthood. Keers et al. (2017) did not identify independent associations, nor did they identify evidence for an interaction between polygenic score and childhood socioeconomic circumstances.

Multiple explanations exist for why no evidence of an interaction was found in the 1958c and 1970c. Genetic liability and early-life socioeconomic environments each play a role in shaping psychological distress outcomes. However, similar to the findings of key studies in Table 16, the lack of a synergistic interaction between polygenic scores and socioeconomic status suggests that these factors may contribute additively rather than multiplicatively (253,254). This reinforces the idea proposed by previous studies that while socioeconomic disadvantage and genetic risk both increase the likelihood of psychological distress; their combined effect may not exponentially increase distress levels.

It should be noted that societal conditions can change over time, and their effects may be cumulative or dependent on critical periods in life (e.g., childhood adversity plus adulthood disadvantage). Capturing these dynamics in a study is difficult. The first analysis exposure variable was binarised father's social class, which may not capture the full extent of the environmental childhood conditions with enough nuance. Therefore, another analysis was run to test if there was a social gradient.

A social gradient in distress was found, whereby the independent associations between those with the least advantage in society had higher predicted psychological distress scores. The findings corroborate the systematic meta-review, identifying that the social gradient is consistent across all domains (324). There was tentative evidence of an interaction for the IV Skilled manual category, which aligns with prior findings (325,340). However, these prior findings did not focus solely on childhood or parental socioeconomic circumstances, and this result should be interpreted with caution.

In study 3, the differential hypothesis is being tested to see if father's social class moderates genetic effects. The finding that genetic influences are strongest among children of skilled non-manual fathers may be because, as the environment is more advantaged and homogenous for this category, genetics has a stronger signal for those higher up the social strata. Yet, this pattern does not fully confirm the differential susceptibility hypothesis, as there is no evidence of an interaction between the polygenic score and the top two most advantaged categories. do not propose that genetic variants conferring environmental sensitivity manifest full effects only in supportive contexts where individuals can select environments matching their genetic predispositions.

However, sensitivity analyses revealed that interaction effects for the Skilled manual category fully attenuated at permissive p-value thresholds, while becoming significant in the Partially Skilled category. The lack of interaction in the risk score analyses contrasts with 6-category models, suggesting that these findings may represent statistical artefacts due to multiple comparisons between different subgroups rather than robust gene-environment interactions. These inconsistent findings were part of the motivation for conducting the meta-analysis.

To contextualise, these studies tested the index of multiple deprivation and a composite of lifecourse socioeconomic status rather than solely focusing on childhood indicators as the current study did. There is also conflicting evidence whereby the same study that had evidence of an interaction between the lifecourse composite of socioeconomic circumstances, did not find evidence of an interaction with solely

childhood socioeconomic circumstance. The current study's sensitivity analyses using stricter and more permissive p-value thresholds for the polygenic score showed that the interaction of the IV Skilled manual category fully attenuated at the most permissive threshold of 1 and became significant in the V Partially Skilled category. The effect change suggests that the interaction findings may not be robust. Furthermore, the lack of interaction in the ridit score analysis contrasts with the 6-category model and underscores the need for caution.

The lack of evidence for cohort differences in the interaction between polygenic scores and father's social class aligns with prior research, which similarly found no evidence of cohort-specific interactions between socioeconomic circumstances and polygenic scores (253). However, this finding contrasts with Machlitt-Northen et al. (2023), who reported cohort-specific differences in the relationship between polygenic scores and socioeconomic factors (326). This discrepancy may stem from methodological differences, as Machlitt-Northen et al. (2023) utilised datasets and measures that were not directly comparable as those in the current study. For instance, differences in the operationalisation of socioeconomic status, the cohorts' composition, and the studied populations' genetic architecture could explain the divergent findings. As such, the observed differences may reflect variations in measurement or study design rather than true cohort-specific effects. This highlights the current study's contribution to research in employing harmonised measures and comparable samples to better disentangle the role of cohort differences in gene-environment interactions.

The current study's lack of cohort interactions might be due to the use of a GWAS that relied on a phenotype of onset instead of disease progression over time (371). A GWAS of onset captures genetic variants associated with the initiation of psychological distress but may not account for genetic factors influencing the progression or chronicity of distress over time. If socioeconomic circumstances play a more prominent role in shaping the trajectory of distress rather than its onset, the interaction may not manifest (371). As there is evidence of higher mean levels of distress for the 1970c, this may not be attributable to genetic factors but rather driven by variation in the environment, such as societal conditions. Societal conditions are often interconnected and dynamic, making it difficult to isolate their specific effects or interactions with genetic factors.

The sex-specific analysis examined how genetic liability and socioeconomic background relate to psychological distress in males and females. However, there was no evidence of a sex difference when stratified, which was consistent with prior literature (317). This finding aligns with the results from Kosciuszko et al. (2023) and Lam et al. (2019), who also found that genetic and socioeconomic influences on psychological distress are stable across demographic groups, including sex (253,349). Prior research on sex differences in the common autosomal genetic architecture of distress phenotypes found that the effect sizes were small and unlikely to account for the observed sex-differentiated outcomes fully (317). While females in both cohorts showed higher average levels of psychological distress, this difference may be attributed to factors beyond those measured by the current study's socioeconomic circumstances.

All of these results should be interpreted with the possibility that polygenic scores can be confounded by assortative mating and population stratification (372). These

confounding factors can influence the observed associations between polygenic scores and socioeconomic outcomes, potentially biasing the results (224,372).

Population stratification refers to differences in allele frequencies between subgroups of a population due to shared ancestry or systematic cultural, geographic, or social divisions, which can confound genetic association studies. All analyses included principal components to help account for population stratification to the best of my ability.

Assortative mating is the non-random pattern of individuals choosing mates with similar traits, such as socioeconomic status or education, which can influence genetic similarity within families. As a result, genetic factors can confound socioeconomic status (373). Prior research revealed that genetic factors had the most substantial impact on education and the least on income. Regardless of the socioeconomic measure employed, genetic influences were consistently strongest amongst individuals from the most privileged socioeconomic backgrounds (374).

If genetic factors are more pronounced in privileged socioeconomic groups, this could amplify the observed genetic effects on outcomes like psychological distress within these groups, creating a misleading impression of stronger genetic influence (375). Similarly, if socioeconomic circumstances are influenced by genetic factors, it can act as both a mediator and confounder, complicating the interpretation of the independent effects of socioeconomic circumstances and polygenic scores (375). For these analyses, the apparent relationships between polygenic scores, socioeconomic gradients, and psychological distress may partly reflect underlying genetic influences on socioeconomic circumstances rather than direct effects of genetic liability on distress (375). Without careful adjustment for population stratification and accounting for these dynamics, the findings could overestimate or misattribute the role of genetic factors across different socioeconomic contexts (224,372,375).

Genetic confounding could occur when parental mental health impacts polygenic scores as the SNPs inherited come from the parents, it is difficult to separate individual versus parental influence on mental health via passive gene-environment correlation (87). There is known literature linking parental mental health and offspring mental health outcomes. Father's social class could be related to their mental health, for example, if they become economically inactive due to poor mental health. The contribution of indirect genetic effects to psychological distress phenotypes is known to shrink total genetic effects estimates by ~20% (61,376).

The polygenic score used in this analysis is intended to capture the total genetic contribution to psychological distress. However, it is an imperfect proxy as it does not differentiate between direct genetic effects and genetic nurture effects, the latter as parental genetic characteristics which indirectly influence offspring outcomes through environmental pathways (e.g., parenting behaviours or parental mental health). Future research may benefit from employing within-family designs, which could help mitigate these confounding influences and provide more precise estimates of gene-environment interactions (60,377). However, within-family designs are not without their own limitations, as finding datasets with sufficient power is a challenge.

Finally, a key reason for unobserved interactions across most of the research questions could be insufficient statistical power, further explored in limitations below. If the effect size were larger (i.e. a magnitude of 0.30 instead of 0.01), there would have been sufficient power, however small effect sizes were assumed. A study comprising a sample of ~74,000 participants from UK Biobank detected a multiplicative interaction (340).

6.4.3 Strengths & Limitations

This study's strengths include prospectively ascertained measures in both cohorts, which were chosen and constructed to be as comparable as possible. For example, the construction of the polygenic scores was standardised between the cohorts, as it was restricted to overlapping SNPs, and neither cohort was present in the discovery sample of the GWAS used. Furthermore, the other benefit is utilising a longitudinal design as the outcome variable. Psychological distress was collected at multiple ages in each cohort, using the same questionnaire for measurement at each time point. The selection of the father's social class, in lieu of parental income, or education, allowed for comparisons between the cohorts. This is because income was not measured accurately in the 1970c or at all in the 1958, neither do they have detailed wealth variables that are reliably comparable (187,188). Data collection spans from 1981 to 2016 across two generationally different cohorts (Baby Boomers versus Generation X), enabling cross-cohort comparison. Sensitivity analyses using more conservative and more liberal p-value thresholds for polygenic score showed that the estimates were largely comparable no matter the threshold specified.

However, this study was not without limitations. The first was the low statistical power to detect an interaction. Large statistical power is required to detect gene-environment interactions. These typically require a substantially larger sample size compared to detecting main effects. Estimates suggest that sample sizes may need to be between four and sixteen times larger to reliably identify interactions, with some suggesting a minimum of 300,000 participants if the variance explained by polygenic scores is less than 1% (378). A study using UK Biobank and ~74,000 participants detected an interaction between the index of material deprivation and polygenic scores (340). This highlights the issue of power in this and similar studies. Post-hoc power analysis found that each research question in the present study was well-powered to detect main effects but was underpowered to detect evidence of interaction (See Supplementary Table 12 & Supplementary Figure 5). The current study adds to the literature by both providing new evidence from two cohorts (1958c and 1970c) and synthesising existing evidence. Future studies may include the present study as part of a meta-analysis – as explored in research question 5.

Additionally, the present study may be subject to sampling and selection biases, particularly due to health survivorship and the tendency for more socioeconomically advantaged participants to remain in studies with lower attrition rates (67). This potential bias could impact the generalisability of the findings, as the sample may not fully represent the broader population. However maximum likelihood estimation as used as part of the multilevel modelling approach to try to mitigate this.

A further limitation of Study 3 is that childhood socioeconomic position was operationalised using father's occupational social class only, rather than employing

the dominance method (37). While this approach was consistent with historical conventions in the 1958 and 1970 cohorts and aided cross-cohort comparison, it may have introduced bias in two respects. First, children from households without a resident father, or with fathers not in paid employment, were underrepresented. Second, by not incorporating mother's social class data, our measure may underestimate family resources in households where mothers' occupations represented the higher social class or where participants were dropped for not having fathers ' social class information. Consequently, social gradients in psychological distress may be more conservatively estimated compared to studies adopting the dominance method.

Finally, although this study was the first to conduct a meta-analysis to test the relationship between socioeconomic circumstances and polygenic scores on psychological distress. It should be interpreted cautiously due to the following limitations. Qi et al. (2024) study from UK Biobank contributed disproportionately to the pooled effect size (e.g., up to 93.3% weight in one analysis) (340). This overrepresentation reduces the influence of smaller datasets and may bias the results toward findings from the UK Biobank, which has unique demographic and a highly selected socioeconomically advantaged sample (67).

Although the meta-analyses included multiple socioeconomic circumstance measures (e.g., father's social class, years of schooling, composite SES indicators), the measures were not uniformly applied across all datasets. Variations in the definition and operationalisation of socioeconomic circumstance could limit comparability. For instance, "years of schooling" in ELSA and the "Index of Multiple Deprivation" (IMD) in UK Biobank measure socioeconomic circumstance differently, potentially capturing distinct aspects of the construct (168). Similarly, some studies only capture adulthood socioeconomic circumstances rather than childhood. These factors led to testing the three versions of the meta-analysis with increasing restrictions on which studies were included based on how closely they resembled the current study's specifications.

6.4.4 Conclusion

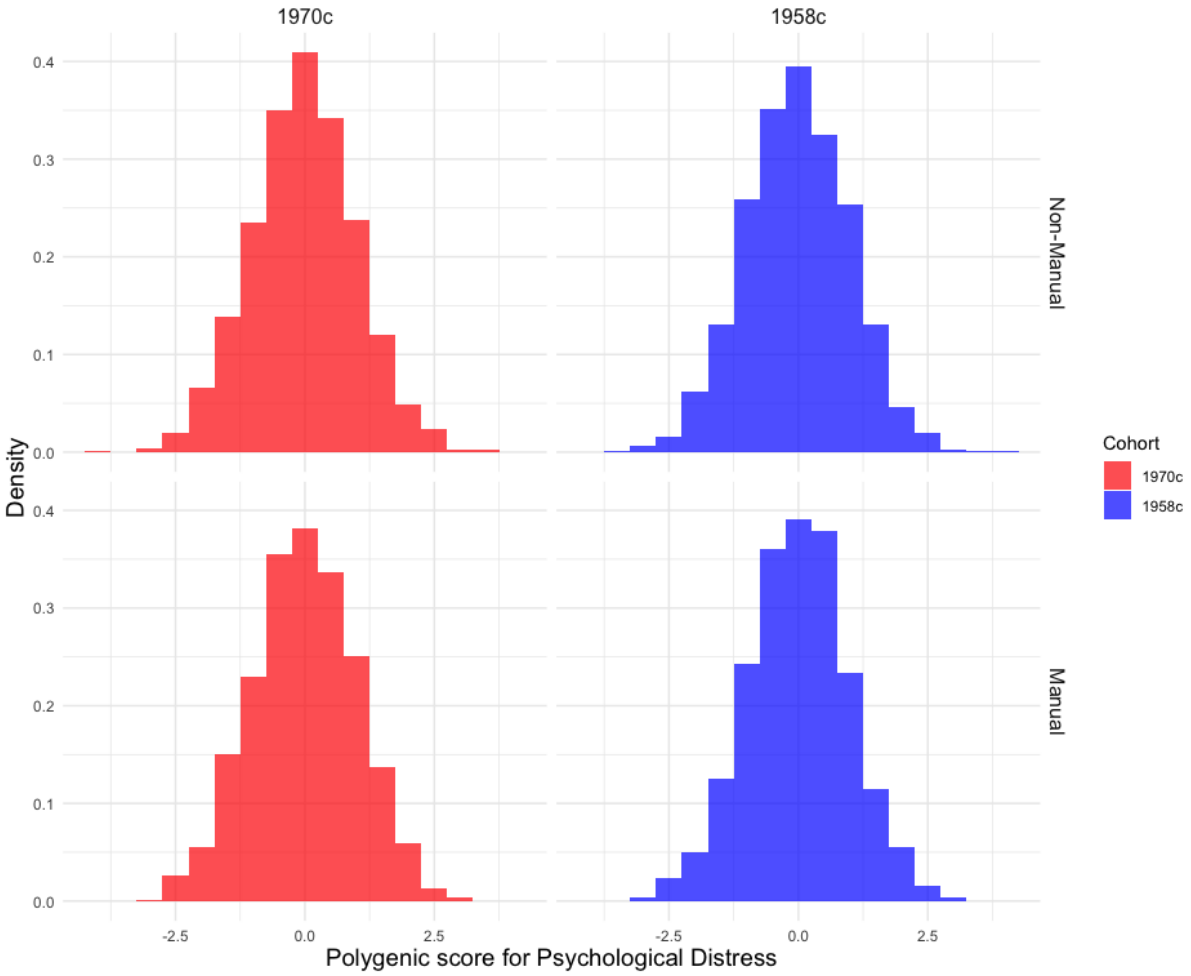
Overall, the associations of fathers' social class and polygenic scores with psychological distress are relatively stable, regardless of the cohort or sex. There was weak evidence for an interaction in one category of the father's social class. Yet the meta-analysis indicates some evidence of an interaction between PGS and socioeconomic circumstances.

The results from this study, taken collectively, suggest that other cohort- or sex-specific factors may influence the average distress levels. As without an interaction identified and with the low amount of variance explained by the polygenic score (~2%), the variation in distress scores that are higher in the 1970s and females must be more attributed to other environmental factors that were not tested within this study. These environmental factors may potentially be exclusive to each cohort or sex and may be both systematic (i.e. variables observed that have not been tested here) or stochastic contributions.

6.5 Study 3 Summary

- Manual father's social class and higher polygenic scores were both independently related to higher psychological distress scores, irrespective of cohort or sex.
- Few studies have examined the social gradient interaction with polygenic scores; this study found that distress outcomes worsen progressively across the Registrar General's six-class framework, with some evidence of interaction in the IV partially skilled category.
- While there was no evidence for an interaction with manual father's social class, there was tentative evidence for an interaction when pooling estimates from all available studies in a meta-analysis.

6.6 Supplementary Material



Supplementary Figure 4: Histograms of polygenic score density by father's social class and by cohort in the 1958c and 1970c

Supplementary Table 10: Mean and standard deviation of the Malaise Inventory Score of those with genetic data by father's social class in 1958c and 1970c

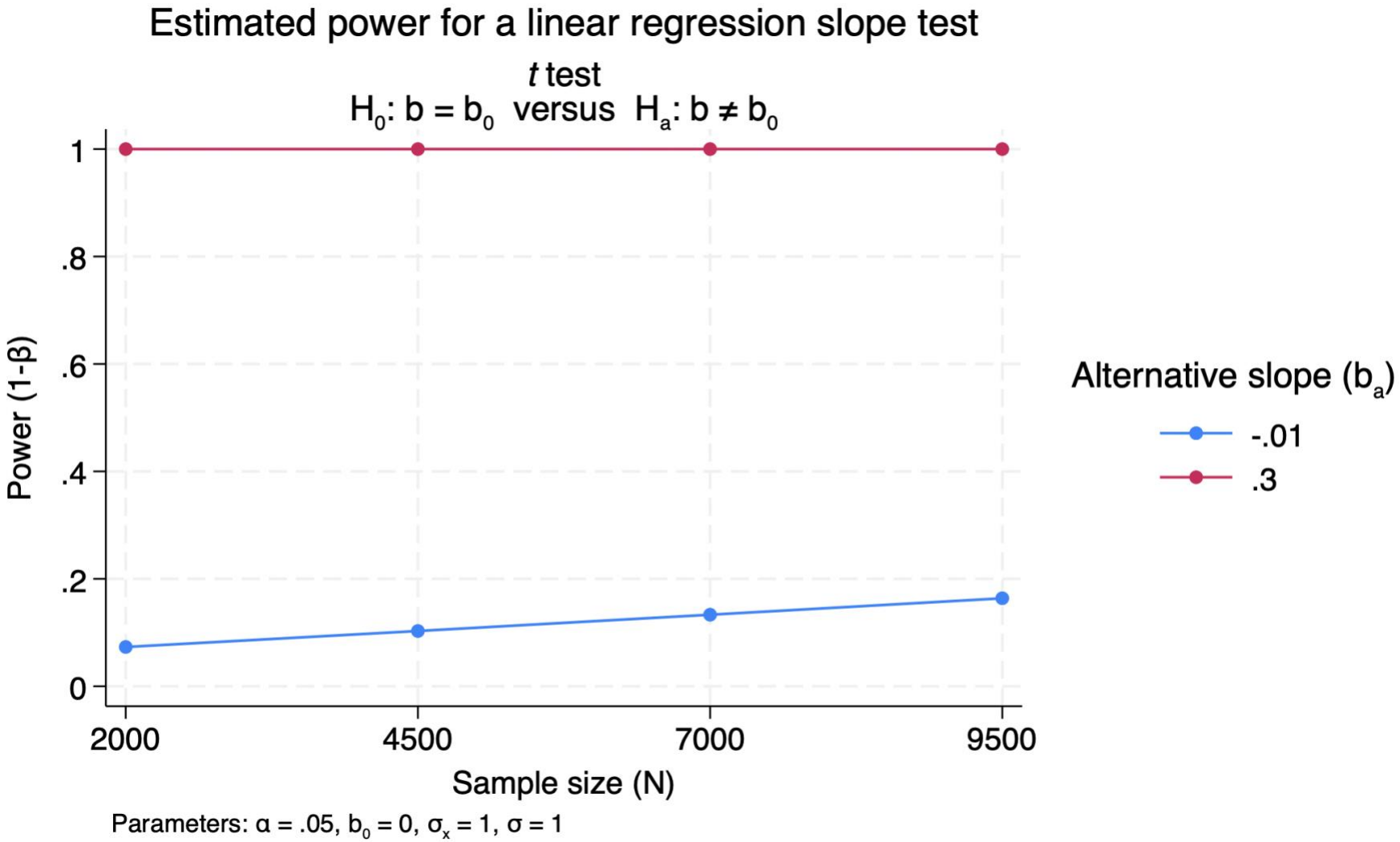
		Manual			Non-Manual		
	Age	Mean	SD	N	Mean	SD	N
1958c	23	1.19	1.51	2,947	1.02	1.39	1,745
	33	0.96	1.51	2,988	0.79	1.25	1,798
	42	1.49	1.75	3,257	1.47	1.68	1,930
	50	1.36	1.85	2,919	1.29	1.75	1,805
1970c	26	1.64	1.68	1,557	1.51	1.57	1,405
	30	1.41	1.62	1,927	1.34	1.58	1,615
	34	1.46	1.70	1,827	1.44	1.73	1,564
	42	1.49	1.75	1,820	1.47	1.68	1,571
	50	1.68	2.06	1,993	1.51	1.94	1,700

Supplementary Table 11: Mean and standard deviation of the Malaise Inventory Score of those with genetic data by 6-category father's social class in 1958c and 1970c

		I-Professional			II-Managerial			III-Skilled Non-manual			IV-Skilled Manual			V-Partially Skilled			VI-Unskilled		
	Age	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
1958c	23	0.84	1.21	243	0.98	1.43	976	0.98	1.35	524	1.18	1.47	1967	1.25	1.56	648	1.40	1.69	341
	33	0.62	1.05	248	0.78	1.28	1009	0.85	1.31	536	0.98	1.57	2005	0.98	1.49	655	1.16	1.63	343
	42	1.32	1.57	266	1.40	1.63	1087	1.32	1.53	574	1.51	1.78	2182	1.53	1.81	708	1.68	1.91	380
	50	1.25	1.67	256	1.36	1.8	1012	1.29	1.80	533	1.46	1.96	1954	1.49	1.92	636	1.60	1.97	341
1970c	26	1.35	1.51	203	1.53	1.62	850	1.52	1.68	352	1.72	1.67	1082	1.83	1.72	325	1.65	1.52	150
	30	1.21	1.46	236	1.40	1.62	988	1.24	1.50	391	1.45	1.66	1342	1.54	1.72	401	1.80	1.78	184
	34	1.38	1.70	237	1.48	1.71	941	1.32	1.66	386	1.58	1.84	1277	1.56	1.80	380	1.56	1.78	170
	42	1.64	1.82	225	1.64	1.78	960	1.61	1.8	386	1.75	1.90	1256	1.85	2.06	386	1.97	1.97	178
	50	1.48	1.79	244	1.66	1.99	1044	1.44	1.95	412	1.73	2.13	1376	1.86	2.21	419	1.65	1.93	198

Supplementary Table 12: Post-hoc power calculations research questions 1 and 2 from study 3

	RQ1	RQ2 Cohort		RQ2 Sex	
	Binary	1958c	1970c	Males	Females
Sample Size	8,923	5,242	3,681	4,623	4,300
Main Effect Size of predictor 1 (Polygenic)	0.25	0.24	0.25	0.22	0.27
Main effect size of predictor 2 (SES)	0.20	0.20	0.21	0.14	0.30
Interaction Effect Size	-0.01	-0.03	0.01	-0.02	0.03
Power Calculation for Main Effect	100%	100%	100%	100%	100%
Power Calculation for Interaction	15%	58%	9%	27%	82%



Supplementary Figure 5: Estimated power of main effect (0.30) and interaction effect (-0.01) across sample sizes of 2,000-9,500. Lines indicate the effect size (red = 0.03, blue = -0.01)

7 Discussion

This thesis investigated genetic and social contributions to psychological distress across adulthood using data from two British Birth Cohorts; to date, few studies have assessed the polygenic contribution of psychological distress across adulthood using representative samples. This thesis examined age-, cohort- and sex-related differences in genetic associations alongside potential gene-environment interactions (GxE) between a polygenic score for distress and socioeconomic position.

7.1 Summary of Thesis Findings

The empirical component of this thesis comprised three interconnected studies, each building upon one another to add to the understanding of how genetic liability for psychological distress operates across adulthood and across social contexts.

Study 1 examined age-related differences in the association between polygenic score for distress and phenotypic psychological distress from ages 23 to 50 in the 1958 cohort. This research adds to the recent literature of longitudinal studies examining the contribution of genetic factors to distress-related phenotypes by age. The findings revealed a positive association between polygenic scores and psychological distress that persisted across adulthood, with the strength of this association increasing slightly with age. The variance explained by polygenic scores increased slightly across age. This modest but increasing genetic contribution suggested that while genetic factors play a consistent role in psychological distress, their influence may become more pronounced with age, raising questions about whether this pattern would be observed in cohorts born more recently which have experience higher distress levels.

Building on these findings, Study 2 examined whether the polygenic score for psychological distress is associated with phenotypic psychological distress by age using the 1958 National Child and Development Study (1958c) from the previous chapter and the 1970 British Cohort Study (1970c). This study provides evidence that the association between polygenic scores for psychological distress and observed psychological distress remains persistent from early (~ age 23) to middle (~ age 50) adulthood in both cohorts. Each cohort shows an increasing association and greater explanatory power of genotypic effects with age. There was some evidence of a cohort difference at younger age points. However, the cohorts converged by late adulthood

Study 3 assessed whether father's social class interacts with polygenic scores in its association with adulthood psychological distress in the 1958c and 1970c. The study found the main effects of fathers' social class and polygenic scores with psychological distress are relatively stable, regardless of the cohort or sex. There was weak evidence for an interaction in the IV Skilled manual category of the father's social class. When meta-analysing the current results with previously published estimates from other samples, there was small evidence of a multiplicative interaction between polygenic scores and socioeconomic circumstance on psychological distress. The results from this study, taken collectively, suggest that other cohort- or gender-specific factors may influence the difference in average distress levels. This means there are other environmental factors, potentially exclusive to each cohort or gender, that influence

the difference in mean levels, whereby the 1970c and females have higher average distress levels.

7.2 Discussion of Findings

Across the three studies, different theoretical frameworks were applied to test the role of genetic and environmental influences on psychological distress. In Study 1, the Selection, Optimisation, and Compensation (SOC) model predicted that genetic effects would diminish with age as biological plasticity declined and individuals increasingly relied on compensatory social and cultural resources. However, the findings showed small increases in both polygenic heritability and penetrance with age, a pattern inconsistent with SOC. In Study 2, social control theory proposed that genetic effects would vary across cohorts due to historical differences in distal social environments. This expectation was partly supported: while polygenic penetrance remained stable, polygenic heritability was greater in the 1958 cohort than in the 1970 cohort, in line with the hypothesis that broader social contexts shape the expression of distress. Finally, Study 3 tested the differential susceptibility hypothesis, which suggested that genetic sensitivity to distress would be amplified under both advantaged and disadvantaged paternal social class conditions. The lack of strong evidence for a gene × father's social class interaction meant that the empirical study results did not support the differential susceptibility hypothesis.

Table 27: Summary of primary theoretical frameworks and hypothesis for each study

Study	Primary theoretical framework	Hypothesis	Study Finding
1	Baltes' Selection, Optimisation and Compensation (SOC) model	The effect of the polygenic score for psychological distress will diminish with age**,** reflecting age-related declines in biological plasticity and increasing reliance on compensatory social and cultural resources.	Both polygenic heritability and polygenic penetrance showed small increases with age. This finding does not fit with SOC model hypothesis.
2	Social Control Theory	Genetic effects on psychological distress will differ by cohort, reflecting historical differences in distal social environments. Specifically, variation in social norms and institutional structures across cohorts is expected to moderate the expression of genetic influences.	Results indicated that polygenic penetrance was similar across cohorts, while polygenic heritability was greater in the 1958c than the 1970c. The social control theory does fit with this study's findings.
3	Differential susceptibility hypothesis	Genetic effects on psychological distress will be moderated by paternal social class as a proximal environmental factor, with individuals carrying higher genetic scores showing greater sensitivity (for better or worse) to socio-economic conditions.	As there was a lack of strong evidence of an interaction, the findings from this study do not fit the differential susceptibility hypothesis.

This thesis' main contribution can be contextualised by comparison with prior literature. There have been secular increase in psychological distress in multiple countries over the past decades, including Denmark, the U.K. and U.S. (379) (102,124,123,268). In the U.K. context, the trend of increasing distress levels is higher in the more recent 2000c cohort (102). However, even in previous cohorts there is a trend of younger cohorts having increasingly higher mean psychological distress (124). As evidenced by Gondek et al. 2021, which utilises the same sample as this thesis, the British birth cohorts (124). This thesis reiterated these findings, with higher mean levels of distress in the 1970c at every age point compared to the 1958c. The premise of this thesis was to examine whether the environmental and genetic correlates of psychological distress had changing contributions by age, cohort and sex-stratified analyses.

Gene-by-cohort evidence has found conflicting contributions in different contexts (75,306). Two population-wide studies found more recent cohorts have higher levels of distress and that mental distress heritability estimates are increasing (306,379). Whereas in the USA context Conley et al. 2016 found no difference in the polygenic penetrance to depression between cohorts born between 1920-1950. A methodologically similar study using ~400,000 participants from the UK Biobank identified gene-by-cohort interactions on behavioural traits (albeit not mental health related) (380). These differences in gene-by-cohort patterns could be hypothesised to be due to societal conditions – for example, changes in resource availability, policy, and norms can alter whether genetic effects strengthen or weaken over time

Members of the 1958 cohort came of age in the 1970s and 1980s—a time marked by a robust post-war social safety net that gradually yielded to economic turbulence—while the 1970 cohort reached adulthood amid rising neoliberal policies and a worsening labour market in the late 1980s and 1990s (139,381). These differing socio-economic contexts likely influenced stress exposures. This is potentially due to changing environmental factors, i.e., a 'depressogenic' environment emerges which the 1970c experienced more of during entry to adulthood and their middle age (124,150,301,382).

The shift from an industrial to a service-based economy in the late 20th century, coupled with increased educational attainment and women's labour market participation, may have contributed to a more distressing environment, particularly for the 1970 cohort (301) (139). As absolute living standards increased, self-reported quality of life may have decreased, and inequalities increased for both cohorts, but more so within the 1970 cohort (296).

This thesis utilised the social context and differential susceptibility models to test the relationship between genes and environments. As examined in study 2, the social control model suggests that restrictive environments may suppress genetic differences by elevating distress across genotypes. In contrast the differential susceptibility model, investigated in study 3, proposes that environments can magnify genetic effects, allowing individuals with higher polygenic scores to experience more extreme outcomes both positive and negative. These empirically tested model help go some way to contextualise this thesis' results.

Study 3 was likely underpowered to detect a GxE in the datasets included in this thesis but found tentative evidence for an interaction (albeit small) when meta-analysing across multiple studies. This gives some evidence to support the differential susceptibility hypothesis. Study 2 found both the 1958c and 1970c show an increasing association and greater explanatory power of genotypic effects between ages 23-50, suggesting that the influence of genetic factors may become more pronounced with age. While there was some evidence of a cohort difference at younger age points, the cohorts converged by older adulthood. The increasing genetic influence by age in both cohorts indicate partial agreement with the social control model. It may be the environment was more restrictive at earlier age points suppressing genetic expression, while convergence at older age shows social control diminished.

Overall, this thesis adds to the evidence that genetic correlates of psychological distress slightly increase by age and are stable across cohort in the UK context.

7.3 Discussion of Broader Themes

In the following sections, three key themes are drawn from this thesis' findings and discussed in relation to advancing knowledge in epidemiology. These are: (1) time as the third dimension of variability, (2) widening how we conceptualise the environment and (3) the interdependence of nature and nurture.

7.3.1 Theme 1: Time, the third dimension of variability in phenotypes

Across the thesis, the role time plays in psychological distress was explored in each study. In both cohorts, there was an increase in the explanatory power of the polygenic score as participants aged. Historically, phenotypes have been assessed by isolating the proportional contribution of genetics and the environment i.e. the nature versus nurture debate (77,83,84). This thesis contributes to the literature that time is necessary for a phenotype's trajectory, and this process contributes a third dimension of variability.

As outlined in the introduction (Section 1), theoretical shifts in genetics and developmental biology have emphasised the role time plays in the coalescence of genes and environments. Early biometric models, such as Fisher's additive genetic framework, conceptualised genetic and environmental contributions as discrete and independent sources of variance (77). However, Hogben (1933) introduced the "norm of reaction", highlighting how individuals with identical genotypes can exhibit different phenotypic outcomes depending on environmental conditions (86). Building on this, Feldman & Lewontin (1975) argued that phenotypes must be understood as dynamic developmental processes rather than fixed genetic potentials (383). These perspectives necessitated a reframing of the gene + environment equation to include time as the third portion of variability.

Boyce et al. (2020) argue that time operates at multiple biological and social scales, from epigenetic modifications occurring within hours to developmental transitions unfolding over decades (90). Twin and molecular research had predominantly found

that the genetic contribution to psychological distress is stable as people age (107–111,239,252,384). However, one study utilising the 1958c found increasing genetic influences on psychological distress between ages 23 and 33 and then stability between ages 33 and 42 (257). When they included data from childhood and adolescence into adulthood there was a decrease between age 7 to 11 and then steady increase in the correlation between genetic factors and depression phenotype from 11 up to 33.

Individual variability in psychological distress reflects not only genetic predisposition and environmental exposure, but also dynamic developmental processes, including epigenetic modifications that alter gene expression in response to environmental inputs and stochastic neurodevelopmental events that create individual differences independent of genetic sequence variation (54). The modest predictive power of polygenic scores in this thesis may reflect these unmeasured biological processes, though alternative explanations, including rare genetic variants, gene-gene interactions, and measurement error, cannot be excluded. These unmeasured developmental processes mean that polygenic scores may represent only the static genetic foundation upon which dynamic epigenetic and stochastic mechanisms operate. Future research may benefit from incorporating longitudinal epigenetic profiling alongside genetic analysis to capture the full biological architecture of psychological distress across development, and that current modest effect sizes may reflect fundamental limitations of static genetic approaches rather than weak genetic influences per se.

Understanding genetic influences on psychological distress requires moving beyond static models of heritability to a framework that recognises development as a dynamic, bidirectional process. Therefore, this thesis contributes original research, as it utilised longitudinal data to try and understand the role time plays in the psychological distress phenotype. Genes are not expressed in isolation; rather, they shape behaviour while simultaneously being upregulated by behaviour and experience (epigenetics) (83). This thesis has demonstrated that the correlation between genetic factors and psychological distress became more pronounced by age, a pattern that aligns with developmental systems theory, as proposed by Gottlieb and Waddington (386–389). Their work suggests that phenotypic outcomes arise from an ongoing interaction between genetic propensity and environmental exposures, with Bronfenbrenner's ecological systems theory further highlighting how these interactions unfold across multiple levels of social and environmental context throughout the lifecourse (97,112,390).

As individuals develop, they become increasingly embedded within stable patterns of behaviour, social environments, and self-reinforcing experiences that structure the expression of their genetic propensity (87). This may explain why heritability estimates for psychological traits increase with age — because individuals become more consistent in how they experience and respond to the world, reinforcing gene-environment correlations (87). This perspective challenges deterministic interpretations of genetic propensity, emphasising that genotypes are expressed within specific environmental and developmental contexts.

7.3.2 Theme 2: Widening how we conceptualise the environment

The role of the environment in shaping the psychological distress phenotype has been central to this thesis. Across the thesis, the conceptualisation of the environment has encompassed cohort membership and father's social class. The latter study was the first to utilise these datasets to investigate cohort differences in whether early-life socioeconomic position moderates the effect of polygenic liability on distress, testing both independent and synergistic contributions.

Traditional gene-environment interaction research primarily focused on acute stressors, such as major life events or short-term exposures. While acute stress can trigger distress responses, lifecourse epidemiology highlights the cumulative and chronic nature of disadvantage (3). As per the rationale in the introduction of the thesis, the current thesis focused on the conceptualisation of chronic stressors. Chronic stressors, including prolonged socioeconomic adversity, structural inequalities, and sustained social exclusion, may better explain variations in distress over time (10,391). We conceptualised cohort and father's social class—a measure of structural socioeconomic positioning during childhood—within this framework.

To better understand the role the environment can play in phenotypic changes, we can look to the Flynn Effect (392). The Flynn Effect identifies rising heritability estimates of cognitive traits over the 20th century (392). Yet, this effect cannot be attributed to genetic change, given the insufficient time for evolutionary processes to account for such shifts (392). Therefore, fluctuations in heritability may reflect broader environmental change or gene-environment interplay rather than genetic evolution.

Beyond individual-level exposures, this thesis highlights the broader socioeconomic landscape that shapes distress across generations, predominantly in Study 2. Cohort differences were tested based on the observation that the 1970 cohort had higher average distress scores than the 1958 cohort, yet participants were born 12 years apart. Therefore, if genotypic-phenotypic associations remain similar across the two cohorts, as the findings show, the environment contributed more to the difference in observed psychological distress levels in younger cohorts. This means the environment may be operating in a "depressogenic" way, resulting in a "depressogenic environment."

The conceptualisation of the depressogenic environment extends beyond immediate economic conditions to consider the structural, cultural, and historical contexts that shape risk.

The results across empirical chapters consistently showed that females reported higher average levels of psychological distress compared with males, and that polygenic scores exhibited differences in predictive utility between the sexes. Specifically, polygenic scores explained slightly more variance in distress among women than men, suggesting that the contribution of common genetic variants is stronger in females. In the introduction, it was hypothesised that socially constructed phenomena—such as the accumulation of disadvantage and the gendered distribution of structural strain—are key drivers of this gap. The findings lend some support to this

interpretation: the higher distress levels observed among women may not solely reflect intrinsic biological differences but rather the interaction of genomic vulnerability with cumulative exposure to inequality, discrimination, and structural stressors. This interpretation aligns with lifecourse epidemiological theory, which emphasises how social position and disadvantage accumulate and manifest in health outcomes over time. Taken together, the findings suggest that the structural and cultural position of women in society may elevate their risk burden for developing and maintaining higher levels of psychological distress, and that the apparent genetic differences observed in sex-stratified analyses may reflect this interplay rather than a fixed biological distinction.

One major shift over the past century has been the transition from an income-based economy to an asset-based economy, altering the mechanisms through which advantage and disadvantage accumulate over time (135). This may mean that inherited social and economic capital may exert greater influence over life outcomes than individual effort or ability as wealth disparities increase (393,394). These structural changes may have reshaped economic security, social mobility, and potentially psychological distress outcomes (333,334).

An aim of this thesis was to examine how early-life socioeconomic factors, specifically father's social class, interact with genetic liability for psychological distress. The limited evidence for gene-environment interaction found in this thesis suggests that the effects of genetic propensity and socioeconomic disadvantage are potentially additive rather than multiplicative. This is in keeping with the results from prior research (253,254,325,341,342,349), though the largest study conducted using UK data (n = ~ 77,000) did identify an interaction (340).

The findings of this thesis highlight the need for a more nuanced conceptualisation of the environment, one that accounts for systematic socioeconomic influences, chronic rather than acute stressors, and the evolving structure of the depressogenic environment.

7.3.3 Theme 3: Interdependence of Nature and Nurture

Across all three studies, a central question emerges: how do nature and nurture interact in shaping psychological distress? This thesis has examined how age, cohort, and socioeconomic context are correlated with distress, demonstrating that polygenic scores account for a small but consistent proportion of variance in this trait. These findings align with Turkheimer's first law of behavioural genetics—that all human traits are heritable (40). While genetic correlations with phenotypes are present, they operate within specific social, historical, and developmental contexts, raising important questions about whether partitioning genetic and environmental contributions is a useful endeavour (77,86).

Gene-environment interactions (GxE) are well-documented in controlled agricultural and animal studies, their application to human behaviour remains methodologically complex (395–397). The non-random distribution of genes across environments means that we cannot assume genetic and environmental independence (224). Gene-

environment correlations arise because individuals select, modify, and are exposed to environments in ways that reflect their genetic propensity (88). This interdependence further complicates efforts to estimate purely environmental or purely genetic influences.

Within this thesis, the limited evidence for gene-environment interactions observed in study 3 may be reflective of these complexities. It may also be due to the study 3 being underpowered to detect a small effect (65,398). If distress is not solely the product of direct genetic effects, but rather the result of complex interaction between genetic propensity and environmental exposure, then standard GxE models may underestimate the true extent of interaction effects. However, study 2 went some way to mitigate this by utilising cohort birth year as an exogenous environmental exposure.

A more complex framework of understanding the interdependence of nature and nurture regarding mental health traits was outlined by Kendler (2005). He argued for a comprehensive biosocial paradigm in psychiatric genetics, advocating for an approach that integrates variance partitioning with mechanistic models of psychiatric risk (399). His vision aligns with the literature overall which suggests that genetic liability does not operate in isolation but is embedded within developmental and social contexts. Understanding distress requires a multilevel approach, incorporating genetic risk, environmental exposures, and lifecourse transitions into a single explanatory framework.

Going one step further, a more recent conceptualisation by Mitchell & Cheney (2024) provides a compelling alternative to the static blueprint model of genetic influence and introduces a generative model of the genome (400). Rather than viewing genes as what can be easily mistaken as a “deterministic” blueprint, they propose that the genome functions as a generative model, dynamically interacting with environmental inputs to shape developmental outcomes. This perspective helps explain why polygenic scores only capture a modest proportion of variance in distress—genetic influences operate within a self-organising, responsive biological system. This model aligns with emerging work in biosocial epidemiology, emphasising that genes alone do not determine outcomes but instead provide probabilistic constraints on development, contingent on environmental context.

7.4 Strengths & Limitations

This section provides an overview of this thesis's key strengths and limitations. While each study’s specific strengths and limitations have been discussed in its respective chapters, this section reflects on the broader methodological and conceptual issues.

7.4.1 Strengths

The strength of this thesis lies in it being the first study to use both the rich phenotypic data of the 1970c and 1958c alongside the genotypic data to interrogate genetic contributions to psychological distress phenotype longitudinally. These data enabled me to ask relatively novel research questions.

One of the main strengths of the data is its longitudinal nature, which allows for the examination of changes in genetic and environmental influences by age and cohort. This is particularly important in the study of psychological distress, where cross-sectional analyses may obscure developmental patterns and fail to capture the dynamic interplay between genes and environmental exposures (401,402). Incorporating repeated measures enables the investigation of how the correlation between genetic factors and distress unfold across different stages of life and how environmental factors might modify it (403).

Another key strength is the use of harmonisable measures in each cohort (and at multiple time points), ensuring greater comparability of the outcome (psychological distress) and key environmental exposures, such as socioeconomic position (404–407). This was a key limitation of prior studies in the area (307). Differences in measurement can introduce bias and limit the ability to make valid comparisons, particularly when examining the impact of early-life conditions on later-life mental health. Using consistent measures strengthens the robustness of findings and provides greater confidence in the observed associations, as differences are less likely to be driven by measurement differences, albeit not entirely removing measurement error.

A further strength is the harmonisation of the genetic exposure. This thesis employed consistent quality control, identified overlapping SNPs across cohorts, and standardized polygenic scores in a pooled dataset, ensuring valid cross-cohort, cross-age, and cross-sex comparisons. All of which helped to mitigate bias.

The thesis also benefits from a rigorous approach to handling missing data, a common challenge in longitudinal research. For the linear analyses in Studies 1 and 2, multiple imputation by chained equations (MICE) was implemented due to its flexibility in handling different variable types and non-monotone missing patterns across continuous and categorical measures (230,261). Studies 2 and 3 utilised maximum likelihood estimation within the mixed-effects models, which is particularly advantageous for dynamic longitudinal modelling as it can accommodate time-varying covariates and directly model individual trajectories without requiring a complete dataset at each time point (231,236,408). The robustness of both approaches was evaluated through sensitivity analyses comparing findings with complete case sample estimates.

7.4.2 Limitations

While this thesis contributes to the understanding of the interplay between genetic and environmental influences on psychological distress, several limitations should be acknowledged.

One of the primary limitations is that correlation does not imply causation, a constraint that applies to all findings in this thesis. Polygenic scores are statistical constructs derived from genome-wide association studies (GWAS) and do not capture causal biological mechanisms (409,410). Instead, they serve as probabilistic indicators of

genetic liability within a specific population. This makes them useful prediction tools but inadequate for establishing biological pathways to psychological distress. This is due to the aggregated nature of polygenic scores which obscure which specific genetic variants drive any effects creating a "black box". Accordingly, the aim of this thesis was not to establish causality. The findings from this thesis instead add foundational knowledge about the genetic and environmental correlates of psychological distress across adulthood, enabling a more complete epidemiological picture for future research.

A key limitation of this thesis is its reliance on polygenic scores derived from current genome-wide association studies (GWAS), which have relatively modest predictive power (65,410). Across the thesis studies, the polygenic scores explain only a small proportion of variance (low R^2) in psychological distress phenotypes, limiting statistical power. This was particularly a problem in Study 3, where detecting gene-environment interactions requires substantial statistical power, but less so in Studies 1-2. As GWAS sample sizes continue to increase, future research will benefit from more powerful polygenic scores, potentially uncovering effects that could not be detected in the present work.

While this thesis appropriately used European ancestry samples that match the ancestry of the discovery GWAS, our findings may not generalise to other ancestral groups (411). Heritability is a population-specific measure contingent on a given time, place, and environmental context and cannot be extrapolated to infer universal genetic influences on psychological distress (41). For the depressive symptom measurement trait in the GWAS catalog, only 16 of 96 studies included ancestries beyond European populations (267). This reflects a broader pattern where European individuals account for approximately 16% of the global population but represent around 80% of GWAS samples (412). Future research should test whether the estimates observed in this thesis replicate in more diverse populations. Without such efforts, the applicability of findings will remain limited, and the risk of exacerbating health disparities through Eurocentric genetic research will persist (412).

Another important limitation is the reliance on samples of unrelated individuals, which prevents distinguishing between direct genetic effects and indirect genetic influences. Recent research using within-family designs demonstrates that polygenic scores for psychological traits shrink substantially in their explanatory power when comparing siblings or parent-offspring pairs to estimates derived from unrelated individuals (376). For psychological traits, within-family estimates are typically ~20% smaller than those from unrelated individuals, suggesting that a significant proportion of genetic associations may reflect indirect genetic effects, including parental influences on offspring environments (genetic nurture) or other forms of gene-environment correlation (82,376).

Within-family designs offer three key advantages: they control for (1) indirect genetic effects through relatives, (2) assortative mating effects, and (3) population stratification (413). However, it's important to note that these designs are not a complete solution, as recent research found that they do not disaggregate all direct and non-direct effects, particularly those extending beyond the immediate nuclear family (414,415). While this limitation affects the interpretation of the mechanisms underlying our findings, it does

not invalidate the observed associations but suggests caution in attributing them solely to direct biological pathways. This may be a function of the distress phenotype being complex and biosocial. In contrast, traits such as height, which exhibit similar estimates in both unrelated individuals and within-family comparisons, appear to have stronger direct genetic effects (61).

7.5 Future Research Directions

The future research directions informed by the findings of this thesis are outlined below. First, more contexts could be tested. Second, other variance contributors could be tested, i.e. expanding the environmental exposure selection. Finally, different approaches to phenotyping mental health, including symptom-level analyses and modelling of distress trajectories to determine genetic contributions to different patterns of distress development.

Inclusion of more time points and contexts

Future studies can extend the age range used in the thesis when future data time points of both the 1958c and 1970c are released. The 1970c age 52 will be released by the spring of 2025, meaning a better comparator point for age 50 in 1958c (416). Equally, beyond more age points, adding other representative British birth cohorts such as the 1946c (417). This expansion to include the 1946c would necessitate a broader measurement strategy for psychological distress, as they do not have the same phenotypic measure for psychological distress (404). Therefore this would require an assumption that the different questionnaires all measure the same underlying construct of psychological distress.

Further research comparing a wider age range might be beneficial. While this study focused on adulthood using strictly comparable distress measures, future research could include childhood and adolescence. This would once again require a broader measurement strategy and the assumption that there is a general psychopathology genetic contribution that underlies internalising disorders from childhood across adulthood (418). Prior meta-analysis established some continuity between the genetic contribution from childhood into adulthood (239). This could be tested using either the Next Steps cohort (1990c) or the Millennium Cohort Study (2000c) (419,420).The 2000c exhibit even higher average levels of psychological distress than the 1970 cohort, therefore would be a good further cohort comparator (102,124).

As age and cohort trends in distress are not the same in all countries (421). Future research would benefit from expanding analyses to include other longitudinal cohorts across different countries. Cross-country comparison enables an investigation as to whether genetic contributions change in lockstep with population changes in distress. The Health and Retirement Study based in the USA was made to be highly comparable to the English Longitudinal Study of Ageing with similar measures and data collection methods (422,423). Both have polygenic scores and follow adults between the ages of 50-95 years old. Such cross-country comparisons could illuminate how biosocial interactions manifest across different cultural contexts and demographic groups.

Other contributions to variance in psychological distress

Future research should explore alternative approaches to understanding the unexplained variance in psychological distress beyond polygenic scores and social class, which together account for only a modest portion of the total variance – as found in this thesis. This unexplained variance likely encompasses additional genetic factors, including gene-environment interactions and alternative polygenic constructs such as those derived from the 'p-factor' of general psychopathology (111,418,424,425). Non-shared environmental influences within families may also contribute significantly to psychological distress, though the current datasets used in this thesis cannot fully capture these effects (80,415). Identifying these additional contributors would provide a more comprehensive understanding of the aetiology of psychological distress.

As not all sources of variance in psychological distress are systematic. Some arise from stochastic biological processes, such as epigenetic modifications, neurodevelopmental variability, and unpredictable life events. While systematic contributions—such as socioeconomic disadvantage, cohort effects, and cumulative stress exposure—can be studied using structured models, stochastic influences introduce random variability in outcomes, even among individuals with identical genetic and environmental backgrounds (343). While polygenic scores capture average effects, they fail to account for individual trajectories shaped by chance events, epigenetic drift, or developmental noise (409,410). Recognising this stochasticity is critical for advancing our understanding of psychological distress, one that acknowledges the limits of predictability in human development.

Different approaches to phenotyping mental health

Future research will continue to use longitudinal data and integrate polygenic scores (256). While this thesis was guided by existing literature on mean levels of distress, there is potential in developing a trajectory-based approaches (114,123); do polygenic scores predict particular patterns of mental health over the lifecourse? The strengths of the current thesis are the comparability of measures and phenotypes across time. If associations are identified between polygenic scores and trajectories, then it works on the assumption that these trajectories are capturing an underlying psychological distress homogenous pattern (115,426).

Of particular interest would be the examination of persistent or increasing trajectories of distress, which age-specific and cohort-specific factors might influence (150,427). This approach may be helpful as mean-level symptom changes may obscure different patterns of symptoms over time. This approach could help groups at risk of maintaining or developing elevated distress levels over time, potentially informing more targeted interventions. Moreover, understanding how genetic factors operate across different developmental stages could provide insights into critical periods where environmental interventions might be most effective. However, this approach is not without limitations, with the need for large sample sizes to meaningfully disaggregate groups of lifecourse trajectories meaning it is not a panacea.

While the current study utilised sum scores of psychological distress, disaggregating these into symptom-level analyses could reveal differing patterns of genetic and environmental contributions (428,429). This is due to symptom reporting changes across the lifecourse, whereby at older ages more somatic symptoms are reported (129,430). However, the current thesis did not undertake this approach due to measurement reliability issues with single items and binary response options lowering statistical power. The aggregate approach taken in this thesis was better for the measure available in these cohorts and provided greater statical power while also minimised the burden of multiple comparisons.

Yet, a more granular approach to phenotype analysis represents another direction for future research. This approach could help identify specific mechanisms through which genetic propensity influences aspects of psychological distress (431). Currently, such analyses are constrained by the limitations of available genome-wide association studies (GWAS), particularly their sample sizes and the restricted availability of summary statistics for individual distress symptoms, such as those conducted by Thorp et al. in the UK Biobank (13). However, as larger GWAS become available and summary statistics are more widely shared, symptom-level analyses could provide valuable insights into the specific pathways through which genetic and environmental factors combine to influence psychological distress. This more refined approach could help bridge the gap between statistical associations and biological mechanisms, potentially informing more targeted therapeutic approaches.

Reflection on Theoretical Frameworks

Taken together, these findings highlight the need for refinement of existing theoretical models in biosocial epidemiology.

The inconsistency with the SOC model suggests that genetic influences on distress may persist, or even strengthen, across adulthood, challenging assumptions that compensatory cultural resources can fully offset biological vulnerability with age. Future research should therefore consider more dynamic models of psychological distress that integrate both genetic plasticity and time-varying environmental variables across the lifecourse.

The partial support for social control theory refocuses the importance of historical and structural context in shaping the expression of genetic influences. Extending this work to cross-national or cross-generational comparisons could provide stronger tests of how distal environments constrain or amplify genetic effects.

Finally, the lack of evidence for differential susceptibility in Study 3 cautions against assuming universal gene × environment interactions and points instead to the need for large-scale, harmonised datasets to detect robust interaction effects and further meta-analysis of more studies.

Future studies should also examine whether alternative operationalisations of socioeconomic position, or more proximal measures of family environment, yield clearer evidence of moderation.

8 Conclusion

This thesis provides new insights into the interplay between genetic and environmental correlates of psychological distress across adulthood, using longitudinal data from two British birth cohorts. The findings demonstrate that while polygenic scores for psychological distress explain a small but increasing proportion of variance by age, their effects remain persistent yet modest compared to broader socioeconomic and cohort-related influences. These results add to the evidence conceptualising psychological distress as a dynamic phenotype which is shaped by independent developmental and environmental contexts rather than as a static trait with a fixed genetic contribution.

This thesis also highlights the role of structural inequalities in shaping psychological distress, demonstrating that early-life socioeconomic position exerts independent effects, with only weak evidence for interaction effects with genetic liability. This suggests that rather than genetic liability amplifying social gradients in distress, it may be that environmental conditions function as independent population-level differences in mental health outcomes. As the historical shifts in economic structures, employment conditions, and social mobility between the 1958 and 1970 cohorts, these findings underscore the importance of situating genetic factors within their broader societal context. However, given the limited statistical power of the analysis in Study 3 and the relatively small proportion of variance explained by polygenic scores currently, these findings should be interpreted with caution.

Overall, this thesis adds to the evidence that psychological distress is shaped by intersecting biological and environmental factors that evolve by age and cohort, necessitating a lifecourse-informed interdisciplinary approach. Future research should continue integrating longitudinal designs and biosocial frameworks to capture the complexity of gene-environment interplay better. Expanding analyses to diverse populations is particularly critical, as most research in this field remains concentrated in European ancestry from high-income countries.

9 References

1. Susser E, Schwartz S, Morabia A, Bromet EJ. Psychiatric Epidemiology: Searching for the Causes of Mental Disorders. Oxford: Oxford University Press; 2006.

2. Campbell H, Anderson N. Genetic Epidemiology [Internet]. Second Edi. Vol. 3, International Encyclopedia of Public Health. Elsevier; 2016. 248–252 p. Available from: http://dx.doi.org/10.1016/B978-0-12-803678-5.00172-7

3. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life Course Epidemiology. Journal of Epidemiology and Community Health. 2003;57:778–83.

4. Barone C, Hertel FR, Smallenbroek O. The rise of income and the demise of class and social status? A systematic review of measures of socio-economic position in stratification research. Research in Social Stratification and Mobility. 2022 Apr;78:100678.

5. Chauvel L, Leist AK. Socioeconomic hierarchy and health gradient in Europe: the role of income inequality and of social origins. Int J Equity Health. 2015 Dec;14(1):132.

6. Corna LM. A life course perspective on socioeconomic inequalities in health: A critical review of conceptual frameworks. Advances in Life Course Research. 2013 June;18(2):150–9.

7. Pickett KE, Wilkinson RG. Inequality: an underacknowledged source of mental illness and distress. Br J Psychiatry. 2010 Dec;197(6):426–8.

8. Lantz PM, House JS, Mero RP, Williams DR. Stress, Life Events, and Socioeconomic Disparities in Health: Results from the Americans' Changing Lives Study. J Health Soc Behav. 2005 Sept;46(3):274–88.

9. Gadalla TM. Determinants, correlates and mediators of psychological distress: A longitudinal study. Social Science & Medicine. 2009 June;68(12):2199–205.

10. McGonagle KA, Kessler RC. Chronic stress, acute stress, and depressive symptoms. American J of Comm Psychol. 1990 Oct;18(5):681–706.

11. Harris KM, McDade TW. The Biosocial Approach to Human Development, Behavior, and Health Across the Life Course. rsf. 2018 Apr;4(4):2–26.

12. Drapeau A, Marchand A, Beaulieu-Prevost D. Epidemiology of Psychological Distress. In: LAbate L, editor. Mental Illnesses - Understanding, Prediction and Control [Internet]. InTech; 2012 [cited 2023 Aug 14]. Available from: http://www.intechopen.com/books/mental-illnesses-understanding-prediction-and-control/epidemiology-of-psychological-distress

13. Thorp JG, Campos AI, Grotzinger AD, Gerring ZF, An J, Ong JS, et al. Symptom-level modelling unravels the shared genetic architecture of anxiety and depression. Nat Hum Behav. 2021 Apr 15;5(10):1432–42.

14. Kessler RC, Gruber M, Hettema JM, Hwang I, Sampson N, Yonkers KA. Co-morbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow-up. Psychol Med. 2008 Mar;38(3):365–74.

15. Telles-Correia D, Marques JG. Melancholia before the twentieth century: fear and sorrow or partial insanity? Front Psychol [Internet]. 2015 Feb 3 [cited 2022 July 26];6. Available from: http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.00081/abstract

16. Burton R. The Anatomy of Melancholia. London; 1628.

17. Bright T. Treatise of Melancholy. London; 1612.

18. Jansson Å. From Melancholia to Depression: Disordered Mood in Nineteenth-Century Psychiatry [Internet]. Cham: Springer International Publishing; 2021 [cited 2022 Nov 10]. Available from: http://link.springer.com/10.1007/978-3-030-54802-5

19. Berrios GE. Melancholia and Depression During the 19th Century: a Conceptual History. Br J Psychiatry. 1988 Sept;153(3):298–304.

20. Radden J. The nature of melancholy: from Aristotle to Kristeva. New York (N.Y.): Oxford university press; 2002.

21. Kendler KS. The Phenomenology of Major Depression and the Representativeness and Nature of DSM Criteria. AJP. 2016 Aug 1;173(8):771–80.

22. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders [Internet]. Fifth Edition. American Psychiatric Association; 2013 [cited 2024 Oct 21]. Available from: https://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596

23. World Health Organization. The ICD-10 Classification of Mental and Behavioral Disorders. Clinical descriptions and diagnostic guidelines. Geneva, Switzerland: World Health Organization; 1992.

24. Mirowsky J, Ross CE. Measurement for a Human Science. Journal of Health and Social Behavior. 2002;43(2):152–70.

25. Fried EI. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. Journal of Affective Disorders. 2017;208(September 2016):191–7.

26. Mitchell BL, Thorp JG, Wu Y, Campos AI, Nyholt DR, Gordon SD, et al. Polygenic Risk Scores Derived From Varying Definitions of Depression and Risk of Depression. JAMA psychiatry. 2021;1–9.

27. Piazza GG, Allegrini A, Eley TC, Epskamp S, Fried EI, Isvoranu AM, et al. Unweaving the web: Polygenic influences on networks of psychopathology symptoms [Internet]. PsyArXiv; 2023 Feb [cited 2023 Apr 14]. Available from: https://osf.io/6mzxd

28. Kendler KS. Major Depression and Generalized Anxiety Disorder: Same Genes, (Partly) Different Environments? Arch Gen Psychiatry. 1992 Sept 1;49(9):716.

29. Mei L, Gao Y, Chen M, Zhang X, Yue W, Zhang D, et al. Overlapping common genetic architecture between major depressive disorders and anxiety and stress-related disorders. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2022 Mar;113:110450.

30. Jermy B, Howard DM, Glanville K, Hagenaars SP, Coleman JRI, Breen G, et al. Using major depression polygenic risk scores to explore the depressive symptom continuum. Psychological medicine. 2020;1–10.

31. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. 2018 Nov;392(10159):1789–858.

32. World Health Organization. Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level: Report by the Secretariat. [Internet]. World Health Organisation; 2011 [cited 2023 Mar 30]. Available from: apps.who.int/gb/ebwha/pdf_files/EB130/B130_9-en.pdf

33. Clark LA, Cuthbert B, Lewis-Fernández R, Narrow WE, Reed GM. Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). Psychological Science in the Public Interest. 2017;18(2):72–145.

34. Kaufman J, Charney D. Comorbidity of mood and anxiety disorders. Depress Anxiety. 2000;12(S1):69–76.

35. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. The Lancet. 2018 Oct;392(10157):1553–98.

36. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. The Lancet. 2007 Sept;370(9590):878–89.

37. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. The Lancet. 2007 Sept;370(9590):859–77.

38. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: Review and meta-analysis. American Journal of Psychiatry. 2000;157(10):1552–62.

39. Plomin R, Haworth CMA, Davis OSP. Common disorders are quantitative traits. Nat Rev Genet. 2009 Dec;10(12):872–8.

40. Turkheimer E. Three laws of behavior genetics and what they mean. Current Directions in Psychological Science. 2000;9(5):160–4.

41. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era - Concepts and misconceptions. Nature Reviews Genetics. 2008;9(4):255–66.

42. Polderman TJC, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nature Genetics. 2015;47(7):702–9.

43. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002 Nov;3(11):872–82.

44. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. American Journal of Human Genetics. 2017;101(1):5–22.

45. Alberts B, editor. Molecular biology of the cell. 4th ed. New York: Garland Science; 2002. 1 p.

46. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends in Genetics. 2001 Sept;17(9):502–10.

47. Hernandez RD, Uricchio LH, Hartman K, Ye C, Dahl A, Zaitlen N. Ultrarare variants drive substantial cis heritability of human gene expression. Nat Genet. 2019 Sept;51(9):1349–55.

48. Young AS, Martin HC. Discovering genes that affect cognitive ability. Trends in Genetics. 2023 Nov;39(11):810–2.

49. Tian R, Ge T, Kweon H, Rocha DB, Lam M, Liu JZ, et al. Whole-exome sequencing in UK Biobank reveals rare genetic architecture for depression. Nat Commun. 2024 Feb 26;15(1):1755.

50. Noordam R, Direk N, Sitlani CM, Aarts N, Tiemeier H, Hofman A, et al. Identifying genetic loci associated with antidepressant drug response with drug–gene interaction models in a population-based study. Journal of Psychiatric Research. 2015 Mar;62:31–7.

51. Baselmans BML, Jansen R, Ip HF, van Dongen J, Abdellaoui A, van de Weijer MP, et al. Multivariate genome-wide analyses of the well-being spectrum. Nature Genetics. 2019;51(3):445–51.

52. Sullivan PF, De Geus EJC, Willemsen G, James MR, Smit JH, Zandbelt T, et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. Mol Psychiatry. 2009 Apr;14(4):359–75.

53. McIntosh AM, Lewis CM, Mark J Adams for the Psychiatric Genomics Consortium Major Depressive Disorder Working Group. Genome-wide study of half a million individuals with major depression identifies 697 independent associations, infers causal neuronal subtypes and biological targets for novel pharmacotherapies [Internet]. 2024 [cited 2024 Oct 25]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.04.29.24306535

54. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013 Sept;45(9):984–94.

55. Levey DF, Stein MB, Wendt FR, Pathak GA, Zhou H, Aslan M, et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nature Neuroscience. 2021;24(July).

56. Andreassen OA, Hindley GFL, Frei O, Smeland OB. New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications. World Psychiatry. 2023 Feb;22(1):4–24.

57. Tropf FC, Lee SH, Verweij RM, Stulp G, Van Der Most PJ, De Vlaming R, et al. Hidden heritability due to heterogeneity across seven populations. Nature Human Behaviour. 2017;1(10):757–65.

58. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct;461(7265):747–53.

59. Young AI, Frigge ML, Gudbjartsson DF, Thorleifsson G, Bjornsdottir G, Sulem P, et al. Relatedness disequilibrium regression estimates heritability without environmental bias. Nat Genet. 2018 Sept;50(9):1304–10.

60. Cheesman R, Eilertsen EM, Ahmadzadeh YI, Gjerde LC, Hannigan LJ, Havdahl A, et al. How important are parents in the development of child anxiety and depression? A genomic analysis of parent-offspring trios in the Norwegian Mother Father and Child Cohort Study (MoBa). BMC Med. 2020 Dec;18(1):284.

61. Tan T, Jayashankar H, Guan J, Nehzati SM, Mir M, Bennett M, et al. Family-GWAS reveals effects of environment and mating on genetic associations [Internet]. 2024 [cited 2024 Oct 21]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.10.01.24314703

62. Greenwald AG, Banaji MR, Nosek BA. Statistically small effects of the Implicit Association Test can have societally large effects. Journal of Personality and Social Psychology. 2015 Apr;108(4):553–61.

63. Götz FM, Gosling SD, Rentfrow PJ. Small Effects: The Indispensable Foundation for a Cumulative Psychological Science. Perspect Psychol Sci. 2022 Jan;17(1):205–15.

64. Choi SW, Mak TSH, O'Reilly PF. A guide to performing Polygenic Risk Score analyses. bioRxiv. 2018;2:11–3.

65. Dudbridge F. Power and Predictive Accuracy of Polygenic Risk Scores. Wray NR, editor. PLoS Genet. 2013 Mar 21;9(3):e1003348.

66. Pirastu N, Cordioli M, Nandakumar P, Mignogna G, Abdellaoui A, Hollis B, et al. Genetic analyses identify widespread sex-differential participation bias. Nat Genet. 2021 May;53(5):663–71.

67. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault JB, Kutalik Z. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. Nat Hum Behav. 2023 Apr 27;7(7):1216–27.

68. Pingault J, Allegrini AG, Odigie T, Frach L, Baldwin JR, Rijsdijk F, et al. Research Review: How to interpret associations between polygenic scores, environmental risks, and phenotypes. Child Psychology Psychiatry. 2022 Oct;63(10):1125–39.

69. Sud A, Horton RH, Hingorani AD, Tzoulaki I, Turnbull C, Houlston RS, et al. Realistic expectations are key to realising the benefits of polygenic scores. BMJ. 2023 Feb 28;e073149.

70. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019 July 25;10(1):3328.

71. Janssens ACJW. Validity of polygenic risk scores: are we measuring what we think we are? Human Molecular Genetics. 2019 Nov 21;28(R2):R143–50.

72. Morris TT, Davies NM, Hemani G, Smith GD. Population phenomena inflate genetic associations of complex social traits. Sci Adv. 2020 Apr 17;6(16):eaay0328.

73. Bigdeli TB, Voloudakis G, Barr PB, Gorman BR, Genovese G, Peterson RE, et al. Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia, Bipolar Disorder, and Depression Among Adults in the US Veterans Affairs Health Care System. JAMA Psychiatry. 2022 Nov 1;79(11):1092.

74. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet. 2013 Oct;132(10):1077–130.

75. Conley D, Laidley TM, Boardman JD, Domingue BW. Changing Polygenic Penetrance on Phenotypes in the 20th Century Among Adults in the US Population. Sci Rep. 2016 July 26;6(1):30348.

76. Kendler KS, Greenspan RJ. The Nature of Genetic Influences on Behavior: Lessons From “Simpler” Organisms. AJP. 2006 Oct;163(10):1683–94.

77. Tabery J. R. A. Fisher, Lancelot Hogben, and the Origin(s) of Genotype–Environment Interaction. J Hist Biol. 2008 Dec;41(4):717–61.

78. Plomin R, Daniels D. Why are children in the same family so different from one another? Behav Brain Sci. 1987 Mar;10(1):1–16.

79. Turkheimer E, Waldron M. Nonshared environment: A theoretical, methodological, and quantitative review. Psychological Bulletin. 2000;126(1):78–108.

80. Tikhodeyev ON, Shcherbakova OV. The Problem of Non-Shared Environment in Behavioral Genetics. Behavior Genetics. 2019;49(3):259–69.

81. Daw J, Guo G, Harris KM. Nurture net of nature: Re-evaluating the role of shared environments in academic achievement and verbal intelligence. Social Science Research. 2015 July;52:422–39.

82. Kong A, Thorleifsson G, Frigge ML, Vilhjálmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. Science. 2018;428(January):424–8.

83. Plomin R. Beyond Nature versus Nurture. In: Hall LL, editor. Genetics and Mental Illness [Internet]. Boston, MA: Springer US; 1996 [cited 2025 Feb 5]. p. 29–50. Available from: http://link.springer.com/10.1007/978-1-4899-0170-5_3

84. Galton F. Inquiries into human faculty and its development. [Internet]. New York: MacMillan Co; 1883 [cited 2025 Feb 5]. Available from: https://content.apa.org/books/14178-000

85. Turkheimer E. Understanding the nature–nurture debate. Cambridge: Cambridge University Press; 2024. 1 p. (Understanding life).

86. Hogben L. The limits of applicability of correlation technique in human genetics. Journ of Genetics. 1933 Aug;27(3):379–406.

87. Jaffee SR, Price TS. Gene–environment correlations: a review of the evidence and implications for prevention of mental illness. Mol Psychiatry. 2007 May 1;12(5):432–42.

88. Plomin R, DeFries JC, Loehlin JC. Genotype-environment interaction and correlation in the analysis of human behavior. Psychol Bull. 1977 Mar;84(2):309–22.

89. Belsky J, Hartman S. Gene-environment interaction in evolutionary perspective: differential susceptibility to environmental influences. World Psychiatry. 2014 Feb;13(1):87–9.

90. Boyce WT, Sokolowski MB, Robinson GE. Genes and environments, development and time. Proc Natl Acad Sci USA. 2020 Sept 22;117(38):23235–41.

91. Belsky J, Pluess M. Beyond diathesis stress: Differential susceptibility to environmental influences. Psychological Bulletin. 2009;135(6):885–908.

92. Mills MC, Tropf FC. Sociology, Genetics, and the Coming of Age of Sociogenomics. Annu Rev Sociol. 2020 July 30;46(1):553–81.

93. Boyce WT, Ellis BJ. Biological sensitivity to context: I. An evolutionary–developmental theory of the origins and functions of stress reactivity. Develop Psychopathol [Internet]. 2005 June [cited 2024 Mar 19];17(02). Available from: http://www.journals.cambridge.org/abstract_S0954579405050145

94. Monroe SM, Simons AD. Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders. Psychological Bulletin. 1991;110(3):406–25.

95. Caspi A, Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, et al. Influence of Life Stress on Depression : Moderation by a Polymorphism in the 5-HTT Gene. 2012;386(2003).

96. Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, et al. Role of Genotype in the Cycle of Violence in Maltreated Children. Science. 2002 Aug 2;297(5582):851–4.

97. Bronfenbrenner U, Ceci SJ. Nature-nurture reconceptualised: A bio-ecological model. Psychological Review. 1994;10(4):568–86.

98. Engzell P, Tropf FC. Heritability of education rises with intergenerational mobility. Proc Natl Acad Sci USA. 2019 Dec 17;116(51):25386–8.

99. Shanahan MJ, Hofer SM. Social Context in Gene–Environment Interactions: Retrospect and Prospect. The Journals of Gerontology: Series B. 2005 Mar 1;60(Special_Issue_1):65–76.

100. Rutter M, Pickles A. Person-environment interactions: Concepts, mechanisms, and implications for data analysis. In: Wachs TD, Plomin R, editors. Conceptualization and measurement of organism-environment interaction [Internet]. Washington: American Psychological Association; 1991 [cited 2024 Mar 27]. p. 105–41. Available from: https://content.apa.org/books/10100-006

101. Dick DM. Gene-Environment Interaction in Psychological Traits and Disorders. Annu Rev Clin Psychol. 2011 Apr 27;7(1):383–409.

102. Armitage JM, Kwong ASF, Tseliou F, Sellers R, Blakey R, Anthony R, et al. Cross-cohort change in parent-reported emotional problem trajectories across childhood and adolescence in the UK. The Lancet Psychiatry. 2023 July;10(7):509–17.

103. Hang S, Jost GM, Guyer AE, Robins RW, Hastings PD, Hostinar CE. Understanding the development of chronic loneliness in youth. Child Dev Perspectives. 2023 Dec 22;cdep.12496.

104. Alexander D. Genes, Determinism and God [Internet]. 1st edn. Cambridge University Press; 2017 [cited 2025 Feb 5]. Available from: https://www.cambridge.org/core/product/identifier/9781316493366/type/book

105. Sallis H, Evans J, Wootton R, Krapohl E, Oldehinkel AJ, Davey Smith G, et al. Genetics of depressive symptoms in adolescence. BMC Psychiatry. 2017;17(1):1–8.

106. Hannigan LJ, Walaker N, Waszczuk MA, McAdams TA, Eley TC. Aetiological Influences on Stability and Change in Emotional and Behavioural Problems across Development: A Systematic Review. Journal of Experimental Psychopathology. 2017;a4(1):52–108.

107. Nivard MG, Dolan CV, Kendler KS, Kan KJ, Willemsen G, Van Beijsterveldt CEM, et al. Stability in symptoms of anxiety and depression as a function of

genotype and environment: A longitudinal twin study from ages 3 to 63 years. Psychological Medicine. 2015;45(5):1039–49.

108. Baselmans BML, Willems YE, van Beijsterveldt CEM, Ligthart L, Willemsen G, Dolan CV, et al. Unraveling the genetic and environmental relationship between well-being and depressive symptoms throughout the lifespan. Frontiers in Psychiatry. 2018;9(JUN):1–12.

109. Torvik FA, Gustavson K, Ystrom E, Rosenström TH, Gillespie N, Reichborn-Kjennerud T, et al. Continuity of genetic and environmental influences on clinically assessed major depression from ages 18 to 45. Psychol Med. 2019 Nov;49(15):2582–90.

110. Nes RB, Røysamb E, Reichborn-Kjennerud T, Harris JR, Tambs K. Symptoms of Anxiety and Depression in Young Adults: Genetic and Environmental Influences on Stability and Change. Twin Res Hum Genet. 2007 June 1;10(3):450–61.

111. Gillespie NA, Kirk KM, Evans DM, Heath AC, Hickie IB, Martin NG. Do the Genetic or Environmental Determinants of Anxiety and Depression Change with Age? A Longitudinal Study of Australian Twins. Twin Res. 2004 Feb 1;7(1):39–53.

112. Bronfenbrenner U. Environments in developmental perspective: Theoretical and operational models. In: Friedman SL, Wachs TD, editors. Measuring environment across the life span: Emerging methods and concepts [Internet]. Washington: American Psychological Association; 1999 [cited 2023 Nov 23]. p. 3–28. Available from: http://content.apa.org/books/10317-001

113. Baltes PB. On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. American Psychologist. 1997 Apr;52(4):366–80.

114. Kwong ASF, Manley D, Timpson NJ, Pearson RM, Heron J, Sallis H, et al. Identifying Critical Points of Trajectories of Depressive Symptoms from Childhood to Young Adulthood. Journal of Youth and Adolescence. 2019;48(4):815–27.

115. Colman I, Ploubidis GB, Wadsworth MEJ, Jones PB, Croudace TJ. A Longitudinal Typology of Symptoms of Depression and Anxiety Over the Life Course. Biological Psychiatry. 2007;62(11):1265–71.

116. Colman I, Ataullahjan A. Life course perspectives on the epidemiology of depression. Canadian Journal of Psychiatry. 2010;55(10):622–32.

117. Ryder NB. The Cohort as a Concept in the Study of Social Change. American Sociological Review. 1965 Dec;30(6):843.

118. Fosse E, Winship C. The Anatomy of Cohort Analysis: Decomposing Comparative Cohort Careers. Sociological Methodology. 2023 Aug;53(2):217–68.

119. Yang Y, Land KC. A Mixed Models Approach to the Age-Period-Cohort Analysis of Repeated Cross-Section Surveys, with an Application to Data on Trends in Verbal Test Scores. Sociological Methodology. 2006 Aug;36(1):75–97.

120. Bell A, Jones K. The impossibility of separating age, period and cohort effects. Social Science & Medicine. 2013 Sept;93:163–5.

121. Rohrer JM. Thinking Clearly About Age, Period, and Cohort Effects. Advances in Methods and Practices in Psychological Science. 2025 Apr;8(2):25152459251342750.

122. Bell A. Age period cohort analysis: a review of what we should and shouldn't do. Annals of Human Biology. 2020 Feb 17;47(2):208–17.

123. Bell A. Life-course and cohort trajectories of mental health in the UK, 1991–2008 – A multilevel age–period–cohort analysis. Social Science & Medicine. 2014 Nov;120:21–30.

124. Gondek D, Bann D, Patalay P, Goodman A, McElroy E, Richards M, et al. Psychological distress from early adulthood to early old age: Evidence from the 1946, 1958 and 1970 British birth cohorts. Psychological Medicine. 2021;(May).

125. Ploubidis GB, Sullivan A, Brown M, Goodman A. Psychological distress in mid-life: evidence from the 1958 and 1970 British birth cohorts. Psychol Med. 2017 Jan;47(2):291–303.

126. Blanchflower DG, Oswald AJ. Is Well-being U-Shaped over the Life Cycle ? 2008;

127. Spiers N, Brugha TS, Bebbington P, McManus S, Jenkins R, Meltzer H. Age and birth cohort differences in depression in repeated cross-sectional surveys in England: the National Psychiatric Morbidity Surveys, 1993 to 2007. Psychol Med. 2012 Oct;42(10):2047–55.

128. Sacker A, Wiggins RD. Age–period–cohort effects on inequalities in psychological distress, 1981–2000. Psychol Med. 2002 Aug;32(6):977–90.

129. Musliner KL, Munk-Olsen T, Eaton WW, Zandi PP. Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. Journal of Affective Disorders. 2016 Mar;192:199–211.

130. Oerlemans AM, Wardenaar KJ, Raven D, Hartman CA, Ormel J. The association of developmental trajectories of adolescent mental health with early-adult functioning. Santana GL, editor. PLoS ONE. 2020 June 10;15(6):e0233648.

131. Thern E, Matilla-Santander N, Hernando-Rodriguez JC, Almroth M, Bodin T. Precarious employment in early adulthood and later mental health problems: a register-linked cohort study. J Epidemiol Community Health. 2023 Dec;77(12):755–61.

132. Lachman ME. Development in Midlife. Annu Rev Psychol. 2004 Feb 1;55(1):305–31.

133. Coyle CE, Dugan E. Social Isolation, Loneliness and Health Among Older Adults. J Aging Health. 2012 Dec;24(8):1346–63.

134. Carstensen LL, Fung HH, Charles ST. Socioemotional Selectivity Theory and the Regulation of Emotion in the Second Half of Life. Motivation and Emotion. 2003;27(2):103–23.

135. Bambra C. The U-Shaped Curve of Health Inequalities Over the 20th and 21st Centuries. International Journal of Social Determinants of Health and Health Services. 2024 July;54(3):199–205.

136. Frijters P, Beatton T. The mystery of the U-shaped relationship between happiness and age. Journal of Economic Behavior and Organization. 2012;82(2–3):525–42.

137. Blanchflower D, Bryson A, Xu X. The Declining Mental Health Of The Young And The Global Disappearance Of The Hump Shape In Age In Unhappiness [Internet]. Cambridge, MA: National Bureau of Economic Research; 2024 Apr [cited 2024 Oct 24] p. w32337. Report No.: w32337. Available from: http://www.nber.org/papers/w32337.pdf

138. Brown RL, Richman JA, Rospenda KM. Economic Stressors and Psychological Distress: Exploring Age Cohort Variation in the Wake of the Great Recession. Stress and Health. 2017;33(3):267–77.

139. Ferri E, Bynner J, Wadsworth M, editors. Changing Britain, changing lives: three generations at the turn of the century. London: Inst. of Education, Univ. of London; 2003. 344 p. (Bedford Way papers).

140. Schwandt H, Von Wachter T. Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets. Journal of Labor Economics. 2019 Jan;37(S1):S161–98.

141. Bentley R, Baker E, Ronald R, Reeves A, Smith SJ, Simons K, et al. Housing affordability and mental health: an analysis of generational change. Housing Studies. 2022 Nov 26;37(10):1842–57.

142. Our World in Data. Our World in Data: Life expectancy at birth. 2023. Our World in Data: Life expectancy at birth. Available from: https://ourworldindata.org/grapher/life-expectancy?country=%7EGBR

143. Bank of England. Millennium of Macro: Hours Worked since 1700 table A.54. [Internet]. Available from: https://www.bankofengland.co.uk/statistics/research-datasets

144. Nationwide. Nationwide UK House Price Index [Internet]. Available from: https://www.nationwidehousepriceindex.co.uk/download/uk-house-prices-since-1952/

145. Office for National Statistics. Long-term trends in UK employment: 1861 to 2018 [Internet]. 2019. Available from:

https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/compendium/economicreview/april2019/longtermtrendsinukemployment1861to2018

146. Bolton P. Education: Historical statistics [Internet]. House of Commons Library; 2012. Available from: https://researchbriefings.files.parliament.uk/documents/SN04252/SN04252.pdf

147. Clark T, Goodman A. Living standards under Labour [Internet]. 2001. Available from: https://ifs.org.uk/sites/default/files/2022-07/Living-standards-under-labour.pdf

148. Marmot M. Health equity in England: the Marmot review 10 years on. BMJ. 2020 Feb 24;m693.

149. Hall PA. Great Britain: The Role of Government and the Distribution of Social Capital. In: Putnam RD, editor. Democracies in Flux [Internet]. 1st edn Oxford University PressNew York; 2002 [cited 2024 Mar 28]. p. 21–58. Available from: https://academic.oup.com/book/8126/chapter/153593089

150. Moulton V, Sullivan A, Goodman A, Parsons S, Ploubidis GB. Adult life-course trajectories of psychological distress and economic outcomes in midlife during the COVID-19 pandemic: evidence from the 1958 and 1970 British birth cohorts. Soc Psychiatry Psychiatr Epidemiol. 2023 May;58(5):779–94.

151. Thompson EJ, Stafford J, Moltrecht B, Huggins CF, Kwong ASF, Shaw RJ, et al. Psychological distress, depression, anxiety, and life satisfaction following COVID-19 infection: evidence from 11 UK longitudinal population studies. The Lancet Psychiatry. 2022 Nov;9(11):894–906.

152. Kuehner C. Why is depression more common among women than among men? The Lancet Psychiatry. 2017 Feb;4(2):146–58.

153. Salk RH, Hyde JS, Abramson LY. Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. Psychological Bulletin. 2017 Aug;143(8):783–822.

154. Twenge JM, Nolen-Hoeksema S. Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: A meta-analysis. Journal of Abnormal Psychology. 2002 Nov;111(4):578–88.

155. Kuehner C. Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. Acta Psychiatr Scand. 2003 Sept;108(3):163–74.

156. Butler J. Gender trouble: feminism and the subversion of identity. New York: Routledge; 2006. 236 p. (Routledge classics).

157. Crystal S, Shea DG, Reyes AM. Cumulative Advantage, Cumulative Disadvantage, and Evolving Patterns of Late-Life Inequality. GERONT. 2016 Mar 30;gnw056.

158. Kratz F, Patzina A. Endogenous selection bias and cumulative inequality over the life course: Evidence from educational inequality in subjective well-being. European Sociological Review. 2020;36(3):333–50.

159. Kendler KS, Thornton LM, Prescott CA. Gender Differences in the Rates of Exposure to Stressful Life Events and Sensitivity to Their Depressogenic Effects. AJP. 2001 Apr 1;158(4):587–93.

160. Khramtsova EA, Davis LK, Stranger BE. The role of sex in the genomics of human complex traits. Nat Rev Genet. 2019 Mar;20(3):173–90.

161. Kropp DR, Hodes GE. Sex differences in depression: An immunological perspective. Brain Research Bulletin. 2023 May;196:34–45.

162. Oyola MG, Handa RJ. Hypothalamic–pituitary–adrenal and hypothalamic–pituitary–gonadal axes: sex differences in regulation of stress responsivity. Stress. 2017 Sept 3;20(5):476–94.

163. Blokland GAM, Grove J, Chen CY, Cotsapas C, Tobet S, Handa R, et al. Sex-Dependent Shared and Nonshared Genetic Architecture Across Mood and Psychotic Disorders. Biological Psychiatry. 2022 Jan;91(1):102–17.

164. Silveira PP, Pokhvisneva I, Howard DM, Meaney MJ. A Sex-Specific Genome-Wide Association Study of Depression Phenotypes in UK Biobank. 2022 Mar 31 [cited 2022 June 27]; Available from: http://medrxiv.org/lookup/doi/10.1101/2022.03.30.22273201

165. Thomas JT, Thorp JG, Huider F, Grimes PZ, Wang R, Youssef P, et al. Sex-stratified genome-wide association meta-analysis of major depressive disorder. Nat Commun. 2025 Aug 26;16(1):7960.

166. Campbell OLK, Bann D, Patalay P. The gender gap in adolescent mental health: A cross-national investigation of 566,829 adolescents across 73 countries. SSM - Population Health. 2021;13(January):100742.

167. Glymour MM, Avendano M, Kawachi I. Socioeconomic Status and Health. In: Social Epidemiology. 2nd edn Oxford: Oxford University Press; 2014.

168. Galobardes B, Lynch J, Smith GD. Measuring socioeconomic position in health research. British Medical Bulletin. 2007 Feb 6;81–82(1):21–37.

169. Bartley M. Health inequality: an introduction to concepts, theories and methods. Second edition. Cambridge, UK Malden, MA, USA: polity; 2017. 244 p.

170. Bourdieu P, Champagne P, Poupeau F, Rivière MC, Bourdieu P. Forms of capital: lectures at the Collège de France (1983-1984). Duval J, editor. Cambridge, UK Medford, MA: Polity; 2021. 380 p. (General sociology).

171. Marx K, Fowkes B, Fernbach D, Engels F, Mandel E, Marx K. Capital: a critique of political economy; v.1. London New York, N.Y., USA: Penguin Books in association with New Left Review; 1990. 1141 p. (Penguin classics).

172. Weber M, Parsons T. The theory of social and economic organization. First Free Press paperback edition 1964. London [England]: Collier Macmillan Publishers; 1964. 1 p. (A Free Press paperback).

173. Kröger H, Pakpahan E, Hoffmann R. What causes health inequality? A systematic review on the relative importance of social causation and health selection. Eur J Public Health. 2015 Dec;25(6):951–60.

174. Phelan JC, Link BG, Tehranifar P. Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. Journal of Health and Social Behavior. 2010;51(1_suppl):S28–40.

175. Chandola T, Bartley M, Sacker A, Jenkinson C, Marmot M. Health selection in the Whitehall II study, UK. Social Science & Medicine. 2003 May;56(10):2059–72.

176. Theodossiou I, Zangelidis A. The social gradient in health: The effect of absolute income and subjective social status assessment on the individual's health in Europe. Economics & Human Biology. 2009 July;7(2):229–37.

177. Hertzman C, Boyce T. How experience gets under the skin to create gradients in developmental health. Annual Review of Public Health. 2010;31:329–47.

178. Pickett KE, Wilkinson RG. Income inequality and health: A causal review. Social Science & Medicine. 2015 Mar;128:316–26.

179. Präg P, Mills M, Wittek R. Income and Income Inequality as Social Determinants of Health: Do Social Comparisons Play a Role? European Sociological Review. 2014 Apr;30(2):218–29.

180. Kawachi I. Social Ties and Mental Health. Journal of Urban Health: Bulletin of the New York Academy of Medicine. 2001 Sept 1;78(3):458–67.

181. Orton LC, Pennington A, Nayak S, Sowden A, Petticrew M, White M, et al. What is the evidence that differences in 'control over destiny' lead to socioeconomic inequalities in health? A theory-led systematic review of high-quality longitudinal studies on pathways in the living environment. Journal of Epidemiology and Community Health. 2019;73(10):929–34.

182. Angelini V, Howdon DDH, Mierau JO. Childhood Socioeconomic Status and Late-Adulthood Mental Health: Results From the Survey on Health, Ageing and Retirement in Europe. The Journals of Gerontology: Series B. 2019 Jan 1;74(1):95–104.

183. Susser E, Koenen K, Rudenstine S, Galea S. Applying a lifecourse perspective to depression. In: A Life Course Approach to Mental Disorders. Oxford: Oxford University Press; 2014.

184. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. Nat Genet. 2022 Apr;54(4):437–49.

185. Akimova ET, Wolfram T, Ding X, Tropf FC, Mills MC. Polygenic prediction of occupational status GWAS elucidates genetic and environmental interplay in intergenerational transmission, careers and health in UK Biobank. Nat Hum Behav [Internet]. 2024 Dec 23 [cited 2025 Feb 5]; Available from: https://www.nature.com/articles/s41562-024-02076-3

186. Andreu-Bernabeu Á, González-Peñas J, Arango C, Díaz-Caneja CM. Socioeconomic status and severe mental disorders: a bidirectional multivariable Mendelian randomisation study. BMJ Ment Health. 2023 Nov;26(1):e300821.

187. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). International Journal of Epidemiology. 2006;35(1):34–41.

188. Sullivan A, Brown M, Hamer M, Ploubidis GB. Cohort Profile Update: The 1970 British Cohort Study (BCS70). International Journal of Epidemiology. 2023 June 6;52(3):e179–86.

189. Bridges EC, Rayner NW, Mountford HS, Bates TC, Luciano M. Longitudinal Reading Measures and Genome Imputation in the National Child Development Study: Prospects for Future Reading Research. Twin Res Hum Genet. 2023 Feb;26(1):10–20.

190. Shireby G, Morris TT, Wong A, Chaturvedi N, Ploubidis GB, Fitzsimmons E, et al. Data Resource Profile: Genomic Data in Multiple British Birth Cohorts (1946-2001)—Health, Social, and Environmental Data from Birth to Old Age [Internet]. 2024 [cited 2025 Feb 5]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.11.06.24316761

191. U.K. Data Service. National Child Development Study [Internet]. 2022. Available from: https://ukdataservice.ac.uk

192. Rutter M, Tizard J, Whitmore K. Education, Health And Behaviour. London: Longman; 1970. London: Longman; 1970.

193. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 June;30(6):473–83.

194. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Short Warwick–Edinburgh Mental Well-being Scale [Internet]. 2021 [cited 2025 Apr 1]. Available from: https://doi.apa.org/doi/10.1037/t80221-000

195. Spitzer RL, Kroenke K, Williams JBW, Löwe B. Generalized Anxiety Disorder 7 [Internet]. 2011 [cited 2025 Apr 1]. Available from: https://doi.apa.org/doi/10.1037/t02591-000

196. Kroenke K, Spitzer RL, Williams JBW. Patient Health Questionnaire-9 [Internet]. 2011 [cited 2025 Apr 1]. Available from: https://doi.apa.org/doi/10.1037/t06165-000

197. McGee R, Williams S, Silva PA. An evaluation of the Malaise inventory. Journal of Psychosomatic Research. 1986 Jan;30(2):147–52.

198. Brodman K. The Cornell Medical Index: An Adjunct to Medical Interview. JAMA. 1949 June 11;140(6):530.

199. Rodgers B, Pickles A, Power C, Collishaw S, Maughan B. Validity of the Malaise Inventory in general population samples. Social Psychiatry and Psychiatric Epidemiology. 1999 June 25;34(6):333–41.

200. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951 Sept;16(3):297–334.

201. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ. 2018 Dec;48(6):1273–96.

202. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Medical Education. 2011 June 27;2:53–5.

203. Hirst MA, Bradshaw JR. Evaluating the Malaise inventory: A comparison of measures of stress. Journal of Psychosomatic Research. 1983 Jan;27(3):193–9.

204. Ploubidis GB, McElroy E, Moreira HC. A longitudinal examination of the measurement equivalence of mental health assessments in two british birth cohorts. Longitudinal and Life Course Studies. 2019;10(4):471–89.

205. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience. 2019 July 1;8(7):giz082.

206. Goldberg D. Manual of the general health questionnaire. Nfer Nelson; 1978.

207. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population.

208. Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. The American Journal of Human Genetics. 2019 Dec;105(6):1213–21.

209. Pain O, Glanville KP, Hagenaars SP, Selzam S, Fürtjes AE, Gaspar HA, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. Plagnol V, editor. PLoS Genet. 2021 May 4;17(5):e1009021.

210. Shifman S. Linkage disequilibrium patterns of the human genome across populations. Human Molecular Genetics. 2003 Apr 1;12(7):771–6.

211. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nature Neuroscience. 2019;22(3):343–52.

212. Kwong ASF, Morris TT, Pearson RM, Timpson NJ, Rice F, Stergiakouli E, et al. Polygenic risk for depression, anxiety and neuroticism are associated with the severity and rate of change in depressive symptoms across adolescence. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2021;

213. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. 2020;

214. Bland R. Measuring 'Social Class': A Discussion of the Registrar-General's Classification. Sociology. 1979 May;13(2):283–91.

215. Wood N, Bann D, Hardy R, Gale C, Goodman A, Crawford C, et al. Childhood socioeconomic position and adult mental wellbeing: Evidence from four British birth cohort studies. Ergin I, editor. PLoS ONE. 2017 Oct 25;12(10):e0185798.

216. Szreter SRS. The Genesis of the Registrar-General's Social Classification of Occupations. The British Journal of Sociology. 1984 Dec;35(4):522.

217. Barra Roantree, Kartik Vira. The rise and rise of women's employment in the UK: IFS Briefing Note BN234 [Internet]. Institute for Fiscal Studies; Report No.: IFS Briefing Note BN234. Available from: https://ifs.org.uk/sites/default/files/output_url_files/BN234.pdf

218. Gagné T, Sacker A, Schoon I. Changes in Patterns of Social Role Combinations at Ages 25–26 among Those Growing Up in England between 1996 and 2015–16: Evidence from the 1970 British Cohort and Next Steps Studies. J Youth Adolescence. 2021 Oct;50(10):2052–66.

219. Erikson R. Social Class of Men, Women and Families. Sociology. 1984 Nov;18(4):500–14.

220. Penn R, Rose M, Rubery J, editors. Skill and occupational change. Oxford ; New York: Oxford University Press; 1994. 365 p. (Social change and economic life initiative).

221. Elias, Peter. Social class and the standard occupational classification. In: Rose, D., (ed.) .Swindon: Economic and Social Research Council, pp. 40-64. 1995. (A report on phase 1 of the ESRC review of OPCS social classifications).

222. Galea S, Tracy M. Participation Rates in Epidemiologic Studies. Annals of Epidemiology. 2007 Sept;17(9):643–53.

223. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006;38(8):904–9.

224. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. Current Protocols in Human Genetics. 2017;95(1):1.22.1-1.22.23.

225. Privé F, Luu K, Blum MGB, McGrath JJ, Vilhjálmsson BJ. Efficient toolkit implementing best practices for principal component analysis of population genetic data. Schwartz R, editor. Bioinformatics. 2020 Aug 15;36(16):4449–57.

226. Steele F. Multilevel Modelling of Repeated Measures Data [Internet]. Bristol; 2014. (LEMMA VLE Module 15). Available from: http://www.bristol.ac.uk/cmm/learning/course.html

227. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine. 2011 Feb 20;30(4):377–99.

228. Rubin D. Inference and missing data. Biometrika. 1976;63:581–92.

229. Little RJA, Rubin DB. Statistical analysis with missing data. Third edition. Hoboken, NJ: Wiley; 2020. 449 p. (Wiley series in probability and statistics).

230. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psych Res. 2011 Mar;20(1):40–9.

231. Silverwood R, Narayanan M, Dodgeon B, Ploubidis GB. Handling missing data in the National Child Development Study: User guide. London: UCL Centre for Longitudinal Studies; 2021.

232. Centre for Longitudinal Studies. NCDS response and handling missing data [Internet]. 2024. Available from: https://cls.ucl.ac.uk/cls-studies/1958-national-child-development-study/ncds-response-and-handling-missing-data/

233. Rubin DB. Multiple Imputation for Nonresponse in Surveys [Internet]. 1st edn. Wiley; 1987 [cited 2024 Oct 2]. (Wiley Series in Probability and Statistics). Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316696

234. StataCorp. StataCorp. 2021.

235. Mostafa T, Wiggins RD. Handling attrition and non-response in the 1970 British Cohort Study. Centre for Longitudinal Studies; 2014. Report No.: 978-1-906929-79–4.

236. Katsoulis M, Narayanan M, Dodgeon B, Ploubidis G, Silverwood R. A data driven approach to address missing data in the 1970 British birth cohort [Internet]. 2024 [cited 2024 Sept 15]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.02.01.24302101

237. Steele, F. Multilevel models for longitudinal data. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2008;171(1):5–19.

238. Kwong ASF, López-López JA, Hammerton G, Manley D, Timpson NJ, Leckie G, et al. Genetic and Environmental Risk Factors Associated with Trajectories of Depression Symptoms from Adolescence to Young Adulthood. JAMA Network Open. 2019;2(6):1–14.

239. Akingbuwa WA, Hammerschlag AR, Jami ES, Allegrini AG, Karhunen V, Sallis H, et al. Genetic Associations Between Childhood Psychopathology and Adult Depression and Associated Traits in 42 998 Individuals: A Meta-analysis. JAMA Psychiatry. 2020 July 1;77(7):715.

240. Nivard MG, Gage SH, Hottenga JJ, van Beijsterveldt CEM, Abdellaoui A, Bartels M, et al. Genetic Overlap Between Schizophrenia and Developmental Psychopathology: Longitudinal and Multivariate Polygenic Risk Prediction of

Common Psychiatric Traits During Development. Schizophrenia Bulletin. 2017 Oct 21;43(6):1197–207.

241. Rice F, Riglin L, Thapar AK, Heron J, Anney R, O'Donovan MC, et al. Characterizing Developmental Trajectories and the Role of Neuropsychiatric Genetic Risk Variants in Early-Onset Depression. JAMA Psychiatry. 2019;76(3):306–13.

242. Akingbuwa WA, Hammerschlag AR, Bartels M, Middeldorp CM. Systematic Review: Molecular Studies of Common Genetic Variation in Child and Adolescent Psychiatric Disorders. Journal of the American Academy of Child & Adolescent Psychiatry. 2022 Feb;61(2):227–42.

243. Halldorsdottir T, Piechaczek C, De Matos APS, Czamara D, Pehl V, Wagenbuechler P, et al. Polygenic risk: Predicting depression outcomes in clinical and epidemiological cohorts of youths. American Journal of Psychiatry. 2019;176(8):615–25.

244. Cheesman R, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Purves KL, Pingault JB, Breen G, Rijsdij k F, et al. Extracting stability increases the SNP heritability of emotional problems in young people. Transl Psychiatry. 2018 Dec;8(1):223.

245. Jansen PR, Polderman TJC, Bolhuis K, Van Der Ende J, Jaddoe VWV, Verhulst FC, et al. Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. Child Psychology Psychiatry. 2018 Jan;59(1):39–47.

246. Su J, Kuo SIC, Bucholz KK, Edenberg HJ, Kramer JR, Schuckit M, et al. Understanding Mechanisms of Genetic Risk for Adolescent Internalizing and Externalizing Problems: The Mediating Role of Parenting and Personality. Twin Res Hum Genet. 2018 Aug;21(4):310–21.

247. Taylor MJ, Martin J, Lu Y, Brikell I, Lundström S, Larsson H, et al. Association of Genetic Risk Factors for Psychiatric Disorders and Traits of These Disorders in a Swedish Population Twin Sample. JAMA Psychiatry. 2019 Mar 1;76(3):280.

248. Ensink JBM, De Moor MHM, Zafarmand MH, De Laat S, Uitterlinden A, Vrijkotte TGM, et al. Maternal environmental risk factors and the development of internalizing and externalizing problems in childhood: The complex role of genetic factors. American J of Med Genetics Pt B. 2020 Jan;183(1):17–25.

249. Riglin L, Thapar AK, Leppert B, Martin J, Richards A, Anney R, et al. Using Genetics to Examine a General Liability to Childhood Psychopathology. Behav Genet. 2020 July;50(4):213–20.

250. Kwong ASF, Morris TT, Pearson RM, Timpson NJ, Rice F, Stergiakouli E, et al. Polygenic risk for depression, anxiety and neuroticism are associated with the severity and rate of change in depressive symptoms across adolescence. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2021;

251. Gao B, Song S, Guo J. Associations between life's simple 7 and incident depression among adults aged 50 years and older: A 15-year cohort study. Psychiatry Research. 2023 Feb;320:115046.

252. Schultebraucks K, Choi KW, Galatzer-Levy IR, Bonanno GA. Discriminating Heterogeneous Trajectories of Resilience and Depression after Major Life Stressors Using Polygenic Scores. JAMA Psychiatry. 2021;78(7):744–52.

253. Kosciuszko M, Steptoe A, Ajnakina O. Genetic propensity, socioeconomic status, and trajectories of depression over a course of 14 years in older adults. Translational Psychiatry. 2023;13(68):17.

254. Stringa N, Milaneschi Y, van Schoor NM, Suanet B, van der Lee S, Holstege H, et al. Genetic Liability for Depression, Social Factors and Their Interaction Effect in Depressive Symptoms and Depression Over Time in Older Adults. The American Journal of Geriatric Psychiatry. 2020 Aug;28(8):844–55.

255. Kendler KS, Gardner CO. Genetic and environmental influences on last-year major depression in adulthood: a highly heritable stable liability but strong environmental effects on 1-year prevalence. Psychol Med. 2017 July;47(10):1816–24.

256. Pingault JB, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nat Rev Genet. 2018 Sept;19(9):566–80.

257. Riglin L, Collishaw S, Richards A, Thapar AK, Rice F, Maughan B, et al. The impact of schizophrenia and mood disorder risk alleles on emotional problems: investigating change from childhood to middle age. Psychol Med. 2018 Oct;48(13):2153–8.

258. Streiner DL. Breaking up is Hard to Do: The Heartbreak of Dichotomizing Continuous Data. Can J Psychiatry. 2002 Apr;47(3):262–6.

259. Deyi BA, Kosinski AS, Snapinn SM. Power considerations when a continuous outcome variable is dichotomized. Journal of Biopharmaceutical Statistics. 1998 Jan 1;8(2):337–52.

260. Rodríguez MR, Nuevo R, Chatterji S, Ayuso-Mateos JL. Definitions and factors associated with subthreshold depressive conditions: a systematic review. BMC Psychiatry. 2012 Dec;12(1):181.

261. Li P, Stuart EA, Allison DB. Multiple Imputation: A Flexible Tool for Handling Missing Data. JAMA. 2015 Nov 10;314(18):1966.

262. Cerdá M, Sagdeo A, Johnson J, Galea S. Genetic and environmental influences on psychiatric comorbidity: A systematic review. Journal of Affective Disorders. 2010 Oct;126(1–2):14–38.

263. Plomin R, DeFries JC. Origins of individual differences in infancy: the Colorado adoption project. Orlando, Fla.: Acad. Pr; 1985. 406 p. (Developmental psychology series).

264. Abdellaoui A, Verweij KJH. Dissecting polygenic signals from genome-wide association studies on human behaviour. Nat Hum Behav. 2021 May 13;5(6):686–94.

265. Plomin R, DeFries JC, Knopik VS, Neiderhiser JM. Top 10 Replicated Findings From Behavioral Genetics. Perspect Psychol Sci. 2016 Jan;11(1):3–23.

266. Wolf S, Melo D, Garske KM, Pallares LF, Lea AJ, Ayroles JF. Characterizing the landscape of gene expression variance in humans. Cai JJ, editor. PLoS Genet. 2023 July 6;19(7):e1010833.

267. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. Cell. 2019 Mar;177(1):26–31.

268. Gondek D, Lacey RE, Blanchflower DG, Patalay P. How is the distribution of psychological distress changing over time? Who is driving these changes? Analysis of the 1958 and 1970 British birth cohorts. Soc Psychiatry Psychiatr Epidemiol. 2022 May;57(5):1007–16.

269. Tomitaka S, Kawasaki Y, Ide K, Akutagawa M, Ono Y, Furukawa TA. Distribution of psychological distress is stable in recent decades and follows an exponential pattern in the US population. Sci Rep. 2019 Aug 19;9(1):11982.

270. Keyes KM, Nicholson R, Kinley J, Raposo S, Stein MB, Goldner EM, et al. Age, Period, and Cohort Effects in Psychological Distress in the United States and Canada. American Journal of Epidemiology. 2014 May 15;179(10):1216–27.

271. Keyes KM, Gary D, O'Malley PM, Hamilton A, Schulenberg J. Recent increases in depressive symptoms among US adolescents: trends from 1991 to 2018. Soc Psychiatry Psychiatr Epidemiol. 2019 Aug;54(8):987–96.

272. Beller J. Age-period-cohort analysis of depression trends: are depressive symptoms increasing across generations in Germany? Eur J Ageing. 2022 Dec;19(4):1493–505.

273. Twenge JM, Cooper AB, Joiner TE, Duffy ME, Binau SG. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. Journal of Abnormal Psychology. 2019 Apr;128(3):185–99.

274. Okui T. An age-period-cohort analysis for prevalence of common psychiatric disorders in Japan, 1999–2017. Soc Psychiatry Psychiatr Epidemiol. 2021 Apr;56(4):639–48.

275. Perreault C. The Pace of Cultural Evolution. Mesoudi A, editor. PLoS ONE. 2012 Sept 14;7(9):e45150.

276. Gondek D, Lacey RE, Blanchflower DG, Patalay P. How is the distribution of psychological distress changing over time? Who is driving these changes? Analysis of the 1958 and 1970 British birth cohorts. 2021 June 1 [cited 2023 Sept 16]; Available from: http://medrxiv.org/lookup/doi/10.1101/2021.05.31.21257562

277. Kirkbride JB, Anglin DM, Colman I, Dykxhoorn J, Jones PB, Patalay P, et al. The social determinants of mental health and disorder: evidence, prevention and recommendations. World Psychiatry. 2024 Feb;23(1):58–90.

278. Zuckerman M, Li C, Diener EF. Societal Conditions and the Gender Difference in Well-Being: Testing a Three-Stage Model. Pers Soc Psychol Bull. 2017 Mar;43(3):329–36.

279. Homan P. Structural Sexism and Health in the United States: A New Perspective on Health Inequality and the Gender System. Am Sociol Rev. 2019 June;84(3):486–516.

280. McMunn A, Lacey R, Worts D, Kuh D, McDonough P, Sacker A. Work-family life courses and psychological distress: Evidence from three British birth cohort studies. Advances in Life Course Research. 2021 Dec;50:100429.

281. Xue B, McMunn A. Gender differences in unpaid care work and psychological distress in the UK Covid-19 lockdown. Tran TD, editor. PLoS ONE. 2021 Mar 4;16(3):e0247959.

282. Tao W, Janzen BL, Abonyi S. Gender, Division of Unpaid Family Work and Psychological Distress in Dual-Earner Families. Clin Pract Epidemiol Ment Health. 2010 June 18;6(1):36–46.

283. Patalay P, Fitzsimons E. Correlates of Mental Illness and Wellbeing in Children: Are They the Same? Results From the UK Millennium Cohort Study. Journal of the American Academy of Child and Adolescent Psychiatry. 2016;55(9):771–83.

284. Piketty T, Goldhammer A. Capital in the twenty-first century. Cambridge Massachusetts: The Belknap Press of Harvard University Press; 2014. 685 p.

285. Bambra C, Barr B, Milne E. North and South: Addressing the English health divide. Journal of Public Health (United Kingdom). 2014;36(2):183–6.

286. Dorling D. Shattered nation: inequality and the geography of a failing state. Brooklyn, NY: Verso Books; 2023.

287. Case A, Deaton A. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. Proc Natl Acad Sci USA. 2015 Dec 8;112(49):15078–83.

288. Mackenbach JP. Health inequalities: persistence and change in European welfare states. First edition. Oxford: Oxford University Press; 2019. 240 p.

289. Patel V, Burns JK, Dhingra M, Tarver L, Kohrt BA, Lund C. Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms. World Psychiatry. 2018;17(1):76–89.

290. Tibber MS, Walji F, Kirkbride JB, Huddy V. The association between income inequality and adult mental health at the subnational level—a systematic review. Soc Psychiatry Psychiatr Epidemiol. 2022 Jan;57(1):1–24.

291. Ribeiro WS, Bauer A, Andrade MCR, York-Smith M, Pan PM, Pingani L, et al. Income inequality and mental illness-related morbidity and resilience: a systematic review and meta-analysis. The Lancet Psychiatry. 2017 July;4(7):554–62.

292. Lund C, Brooke-Sumner C, Baingana F, Baron EC, Breuer E, Chandra P, et al. Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. The Lancet Psychiatry. 2018;5(4):357–69.

293. World Bank. World Bank Gini Index [Internet]. 2024. Available from: https://data.worldbank.org/indicator/SI.POV.GINI?locations=GB&most_recent_year_desc=true&view=map&year=1989

294. Tom Clark, Mike Elsby, Sarah Love. Twenty-five years of falling investment? Trends in capital spending on public services [Internet]. Institute for Fiscal Studies; 2001. Report No.: Briefing Note No. 20. Available from: https://ifs.org.uk/sites/default/files/output_url_files/bn20.pdf

295. Sullivan A, Bann D. Generation X enters middle age. LLCS [Internet]. 2015 Apr 30 [cited 2024 Mar 12];6(2). Available from: http://www.llcsjournal.org/index.php/llcs/article/view/351/368

296. Coyle D, Nakamura L. Toward a Framework for Time Use, Welfare, and Household-Centric Economic Measurement [Internet]. Federal Reserve Bank of Philadelphia; 2019 Feb [cited 2024 Mar 22] p. 19–11. (Working paper (Federal Reserve Bank of Philadelphia)). Report No.: 19–11. Available from: https://www.philadelphiafed.org/-/media/frbp/assets/working-papers/2019/wp19-11.pdf

297. Niedzwiedz CL, Katikireddi SV, Pell JP, Mitchell R. Life course socio-economic position and quality of life in adulthood: a systematic review of life course models. BMC Public Health. 2012 Dec;12(1):628.

298. Sen A. Mortality as an Indicator of Economic Success and Failure. The Economic Journal. 1998 Jan 1;108(446):1–25.

299. Mousteri V, Daly M, Delaney L. Underemployment and psychological distress: Propensity score and fixed effects estimates from two large UK samples. Social Science & Medicine. 2020 Jan;244:112641.

300. Barbalat G, Franck N. Ecological study of the association between mental illness with human development, income inequalities and unemployment across OECD countries. BMJ Open. 2020;10(4):1–7.

301. Office for National Statistics. Transition from a manufacturing to service led labour market over past 170 years [Internet]. 2015 [cited 2024 Mar 26]. Available from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/transitionfromamanufacturingtoserviceledlabourmarketoverpast170years/2015-08-06

302. Rowthorn R. Kalecki Centenary Lecture the Political Economy of Full Employment in Modern Britain. Oxf Bull Econ Stat. 2000 May;62(2):139–73.

303. Larsen CA. The Rise and Fall of Social Cohesion: The Construction and De-construction of Social Trust in the US, UK, Sweden and Denmark [Internet]. Oxford University Press; 2013 [cited 2024 Mar 28]. Available from: https://academic.oup.com/book/1554

304. United Nations. Social cohesion concept and measurement. Geneva: United Nations; 2023.

305. McGowan VJ, Akhter N, Halliday E, Popay J, Kasim A, Bambra C. Collective control, social cohesion and health and well-being: baseline survey results from the communities in control study in England. Journal of Public Health. 2021;1–9.

306. Athanasiadis G, Meijsen JJ, Helenius D, Schork AJ, Ingason A, Thompson WK, et al. A comprehensive map of genetic relationships among diagnostic categories based on 48.6 million relative pairs from the Danish genealogy. Proc Natl Acad Sci USA. 2022 Feb 8;119(6):e2118688119.

307. Machlitt-Northen S, Keers R, Munroe P, Howard D, Pluess M. Gene–Environment Correlation over Time: A Longitudinal Analysis of Polygenic Risk Scores for Schizophrenia and Major Depression in Three British Cohorts Studies. Genes. 2022 June 24;13(7):1136.

308. Zhao L, Han G, Zhao Y, Jin Y, Ge T, Yang W, et al. Gender Differences in Depression: Evidence From Genetics. Front Genet. 2020 Oct 15;11:562316.

309. Uddin M, Koenen KC, De Los Santos R, Bakshis E, Aiello AE, Galea S. Gender differences in the genetic and environmental determinants of adolescent depression. Depress Anxiety. 2010 Mar 24;27(7):658–66.

310. Border R, Johnson EC, Evans LM, Smolen A, Berley N, Sullivan PF, et al. No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. AJP. 2019 May;176(5):376–87.

311. Hannigan LJ, Walaker N, Waszczuk MA, McAdams TA, Eley TC. Aetiological Influences on Stability and Change in Emotional and Behavioural Problems across Development: A Systematic Review. Psychopathology Review. 2017 Mar;a4(1):52–108.

312. Mayhew AJ, Meyre D. Assessing the Heritability of Complex Traits in Humans: Methodological Challenges and Opportunities. CG [Internet]. 2017 July 26 [cited 2024 Mar 28];18(4). Available from: http://www.eurekaselect.com/150709/article

313. Gale CR, Deary IJ, Stafford M. A life course approach to psychological and social wellbeing. In: Kuh D, Cooper R, Hardy R, Richards M, editors. A Life Course Approach to Healthy Ageing [Internet]. Oxford University Press; 2013 [cited 2022 Mar 9]. p. 46–62. Available from:

https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199656516.001.0001/acprof-9780199656516-chapter-4

314. Jiang X, Zhang MJ, Zhang Y, Durvasula A, Inouye M, Holmes C, et al. Age-dependent topic modelling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk [Internet]. Genetic and Genomic Medicine; 2022 Oct [cited 2023 May 15]. Available from: http://medrxiv.org/lookup/doi/10.1101/2022.10.23.22281420

315. Harder A, Nguyen TD, Pasman JA, Mosing MA, Hägg S, Lu Y. Genetics of age-at-onset in major depression. Transl Psychiatry. 2022 Dec;12(1):124.

316. Hewitt J, Walters M, Padmanabhan S, Dawson J. Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease. BMJ Open. 2016 Mar;6(3):e009161.

317. Martin J, Khramtsova EA, Goleva SB, Blokland GAM, Traglia M, Walters RK, et al. Examining Sex-Differentiated Genetic Effects Across Neuropsychiatric and Behavioral Traits. Biological Psychiatry. 2021 June;89(12):1127–37.

318. Li L, Hardy R, Kuh D, Power C. Life-course body mass index trajectories and blood pressure in mid life in two British birth cohorts: stronger associations in the later-born generation. Int J Epidemiol. 2015 June;44(3):1018–26.

319. Conley D, Fletcher J. The genome factor: what the social genomics revolution reveals about ourselves, our history, and the future. Princeton Oxford: Princeton University Press; 2017. 282 p.

320. Huitfeldt A. Is caviar a risk factor for being a millionaire?: Table 1. BMJ. 2016 Dec 9;i6536.

321. Hazell M, Thornton E, Haghparast-Bidgoli H, Patalay P. Socio-economic inequalities in adolescent mental health in the UK: Multiple socio-economic indicators and reporter effects. SSM - Mental Health. 2022 Dec;2:100176.

322. Alegría M, NeMoyer A, Falgàs Bagué I, Wang Y, Alvarez K. Social Determinants of Mental Health: Where We Are and Where We Need to Go. Current Psychiatry Reports. 2018;20(11).

323. Dohrenwend BP, Levav I, Shrout PE, Schwartz S, Naveh G, Link BG, et al. Socioeconomic Status and Psychiatric Disorders: The Causation-Selection Issue. Science. 1992 Feb 21;255(5047):946–52.

324. Dougall I, Vasiljevic M, Wright JD, Weick M. How, when, and why is social class linked to mental health and wellbeing? A systematic meta-review. Social Science & Medicine. 2024 Feb;343:116542.

325. Keers R, Pluess M. Childhood quality influences genetic sensitivity to environmental influences across adulthood: A life-course Gene × Environment interaction study. Dev Psychopathol. 2017 Dec;29(5):1921–33.

326. Machlitt-Northen S, Keers R, Munroe PB, Howard DM, Pluess M. Polygenic risk scores for schizophrenia and major depression are associated with socio-economic indicators of adversity in two British community samples. Transl Psychiatry. 2022 Nov 14;12(1):477.

327. Leete R, Fox J. Registrar General's Social Classes: Origins and Uses. Population Trends. 1977;8:1–7.

328. Kok AAL, Huisman M, Cooper R, Cosco TD, Deeg DJH, Kuh D, et al. Lifetime trajectories of socio-economic adversity and their associations with psychosocial factors and attitudes towards social class. Longitudinal and Life Course Studies. 2020;11(1):81–104.

329. Goldthorpe JH. Social class mobility in modern Britain: changing structure, constant process. JBA. 2016 Aug 18;4:89–111.

330. Marshall G, Newby H, Rose D, Vogler C. Social Class in Modern Britain [Internet]. 0 edn. Routledge; 2005 [cited 2024 Dec 3]. Available from: https://www.taylorfrancis.com/books/9781134858941

331. Galobardes B. Indicators of socioeconomic position (part 1). Journal of Epidemiology & Community Health. 2006 Jan 1;60(1):7–12.

332. Bukodi E, Goldthorpe JH, Waller L, Kuha J. The mobility problem in Britain: new findings from the analysis of birth cohort data: The mobility problem in Britain. The British Journal of Sociology. 2015 Mar;66(1):93–117.

333. Erikson R, Goldthorpe JH. Has social mobility in Britain decreased? Reconciling divergent findings on income and class mobility: Has social mobility in Britain decreased? The British Journal of Sociology. 2010 June;61(2):211–30.

334. Jiménez-Solomon O, Garfinkel I, Wall M, Wimer C. When money and mental health problems pile up: The reciprocal relationship between income and psychological distress. SSM - Population Health. 2024 Mar;25:101624.

335. Gustavson K, Von Soest T, Karevold E, Røysamb E. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. BMC Public Health. 2012 Dec;12(1):918.

336. Jerrim. Measuring socio-economic background using administrative data. What is the best proxy available? [Internet]. UCL: Social Research Institute; 2020. (Quantitative Social Science Working Paper No. 20-09). Available from: http://repec.ioe.ac.uk/REPEc/pdf/qsswp2009.pdf

337. Currie J. Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. Journal of Economic Literature. 2009 Mar 1;47(1):87–122.

338. Dodgeon B, Morris T, Crawford C, Parsons S, Vignoles A, Oldfield Z, et al. CLOSER work package 2: Harmonised socio-economic measures user guide. London: CLOSER. 2019;(August).

339. Assary E, Vincent J, Machlitt-Northen S, Keers R, Pluess M. The Role of Gene-Environment Interaction in Mental Health and Susceptibility to the Development of Psychiatric Disorders. In: Teperino R, editor. Beyond Our Genes [Internet]. Cham: Springer International Publishing; 2020 [cited 2024 Apr 27]. p. 117–38. Available from: http://link.springer.com/10.1007/978-3-030-35213-4_7

340. Qi X, Yang J, Liu L, Hao J, Pan C, Wen Y, et al. Socioeconomic inequalities, genetic susceptibility, and risks of depression and anxiety: A large-observational study. Journal of Affective Disorders. 2024 Dec;367:174–83.

341. Agerbo E, Trabjerg BB, Børglum AD, Schork AJ, Vilhjálmsson BJ, Pedersen CB, et al. Risk of Early-Onset Depression Associated with Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. JAMA Psychiatry. 2021;78(4):387–97.

342. Hoang CT, Amin V, Behrman JR, Kohler HP, Kohler IV. Heterogenous trajectories in physical, mental and cognitive health among older Americans: Roles of genetics and life course contextual factors. SSM - Population Health. 2023 Sept;23:101448.

343. Smith GD. Epidemiology, epigenetics and the 'Gloomy Prospect': Embracing randomness in population health research and practice. International Journal of Epidemiology. 2011;40(3):537–62.

344. Rothman KJ, Greenland S, Walker AM. Concepts of Interaction. American Journal of Epidemiology. 1980 Oct;112(4):467–70.

345. Avison WR, Turner RJ. Stressful Life Events and Depressive Symptoms: Disaggregating the Effects of Acute Stressors and Chronic Strains. Journal of Health and Social Behavior. 1988 Sept;29(3):253.

346. Assary E, Krebs G, Eley TC. Practitioner Review: Differential susceptibility theory: might it help in understanding and treating mental health problems in youth? Child Psychology Psychiatry. 2023 Aug;64(8):1104–14.

347. Duncan LE, Keller MC. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. AJP. 2011 Oct;168(10):1041–9.

348. Becker J, Burik CAP, Goldman G, Wang N, Jayashankar H, Bennett M, et al. Resource profile and user guide of the Polygenic Index Repository. Nat Hum Behav. 2021 June 17;5(12):1744–58.

349. Lam JR, Tyler J, Scurrah KJ, Reavley NJ, Dite GS. The Association between Socioeconomic Status and Psychological Distress: A Within and Between Twin Study. Twin Research and Human Genetics.

350. Sheikh MA, Abelsen B, Olsen JA. Differential Recall Bias, Intermediate Confounding, and Mediation Analysis in Life Course Epidemiology: An Analytic Framework with Empirical Example. Front Psychol [Internet]. 2016 Nov 23 [cited 2024 Nov 28];7. Available from: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01828/full

351. Zhang A, Gagné T, Walsh D, Ciancio A, Proto E, McCartney G. Trends in psychological distress in Great Britain, 1991–2019: evidence from three representative surveys. J Epidemiol Community Health. 2023 July;77(7):468–73.

352. Bolton P, Lewis. Equality of access and outcomes in higher education in England [Internet]. House of Commons Library; 2024. Available from: https://researchbriefings.files.parliament.uk/documents/CBP-9195/CBP-9195.pdf

353. Howarth J. Women in Britain: voices and perspectives from twentieth-century history. London: I.B. Tauris; 2019.

354. Carnevali F, Strange JM, editors. 20th Century Britain: Economic, Cultural and Social Change [Internet]. 0 edn. Routledge; 2014 [cited 2024 Dec 5]. Available from: https://www.taylorfrancis.com/books/9781317868378

355. Sweeting H, Bhaskar A, Benzeval M, Popham F, Hunt K. Changing gender roles and attitudes and their implications for well-being around the new millennium. Soc Psychiatry Psychiatr Epidemiol. 2014 May;49(5):791–809.

356. Chauvel L, Bar Haim E, Hartung A, Murphy E. Rewealthization in twenty-first century Western countries: the defining trend of the socioeconomic squeeze of the middle class. J Chin Sociol. 2021 Dec;8(1):4.

357. Price T, McGowan V, Visram S, Wildman J, Bambra C. “They’re not mentally ill, their lives are just shit”: Stakeholders’ understanding of deaths of despair in a deindustrialised community in North East England. Health & Place. 2024 Nov;90:103346.

358. Van Hootegem A, Rogne AF, Lyngstad TH. Heritability of class and status: Implications for sociological theory and research. Research in Social Stratification and Mobility. 2024 Aug;92:100940.

359. Marioni RE, Davies G, Hayward C, Liewald D, Kerr SM, Campbell A, et al. Molecular genetic contributions to socioeconomic status and intelligence. Intelligence. 2014 May;44:26–32.

360. Machlitt-Northen S, Keers R, Munroe PB, Howard DM, Trubetskoy V, Pluess M. Polygenic scores for schizophrenia and major depression are associated with psychosocial risk factors in children: evidence of gene–environment correlation. Child Psychology Psychiatry. 2022 Oct;63(10):1140–52.

361. Agerbo E, Trabjerg BB, Børglum AD, Schork AJ, Vilhjálmsson BJ, Pedersen CB, et al. Risk of Early-Onset Depression Associated With Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. JAMA Psychiatry. 2021 Apr 1;78(4):387.

362. Marees AT, Smit DJA, Abdellaoui A, Nivard MG, Van Den Brink W, Denys D, et al. Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits. Nat Hum Behav. 2021 Mar 8;5(8):1065–73.

363. On the use and interpretation of certain test criteria for purposes of statistical inference. Part I. In: Joint Statistical Papers [Internet]. University of California Press; 1967 [cited 2024 Dec 5]. p. 1–66. Available from: https://www.degruyter.com/document/doi/10.1525/9780520339897-003/html

364. Burnham KP, Anderson DR. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research. 2004 Nov;33(2):261–304.

365. Cox DR. Interaction. International Statistical Review / Revue Internationale de Statistique. 1984 Apr;52(1):1.

366. Mackenbach JP, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: An overview of available measures illustrated with two examples from Europe. Social Science & Medicine. 1997 Mar;44(6):757–71.

367. Sergeant JC. Relative index of inequality: definition, estimation, and inference. Biostatistics. 2005 Aug 3;7(2):213–24.

368. Egger M, Smith GD. Meta-analysis: Potentials and promise. BMJ. 1997 Nov 22;315(7119):1371–4.

369. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011 Feb 10;342(feb10 2):d549–d549.

370. Von Hippel PT. The heterogeneity statistic I2 can be biased in small meta-analyses. BMC Med Res Methodol. 2015 Dec;15(1):35.

371. Yang Z, Pajuste FD, Zguro K, Cheng Y, Kurant DE, Eoli A, et al. Limited overlap between genetic effects on disease susceptibility and disease survival [Internet]. Genetic and Genomic Medicine; 2023 Oct [cited 2023 Oct 26]. Available from: http://medrxiv.org/lookup/doi/10.1101/2023.10.10.23296544

372. Torvik FA, Eilertsen EM, Hannigan LJ, Cheesman R, Howe LJ, Magnus P, et al. Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. Nat Commun. 2022 Mar 1;13(1):1108.

373. Erola J, Lehti H, Baier T, Karhula A. Socioeconomic Background and Gene–Environment Interplay in Social Stratification across the Early Life Course. European Sociological Review. 2022 Jan 20;38(1):1–17.

374. Zhou Q, Gidziela A, Allegrini AG, Cheesman R, Wertz J, Maxwell J, et al. Gene-environment correlation: the role of family environment in academic development. Mol Psychiatry [Internet]. 2024 Sept 4 [cited 2024 Dec 4]; Available from: https://www.nature.com/articles/s41380-024-02716-0

375. Pingault JB, Rijsdijk F, Schoeler T, Choi SW, Selzam S, Krapohl E, et al. Genetic sensitivity analysis: Adjusting for genetic confounding in epidemiological associations. Zhu X, editor. PLoS Genet. 2021 June 11;17(6):e1009590.

376. Howe LJ, Nivard MG, Morris TT, Hansen AF, Rasheed H, Cho Y, et al. Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. Nat Genet. 2022 May;54(5):581–92.

377. Wang Z, Grosvenor L, Ray D, Ruczinski I, Beaty TH, Volk H, et al. Estimation of Direct and Indirect Polygenic Effects and Gene-Environment Interactions using Polygenic Scores in Case-Parent Trio Studies [Internet]. 2024 [cited 2024 Dec 4]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.10.08.24315066

378. Kim J, Ziyatdinov A, Laville V, Hu FB, Rimm E, Kraft P, et al. Joint Analysis of Multiple Interaction Parameters in Genetic Association Studies. Genetics. 2019 Feb 1;211(2):483–94.

379. Momen NC, Beck C, Lousdal ML, Agerbo E, McGrath JJ, Pedersen CB, et al. Mental Health Disorder Trends in Denmark According to Age, Calendar Period, and Birth Cohort. JAMA Psychiatry. 2025 Feb 1;82(2):161.

380. Zheng B, Fletcher JM, Song J, Lu Q. Analysis of Sex-Specific Gene-by-Cohort and Genetic Correlation-by-Cohort Interaction in Educational and Reproductive Outcomes Using the UK Biobank Data. J Health Soc Behav. 2024 Sept;65(3):432–48.

381. Roberts MT. Globalization and Neoliberalism: Structural Determinants of Global Mental Health? Humanity & Society. 2021 Nov;45(4):471–508.

382. Arnett JJ, Žukauskienė R, Sugimura K. The new life stage of emerging adulthood at ages 18–29 years: implications for mental health. The Lancet Psychiatry. 2014 Dec;1(7):569–76.

383. Feldman MW, Lewontin RC. The Heritability Hang-Up: The role of variance analysis in human genetics is discussed. Science. 1975 Dec 19;190(4220):1163–8.

384. Hannigan LJ, McAdams TA, Eley TC. Developmental change in the association between adolescent depressive symptoms and the home environment: results from a longitudinal, genetically informative investigation. J Child Psychol Psychiatr. 2017 July;58(7):787–97.

385. Molenaar PCM, Boomsma DI, Dolan CV. A third source of developmental differences. Behav Genet. 1993 Nov;23(6):519–24.

386. Griffiths PE, Tabery J. Developmental Systems Theory. In: Advances in Child Development and Behavior [Internet]. Elsevier; 2013 [cited 2025 Feb 19]. p. 65–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780123979476000039

387. Lux V. With Gottlieb beyond Gottlieb: The Role of Epigenetics in Psychobiological Development. International Journal of Developmental Science. 2013;7(2):69–78.

388. Gottlieb G. Experiential canalization of behavioral development: Theory. Developmental Psychology. 1991 Jan;27(1):4–13.

389. Waddington CH. The Epigenotype. Int J Epidemiol. 2012 Feb;41(1):10–3.

390. Bronfenbrenner U. Developmental Ecology Through Space and Time: A Future Perspective. In: Moen P, Elder GH, Lüscher K, editors. Examining lives in context: Perspectives on the ecology of human development. 1st edn American Psychological Association; 1995. p. 619–49.

391. Baum A, Garofalo JP, Yali AM. Socioeconomic status and chronic stress. Does stress account for SES effects on health? Annals of the New York Academy of Sciences. 1999;896:131–44.

392. Trahan LH, Stuebing KK, Fletcher JM, Hiscock M. The Flynn effect: A meta-analysis. Psychological Bulletin. 2014;140(5):1332–60.

393. Filby E. Inheritocracy: It's Time to Talk about the Bank of Mum and Dad. 1st ed. London: Biteback Publishing; 2024. 1 p.

394. Bourquin P, Brewer M, Wernham T. Trends in income and wealth inequalities. Oxford Open Economics. 2024 July 5;3(Supplement_1):i103–46.

395. Huang W, Carbone MA, Lyman RF, Anholt RRH, Mackay TFC. Genotype by environment interaction for gene expression in Drosophila melanogaster. Nat Commun. 2020 Oct 28;11(1):5451.

396. Kang EY, Han B, Furlotte N, Joo JWJ, Shih D, Davis RC, et al. Meta-Analysis Identifies Gene-by-Environment Interactions as Demonstrated in a Study of 4,965 Mice. Gibson G, editor. PLoS Genet. 2014 Jan 9;10(1):e1004022.

397. Napier JD, Heckman RW, Juenger TE. Gene-by-environment interactions in plants: Molecular mechanisms, environmental drivers, and adaptive plasticity. The Plant Cell. 2023 Jan 2;35(1):109–24.

398. Ma J, Thabane L, Beyene J, Raina P. Power Analysis for Population-Based Longitudinal Studies Investigating Gene-Environment Interactions in Chronic Diseases: A Simulation Study. Olivier J, editor. PLoS ONE. 2016 Feb 22;11(2):e0149940.

399. Kendler KS. Psychiatric Genetics: A Methodologic Critique. AJP. 2005 Jan 1;162(1):3–11.

400. Mitchell KJ, Cheney N. The Genomic Code: The genome instantiates a generative model of the organism [Internet]. arXiv; 2024 [cited 2025 Feb 19]. Available from: https://arxiv.org/abs/2407.15908

401. Xu Z, Shen X, Pan W, for the Alzheimer's Disease Neuroimaging Initiative. Longitudinal Analysis Is More Powerful than Cross-Sectional Analysis in Detecting Genetic Association with Neuroimaging Phenotypes. Chen L, editor. PLoS ONE. 2014 Aug 6;9(8):e102312.

402. Talens RP, Christensen K, Putter H, Willemsen G, Christiansen L, Kremer D, et al. Epigenetic variation during the adult lifespan: cross-sectional and

longitudinal data on monozygotic twin pairs. Aging Cell. 2012 Aug;11(4):694–703.

403. Kristjansson SD, Kircher JC, Webb AK. Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling. Psychophysiology. 2007 Sept;44(5):728–36.

404. Mcelroy E, Villadsen A, Patalay P, Goodman A, Richards M, Northstone K, et al. Harmonisation and measurement properties of mental health measures in six British cohorts. London: CLOSER; 2021.

405. Ploubidis GB. Harmonisation of mental health measures in British birth cohorts [Internet]. London: CLOSER. 2021 [cited 2021 May 31]. Available from: https://www.closer.ac.uk/research-fund-2/data-harmonisation/harmonisation-mental-health-measures-british-birth-cohorts/

406. Adhikari K, Patten SB, Patel AB, Premji S, Tough S, Letourneau N, et al. Data harmonization and data pooling from cohort studies: a practical approach for data management. Int J Popul Data Sci. 2021;6(1):1680.

407. Cheng C, Messerschmidt L, Bravo I, Waldbauer M, Bhavikatti R, Schenk C, et al. A General Primer for Data Harmonization. Sci Data. 2024 Jan 31;11(1):152.

408. Mazza GL, Enders CK, Ruehlman LS. Addressing Item-Level Missing Data: A Comparison of Proration and Full Information Maximum Likelihood Estimation. Multivariate Behavioral Research. 2015 Sept 3;50(5):504–19.

409. Rosenberg NA, Edge MD, Pritchard JK, Feldman MW. Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. Evolution, Medicine, and Public Health. 2019 Jan 1;2019(1):26–34.

410. Plomin R, Von Stumm S. Polygenic scores: prediction versus explanation. Mol Psychiatry. 2022 Jan;27(1):49–52.

411. Wang JY, Lin N, Zietz M, Mares J, Narasimhan VM, Rathouz PJ, et al. Three Open Questions in Polygenic Score Portability [Internet]. 2024 [cited 2025 Feb 21]. Available from: http://biorxiv.org/lookup/doi/10.1101/2024.08.20.608703

412. Nat Genet. Genetics for all. Nat Genet. 2019 Apr;51(4):579–579.

413. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GÅ, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nature Communications. 2020;11(1):1–13.

414. Nivard MG, Belsky D, Harden KP, Baier T, Ystrom E, van Bergen E, et al. Neither nature nor nurture: Using extended pedigree data to elucidate the origins of indirect genetic effects on offspring educational outcomes. [Internet]. PsyArXiv; 2022 Feb [cited 2023 Jan 17]. Available from: https://osf.io/bhpm5

415. Nivard MG, Belsky DW, Harden KP, Baier T, Andreassen OA, Ystrøm E, et al. More than nature and nurture, indirect genetic effects on children's academic

achievement are consequences of dynastic social processes. Nat Hum Behav. 2024 Jan 15;8(4):771–8.

416. Centre for Longitudinal Studies. Centre for Longitudinal Studies: 1970. 2025. 1970 British Cohort Study Age 51 data now available. Available from: https://cls.ucl.ac.uk/1970-british-cohort-study-age-51-data-now-available/

417. Kuh D, Pierce M, Adams J, Deanfield J, Ekelund U, Friberg P, et al. Cohort Profile: Updating the cohort profile for the MRC National Survey of Health and Development: a new clinic-based data collection for ageing research. International Journal of Epidemiology. 2011 Feb 1;40(1):e1–9.

418. Allegrini AG, Cheesman R, Rimfeld K, Selzam S, Pingault J, Eley TC, et al. The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. Child Psychology Psychiatry. 2020 Jan;61(1):30–9.

419. Connelly R, Platt L. Cohort profile: UK Millennium Cohort Study (mcs). International Journal of Epidemiology. 2014;43(6):1719–25.

420. Wu AFW, Henderson M, Brown M, Adali T, Silverwood RJ, Peycheva D, et al. Cohort Profile: Next Steps—the longitudinal study of people in England born in 1989–90. International Journal of Epidemiology. 2024 Oct 13;53(6):dyae152.

421. Botha F, Morris RW, Butterworth P, Glozier N. Generational differences in mental health trends in the twenty-first century. Proc Natl Acad Sci USA. 2023 Dec 5;120(49):e2303781120.

422. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort Profile: The English Longitudinal Study of Ageing. International Journal of Epidemiology. 2013 Dec 1;42(6):1640–8.

423. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). International Journal of Epidemiology. 2014 Apr 1;43(2):576–85.

424. Lee LO, Gatz M, Pedersen NL, Prescott CA. Anxiety trajectories in the second half of life: Genetic and environmental contributions over age. Psychology and Aging. 2016 Feb;31(1):101–13.

425. Petkus AJ, Gatz M, Reynolds CA, Kremen WS, Wetherell JL. Stability of Genetic and Environmental Contributions to Anxiety Symptoms in Older Adulthood. Behav Genet. 2016 July;46(4):492–505.

426. Herle M, Micali N, Abdulkadir M, Loos R, Bryant-Waugh R, Hübel C, et al. Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. Eur J Epidemiol. 2020 Mar;35(3):205–22.

427. Moulton V, Sullivan A, Patalay P, Fitzsimons E, Henderson M, Bann D, et al. Association between psychological distress trajectories from adolescence to midlife and mental health during the pandemic: evidence from two British birth cohorts. Psychol Med. 2023 Oct;53(14):6547–59.

428. McNeish D, Wolf MG. Thinking twice about sum scores. Behav Res. 2020 Dec;52(6):2287–305.

429. Fried EI, Nesse RM. Depression sum-scores don’t add up: why analyzing specific depression symptoms is essential. BMC Med. 2015 Dec;13(1):72.

430. Schaakxs R, Comijs HC, Lamers F, Beekman ATF, Penninx BWJH. Age-related variability in the presentation of symptoms of major depressive disorder. Psychol Med. 2017 Feb;47(3):543–52.

431. Piazza GG, Allegrini AG, Eley TC, Epskamp S, Fried E, Isvoranu AM, et al. Polygenic Scores and Networks of Psychopathology Symptoms. JAMA Psychiatry [Internet]. 2024 June 12 [cited 2024 June 25]; Available from: https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2819863