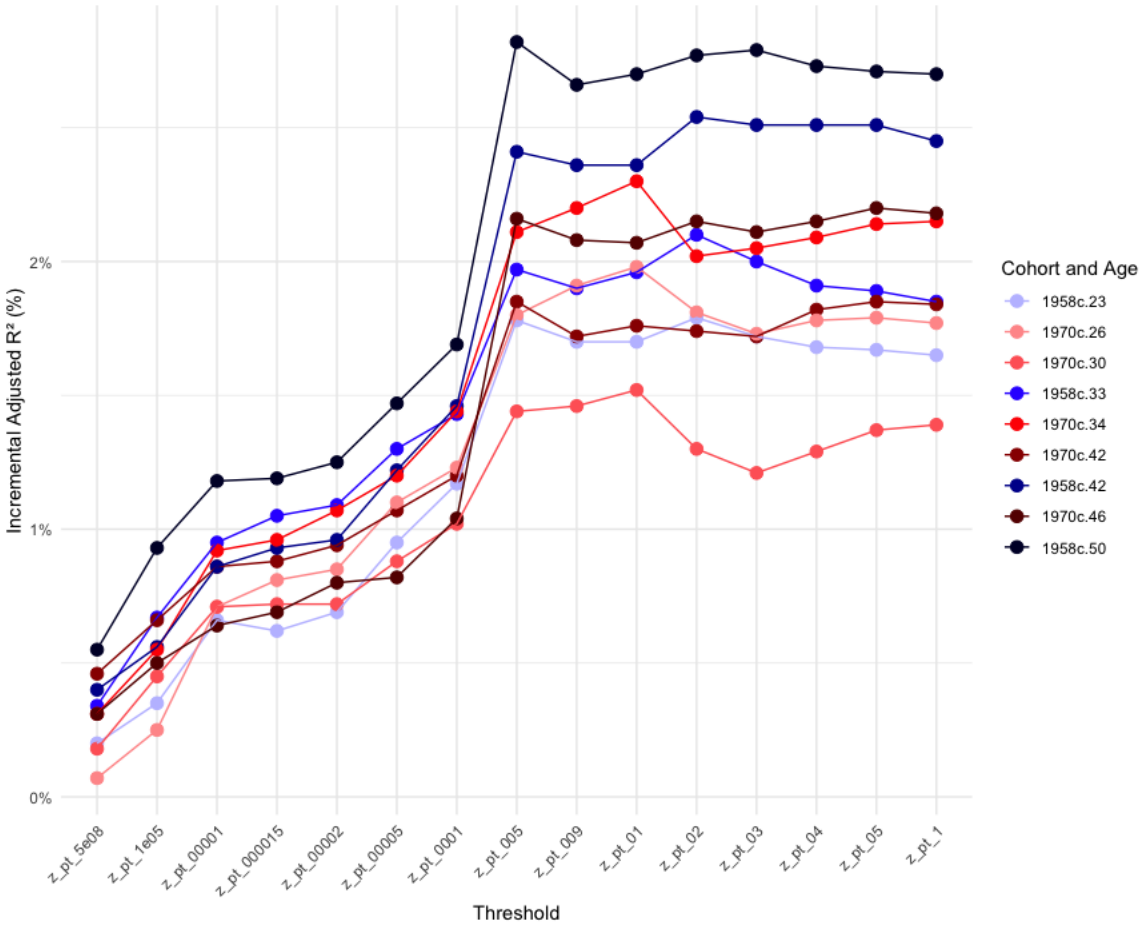
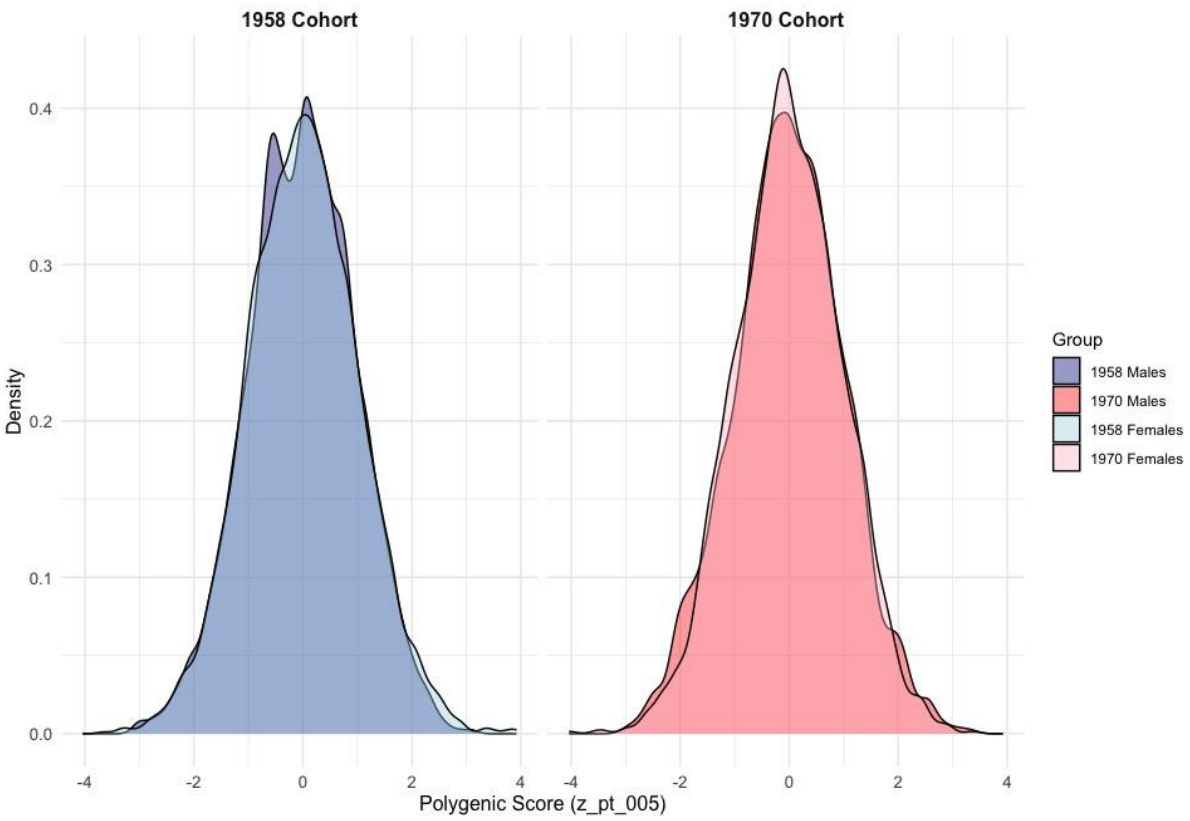
Supplementary Materials

Supplementary Table 1: Mean and standard deviation of the malaise inventory score overall and by sex in 1958c and 1970c

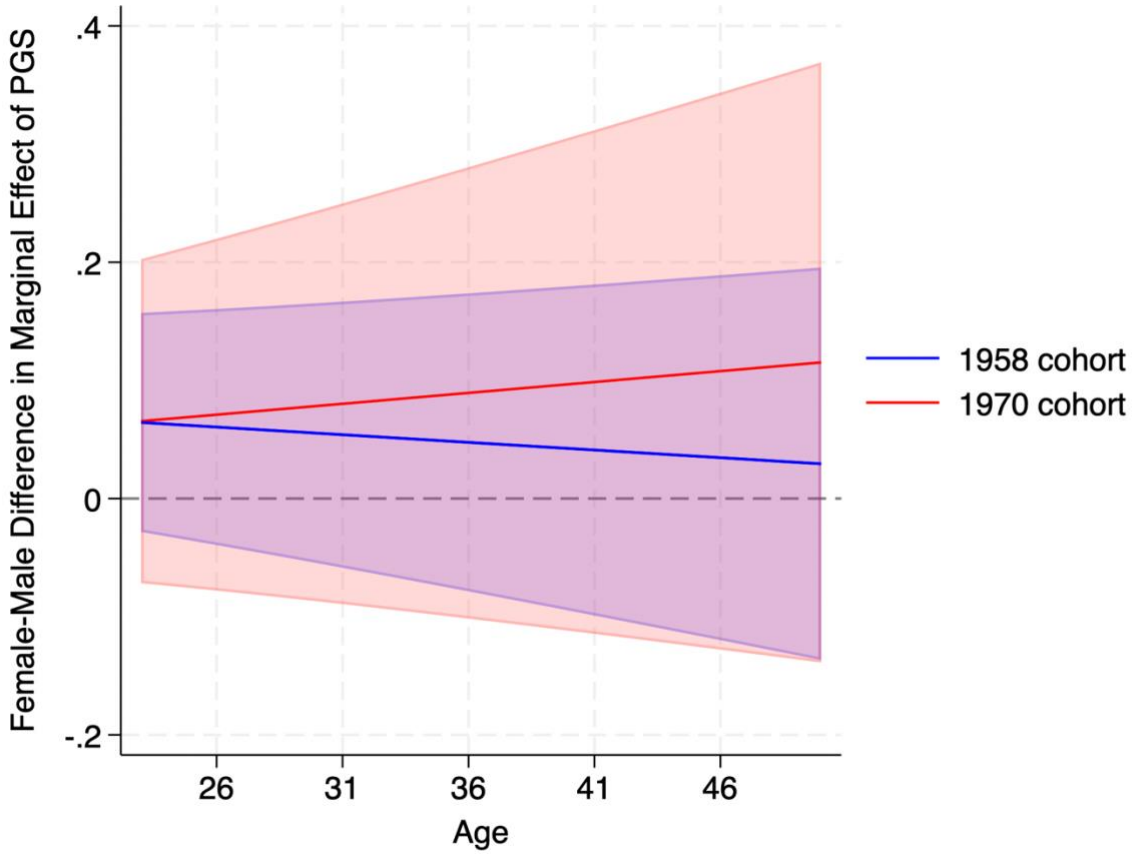
		Overall		Female		Male	
Cohort	Age	Mean	SD	Mean	SD	Mean	SD
1958	23	1.15	1.50	1.54	1.66	0.75	1.19
1970	26	1.63	1.65	1.97	1.71	1.20	1.48
1970	30	1.43	1.65	1.67	1.72	1.15	1.51
1958	33	0.94	1.46	1.22	1.63	0.64	1.20
1970	34	1.52	1.77	1.76	1.84	1.24	1.65
1958	42	1.49	1.74	1.80	1.79	1.22	1.63
1970	42	1.74	1.89	1.85	1.90	1.30	1.69
1970	46	1.68	2.06	1.90	2.13	1.44	1.94
1958	50	1.42	1.89	1.73	2.04	1.12	1.68



Supplementary Figure 1: Specification curve plot 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 1958c and 1970c



Supplementary Figure 2: Density plots of polygenic scores by sex in each cohort



Supplementary Figure 3: Difference in the marginal effects of polygenic scores between males and females in the 1958 and 1970 cohorts

Supplementary Methods

Genetic data, imputation, and quality control

1958 cohort

Genotype data for the 1958 National Child Development Study were collected during the biomedical sweep (ages 44–45). Genotyping was conducted across multiple arrays and subjected to standard quality control procedures prior to imputation, including exclusion of individuals with high missingness (>2%), sex discordance, excess heterozygosity (>3 SD), and relatedness (KING cutoff 0.0884), as well as exclusion of variants with high missingness (>3%), Hardy–Weinberg equilibrium P < 1×10^{-6}, or minor allele frequency <1%.

Cleaned genotype data were phased using Eagle2 and imputed to the TOPMed reference panel using Minimac4. Post-imputation filtering excluded variants with INFO < 0.8 and multi-allelic variants. Ancestry inference was performed using the GenoPred pipeline, merging cohort data with 1000 Genomes Phase 3 reference samples, followed by LD pruning and elastic-net classification into super-population groups. Analyses were restricted to individuals of European ancestry, identified based on principal component distributions.

The final quality-controlled imputed dataset comprised 6,396 individuals and 7,545,708 variants, provided in PLINK binary format (genome build: GRCh38). Full details of genotyping, quality control, and imputation procedures are described in Bridges et al. (2023) .

1970 cohort

Genotype data for the 1970 British Cohort Study were collected during the biomedical sweep (ages 46–48) using the Illumina Global Screening Array. Standard quality control procedures were applied prior to imputation, including exclusion of individuals based on missingness (>2%), sex discordance, excess heterozygosity, and relatedness (KING cutoff 0.0884), and exclusion of variants with high missingness (>3%), Hardy–Weinberg equilibrium P < 1×10^{-6}, or minor allele frequency <1%.

Following QC, genotypes were phased using Eagle2 and imputed to the TOPMed reference panel using Minimac4. Genome build conversion to GRCh38 was performed within the TOPMed pipeline. Post-imputation filtering excluded variants with INFO < 0.8 and multi-allelic variants. Ancestry inference was conducted using GenoPred in conjunction with 1000 Genomes Phase 3 reference data, and analyses were restricted to individuals of European ancestry.

The final quality-controlled imputed dataset comprised 5,598 individuals and 8,640,849 variants (genome build: GRCh38), provided in PLINK binary format.

Polygenic Score Method

The PGS for depressive symptom measurement, which is a proxy for psychological distress phenotype, was generated using summary statistics from the Baselmans et al. 2019 Genome-Wide Association Studies (GWAS) for depressive symptom measurement (1,067,913 individuals; total SNPs: 4,310,706). 1958c and 1970c were not included in the GWAS. We used PRSice2 to generate the polygenic score (28,29). Pruning was set to 250kb, as this distance is typically used to limit the inclusion of

SNPs in linkage disequilibrium with one another, thereby ensuring the polygenic score captures independent genetic signals. 15 thresholds were included from $5x10^{-8}$ to 1. At the p-value threshold of 1, the polygenic score contained 93,962 SNPs overlapping between 1958c and 1970c. Prior research used a threshold between 0.005 and 1 for a similar phenotype (30). See below the sensitivity analyses used to assess the impact of the different p-value thresholds on the variance explained by the polygenic score. The primary analysis exposure is the standardised overlapping SNP polygenic scores for psychological distress at a threshold of 0.005, which included 24,166 SNPs in both cohorts. The polygenic score was standardised across both cohorts (rather than separately) to have a mean of 0 and a standard deviation of 1; thus, a higher polygenic score indicates greater liability to psychological distress.